

**Studies on the biosynthesis of complex natural products
from myxobacteria**

Dissertation
zur Erlangung des Grades des
Doktors der Naturwissenschaften (Dr.rer.nat.)
der Naturwissenschaftlichen-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von

Kathrin Buntin

Saarbrücken

2010

Tag des Kolloquiums:	27.Mai. 2010
Dekan:	Prof. Dr.-Ing. Stefan Diebels
Berichterstatter:	Prof. Dr. Rolf Müller Prof. Dr. Rita Bernhard
Vorsitz:	Prof. Dr. Alexandra Kiemer
Akad. Mitarbeiter:	Dr. Carsten Volz

**Fallen ist keine Schande, liegen bleiben schon.
(Sprichwort)**

Danksagung

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Rolf Müller für die Überlassung der interessanten Themen und die Aufnahme in seine Arbeitsgruppe. Auch für seine Unterstützung, sein Engagement und seine jederzeit offene Tür möchte ich mich bedanken. Ein Dank geht auch an Prof. Dr. Rita Bernhardt für die Übernahme des Koreferates.

Des Weiteren möchte ich mich insbesondere bei Dr. Kira J. Weissman bedanken für ihre großartige Unterstützung und die äußerst produktive Zusammenarbeit. Die vielen wissenschaftlichen Diskussionen mit ihr haben meine Arbeit in vielen Bereichen vorangebracht und mich stets motiviert auch schwierigen Fragestellungen nachzugehen. Auch für das Korrekturlesen sämtlicher Manuskripte und meiner Doktorarbeit möchte ich mich bei ihr bedanken.

Ein Dank geht auch an Dr. Herbert Irschik vom HZI in Braunschweig, der mich in die Geheimnisse der Konjugation der Sorangien eingeweiht und mich auch aus der Ferne mit hilfreichen Tipps bezüglich Kultivierung und Handhabung dieser Myxobakterien unterstützt hat. Auch bedanken möchte ich mich bei Shwan Rachid für die anfängliche Betreuung und die Hilfe bei der Cosmidbank.

Ich möchte mich bei Eva Luxenburger bedanken für ihren analytischen Beistand auch nach der 1000. Messung, Dr. Daniel Krug für die ganz schwierigen Fälle, Prof. Dr. Helge Bode für die Hilfe bei der NMR-Strukturaufklärung und Ole Revermann und Thorsten Klefisch für die hilfreichen Tipps bei diversen Aufreinigungen.

Ein Dankeschön geht auch an meine Diplomandin Katja Gemperlein für die nette Zusammenarbeit und Dr. Yanyan Li für die Hilfe in proteinchemischen Fragestellungen.

Bedanken möchte ich mich natürlich auch bei der gesamten Myxoarbeitsgruppe und den ehemaligen Mitgliedern für die gute Zusammenarbeit und die freundliche Atmosphäre, die das Arbeiten sehr angenehm gemacht haben. Dabei geht ein besonderer Dank an meine Kollegen im Büro, für die vielen lustigen Momente und die angeregten Diskussionen.

Zuletzt geht auch ein ganz großes Dankeschön an meine Familie, die immer an mich geglaubt hat, hinter mir stand und mich unterstützt hat. Und natürlich gilt auch ein besonderer Dank Alexander Brachmann für die vielen wunderbaren Augenblicke, die das Leben schön machen.

Vorveröffentlichungen der Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung des Fachbereichs, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

Publikationen:

Bock, M., **Buntin, K.**, Müller, R., Kirschning, A. (2008) Stereochemical determination of thuggacins A–C, highly active antibiotics from the myxobacterium *Sorangium cellulosum*. *Angew. Chem. Int. Ed. Engl.*, **47**, 2308–2311.

Buntin, K., Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., Müller, R. (2008) Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.*, **47**, 4595–4599.

Buntin, K., Irschik, H., Weissman, K.J., Luxenburger, E., Blöcker, H., Müller, R. (2010) Biosynthesis of the anti-tuberculosis macrolide thuggacins in myxobacteria: comparative cluster analysis reveals the basis for natural product structural diversity. *Chem. Biol.*, **17** (4): 342-56

Buntin, K., Weissman, K.J., Müller, R. (2010) An unusual thioesterase “chaperones” isochromanone ring formation in ajudazol biosynthesis. *ChemBioChem*, **11** (8):1137-46

Tagungsbeiträge:

Buntin, K., Rachid, S., Kochems, I., Müller, R. October 2006. Identification of the gene cluster involved in the biosynthesis of the antimicrobial secondary metabolite ajudazol in *Chondromyces crocatus*. International workshop: *Biology of Bacteria Producing Natural Products*, Tübingen, Germany (poster presentation).

Buntin, K., Rachid, S., Weissman, K.J., Müller, R. October 2007. New insights into the biosynthesis of the antifungal secondary metabolite ajudazol in *Chondromyces crocatus* Cm c5. International workshop: *Biology of Bacteria Producing Natural Products*, Saarbrücken, Germany (poster presentation).

Buntin, K., Rachid, S., Weissman, K.J., Müller, R. September 2008. Production of the highly antifungal ajudazols in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. International workshop: *Biology of Bacteria Producing Natural Products*, Berlin, Germany (oral presentation).

Zusammenfassung

Viele myxobakterielle Naturstoffe haben strukturelle Besonderheiten und/oder weisen eine neuartige biologische Aktivität auf. Im Rahmen dieser Arbeit wurde unter anderem die Biosynthese von Ajudazol und Thuggacin in *Chondromyces crocatus* Cm c5 und von Thuggacin in *Sorangium cellulosum* So ce 895 aufgeklärt. Für die Biosynthese von Ajudazol konnte gezeigt werden, dass die Thioesterase (TE) Domäne am Ende des Multienzymkomplexes die Bildung des Isochromanonrings unterstützt und auf Grund der katalysierten Reaktion eine neue Klasse von TE Domänen bildet. Des Weiteren wurde die Rolle zweier P450 Enzyme, die an der Modifikation des Ajudazolgrundgerüsts beteiligt sind nachgewiesen.

C. crocatus und *S. cellulosum* produzieren strukturell unterschiedliche Varianten des antimycotischen Sekundärstoffes Thuggacin. Durch die Identifizierung, Charakterisierung und dem anschließendem Vergleich der beiden Biosynthesewege konnten die entsprechenden enzymatischen Mechanismen, die zu den strukturellen Unterschieden führen, aufgeklärt werden. So konnte gezeigt werden, dass im *S. cellulosum* Biosyntheseweg ein zu Crotonyl-CoA-Reduktasen/Carboxylasen homologes Enzym für den Einbau der Hexyl-Seitenkette verantwortlich ist. Die unterschiedlichen Hydroxylierungen beruhen vermutlich auf der variablen Aktivität einer FMN abhängigen Monooxygenase, deren Funktion in *C. crocatus* nachgewiesen wurde.

In weiteren Experimenten wurden die Entstehung der Pyrrolstartereinheit in der Leupyrrinbiosynthese in *Sorangium cellulosum* So ce690 und die Bildung des Disorazol Dilactons in *Sorangium cellulosum* So ce12 biochemisch untersucht.

Abstract

Many myxobacterial natural products display unusual structural features and/or a novel bioactivity, making these secondary metabolites attractive targets for study. This thesis describes the elucidation of the biosynthetic pathways to the ajudazols and thuggacins in *Chondromyces crocatus* Cm c5 and the thuggacins in *Sorangium cellulosum* So ce895. Using experiments *in vitro* and *in vivo*, we have shown that the thioesterase (TE) domain of the ajudazol assembly line “chaperones” the formation of the characteristic isochromanone ring, thus placing it within a novel class of TE enzymes. In addition, we have demonstrated the involvement of two P450 enzymes in post-assembly line modification of ajudazol.

Distinct variants of the anti-tuberculosis macrolide thuggacins are produced by *C. crocatus* and *S. cellulosum*. The basis for this architectural diversity has been elucidated by identifying the biosynthetic pathways in both myxobacteria, and comparing them in detail. In the course of this analysis, a crotonyl-CoA reductase/carboxylase homologue was discovered in the *S. cellulosum* thuggacin gene cluster, and was shown to participate in the assembly of an unusual hexyl side chain. Moreover, evidence was provided that the distinct pattern of hydroxylation observed in the *C. crocatus* and *S. cellulosum* thuggacins may result from variable action of a conserved FMN-dependent monooxygenase.

Finally, we report experiments to probe the formation of the pyrrole starter unit in leupyrrin biosynthesis in *Sorangium cellulosum* So ce690, and the biochemical investigation of the mechanism by which the disorazol dilactone structure is generated in *Sorangium cellulosum* So ce12.

TABLE OF CONTENTS

Zusammenfassung	IV
Abstract	V
Table of contents	VI
Introduction	
1. Natural products in drug discovery	1
2. Myxobacteria: a promising source for novel natural products	2
2.1 Myxobacterial natural products	4
3. Polyketide and nonribosomal peptide biochemistry	6
3.1 PKS biochemistry	7
3.2 NRPS biochemistry	10
3.3 PKS/NRPS hybrids	13
3.4 Type I and type II TEs: chain release and proof-reading	14
3.5 Post-assembly line modifications	16
4. Outline of the work	18
Chapter 1	
Production of the antifungal isochromanone ajudazols A and B in <i>Chondromyces crocatus</i> Cm c5: biosynthetic machinery and cytochrome P450 modifications	22
Chapter 2	
An unusual thioesterase “chaperones” isochromanone ring formation in ajudazol biosynthesis	23
Chapter 3	
Stereochemical determination of thuggacins A-C, highly active antibiotics from the myxobacterium <i>Sorangium cellulosum</i>	24
Chapter 4	
Biosynthesis of the anti-tuberculosis macrolide thuggacins in myxobacteria: comparative cluster analysis reveals the basis for natural product structural diversity	25
Chapter 5	
Studies on the disorazol TE	26
Chapter 6	
Generation of the starter unit in leupyrrin biosynthesis	36

Chapter 7

The origin of the uncommon extender unit in Soce thuggacin biosynthesis 43

Discussion

1. General summary of this work 51

2. Biosynthesis of ajudazol in *C. crocatus* Cm c5 52

3. Biosynthesis of thuggacins in *C. crocatus* Cm c5 and *S. cellulosum* So ce 895 59

References 65

Author's effort in publications from Chapter 1-4 76

Introduction

1. Natural products in drug discovery

Microorganisms produce an enormous number of secondary metabolites which exhibit a wide variety of biological activities useful to man, acting as anti-bacterial, anti-fungal, anti-parasitic and anti-cancer agents. Furthermore, these molecules are also used in veterinary medicine as antibiotics and antiparasitic drugs, and in agriculture as plant growth regulators, herbicides and insecticides ^[1]. In addition to their diverse bioactivities, these so-called ‘natural products’ are characterized by a depth of structural diversity which is not available from alternative sources. Both of these features explain the continuing dominance of natural products as lead structures for the development of novel drugs across many therapeutic areas.

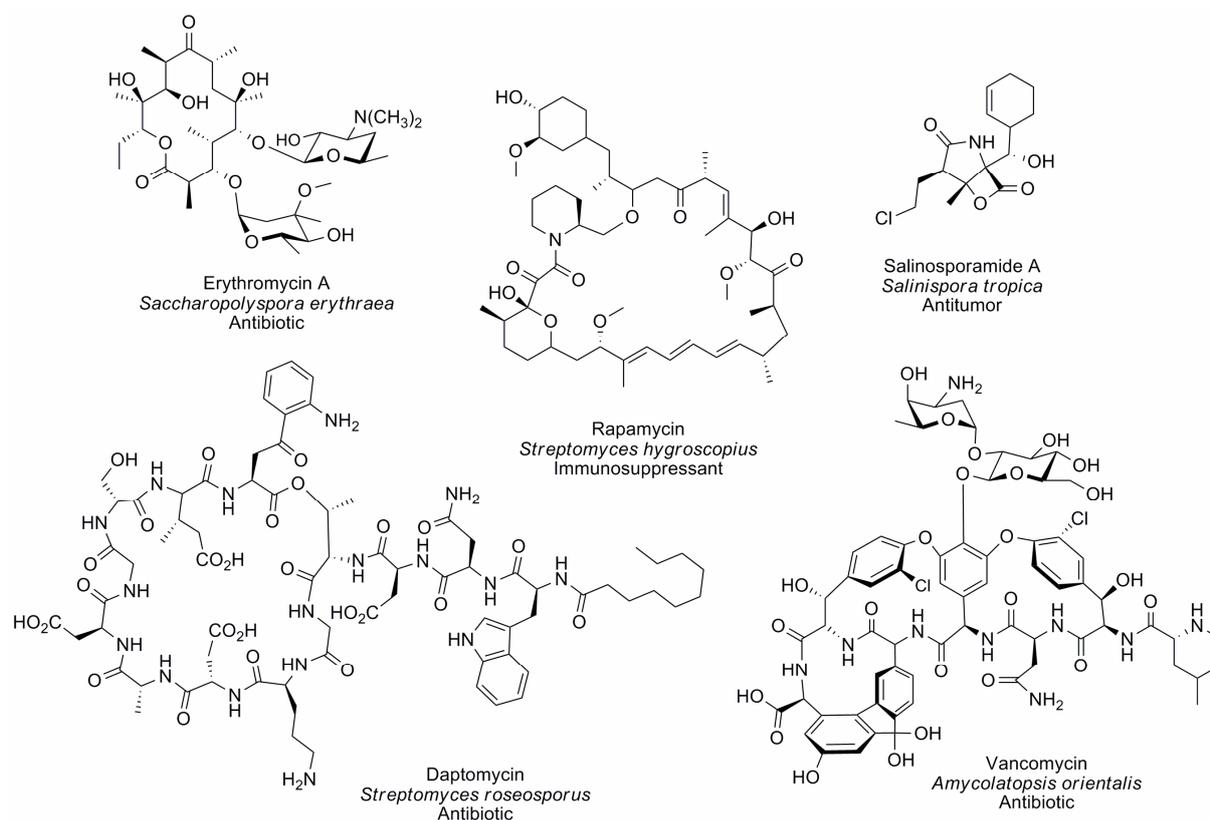


Figure 1. Examples of natural products derived from bacteria which are used in the clinic or are in clinical trials. The name, natural producing organism and pharmaceutical application are indicated below each structure.

In fact, approximately 50% of the pharmaceuticals in clinical use are natural products or their derivatives (Examples are presented in **Figure 1**) ^[2;3]. Despite this significant success, the need for novel natural products for drug development continues. Drivers for drug discovery

include the emergence of multi-drug resistance among the latest generation of pathogens, the lack of highly effective therapeutics for many diseases, including tuberculosis, and side effects associated with some current medicines ^[1;4]. Although the big pharmaceutical companies have largely abandoned their programs in natural product research ^[5] new natural products with unique features are discovered every year by small biotechnology firms and academic researchers.

Among the natural producers of secondary metabolites, besides plants and fungi, microorganisms continue to be a promising source for novel metabolites, as only a minor proportion of prokaryotic diversity has been examined to date ^[1]. Efforts have been made to identify new groups of microbial producers by analysing bacteria that are difficult to cultivate (e.g. cyanobacteria) ^[6] or are pathogenic (e.g. the genera *Nocardia*) ^[7]. In addition, underexplored classes of bacteria, such as those from marine environments and the myxobacteria ^[4;8], have come into focus as promising producers of compounds with both unique structures and bioactivities.

2. Myxobacteria – a promising source for novel natural products

Myxobacteria are obligate, aerobic Gram-negative mesophilic δ -proteobacteria which are commonly isolated from soil, the bark of trees, decaying plant materials and the dung of herbivores ^[9]. However, the discovery of novel myxobacterial species from moderately halophilic soil as well as marine environments, illustrates that they can adapt to a wider variety of environmental conditions ^[10;11]. All known myxobacteria are united in the order *Myxococcales*, which can be further divided into the three suborders *Cystobacterineae*, *Sorangineae* and *Nannocystineae* ^[12]. More than 7500 strains within the order *Myxococcales* have been already isolated by research groups at the Helmholtz Center for Infection Research (Braunschweig, Germany; formerly German Research Center of Biotechnology (GBF)) and novel strains, species and even families are continually being discovered ^[9;13].

The vegetative cells of myxobacteria are typically rod-shaped and often rather large (4–12 μm long and 0.7–1.2 μm wide). When cultivated in liquid media, for most myxobacterial strains dispersed growth can be obtained after several passages in specific media except for strains from the genera *Chondromyces*, *Polyangium* (suborder *Sorangineae*) and *Nannocystis* (suborder *Nannocystineae*). The cells of these genera exhibit a strong tendency to stick together, and thus grow preferentially in clumps rather than as independent cells (**Figure 2A**) ^[14]. On solid surfaces, myxobacteria can glide in coordinated swarms, while under starvation

conditions they aggregate to form characteristic multi-cellular structures called fruiting bodies (Figure 2B+C) [15].

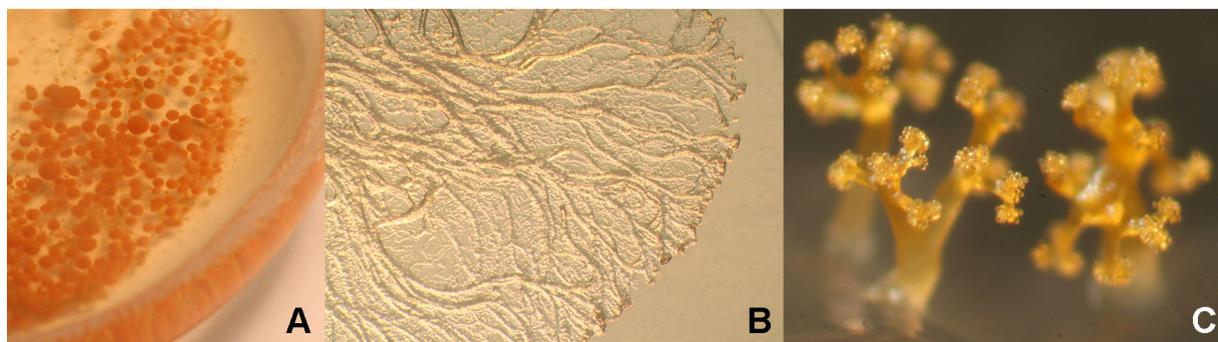


Figure 2. (A): Cell clumps of *Chondromyces crocatus* Cm c5 in liquid culture. (B): Swarming behavior of *Sorangium cellulosum* SB So018 on an agar plate. (C): Fruiting bodies of *Chondromyces crocatus* SB Cm010.

The shape and morphology of fruiting bodies can vary highly between different genera, with the genus *Chondromyces* producing the most sophisticated designs [14]. At the end of the myxobacterial life cycle, myxospores are formed within the mature fruiting body. These structures can withstand environmental extremes, allowing the bacteria to persist for long periods in a dormant state [10]. When suitable nutritional and environmental conditions are restored, the spores germinate and retransform into viable cells.

Myxobacteria are also often described as ‘micropredators’, as they are able to feed on other living organisms such as other bacteria and yeast [10]. For this, they excrete a variety of lytic enzymes that catalyze lysis of the “victim” cells and also subsequently digest the released proteins, lipids and nucleic acids. *Sorangium cellulosum* is particularly notable in this context, as it not only feeds on other cells, but it is the sole myxobacterial species that can degrade cellulose [14;16]. The predatory and social behaviors of myxobacteria have interested microbiologists for decades, but more recently these bacteria have come into focus as multi-producers of natural products. At least 100 core structures and more than 500 derivatives have been identified to date [15], and every year new compounds are discovered [9]. Many of these characterized natural products display novel structural features, and modes-of-action which tend to differ from those of other secondary metabolites [20-22]. Both attributes make these compounds attractive candidates for new lead structures in drug development.

Prior to genome sequencing, the myxobacterial strains *Myxococcus xanthus* DK1622 and *Sorangium cellulosum* So ce56 were not known as prominent producers of secondary metabolites. However, analysis of their genomes, among the largest yet discovered in bacteria, showed that they harbor 18 and 17 secondary metabolite loci, respectively [17;18]. As

large genomes appears to be a characteristic of many myxobacteria^[17], and there is a strong correlation between genome size and the number of secondary metabolite gene clusters in a microorganism^[19], these findings reinforce the idea that myxobacteria are a promising future source of novel natural products.

2.1 Myxobacterial natural products

The first myxobacterial natural product to be applied for clinical use was a semi-synthetic epothilone B derivative named Ixabepilone® (Ixempra), which was approved in the United States for chemotherapy against breast cancer in 2007 (**Figure 3**). Epothilone interacts with the cytoskeleton of eukaryotic cells by binding to β -tubulin inducing microtubule polymerization^[23]. The resulting suppression of microtubule dynamics leads to the arrest of the cell cycle at the G2/M transition, followed by cell death via apoptosis^[24]. As cancer cells are highly dependent on microtubule function to support their rapid cell division, the tubulin system represents an attractive target for specifically attacking tumor cells^[25]. Additional myxobacterial compounds that interact with the cytoskeleton include tubulysin and disorazol (**Figures 3 and 11**), which destabilize their tubulin target^[26;27] and rhizopodin (**Figure 3**), which binds to actin and prevents its polymerization^[28]. All three compounds are potential candidates for use in cancer treatment, with several tubulysin derivatives already progressed into pre-clinical trials (http://www.innovations-report.de/html/berichte/bildung_wissenschaft/bericht-17016.html). In the case of disorazol a strategy was developed to specifically deliver the agent to cancer cells. A variety of cancers, including ovarian, breast and prostate, express receptors for luteinizing hormone releasing hormone (LHRH). Thus, in theory, attaching a cytotoxic agent to a LHRH peptide agonist provides a means to achieve targeted chemotherapy, minimizing potential side effects of these compounds^[28b]. Proof-of-principle has been demonstrated recently with disorazol, which was shown to induce apoptosis of cancer cells following internalization^[28b].

Another interesting family of myxobacterial natural products are soraphens from *Sorangium cellulosum* (**Figure 3**). Due to its strong anti-fungal effects, soraphen A was first investigated as a plant protective agent^[29]. However, development of the metabolite was abandoned when the compound's teratogenic potential was discovered. Nonetheless, these early studies revealed important insights into soraphen's mode-of-action. Soraphen A interacts directly with eukaryotic acetyl-CoA carboxylases (ACC) by disrupting the oligomerization of these enzymes and thereby inhibiting their activity^[30;31]. Notably, acetyl-CoA carboxylases are of

critical importance for fatty acid synthesis, as they furnish the central building block malonyl-CoA. Determining the mode-of-action of soraphen brought the acetyl-CoA carboxylases into focus as potential targets in the treatment of both cancer^[32] and obesity^[33] treatment.

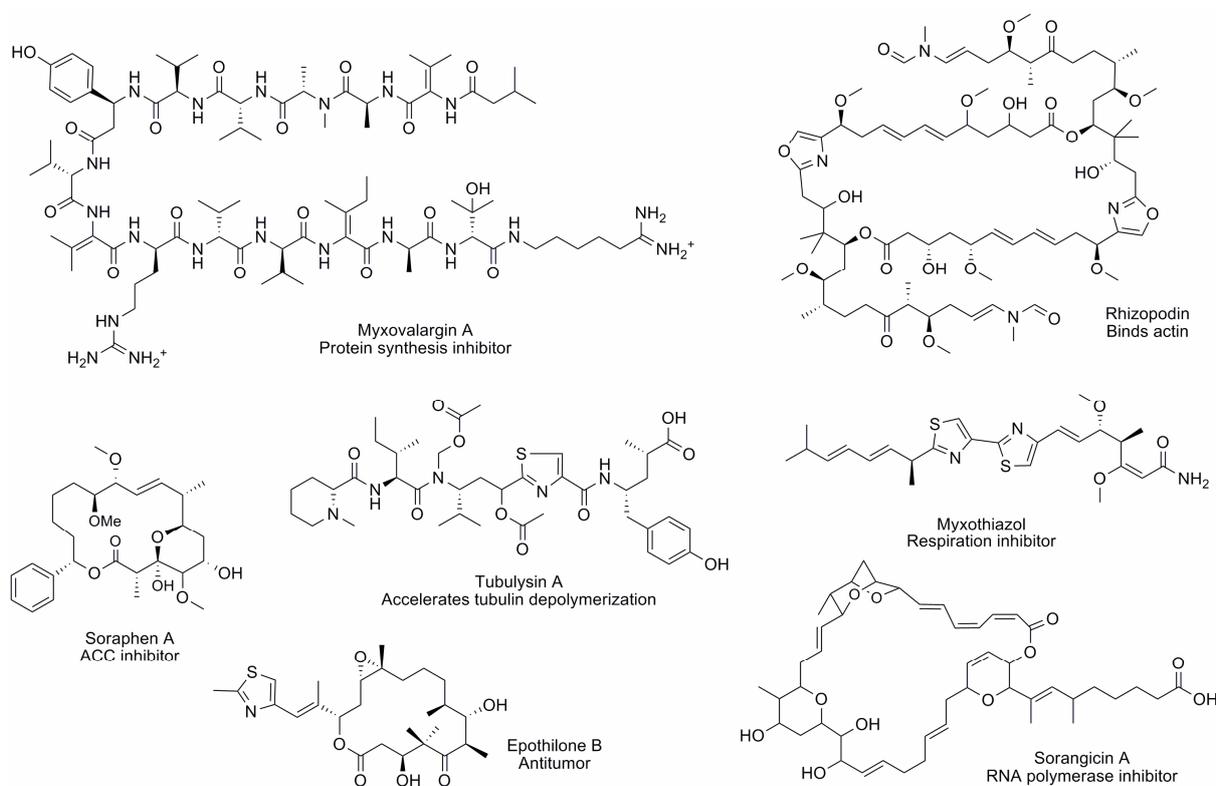


Figure 3. A selection of myxobacterial secondary metabolites, illustrating both their structural diversity and their varied biological activities. The compound name and mode-of-action are given below each structure.

The two myxobacterial strains *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5, both produce structural variants of the novel antibacterial macrolide thuggacin (**Figure 10**). Thuggacins have been shown to effectively inhibit the bacterial respiratory chain in several Gram-positive bacteria, including clinical isolates of *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB)^[34;35]. This bioactivity is particularly promising as it represents an alternative mode of action compared to the current first- and second-line antibiotics in TB treatment which mainly interact with RNA, DNA and cell wall synthesis. The bacterial respiratory chain is a particularly promising target for new chemotherapeutics, as it appears to be essential for both replicating and non-replicating mycobacteria^[36]. Other targets of myxobacterial secondary metabolites include the mitochondrial electron transport chain (e. g. myxothiazol, melithiazol)^[37;38], eubacterial RNA polymerases (e. g.

myxopyronin/corralopyronin, sorangicin)^[39-41] and protein synthesis (e. g. myxovalargin)^[42] (**Figure 3**).

Despite the diversity of structures, the majority of known myxobacterial secondary metabolites are linear or cyclic polyketides (PKs) and non-ribosomal polypeptides (NRPs), which are constructed by a common biosynthetic logic. In fact, more than 50% of identified compounds contain both PK and NRP elements, and therefore are termed hybrid PK-NRP metabolites^[20]. In contrast, such mixed natural products are quite rare in other bacterial secondary metabolite producers such as the actinomycetes, which synthesize preferentially pure PK or NRP compounds^[20].

3. Polyketide and nonribosomal peptide biochemistry

PK and NRP natural products are assembled on multienzymes called polyketide synthases (PKS), and nonribosomal peptide synthetases (NRPS), respectively^[43;44]. PKSs, NRPSs and their hybrids, are large multifunctional enzyme complexes that utilize monomeric building blocks such as short acyl-CoA esters and amino acids to build molecules of high complexity^[45;46]. The mode of operation of these multienzymes is often described as an assembly line, as they display a modular organization in which each module is responsible for the incorporation of one building block into the growing product chain^[47;48]. In addition, each module can be further subdivided into domains, which represent the enzymatic units that are responsible for the individual steps of loading, condensation and subsequent (though optional) modification of a specific extender unit.

The genes encoding the biosynthetic machineries are typically clustered together in the microbial genomes, and are often co-localized with transcriptional regulators and genes for self-resistance^[49]. In some cases, there is a one-to-one correspondence between the domains present in the PKS/NRPS and the set of biosynthetic transformations which occurs, an observation referred to as ‘colinearity’^[50]. Based on this colinear principle, modules and domains that are involved in the biosynthesis of a particular natural product can be predicted by analyzing the structure of the compound. This ‘retrobiosynthetic analysis’ approach can significantly enable the search for the corresponding gene cluster. Conversely, if the cluster is already identified but its product is unknown, the organization of the assembly line can be used to predict some aspects of the metabolite’s basic backbone structure.

The operation of PKS and NRPS systems is also described as a ‘multiple carrier thio-template mechanism’^[51;52]. The central feature of this model is that during the assembly process, all

substrates, intermediates and products remain covalently tethered via a thioester-linkage to a carrier protein (CP) domain within a chain-extension module.

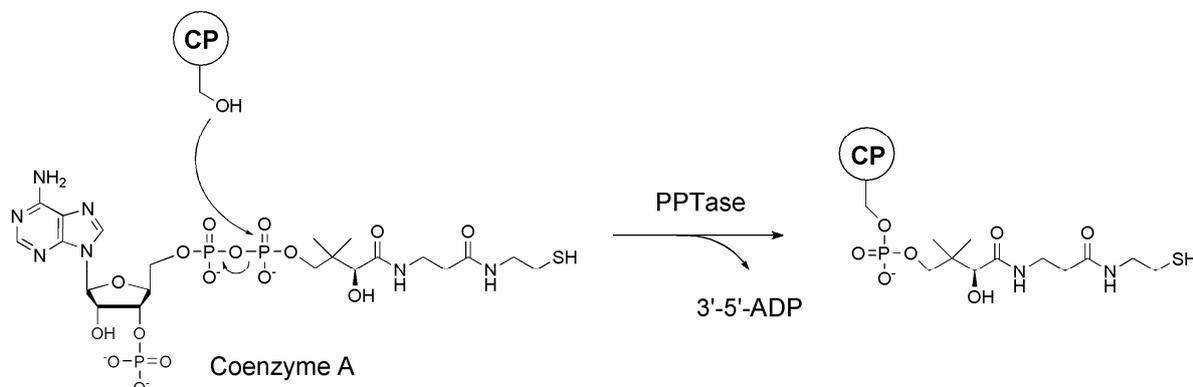


Figure 4. Post-translational modification of a CP domain by a PPTase. The PPTase catalyzes the transfer of phosphopantetheine from coenzyme A to a conserved serine in the CP domain, converting the inactive *apo* CP to the active *holo* form.

Therefore, in order to produce a functional enzymatic assembly line, each CP has to be post-translationally modified by a phosphopantetheinyltransferase (PPTase) enzyme. PPTases are specialized, discrete enzymes that act in *trans* to transfer the 4'-phosphopantetheine (Pant) moiety of CoA to conserved serine residues of the CPs, converting the domains from their inactive *apo* to their active *holo* forms (**Figure 4**)^[46,53]. During the biosynthesis, the intermediates are bound as thioesters to the terminal thiol group of this phosphopantetheine arm, which activates the substrates for Claisen or amide bond condensation. In addition, the covalent linkage between substrate and the phosphopantetheine arm allows the intermediates to be shepherded between the individual active sites, without diffusing into the cellular medium. This sequestration provides an effective means to stabilize the reactive intermediates, as well as a kinetic advantage to the overall biosynthesis.

3.1 PKS biochemistry

The biosynthesis of a polyketide backbone requires three essential core domains, an acyltransferase (AT), a ketosynthase (KS) and an acyl carrier protein (ACP). The AT domain selects the specific extender unit, the KS domain catalyzes the condensation between the extender unit and the growing chain, and the ACP is responsible for delivering the intermediate to the various catalytic domains. A typical PKS module contains all three domains, with the exception of initiation modules, which can be organized in two different ways. One type of starter module contains all three core domains. However, the KS domain of this initiation module is catalytically inactive for condensation as its active site cysteine has

been replaced by a glutamine (this type of KS is therefore referred to as a KS^Q domain). As the starter module typically provides a monocarboxylic acid for the first elongation step, the KS^Q domain catalyzes decarboxylation of the (methyl)malonyl-S-ACP to generate the required building block (e.g. propionate/acetate). The second type of loading module is a didomain, consisting of an AT and an ACP. In this case, the AT domain directly selects short chain monocarboxylic acids such as acetyl- and propionyl-CoA as starter units. In addition, certain initiation modules recruit less common building blocks, such as aromatic and branched-chain carboxylic acids [54-57]. In either case, however, the starter unit is ultimately delivered to the KS domain of the first chain extension module.

In chain extension modules, AT domains typically exhibit a preference for dicarboxylic acids such as malonyl-CoA and methylmalonyl-CoA. In some cases, however, uncommon extender units are recognized, such as hydroxymalonyl-ACP, methoxymalonyl-ACP and aminomalonyl-ACP [58-61]. The corresponding substrates are selected by the AT domain while they are attached to the ACP, and are then transferred to a carrier protein within the multienzyme. Sequence analysis of numerous AT domains has revealed conserved amino acid motifs which can be correlated to substrate specificity [62-64]. The prediction of substrate specificity based on these conserved residues is relatively reliable, but in some cases, the identified code cannot be related with confidence to a certain extender unit, which can be regularly observed in myxobacterial systems [56;57;65].

After selection of a specific extender unit, the AT domain transfers it to the Ppant moiety of the ACP domain within its own module. The ACP-bound extender unit is then decarboxylated and the resulting enolate attacks the upstream KS-bound acyl thioester. This KS-catalyzed condensation results in a β -keto-acyl-ACP intermediate which is extended by one C₂ unit, and which is covalently bound to the ACP domain of the extension module (**Figure 5A**). At this stage, the intermediate can be transferred to the KS domain of the following module and participate in another chain elongation step (**Figure 5A**).

The structural diversity of PKs can be enhanced by the action of modifying domains, which are optionally present in certain modules and act directly after the chain elongation step [50;66]. Depending on whether ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains are incorporated in a particular module, various extents of redox adjustment of the β -keto functionality can result. Initially the KR domain catalyzes the stereospecific and NADPH-dependent reduction of the β -keto function to the β -hydroxy intermediate. Subsequently the DH carries out dehydration, leading to an α,β -enoyl intermediate, which can ultimately be reduced to the fully saturated acyl chain by the ER domain, with further

consumption of NADPH (**Figure 5B**). A full ‘reductive loop’, i.e. all three enzymatic domains acting in tandem, is not necessarily present in every PKS module^[50]. Consequently the partial or full omission of the sequential ketoreduction, dehydration and enoyl reduction reactions during polyketide assembly in a

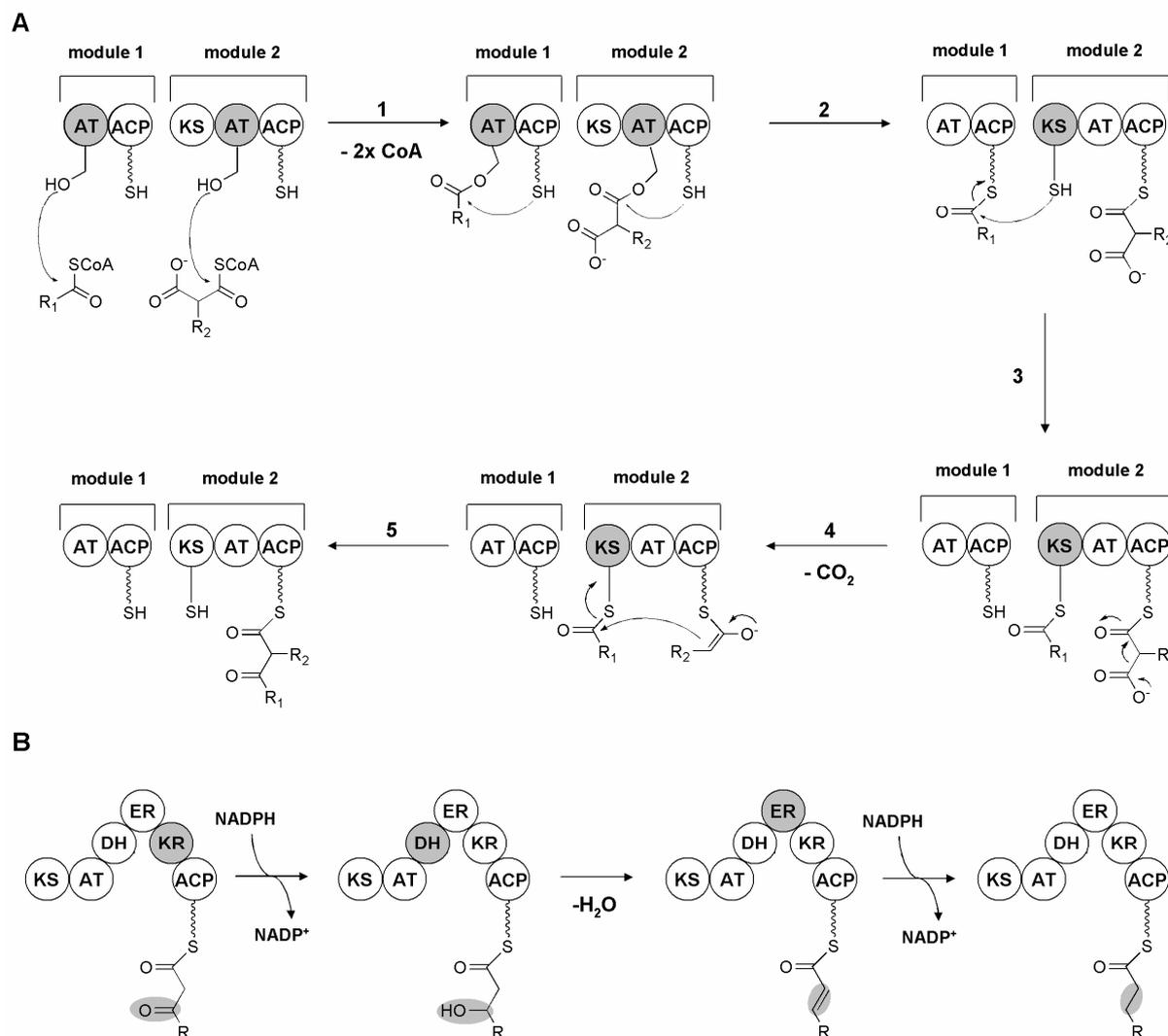


Figure 5. Schematic overview of PKS biochemistry. (A): AT domains from module 1 (initiation module) and module 2 (first elongation module) select their respective acyl-CoA monomers (1) and catalyze the transthiolation of these substrates to the downstream ACP domains (2). In the next step, the intermediate from module 1 is transferred to the KS of module 2 (3). Subsequently decarboxylative condensation catalyzed by the KS domain occurs to form the C–C bond between the upstream acyl thioester and the downstream enolate (4+5) $R_1 = \text{CH}_3$ or a less common starter unit (see text) $R_2 = \text{H}, \text{CH}_3$. **(B):** Optional β -carbon processing catalyzed by the KR, DH and ER domains. R = remaining polyketide chain. Domains that are involved in a specific catalytic step are colored in grey.

highly programmed manner is one of the biosynthetic features underlying the structural diversity of PK products.

Recent protein sequence analysis of KR domains led to the identification of several conserved motifs that allow the stereochemistry (‘A’- or ‘B-type’) of the resulting hydroxyl moiety to be

predicted [67;68]. If the KR is followed by a DH domain a correlation between B-type ketoreduction and the formation of a *trans* double bond by *syn* dehydration and for A-type ketoreduction and the formation of a *cis* double bond by the same mechanism could be observed [67;68]. Of course, as for all *in silico* predictions, this correlation is not absolute [69]. Furthermore, the final stereochemistry of saturated, methyl branched sites can be predicted from a distinct specificity motif within the ER domains [70]. With these recent discoveries, it is now possible to predict the absolute configuration of a product from the gene sequences.

The termination module is the last module of the assembly line and is responsible for the release of the full-length acyl chain. This process of chain termination is typically catalyzed by a thioesterase (TE) domain, whose function is described in more detail in **section 3.4**.

As an aside, it's worth mentioning that in addition to type I modular PKSs, two other types of PKS systems are known in nature. Type II systems consist of a set of discrete and usually monofunctional enzymes that form a multienzyme complex, and are used iteratively to produce polycyclic, aromatic metabolites [66]. In type III systems, the catalytic activity of a single active site accomplishes the complete biosynthesis, which includes decarboxylation, condensation, cyclization and aromatization reactions [71;72]. Another fundamental difference between type III systems and the others, is that they use CoA-esters instead of ACP bound intermediates as substrates.

3.2 NRPS biochemistry

By analogy to PKS biosynthetic logic, a minimal NRPS module also consists of three core domains: adenylation (A), condensation (C) and peptidyl carrier protein (PCP). NRP biosynthesis is initiated by an A domain, which selects a specific amino acid (**Figure 6A**). The specificity of A domains is not restricted to the 20 proteinogenic amino acids, as they can also recognize and activate a much wider variety of nonproteinogenic amino and aryl acids as monomer building blocks [46]. This biosynthetic feature contributes to the high structural diversity within this class of natural compounds.

The crystal structure of PheA from gramicidin S synthetase [73] coupled with analysis of primary sequences from various A domains, led to the identification of 8–10 amino acids residues within the substrate binding pocket that constitute the major determinants of A domain substrate specificity [73-75]. This ‘nonribosomal code’ is a useful tool to predict the substrate specificity of an A domain, even if the natural product is unknown. However, it is

not possible to make a confident specificity prediction for all A domains, as in some cases the extracted code cannot be assigned to a certain amino acid.

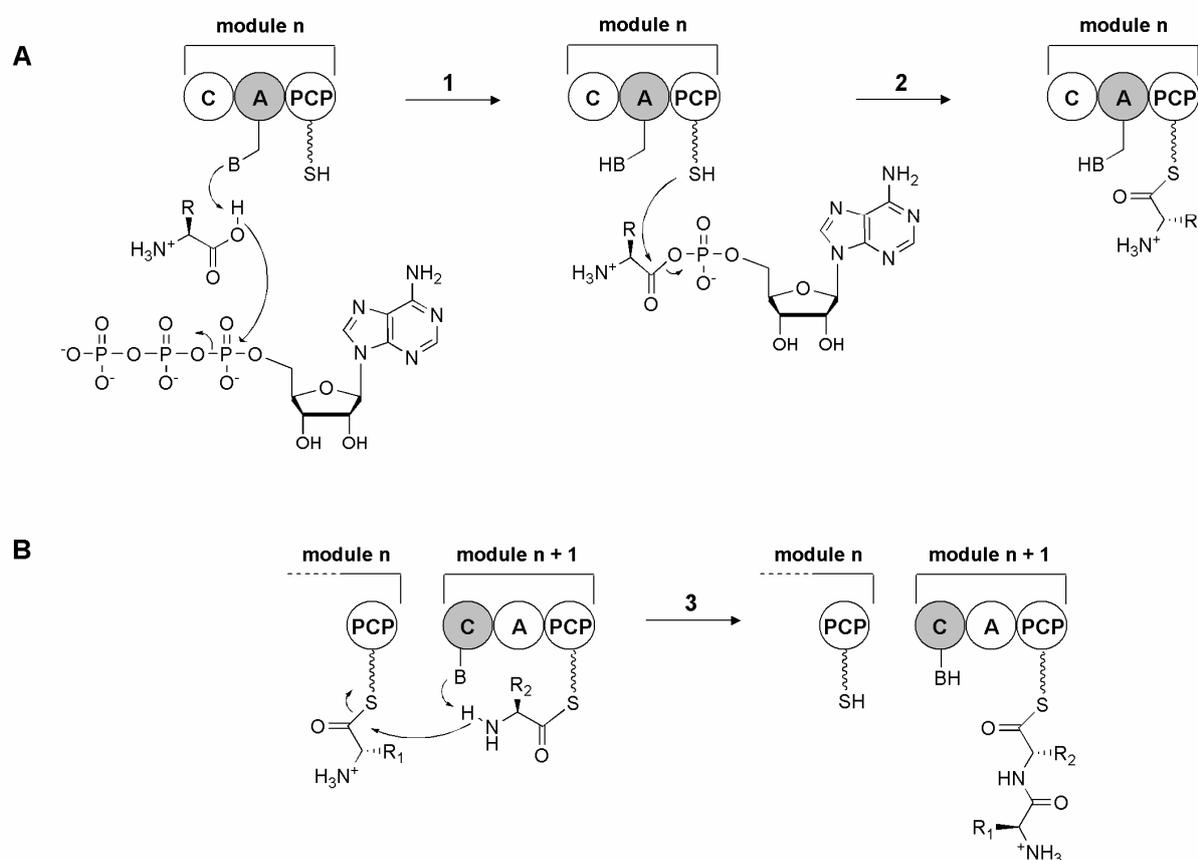


Figure 6. Schematic overview of NRPS biochemistry. (A): Selection and activation of a specific amino acid by the A domain. ATP is consumed in the activation reaction (1), and the aminoacyl group of the obtained aminoacyl-AMP is transferred to the PCP domain (2). (B): The C domain catalyzes a condensation reaction (3) which results in formation of a peptide bond between two adjacent aminoacyl-S-PCP intermediates. Catalytic domains involved in a particular reaction are indicated in grey.

In addition to its function in selecting a specific residue, each A domain also activates the amino acid to its aminoacyl-adenylate by consuming ATP (**Figure 6A**). In the next step, the amino acid is transferred to the peptidyl carrier protein (PCP) domain. The subsequent condensation reaction between amino acyl substrates tethered to PCPs of adjacent modules is catalyzed by the condensation (C) domain. The amino group of the downstream aminoacyl-S-PCP performs a nucleophilic attack on the acyl group of the upstream peptidyl-S-PCP, leading to the generation of a new peptide bond (**Figure 6B**). As a result, the peptide intermediate is elongated by one amino acid and remains covalently linked to the PCP of the downstream module. Additional elongation steps can follow until the end of the assembly line is reached. As with PKS machinery, NRPS modules can also contain optional domains that increase the structural variety of the final products. A notable example is a variant of the C domain, the

heterocyclization (HC) domain. In addition to peptide bond formation, this domain catalyzes the heterocyclization of functional side chains of the amino acids cysteine, serine and threonine with the peptide backbone. The obtained five-membered hydrolytically labile thiazoline or oxazoline rings can be further oxidized to stable thiazol or oxazole heterocycles by oxidation (Ox) domains using FMN as cofactor.

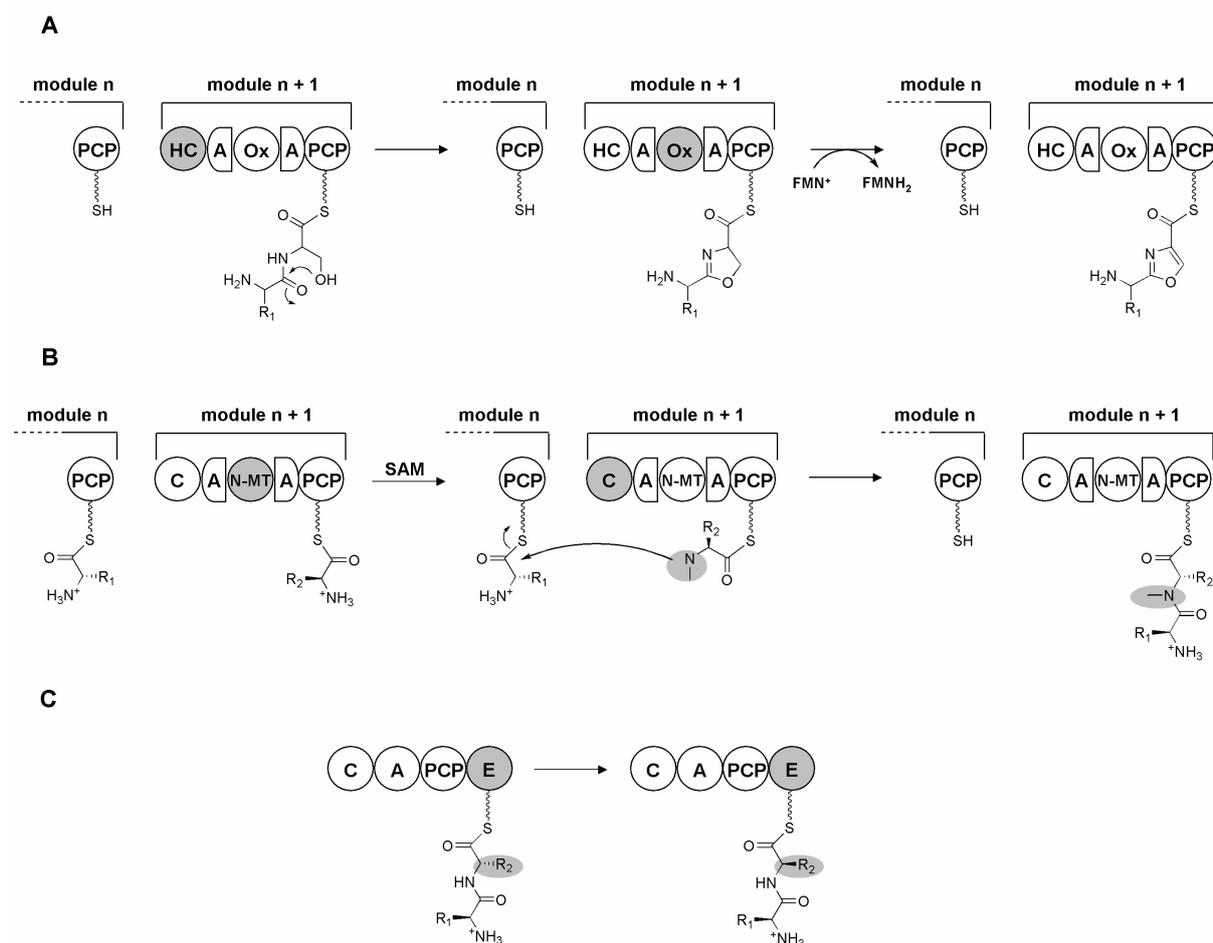


Figure 7. Reactions of various modifying catalytic domains. (A): Formation of an oxazole ring catalyzed by HC and Ox domains. After catalyzing the condensation between two aminoacyl intermediates, the HC domain facilitates the attack of an internal nucleophilic (in this case, a serine) on the adjacent carbonyl, followed by dehydration to yield the oxazoline. Subsequently an Ox domain can act on the oxazoline using FMN as hydrogen acceptor, converting it to the fully-reduced aromatic oxazole. (B): *N*-methylation catalysed by an *N*-MT domain. The methylation of the aminoacyl-*S*-PCP amine occurs prior to the condensation, and requires *S*-adenosylmethionine (SAM) as co-factor. (C): *E*-domain catalyzed epimerization occurs after amide bond formation. Domains that participate in a specific reaction are shaded in grey.

Ox domains have been reported frequently in modules that already contain a HC domain, and can be integrated into the corresponding A domain (**Figure 7A**) or located downstream of the module's integral PCP^[53]. Many NRPS also contain methyltransferase (MT) domains which are responsible for *N*- or *C*- methylation of amino acids residues, modifications which prevent premature proteolytic breakdown of the peptides^[53]. MT domains transfer the methyl group from *S*-adenosylmethionine (SAM) to the amino group or carbon of the respective aminoacyl-

S-PCP intermediate, before the condensation with the upstream peptidyl-*S*-PCP occurs (**Figure 7B**). Furthermore, some NP products contain D-amino acids, which is notable as D-amino acids only occur rarely in the microbial producers. Three different mechanisms have been discovered for how NRPSs can generate D-amino acids and integrate them into their product structures. The simplest way is to use a D-amino acid-selective A domain, as for example in cyclosporine biosynthesis^[76]. Here, the required D-amino acid is generated by an external racemase^[77]. Alternatively, a subtype of C domain as observed in arthrofactin biosynthesis displays dual activities, catalyzing both condensation and epimerization of L-amino acids^[78]. However, the most common method involves epimerization of a PCP-tethered L-amino acid, catalyzed by an epimerization (E) domain which is integrated into the respective module. This catalytic domain acts in *cis* on aminoacyl-*S*-PCP and peptidyl-*S*-PCP intermediates during initiation and elongation, respectively (**Figure 7C**).

When the last module is reached, the final peptide chain is typically transferred to a TE domain, and released from the assembly line (see also **section 3.4**).

3.3 PKS-NRPS hybrids

Hybrid PK-NRP natural products are derived from both amino acids and carboxylic acids. The corresponding biosynthetic machineries can be divided into two classes according to the mechanisms by which the different building blocks are incorporated into the final product^[79;80]. The first class does not involve functional interactions between NRPS and PKS modules. Consequently the peptide and polyketide moieties are synthesized independently and are coupled afterwards by a discrete enzyme, as for example in coronatine biosynthesis^[81-83]. The second class, which includes most of the myxobacterial hybrid products, involves a functional interaction between NRPS and PKS modules, i.e. a PKS-bound growing ketide chain is directly elongated by a NRPS module or vice versa. The functionality of mixed systems depends on several events. First, a PPTase with broad substrate specificity for both ACPs and PCPs is needed to convert both types of CP domains into their active forms. Most of the PPTases analyzed to date, including EntD and ACPS from *E. coli*, show high specificity towards a certain type of CP^[84]. But with the biochemical characterization of Sfp, Svp and the myxobacterial MtaA, three PPTases have been discovered that efficiently 4'-phosphopantetheinylate both ACPs and PCPs, as required for the functioning of hybrid PKS-NRPS systems^[85-87].

Another point to be considered is the transfer of intermediates across NRPS-PKS and PKS-NRPS interfaces, respectively. At a PKS-NRPS interface, the C domain must condense a

ketide chain on an upstream ACP domain with an amino acid extender unit tethered to the downstream PCP. In contrast, at NRPS-PKS interfaces, the KS domain must first accept an upstream peptidyl intermediate for transacylation onto its own active site, and subsequently catalyze the condensation reaction with a downstream acyl-S-ACP. Consequently ACP/PCP and C/KS are the critical domains for the functional interaction in hybrid PKS/NRPS systems. They have to recognize and cooperate with domains from the other type of system on the one side and on the other side they also have to accept and catalyze the corresponding substrates. Obviously, intermodular communication in hybrid systems is essential for the transfer of the growing chain over the mixed interfaces. This process is also facilitated by sequence regions that either physically join the PKS and NRPS modules together within subunits ('linkers'), or allow them to interact non-covalently ('docking domains') across intersubunit junctions.

3.4 Type I and Type II TEs: chain release and proof-reading

In most cases, the last module of both PKS and NRPS assembly lines contains a C-terminal type I thioesterase (TE) domain which is responsible for release of the full-length product. Here, the intermediate is transacylated from the Ppant arm of the final carrier protein onto the active site serine of the terminal TE domain. The TE domain then catalyzes the controlled off-loading of the product (**Figure 9A**). One mode of chain release is hydrolysis, in which the final intermediate is transferred to water as external nucleophile. Hydrolysis is the fate, for example, in the release of the myxobacterial metabolites spirangien and DKxanthene by their respective assembly lines ^[92;93]. Alternatively, and more frequently, the TE catalyzes an intramolecular nucleophilic attack of a suitable internal nucleophile (e.g., -OH or -NH₂) on the product chain, releasing a macrocyclic compound ^[94]. A large variety of products can result from this reaction, based on the nature of the acyl chain (polyketide, polypeptide or hybrid), the location of the nucleophile which is used for the cyclization (to generate simple rings or more complex branched structures ^[95]), and the type of linkage. For example, if an internal hydroxyl is used to close the ring, the result is a macrolactone such as in the polyketide erythromycin A ^[96]. Alternatively, the selection of an amine as internal nucleophile can result in an intramolecular peptide bond, as in the polyketide vicenistatin ^[97]. Furthermore, macrocycles can be also formed by the cyclooligomerization of several intermediate chains, as in the case of gramicidin S ^[98], enterobactin ^[99] and disorazol ^[100]. In the case of gramicidin S, a ten-membered ring is formed by the head-to-tail joining of two identical pentapeptidyl units ^[98]. The ability of TEs to catalyse a variety of chain release reactions also raises the interest in their potential as useful biocatalysts. It could be shown,

that NRPS TE domains expressed as stand-alone proteins efficiently catalyze macrocyclization of linear peptides, which vary significantly in structure from their native substrates^[101-103]. Considering the large number of TE domains found to date, these enzymes might find utility in future in the synthesis and diversification of cyclic peptides and polyketides.

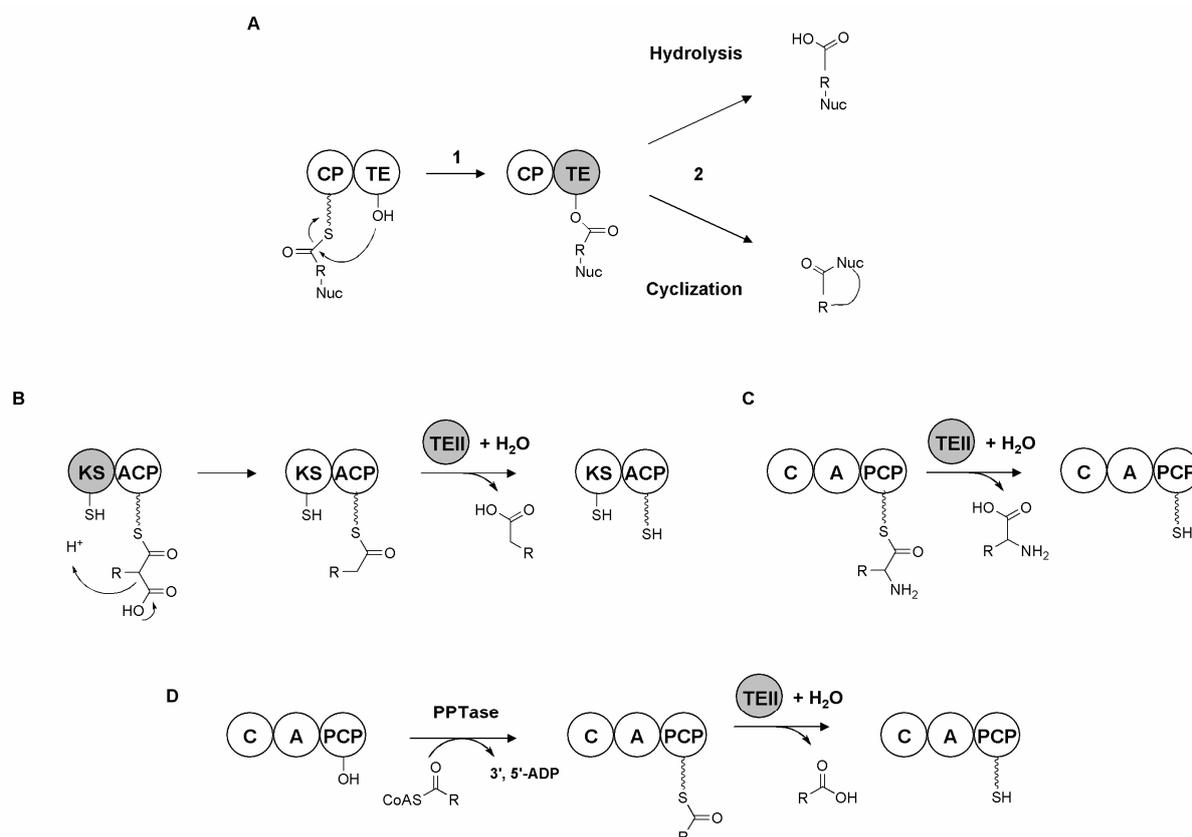


Figure 9. Role of type I and type II TE domains in natural product synthesis. (A): Transfer of the final acyl chain from the phosphopantetheinyl arm of the last ACP onto the active site serine of the type I TE domain (1). Depending on the character of the TE, the acyl-enzyme intermediate can undergo hydrolysis and be released as free acid or experience intramolecular cyclization to form a macrocyclic product. **(B):** Aberrant decarboxylation of the chain extender unit catalyzed by the KS domain. The resulting acyl group (e.g. acetate or propionate) is attached to the ACP and blocks the assembly line. The discrete type II TE catalyzes hydrolytic release of the acyl chain, restoring the activity of the ACP domain. **(C):** Removal of a wrongly attached amino acid from the PCP by TE II. **(D):** Erroneous modification of a PCP with an acyl-CoA derivative due to the relaxed selectivity of the PPTase. To regenerate the inactive NRPS, the TE II hydrolyzes the thioester and leaves only the 4'-phosphopantetheine cofactor on the PCP. Domains involved in a particular process are colored in grey.

In addition, some biosynthetic gene clusters also contain a second TE domain. These discrete type II TEs are not absolutely essential for secondary metabolite production by the modular megasynthases, as various disruption experiments on type II TEs resulted in a significant decrease but not in complete abolishment of product yields^[104-106]. It was therefore postulated that type II TEs improve the efficiency of product formation by their associated NRPS/PKS

systems by regenerating blocked assembly lines. In PKS systems, ACPs can be stalled by short fatty acyl thioesters that are derived by aberrant KS-catalyzed decarboxylation of the chain extender units in the absence of an appropriate acceptor chain. The TEII would thereupon interact with the incorrectly acylated ACP by transferring the corresponding acyl group to its active site serine (**Figure 9B**). The regenerated ACP is then able to continue its normal operations, while the TEII releases the acyl group via hydrolysis ^[107].

In NRPS systems, TEIIs are assumed to recover the activity of blocked NRPS modules in two ways. During the conversion from *apo* to *holo* NRPS, the PPTase may transfer an acyl-4'-phosphopantetheine moiety instead of a 4'-phosphopantetheine group onto a PCP. This mistake would lead to a misprimed NRPS that is inactive, as one of its 4'-PPant arms is blocked by an acyl group. It has been shown that TEIIs can regenerate misprimed NRPS by hydrolyzing these short chain acyl CoA-groups (**Figure 9D**) ^[108]. It seems likely that malfunctioning PPTases could also cause blockages in PKS systems, but this mechanism has not yet been directly demonstrated. Additionally, type II TEs in NRPS systems can also restore the activity of modules that are blocked with unprocessed aminoacyl intermediates. These incorrectly loaded amino acids are recognized by the TEII and consequently released by hydrolysis from the corresponding PCP (**Figure 9C**) ^[108;109]. In mixed PKS-NRPS systems, type II TEs are required to be more promiscuous, as they have to recognize misprimed ACPs as well as PCPs, and the corresponding aberrant intermediates ^[110].

3.5 Post-assembly line modifications

After release from the multienzyme complex, many secondary metabolites are chemically modified by discrete enzymes which are also encoded in the biosynthetic gene clusters ^[111;112]. These modifications modulate the physicochemical properties such as hydrophobicity and/or binding properties of the released product scaffold, and are often responsible for conferring bioactivity on the natural products. In actinomycetes, PKs are often decorated with additional moieties such as sugars, hydroxyls and methyl groups ^[50;113]. In contrast, NRP products are typically modified during assembly of their scaffold by *cis* acting domains (see **part 3.2**). Nevertheless glycosylations and oxidative cross-linking by *trans* acting enzymes, have been observed in NRPS systems ^[111].

Post-assembly line modifications, particularly glycosylations, are quite rare in myxobacteria ^[20;21]. Only a few glycosylated myxobacterial compounds have been identified to date, among them chivosazol and sorangicin ^[41;114]. In addition, halogenation reactions which are quite common for natural products from marine microbes are also rather unusual in myxobacterial

compounds ^[8]. However, two halogenases involved in the biosynthesis of two different myxobacterial natural products were described recently. One of these enzymes chlorinates the 2-position of a tryptophan residue in the chondramides, and therefore represents an example of a rare tryptophan 2-halogenase ^[115]. The second halogenase was discovered in the chondrochloren gene cluster and is FAD-dependent. The natural substrate of this halogenase could not be identified yet. However it was assumed that it accepts carrier-bound substrates ^[116]. Furthermore, hydroxylations, methylations and acylations of myxobacterial natural products have been observed ^[44].

Enzymes which are involved in post-assembly line modifications have been also targeted for their potential as biocatalysts ^[117]. They often display novel reaction mechanisms, which can be potentially used for the synthesis and derivatization of pharmaceutical compounds and drug intermediates ^[117]. For example studies of glycosyltransferases were carried out to define their biocatalytic potential ^[118;119]. In addition experiments with P450 enzymes showed that these enzymes are able to hydroxylate next to their native PKS substrates also non-native substrates and are consequently good candidates for bio-engineering ^[120].

4. Outline of this work

The overall goal of the work described in this thesis was to carry out studies on the biosynthesis of several myxobacterial secondary metabolites and consequently increase our knowledge about both modular assembly lines and post-assembly line modifications. Specific aims included a comparative analysis of thuggacin biosynthesis in *Chondromyces crocatus* Cm c5 and *Sorangium cellulosum* So ce895, and studying several aspects of the assembly of three other metabolites, ajudazol (*Chondromyces crocatus* Cm c5), disorazol (*Sorangium cellulosum* So ce12) and leupyrrin (*Sorangium cellulosum* So ce690).

The thuggacins (**Figure 10**) are macrolide antibiotics which were identified recently from the myxobacterial strains *Chondromyces crocatus* Cm c5 and *Sorangium cellulosum* So ce895 [34;35]. The most striking structural difference between the set of compounds in each strain is the branching functionality at carbon C2. *C. crocatus* Cm c5 thuggacins (Cmc-thuggacins) incorporate a methyl moiety at this position whereas the *S. cellulosum* So ce895 thuggacins (Soce-thuggacins) contain an uncommon hexyl side chain. The only other known secondary metabolite with such a hexyl-branching is cinnabaramide from the *Streptomyces* strain JS360 [121]. In addition, the pattern of modification by hydroxylation also differs between the strains. Soce-thuggacins incorporate a hydroxyl group at carbon C20 which is not present in the Cmc compounds, while in contrast, the C32 hydroxyl moiety of the Cmc thuggacins is absent in the Soce counterparts. Ring-size variants are found in both strains, but these derivatives likely arise from spontaneous rearrangement during the isolation process and are thus not biologically relevant [35]. Interestingly, the ring configuration does not appear to influence the compounds' bioactivity, as Soce-thuggacins A and B are equally active [35]. In addition to these notable structural features, the thuggacins show promising biological activity. The compounds were found to inhibit the respiratory chain in several Gram-positive bacteria, including clinical isolates of *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) [34;35]. As TB remains the second most lethal infectious disease worldwide and drug-resistant TB strains are increasingly emerging, the demand for novel therapeutics has increased [122]. Chemotherapeutics such as the thuggacins which target the bacterial respiratory chain are of particular interest, as they also affect dormant bacteria [123]. As the thuggacin gene clusters had not been identified in either strain, the first goal was to locate the biosynthetic loci by screening the appropriate cosmid libraries, coupled with gene inactivation experiments in *C. crocatus* Cm c5 and *S. cellulosum* So ce895 to prove the cluster identities. For *C. crocatus* Cm c5, a cosmid library was already available [115] and methods for genetic

modifications were established, whereas for *S. cellulsum* So ce895, the project required both generating a cosmid library and developing methods for genetic manipulation of the strain.

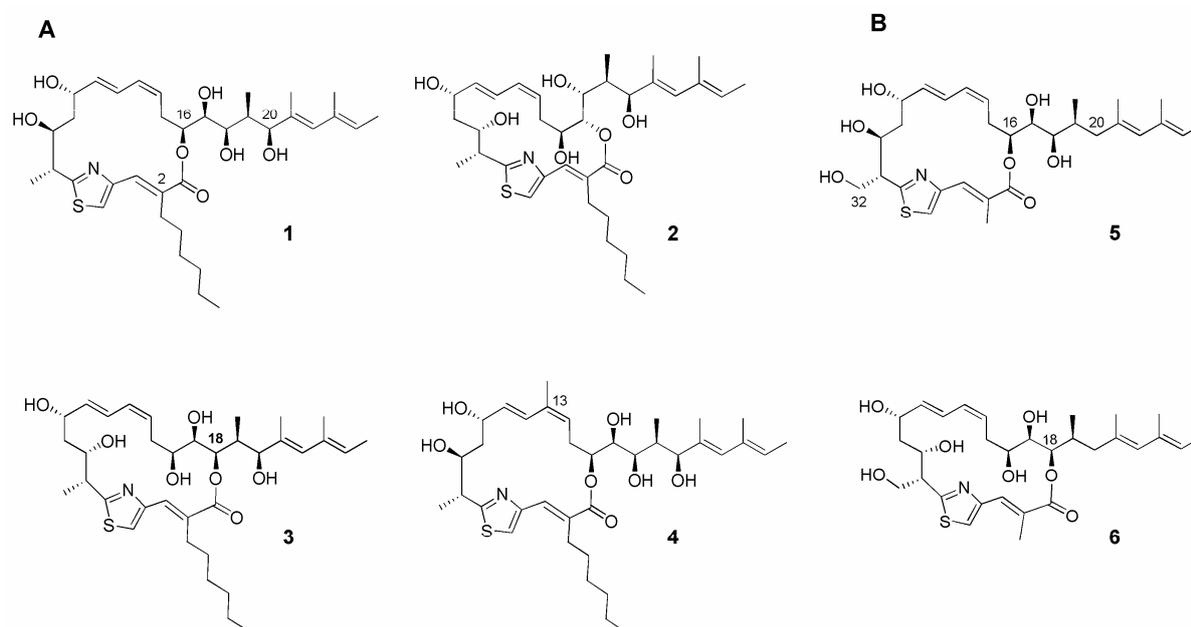


Figure 10. Structures of the thuggacins produced by *Sorangium cellulsum* So ce895 and *Chondromyces crocatus* Cm c5. (A) Thuggacins derived from *S. cellulsum* So ce895: Soce-thuggacin A (1), Soce-thuggacin B (2), Soce-thuggacin C (3), and the minor metabolite 13-methyl-thuggacin A (4). Compounds 2 and 3 are derived from spontaneous rearrangement of 1. (B) Thuggacins isolated from *C. crocatus* Cm c5: Cmc-thuggacin A (5) and Cmc-thuggacin C (6). 6 is derived from 5 during isolation.

Once the clusters were identified, sequenced and annotated, both biosynthetic machineries were compared in detail to elucidate the origin of the structural variations in the metabolites derived from both strains. For this, post PKS/NRPS genes that were putatively involved in this process were inactivated in the appropriate strain, and where possible, the corresponding recombinant proteins were analyzed *in vitro* via an appropriate biochemical assay. As the thuggacin yields in *C. crocatus* Cm c5 are quite low, an additional aim was to increase production by inserting a strong constitutive promoter directly in front of the biosynthetic gene cluster by homologous recombination. The obtained data and their analysis are reported in **Chapter 3, 4 and 7**.

The antifungal ajudazols A and B (**Figure 11**) are potent inhibitors of mitochondrial electron transport, and were also isolated from the natural product multi-producer *C. crocatus* Cm c5 [124;125]. Ajudazols are novel isochromanone derivatives which incorporate an extended side chain containing a heterocyclic oxazole ring, a *Z,Z*-diene and a 3-methoxybutenoic acid amide. Ajudazol A, the major metabolite, contains an *exo*-methylene functionality at carbon C-15 instead of the methyl group found in ajudazol B. The aim of this project was to annotate

the corresponding gene cluster, which was initially identified by Dr. Shwan Rachid, and to use this information to develop a detailed biosynthetic hypothesis. Furthermore, the roles of two cytochrome P450-encoding genes involved in post-assembly line modifications were investigated. For this, the genes in *C. crocatus* Cm c5 were inactivated, and the structures of the resulting products were analysed by HPLC-MS and NMR, in order to determine the function of the enzymes. The results of this study are presented in **Chapter 1**.

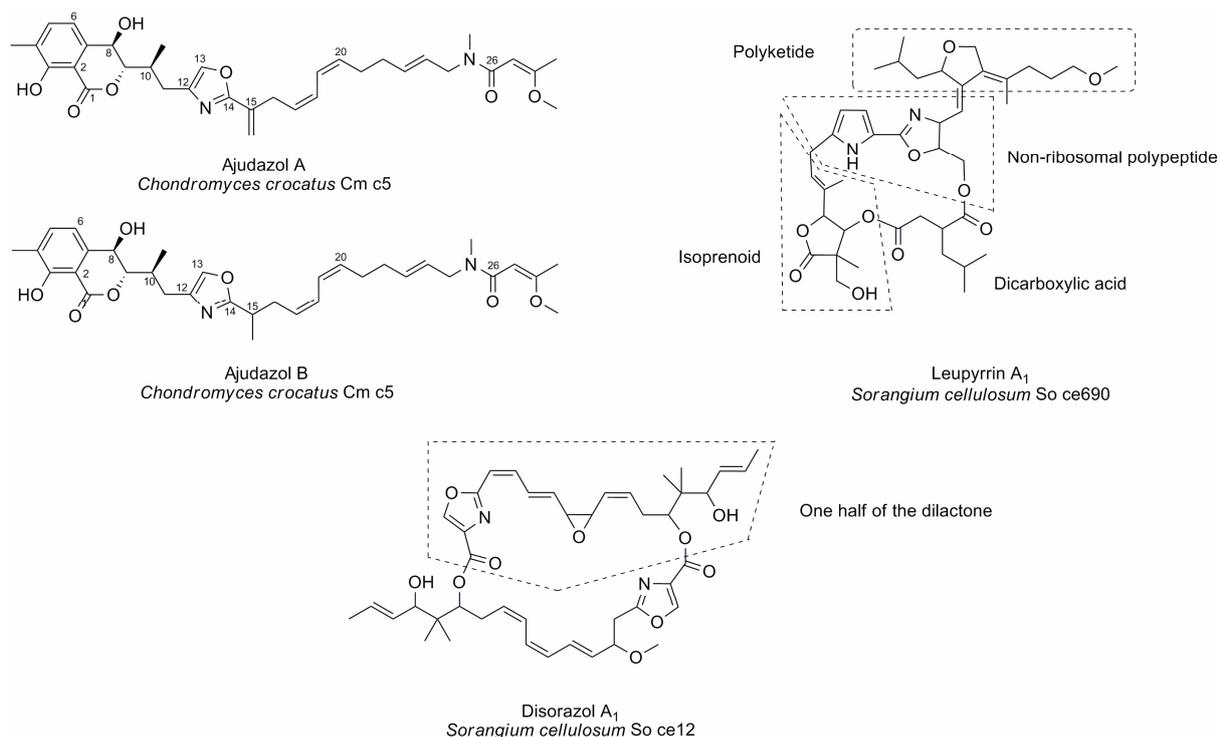


Figure 11. Structures of ajudazol A and B produced by *Chondromyces crocatus* Cm c5, leupyrrin A₁ from *Sorangium cellulosum* So ce690 and disorazol A₁ derived from *Sorangium cellulosum* So ce12.

Another goal of the ajudazol project was to investigate the formation of the isochromanone ring. The isochromanone ring system is quite uncommon for a modular type I PKS, as biosynthesis of aromatic structures by bacteria is typically accomplished by type II or type III PKS systems. Other rare examples for bacterial type I PKS products with aromatic moieties are stigmatellin and lasalocid. In the stigmatellin pathway a novel cyclization domain which is located at the C-terminal end of the assembly line appears to catalyze cyclization and aromatization to yield the final chromone ring^[126]. The mode of formation of the benzenoid ring in lasalocid biosynthesis remains unknown, although the bacterial type I PKS incorporates a typical type I TE domain^[127]. Therefore the question emerged how the isochromanone ring system is formed by the ajudazol PKS-NRPS, which terminates in a TE domain. The role of the ajudazol TE in the ring formation process was investigated using

experiments both *in vitro* and *in vivo*. The obtained data and their analysis are reported in **Chapter 2**.

In addition, the role of the C-terminal TE domain in disorazol biosynthesis was investigated. Disorazol (**Figure 10**) consists of two nearly identical monomers that are fused together at the end of the assembly line to form the characteristic dilactone core structure ^[100]. The aim of this project was to express the TE domain heterologously and to examine *in vitro* its function in macrodiolide formation. The results of this project are summarized in **Chapter 5**.

The final project targeted the leupyrrin family of compounds, which were isolated from several *Sorangium cellulosum* strains ^[128]. Leupyrrins are an elegant example of the extent of structural variety within myxobacterial compounds (**Figure 10**): in addition to PK and NRP building blocks, the molecules also incorporate an isoprenoid unit as well as a dicarboxylic acid ^[129]. The majority of the biosynthetic gene cluster responsible for leupyrrin assembly had already been identified by Maren Kopp ^[130]. Therefore, as follow-up work, *in vitro* studies were performed to investigate a specific step in the biosynthesis, the generation of the starter unit pyrrole carboxylic acid. This study is presented in **Chapter 6**.

Chapter 1

Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications

Buntin, K., Rachid, S., Scharfe, M., Blöcker, H., Weissman, K.J., Müller, R. (2008)

Angew Chem Int Ed Engl.; 47(24): 4595-99

This article is available online at:

<http://dx.doi.org/10.1002/anie.200705569>

Chapter 2

An unusual thioesterase “chaperones” isochromanone ring formation in ajudazol biosynthesis

Buntin, K., Weissman, K.J., Müller, R. (2010)

ChemBioChem, 11(8): 1137-46

This article is available online at:

<http://dx.doi.org/10.1002/cbic.200900712>

Chapter 3

Stereochemical determination of thuggacin A-C, highly active antibiotics from the myxobacterium *Sorangium cellulosum*

Bock, M., Buntin, K., Müller, R., Kirschning, A. (2008)

Angew Chem Int Ed Engl.; 47(12): 2308-11

This article is available online at:

<http://dx.doi.org/10.1002/anie.200704897>

Chapter 4

Biosynthesis of the anti-tuberculosis macrolide thuggacins in myxobacteria: comparative cluster analysis reveals the basis for natural product structural diversity

Kathrin Buntin, Herbert Irschik, Kira J. Weissman, Eva Luxenburger, Helmut Blöcker, and Rolf Müller (2010)

Chem. Biol., 17(4):342-56

This article is available online at:

<http://dx.doi.org/10.1016/j.chembiol.2010.02.013>

Chapter 5

Studies on the disorazol TE

Introduction

The bioactivity of several myxobacterial natural products is based on their interaction with the cytoskeleton of cells from higher organisms. Compounds with this mode-of-action are of particular interest in the pharmaceutical industry, as they can potentially be applied in cancer therapy. Disorazol, which is produced by the myxobacterial strain *Sorangium cellulosum* So ce12, inhibits the polymerization of tubulin and induces apoptosis at picomolar concentrations [131]. The disorazol family of metabolites consists of 29 derivatives that on the basis of their structural characteristics, can be sub-divided into 7 structural groups and 3 additional metabolites [132]. All disorazols are macrocyclic dilactones of a 4-oxazolecarboxylic acid that is linked with a C₁₅ chain in position 4, except for disorazol Z which incorporates a shorter, C₁₃ chain. The diversity of the disorazol family arises by variable modification of this chain, including altering the configuration and location of double bonds, and the presence or absence of epoxide, hydroxyl and methoxyl functionalities. In addition, ring-expanded disorazols have been observed [132].

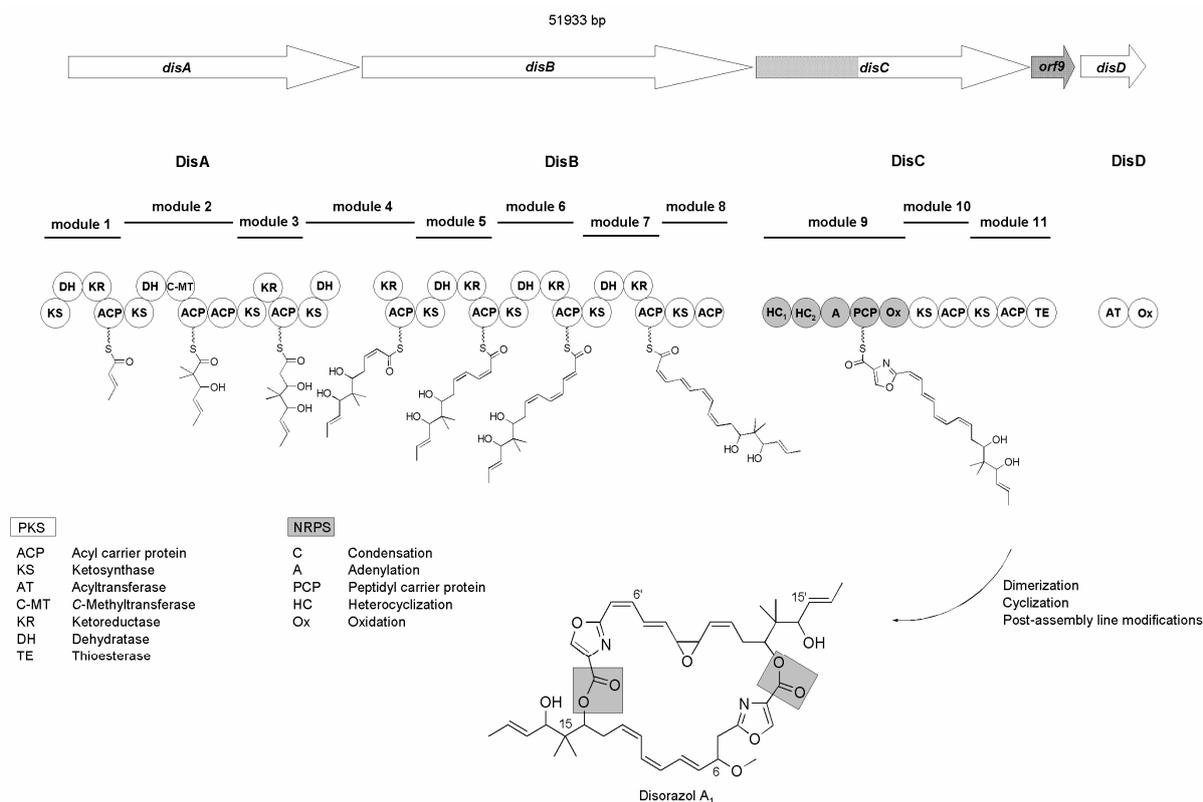


Figure 1. Schematic representation of disorazol biosynthesis in *Sorangium cellulosum* So ce12. The modules are used iteratively to generate the final product. Shaded in grey are the two ester linkages that are formed during dimerization of the two intermediate chains (scheme according to Kopp *et al.* [100]).

Recently the hybrid PKS/NRPS gene cluster responsible for disorazol biosynthesis was identified, and the pathway was partially elucidated [100;133]. The assembly line incorporates 10

PKS and one NRPS module, although three of the PKS modules (module 8, 10 and 11) are proposed to be inactive as their KS domains lack a conserved active site histidine. AT domains are uniformly missing from the PKS modules, and instead, the extender unit malonyl-CoA is provided to the ACP domains by the discrete *trans* acting AT domain encoded on *disD*. Thus the disorazol PKS-NRPS is a '*trans* AT' system. Furthermore, analysis of the catalytic domains in modules 1–7 led to the assumption that these modules are responsible for the formation of one monomer of the disorazol dilactone (**Figure 1**). Following module 7, the chain-extension intermediate is probably directly transferred to the NRPS module (module 9) and the inactive module 8 is skipped. After the incorporation of the amino acid serine, the necessary chain length is reached. However, the NRPS module is followed by two additional PKS modules. To account for this organization, it has been proposed that these modules do not catalyze chain extension, but instead pass the intermediate along until the C-terminal TE domain is reached ^[133].

As the disorazol biosynthetic machinery only contains the number of modules required for biosynthesis of one monomer within the bislactone, it is likely that the modules are used twice to generate the final product. Consequently, the TE domain is likely to be involved in the dimerization of two monomeric chains to yield the final macrodiolide structure. However, to attempt to demonstrate directly this function of the disorazol TE, the domain was expressed heterologously and *in vitro* studies were performed with a substrate which mimicked one simplified monomer of the disorazol Z molecule, derivatized as its *N*-acetylcysteamine (NAC) thioester (**Figure 2B**).

Experimental procedures

Cloning of pET28bDisTE

Disorazol TE was PCR amplified from BAC D17 ^[100] using the primers DisTEfwdI (5'-tgagc**CATATG**agcaacggcgcagccccggca-3') and DisTErev (5'-acctg**GCGGCCGC**ctcatgaaagccctcgcgacgtg-3'). The forward primer contained an introduced *NdeI* restriction site while the reverse primer was designed to include a *NotI* site downstream of the stop codon (restriction sites are shown in bold and underlined). The PCR product was cloned into pJET1.2 (Fermentas), digested with *NdeI* and *NotI*, and subsequently ligated into the expression vector pET-28b+ (Novagen), previously digested with *NdeI* and *NotI*. The obtained expression construct was designated as pET28b-DisTE and verified by sequencing.

Heterologous expression and purification of DisTE

The expression construct pET28b-DisTE was transformed into the strain *E. coli Rosetta* BL21 (DE3)pLysS/RARE (Novagen). Expression was carried out in LB medium (200 mL) containing kanamycin sulfate (40 $\mu\text{g mL}^{-1}$) and chloramphenicol (20 $\mu\text{g mL}^{-1}$) at 37 °C. Protein expression was induced at $A_{600} = 0.8\text{--}1.0$ by addition of isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.2 mM. After induction, the cells were cultivated at 16 °C over night. Cells were then harvested by centrifugation at 15344 g at 4 °C. Purification of the protein was carried out at 4 °C using an ÄktaPrime Purification System (GE Healthcare). The cell pellet was resuspended in buffer A (20 mM Tris (pH 7.8), 200 mM NaCl, 10% glycerol, 10 mM imidazole; 20 mL). The cells were broken by three passes through a French Press (1000 psi) and the insoluble material was sedimented from the lysate by centrifugation at 15344 g at 4 °C. The lysate was filtered through a 1.2 μm syringe filter (PALL[®]), before being applied to a HisTrap[™] HP column (1 mL; GE Healthcare). All steps of the purification were carried out at a flow rate of 1 mL min⁻¹. The protein extract (20 mL) was loaded onto the column after an equilibration step with buffer A (20 mL). After loading, the column was washed with buffer A (20 mL), and then the proteins were eluted using a stepwise gradient with buffer B (buffer A + 500 mM imidazole) to give concentrations of 60, 100, 200, 300 and 500 mM imidazole. Elution of the proteins was monitored by recording the absorbance at 280 nm. Appropriate fractions were analyzed by SDS-PAGE. The fractions containing the recombinant protein were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column (GE Healthcare) into storage buffer (50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10% glycerol, 2 mM DTT). Purified protein was then flash frozen in liquid nitrogen and stored at -80 °C. Typically, 1.7 mg of purified protein was obtained from 200 mL cell culture.

Enzyme assay

The reaction mixture (120 μL) contained buffer (50 mM Tris (pH 8.0), 150 mM NaCl), 5 μg NAC thioester (synthesized by R. Schäckel and M. Kalesse, University of Hannover.) and 0.75 μg purified DisTE. The assay was incubated overnight at 30 °C and quenched by the addition of 120 μL MeOH. After centrifugation, the supernatant was dried by evaporation and then redissolved in 100 μL MeOH. Finally, the sample was analyzed using high resolution mass spectrometry. High-resolution measurements were performed on an Accela UPLC-system (Thermo-Fisher) coupled to an LTQ-Orbitrap (linear trap-FT-Orbitrap combination) operating in positive ionization mode. Compounds were separated on a BEH RP-C₁₈ column (Waters; 50 \times 2 mm, 1.7 μm particle size, flow 0.6 mL min⁻¹), using a solvent system consisting of water and acetonitrile, both containing 0.1% formic acid. The following gradient

was applied: 5–95% acetonitrile over 9 min. The UPLC-system was coupled to the Orbitrap by a Triversa Nanomate (Advion), a chip-based nano-ESI interface.

Results

In order to demonstrate the role of DisTE in the dimerization process, we aimed to evaluate the activity of the domain towards a synthetic NAC thioester which mimics one monomer of the dilactone. The expression construct for DisTE was designed to incorporate the majority of the linker between the TE and the upstream ACP domain (starting 5 amino acids C-terminal to the ACP). This strategy had already proven successful for the expression of the terminal TE domain from the erythromycin PKS, which is also located downstream of an ACP domain [134]. The gene was cloned into the vector pET28b+ in order to obtain a N-terminally His₆-tagged protein, and expression was carried out in *E. coli* RosettaTM BL21(DE3)pLysS/RARE cells. Recombinant DisTE was then purified by affinity chromatography (**Figure 2A**), and the identity of the protein was confirmed by MALDI-MS analysis.

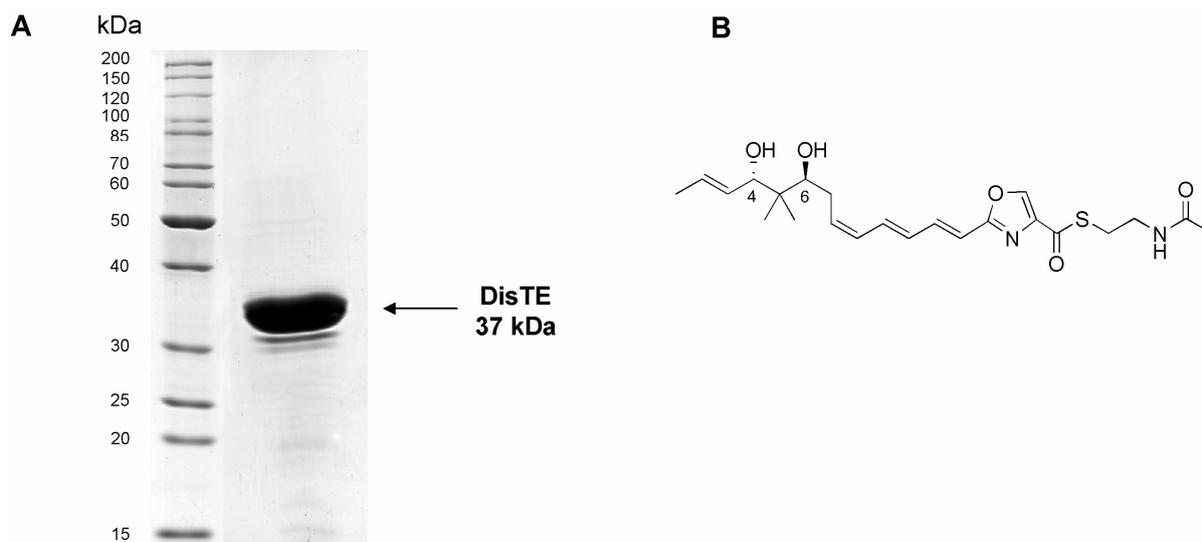


Figure 2. (A): SDS-PAGE analysis of the DisTE domain (calculated molecular weight: 37.2 kDa). The left lane contains molecular weight marker. **(B):** NAC thioester used in this study.

The available NAC-ester (provided by collaboration partners Romy Schäckel and Markus Kalesse, University of Hannover) represents a simplified monomer of the natural product disorazol Z (**Figure 2B**). For each assay, the TE (0.75 μ g) was incubated with the linear NAC-monoester (5 μ g) in assay buffer, and analysis of the reaction products was carried using HPLC-MS (**Figure 3**). The major compound that was detected had an accurate mass of m/z $[M+H]^+ = 348.18072$ (retention time (r.t.) = 5.7 min). This mass is consistent with the molecular formula C₁₉H₂₅NO₅ (calc'd. $[M+H]^+ = 348.18055$, $\Delta = 0.5$ ppm), which

corresponds to the hydrolytic product of the NAC-monoester, i.e. the free acid of the model substrate. In addition, minor amounts of two compounds with the accurate mass of m/z $[M+H]^+ = 778.37424$ (r.t. = 7.7–7.8 min) were detected.

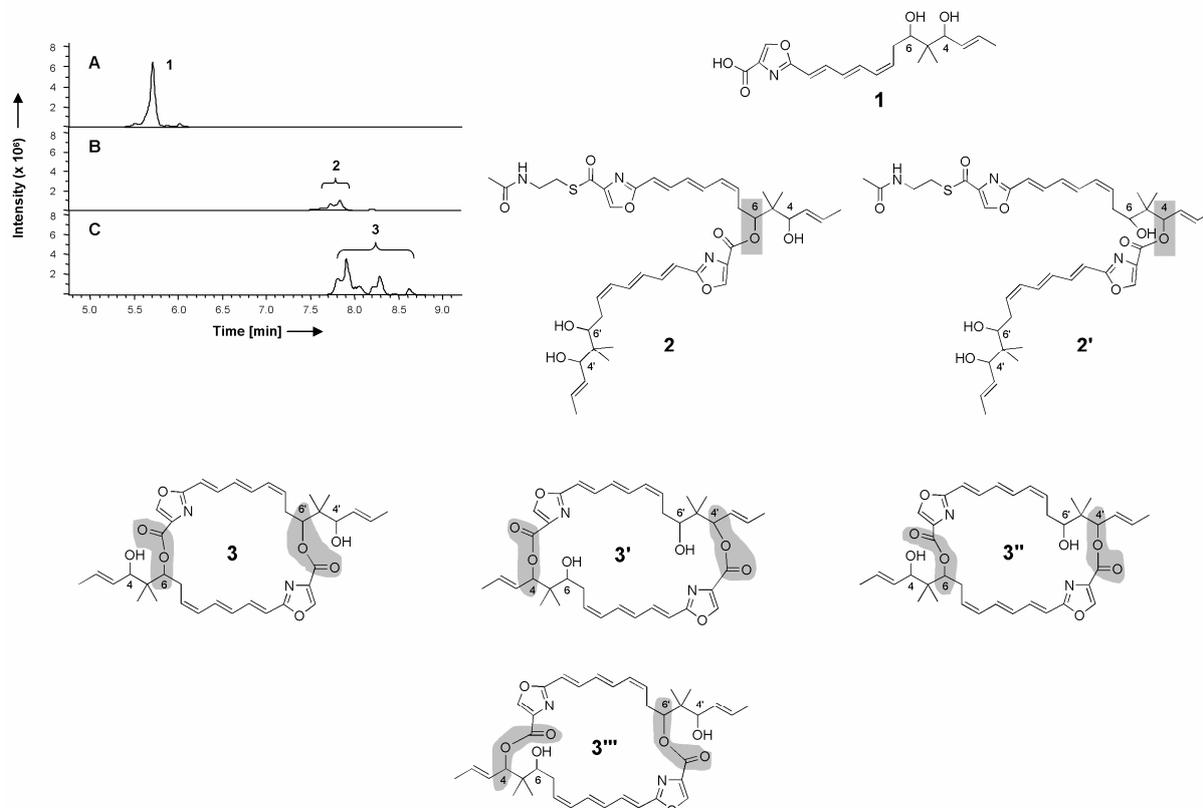


Figure 3. High resolution HPLC-MS analysis of products formed by DiSTE. (A): Extracted ion chromatogram EIC of m/z $[M+H]^+ = 348.180$. A compound (r.t. 5.7 min) with a mass consistent with the hydrolytic product of the NAC-monoester (structure 1) (calculated m/z $[M+H]^+$ for $C_{19}H_{25}NO_5 = 348.18055$; found: 348.18072; $\Delta = 0.5$ ppm) was detected. (B): EIC of m/z $[M+H]^+ = 778.374$. Two compounds (r.t.= 7.7–7.8 min) corresponding to the mass of linear dimer tethered to the NAC moiety (most likely structures 2 and 2') (calculated m/z $[M+H]^+$ for $C_{42}H_{56}N_3O_9S = 778.37318$; found: 778.37424; $\Delta = 1.363$ ppm) were detected. (C): EIC of m/z $[M+H]^+ = 659.333$. Four compounds (r.t.= 7.7–8.7 min) having a mass consistent with the cyclic dimer (structures 3, 3', 3'' and 3''') (calculated m/z $[M+H]^+$ for $C_{38}H_{46}N_2O_8 = 659.33269$; found: 659.33301; $\Delta = 0.5$ ppm) were detected. The alternative esters that can be formed depending on the hydroxyl group used are shown in grey.

The corresponding sum formula $C_{42}H_{56}N_3O_9S$ (calc'd. $[M+H]^+ = 778.37318$, $\Delta = 1.363$ ppm) is consistent with a linear dimer of the substrate tethered on one end to the NAC (linear dimeric NAC-thioester). Formation of two linear dimeric NAC-thioesters (2 and 2') likely occurs by alternative use of the two nucleophilic hydroxyl moieties (C4 or C6; linear monoester numbering) during installation of the ester bond.

Furthermore, several compounds (r.t. = 7.7–8.7 min) with an accurate mass of m/z $[M+H]^+ = 659.33301$ were identified. The predicted sum formula for these compounds, $C_{38}H_{46}N_2O_8$, (calc'd. $[M+H]^+ = 659.33269$, $\Delta = 0.5$ ppm) tallies with the cyclic dilactone structure. These

structures likely correspond to dilactones of varying size, whose formation again depends on the particular hydroxyl moiety used to create the ester linkage (**Figure 3**). As further evidence for the role of the DisTE in generating these products, none of the compounds were formed in a negative control containing boiled DisTE (data not shown). The comparison of the integrated peak areas from the different products resulted in following approximate ratios: free acid to linear dimer 4.5:1 and free acid to cyclic products 1.5:1.

Discussion

Several non-ribosomal peptide natural products consist of dimerized or trimerized monomeric chains, including gramicidin S^[98] and enterobactin^[99]. In addition such structures can also be generated by hybrid PKS-NRPS machineries, as for example in rhizopodin biosynthesis^([135] and D. Pistorius, unpublished). The monomeric chains are produced by iterative use of the assembly line, while the cyclo-oligomerization is catalyzed by the TE domain at the C-terminal end of the biosynthetic machinery^[136]. Consequently, in these systems, TE domains are of particular interest as they are somehow able to ‘count’ the appropriate number of linked monomers, before releasing the final product by intramolecular cyclization.

This iterative mode of biosynthesis has been demonstrated directly for the gramicidin (GrsB) and enterobactin (EntF) TEs, which produce dimerized and trimerized products, respectively^[136;137]. After the first chain is transferred to the TE, the TE selects a nucleophilic residue in the substrate tethered to the upstream PCP, and uses it to attack the acyl terminus of the chain bound at its active site serine. This step yields a linear dimer attached to the PCP domain. The dimeric intermediate is then transferred to the TE, at which point the TE either catalyzes intramolecular cyclization/release, or forms a trimeric species in cooperation with the upstream PCP, before releasing the final product^[136;137]. By constructing molecules in this way the ‘economic’ advantage for the bacterium is obvious. Instead of using a very large assembly line that generates the natural product in one run, a shorter and energetically cheaper biosynthetic machinery is used several times in series to generate one molecule. However this system requires the evolution of a specialized TE capable of catalyzing the required reactions. The mixed PKS-NRPS system for disorazol biosynthesis also appears to use this mechanism, as the natural product consists of a cyclic dimer of two almost identical monomeric chains, although the corresponding gene cluster only contains the number of modules appropriate to generate chains with the length of one monomer. The structural differences between the monomeric chains in the final molecule are postulated not to be generated during the

assembly of the natural product on the biosynthetic machinery but after the release of the core molecule by post assembly line modifications.

To show directly that DisTE has a similar function as GrsBTE and EntFTE, the domain was expressed heterologously and tested for its ability to form the dimeric cyclic product with a NAC-thioester mimicking one half of the disorazol Z molecule. A simplified version of the disorazol Z monomer was chosen as it represents the disorazol with the shortest C-chain and is therefore easier to synthesize. The major product of this experiment was the free acid of the linear monomeric NAC-thioester, which likely arose from DisTE-catalyzed hydrolysis. However, cyclic dimeric products as well as linear dimeric NAC-thioesters were also detected in minor amounts, whereas the ratio between free acid and cyclic dimeric products is approximately 1.5:1 and the ratio between free acid and linear dimeric products is 4.5:1. The cyclic derivatives with identical masses likely correspond to differently-sized macrodiolide rings, as each monomeric-NAC-thioester has two free hydroxyl groups (C2 and C6) that can potentially participate in ester bond formation (**Figure 3**). Indeed, variously-sized macrodiolide rings, representing all possible combinations of ester linkages between the free hydroxyl groups were also observed among the 29 natural disorazol derivatives, i.e. ester bonds are formed between either the outer or inner hydroxyl groups but also all mixed possibilities (outer and inner hydroxyl group are linked with each other via an ester bond) could be detected ^[132]. In addition, alternative use of hydroxyl nucleophiles also likely accounts for formation of the two different linear dimeric NAC-thioesters. Based on the obtained data, it appears that in general the DisTE shows relatively broad specificity towards the different nucleophiles in the *in vitro* assay. These findings are supported by the various-sized disorazols produced *in vivo*. However, a certain preference of the DisTE for ester bond formation between the two inner hydroxyl groups can be assumed as the major natural product is disorazol A₁ (**Figure 1**) ^[132].

Characterization of the excised tyrocidine TE domain also yielded a similar ratio of hydrolytic versus cyclized products ^[103]. In contrast, when such an iteratively-acting TE was expressed as a fusion protein with the corresponding upstream PCP, as for GrsB PCP-TE, the amount of hydrolytic product decreased relative to the yields of linear dimerization and cyclic products ^[137]. In our assay, the monomeric NAC-thioester, a mimic of the acyl-CP, is provided *in trans*, as in the study with the tyrocidine domain. In the first encounter between NAC-thioester and DisTE, the intermediate chain is transferred from the NAC onto the TE domain. To generate the dimeric cyclic product, a second monomeric chain has to be presented to the DisTE. However, in contrast to the *cis*-acting PCP-TE systems in which the required monomer is

assembled on the adjacent PCP domain, performing the reaction *in trans* with discrete species, means that the second monomer will have to reach the TE active site by diffusion. Evidently, the diffusive delivery of the second NAC-thioester is too slow to guarantee an efficient encounter in time between the intermediate-tethered TE and the second monomeric NAC-thioester. Consequently the intermediate bound to DisTE is more likely to be released by hydrolysis than to dimerize with the second NAC-thioester. Taken together, our results show that it is essential for a successful dimerization process that the TE domain already loaded with one substrate chain is directly supplied with the second chain by an upstream CP domain, a result consistent with the earlier literature^[136].

A linear dimer of gramicidin – the product corresponding to the linear dimeric NAC thioesters seen with DisTE – was observed during characterization of GrsB using a monomeric pentapeptidyl-SNAC as substrate^[137]. The product only retains its NAC-functionality when the DisTE catalyzes ester bond formation between one of the two free hydroxyl moieties of a linear NAC monoester and the acyl terminus of the TE-bound monomer. If the reaction were to occur in the opposite manner, that is by attack of a free hydroxyl group of the TE tethered monomer onto the acyl terminus of the NAC-monomer, followed by hydrolysis, the products would be linear dimers of the free acids (e.g. NAC-free). As such products were not observed in these experiments, the DisTE must catalyze cyclization by the conventional mechanism.

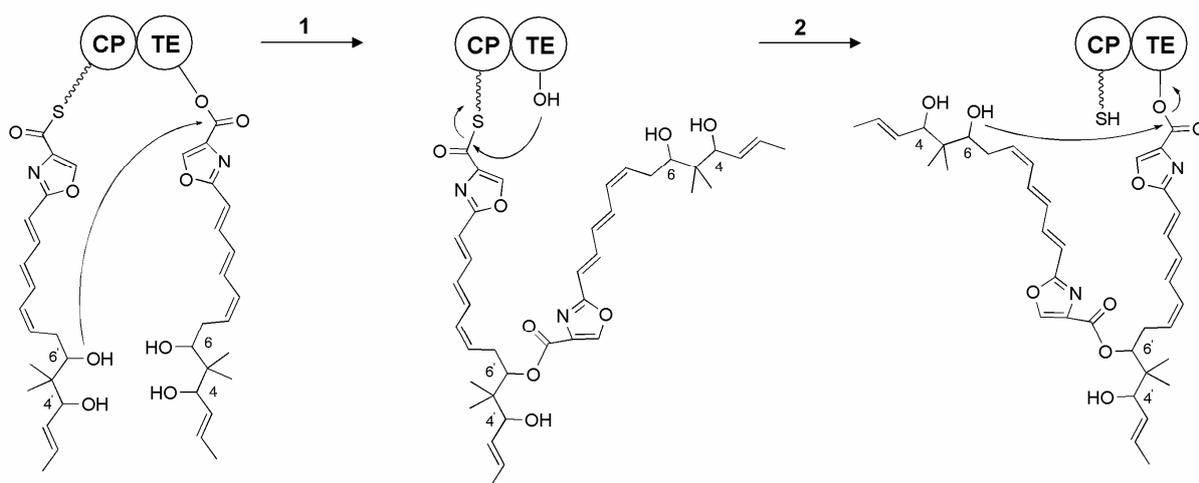


Figure 4. Proposed mechanism for the formation of disorazol. First, one of the two hydroxyl groups from the CP-tethered acyl-chain nucleophilically attacks the acyl terminus of the intermediate bound to the TE domain. In the second step, the CP-bound dimer is transferred to the active site serine of the TE, which then catalyzes the cyclization reaction, with concomitant release of the final product.

By analogy with the results obtained with GrsB^[137], we propose the following overall mechanism for the linkage and cyclization of two monomeric chains by DisTE: initially one

monomeric chain is transferred onto the active site serine of DisTE from the adjacent CP. In the next step, one of the two hydroxyl groups of a second upstream CP-tethered monomer is used to form an ester bond with the acyl terminus of the TE-bound monomer, resulting in a dimer covalently linked to the CP domain. Subsequently the dimer is transferred to the active site serine of the TE, which then catalyzes the cyclization reaction to generate the final bislactone product (**Figure 4**).

In the disorazol assembly line, several CP domains are possible candidates for involvement in the formation of the cyclic dimer. As the last building block is incorporated by module 9, it is possible that the PCP domain present in this module directly interacts with DisTE. On the other hand, both ACPs of the non-elongating modules 10 and 11 contain the active site serine necessary for the domains to be functional^[133]. Therefore, an alternative possibility is that the intermediate is handed off to these modules, and ultimately the ACP domain from module 11 and the adjacent TE cooperate to catalyze the three steps of cyclodimerization. In fact, the latter mechanism seems more likely, as ACP₁₁ and the TE domain are covalently linked to each other, which should facilitate a significantly more efficient interaction.

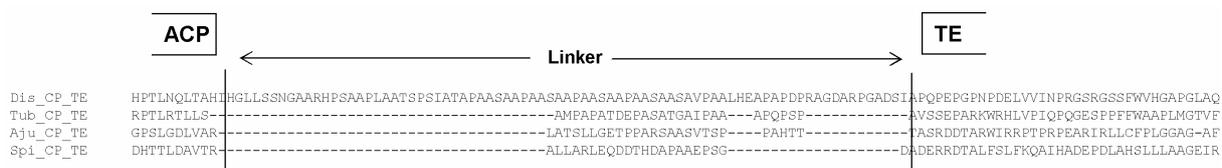


Figure 4. Sequence alignment of the linker region between ACP and TE of different PKS/NRPS hybrid systems (tubulysin, ajudazol, disorazol) and one PKS system (spirangien). The vertical bars mark the end of the ACPs and the start of the TEs.

In addition, the linker region between ACP₁₁ and the TE domain is exceptionally long compared to other myxobacterial ACP-TE junctions (**Figure 5**). A similar observation was made for GrsBTE, and it was speculated that the extended linker region may be important for the mobility of the TE^[137]. Indeed, a high degree of mobility is necessary as the enzyme has to access the multiple substrate positions on the CP-bound intermediate chains.

Chapter 6

Generation of the starter unit in leupyrrin biosynthesis

Introduction

The antifungal leupyrrins are produced by several myxobacterial strains, including *Sorangium cellulosum* So ce690 [138]. Extensive feeding studies revealed that the leupyrrins consists of a PK as well as a NRP portion, but also incorporate an isoprenoid unit and a dicarboxylic acid [129]. One amino acid that is integrated into the leupyrrin scaffold is L-proline which forms the pyrrole moiety representing the starter unit of leupyrrin biosynthesis. Pyrrole moieties derived from L-proline are found in a number of structurally diverse natural products, including coumermycin A₁ [139] and Dkxanthene [93]. The mechanism for the conversion of the pyrrolidine ring of L-proline to a pyrrole has been elucidated for the biosynthesis of clorobiocin, undecylprodigiosin, but also several other biosynthetic pathways [140-143].

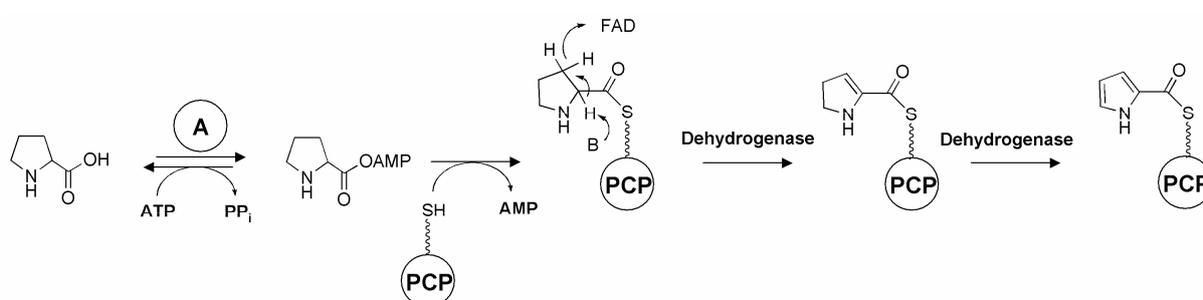


Figure 1. Schematic representation of pyrrolyl-2-carboxyl-S-PCP synthesis as shown for clorobiocin, coumermycin A₁, and undecylprodigiosin biosynthesis.

Typically in these systems three distinct enzymes work together. First an A domain recognizes and activates L-proline and subsequently transfers it to a stand-alone PCP domain to generate the L-prolyl-S-PCP intermediate. The subsequent desaturation reactions to generate the pyrrolyl-2-carboxyl-S-PCP are carried out by a FMN-dependent dehydrogenase [140;141]. The pyrrole can then be transferred to the adjacent polyketide synthase for further biosynthetic steps (**Figure 1**). Based on this precedent, we hypothesized that the leupyrrin pyrrole moiety would be formed by an analogous mechanism. Therefore, the recently identified gene cluster was searched for possible candidate genes encoding the required set of enzymes [130]. Indeed, this analysis revealed a stand-alone A domain (*leu5*) and a dehydrogenase homologue (*leu6*). For the CP domain, two alternative candidates were found: a discrete ACP domain encoded on *leu7* and a didomain consisting of C and a PCP domain (*leu9*) which also did not play an obvious, alternative role in the biosynthesis (**Figure 2**). In fact, Blast analysis shows that the discrete ACP (and the A and dehydrogenase domains) are most similar to those from the anatoxin system, suggesting that the ACP was most likely involved in the generation of the pyrrol-unit [143]. However as the nonribosomal code of the A

domain was not unambiguous^[130] we aimed to characterize the substrate specificity of the A domain *in vitro*.

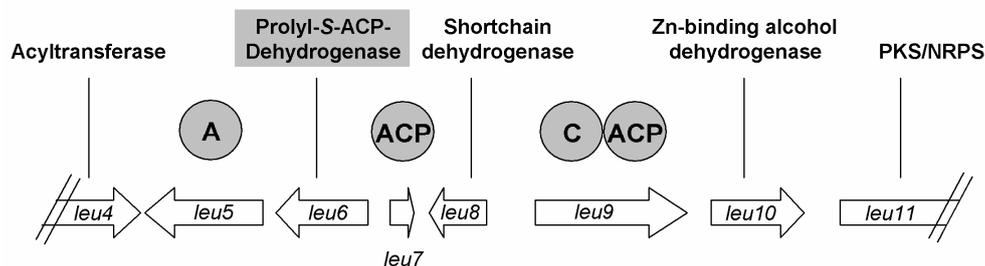


Figure 2. Portion of the leupyrrin gene cluster. Shaded in grey are the proteins that are possibly involved in the formation of the starter unit pyrrolyl-2-carboxyl-S-CP.

Experimental procedures

Cloning of pET28bleuorf5

The gene encoding *orf5* was PCR amplified from the cosmid C9. The primers used for the amplification were Leuorf5fwd (5'-gacacGCGGCCGCtactcgccttcaagc-3') and Leuorf5rev (5'-ctgctCATATGacgtacctgttcacg-3') containing the introduced restriction sites *NotI* and *NdeI*. PCR was carried out with Phusion polymerase (95°C 15 s; 67°C 20s; 72°C 30s; 30 cycles). The PCR product was cloned into pJET1.2 (Fermentas), digested with *NdeI* and *NotI*, and subsequently ligated into the expression vector pET-28b+ (Novagen), previously digested with *NdeI* and *NotI*. The obtained expression construct was designated as pET28bleuorf5 and was verified by sequencing.

Expression and purification of Leuorf5

Expression of pET28bleu5 was carried out in *E. coli* Rosetta BL21 (DE3)pLysS/RARE (Novagen). The cells were grown at 37 °C to an OD₆₀₀ of 0.8–1 in 200 ml LB media supplemented with kanamycin (40 µg ml⁻¹) and chloramphenicol (20 µg ml ml⁻¹). Expression of *orf5* was induced by addition of IPTG to a final concentration of 0.2 mM. Cells were harvested after overnight cultivation at 16 °C by centrifugation (15344 g, 5 min, 4 °C). The cell pellet was resuspended in buffer A (20 mM tris(hydroxymethyl)aminomethane (Tris)-HCl (pH 7.8), 200 mM NaCl, 10 mM imidazole, 10% glycerol) and lysed by three passes through a French press (1000 psi). Cell debris was removed by centrifugation (15344 rpm, 45 min, 4 °C). The lysate was filtered through a 1.2 µm syringe filter (PALL[®]), before being applied to a HisTrap[™] HP column (1 mL; GE Healthcare). Purification of Leuorf5 was carried out with the Äkta Prime system at 4 °C and a flow rate of 1 mL min⁻¹. After

equilibration of the column with buffer A (20 mL), the protein extract was loaded onto the column. The column was then washed with buffer A (20 mL), and the proteins were eluted using a stepwise gradient with buffer B (buffer A + 500 mM imidazole) to give concentrations of 60, 100, 200, 300 and 500 mM imidazole. Elution of the proteins was monitored by recording the absorbance at 280 nm. Appropriate fractions were analyzed by SDS-PAGE. The fractions containing the recombinant protein were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column (GE Healthcare) into storage buffer (50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10% glycerol, 2 mM DTT). A second purification step was carried out with the Äkta Prime using anion exchange chromatography. For this, a 1 mL HiTrapQ HP column (GE Healthcare) was utilized at a constant flow rate (1 mL min⁻¹). The pre-purified protein sample (3.5 mL) was loaded onto the column after an equilibration step with buffer C (50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10% glycerol, 2 mM DTT). After loading (15 mL), the column was washed with buffer C (5 mL) followed by elution of the proteins using a linear gradient with buffer D (buffer C + 1 M NaCl; 0–1 M NaCl over 25 mL). 2 mL fractions were collected and elution of the proteins was monitored by recording the absorbance at 280 nm. Appropriate fractions were analyzed by SDS-PAGE and fractions containing the recombinant protein were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column (GE Healthcare) into storage buffer (50 mM Tris (pH 7.5), 50 mM NaCl, 1mM DTT, 10% glycerol). Leuorf5 was then flash-frozen in liquid nitrogen and stored at -80°C. The protein concentration was determined using the Bradford assay (Bio-Rad), and the identity of the protein was confirmed by MALDI.

ATP-[³²P]PP_i exchange assay for aminoacyl-AMP formation

Reactions (100µL) for determining substrate specificity contained 75 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 100 mM NaCl, 2mM dATP, 10 µM amino acid substrate and 0.1 µCi [³²P] pyrophosphate, and were carried out at 30 °C. The reactions were initiated by addition of Leuorf5 to a final concentration of 50 nM. Reactions were incubated for 1 min 30 sec and then quenched with charcoal suspension (500 µL of 1.2% [w/v] activated charcoal, 0.1 M tetrasodium pyrophosphate, 0.35 M perchloric acid). The charcoal suspension was pelleted by centrifugation, washed twice with quenching buffer lacking charcoal and then resuspended in 500 µL water and submitted for liquid scintillation counting. Varying concentrations of L-proline (0.5, 1, 5, 10, 50, 100, 500 µM and 1mM) were used to measure the kinetic parameters of L-proline activation. For each substrate (L-proline, D-proline, L- and D-pipecolic acid, L-glutamate and glycine) and each L-proline concentration, assays were performed in triplicate.

For the calculation of the kinetic parameters the ongoing radioactivity for 240 nmol [^{32}P] pyrophosphate, which were applied in each assay, was measured in triplicate. Subsequently the average counts per minute (CPM) per nmol were calculated (the result is equivalent to z). After the average blank (measurements of the assay without amino acid in triplicate) was subtracted from the mean values of the different kinetic data points, the obtained values were divided through z . The obtained data is finally divided through the incubation time of the assay (1.5 min) and subsequently the velocity for each substrate concentration is shown in pmol/min. For the final Michaelis-Menten graph substrate concentration is plotted against velocity (pmol/min) using the SigmaBlot software.

Results

To evaluate the proposed role of Leu-orf5 (A domain) in the formation of the starter unit pyrrolyl-2-carboxylic acid, the gene was expressed in recombinant form and was biochemically characterized using an ATP- ^{32}P PP_i exchange assay ^[144]. The expression construct was cloned into the vector pET28b+ in order to obtain a N-terminally His₆-tagged protein. Expression was carried out in *E. coli* RosettaTM BL21(DE3)pLysS/RARE cells and recombinant Leuorf5 was then purified by nickel affinity chromatography followed by ion exchange chromatography (**Figure 3A**). Despite carrying out the procedure at 4 °C, minor degradation of the purified protein was observed.

Leuorf5 is predicted to behave as a free-standing A domain, that catalyzes two half reactions. First L-proline is activated to L-prolyl-AMP using ATP, and then the activated aminoacyl moiety is transferred to the CP. The first half reaction is typically assayed by amino-acid-dependent exchange of radioactivity from $^{32}\text{PP}_i$ into ATP. We first evaluated the specificity of this reaction using an end-point assay, with the following substrates: L-proline, D-proline, L- and D-pipecolic acid, L-glutamate and glycine. As shown in **Figure 3B**, only proline was activated to a significant extent in this time period (all others were < 2%). Full kinetic characterization further confirmed the catalytic efficiency of Leuorf5 towards L-proline ($k_{\text{cat}}/K_{\text{M}} = 5.9 \times 10^8 \text{ s}^{-1}\text{M}^{-1}$; **Figure 3C**).

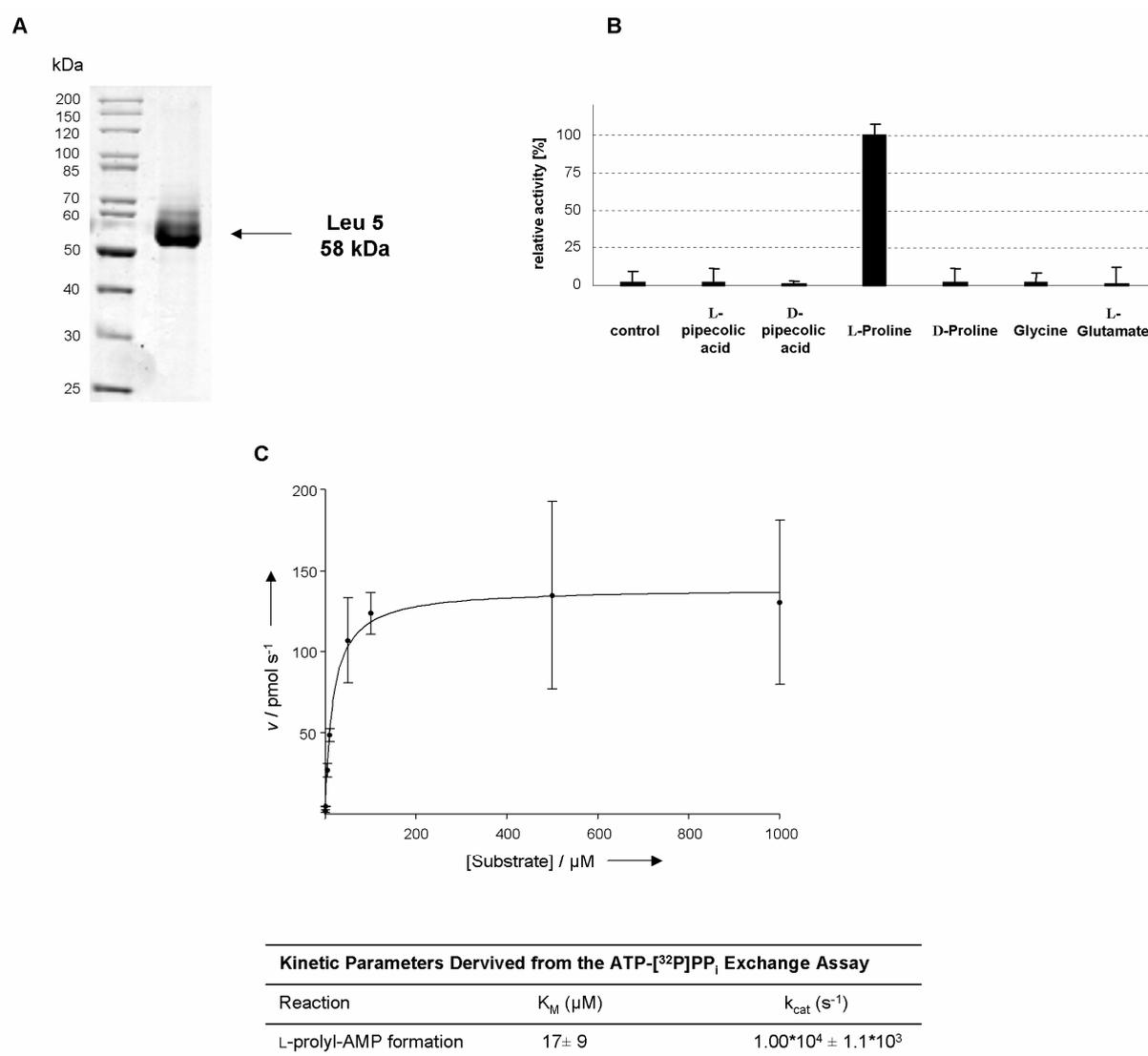


Figure 3. (A): SDS-PAGE analysis of the A domain (Leuorf5) (calculated molecular weight: 58.87 kDa). The left lane contains the molecular weight marker. (B): Relative substrate specificity determined by ATP-³²P]PP_i exchange assay catalysed by Leuorf5. Data were derived from 1.5 min end-point assays. (C): Kinetic characterization of Leuorf5. Shown are [substrate] vs. velocity data fit to the Michaelis-Menten equation, and the determined kinetic parameters.

Discussion

The scaffold of the natural product leupyrrin incorporates a pyrrole moiety which is derived from L-proline, and represents the starter unit of the biosynthesis. For the generation of the pyrrolyl-2-carboxylic acid starter unit a biosynthetic pathway was proposed that involves three distinct enzymes, an A domain, a CP domain and a dehydrogenase. The first enzyme of this pathway is the free-standing A domain (encoded by *leu5*) which was biochemically characterized in this study. Using kinetic analysis, we have shown that Leuorf5 specifically activates L-proline to L-prolyl-AMP by consuming ATP, demonstrating that it catalyzes the first half reaction in the pathway to pyrrolyl-2-carboxyl-*S*-PCP. Furthermore, comparison of the catalytic efficiency (k_{cat}/K_M) of LeuOrf5 to other L-proline activating stand-alone A-

domains (e.g. CloN4 ($1.5 \times 10^6 \text{ s}^{-1}\text{M}^{-1}$) from clorobiocin biosynthesis, CouN4 ($1.2 \times 10^5 \text{ s}^{-1}\text{M}^{-1}$) from the coumermycin A₁ pathway, orf11 ($9.9 \times 10^6 \text{ s}^{-1}\text{M}^{-1}$) from undecylprodigiosin biosynthesis and PltF ($3.9 \times 10^7 \text{ s}^{-1}\text{M}^{-1}$) from the pyoluteorin pathway) showed that Leuorf5 ($5.9 \times 10^8 \text{ s}^{-1}\text{M}^{-1}$) exhibits a significant higher catalytic efficiency than the other A domains towards this substrate ^[140;141]. This effect is largely due to the low K_M value ($17 \pm 9 \text{ }\mu\text{M}$) measured for Leuorf5, as the determined k_{cat} for Leuorf5 ($1.00 \times 10^4 \pm 1.1 \times 10^3 \text{ s}^{-1}$) is within the same range as orf11 ($1.02 \times 10^4 \pm 4.6 \times 10^2 \text{ s}^{-1}$) and PltF ($1.99 \times 10^4 \pm 5.0 \times 10^2 \text{ s}^{-1}$) ^[140;141].

Biochemical studies to elucidate the remaining steps in starter unit biosynthesis were carried out by a diploma student, Katja Gemperlein ^[145], under my guidance. These experiments revealed that the activated amino acid is transferred to the stand-alone ACP (Leu7) and not to the C-PCP bidomain (Leu9), and that the subsequent desaturation reactions are catalyzed by the dehydrogenase homolog Leu7, as predicted. Thus, together, we have elucidated the full pathway to the pyrrole moiety in leupyrrin biosynthesis.

Chapter 7

The origin of the uncommon extender unit in Soce thuggacin biosynthesis

Introduction

The thuggacins were isolated from two myxobacterial species, *Chondromyces crocatus* Cm c5 and *Sorangium cellulosum* So ce895. The most obvious difference between the set of compounds in each strain is the branching functionality at C-2, a methyl in the case of the *C. crocatus* Cm c5 thuggacins versus a hexyl side chain in the case of the *S. cellulosum* So ce895 thuggacins (**Figure 1A**). Identification, analysis and comparison of the corresponding gene clusters, revealed that the final PKS modules of each assembly line are responsible for the structural variance at the C-2 position between Soce and Cmc thuggacins (**Chapter 4**). While the AT domain from the corresponding Cmc module accepts methylmalonyl-CoA as a building block, it was proposed that the corresponding Soce AT recognizes an uncommon extender unit.

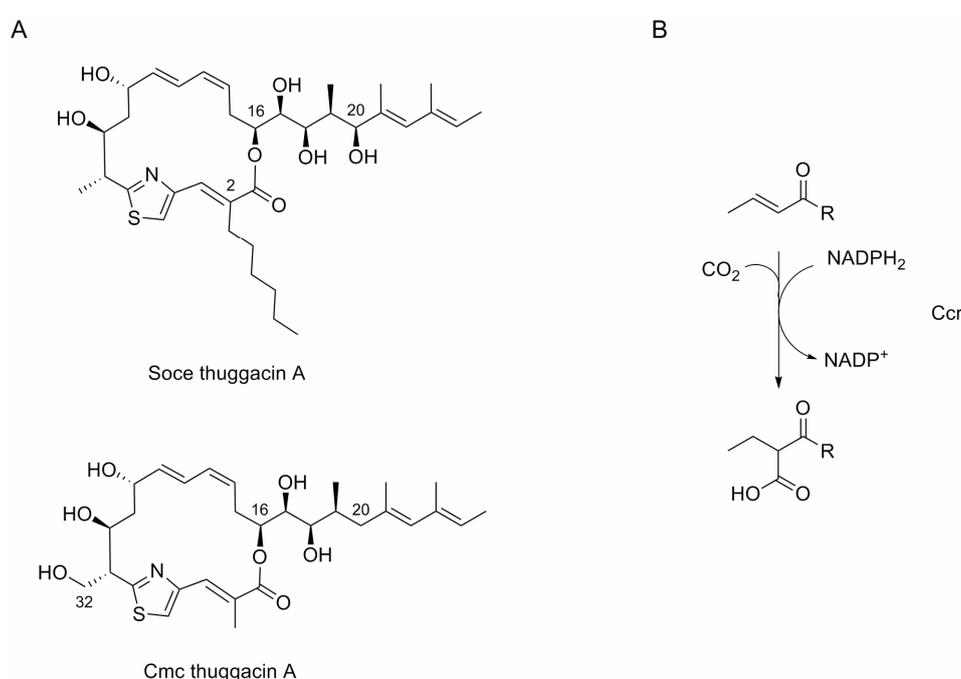


Figure 1. (A): Thuggacins produced by *Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5. (B): Reductive decarboxylation catalyzed by CCR from *Rhodobacter sphaeroides* ^[146].

Based on known pathways for the biosynthesis of other unusual extender units, several possibilities for the source of the extender unit were considered. For example hydroxymalonate, aminomalonate and methoxymalonate are generated in a multi-step reaction from 1,3-bisphosphoglycerate while the intermediates are tethered to an ACP ^[59;61;147]. Subsequently the mature ACP-bound extender is recognized by the respective AT domain. On the other hand, the extender unit ethylmalonyl-CoA has been shown to be derived from a NADPH-dependent reaction catalyzed by a crotonyl-CoA reductase (CCR) ^[148]. These enzymes catalyse reductive carboxylation of unsaturated precursors to saturated products as illustrated for the CCR from *Rhodobacter sphaeroides* (**Figure 1B**) ^[148]. Consequently, we

considered 2-carboxy-octanoyl-CoA and 2-carboxy-octanoyl-ACP as possible candidates for the Soce thuggacin extender units. These building blocks could be derived from octenoyl-CoA or octenoyl-ACP precursors which are intermediates of fatty acid degradation and synthesis, respectively, making them readily available in the cell.

Detailed annotation of the gene cluster revealed one candidate enzyme to form the extender unit, TgaD. The gene *tgaD* is located at one end of the Soce thuggacin gene cluster, and although it shows highest homology to alcohol dehydrogenases (e.g. 46% identity, 61% similarity to BadC from *Streptomyces roseosporus*), it also exhibits convincing similarity to crotonyl-CoA carboxylases/reductase enzymes (CCR) (e.g. 23% identity, 40% similarity to the CCR of *Caulobacter* sp. K31). Consistent with its role in the biosynthesis, inactivation of *tgaD* in *S. cellulosum* So ce895 abolished thuggacin production (**Chapter 4**). However, to provide further evidence for its function as a CCR, and to discriminate between the different precursor substrates, TgaD was obtained in recombinant form from *E. coli* and characterized *in vitro*.

Experimental procedures

Cloning of pSUMOck4TgaD

Gene *tgaD* was PCR amplified from cosmid DN15 (**Chapter 4**) using the primers SUMOck4fwdI (5'-acgcg**GGATCC**atgtacacgactgcggttgct-3') and SUMOck4rev (5'-cgagc**GAATTC**tcacggcgagttgacgcgata-3'). The forward primer contains an artificial *Bam*HI restriction site while the reverse primer was designed to introduce an *Eco*RI site downstream of the stop codon (restriction sites shown in bold and underlined). The PCR product was cloned into pJET1.2 (Fermentas), digested with *Bam*HI and *Eco*RI, and subsequently ligated into the expression vector pSUMOck4 (pSUMO from Lifesensors modified by Carsten Kegler, unpublished), previously digested with *Bam*HI and *Mun*I. The obtained expression construct was designated as pSUMOck4TgaD and verified by sequencing.

Heterologous expression and purification of TgaD

Following several unsuccessful expression experiments with pET28b and pGEX expression constructs and coexpression of various chaperone systems, the expression construct pSUMOck4TgaD co-transformed with the chaperone plasmid pG-KJE8 (TAKARA Bio Inc.) into the expression strain *E. coli* BL21 (DE3)pLysS/RARE (Novagen) resulted in soluble protein expression. Expression was carried out in LB medium (3 × 200 mL) containing kanamycin sulfate (40 µg mL⁻¹) and chloramphenicol (20 µg mL⁻¹) at 37 °C. Chaperone expression was induced at the time of inoculation of the expression culture by adding

arabinose (2 mg mL⁻¹) and tetracycline (5 ng mL⁻¹). Expression of TgaD was then induced at $A_{600} = 0.5\text{--}0.7$ by addition of IPTG to a final concentration of 0.2 mM. After induction, the cells were cultivated at 16 °C over night. Cells were then harvested by centrifugation at 15344 g at 4 °C. The cell pellet was resuspended in buffer A (20 mM Tris (pH 7.8), 200 mM NaCl, 10% glycerol, 10 mM imidazole; 15 mL). The cells were broken by three passes through a French Press (1000 psi), and the insoluble material was sedimented from the lysate by centrifugation at 15344 g at 4 °C. The protein extract was then incubated with nickel sepharose (GE Healthcare; 500 μ L) for 1.5 hours with slow, constant rotation at 4 °C. Subsequently the sepharose was centrifuged (4000 rpm, 15 min; 4 °C), the supernatant was discarded, and then the nickel sepharose was washed twice with 7 mL of buffer B (buffer A + 60 mM imidazole) for 20 min. For elution of the recombinant protein, the sepharose was incubated for 30 min with 2.5 mL buffer C (buffer A + 500 mM imidazole) and then pelleted (4000 rpm, 15 min; 4 °C). The supernatant from the elution step was then analyzed by SDS-PAGE together with the supernatants from the other purification steps. The elution fraction containing the recombinant protein was desalted using a PD-10 column (GE Healthcare) into storage buffer (50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 10% glycerol, 2 mM DTT). To remove the SUMO-tag, the purified protein was incubated with SUMO Protease 2 (1 unit enzyme to 100 μ g protein; LifeSensors) at 30 °C for 1 h, followed by incubation overnight at 4 °C. Removal of the tag was followed by SDS-PAGE. The purified protein was then flash frozen in liquid nitrogen and stored at -80 °C.

Synthesis of octenoyl-CoA and octanoyl-SNAC

Synthesis of octenoyl-CoA and octanoyl-SNAC was performed by Angelika Ullrich and Katharina Schulz.

Purification of octenoyl-CoA

Crude, synthetic octenoyl-CoA was purified using the Waters autopurification system operating in positive ionization mode. Separation was achieved using an XBridgePrep C₁₈ column (Waters; 19 \times 150 mm, 5 μ m particle size, flow 25 mL min⁻¹) with a solvent system consisting of solvent A (20 mM ammonium formate (pH 6.0)) and solvent B (1:1 solvent A: MeOH). The following gradient was applied: 40% B for 4 min, 40–60% B over 4 min, 60–100% B over 0.5 min, 100% B for 3 min, 100–40% B over 0.5 min, and 4 min 40% B. Compounds were detected by diode array and ESI-MS analysis. The purified product was subsequently desalted using a Sephadex column. The structure of the product was verified by NMR (performed by Angelika Ullrich).

Crotonyl-CoA reductase assay

The reaction mixture (60 μ L) contained 100 mM Tris-HCl buffer (pH 7.9), 4 mM NADPH, 33 mM NaHCO₃, 2 μ g purified TgaD and 2 mM substrate. The following compounds were used as substrates: crotonyl-CoA, octenoyl-CoA, octanoyl-CoA, octenoyl-SNAC and octenoic acid. Additionally the cofactors NAD⁺ (1 mM), NADP⁺ (1 mM) and FAD⁺ (200 μ M) with or without *E. coli* oxidoreductase (received from YanYan Li) as appropriate were added to assays with octanoyl-CoA as substrate. The reaction was started by adding the substrate. Negative controls were performed by boiling the reaction mixture for 10 min at 95 °C before the substrate was added. Assays were incubated overnight at 30 °C. Reactions with CoA-esters as substrates were stopped by addition of 1% TCA, while the others were halted by adding 70 μ L MeOH. After centrifugation, the assays were analyzed using HPLC-MS. Standard analysis of all assays was performed on a HPLC-DAD system (Agilent 1100) coupled to an HCTultra ESI-MS ion trap instrument (Brucker Daltonics) operating in positive ionization mode. For assays using CoA-esters as substrates, separation was achieved using a HydroRP column at 35 °C (Phenomenex; 150 \times 2 mm, 4 μ m particle size, flow 0.4 mL min⁻¹), with a solvent system consisting of solvent A (20 mM ammonium formate (pH 6.0)) and solvent B (1:1 solvent A: MeOH). The following gradient was applied: 5–100% solvent B over 20 min. For the other assays a Luna RP-C₁₈ column (Phenomenex; 125 \times 2 mm, 2.5 μ m particle size, flow 0.4 mL min⁻¹) was used, with a solvent system consisting of water and acetonitrile, both containing 0.1% formic acid. The following gradient was applied: 5–95% acetonitrile over 20 min. Compounds were detected in both cases by diode array and ESI-MS analysis. High-resolution measurements were performed on an Accela UPLC-system (Thermo-Fisher) coupled to an LTQ-Orbitrap (linear trap-FT-Orbitrap combination) operating in positive ionization mode. Compounds were separated on a BEH RP-C₁₈ column (Waters; 50 \times 2 mm, 1.7 μ m particle size, flow 0.6 mL min⁻¹), using a solvent system consisting of water and acetonitrile, both containing 0.1% formic acid. The following gradient was applied: 5–95% acetonitrile over 9 min. The UPLC-system was coupled to the Orbitrap by a Triversa Nanomate (Advion), a chip-based nano-ESI interface.

Results

To demonstrate the function of TgaD in the formation of the uncommon extender unit for thuggacin biosynthesis, we biochemically characterized recombinant TgaD. For this, the gene was cloned into vector pSUMOck4 (C. Kegler, unpublished) in order to obtain an N-terminally His₆-tagged protein, and soluble expression was facilitated by coexpression of the

chaperones DnaK, DnaJ, GrpE, GroES and GroEL in *E. coli* BL21(DE3)pLysS/RARE cells [149;150]. The recombinant protein was then purified to homogeneity by affinity chromatography, and the identity of TgaD was confirmed by MALDI-MS analysis. To evaluate whether the SUMO tag had an effect on activity, untagged TgaD was obtained by proteolytic release of the tag using the recommended protease (**Figure 2A**).

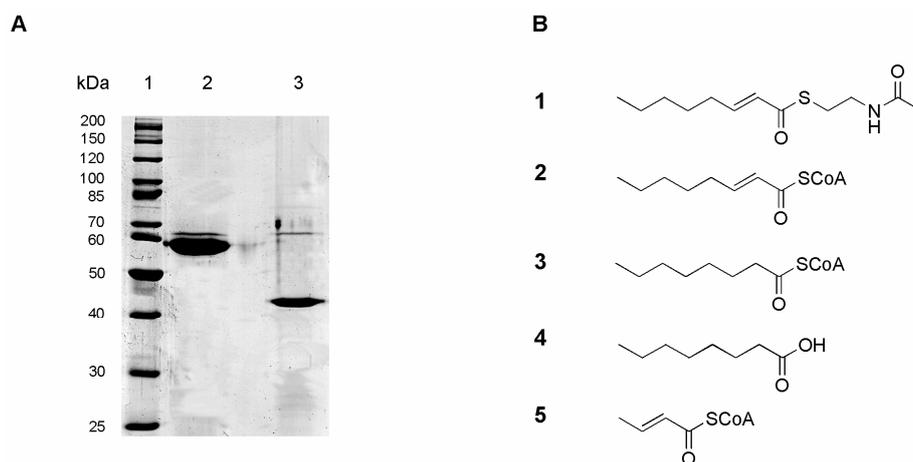


Figure 2. (A): SDS-PAGE analysis of the TgaD. Lane 1 contains molecular weight marker. Lane 2 shows the recombinant protein with attached SUMO tag (calculated molecular weight: 54.72 kDa) and lane 3 contains TgaD after cleavage of the SUMO tag (calculated molecular weight: 42.7kDa). **(B):** Substrates used in the assay: octenoyl-SNAC (1), octenoyl-CoA (2), octanoyl-CoA (3), octanoic acid (4) and crotonyl-CoA (5).

To test the capability of TgaD to catalyze reductive carboxylation, and to identify its natural substrate, several different substrates were tested. First of all octenoyl-CoA and octenoyl-ACP (**Figure 6B**), mimicked in the assay by octenoyl-NAC-thioester, were used in the assay. In addition, also octanoyl-CoA (**Figure 6B**) was utilized as substrate, as it could be not excluded that TgaD, due to its high homology to alcohol dehydrogenases, catalyzes both dehydrogenation and subsequent reductive carboxylation. Further substrates were crotonyl-CoA, which was shown to be the substrate of CCR from *Rhodobacter sphaeroides* [148] and octenoic acid (**Figure 6B**). The recombinant protein with and without attached SUMO tag was incubated with the various substrates in the presence of NADPH and bicarbonate as a source of carbon dioxide, overnight at 30 °C. In assays with octanoyl-CoA as substrate, various other cofactors (NAD^+ , NADP^+ and FAD^+) were tested in combination with a recombinant NADH oxidoreductase from *E. coli* (obtained from Yanyan Li). After quenching the reactions, analysis was carried by HPLC-MS and high resolution LC-MS. In assays with octenoyl-CoA and crotonyl-CoA as substrates, compounds with the masses m/z $[\text{M}+\text{H}]^+ = 938.2$ (retention time (r.t.) = 17.4 min) and m/z $[\text{M}+\text{H}]^+ = 882.1$ (r.t. = 8.8 min) respectively were detected, which were absent in the corresponding negative controls (the same assay

composition but with boiled, inactive enzyme) (**Figure 3A and 3B**). These masses are consistent with the respective carboxylated products, 2-carboxy-octanoyl-CoA and ethylmalonyl-CoA.

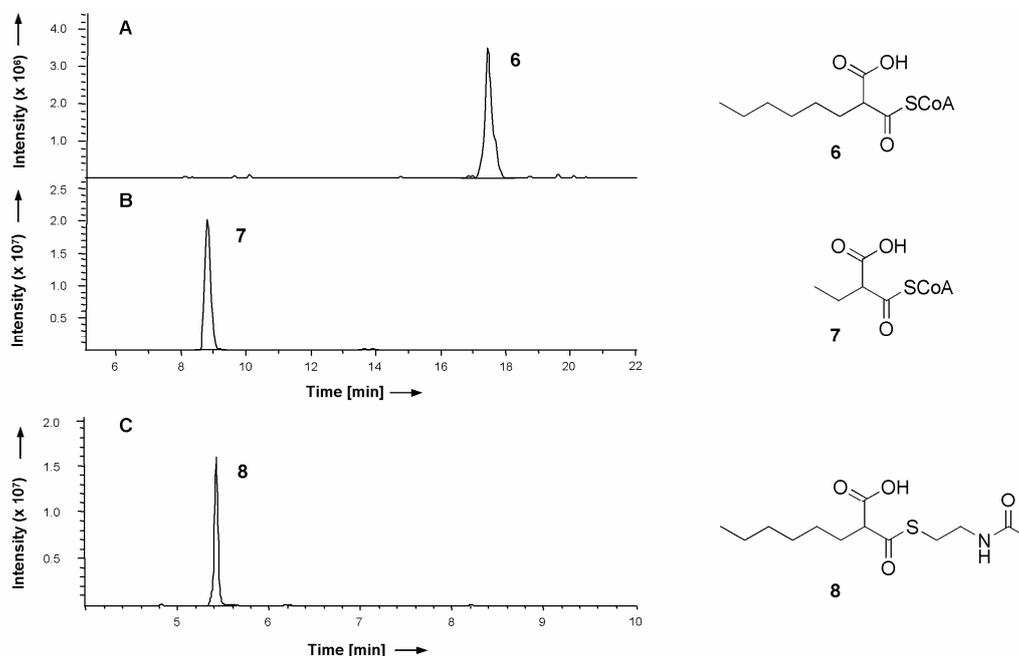


Figure 3. HPLC-MS analysis of different products generated by TgaD. (A): HPLC-MS analysis of an assay with octenoyl-CoA as substrate. The extracted ion chromatogram (EIC) of m/z $[M+H]^+ = 938.0$ is shown. A compound (r.t. = 17.4 min) with the mass m/z $[M+H]^+ = 938.2$ corresponding to 2-carboxy-octanoyl-CoA (6) was detected. (B): HPLC-MS analysis of an assay with crotonyl-CoA as substrate. The EIC of m/z $[M+H]^+ = 882$ is shown. A compound (r.t. = 8.8 min) with the mass m/z $[M+H]^+ = 882.1$ consistent with ethylmalonyl-CoA (7) was detected. (C): High resolution HPLC-MS analysis of an assay with octenoyl-SNAC as substrate. The extracted ion chromatogram (EIC) of m/z $[M+H]^+ = 290.0$ is shown. A compound (r.t. = 5.4 min) with the mass m/z $[M+H]^+ = 290.1$ corresponding to 2-carboxy-octanoyl-SNAC (8) was detected.

Furthermore, in reactions with octenoyl-NAC thioester as substrate, a new compound with the accurate mass m/z $[M+H]^+ = 290.14209$ (r.t. = 5.4 min) was identified by high resolution LC-MS, which was not present in the appropriate negative controls (**Figure 3C**). The predicted sum formula for this compound $C_{13}H_{24}NO_4S$ (calcd. $[M+H]^+ = 290.14206$, $\Delta = 0.1$ ppm) is consistent with the corresponding carboxylated product, 2-carboxy-octanoyl-NAC thioester. In contrast, in assays that were performed with octanoyl-CoA and octenoic acid as substrates, no carboxylated products were observed. Assays carried out with both cleaved and uncleaved TgaD gave identical results (data not shown), demonstrating that the SUMO tag does not influence the catalytic activity of the recombinant protein.

Discussion

Soce thuggacins contain an uncommon hexyl side chain at carbon C-2 that is probably derived from an unusual extender unit. Previous sequence analysis and inactivation experiments (**Chapter 4**) supported a possible role for TgaD in the generation of this building block. Here we present evidence obtained *in vitro* that TgaD catalyzes reductive carboxylation.

As candidate substrates for TgaD, we evaluated octenoyl-CoA and octenoyl-NAC thioester, a simplified mimic of octenoyl-ACP. However, as the sequence analysis (**Chapter 4**) revealed that TgaD may be undergoing an evolutionary transition from a dehydrogenase to a reductase/carboxylase, we considered the possibility that it may also carry out dehydrogenation prior to reductive carboxylation. Thus, we also tested octanoyl-CoA as a possible substrate. Analysis by LC-MS revealed that both octenoyl-CoA and octenoyl-NAC thioester were reductively carboxylated by TgaD, while octanoyl-CoA was not converted to its carboxylated equivalent, even in the presence of various other cofactors (NAD⁺, NADP⁺ and FAD⁺) and an oxidoreductase. Furthermore, TgaD exhibited a clear preference for octenoyl-CoA over octenoyl-NAC thioester, as the minor 2-carboxy-octanoyl-NAC thioester product was only detectable by high-resolution MS. The only structural difference between the two substrates is the CoA moiety which is absent in the NAC thioester and seems to contribute to substrate recognition by TgaD. However, these data alone do not allow us to rule out preferred use of 2-octenoyl-ACP as substrate, as it shares the phosphopantetheine arm of octenoyl-CoA. Consequently also ACP bound octenoyl should be tested as a substrate.

We also explored the tolerance of TgaD to the alternative substrates octenoic acid, and crotonyl-CoA. A carboxylated product was not detected for octenoic acid. The reaction did occur, however, with crotonyl-CoA, to generate the expected ethylmalonyl-CoA. In fact, ethylmalonyl-CoA was produced in apparently higher yields than 2-carboxy-octanoyl-CoA as judged by the respective chromatograms. But if crotonyl-CoA is really a better substrate than octenoyl-CoA, has to be elucidated in extensive kinetic studies with both substrates.

Taken together, these data strongly suggest that TgaD represents a new member of the family of CCR enzymes, which generate uncommon carboxylated extender units for PKS systems. Another member of this family is for example SalG from the salinosporamide synthase in the marine bacterium *Salinispora tropica*, that reductively carboxylates halogenated crotonyl-CoA and propylmalonyl-CoA [151;152]. However, to the best of our knowledge, TgaD is the first enzyme to be reported which is capable of reductively carboxylating a long acyl chain.

Final discussion

1. General summary of this work

The present thesis deals with the identification and characterization of natural product biosynthetic pathways in myxobacteria with special focus on post-assembly line modifications and the biochemical analysis of specific pathway enzymes. Detailed molecular, bioinformatic and analytical studies were carried out in order to elucidate the biosynthesis of the secondary metabolites ajudazol and thuggacin in *Chondromyces crocatus* Cm c5 and thuggacin in *Sorangium cellulosum* So ce895.

We have shown that the TE domain at the end of the ajudazol megasynthase ‘chaperones’ isochromanone ring formation, and is therefore a member of a novel class of TE domains. In addition, the involvement of two P450 enzymes in ajudazol post-assembly line modifications was demonstrated. One of the two P450s catalyzes a desaturation to give the ajudazol A *exo*-methylene, an unusual reaction for this type of enzyme.

Comparative analysis of two thuggacin clusters from different myxobacterial strains provided insights into the evolutionary development of biosynthetic gene clusters, and revealed the mechanistic basis for the structural diversity of the two thuggacin compounds. In the course of this study, evidence was also provided that a crotonyl-CoA-reductase homologue participates in the formation of the hexyl-side chain present in *S. cellulosum* thuggacins. Furthermore the different hydroxylation pattern might be due to the variable action of a FMN-dependent monooxygenase, whose function was proven in the case of *C. crocatus* by gene inactivation. In addition it was shown that the insertion of a strong constitutive promoter in front of the thuggacin gene cluster in *C. crocatus*, significantly increased production of the metabolites.

Finally, the biochemical and catalytic characterization *in vitro* of selected enzymes from leupyrrin and disorazol biosynthesis revealed novel enzymatic transformations that contribute to the large structural variety of myxobacterial natural products.

This section represents a summarized discussion of ajudazol and thuggacin biosynthesis, as the other results were already discussed in detail in **Chapters 5** and **6**.

2. Biosynthesis of the ajudazols in *C. crocatus* Cm c5

The highly antifungal ajudazols are generated by the myxobacterial strain *Chondromyces crocatus* Cm c5, which is known for its potential to produce secondary metabolites with unique structural elements. The ajudazols represent a novel class of natural products as they are new isochromanone derivatives that incorporate an extended side chain containing an oxazole ring, a *Z,Z* diene, and a 3-methoxybutenoic acid amide. Two derivatives of ajudazol have been identified to date: the major metabolite ajudazol A which incorporates an *exo*-methylene functionality at C15, and ajudazol B, which instead has a methyl group at this position. Based on ‘retrobiosynthetic analysis’, we predicted that a mixed PKS/NRPS system would be responsible for ajudazol biosynthesis. The characterization of such hybrid systems is of particular interest, as the two different types of multienzymes have to collaborate with each other to construct the final product.

Indeed, the identification and analysis of the corresponding gene cluster revealed the expected hybrid PKS-NRPS megasynthase. The ajudazol assembly line exhibits a high colinearity between the gene order, module composition and the required biosynthetic steps (**Figure 1**). The only exceptions to this colinearity are the first two modules of the biosynthesis, which are encoded on the last two genes (*ajuK* and *ajuL*) of the biosynthetic gene cluster. Ajudazol biosynthesis starts with a module containing domains for both chain initiation as well as the first round of chain extension. This intermixed loading module/module 1 architecture is frequently found in myxobacterial systems (e.g. chivosazol^[69] and myxalamid^[57]), and is also present in the two thuggacin clusters described in this thesis (**Chapter 4**). The remaining steps in the assembly of ajudazol backbones can be correlated straightforwardly to the modules of the subunits AjuA–AjuH, and indeed the predicted substrate specificity of the corresponding AT and A domains is in complete agreement with the final structure of the ajudazol backbone. Furthermore, the predicted complement of modifying domains is present in all modules with the exception of 3, 5 and 12, in which a dehydratase activity is absent due to a missing or inactive DH domain. It is assumed that these missing activities are complemented by the iterative action of DH domains in downstream modules, which was experimentally proven for epothilone biosynthesis^[153] and afterwards postulated for several other myxobacterial PKS systems^[69;92;154].

The two NRPS modules that incorporate the amino acids glycine and serine respectively both contain modifying domains (*N*-MT and Ox domain) integrated into their A domains (**Figure 1**).

Typically, modifying NRPS domains are located either downstream of the A domain as in the case of the *N*-MT domains or adjacent to the PCP, as for Ox domains [112]. In contrast, in myxobacterial mixed NRPS-PKS systems, these tailoring domains are often found inserted within the primary sequence of A domains. Additional examples of this organization include the *N*-MT domains in the tubulysin pathway [155], and the Ox domain of myxothiazol biosynthesis [56].

After the last condensation step on the ajudazol assembly line, the isochromanone ring is formed and the final product is released. In type I PKS-NRPS systems, type I TEs typically catalyze either hydrolysis to release a free acid or intramolecular cyclization to generate a lactone structure [94]. However, additional release mechanisms have been identified in type I systems. For example, the formation of the macrolactams rifampicin [156] and ansatrienin [157], is proposed to be catalyzed by discrete amide synthases instead of TE domains. Furthermore, C-terminal reductase domains were identified in biosynthetic pathways including the myxochelin assembly line [158-160], which are responsible for reductive release of the mature intermediate, resulting in aldehydes that can be reductively transaminated or further reduced to the alcohol [161;162]. Nonetheless, the generation of aromatic moieties during chain release is rare in modular type I PKS-NRPS, and instead, aromatic rings are typically produced in bacteria by type II or type III PKS systems, with type II systems employing cyclases and aromatases that direct a controlled cyclisation process [66]. Along with the chromone ring of stigmatellin [126] and the benzenoid ring of lasalocid [127], ajudazol represents one of the few examples of bacterial type I PKS-derived aromatic moieties. While the cyclization and aromatization reactions in stigmatellin biosynthesis are thought to be catalyzed by a novel type of cyclase domain located at the end of the assembly line [126], the ajudazol and the lasalocid assembly line terminate in TE domains, suggesting potentially new functions for these enzymes.

Indeed, the present work revealed that the ajudazol TE (AjuTE) represents a novel type of TE involved in the formation of the isochromanone ring. Detailed sequence and phylogenetic analysis demonstrated that AjuTE most closely resembles type II hydrolytic TE domains (**Figure 2**). Based on this *in silico* analysis, we have suggested that AjuTE was once a stand-alone type II TE domain, but was fused genetically to the end of the ajudazol assembly line. Due to this covalent linkage, AjuTE can no longer act as a discrete enzyme and carry out its proof-reading function, which normally consists in removing aberrant acyl groups from the carrier proteins.

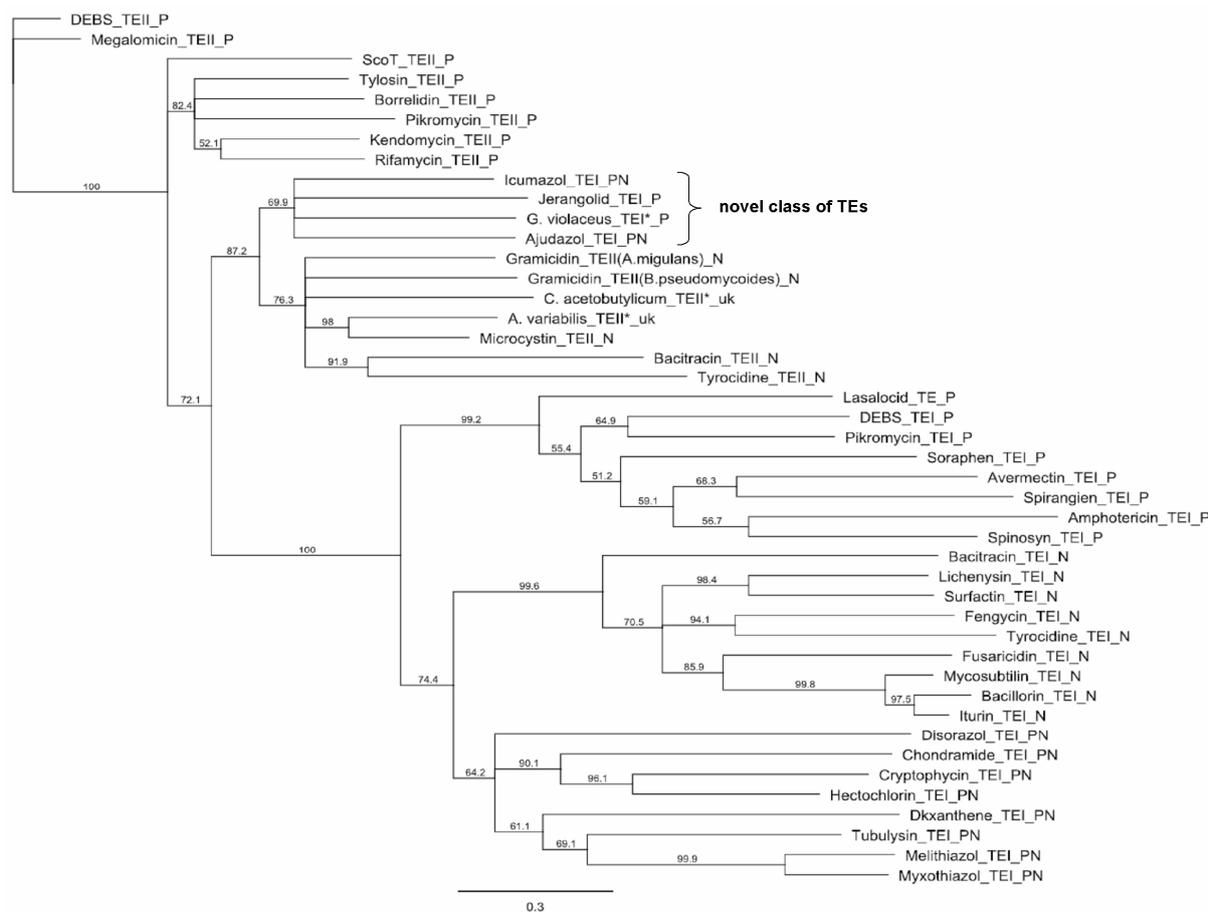


Figure 2. Phylogenetic tree analysis of various type I and type II TEs using neighbourhood joining. Indicated are the type of assembly line (PKS (P), NRPS (N) or hybrid (PN)) and if the TE is a distinct enzyme (TEI) or part of the multienzyme complex (TEI). Uk indicates that the corresponding natural product is unknown. Sequence information of icumazol TEI was kindly provided by N. Luniak (N. Luniak unpublished). The novel class of TE identified in this work is highlighted.

In support of the hypothesis of an original type II TE function, experiments *in vitro* with substrates typical for proof-reading TEIIs showed that AjuTE behaved more similarly to hydrolytic type I and II TEs. Nonetheless, inactivation of AjuTE by mutagenesis of the active site serine led to a reduction of ajudazol production, clearly indicating a role for the domain in formation of the isochromanone when the intermediate is bound to its active site. Thus, despite the fact that AjuTE may have originally functioned as a discrete type II TE, it seems now to play an active role in the cyclization process.

However, the exact mechanism by which the domain catalyses ring formation remains to be elucidated. Initially, the full-length intermediate chain containing all functionalities required for ring formation is transferred to the active site serine of AjuTE. AjuTE could then direct the attack of the C9 hydroxyl group on the acyl ester, so that the macrolactone is generated. The subsequent aromatization of the second ring could then occur spontaneously (**Figure 3A**). Alternatively, the aromatic ring may be generated spontaneously, while the intermediate is still bound to the TE, and in the second step, AjuTE catalyzes ester bond formation between

the C9 hydroxy group and the acyl ester (**Figure 3B**). In the third model, AjuTE simply provides a favourable environment that channels the tethered intermediate into a reactive conformation, allowing spontaneous ring formation and aromatization (**Figure 3C**). The ‘chaperoning’ function of AjuTE in the third mechanism resembles the role of cyclase domains in type II systems ^[66].

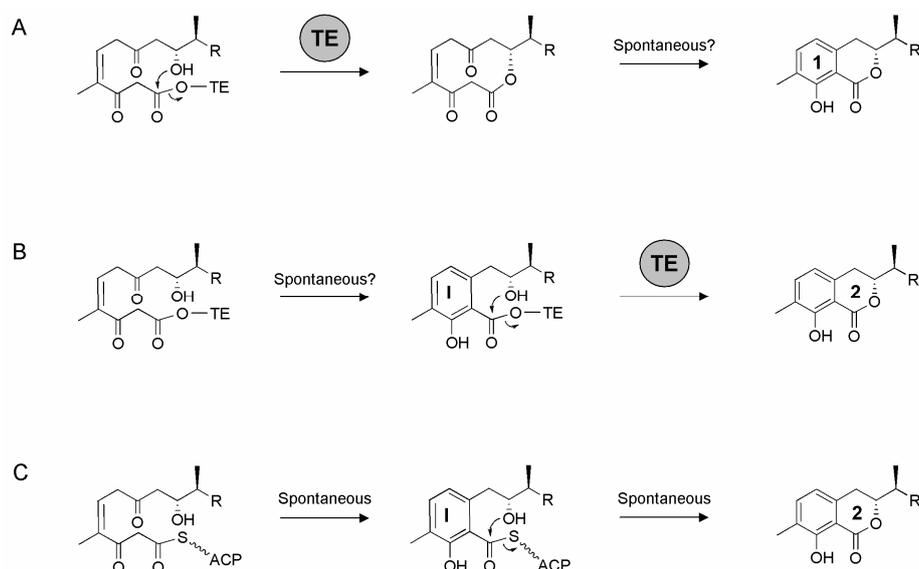


Figure 3. Proposed models for isochromaone ring formation. (A): TE catalyses actively macrolactonization and release from the assembly line. Subsequently ring 1 is formed spontaneously. (B): initially spontaneous formation of ring 1 occurs while the intermediate is still tethered to the TE. In the second step AjuTE catalyses formation of ring 2 and thus release of the final product. (C): AjuTE creates a favorable environment that allows spontaneous ring formation and aromatization.

Taken together AjuTE appears to be a member of a newly-discovered class of TE domains. Its closest homologue, the TE domain from the jerangolid PKS cluster, exhibited a similar catalytic behaviour in the *in vitro* assays. Indeed, in its native context, JerTE would catalyze the formation of a six-membered ring using an internal hydroxyl nucleophile. This reaction is similar to the formation of ring 2 in the ajudazol isochromanone ring ^[163]. Further candidate family members are the TE domain from a cryptic PKS cluster of *Gloeobacter violaceus*, and the TE from icumazol biosynthesis, whose biosynthetic gene cluster was recently discovered (N. Luniak, unpublished data). While the product from the cryptic cluster has still to be identified, it is notable that icumazol also contains an isochromanone ring in its scaffold (**Figure 4**).

The biosynthesis of ajudazol A and B also requires several post-PKS reactions in order to install the hydroxyl moiety at C8 and the C15 *exo*-methylene functionality of ajudazol A. The inactivation of the two P450 enzyme encoding genes *ajuI* and *ajuJ* in *C. crocatus* Cm c5 and

the subsequent analysis of their derivatives led to a proposal for the sequence of post-PKS tailoring reactions (**Figure 5**).

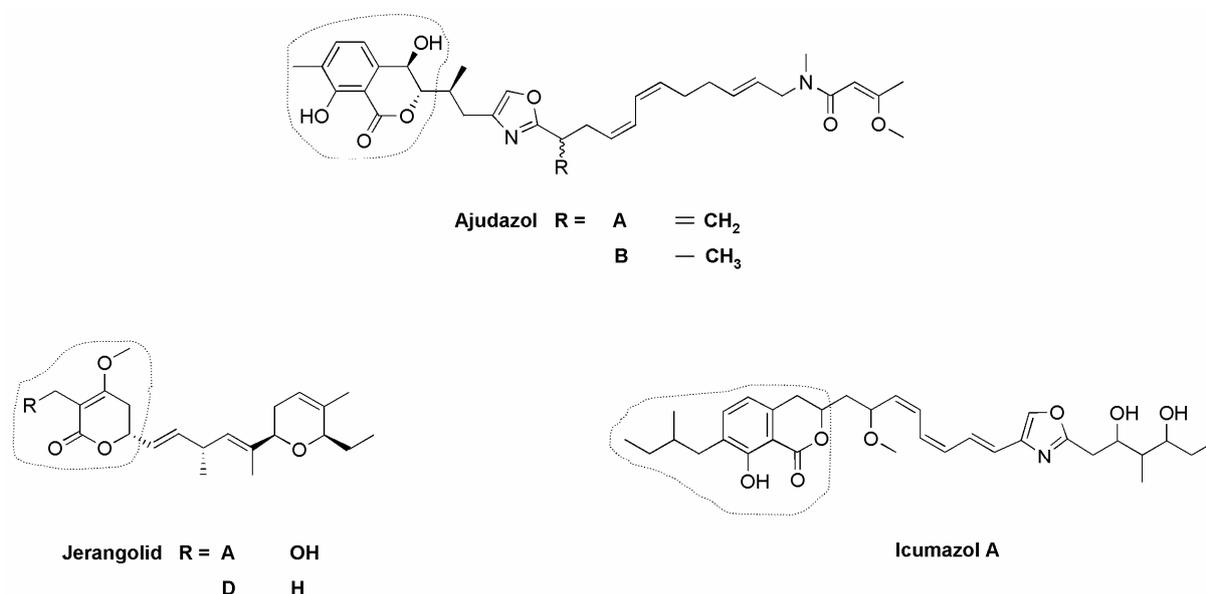


Figure 4. Structures of the natural products ajudazol, jerangolid and icumazol A. The generation of the 6-membered ring in jerangolid and the isochromanone ring in icumazol (according structural moieties are marked) probably occurs by a mechanism similar to the formation of the ajudazol isochromanone, as the TE domains of both assembly lines cluster together with AjuTE in the phylogenetic tree.

The post-assembly modifications commence following release of deshydroxy ajudazol B from the assembly line. Deshydroxy ajudazol B is a substrate for both P450 enzymes, AjuJ and AjuI. If AjuJ acts first, the C8 position is hydroxylated, resulting in the final product ajudazol B. As soon as ajudazol B is generated, it is no longer recognized as a substrate by AjuI. Conversely, if the P450 enzyme AjuI operates first, desaturation of the C15 *exo*-methyl functionality occurs and the intermediate deshydroxy ajudazol A is produced. Deshydroxy ajudazol A is finally converted to ajudazol A by AjuJ, by addition of the hydroxyl group at C8. As the proposed mechanism excludes a conversion of ajudazol B to A and vice versa, this also explains the generation of the two products at a specific ratio in the natural producer *C. crocatus*. While P450-mediated hydroxylation is a common post-PKS tailoring reaction^[164], the desaturation catalyzed by AjuJ is, to our knowledge, the first example of this reaction from bacterial metabolism. To date, P450-catalyzed desaturation has only been reported during flavone biosynthesis in plants^[165], in the biosynthesis of fungal metabolites^[166], and from eukaryotic primary metabolism^[167].

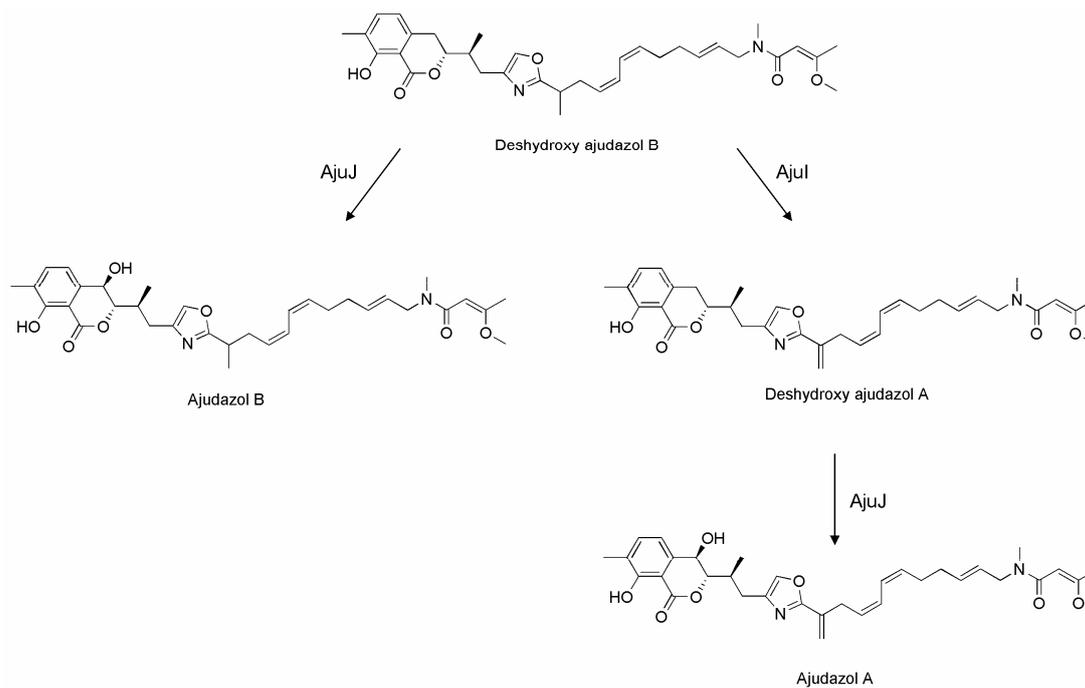


Figure 5. Post-assembly line modifications in ajudazol biosynthesis. The initial product released from the biosynthetic machinery is deshydroxy ajudazol B. Deshydroxy ajudazol B is a substrate for both P450 enzymes AjuJ and AjuI. If AjuJ acts first, ajudazol B is produced, which is no longer recognized as a substrate by AjuI. If AjuI acts first, deshydroxy ajudazol A is produced, which is subsequently hydroxylated by AjuJ to generate ajudazol A.

3. Thuggacin biosynthesis

The thuggacins are macrolide antibiotics that effectively inhibit the bacterial respiratory chain in several Gram-positive bacteria, including clinical isolates of *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) [34;35]. The activity against *M. tuberculosis* is particularly promising, as it represents an alternative mode-of-action relative to currently used TB chemotherapeutics which interact with RNA, DNA and cell wall synthesis. In addition, by targeting energy metabolism, both replicating and non-replicating mycobacteria are affected, which also represents an advantage of thuggacin over other TB drugs. Thuggacins have been isolated from two myxobacterial species, *Chondromyces crocatus* Cm c5 (Cmc-thuggacins) and *Sorangium cellulosum* So ce895 (Soce-thuggacins). The two sets of compounds show some structural differences. While the Soce thuggacins exhibit an uncommon hexyl-side-chain as a C2-branching functionality, the Cmc thuggacins incorporate a methyl group at this position. In addition, the hydroxylation pattern differs between the sets of metabolites: Soce-thuggacins incorporate a hydroxyl moiety at C20 which is absent in the Cmc compounds, while the C32 hydroxyl of Cmc-thuggacins is not present in the Soce counterparts. To elucidate the biosynthetic steps which are responsible for the structural differences between the otherwise identical compounds, the corresponding biosynthetic pathways were identified and analyzed.

Both hybrid PKS-NRPS assembly lines (**Figure 6**), which generate the thuggacin backbone, contain an identical number of modules (loading module + 10 PKS modules + 1NRPS module) and display a high colinearity between the required biosynthetic steps and the complement of modules, which is uncommon for myxobacterial systems [21]. Therefore, it was possible to correlate small differences between the two assembly lines to the structural disparities between the Soce- and Cmc-thuggacins.

The C20 hydroxyl group that is present in Soce-thuggacins but absent in Cmc-thuggacins arises from the divergence in catalytic steps catalyzed by modules 2 and 3 in the respective assembly lines. In the Cmc-cluster, modules 2 (KR and DH domains present) and 3 (KR, DH and ER domains present) contain all required domains that are necessary to generate the observed functionalities at the appropriate positions, i.e. the double bond between C21 and C22 and the full reduction of the β -keto group at C20. In contrast, module 2 of the Soce assembly line, which should also generate a double bond between C21 and C22, is missing the necessary KR domain and only incorporates a DH domain (**Figure 6**). As the DH₂ domain cannot act without previous ketoreduction, it is likely that the intermediate is not processed and therefore the required double bond is not generated at this stage. However, module 3 of

the Soce cluster integrates a redundant DH domain, as only a KR domain is required to produce the observed hydroxyl moiety. Therefore, we hypothesize that the DH of module 3 acts on the chain extension functionality generated by module 2. Thus, initially the unreduced intermediate generated by module 2 could be directly transferred to the ACP₃ of module 3 without undergoing chain extension, as postulated for leinamycin biosynthesis^[154]. Both KR₃ and DH₃ would then act on the intermediate to produce the required *trans* double bond. Afterwards, the intermediate would be passed to the upstream KS₃ domain for chain extension. The obtained β -keto group of the tetraketide would then only be modified by the KR₃ to yield the C20 hydroxyl moiety of Soce-thuggacin. However, the mechanism by which the DH₃ can distinguish between the two different intermediates and dehydrate only at the diketide stage remains unclear. Alternatively, both domains KR₃ and DH₃ are used in the third round of chain extension and the obtained double bond is rehydrated at a later stage during the biosynthesis. The need for domains acting out of sequence has been observed for several other myxobacterial systems^[92;154;168], but this is to our knowledge the first example where two domains show this behavior.

The second structural difference in the backbone structure could be correlated to the final PKS module of each assembly line, which seems to be responsible for the alternative branching at C2. *In silico* analysis of the respective AT domains revealed that AT₁₁ from the Cmc module accepts methylmalonyl-CoA as a building block, which is consistent with the methyl branching at C2. In contrast, no prediction could be made for Soce AT₁₁, and it was assumed that this AT may recognize an uncommon extender unit. Based on known pathways for the biosynthesis of other unusual extender units, several possibilities for the origin of this extender unit were considered. It was shown for hydroxymalonate, aminomalonate and methoxymalonate that these extender units are delivered as ACP-bound species to the appropriate AT domain^[59;61;147]. Subsequently the AT domain recognizes the extender unit and transferred it to a CP in the assembly line. On the other hand, the extender unit ethylmalonyl-CoA has been shown to be derived from crotonyl-CoA. This NADPH-dependent reaction is catalysed by a crotonyl-CoA-carboxylase (CCR), which reductively carboxylates crotonyl-CoA to ethylmalonyl-CoA^[148].

The reductive carboxylation of unsaturated precursors to saturated products was initially shown for the CCR from *Rhodobacter sphaeroides*^[146;148] and this novel class of carboxylases was recently expanded by the identification and characterization of the CCR SalG from *Salinispora tropica*^[151;152]. In addition, most of the ‘normal’ extender units are CoA tethered, which would also argue for a CoA derived extender unit.

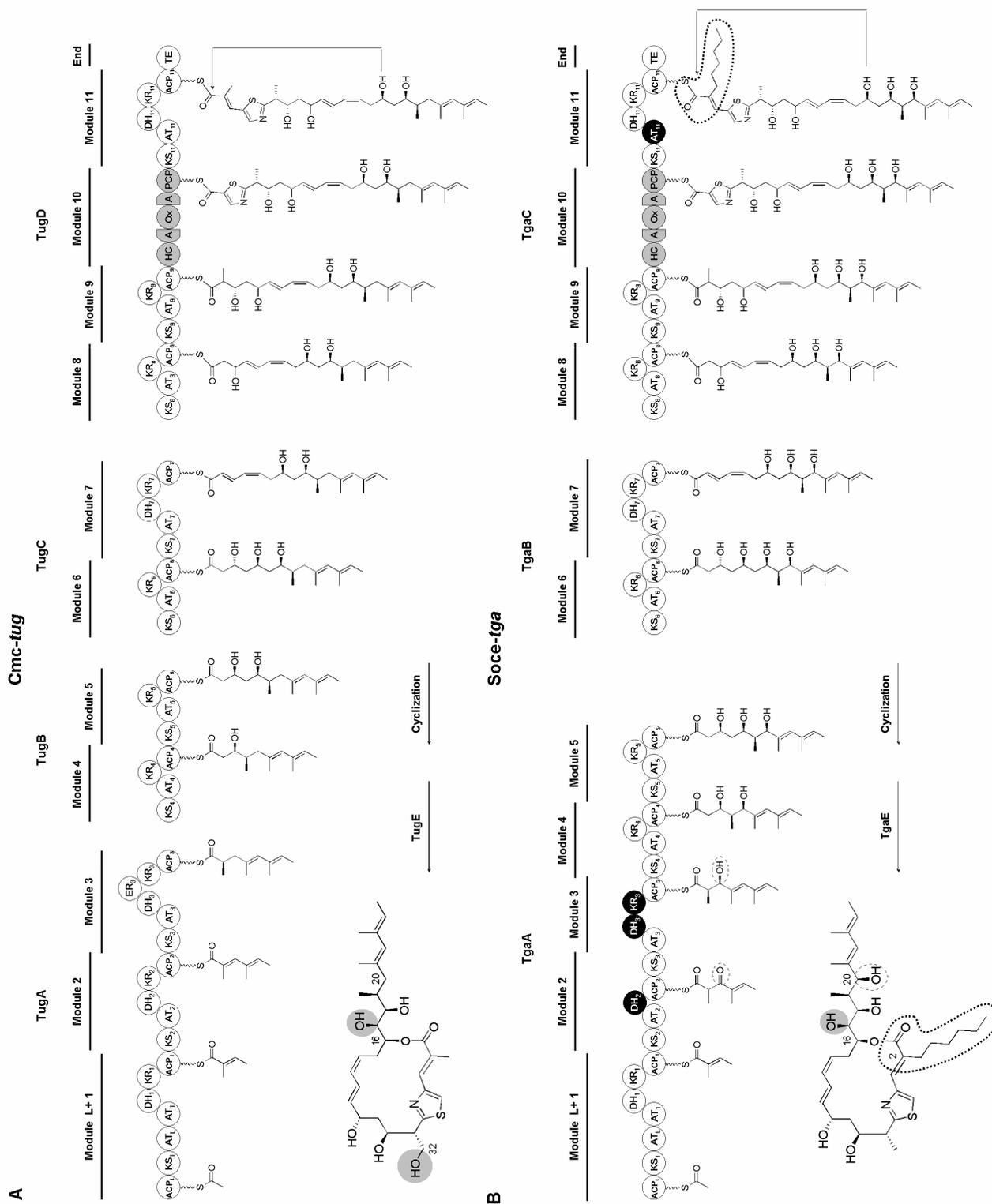


Figure 6. Biosynthetic models for thuggacin production in *C. crocatus* Cm c5 (A) and in *S. cellulose* So ce895 (B). The differences in the assembly lines are highlighted.

Considering both possibilities, 2-carboxy-octanoyl-CoA and 2-carboxy-octanoyl-ACP were suggested as likely candidates for the uncommon Soce-extender unit. These extender units could be generated from either octenoyl-CoA or octenoyl-ACP precursors, which are intermediates of fatty acid degradation or synthesis, and so are continuously available in the cell. Analysis of the Soce gene cluster revealed one candidate enzyme which might be involved in the formation of the extender unit, TgaD. The gene *tgaD* is located upstream of the Soce biosynthetic gene cluster and although it shows highest homology to alcohol dehydrogenases, it also exhibits similarity to CCRs. Inactivation of *tgaD* in *S. cellulosum* So ce895 led to abolishment of thuggacin production, providing the first evidence that TgaD catalyzes reductive carboxylation to yield the uncommon extender unit. To confirm its role, and to attempt to discriminate between the different precursor substrates, TgaD was expressed heterologously as a N-terminal His-tagged protein and characterized *in vitro*. The ability of TgaD to catalyse reductive carboxylation was tested with several different substrates. But only in case of octenoyl-CoA, octenoyl-SNAC and crotonyl-CoA the respective carboxylated products were observed (**Chapter 7**). Furthermore TgaD favors octenoyl-CoA over octenoyl-NAC thioester, whereas the latter substrate mimics 2-octenoyl-ACP. As these two substrates only differ in their CoA moiety this preference does not absolutely exclude 2-octenoyl-ACP as favoured substrate as it shares the phosphopantetheine arm of octenoyl-CoA. Taken together, the obtained data clearly shows that TgaD catalyses reductive decarboxylation and therefore represents a new member of the family of CCR enzymes, which generate uncommon carboxylated extender units for PKS systems.

The last structural difference between the thuggacins is the hydroxyl group at C32 which is present in the Cmc-thuggacins but absent in the Soce counterparts. This hydroxyl moiety is added along with the shared hydroxyl group at C17, following release of the intermediates from the biosynthetic machinery. Despite the fact that different numbers of hydroxylations are observed in the two pathways, only a single candidate for catalyzing the required reactions was discovered in both gene clusters. *TugE* (Cmc) and *tgaE* (Soce) respectively are located upstream of their PKS-NRPS megasynthases, show high mutual sequence homology, and share a closest homologue in the public database, the FMN-dependent alkanal monooxygenase of *Parvibaculum lavamentivorans*. Inactivation of *tugE* in *C. crocatus* abolished Cmc-thuggacin A and C production. In place of these metabolites, we instead identified di-deshydroxy thuggacin (**1**), lacking both hydroxy groups at C17 and C32, and a mono-deshydroxy thuggacin. Considering the possibility that these two new derivatives are intermediates on the pathway to the mature thuggacins, we reanalysed the wild type extract.

Indeed we were able to detect both compounds in the wild type and yet a second mono-deshydroxy thuggacin. Structure elucidation of di-deshydroxy thuggacin based on high resolution fragmentation pattern unambiguously showed the absence of the hydroxyl groups at C17 and C32. However for the two mono-deshydroxy thuggacin a definitive assignment of the site of monohydroxylation was despite different fragmentation patterns of the derivatives not possible.

The performed experiment clearly shows that Cmc TugE catalyses hydroxylation of a precursor thuggacin in *C. crocatus* and as Cmc TugE and Soce TgaA share more than 80% sequence identity and sequence analysis of Soce TgaA revealed no obvious evidence for inactivity we predict that Soce TgaA is involved in post-assembly line modification of Soce-thuggacins. As Soce TgaA can be only responsible for the addition of the hydroxyl group at C17 we subsequently assume that Cmc TugE also catalyses C17 hydroxylation.

By considering the post-assembly line modifications in *S. cellulosum* thuggacin biosynthesis in combination with the obtained data allows us to propose two possible models for the hydroxylations of Cmc-thuggacins. In the first model (**Figure 7B**) Cmc TugE carries out both hydroxylations. Initially it hydroxylates the position at C17 resulting in C17 mono-hydroxylated thuggacin (**2**) and in the second step it adds the hydroxyl group to C32 to yield the mature thuggacins. The presence of minor amounts of C32 mono-hydroxylated thuggacin (**3**) in the mutant and in the wild type are probably due to an additional, promiscuous oxygenase encoded elsewhere in the genome. In the alternative model (**Figure 7C**) the hydroxylations are catalysed by two different enzymes. First Cmc TugE carries out C17 hydroxylation and the subsequent C32 modification is accomplished by a second enzyme located at a different place in the genome. However in this model we assume that the second enzyme has a certain activity towards (**1**), which results in the generation of (**3**) in mutant and wild type. Although flavin-dependent monooxygenases are typically regiospecific^[169] the first model requires a relaxed regiospecificity of Cmc TugE which is more reminiscent of P450 hydroxylases from secondary metabolism^[170].

However, it remains unclear in the first case why the Cmc TugE homologue Soce TgaE only hydroxylates at C17, and does not also act at C32. One reason could be that although Soce TgaE and Cmc TugE are highly similar, minor differences within the active sites may only permit C17 hydroxylation in Soce-thuggacins. Alternatively the unique functionalities of the Soce-thuggacin, i.e. the hexyl-side chain and the C20 hydroxyl moiety, may only allow the substrate to adopt an orientation within the Soce TgaE active site which renders C17 accessible to the catalytic machinery. If the C32 hydroxylation is instead carried out by a

second enzyme the lack of modification may be due to the absence of this enzyme in *S. cellulosum*.

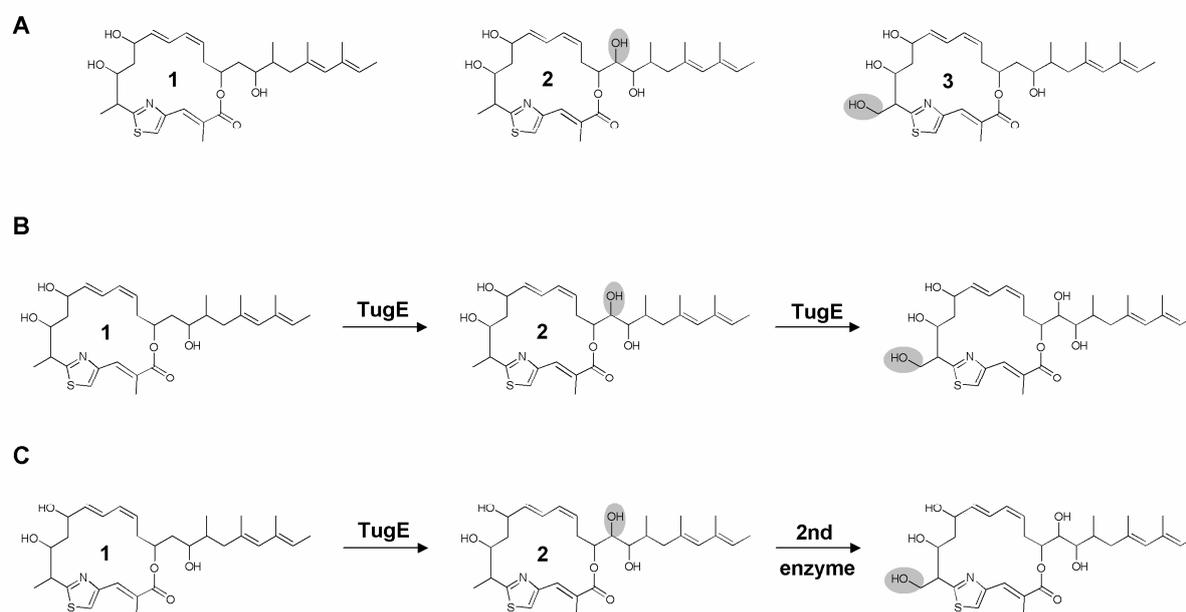


Figure 7. (A) Identified Cmc-thuggacin derivatives: all three compounds are produced by *C. crocatus* wild type, *tugE* mutant generates only compound (1) and one of the two mono-deshydroxy thuggacins. (B) In the first hydroxylation model *TugE* acts twice. First it adds the hydroxyl group to C17 and subsequently to C32. (C) In the second model *TugE* just catalyze C17 hydroxylation, while the C32 modification is carried out by a second enzyme.

Finally, comparative phylogenetic analysis of the two biosynthetic machineries allowed us to develop two models for the evolution of the thuggacin clusters. The phylogenetic studies were carried out with KS domains, as these are the most conserved domains within PKS systems and in general do not participate in recombinatorial exchange^[171]. Therefore they are the domains best suited to trace the evolution of a PKS system^[172]. As the thuggacin KS domains show highest homology to *Streptomyces* KSs – the first time this observation has been made for any myxobacterial biosynthetic pathway – it appears that a progenitor thuggacin gene cluster may have been acquired from a Streptomycete by horizontal gene transfer. In the first model a smaller ancestral thuggacin cluster present in both strains may have been differently expanded by module duplication to generate the present-day assembly lines. In this case the two thuggacins clusters represent good candidates for convergent evolution, as they appear to have independently developed towards a similar multienzyme organization. Alternatively the same full length progenitor gene cluster could have been acquired by both strains or their ancestors and subsequent gene conversion events coupled with deletion and exchange of single domains lead to the development of the present-day clusters.

Reference List

1. Demain AL: **From natural products discovery to commercialization: a success story.** *J Ind Microbiol Biotechnol* 2006, **33**:486–495.
2. Newman DJ, Cragg GM: **Natural products as sources of new drugs over the last 25 years.** *J Nat Prod* 2007, **70**:461–477.
3. Demain AL: **Pharmaceutically active secondary metabolites of microorganisms.** *Appl Microbiol Biotechnol* 1999, **52**:455–463.
4. Fischbach MA, Walsh CT: **Antibiotics for emerging pathogens.** *Science* 2009, **325**:1089–1093.
5. Nathan C: **Antibiotics at the crossroads.** *Nature* 2004, **431**:899–902.
6. Harada K: **Production of secondary metabolites by freshwater cyanobacteria.** *Chem Pharm Bull (Tokyo)* 2004, **52**:889–899.
7. Behal V: **Alternative sources of biologically active substances.** *Folia Microbiologica* 2003, **48**:563–571.
8. Gulder TAM, Moore BS: **Chasing the treasures of the sea - bacterial marine natural products.** *Curr Opin Microbiol* 2009, **12**:252–260.
9. Garcia RO, Krug D, Müller R: **Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods.** *Methods Enzymol* 2009, **458**:59–91.
10. Dawid W: **Biology and global distribution of myxobacteria in soils.** *FEMS Microbiol Rev* 2000, **24**:403–427.
11. Iizuka T, Jojima Y, Fudou R, Yamanaka S: **Isolation of myxobacteria from the marine environment.** *FEMS Microbiol Lett* 1998, **169**:317–322.
12. Reichenbach H: **Order VIII. Myxococcales. Tchan, Pochon and Pre'vot 1948, 398AL.** In *Bergey's manual of systematic bacteriology*, edn 2nd, vol. 2, part C. Edited by Brenner DJ, Krieg NR, Staley JT. Springer; 2005:1059–1144.
13. Garcia RO, Reichenbach H, Ring MW, Müller R: ***Phaselicystis flava* gen. nov., sp. nov., an arachidonic acid-containing soil myxobacterium, and the description of *Phaselicystidaceae* fam. nov.** *I J Syst Evol Microbiol* 2008, **59**:1524–1530.
14. Shimkets L, Dworkin M, Reichenbach H: **The Myxobacteria.** In *The Prokaryotes*, edn 3. Edited by Dworkin M. Berlin: Springer; 2006:31–115.
15. Reichenbach H: **Myxobacteria, producers of novel bioactive substances.** *J Ind Microbiol Biotechnol* 2001, **27**:149–156.
16. Gerth K, Pradella S, Perlova O, Beyer S, Müller R: **Myxobacteria: proficient producers of novel natural products with various biological activities - past and future biotechnological aspects with the focus on the genus *Sorangium*.** *J Biotechnol* 2003, **106**:233–253.
17. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, Bartels D, Bekel T, Beyer S, Bode E, Bode HB, Bolten CJ, Choudhuri JV, Doss S, Elnakady YA, Frank B, Gaigalat L, Goesmann A, Groeger C, Gross F, Jelsbak L, Jelsbak L, Kalinowski J, Kegler C, Knauber T, Konietzny S, Kopp M, Krause L, Krug D, Linke B, Mahmud T, Martinez-Arias R, McHardy AC, Merai M, Meyer F, Mormann S, Munoz-Dorado J, Perez J, Pradella S, Rachid S, Raddatz G, Rosenau F, Rückert C, Sasse F, Scharfe M, Schuster SC, Suen G, Treuner-Lange A, Velicer GJ, Vorhölter FJ, Weissman KJ, Welch RD, Wenzel SC, Whitworth DE, Wilhelm S, Wittmann C, Blöcker H, Pühler A, Müller R: **Complete genome sequence of the myxobacterium *Sorangium cellulosum*.** *Nat Biotechnol* 2007, **25**:1281–1289.

-
18. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen J, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB: **Evolution of sensory complexity recorded in a myxobacterial genome.** *P Natl Acad Sci USA* 2006, **103**:15200–15205.
 19. Wenzel SC, Müller R: **The impact of genomics on the exploitation of the myxobacterial secondary metabolome.** *Nat Prod Rep* 2009, **26**:1385–1407.
 20. Weissman KJ, Müller R: **A brief tour of myxobacterial secondary metabolism.** *Bioorg Med Chem* 2009, **17**:2121–2136.
 21. Wenzel SC, Müller R: **Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry.** *Nat Prod Rep* 2007, **24**:1211–1224.
 22. Bode HB, Müller R: **Analysis of myxobacterial secondary metabolism goes molecular.** *J Ind Microbiol Biotechnol* 2006, **33**:577–588.
 23. Goodin S, Kane MP, Rubin EH: **Epothilones: Mechanism of action and biologic activity.** *J Clin Oncol* 2004, **22**:2015–2025.
 24. Bollag DM, McQueney PA, Zhu J, Hensens O, Koupal L, Liesch J, Goetz M, Lazarides E, Woods M: **Epothilones, a new class of microtubule-stabilizing agents with taxol-like mechanism of action.** *Cancer Res* 1995, **55**:2325–2333.
 25. Reichenbach H, Höfle G: **Discovery and development of the epothilones - A novel class of antineoplastic drugs.** *Drugs in R&D* 2008, **9**:1–10.
 26. Khalil MW, Sasse F, Lünsdorf H, Elnakady YA, Reichenbach H: **Mechanism of action of tubulysin, an antimetabolic peptide from myxobacteria.** *ChemBioChem* 2006, **7**:678–683.
 27. Elnakady YA, Sasse F, Lünsdorf H, Reichenbach H: **Disorazol A1, a highly effective antimetabolic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells.** *Biochem Pharmacol* 2004, **67**:927–935.
 28. Hagelueken G, Albrecht SC, Steinmetz H, Jansen R, Heinz DW, Kalesse M, Schubert WD: **The absolute configuration of rhizopodin and its inhibition of actin polymerization by dimerization.** *Angew Chem Int Ed Engl* 2009, **48**:595–598.
 - 28b. Günther E.G., Paulini K.W., Alcher B., Baasner S., Schmitt P., Teifel M., Schäfer O., Grundker C., Fister S: **Targeting disorazol Z to LHRH-receptor positive tumors by the cytotoxic conjugate AEZS-125.** Poster at the experimental biology conference in San Diego 2008.
 29. Reichenbach H, Höfle G: **Discovery of a new antifungal mechanism of action: soraphen - an almost-success story.** In *Scientific Annual Report*. Braunschweig: Gesellschaft für Biotechnologische Forschung mbH; 1994:5–22.
 30. Pridzun L, Sasse F, Reichenbach H: **Inhibition of fungal acetyl-CoA carboxylase: A novel target discovered with the myxobacterial compound soraphen.** In *Antifungal agents*. Edited by Dixon GK, L.G. C, Hollomon DW. Oxford, UK: BIOS Scientific Publishers Ltd; 1995:99–109.
 31. Vahlensieck HF, Pridzun L, Reichenbach H, Hinnen A: **Identification of the yeast ACC1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A.** *Curr Genet* 1994, **25**:95–100.
 32. Beckers A, Organe S, Tinunermans L, Scheys K, Peeters A, Brusselmans K, Verhoeven G, Swinnen JV: **Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells.** *Cancer Res* 2007, **67**:8180–8187.
 33. Shen Y, Volrath SL, Weatherly SC, Elich TD, Tong L: **A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme a carboxylase by soraphen A, a macrocyclic polyketide natural product.** *Mol Cell* 2004, **16**:881–891.
-

-
34. Irschik H, Reichenbach H, Höfle G, Jansen R: **The thuggacins, novel antibacterial macrolides from *Sorangium cellulosum* acting against selected gram-positive bacteria.** *J Antibiot* 2007, **60**:733–738.
 35. Steinmetz H, Irschik H, Kunze B, Reichenbach H, Höfle G, Jansen R: **Thuggacins, macrolide antibiotics active against *Mycobacterium tuberculosis*: Isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis.** *Chemistry* 2007, **13**:5822–5832.
 36. Xie Z, Siddiqi N, Rubin EJ: **Differential antibiotic susceptibilities of starved *Mycobacterium tuberculosis* isolates.** *Antimicrob Agents Chemother* 2005, **49**:4778–4780.
 37. Gerth K, Irschik H, Reichenbach H, Trowitzsch W: **Myxothiazol, an antibiotic from *Myxococcus fulvus* (myxobacterales). I. Cultivation, isolation, physico-chemical and biological properties.** *J Antibiot* 1980, **33**:1474–1479.
 38. Sasse F, Böhlendorf B, Herrmann M, Kunze B, Forche E, Steinmetz H, Höfle G, Reichenbach H, Hermann M: **Melithiazols, new beta-methoxyacrylate inhibitors of the respiratory chain isolated from myxobacteria. Production, isolation, physico-chemical and biological properties.** *J Antibiot* 1999, **52**:721–729.
 39. Irschik H, Jansen R, Höfle G, Gerth K, Reichenbach H: **The coralopyronins, new inhibitors of bacterial RNA synthesis from myxobacteria.** *J Antibiot* 1985, **38**:145–152.
 40. Irschik H, Gerth K, Höfle G, Kohl W, Reichenbach H: **The myxopyronins, new inhibitors of bacterial RNA synthesis from *Myxococcus fulvus* (Myxobacterales).** *J Antibiot* 1983, **36**:1651–1658.
 41. Irschik H, Jansen R, Gerth K, Höfle G, Reichenbach H: **The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria.** *J Antibiot* 1987, **40**:7–13.
 42. Irschik H, Reichenbach H: **The mechanism of action of myxovalargin A, a peptide antibiotic from *Myxococcus fulvus*.** *J Antibiot* 1985, **38**:1237–1245.
 43. Wenzel SC, Müller R: **Myxobacteria--'microbial factories' for the production of bioactive secondary metabolites.** *Mol Biosyst* 2009, **5**:567–574.
 44. Wenzel SC, Müller R: **The biosynthetic potential of myxobacteria and their impact on drug discovery.** *Curr Opin Drug Discov Devel* 2009, **12**:220–230.
 45. Walsh CT: **Polyketide and nonribosomal peptide antibiotics: modularity and versatility.** *Science* 2004, **303**:1805–1810.
 46. Finking R, Marahiel MA: **Biosynthesis of nonribosomal peptides.** *Annual Review of Microbiology* 2004, **58**:453–488.
 47. Weissman KJ: **Polyketide synthases: mechanisms and models.** *Ernst Schering Res Found Workshop* 2005, 43–78.
 48. Schwarzer D, Finking R, Marahiel MA: **Nonribosomal peptides: From genes to products.** *Nat Prod Rep* 2003, **20**:275–287.
 49. Stratigopoulos G, Bate N, Cundliffe E: **Positive control of tylosin biosynthesis: pivotal role of TyIR.** *Mol Microbiol* 2004, **54**:1326–1334.
 50. Staunton J, Weissman KJ: **Polyketide biosynthesis: a millennium review.** *Nat Prod Rep* 2001, **18**:380–416.
 51. Schwarzer D, Marahiel MA: **Multimodular biocatalysts for natural product assembly.** *Naturwissenschaften* 2001, **88**:93–101.
 52. Stein T, Vater J, Kruft V, Otto A, Wittmann-Liebold B, Franke P, Panico M, McDowell R, Morris HR: **The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates.** *J Biol Chem* 1996, **271**:15428–15435.
-

-
53. Fischbach MA, Walsh CT: **Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms.** *Chem Rev* 2006, **106**:3468–3496.
 54. Ligon J, Hill S, Beck J, Zirkle R, Molnar I, Zawodny J, Money S, Schupp T: **Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26.** *Gene* 2002, **285**:257–267.
 55. Ikeda H, Nonomiya T, Omura S: **Organization of biosynthetic gene cluster for avermectin in *Streptomyces avermitilis*: analysis of enzymatic domains in four polyketide synthases.** *J Ind Microbiol Biotechnol* 2001, **27**:170–176.
 56. Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, Brandt P, Blöcker H, Höfle G, Beyer S, Müller R: **New lessons for combinatorial biosynthesis from myxobacteria: the myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1.** *J Biol Chem* 1999, **274**:37391–37399.
 57. Silakowski B, Nordsiek G, Kunze B, Blöcker H, Müller R: **Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: The myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15.** *J Biol Chem* 2001, **8**:59–69.
 58. Reeves CD, Chung LM, Liu Y, Xue Q, Carney JR, Revill WP, Katz L: **A new substrate specificity for acyl transferase domains of the ascomycin polyketide synthase of *Streptomyces hygroscopicus*.** *J Biol Chem* 2002, **277**:9155–9159.
 59. Chan YA, Boyne MT, Podevels AM, Klimowicz AK, Handelsman J, Kelleher NL, Thomas MG: **Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units.** *P Natl Acad Sci USA* 2006, **103**:14349–14354.
 60. Carroll BJ, Moss SJ, Bai LQ, Kato Y, Toelzer S, Yu TW, Floss HG: **Identification of a set of genes involved in the formation of the substrate for the incorporation of the unusual "glycolate" chain extension unit in ansamitocin biosynthesis.** *J Am Chem Soc* 2002, **124**:4176–4177.
 61. Wenzel SC, Williamson RM, Grünanger C, Xu J, Gerth K, Martinez RA, Moss SJ, Carroll BJ, Grond S, Unkefer CJ, Müller R, Floss HG: **On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of "glycolate" units into ansamitocin and soraphen A.** *J Am Chem Soc* 2006, **128**:14325–14336.
 62. Del Vecchio F, Petkovic H, Kendrew SG, Low L, Wilkinson B, Lill R, Cortes J, Rudd BA, Staunton J, Leadley PF: **Active-site residue, domain and module swaps in modular polyketide synthases.** *J Ind Microbiol Biotechnol* 2003, **30**:489–494.
 63. Haydock SF, Aparicio JF, Molnar I, Schwecke T, König A, Marsden AFA, Galloway IS, Staunton J, Leadley PF: **Divergent structural motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acylcarrier protein transacylase domains in the modular polyketide synthases.** *FEBS Lett* 1995, **374**:246–248.
 64. Yadav G, Gokhale RS, Mohanty D: **Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases.** *J Mol Biol* 2003, **328**:335–363.
 65. Molnar I, Schupp T, Ono M, Zirkle R, Milnamow M, Nowak-Thompson B, Engel N, Toupet C, Stratmann A, Cyr DD, Gorlach J, Mayo JM, Hu A, Goff S, Schmid J, Ligon JM: **The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90.** *Chem Biol* 2000, **7**:97–109.
 66. Hertweck C: **The biosynthetic logic of polyketide diversity.** *Angew Chem Int Ed Engl* 2009, **48**:4688–4716.
 67. Caffrey P: **Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases.** *ChemBioChem* 2003, **4**:654–657.
-

-
68. Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R: **A model of structure and catalysis for ketoreductase domains in modular polyketide synthases.** *Biochemistry-US* 2003, **42**:72–79.
69. Perlova O, Gerth K, Hans A, Kaiser O, Müller R: **Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56.** *J Biotechnol* 2006, **121**:174–191.
70. Kwan DH, Sun Y, Schulz F, Hong H, Popovic B, Sim-Stark JC, Haydock SF, Leadlay PF: **Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases.** *Chem Biol* 2008, **15**:1231–1240.
71. Austin MB, Noel JP: **The chalcone synthase superfamily of type III polyketide synthases.** *Nat Prod Rep* 2003, **20**:79–110.
72. Schröder J: **A family of plant-specific polyketide synthases: facts and predictions.** *Trends Plant Sci* 1997, **2**:373–378.
73. Conti E, Stachelhaus T, Marahiel MA, Brick P: **Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.** *EMBO J* 1997, **16**:4174–4183.
74. Stachelhaus T, Mootz HD, Marahiel MA: **The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.** *Chem Biol* 1999, **6**:493–505.
75. Challis GL, Ravel J, Townsend CA: **Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.** *Chem Biol* 2000, **7**:211–224.
76. Weber G, Leitner E: **Disruption of the cyclosporin synthetase gene of *Tolypocladium niveum*.** *Curr Genet* 1994, **26**:461–467.
77. Hoffmann K, Schneider-Scherzer E, Kleinkauf H, Zocher R: **Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis.** *J Biol Chem* 1994, **269**:12710–12714.
78. Balibar CJ, Vaillancourt FH, Walsh CT: **Generation of D-amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains.** *Chem Biol* 2005, **12**:1189–1200.
79. Du L, Shen B: **Biosynthesis of hybrid peptide-polyketide natural products.** *Curr Opin Drug Discov Devel* 2001, **4**:215–228.
80. Du LH, Sanchez C, Shen B: **Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules.** *Metab Eng* 2001, **3**:78–95.
81. Bender CL, Palmer DA, Peñaloza-Vázquez A, Rangaswamy V, Ullrich M: **Biosynthesis and regulation of coronatine, a non-host-specific phytotoxin produced by *Pseudomonas syringae*.** *Subcell Biochem* 1998, **29**:321–341.
82. Rangaswamy V, Mitchell R, Ullrich M, Bender C: **Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine.** *J Bacteriol* 1998, **180**:3330–3338.
83. Rangaswamy V, Jiralerspong S, Parry R, Bender CL: **Biosynthesis of the *Pseudomonas* polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins.** *P Natl Acad Sci USA* 1998, **95**:15469–15474.
84. Walsh CT, Gehring AM, Weinreb PH, Quadri LE, Flugel RS: **Post-translational modification of polyketide and nonribosomal peptide synthases.** *Curr Opin Chem Biol* 1997, **1**:309–315.
85. Gaitatzis N, Hans A, Müller R, Beyer S: **The *mtaA* gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases.** *J Biochem (Tokyo)* 2001, **129**:119–124.
-

-
86. Quadri LE, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT: **Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.** *Biochemistry-US* 1998, **37**:1585–1595.
 87. Sanchez C, Du LC, Edwards DJ, Toney M, Shen B: **Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin.** *Chem Biol* 2001, **8**:725–738.
 88. Hahn M, Stachelhaus T: **Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides.** *P Natl Acad Sci USA* 2006, **103**:275–280.
 89. Tanovic A, Samel SA, Essen LO, Marahiel MA: **Crystal structure of the termination module of a nonribosomal peptide synthetase.** *Science* 2008, **321**: 659–63
 90. Weissman KJ, Müller R: **Crystal structure of a molecular assembly line.** *Angew Chem Int Ed Engl* 2008, **47**:8344–8346.
 91. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ: **Multienzyme docking in hybrid megasynthetases.** *Nat Chem Biol* 2007, **4**:75–81.
 92. Frank B, Knauber J, Steinmetz H, Scharfe M, Blöcker H, Beyer S, Müller R: **Spiroketal polyketide formation in *Sorangium*: Identification and analysis of the biosynthetic gene cluster for the highly cytotoxic spirangienes.** *Chem Biol* 2007, **14**:221–233.
 93. Meiser P, Weissman KJ, Bode HB, Krug D, Dickschat JS, Sandmann A, Müller R: **DKxanthene biosynthesis – understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism.** *Chem Biol* 2008, **15**:771–781.
 94. Kohli RM, Walsh CT: **Enzymology of acyl chain macrocyclization in natural product biosynthesis.** *Chem Commun* 2003,297–307.
 95. Samel SA, Wagner B, Marahiel MA, Essen LO: **The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide.** *J Mol Biol* 2006, **359**:876–889.
 96. Hopwood DA, Sherman DH: **Molecular genetics of polyketides and its comparison to fatty acid biosynthesis.** *Annu Rev Genet* 1990, **24**:37–66.
 97. Ogasawara Y, Katayama K, Minami A, Otsuka M, Eguchi T, Kakinuma K: **Cloning, sequencing, and functional analysis of the biosynthetic gene cluster of macrolactam antibiotic vicenistatin in *Streptomyces halstedii*.** *Chem Biol* 2004, **11**:79–86.
 98. Kratzschmar J, Krause M, Marahiel MA: **Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases.** *J Bacteriol* 1989, **171**:5422–5429.
 99. Rusnak F, Sakaitani M, Drueckhammer D, Reichert J, Walsh CT: **Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *entF* gene, expression and purification of EntF, and analysis of covalent phosphopantetheine.** *Biochemistry-US* 1991, **30**:2916–2927.
 100. Kopp M, Irschik H, Pradella S, Müller R: **Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes.** *ChemBioChem* 2005, **6**:1277–1286.
 101. Kohli RM, Trauger JW, Schwarzer D, Marahiel MA, Walsh CT: **Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases.** *Biochemistry-US* 2001, **40**:7099–7108.
 102. Trauger JW, Kohli RM, Walsh CT: **Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase.** *Biochemistry-US* 2001, **40**:7092–7098.
-

-
103. Trauger JW, Kohli RM, Mootz HD, Marahiel MA, Walsh CT: **Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase.** *Nature* 2000, **407**:215–218.
104. Xue Y, Zhao L, Liu HW, Sherman DH: **A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity.** *P Natl Acad Sci USA* 1998, **95**:12111–12116.
105. Butler AR, Bate N, Cundliffe E: **Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*.** *Chem Biol* 1999, **6**:287–292.
106. Doi-Katayama Y, Yoon YJ, Choi CY, Yu TW, Floss HG, Hutchinson CR: **Thioesterases and the premature termination of polyketide chain elongation in rifamycin B biosynthesis by *Amycolatopsis mediterranei* S699.** *J Antibiot* 2000, **53**:484–495.
107. Heathcote ML, Staunton J, Leadlay PF: **Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units.** *Chem Biol* 2001, **8**:207–220.
108. Yeh E, Kohli RM, Bruner SD, Walsh CT: **Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated.** *ChemBioChem* 2004, **5**:1290–1293.
109. Schwarzer D, Mootz HD, Linne U, Marahiel MA: **Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases.** *P Natl Acad Sci USA* 2002, **99**:14083–14088.
110. Koglin A, Mofid MR, Lohr F, Schafer B, Rogov VV, Blum MM, Mittag T, Marahiel MA, Bernhard F, Dotsch V: **Conformational switches modulate protein interactions in peptide antibiotic synthetases.** *Science* 2006, **312**:273–276.
111. Samel SA, Marahiel MA, Essen LO: **How to tailor non-ribosomal peptide products - new clues about the structures and mechanisms of modifying enzymes.** *Mol BioSyst* 2008, **4**:387–393.
112. Walsh CT, Chen HW, Keating TA, Hubbard BK, Losey HC, Luo LS, Marshall CG, Miller DA, Patel HM: **Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines.** *Curr Opin Chem Biol* 2001, **5**:525–534.
113. Rix U, Fischer C, Remsing LL, Rohr J: **Modification of post-PKS tailoring steps through combinatorial biosynthesis.** *Nat Prod Rep* 2002, **19**:542–580.
114. Jansen R, Irschik H, Reichenbach H, Höfle G: **Antibiotics from gliding bacteria, LXXX. Chivosazoles A-F: Novel antifungal and cytotoxic macrolides from *Sorangium cellulosum* (Myxobacteria).** *Liebigs Ann Chem* 1997,1725–1732.
115. Rachid S, Krug D, Kochems I, Kunze B, Scharfe M, Blöcker H, Zabriski M, Müller R: **Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5.** *Chem Biol* 2006, **14**:667–681.
116. Buedenbender S, Rachid S, Müller R, Schulz GE: **Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases.** *J Mol Biol* 2009, **385**:520–530.
117. Zhou H, Xie X, Tang Y: **Engineering natural products using combinatorial biosynthesis and biocatalysis.** *Curr Opin Biotechnol* 2008, **19**:590–596.
118. Zhang C, Griffith BR, Fu Q, Albermann C, Fu X, Lee IK, Li L, Thorson JS: **Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions.** *Science* 2006, **313**:1291–1294.
119. Zhang CS, Albermann C, Fu X, Thorson JS: **The *in vitro* characterization of the iterative avermectin glycosyltransferase AveBI reveals reaction reversibility and sugar nucleotide flexibility.** *J Am Chem Soc* 2006, **128**:16420–16421.
-

-
120. Yoon YJ, Beck BJ, Kim BS, Kang HY, Reynolds KA, Sherman DH: **Generation of multiple bioactive macrolides by hybrid modular polyketide synthases in *Streptomyces venezuelae***. *Chem Biol* 2002, **9**:203–214.
121. Stadler M, Bitzer J, Mayer-Bartschmid A, Muller H, et-Buchholz J, Gantner F, Tichy HV, Reinemer P, Bacon KB: **Cinnabaramides A-G: Analogues of lactacystin and salinosporamide from a terrestrial streptomycete**. *J Nat Prod* 2007, **70**:246–252.
122. McKinney JD: ***In vivo veritas*: the search for TB drug targets goes live**. *Nat Med* 2000, **6**:1330–1333.
123. Zhang Y, Post-Martens K, Denkin S: **New drug candidates and therapeutic targets for tuberculosis therapy**. *Drug Discov Today* 2006, **11**:21–27.
124. Jansen R, Kunze B, Reichenbach H, Höfle G: **The ajudazols A and B, novel isochromanones from *Chondromyces crocatus* (Myxobacteria): Isolation and structure elucidation**. *Eur J Org Chem* 2002,917–921.
125. Kunze B, Jansen R, Höfle G, Reichenbach H: **Ajudazols, new inhibitors of the mitochondrial electron transport from *Chondromyces crocatus*. Production, antimicrobial activity and mechanism of action**. *J Antibiot* 2004, **57**:151–155.
126. Gaitatzis N, Silakowski B, Kunze B, Nordsiek G, Blöcker H, Höfle G, Müller R: **The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase**. *J Biol Chem* 2002, **277**:13082–13090.
127. Smith L, Hong H, Spencer JB, Leadlay PF: **Analysis of specific mutants in the lasalocid gene cluster: evidence for enzymatic catalysis of a disfavoured polyether ring closure**. *ChemBioChem* 2008, **9**:2967–2975.
128. Kunze B, Jansen R, Höfle G, Reichenbach H: **Ajudazols, new inhibitors of the mitochondrial electron transport from *Chondromyces crocatus*. Production, antimicrobial activity and mechanism of action**. *J Antibiot* 2004, **57**:151–155.
129. Bode HB, Wenzel SC, Irschik H, Höfle G, Müller R: **Unusual biosynthesis of leupyrrins in the myxobacterium *Sorangium cellulosum***. *Angew Chem Int Ed* 2004, **43**:4163–4167.
130. Kopp M. **Identifizierung und Charakterisierung von Biosynthesegenclustern aus *Sorangium cellulosum***. *Thesis/Dissertation* 2005
131. Irschik H, Jansen R, Gerth K, Höfle G, Reichenbach H: **Disorazol A, an efficient inhibitor of eukaryotic organisms isolated from myxobacteria**. *J Antibiot* 1995, **48**:31–35.
132. Jansen R, Irschik H, Reichenbach H, Wray V, Höfle G: **Disorazoles, highly cytotoxic metabolites from the sorangicin-producing bacterium *Sorangium cellulosum*, strain So ce12**. *Liebigs Ann Chem* 1994,759–773.
133. Carvalho R, Reid R, Viswanathan N, Gramajo H, Julien B: **The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation**. *Gene* 2005, **359**:91–98.
134. Tran L, Tosin M, Spencer JB, Leadlay PF, Weissman KJ: **Covalent linkage mediates communication between ACP and TE domains in modular polyketide synthases**. *ChemBioChem* 2007, **9**:905–915.
135. Sasse F, Steinmetz H, Höfle G, Reichenbach H: **Rhizopodin, a new compound from *Myxococcus stipitatus* (myxobacteria) causes formation of rhizopodia-like structures in animal cell cultures. Production, isolation, physico-chemical and biological properties**. *J Antibiot* 1993, **46**:741–748.
136. Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT: **Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization**. *Chem Biol* 1999, **6**:385–400.
-

-
137. Hoyer KM, Mahlert C, Marahiel MA: **The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization.** *Chem Biol* 2007, **14**:13–22.
138. Bode HB, Irschik H, Wenzel SC, Reichenbach H, Müller R, Höfle G: **The leupyrrins: A structurally unique family of secondary metabolites from the myxobacterium *Sorangium cellulosum*.** *J Nat Prod* 2003, **66**:1203–1206.
139. Pojer F, Li SM, Heide L: **Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics.** *Microbiology* 2002, **148**:3901–3911.
140. Garneau S, Dorrestein PC, Kelleher NL, Walsh C: **Characterization of the formation of the pyrrole moiety during clorobiocin and coumermycin A1 biosynthesis.** *Biochemistry-US* 2005, **44**:2770–2780.
141. Thomas MG, Burkart MD, Walsh CT: **Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis.** *Chem Biol* 2002, **9**:171–184.
142. Nowak-Thompson B, Chaney N, Wing JS, Gould SJ, Loper JE: **Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5.** *J Bacteriol* 1999, **181**:2166–2174.
143. Mejean A, Mann S, Maldiney T, Vassiliadis G, Lequin O, Ploux O: **Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-a in the cyanobacterium *Oscillatoria* PCC 6506 occurs on a modular polyketide synthase initiated by L-proline.** *J Am Chem Soc* 2009, **131**:7512–7513.
144. Keating TA, Suo Z, Ehmann DE, Walsh CT: **Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain.** *Biochemistry-US* 2000, **39**:2297–2306.
145. Gemperlein K. **Biochemische Untersuchungen zur Biosynthese von Leupyrrin aus *Sorangium cellulosum* So ce690.** *Thesis/Dissertation* 2009
146. Erb TJ, Brecht V, Fuchs G, Müller M, Alber BE: **Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase.** *Proc Natl Acad Sci USA* 2009, **106**:8871–8876.
147. Kato Y, Bai L, Xue Q, Revill WP, Yu TW, Floss HG: **Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B.** *J Am Chem Soc* 2002, **124**:5268–5269.
148. Erb TJ, Berg IA, Brecht V, Müller M, Fuchs G, Alber BE: **Synthesis of C-5-dicarboxylic acids from C-2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway.** *P Natl Acad Sci USA* 2007, **104**:10631–10636.
149. Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T: **Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2 in *Escherichia coli*.** *Appl Environ Microbiol* 1998, **64**:1694–1699.
150. Nishihara K, Kanemori M, Yanagi H, Yura T: **Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*.** *Appl Environ Microbiol* 2000, **66**:884–889.
151. Liu Y, Hazzard C, Eustaquio AS, Reynolds KA, Moore BS: **Biosynthesis of salinosporamides from a,b-unsaturated fatty acids: implications for extending polyketide synthase diversity.** *J Am Chem Soc* 2009, **131**:10376–10377.
152. Eustaquio AS, McGlinchey RP, Liu Y, Hazzard C, Beer LL, Florova G, Alhamadsheh MM, Lechner A, Kale AJ, Kobayashi Y, Reynolds KA, Moore BS: **Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine.** *P Natl Acad Sci USA* 2009, **106**:12295–12300.
-

-
153. Tang L, Ward S, Chung L, Carney JR, Li Y, Reid R, Katz L: **Elucidating the mechanism of *cis* double bond formation in epothilone biosynthesis.** *J Am Chem Soc* 2004, **126**:46–47.
154. Tang GL, Cheng YQ, Shen B: **Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase.** *Chem Biol* 2004, **11**:33–45.
155. Sandmann A, Sasse F, Müller R: **Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity.** *Chem Biol* 2004, **11**:1071–1079.
156. August PR, Tang L, Yoon YJ, Ning S, Müller R, Yu TW, Taylor M, Hoffmann D, Kim CG, Zhang X, Hutchinson CR, Floss HG: **Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699.** *Chem Biol* 1998, **5**:69–79.
157. Chen S, von Bamberg D, Hale V, Breuer M, Hardt B, Müller R, Floss HG, Reynolds KA, Leistner E: **Biosynthesis of ansatrienin (mycotrienin) and naphthomycin: identification and analysis of two separate biosynthetic gene clusters in *Streptomyces collinus* Tü 1892.** *E J Biochem* 1999, **261**:98–107.
158. Gaitatzis N, Kunze B, Müller R: ***In vitro* reconstitution of the myxochelin biosynthetic machinery of *Stigmatella aurantiaca* Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases.** *P Natl Acad Sci USA* 2001, **98**:11136–11141.
159. Kopp F, Mahlert C, Grunewald J, Marahiel MA: **Peptide macrocyclization: The reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine.** *J Am Chem Soc* 2006, **128**:16478–16479.
160. Li L, Deng W, Song J, Ding W, Zhao QF, Peng C, Song WW, Tang GL, Liu W: **Characterization of the saframycin A gene cluster from *Streptomyces lavendulae* NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner.** *J Bacteriol* 2008, **190**:251–263.
161. Keating TA, Ehmann DE, Kohli RM, Marshall CG, Trauger JW, Walsh CT: **Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis.** *ChemBioChem* 2001, **2**:99–107.
162. Li Y, Weissman KJ, Müller R: **Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester.** *J Am Chem Soc* 2008, **130**:7554–7555.
163. Julien B, Tian ZQ, Reid R, Reeves CD: **Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis.** *Chem Biol* 2006, **13**:1277–1286.
164. Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, Müller R: **Myxovirescin biosynthesis is directed by hybrid polyketide synthases/ nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl CoA synthases and trans-acting acyltransferases.** *ChemBioChem* 2006, **7**:1206–1220.
165. Akashi T, Fukuchi-Mizutani M, Aoki T, Ueyama Y, Yonekura-Sakakibara K, Tanaka Y, Kusumi T, Ayabe S: **Molecular cloning and biochemical characterization of a novel cytochrome P450, flavone synthase II, that catalyzes direct conversion of flavanones to flavones.** *Plant Cell Physiology* 1999, **40**:1182–1186.
166. Kelkar HS, Skloss TW, Haw JF, Keller NP, Adams TH: ***Aspergillus nidulans* stcL encodes a putative cytochrome P-450 monooxygenase required for bisfuran desaturation during aflatoxin/sterigmatocystin biosynthesis.** *J Biol Chem* 1997, **272**:1589–1594.
167. Ortiz de Montellano PR: **Cytochrome P450: Structure, mechanism, and biochemistry.** New York: Kluwer Academic/Plenum Publishers; 2005.
-

-
168. Buntin K, Rachid S, Scharfe M, Blöcker H, Weissman KJ, Müller R: **Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications.** *Angew Chem Int Ed* 2008, **47**:4595–4599.
169. van Berkel WJH, Kamerbeek NM, Fraaije MW: **Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts.** *J Biotechnol* 2006, **124**:670–689.
170. Xue YQ, Wilson D, Zhao LS, Liu HW, Sherman DH: **Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the pikC-encoded cytochrome P450 in *Streptomyces venezuelae*.** *Chem Biol* 1998, **5**:661–667.
171. Jenke-Kodama H, Borner T, Dittmann E: **Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*.** *Plos Compl Biol* 2006, **2**:1210–1218.
172. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E: **Evolutionary implications of bacterial polyketide synthases.** *Mol Biol Evol* 2005, **22**:2027–2039.

Authors effort in publications from Chapter 1-4

Chapter 1: The author postulated the biosynthetic model, generated the P450 inactivation mutants and performed the purification of deshydroxy ajudazol A. In addition the author participated in the subsequent structure elucidation. Docking domain analysis was done by K. J. Weissman.

Chapter 2: The major part of the work was performed by the author. E. Luxenburger performed the HPLC-MS measurements.

Chapter 3: The author analysed *in silico* the KR domains and predicted the stereochemistry.

Chapter 4: The major part of the work was performed by the author. Docking domain analysis was done by K. J. Weissman. HPLC-MS and high resolution measurements were carried out by E. Luxenburger.