Investigation of bacterial secondary metabolite pathways from *Sorangium cellulosum*

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

von

Katrin Jungmann

Saarbrücken

2015

Tag des Kolloquiums:	22.01.2016
Dekan:	Prof. Dr. Ing. Dirk Bähre
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Rolf W. Hartmann
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Martin Frotscher

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller in der Fachrichtung 8.2, Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät III der Universität des Saarlandes von Januar 2012 bis August 2015.

Alles Wissen und alles Vermehren unseres Wissens
endet nicht mit einem Schlußpunkt,
sondern mit einem Fragezeichen.

Hermann Hesse (1877-1962)

Danksagung

Zunächst möchte ich meinem Doktorvater Prof. Dr. Rolf Müller für das mir entgegengebrachte Vertrauen und die Überlassung herausfordernder und spannender Themen, für seine Unterstützung und die wissenschaftlichen Diskussionen und für die gute Zusammenarbeit danken. Auch bei Prof. Dr. Rolf Hartmann, der diese Arbeit wissenschaftlich begleitet hat, möchte ich mich für die Übernahme des Co-Referates bedanken.

Meinem Betreuer Daniel Krug danke ich dafür, dass er stets Vertrauen in mein Urteil und meine Entscheidungen hatte und mich in meiner Selbstständigkeit unterstützt hat. Ebenso danke ich Dr. Silke Wenzel für die gute und lehrreiche Zusammenarbeit im Maracen Projekt.

Darüber hinaus möchte ich mich bei der ganzen Arbeitsgruppe für die gute Zusammenarbeit bedanken. Bei Dr. Lena Keller für NMR-Messungen, bei Dr. Thomas Hoffmann und Michael Hoffmann, die bei sämtlichen analytischen Fragestellungen eine große Hilfe waren sowie bei Dr. Kirsten Harmrolfs und Dr. David Auerbach für ihre Unterstützung bei der Semi-Synthese. Ein besonderer Dank geht auch an Hilda Sucipto für die zahlreichen Diskussionen und an meine Laborbank-Nachbarn Ullrich Scheid, Alexander von Tesmar und Dr. Alberto Plaza für die gute Zeit. Danke auch an die Freitags-Bier Gruppe für „Guuude Laune“, Spaß und Unfug; ich hoffe wir verbringen noch so manche lustige Stunde miteinander.

Es ist mir ein besonderes Anliegen, mich auch bei meinen Eltern, Martina und Hans-Peter Jungmann zu bedanken, die mich während meiner gesamten Ausbildung unterstützt und mich in schwierigen Situationen ermutigt haben. Euer Rückhalt steht zwischen all diesen Zeilen.

Zu guter Letzt möchte ich auch Henning danken ohne den diese Arbeit so nicht möglich gewesen wäre. Ich danke dir für dein offenes Ohr, all deine Unterstützung und dafür, dass du immer für mich da warst wenn meine Projekte mir zu schaffen machten.

Vorveröffentlichungen der Dissertation

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

Publikationen

Jungmann K., Jansen R., Gerth K., Huch V., Krug D., Fenical W., Müller R. (**2015**) Two of a kind – The biosynthetic pathways of chlorotonil and anthracimycin *ACS Chemical Biology*, Article ASAP, doi: 10.1021/acschembio.5b00523, published online 8 September 2015, reproduced by permission of American Chemical Society

Tagungsbeiträge

Jungmann K., Krug D., Müller R. (**2015**) Two of a kind – The biosyntehtic pathways of chlorotonil and anthracimycin (Poster) *Directing Biosynthesis IV*, John Innes Centre Norwich, UK

Jungmann K., Krug D., Müller R. (**2014**) Deciphering the Chlorotonil biosynthetic pathway in *Sorangium cellulosum* (Poster) *2nd International Conference on Natural Products,* Cape Town, South Africa

Jungmann K., Krug D., Müller R. (**2013**) Deciphering the Chlorotonil biosynthetic pathway in *Sorangium cellulosum* (Poster) *1st European Conference on Natural Products*, Frankfurt am Main, Germany

Jungmann K., Krug D., Müller R. (**2013**) Biosynthesis of new natural products in *Sorangium cellulosum* (Oral Communication) *Summer Symposium of the Interdisciplinary Graduate School of Natural Products Research*, Saarbrücken, Germany

Jungmann K., Krug D., Gemperlein K., Müller R. (**2013**) Chlorotonil & Maracen – Deciphering biosynthetic pathways of secondary metabolites from *Sorangium cellulosum* (Oral Communication) MINAS Symposium “Mikrobielle Naturstoffe”, Warberg, Germany

Weitere Publikationen

Marnix H. Medema, Renzo Kottmann, Pelin Yilmaz, Matthew Cummings, John B. Biggins, Kai Blin […] **Jungmann K.** et al. (**2015**) The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification, *Nature Chemical Biology* **11**, 625–631 (2015) doi:10.1038/nchembio.1890, Published online 18 August 2015

Zusammenfassung

Naturstoffe sind seit jeher eine bedeutende Quelle auf der Suche nach neuen Leitstrukturen, insbesondere vor dem Hintergrund der sich verschärfenden Resistenzentwicklung gegen Antibiotika. Dabei hat es sich bewährt, bisher unerforschte Lebensräume und Mikroorganismen – wie beispielsweise Myxobakterien zu untersuchen. Myxobakterien – ubiquitär vorkommende Gram-negative δ-Proteobakterien - haben in den letzten Jahrzehnten ihre enorme Kapazität zur Produktion strukturell diverser Sekundärmetabolite unter Beweis gestellt.

Aus dem Myxobakterium *Sorangium cellulosum* So ce1525 stammt der neuartige Sekundärmetabolit Chlorotonil A, der neben seinen strukturellen Besonderheiten auch sehr vielversprechende Bioaktivität, insbesondere gegen den problematischen Malariaerreger *Plasmodium falciparum* aufweist. Die Biosynthese dieses Sekundärmetaboliten konnte im Rahmen dieser Dissertation aufgeklärt werden. Zudem erlaubte ein Vergleich mit dem Biosyntheseweg des strukturell ähnlichen Streptomyceten-Sekundärmetaboliten Anthracimycin tiefergehende Einblicke in die zugrundeliegenden Biosyntheseprozesse. Weiterhin konnte der Chlorotonil Biosynthesegencluster mobilisiert und für die Etablierung einer Plattform zur heterologen Expression adaptiert werden. Darüber hinaus wurden auf semi-synthetischem Weg erste Derivate des Naturstoffes erzeugt, die wegweisend für erweiterte Studien mit dem Ziel der Herstellung bioaktiver Chlorotonil-Derivate mit verbesserten pharmakokinetischen Eigenschaften.

Außerdem wurde die Biosynthese der beiden Sekundärmetabolite Maracen und Maracen, die ebenfalls von Vertretern der Gattung *Sorangium* produziert werden und vielversprechende Wirkung gegen Tuberkuloseerreger zeigen, untersucht. Anders als im klassischen Ansatz wurde hierbei mit Methoden der synthetischen Biologie ein artifizieller Gencluster anhand der Arbeitshypothese – Maracen/Maracin Biosynthese durch stufenweise enzymatische Unmwandlung von Eicosapentaensäure - erzeugt. Die durchgeführten Arbeiten ebnen den Weg für Untersuchungen der biosynthetischen Abläufe und ermöglichen gleichzeitig die Etablierung der heterologen Expression des Genclusters.

Summary

Natural products have always been an important source for new lead structures especially in the context of increasing resistance rates against contemporarily available antibiotics. In search of new antiinfectives it has proven to be advantageous to explore the secondary metabolome of yet unknown microorganisms and microoragnisms from ecological niches such as myxobacteria. Myxobacteria, soil-dwelling Gram-negative δ-proteobacteria, have repeatedly demonstrated their enormous potential as prolific produces of structurally diverse secondary metabolites in the last decades.

Chlorotonil A, produced by the myxobacterium *Sorangium cellulosum* So ce 1525, apart from its structural peculiarities shows highly promising bioactivity against the problematic malaria pathogen *Plasmodium falciparum*. The biosynthesis of Chlorotonil A was elucidated whereas a comparison with the biosynthetic pathway of the related streptomycete metabolite anthracimycin allowed for insights into the underlying biosynthetic processes in greater depth. Moreover the chlorotonil biosynthetic gene cluster was successfully mobilized from So ce1525 and subsequently adapted for establishment of a platform for hetrologous expression. Using a semi-synthetic route, also first natural product derivatives, which are giving direction for further studies towards semi-synthesis of chlorotonil derivatives with enhanced pharmacokinetic properties, were generated.

In addition, the biosynthesis of the two secondary metabolites maracen and maracin, produced by several *Sorangium* species and exhibting auspicious anti-tuberculosis activity, was investigated. In derogation of the classical approach, methods of synthetic biology were used to generate an artificial gene cluster based on the working hypothesis. The work accomplished paves the way for extended analysis of the biosynthetic processes and at the same time enables the installation of a system for heterologous expression of the pathway.

Table of Contents

1. Introduction

1.1 Relevance of natural products as contemporary therapeutics

The documented use of natural products as therapeutics reaches back into antiquity, where first records on their application were made. Although the active compounds and their structures were not known, the application and the medical indication of the respective plants were well described. The papyri “Ebers” and “Edwin Smith” which were found in Egypt date back to the 15th century before Christ and in contrast to other documents do not describe mystic magical procedures but contain guidelines for wound treatment and usage of medicinal plants. Also in ancient Rome, willow bark was already well-known for its analgetic and antipyretic effects but its active pharmaceutical ingredient, salicylic acid, remained unknown until the 19th century. The german pharmacologist Johann Andreas Buchner then set the stage for the production of the blockbuster drug acetyl salicylic acid in 1828, by isolating salicin, the active ingredient of willow bark.[2] The history of Aspirin is just one of many examples describing how ancient knowledge of medicinal plants led to successful pharmaceuticals which are still in use today. The history of the malaria therapeutic chinin and the opiate morphin is very similar to that of Aspirin. The respective plant or plant extract was already used for a long time, but only the development of advanced analytical methods and sophisticated protocols in organic chemistry allowed for synthesis of pure compounds and their establishment as successful drug substances.

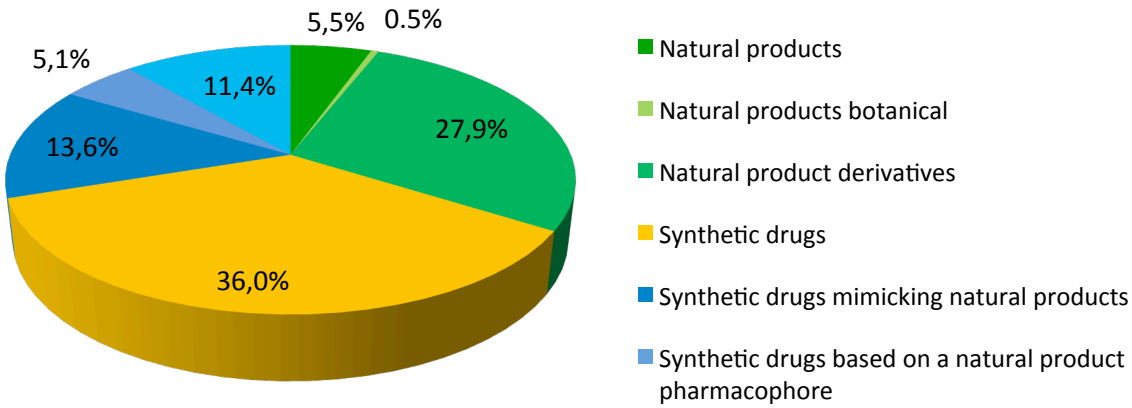


Figure 1: Source of small molecule approved drugs[3]

The capacity of natural products in therapy is also reflected in the origin of newly approved drugs in the last decades; about a third of the compounds are natural products or their derivatives whilst another third is based on pharmacophores from natural products. In total 64 % of all new drugs approved from 1980 to 2010 are either natural products or related to natural products. These numbers underpin the important role of natural products in the drug discovery process. Natural products in comparison to synthetic compounds are evolutionary preselected for biological activity as they often provide selection advantage for their producers.[4] It could be shown that for example that the bacterium *Myxococcus xanthus*, which can prey on other microbes like *Escherichia coli* cells as nutrient source is only able to do so if the antibiotic myxovirescin is produced. Myxovirescin negative mutants of *M. xanthus* were no longer able to grow in the presence of the *E. coli* strain or use it as as nutritional source.[5] Another prominent example is the quorum sensing system of *Pseudomonas aeruginosa*, a Gram-negative pathogen which mostly causes pneumonia and cystic fibrosis. The pathogen produces several quorum sensing molecules which induce biofilm formation and secretion of toxic compounds and thus provide an advantage over other colonizing bacteria.[6] Thus, natural products can be considered as privileged structures while at the same time they provide higher complexity and diversity than found in synthetic libraries[7]. Consequentially, natural products are a promising starting point for the development of new drugs, with novel structural features and new modes of action.

1.2 Anti-infectives and resistance development

Prior to the research work of Robert Koch and Louis Pasteur, the origin of infectious diseases was largely unknown and no therapeutic options for the treatment of infected patients were available. Therefore the patient's recovery strongly depended on his innate immune response and the given living conditions. Despite vaccination against smallpox was already successfully introduced in Europe in the 18th century by the medical doctor Jean de Carro, the origin of the disease remained elusive. It was only when Louis Pasteur pronounced the "Germ Theory" in 1878,[8] and Robert Koch proved the existence of the anthrax pathogen *Bacillus anthracis* (1864)[9] and the tuberculosis pathogen *Mycobacterium tuberculosis* (1881)[10], that microbial pathogens were found to be responsible for outburst of these diseases. Paul Ehrlich who worked in Robert Koch‘s laboratory then contributed to the understanding and the treatment of infectious diseases.

Ehrlich explored chemicals which were able to selectively stain and visualize cells and germs.[11] Based on this principle, he postulated that if there were selective chemicals that stain the germs, one would only have to look for agents that could stain and kill the germ at the same time – so called “magic bullets”. In his search for a syphilis cure, Ehrlich collaborated with Sahachiro Hata who developed a syphilis rat model. When screening a library of chemical compounds, they finally discovered „salvarsan“ the first therapeutic agent against syphilis in 1909 and thus laid the foundation for modern chemotherapy.[12] Also Ehrlichs intense research in vaccines and sera had a great impact on infection research. In 1928, the Scottish bacteriologist Alexander Fleming eventually discovered the first antibiotic, when he observed a penicillium mould contaminating his staphylococcus cultures obviously inhibiting the germs’ growth. Prior to these observations, Fleming already discovered the anti-bacterial effect of lysozyme,[13] but in-depth investigations of the penicillium mould finally led to the discovery of the first broad spectrum antibiotic penicillin.[14] What followed was the finding of several new classes of antibiotics from different natural sources during the so-called “golden era” of antibiotic discovery from 1930 to 1960.[15] Former fatal diseases could now be rapidly and efficiently cured by the use of the newly discovered antibiotics. The large success of anti-infectives during that period even enticed the Surgeon General of the United States of America, William Stewart to state in 1967, that one could close the book of infectious diseases.[16] Following this assumption, no new antibiotics were discovered in the following decades, but known lead structures were altered by the means of medicinal chemistry. This strategy failed promptly, as on the one hand, resistance development proceeded, whilst on the other hand, also several new infectious diseases emerged.[17]

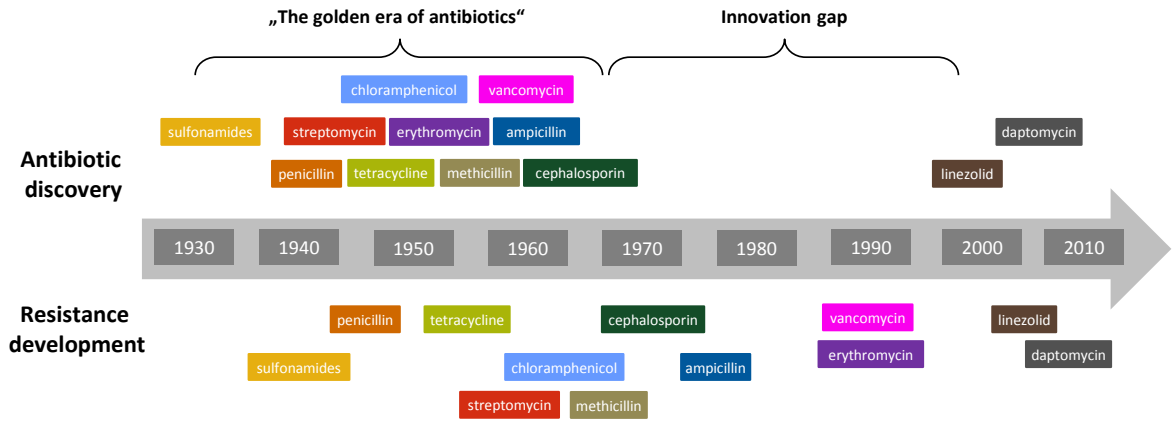


Figure 2: Antibiotic discovery and resistance development. Adapted from Walsh and Wencewizc, 2013[15] and Clatworthy et al. 2007[18]

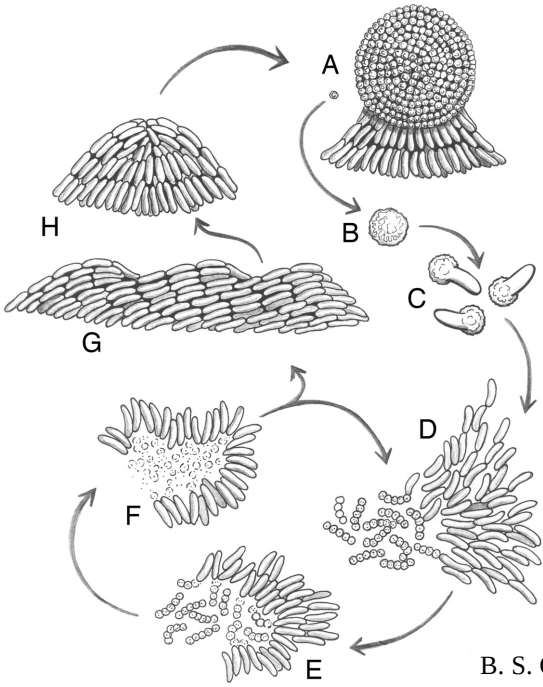
Ever since, emerging antimicrobial resistance against antibiotics commonly used became an increasing problem. The reactions of science and industry to this development came far too late, which has considerably aggravated resistance development. Nowadays, even reserve antibiotics are unable to fight multi-resistant pathogens which are especially problematic in immune suppressed patients that have acquired nosocomial infections. In hospital-acquired infections, the so-called 'ESKAPE bugs', namely *Enterococcus feacium*, *Staphylococcus aureus*, *Klebsiella pneumonia*, *Acinetobacter baumanni*, *Pseudomonas aeruginosa* and *Enterobacter species* play a major role since they 'escape' the antibacterial effects of known drugs through their acquired resistance mechanisms. [19,20] Controlling these pathogens by finding effective therapeutic options is one of the main goals of infection research. But emerging resistance rates and lacking treatment options are not only a problem in first world countries; still there is an immense need for therapeutics against tuberculosis, malaria and so-called neglected diseases, in developing countries. In order to meet these needs, intense research in anti-infectives is necessary as novel therapeutics with new modes of action or new targets are required.

1.3 Myxobacteria as proficient producers of natural products

For decades, microbes such as streptomycetes or fungi are known as proficient producers of bioactive natural products. In 1995, 55 % of the ~12000 antibiotics known were produced by streptomycetes; another 11 % originated from other actinomycetes.[21] The cause for the prolific production of secondary metabolites seen in these bacteria is the highly competitive soil habitat they live in. But due to continuous exploitation of these resources, the isolation of new scaffolds is hampered by frequent re-discovery and re-isolation of known compounds. However it has been shown that the number of secondary metabolite gene clusters present in a certain strain does not necessarily correspond to the number of compounds produced by the respective microorganism. It could thus be concluded that several biosynthetic pathways are silent under the given laboratory conditions meaning that transcription and translation of the biosynthetic machinery does not take place and consequently no corresponding metabolites are produced. The awakening of such silent gene clusters provides an opportunity to access new secondary metabolites, but the more promising approach is to exploit uncharted microbial resources as for example the relatively young field of myxobacterial research.[22] Myxobacteria are soil dwelling Gram-negative δ-proteobacteria, which are characterized by several distinct

features, such as their particular life cycle, their social behavior and the gliding motility. At sufficient nutrient supply, myxobacteria appear as rod shaped vegetative cells of 4-12 μm, which are able to move by gliding over the surface of the substrate. Unlike other bacteria, myxobacteria do not have flagella but they exhibit two motility systems responsible for cell movement. The S-motility (social) describes the movement of cell groups and in contrast to the A-motility (adventurous) of individual single cells, requires cell-cell contact. The S –motility is driven by extension and retraction of pili which are formed at the leading pole of the cells. Those pili extend to bind to exopolysaccharides on the surface of another cell and then move forward by retraction of polysaccharide-bound pili.[23] The mechanisms behind A-motility are not fully understood yet but two models have been proposed. One is that the cells are pushed forward by unilateral secretion and hydration of polysaccharides. The second model involves the binding of uncharacterized protein motors to transient adhesion complexes along the cell which is thought to result in a rotating locomotion.[24]

The life cycle of myxobacteria and also their movement is mainly based on the availability of nutrients. When myxobacterial swarms come across prey organisms or other nutrient sources, they degrade those cells which then serve as substrate for myxobacterial growth. During this process, the so-called “rippling” – a rhythmic movement of the myxobacterial cells- is observed; the cells align and build up waves that travel convergently. When nutrient sources are depleting myxobacteria congregate to form multicellular mounds which develop into fruiting bodies in which – at constant absence of substrates – the cells convert to spherical myxospores. The fruiting bodies measure 20-1000 μm in size and mostly consist of myxospores whilst a smaller fractions of cells (about 10 %) remain undifferentiated and enclosed the fruiting body as a monolayer Myxospores are a non-vegetative dormant cell type, which is highly resistant to extreme environmental factors such as pH value or temperature fluctuations. With ameliorated nutrient supply these spores germinate and the cells re-enter the vegetative growth cycle. [23]



B. S. Goldman et al. PNAS

Figure 3: Depiction of the myxobacterial life cycle of *Myxococcous xanthus* representing a fruiting body (A), spherical myxospores (B), germinating spores (C), rod-shaped vegetative cells in the presence of prey organisms (D), aggregation of vegetative cells (E), alignment of aggregated vegetative cells during prey cell degradation (F), formation of accordion waves (G), aggregation in cell mounds under starvation conditions.

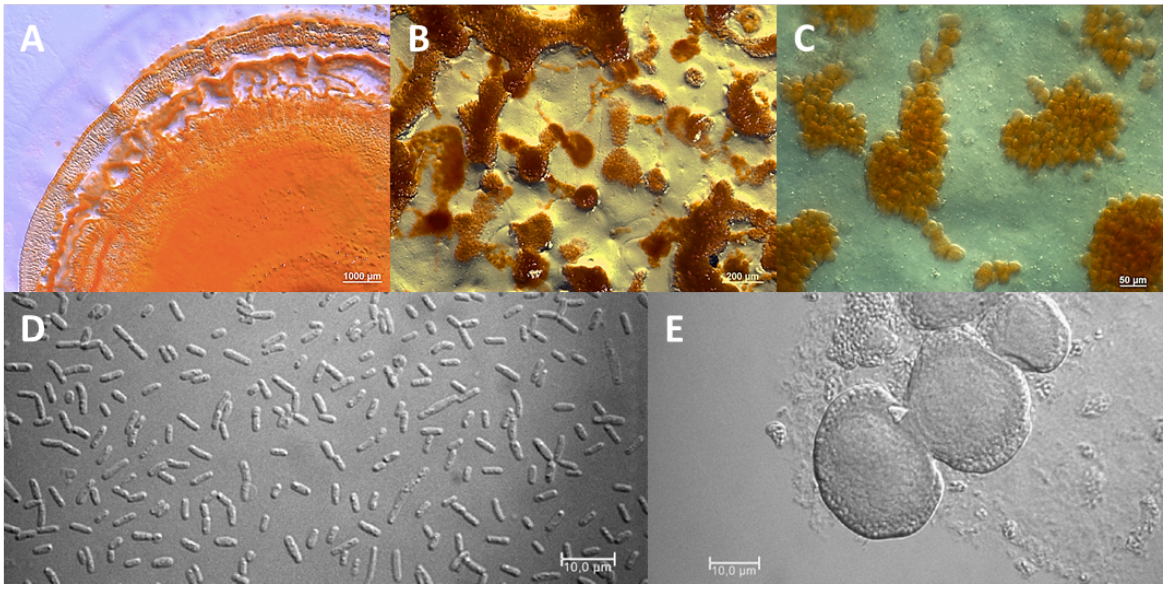
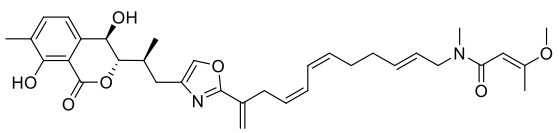
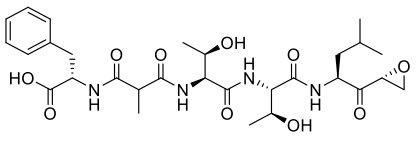


Figure 4 Growth stages of *Sorangium cellulosum* (So ce56) on agar. (pictures taken by Ronald Garcia) A: Deep-agar swarming colony and cellular aggregations. B-C: Clustered sporangioles forming fruiting bodies. D: vegetative cells. E: sporangioles surrounded by spores . A-C, stereophotomicrographs. D and E, differential interference contrast (DIC) photomicrographs.

Besides their complex life cycle myxobacteria also have proven to be proficient producers of secondary metabolites. It is thought that the antibiotic substances produced promote the preying behavior of the bacteria as they kill prey organisms prior to their degradation and also help the rather stationary myxobacterial cell complex to defend its ecological niche against competitors.[25] The capacity for the production of diverse secondary metabolites is also reflected in the genome size of myxobacteria. With an average size of 9-15 Mbp myxobacterial genomes belong to the largest group of bacterial genomes. Especially representatives of the genus *Sorangium* show remarkable genome sizes; amongst them the largest bacterial genome (14.8 Mbp) was found in the epothilon producer strain *Sorangium cellulosum* So0157-2.[26] The average GC content of myxobacterial DNA ranges between 66 – 72 mol% which is similar to the GC content of actinomycete genomes.[25] In contrast to actinomycetes, 8-10 % of the genetic information of myxobacteria are dedicated to genes encoding for secondary metabolite pathways.[1,27] With an average number of 10-20 biosynthetic gene clusters, myxobacteria thus surpass the biosynthetic potential of streptomycetes.[28,29] Already in 1947 the potential of myxobacteria as prolific producers of natural products has been discovered but systematic investigations have only started with the formation of the Helmholtz Center for Infection Research in 1965 (Institute for Molecular Biology, Biochemics and Biophysics at that time).[30] To date more than 100 new natural product core structures and about 600 structural variants have been isolated from myxobacteria;[25,31] most of them exhibiting a basic framework of polyketide or a peptide type. Many of these feature unusual structural peculiarities such as the isochromanone ring in ajudazol or the epoxyketone moiety in macyranone A;[32] very often myxobacterial secondary metabolites also show an unusual or a new mode of action.



Ajudazol A



Macyranone A

Figure 5: Structures of ajudazol A from *Sorangium cellulosum* So ce10 and macyranone A from *Cystobacter fuscus* MCy9118

The most prominent myxobacterial compound by far is the macrolide epothilone which is produced by several *Sorangium* strains.[33] Initially discovered as an anti-fungal

metabolite, a screening for cytotoxic activity revealed that similarly to taxol, epothilon binds β-tubulin and consequently stabilizes the microtubule system which is crucial for permanence and proliferation of cancer cells. [34,35] Epothilone has thus been shown to induce apoptosis in several cancer types.[36] Interestingly, the compound was also found to be active against taxane-resistant tumors, despite taxol and epothilone occupy the same binding pocket, but studies on the binding mode hint that epothilone can tolerate interchanges in the binding pocket better and thus shows lower susceptibility to tumor resistance mechanisms.[37] Due to several disadvantageous PK/PD features of epothilone, the structure was refined leading to the semisynthetic derivative ixabepilone with increased potency, water solubility and plasma stability.[38,39] In 2007 the Food and Drug Administration (FDA) approved the semi-synthetic ixabepilone (Ixempra®) for the treatment of taxol resistant breast cancer.

Apart from epothilone also several other myxobacterial substances interact with the eukaryotic cytoskeleton and interfere with the cell cycle. Tubulysin A, a linear tetrapeptide isolated from *Pyxicoccus fallax* inhibits tubulin polymerization and induces depolimerization of microtubules.[40] The same mechanism of action is described for the macrocyclic polyketide disorazol A which was isolated from *Sorangium cellulosum*.[41] Rhizopodin and chondramid C likewise interact with the cytoskeleton but their cytotoxic effect results from inhibition of actin polymerization.[42,43]

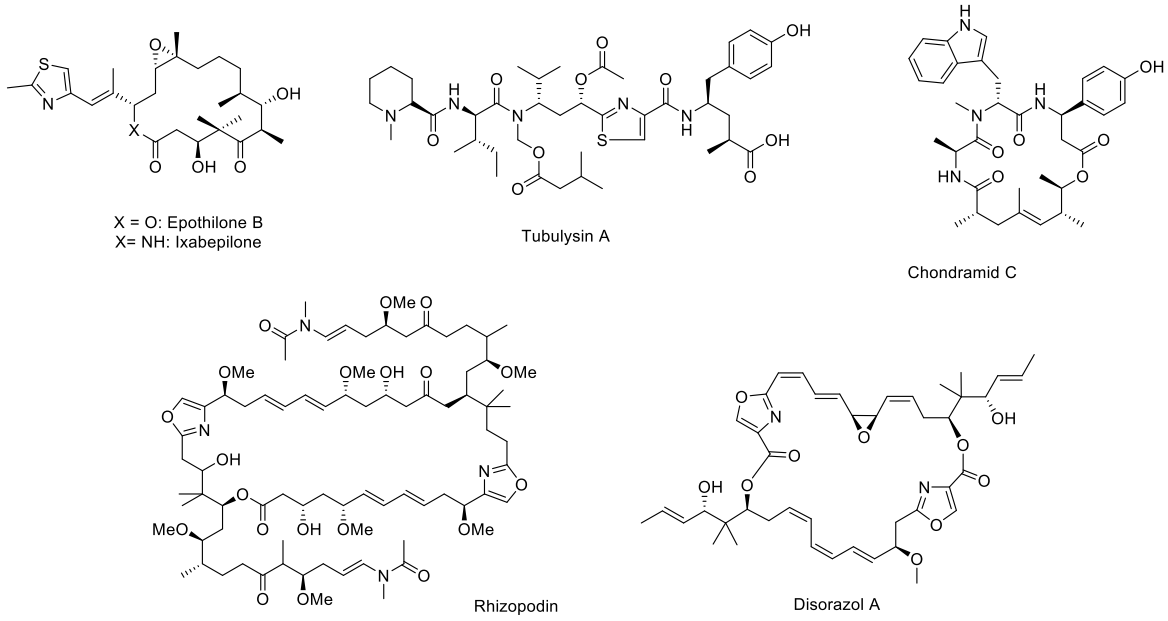


Figure 6: Structures of myxobacterial secondary metabolites interacting with the cytoskeleton

In addition to these cytotoxic substances, also antifungal and antibacterial myxobacterial compounds with various molecular targets have been isolated and investigated. Myxopyronin A,[44] sorangicin A[45] and ripostatin A[46] were found to inhibit the bacterial RNA polymerase; sorangicin A has the same binding site as the anti-tuberculosis drug rifampicin, the β-subunit of the the RNA polymerase, whereas myxopyronin and ripostatin target the so-called switch-region of bacterial RNA polymerase and can thus retain activity against rifampicin resistant pathogens.[47]

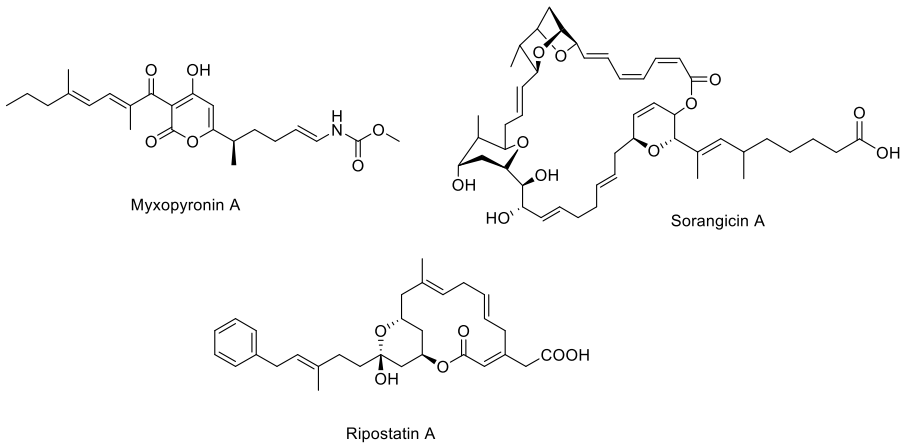


Figure 7 Structures of antibacterial myxobacterial compounds targeting the RNA polymerase

Many myxobacterial compounds show anti-fungal activity as for example myxothiazol[48], myxalamid[49], stigmatellin[50] and crocacin[51]. All four of them target the mitochondrial respiratory chain system.

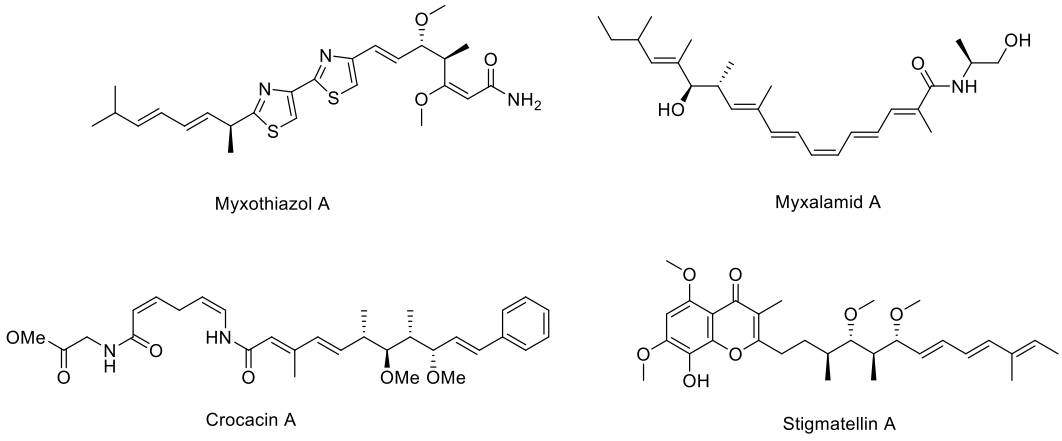


Figure 8: Structures of myxobacterial secondary metabolites exhibiting antifungal activities

Besides a number of secondary metabolites exhibiting antibacterial effects, also immunosuppressants like the cyclic peptide argyrin[52] were isolated from myxobacteria.

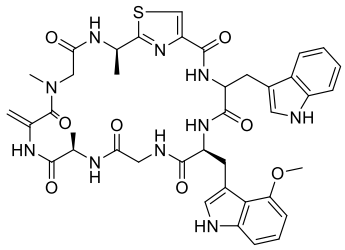


Figure 9: Structure of argyrin A from the myxobacterium Archangia gephyra

In conclusion it can be stated that the great structural variety found in secondary metabolites from myxobacteria as well as the diverse range of bioactivities found for these compounds qualifies myxobacteria as a great source for new bioactive compounds and future lead structures. Yet the potential of myxobacteria is still not exhausted as new strains from ecological niches are frequently isolated and in many cases promising bioactive compounds can be isolated and investigated.[53]

1.4 Natural products biosynthesis

A large number of bacterial secondary metabolites is produced by only two types of megaenzyme complexes: the non-ribosomal peptide synthetases (NRPS) and the polyketide synthases (PKS).[54,55] The genes encoding these pathways are generally – apart from the lately increasing number of reported exemptions - following the colinearity rule, indicating that the order of domains in the consecutive modules is identical to the order of enzymatic steps carried out during biosynthesis.[56,57] This fact makes natural product biosynthesis by PKS and NRPS systems predictable to a certain extent, as with a given structure the required enzyme set can be predicted and vice versa. Nevertheless, a certain vagueness in these predictions remains, as enzymes encoded up- and downstream the respective biosynthetic pathway might be involved in key steps of the biosynthesis. Despite the respective natural products are assembled from different monomeric building blocks, the biosynthetic logic of the two systems - PKS and NRPS - is fairly similar.[58] These megaenzyme complexes consist of so-called modules, whereas every module catalyzes a distinct elongation of the nascent linear natural product with a monomeric building block. Every module contains a set of domains which encode enzymatic functionalities responsible for chain extension and alteration of the resulting intermediate product. There is only a limited set of domains available but the diverse combination options of these functionalities in modules results in remarkable structural diversity; similarly certain modules can be complemented with special enzymatic functionalities giving rise to an even broader structural diversity.[59] The NRPS and PKS natural product assembly lines are thus build up from a certain set of modules which are aligned next to each other like pearls on a string. The strategy employed for chain elongation during biosynthesis is also very similar in the two systems although the chemical reactions involved in chain elongation differ. The linear intermediates of the biosynthesis are tethered to carrier proteins (CP's) which shuttle the intermediates through the assembly line. The inactive apo form of these carrier proteins requires post-translational modification for conversion into the active holo-form. For this, transfer of the 4'-phosphopantetheine moiety of conezmye A onto the active site serine of the carrier proteins is conducted by 4'phosphopantetheine transferases.[60] Due to their relative flexibility, the so activated carrier proteins allow for correct positioning and thus efficient processing of the intermediates by the domains located in the respective module. The monomeric building blocks are as well activated prior to conjunction with the CP—

tethered intermediate. After undergoing the number of elongation steps provided by the modules on the PKS/NRPS enzyme complex, release of the intermediates from the assembly line is also realized in a similar manner. The chain-terminating thioesterase domain catalyzes the off-loading of the linear precursor molecule. Various release modi are known for TE domains as they are able to detach the linear intermediate by hydrolysis leading to a linear natural product or by macrocyclization reactions leading to formation of the respective lactone or lactam form of the compound.[61]

1.4.1 Non- ribosomal peptide synthetases - NRPS

Non ribosomal peptides are synthesized independent from the ribosome and without a nucleic acid template. Instead, multimodular enzyme complexes, the non-ribosomal peptide synthetases (NRPS) are at the same time template and biosynthetic machinery for their production. In principle the NRPS megaenzyme complex consists of several modules, all of which comprise a distinct set of catalytic domains responsible for incorporation and modification of amino acid monomers.[56,62] The sequence of the final peptide product directly results from the number and order of catalytic domains in the modular NRPS assembly line. Every elongation module requires at least a minimal set of domains to perform substrate selection, substrate binding and finally elongation of the growing peptide chain by peptide bond formation. The minimal module consists of an adenylation domain (A domain), a peptidyl carrier protein (PCP domain) and a condensation domain (C domain).[63] For chain extension, the specific substrate is selected by the adenylation domain and subsequently activated; the substrate specificity of these domains is determined by the architecture of their substrate binding pocket where ten amino acid residues are responsible for substrate binding.[64] For activation, the A domain then generates the respective amino acyl adenylate using ATP. This reactive intermediate is afterwards transported to the phosphopantetheinyl arm (Ppant arm) of the acyl carrier protein of the same module. The Ppant arm is a post-translational modification of the peptidyl carrier protein that is attached to a highly conserved serine residue of the domain and thus converts the apo PCP into its holo form. The holo PCP serves as a “swinging arm, which transports the thioester bound biosynthesis intermediates to the catalytic sites of the NRPS domains.[65] In the following biosynthesis step, the PCP bound substrates of two neighboring modules are linked via an amide bond formation catalyzes by the condensation domain. During this condensation step, a

nucleophilic attack onto the activated upstream PCP-bound donor amino acid is carried out by the α-amino acid of the downstream PCP-bound acceptor substrate. After elongation the resulting peptidyl intermediate is transported to the next module for further processing, whilst the vacant PCP domain is available for loading of a new substrate molecule.[66]

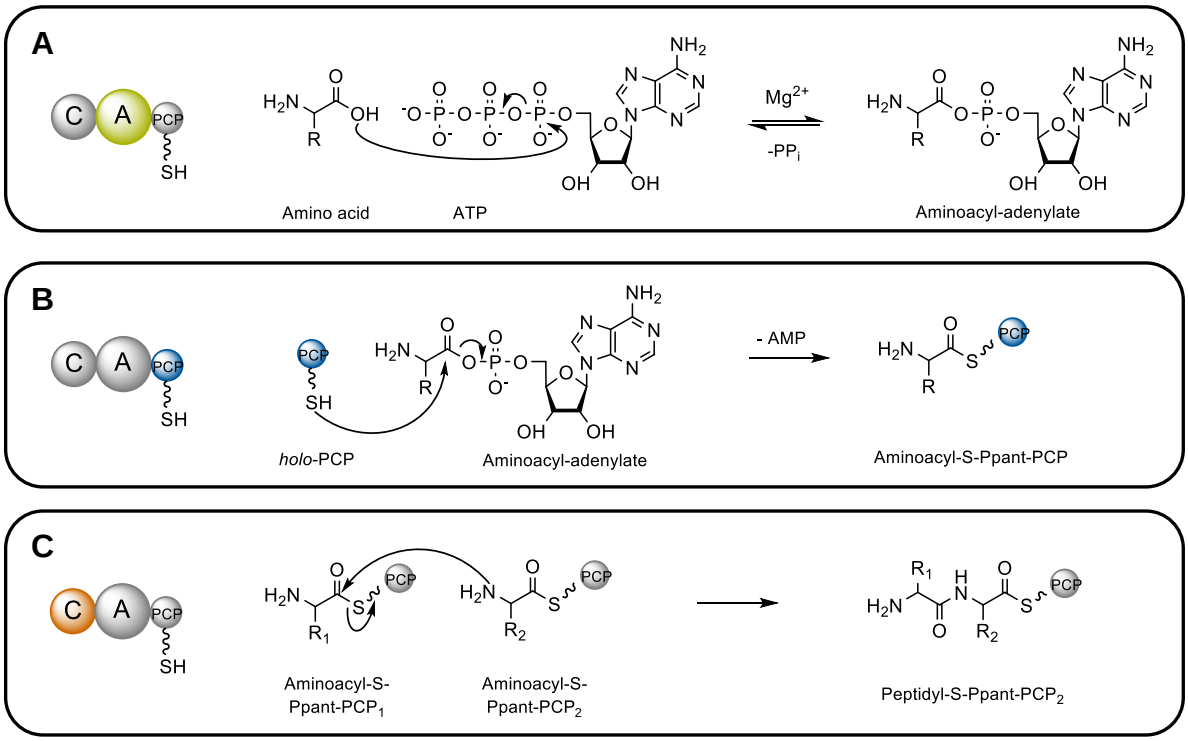


Figure 10: Schematic depiction of peptide biosynthesis catalyzed by non-ribosomal peptide synthetases. A.: activation of amino acids with ATP catalyzed by the A domain. B.: transfer of the activated amino acid onto the ppant arm of the PCP domain. C.: Amide bond formation catalyzed by the C domain.

One elongation circle in an NRPS system thus requires a module comprising three domains: A-PCP-C. Biosynthesis in NRPS starts with a so called loading- or initiation module which comprises of only an A domain and a PCP domain and is responsible for loading of the starter molecule onto the assembly line. Chain termination is usually carried out by a thioesterase domain located downstream the last elongation module. The TE domain is responsible for release of the peptide chain; depending on the reaction catalyzed by the TE either a linear peptide (hydrolysis) or a macrolactam (intermolecular cyclization) is formed.[67,68] The great structural variety seen in NRPS products results from tailoring reactions which can take place either on the NRPS assembly line or after chain termination; also the introduction of unusual building blocks contributes to the diversity of NRPS products.[69,70] For tailoring reactions on the assembly line, there is a set of auxiliary domains which are directly integrated in NRPS modules. The

epimerization domain for example catalyzes the conversion of L-amino acids into the respective D-configured amino acids[71]. For methylation of the peptide, there are two possible scenarios: an N-methylation is carried out with the help of SAM-dependent N-methyltransferases whilst C- methylation requires a C-methyltransferase domain.[72] During C-methylation, the PCP-Ppant bound amino acid is methylated at the α-C atom of of cysteine which can then be converted into the respective α-methyl thiazoline by a cyclization domain. Cyclization domains are able to replace condensation domains in several cases and are known to catalyze heterocycle formation when integrated in modules that select cysteine, serine or threonine residues.[73] Also these heterocycles can further be modified by either oxidation or reduction domains. Oxidation domains convert thiazoline and oxazoline residues into the corresponding thiazoles or oxazoles whilst NADH dependent formation of thiazolidine and oxazolidine rings is carried out by reduction domains. Also a number of post-assembly tailoring reactions is known[74]. The NRPS product can be modified in multiple ways, for example by free standing cytrochrome P450 proteins or a glycosylation with the help of glycosyltransferases can take place[75,76].

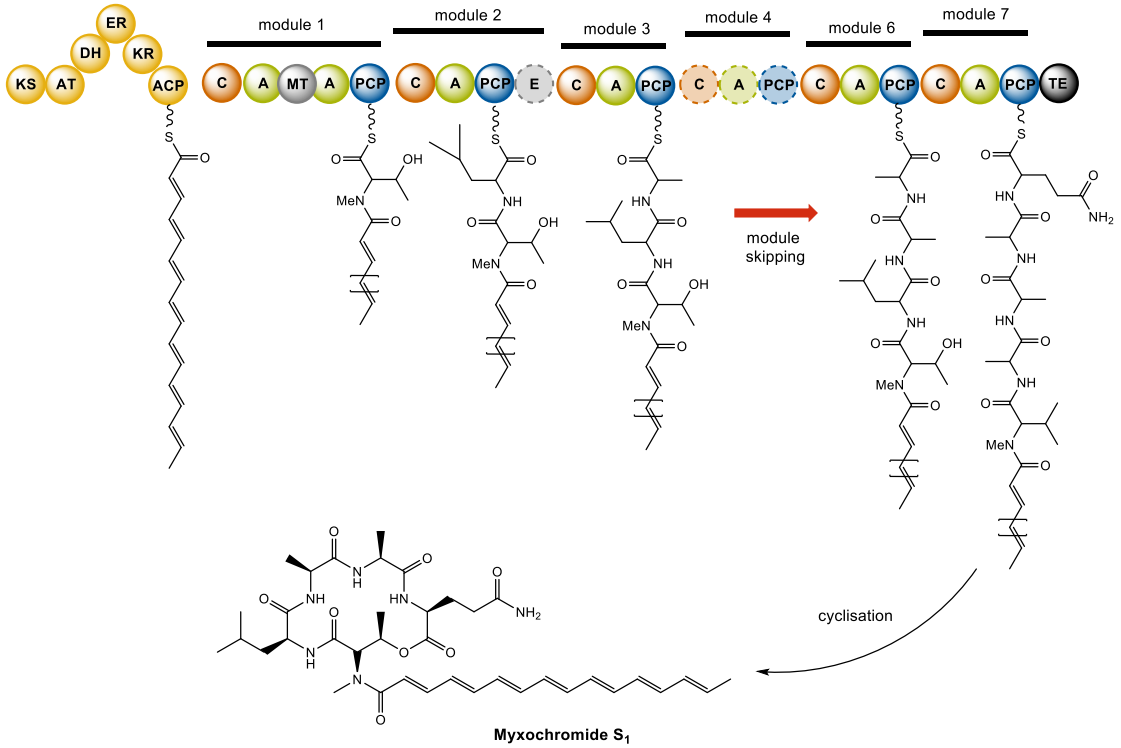


Figure 11 Model for the biosynthesis of myxochromide in *S. aurantiaca* DW4/3-1 (adapted from Wenzel et al., ChemBioChem 2005)[77]. The biosynthetic machinery consists of an iterative type I PKS (see also section 1.5.2.2) and an NRPS: Dashed and faded domains are presumably inactive. The biosynthesis of the NRPS products starts with loading of the polyketide moiety.

1.4.2 Polyketide synthases – PKS

Analogous to non-ribosomal peptides, polyketides are produced following a similar biosynthetic concept. Short chain carboxylic acid precursor molecules are connected on a modular biosynthetic assembly line whilst traversing several tailoring steps catalyzed by specific domains present in the modules. The final linear polyketide is detached from the assembly line in the last module and can undergo cyclization as well as several tailoring reactions. The most straightforward representative of PKS family is the modular type I polyketide synthase which strictly follows the colinearity rule, meaning the sequence of the polyketide corresponds with the domain sequence in the megaenzyme complex. The best studied example for this class of polyketide synthases is the archetypal erythromycin PKS 6-deoxyerythronilid synthase (6-DEBS, see figure 12).[78] In derogation thereof, several other PKS types are known: type II iterative PKS and type II PKs. In type II iterative PKS a single set of enzymatic functionalities is present – only one module so to say – and iteratively catalyzes the composition of the polyketide product. Type III PKs, which are also called chalcone synthases are homodimers of ketosynthase domains that catalyze precursor assembly independent from carrier proteins. In addition to these variants, also system hybrids are known as well as special forms of the three types. One considerable special case are the so called *trans*-AT PKS systems, which are lacking the domain for precursor selection in all of their modules. Several scenarios for the replacement of this functionality are known and discussed in detail below.

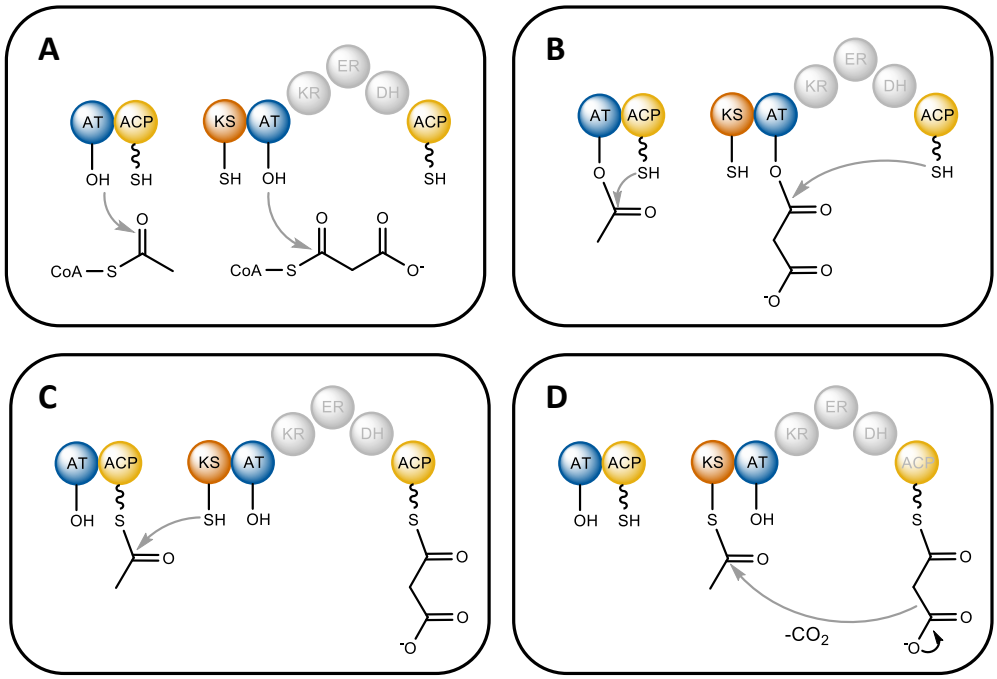


Figure 12: Schematic depiction of reaction steps during polyketide and fatty acid assembly. The AT-ACP di-domain represents a loading module commonly found in PKS systems but not in FAS. A.: selection of activated precursor molecules by AT domains. B.: transesterification of the precursors onto the ppant arm of ACP domains. C.: Passing on of the starter unit onto the KS domain D.: Decarboxylative condensation reaction catalyzed by the KS domain.

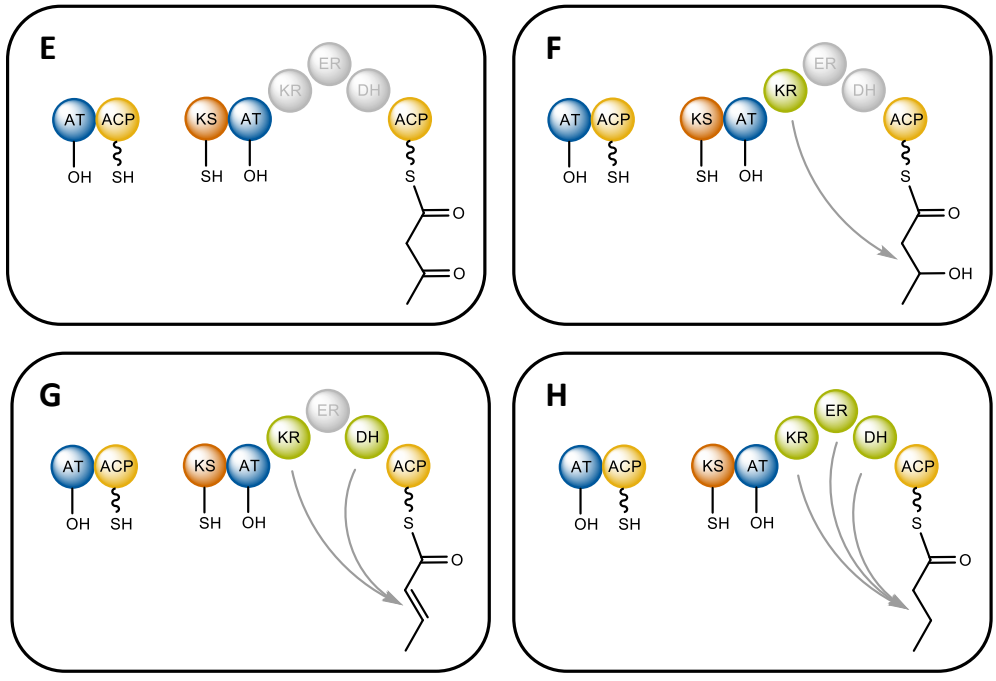


Figure 13: Schematic depiction of β-keto processing proceeding during polyketide biosynthesis. E.: No β-keto processing in absence of reductive domains. F.: Reduction of the keto group to an alcohol catalyzed by a KR domain. G.: Production of an α,β-olefinic moiety upon the successive action of a KR and a DH domain. H.: Fully reduced polyketide intermediate resulting from catalytic activity of a complete reductive loop consisting of KR, DH and ER domain.

1.4.2.1 Type I PKS

In type I PKS systems, the number of functional modules in the biosynthetic machinery reflects the number of building blocks incorporated into the polyketide product.[54] A minimal module of type I PKs consist of an acyltransferase domain (AT), a ketosynthase domain (KS) and an acyl carrier protein (ACP). AT domains in each module are responsible for the selection of extender units as well as for the selection of the starter unit in the loading module. Diverging from non-ribosomal peptide biosynthesis, the selected precursor molecules are not activated under ATP consumption, but pre-activated CoA esters of the respective building block are selected. Malonyl-CoA for example is formed from acetyl-CoA with the help of the acetyl-CoA carboxylase. The transacylation of the activated extender units from CoA onto the phosphopantetheine arm of the ACP domain in the same module is catalyzed by the AT domain. The AT domains usually possess a stringent specificity for one extender unit; most commonly either malonyl-CoA or methylmalonyl-CoA are selected. The incorporation of these extender units takes place via a ping-pong bi-bi mechanism where the formed acyl-AT intermediate undergoes a nucleophilic attack of the thiol group of the phosphopantetheine arm of the ACP domain.[79,80] The alignment and phylogenetic analysis of known AT domains revealed that the domains group according to their substrate specificity even across species barriers.[81] Also a range of unusual extender domains such as ethylmalonyl-CoA, aminomalonyl-CoA, hydroxymalonyl-CoA or ethoxymalonyl-CoA can be selected by AT domains.[82] In the following, the ACP bound precursor molecule is available for extension of the growing polyketide chain. Prior to chain extension, the ACP bound polyketide intermediate of the upstream module is transferred to the conserved cysteine residue of the KS domain. The KS then catalyzes a decarboxylative thioester condensation in a Claisen condensation fashion between the extender unit and the KS-bound intermediate.[83] The result of this condensation is the respective β-keto-acyl-S-ACP intermediate.[84] In addition to the minimal set of catalytic domains (KS-AT-ACP), there is also a set of reducing domains which process the polyketide intermediate at its β-keto position. The optional incorporation of these domains into PKS modules gives rise to the very broad structural diversity seen in polyketides. The so-called “reductive loop” includes a ketoreductase domain (KR), a dehydratase domain (DH) and an enoylreductase domain (ER). The KR domain reduces the β-keto function to the respective β-hydroxyl-intermediate under consumption of NADPH.[85] The resulting stereochemistry of the β-

hydroxyl-intermediate is defined by the KR domain in which sequence motifs and fingerprints provide indications for stereochemistry predictions. A-type KR domains exhibiting a conserved tryptophan produce S-configured alcohols whilst B-type KRs showing an LDD motif produce R-alcohols.[86] Lately the crystal structure of a KR domain from the tylosin pathway in comparison to that of the erythromycin pathway revealed even more motifs and fingerprints responsible for employment of a certain stereochemistry in polyketides.[86,87] Four different types of reduction-competent KR domains are known and their polyketide substrates are guided into the active site through specific key residues. Thus the four types of KR domains are able to form different combination of α-substituents and β-hydroxyl group chiralities. In addition to that two non-reducing KR domains are known which only control the stereochemistry of the α-substituent. The combination of a KR and a DH domain in the reductive loop leads to dehydration of the alcohol formed by the KR, thus the respective α,β-enoyl moiety is formed.[88] If an additional ER domain is present in the same module, the α,β-enoyl moiety is further reduced to a methylene moiety in an NADPH dependent reaction. In this process the enoylreduction influences the stereochemistry at the respective α-carbon.[89] Recapitulatory the β-position of polyketides can appear as a keto function if no reductive loop is present in the module or as hydroxyl-, α,β-enoyl- or methylene moiety depending on the presence and combination of domains in the reductive loop.

The starter unit of type I PKS systems is loaded by a so called loading module which consists either of an AT-ACP di-domain or a KS-AT-ACP tri-domain. In case of an AT-ACP architecture, the AT domain selects and transfers its substrate directly onto the ACP. Also in these loading modules the AT domain exhibits a certain specificity and can load short straight-chain, branched-chain or aromatic monocarboxylic acids. If a KS-AT-ACP tridomian serves as loading module, either malonyl-CoA or methylmalonyl-CoA are selected and transferred to the ACP domain, followed by KS mediated decarboxylation resulting in acetate- or propionate starter units. In this case the KS domain, which was found to be similar to the chain-length factor in fatty acid biosynthesis, belongs to the group of KS^Q domains, in which the active site cysteine is replaced by a highly conserved glutamine. Thus only decarboxylation but no condensation reactions are carried out by the KS in this context.[90]

The termination of the fully elongated polyketide chain and thus its release from the assembly line is commonly carried out by a thioesterase domain. The chain release is carried out in a two-step reaction where the polyketide chain is first loaded onto the TE domain and subsequently released after nucleophilic attack of an exogenous or an intramolecular nucleophile. Attack of an exogenous nucleophile leads to hydrolysis or transesterification whilst attack of intramolecular O-, N-, or C-nucleophiles leads to formation of macrolactones, macrolactames or Claisen like condensation.[91]

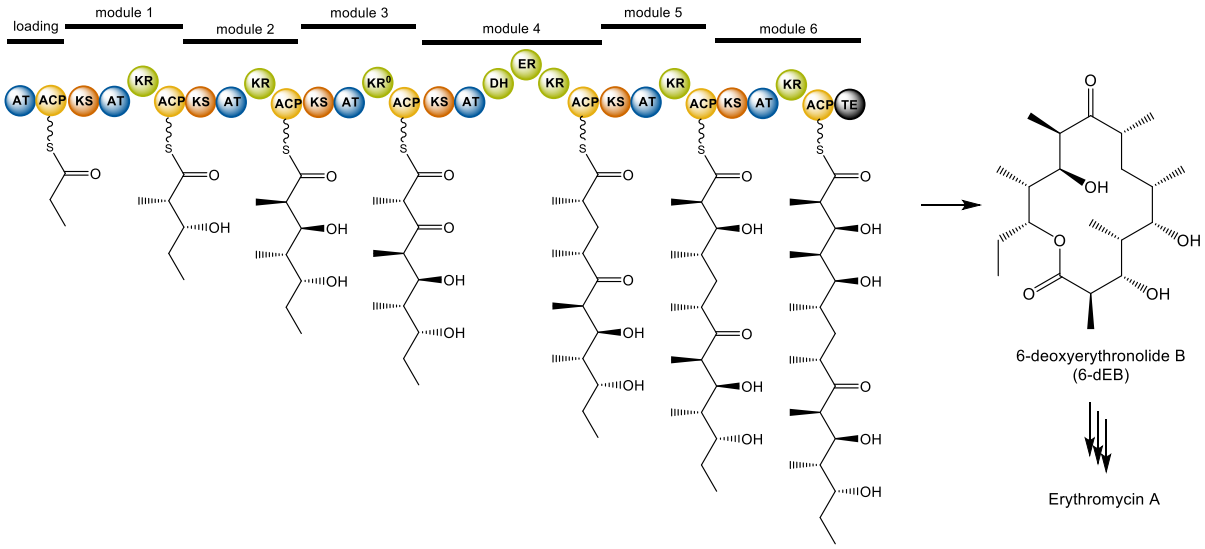


Figure 14 Schematic representation of the polyketide synthase responsible for erythromycin biosynthesis (adapted from Cane, J. Biol. Chem., 2010).[92] On this archetypal type I PKS the erythromycin precursor 6-deoxyerythronolide B is assembled and cyclized. Several post-PKS tailoring steps (methylation and glycosylation recations) lead to the final product erythromycin.

1.4.2.2 Iterative type I PKS

Iterative type I PKS systems are especially found in fungi as for example the lovastatin biosynthetic machinery from *Aspergillus terreus*.[93] However, iterative type I PKS have also been discovered in bacteria where they are responsible for biosynthesis of aromatic metabolites.[77,94,95] depending on the presence or absence of reducing domains, these iterative type I PKS are classified as non-reducing, partially reducing or highly reducing PKSs. Nevertheless, the presence of reducing domains does not imply their usage in every step of the biosynthesis as the degree of reduction can vary in every elongation step

despite the iterative action of the PKSs. The same holds true for methyltransferase domains in the PKS module. [96,97]

1.4.2.3 *trans*-AT PKS

The *trans*-AT PKS is a special case of type I polyketide synthases in which the acyltransferase domain is lacking in every module and is instead replaced by a discrete AT domain[98]. This AT domain is mostly encoded in close proximity to the PKS gene cluster. The bacillaene biosynthetic pathway was the first representative of *trans*-AT PKS;[99,100] however since the gene cluster seemed to be silent and the sequence of the strain was incomplete and hardly accessible at that time it could not be ruled out that the found gene cluster was a non-functional relict. Pederin then emerged as the first polyketide which could be attributed to a trans-AT PKS.[101] During investigations of pederin biosynthesis it was hypothesized for the first time that all modules in the megaenzyme were loaded by only one AT domain, which could finally be proven exemplarily for the leinamycin PKS.[102] It was shown in vitro that the same AT domain is used iteratively to load all modules with extender units which in turn means that only one type of building block can be chosen: malonyl-CoA.[98] AT functionalities can be encoded either as single genes, fused as tandem-AT domains or fused with an oxidoreductase which acts as trans-enoylreductase.[103] When first tandem AT domains from *trans*-AT PKS were reported it was considered, that they were responsible for loading of different extender units. But further investigations revealed that only one of the domains loads extender units whilst the second one carries out proof-reading during biosynthesis. Alignments of the respective AT domains yielded two clades in this context: the malonyltransferase type (AT1) and the acylhydrolase type (AT2). AT domains from the acylhydrolase clade were found to remove stalled extender units from the assembly line.[104,105] While in cis-AT systems (non-iterative type I PKS) the theoretically possible combinations of domains results in maximum of only eight architecturally different modules, a wide variety of different domain sets is known from *trans*-AT PKS. Phylogentic analysis of both systems thereupon revealed that *cis*- and *trans*-AT PKS evolved independently based on fatty-acid synthases.[106] *Cis*-type PKS thus resulted from gene duplication events and further diversification of modules. Contrarily *trans*-AT PKS are rather tessellated systems assembled from multiple gene segments which possibly originate from horizontal gene transfer between different bacterial species. Other than *cis*-

AT PKS, the *trans*-AT systems show deviations from colinearity rule to a greater or lesser extent. This circumstance complicates gene cluster analysis and assignment of chemical scaffolds substantially. On the other hand it could be shown that KS domains from *trans*-AT PKS show rather tight substrate specificity. In a phylogenetic analysis involving the KS domains of 13 different *trans*-AT PKS, those domains processing similar substrates formed distinct clades.[107] This finding also allows substrate prediction for a given set of KS domains when aligned and compared with the surveyed data. Homology models of the KS substrate binding site and docking of substrate structures also revealed several key residues responsible for substrate specificity.[108] As the choice of extender units is limited in *trans*-AT PKS, variations in the resulting polyketides are for example achieved by introduction of α- and β- methyl-branches into the growing polyketide chain. α-methylation is carried out by methyltransferase domains, which are incorporated in the respective module. β-branching however is conducted by a set of enzymes called "β-branching cassette".[109] These cassettes consist of an ACP domain, a ketosynthase, an HMG-CoA synthase and two enoyl-CoA hydratases. Optionally an AT domain as well as tandem-ACP domains can be found in β-branching cassettes. The β-position in contrast to the α-position is electrophilic and thus requires a nucleophilic alkyl source. β-branch incorporation starts with the loading of a malonyl unit onto a free- standing ACP domain catalyzed by the AT domain, followed by decarboxylation by the KS. The ACP-bound acetyl moiety then serves as nucleophile during β-branch incorporation in which the HMG-CoA synthase catalyzes the aldol attack of the acetyl enolate onto the assembly line tethered β-ketothioester. The resulting intermediate is the sequentially dehydrated and decarboxylated by the two enoyl-CoA hydratases to yield the final β-branched polyketide. Especially in this context the function of tandem-ACP domains has been extensively discussed which resulted in several hypotheses for their function. For the bacillaene biosynthesis it has been speculated that the tandem-ACP di-domain serves as a β-branch-acceptor scaffold whilst ACP knockout experiments in the mupirocin PKS hinted that multiplied ACP domains might increase the overall biosynthetic flux through the pathway as they possibly serve as way station for β-branch incorporation.[100,110] In the curacin biosynthetic pathway a dimerized ACP tri-domain was found in a module conducting a rather complicated enzymatic reaction; heterologous expression of several versions of this module containing differing number of ACP domains combined with in vitro assays provided further insights. It could be shown that the presence of all three ACP domains increases the enzymatic reaction efficacy whereupon it was concluded that multiplication

of ACP domains promotes the formation of multienzyme complexes.[111] Despite these interesting findings, it has to be noted that ACP domain multiplication is not unique in trans-AT PKS but has also been reported from PUFA-, naphtopyrone-, and sterigmatocystin biosynthetic gene clusters where their function also is under investigation.[112,113] In summary, it can be said that the exact function of ACP duplications is still elusive, although the current results suggest that the yield in complex enzymatic reactions might be enhanced upon their presence. Another peculiarity seen in *trans*-AT PKS is the presence of dehydration bi-modules in addition to dehydratase domains which carry out dehydration in only one module. Two types of dehydration modules are described; the so called A-type comprises the domains KS-KR-ACP / KS^0-DH-ACP, whereas the second KS^0 is lacking its catalytic histidine residue and hence does not catalyze polyketide chain elongation. The A-type bi-module is responsible for the formation of Z-double bonds; its counterpart, the B-type bi-module introduces conjugates diene moieties in an unknown way whilst exhibiting a KS-KR-ACP / KS-DH-ACP-KR domain arrangement.[114] Also in polyketide biosynthesis by *trans*-AT PKS, a vast variety of post-PKS assembly modifications can be introduced once the molecule is released from the PKS.

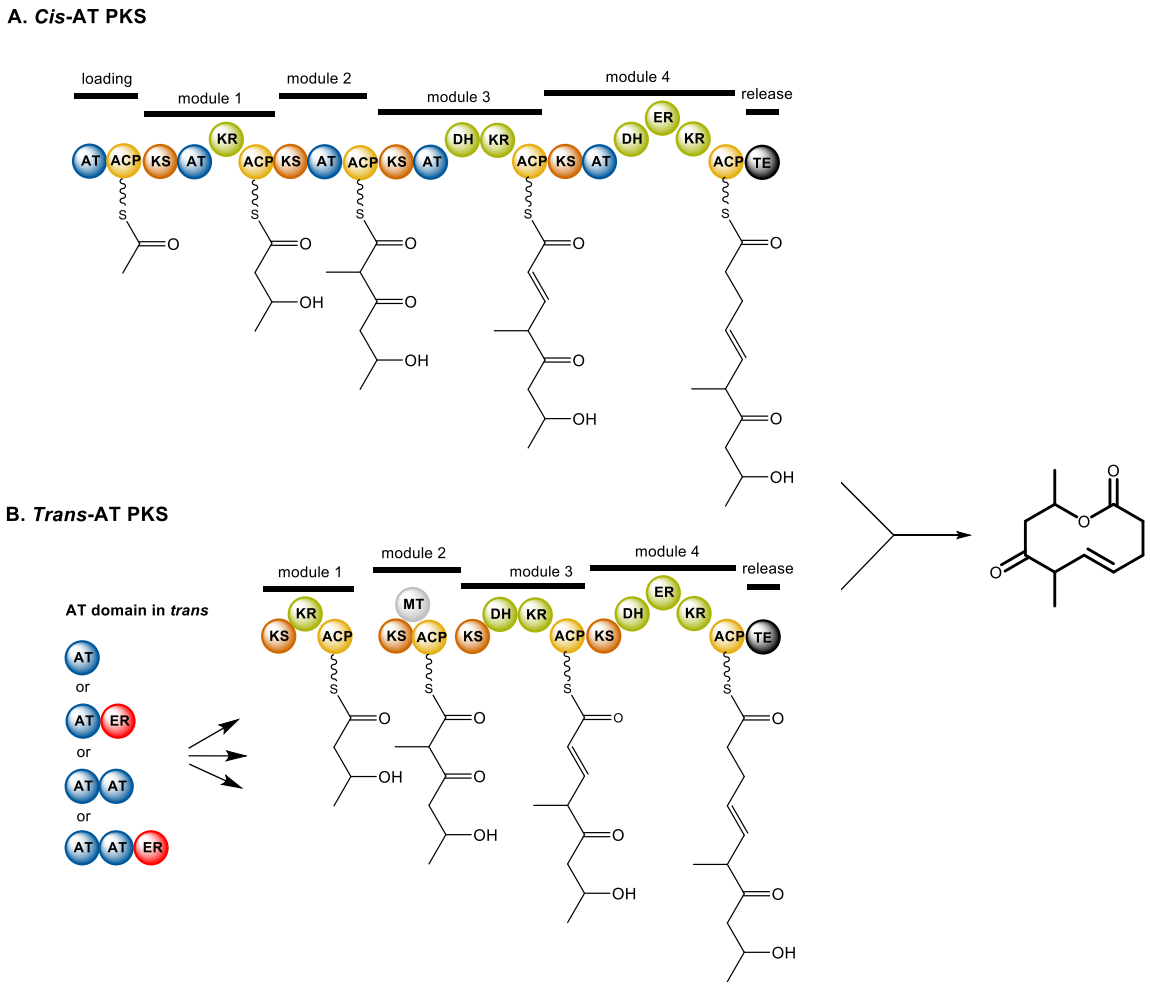


Figure 15 Simplified depiction of a hypothetical *cis*-AT- and a *trans*-AT PKS yielding the same fictional product (adapted from Piel, Nat. Prod. Rep., 2010).[114] The *cis*-AT PKS comprises one AT domain in every module whilst in *trans*-AT PKS a free-standing AT domain loads all modules. *Trans*-AT PKS AT domains can appear as single domains, coupled to ER domains or as tandem AT optionally coupled to ER domains. The origin of the methyl branch incorporated in module two differs in the two systems. In the *cis*-AT PKS PKS methyl branches stems from incorporation of methylmalonyl-CoA selected by the corresponding AT domain; in *trans*-AT PKS a C-methyltransferase domain is needed to introduce the α-methyl branch.

1.4.2.4 Iterative type II PKS

Aside from the rare case of iterative type I PKS systems, the iterative type II PKS are quite common in prokaryotes where they are mainly found in Gram-positive actinomycetes and only few known examples from Gram-negative bacteria were reported.[115,116] Type II PKS are responsible for biosynthesis of polyphenolic natural products which have considerable relevance in the clinic. For example doxorubicin which is widely used in anticancer therapy and tetracycline, a well-established antibiotic are

produced by type II PKS.[117,118] Type II iterative PKS are comprised of several individual enzymes which catalyze iterative decarboxylative condensation of activated malonyl extender units. The minimal PKS needed for this reaction consists of two ketosynthase domains (KS_α and KS_β) and one acyl carrier protein which represents an anchor for the growing polyketide chain. In addition to that, subunits such as ketoreductases, cyclases and aromatases can be part of the PKS and thus define the folding pattern of the polyketide chain. Also post-PKS tailoring reactions by oxygenases, glycosyl- and methyltransferases are known for type II PKS products.[119] The two KS domains of the minimal PKS which form a heterodimer, show high sequence similarities but KS_β in contrast to KS_α is lacking the active site serine which is crucial for polyketide linkage. Thus it seems obvious that the KS_α catalyzes C-C bond formation whilst the not so obvious role of the KS_β domain has been investigated intensely. It was shown that in KS_β of the actinorhodin and the tertacenomycin PKS, the active site serine is mutated to glutamine and that the two domains are involved in loading malonyl-CoA and formation of acetyl-KS from decarboxylation of malonyl-ACP.[90] This process has similarly been described for the condensation-deficient KS_Q found in type I polyketide synthases. The KS_β domain is also called ‘chain length factor’ (CLF) which implies its second important functionality: control of the polyketide chain length. A protein cavity formed in the KS-KS dimer most likely controls the size of the growing polyketide chain; additional results hint that distinct amino acids of the CLF serve as a gatekeeper in the polyketide tunnel and thus determine chain length.[120] For PKS priming in Type II PKS several starter units, such as propionate, (iso)butyrate, malonamate and benzoate are employed whilst for chain extension only malonate is used.[121]

1.4.2.5 Type III PKS

Type III PKS which are members of the chalcone synthase and stilbene synthase superfamily are typically associated with plants but also in bacteria and fungi these biosynthetic machineries have been discovered.[122,123] In higher plants the chalcone synthase initiates the flavonoid biosynthesis by decarboxylative condensation of three acetate units from malonyl-CoA onto the starter unit *p*-coumaryl-CoA which is derived from phenylalanine. The resulting phenylpropanoid triketide then undergoes intramolecular cyclization and aromatization. Type III PKS are structurally and mechanistically diverging from the previously described biosynthetic megaenzymes, as

they do not exhibit a modular structure and they act independent from phosphopantetheinylated ACP domains but use free CoA thioesters as substrates. Type II PKS act as homodimeric enzymes and use the same catalytic site to iteratively perform loading of the acyl-CoA intermediate onto the active site serine and chain extension by decarboxylative condensation of malonyl-CoA building blocks.[124]

1.4.2.6 Fatty acid biosynthesis

Fatty acid biosynthesis and the respective biosynthetic machinery – the fatty acid synthase (FAS) – shows high homology to PKS systems; in fact advances in the field of fatty acid synthases often lead the way for PKS research.[125] Fatty acids are synthesized from simple precursor molecules as well, whereas the underlying enzymology is closely related to polyketide biosynthesis. Iterative condensation of C2 units followed by reduction of the same is carried out until a certain chain length (usually 14, 16 or 18 carbons) is reached. A FAS consists of the domain set KS-ACP-KR-DH-ER-TE plus the so called malonyl-acteyl-transferase (MAT). The MAT domain transfers the starter unit acetyl-CoA and the extender units malonyl-CoA onto the Ppant arm of the ACP domains. Decarboxylative condensation of the starter- and the extender units is catalyzed by the KS domain; the resulting intermediate then runs through a full circle of reduction reactions carried out by the KR-DH-ER domains, yielding a saturated product.[125] This circle is repeated using the same set of enzymes until a fatty acid chain of certain length is obtained. Fatty acid synthases are divided in two classes: type I FAS, which are commonly found in mammals and fungi, exhibit all enzymatic functionalities on one single protein whereas type II FAS from archaea and bacteria use discrete monofunctional enzymes.[126,127] The type I fatty acid synthase can also be found in corynebacteria, mycobacteria and nocardia - the so called CMN-group bacteria – where they produce palmitic acid and in cooperation with type II FAS systems give rise to a broader range of lipids.[81] Saturated fatty acids play an important role in cell wall formation and fluidity and also serve as energy resource, signal molecules or precursors for other metabolites. In addition to saturated fatty acids, so called SAFAs, also branched chain fatty acids, monounsaturated acids MUFAs and polyunsaturated acids, PUFAs are known. Especially the latter ones are becoming increasingly interesting due to their beneficial health effects in humans.[128] The long chain polyunsaturated fatty acids eicosapentaenoic acid (EPA) and docosahexanoic acid (DHA) are essential nutrients but the typical sources for their

supply – such as sea fish – are limited.[129] Primarily, LC-PUFAs result from conversion of saturated fatty acids mediated by oxygen-dependent desaturases and elongases under aerobic conditions.[130] Alternatively, biosynthesis of PUFAs has also been reported from several marine microbes under strictly anaerobic conditions;[131–133] only recently also a PUFA biosynthetic machinery from terrestrial myxobacteria was reported, whereas this pathway has the potential to serve as a platform for large scale production of EPA and DHA.[134] The PUFA synthases present in these bacteria exhibit hybrid architecture with domains that are homologous either to fatty acid synthases or to PKS systems. Biosynthesis of the fatty acids is carried out *de novo* based on acyl-CoA precursors in an iterative process. According to the generally accepted biosynthesis model, the growing PUFA chain is assembled, while attached to an ACP domain via thioester linkage.[135] As already mentioned above, also PUFA synthases stand out through the presence of multiplied ACP domains.[136] It could be exemplarily shown, that every single domain of this multiplett is able to facilitate PUFA biosynthesis.[113] Summarizing the details about PUFA synthases, one might argue that these biosynthetic machineries represent a transition state between FAS and PKS, which has established itself as an independent pathway for LC-PUFA production.

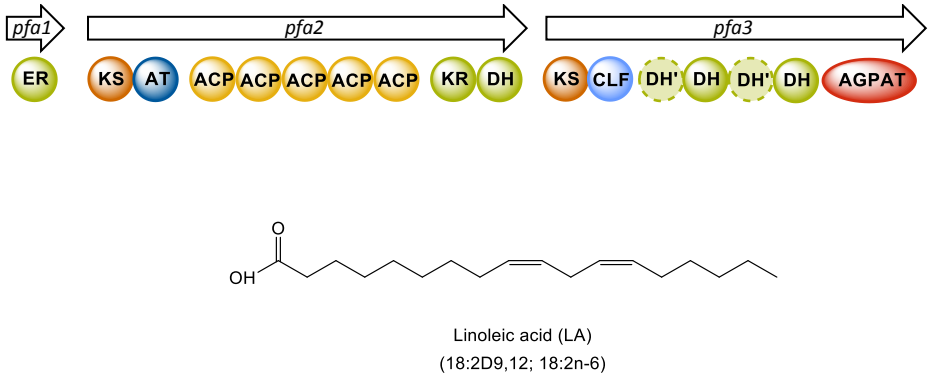


Figure 16 Schematic depiction of a PUFA biosynthetic gene cluster found in *Sorangium cellulosum* So ce56. The PUFA production is this strain is mainly restricted to linoleic acid with only trace amounts (< 1%) of eicosadienoic acid (EDA)- The DH domains shown are either FabA like dehydratases or pseudo DH domains (DH', dashed). The AGPAT domain represents a 1-acylglycerol-3-phosphate-*O*-acyltransferase which is a unique common feature of terrestrial myxobacterial PUFA synthases. Due to the function of this enzyme, - acylation of the 2-position of 1-acylglycerol-3-phosphate during glycerophospholipid biosynthesis – the AGPAT domain is thought to facilitate the direct transfer of the biosynthesized PUFAs into lipids.

1.4.2.7 Summary polyketide & fatty acid biosynthesis

When analyzing the different types of polyketide synthases and fatty acid synthases described in literature, one fact is particularly striking: rigorous assignment of the respective biosynthetic pathway to a certain type (type I PKS, iterative PKS etc.) or restriction of its presumed occurrence in a certain species (fungal PKS) were very often refuted by later observed exceptions to these rules. It is widely accepted, that PKS evolved from fatty acid synthases, but apparently during PKS evolution many transition states and also PKS-NRPS hybrid systems co-evolved. The discovery of *trans*-AT PKS which are seen as a special form of type I PKS complicated the picture, as in these systems the observed deviation from the colinearity rule disagrees with specifications for type I PKS systems. With the increasing amount of data collected about FAS, PKS and all observed special cases, a strict classification, as well as the establishment of specific criteria for assignment to these classes become more and more problematic. The “text book” classification schemes rather serves for rough division of biosynthetic pathways, but only a case-by-case assessment of the respective biosynthetic pathway provides the full picture.[137]

1.5 Synthetic biotechnology approaches for pathway engineering

Myxobacteria have frequently proven to be a rich source for new bioactive natural products; especially the genus *Sorangium* with its typically large genome holds the capacity for biosynthesis of a plethora of structurally diverse secondary metabolites. Despite their potential, the downside of myxobacteria as proficient producer of natural products is the difficult manageability of the strains which is particularly apparent for representatives of the genus *Sorangium.* Sorangium strains in most cases do have low doubling rates and are seldom amenable to genetic modifications; markerless mutations are hardly possible.[138,139] Thus the establishment of platforms for heterologous expression of the desired biosynthetic gene cluster is a key step which offers the opportunity for *in-depth* analysis of biosynthetic processes and also allows for alterations of the respective pathway. Furthermore strain improvement of the heterologous host used for this purpose can alter secondary metabolite production levels and thus facilitate large-scale production of the desired natural product.

The prime prerequisite to establish such a platform for heterologous expression is to mobilize the genetic information needed. There are several options to meet these requirements. In the “classic approach”, genomic DNA of the producer strain is digested and subcloned in order to produce a genomic library. These libraries can then be screened for the desired gene cluster using specific probes. In that way a native version of the gene cluster can be assembled from several smaller DNA fragments. A technique that highly facilitates these intentions is the RedET recombineering which enables defined modifications and also allows for instance the introduction of inducible promoter systems.[140] Using RedET, also variants of the native gene cluster can be produced in order to investigate the role of certain genes for the biosynthetic machinery.

An alternative approach which has gained more and more interest over the last few years, is to produce the genetic information via DNA synthesis. This technique only requires *in silico*-data of the desired biosynthetic pathway and works completely independent from the producer strain; solely good quality sequence data of the respective genomic locus is needed as a template for DNA synthesis. Despite this approach not being trivial at all, as it requires the expertise of several scientific disciplines, it opens the door to numerous options for analysis, modulation and alteration of biosynthetic gene clusters. For example the codon usage of the respective pathway can be adapted to the host’s demands, promoters can be easily exchanged and restriction site engineering allows for fast and straightforward removal or addition of specific genes. The design and construction of gene cluster variants can consequently be planned from the outset. Several reports about successful establishment of heterologous expression systems based on synthetic DNA underline the growing importance of this approach.[141–144]

Both options – assembly of the pathway from genomic DNA or synthesis of DNA fragments containing the desired biosynthesis genes – have their advantages and disadvantages. Whilst the “classic approach” is not as cost-intensive, DNA synthesis provides more flexibility as this technique is for instance not based on naturally occurring restriction sites. On the other hand the maximum size of DNA fragments that can be synthesized is still limited, which necessitates a certain workload for assembly of the pathway from smaller fragments. But the overall goal to design a platform for heterologous expression of a certain biosynthetic pathway can be achieved in both ways.

1.6 Outline of the presented work

The historical view on natural products as well as the currently precarious situation of growing resistance rates against antibiotics and the revival of once eradicated infectious diseases underline the importance of natural products in therapy of human pathogens. Nature offers a plethora of bioactive compounds but much remains to be done until clinical use is possible. Once isolation, structure elucidation and large scale production of a compound of interest is successfully done, its biosynthesis is one of the topics of interest. Looking further ahead, a cell-free biosynthesis factory, amenable to alterations as needed would be the ideal case. Still we are far from this goal, but the elucidation of biosynthesis pathways, the heterologous expression of such pathways in suitable host and the recent advances in synthetic biology contribute to the overall knowledge needed to achieve this aim.

This work is mainly about the myxobacterial polyketide chlorotonil A and its advancement towards enhanced pharmacodynamic and pharmacokinetic properties with the focus on possible clinical use as anti-malaria agent. Chlorotonil A is produced by several *Sorangium cellulosum* strains of which *Sorangium cellulosum* So ce1525 is the main producer with a 77 mg/l production rate. The macrolide compound exhibits an unusual *gem*-dichloro-1,3-dione substructure, which represents a novel structural feature among natural polyketides.[145] An analogous carbon skeleton was found in anthracimycin; this highly similar polyketide was isolated from a marine *Streptomyces* strain and differs from chlorotonil in its stereochemistry and the lacking chlorination pattern.[146] Whilst anthracimycin shows promising anti-bacterial activities against Gram-positive and Gram-negative bacteria, chlorotonil A – in addition to its antibacterial activity - stands out for its excellent anti-malarial activity *in vitro* and *in vivo*.[147,148] It could be shown that with a very early onset, chlorotonil A acts against all blood stages of the disease, while simultaneously being less toxic than the standard treatment with chloroquin. Furthermore, the compound was also active against chloroquin-resistant malaria parasites. The aim of this work was to elucidate the biosynthetic machinery responsible for chlorotonil A production and to create a platform for heterologous expression of the responsible biosynthetic pathway. In addition a detailed comparison with the biosynthetic gene cluster responsible for anthracimycin production in the marine *Streptomyces* strain was carried out. Chapter two and three report the endeavors towards

elucidation of chlorotonil and anthracimycin biosynthesis and explain the strategy for development of a platform for heterologous production of chlorotonil A. One major drawback of chlorotonil A despite its striking bioactivity is the high lipophilicity and the resulting low bioavailability of the compound. This issue was addressed with two major strategies: the development of a formulation suitable for oral chlorotonil administration and the chemical alteration of the compound towards less lipophilic derivatives. Chapter four in this context describes semi-synthesis of two new chlorotonil A congeners and their testing in several anti-bacterial and an anti-malarial assay.

Another topic covered by this thesis is the biosynthesis of the myxobacterial metabolites maracen and maracin in *Sorangium* species. Maracin, an ethinylvinylether and maracen, an α-chlorodivinylether, both isolated from *Sorangium cellulosum* So ce1128 show good activity against mycobacteria which further triggered the interest in the compounds.[149] Due to their structural resemblance to polyunsaturated fatty acids, a linkage between PUFA biosynthesis and the biosynthesis of maracen and maracin was supposed. Previous work on a myxobacterial PUFA synthase corroborated this hypothesis, as *Sorangium* strains showing maracen/maracin production were found to also hold the PUFA synthase gene clusters responsible for EPA and DHA production.[134] Other than the conservative, gene disruption based approach for elucidation of biosynthesis mechanisms, in this case a synthetic biology approach was chosen to gain insights in maracen/maracin biosynthesis. Chapter five describes in-depth analysis of the alleged maracen/maracin biosynthetic gene cluster and reports the constructional design of a construct for heterologous expression of this biosynthetic pathway.

1.7 References

[1] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, et al., *Nat. Biotechnol.* **2007**, *25*, 1281–1289.

[2] O. Lafont, **2007**, 209–216.

[3] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–35.

[4] a L. Demain, a Fang, *Adv. Biochem. Eng. Biotechnol.* **2000**, *69*, 1–39.

[5] Y. Xiao, X. Wei, R. Ebright, D. Wall, *J. Bacteriol.* **2011**, *193*, 4626–4633.

[6] M. J. Kirisits, M. R. Parsek, *Cell. Microbiol.* **2006**, *8*, 1841–1849.

[7] F. E. Koehn, G. T. Carter, *Nat. Rev. Drug Discov.* **2005**, *4*, 206–220.

[8] L. Pasteur, *Bull. Acad. Natl. Med.* **1878**, *2nd series*, 432–453.

[9] R. Koch, Z. Ätiologie, **1881**.

[10] R. Koch, D. Ä. Der Tuberkulose, in (Ed.: J. Schwalbe), Robert Koch-Institut, **1881**.

[11] P. Ehrlich, *Arch. für mikroskopische Anat.* **1877**, *13*, 263–277.

[12] P. Ehrlich, A. Bertheim, *Berichte der Dtsch. Chem. Gesellschaft* **1912**, *45*, 756–766.

[13] A. Fleming, *Proc. R. Soc. London Ser. B* **1922**, *652*, 298–306.

[14] A. Fleming, *Br. J. Exp. Pathol.* **1929**, *10*, 226–236.

[15] C. T. Walsh, T. a Wencewicz, *J. Antibiot. (Tokyo).* **2013**, *67*, 7–22.

[16] B. Spellberg, *Clin. Infect. Dis.* **2008**, *47*, 293–294.

[17] K. E. Jones, N. G. Patel, M. a Levy, A. Storeygard, D. Balk, J. L. Gittleman, P. Daszak, *Nature* **2008**, *451*, 990–993.

[18] A. E. Clatworthy, E. Pierson, D. T. Hung, *Nat. Chem. Biol.* **2007**, *3*, 541–548.

[19] L. B. Rice, *J. Infect. Dis.* **2008**, *197*, 1079–1081.

[20] H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, J. Bartlett, *Clin. Infect. Dis.* **2009**, *48*, 1–12.

[21] O. Kayser, Müller, R.H., Eds. , *Pharmaceutical Biotechnology - Drug Discovery and Clinical Applications*, WILEY-VCH, Weinheim, **2008**.

[22] A. A. Brakhage, V. Schroeckh, *Fungal Genet. Biol.* **2011**, *48*, 15–22.

[23] E. M. F. Mauriello, T. Mignot, Z. Yang, D. R. Zusman, *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 229–249.

[24] D. R. Zusman, A. E. Scott, Z. Yang, J. R. Kirby, *Nat. Rev. Microbiol.* **2007**, *5*, 862–72.

[25] H. Reichenbach, *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 149–156.

[26] K. Han, Z. Li, R. Peng, L. Zhu, T. Zhou, L. Wang, S. Li, X. Zhang, W. Hu, Z. Wu, et al., *Sci. Rep.* **2013**, *3*, 2101.

[27] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, a S. Durkin, J. a Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15200–15205.

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

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

[30] A. E. Oxford, *J. Bacteriol.* **1947**, *53*, 129–138.

[31] A. Plaza, M. Rolf, R. Müller, in *Nat. Prod.*, John Wiley & Sons, Inc., **2014**, pp. 103–124.

[32] K. Buntin, K. J. Weissman, R. Müller, *Chembiochem* **2010**, *11*, 1137–46.

[33] K. Gerth, N. Bedorf, G. Höfle, H. Irschik, H. Reichenbach, *J. Antibiot. (Tokyo).* **1996**, *49*, 560–3.

[34] D. M. Bollag, P. a McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, C. M. Woods, *Cancer Res.* **1995**, *55*, 2325–2333.

[35] S. Goodin, M. P. Kane, E. H. Rubin, *J. Clin. Oncol.* **2004**, *22*, 2015–25.

[36] P. Fumoleau, B. Coudert, N. Isambert, E. Ferrant, *Ann. Oncol.* **2007**, *18*, 9–15.

[37] V. Akbari, S. Moghim, M. R. Mofid, *Avicenna J. Med. Biotechnol.* **2011**, *3*, 167–175.

[38] J. T. Hunt, *Mol. Cancer Ther.* **2009**, *8*, 275–281.

[39] K.-H. Altmann, F. Z. Gaugaz, R. Schiess, *Mol. Divers.* **2011**, *15*, 383–99.

[40] M. W. Khalil, F. Sasse, H. Lünsdorf, Y. a. Elnakady, H. Reichenbach, *ChemBioChem* **2006**, *7*, 678–683.

[41] Y. a. Elnakady, F. Sasse, H. Lünsdorf, H. Reichenbach, *Biochem. Pharmacol.* **2004**, *67*, 927–935.

[42] H. Waldmann, T.-S. Hu, S. Renner, S. Menninger, R. Tannert, T. Oda, H.-D. Arndt, *Angew. Chemie* **2008**, *120*, 6573–6577.

[43] T. M. a Gronewold, F. Sasse, H. Lünsdorf, H. Reichenbach, *Cell Tissue Res.* **1999**, *295*, 121–129.

[44] H. Irschik, K. Gerth, G. Hofle, W. Kohl, H. Reichenbach, A. Mikrobiologie, A. Naturstoffchemie, **1983**, *XXXVI*.

[45] H. Irschik, R. Jansen, K. Gerth, G. Höfle, H. Reichenbach, *J. Antibiot.* **1986**, *XL*, 7–13.

[46] H. Irschik, H. Augustiniak, K. Gerth, G. Höfle, H. Reichenbach, *J. Antibiot. (Tokyo).* **1995**, *48*, 787–792.

[47] J. Mukhopadhyay, K. Das, S. Ismail, D. Koppstein, M. Jang, B. Hudson, S. Sarafianos, S. Tuske, J. Patel, R. Jansen, et al., *Cell* **2008**, *135*, 295–307.

[48] G. Thierbach, H. Reichenbach, *Antimicrob. Agents Chemother.* **1981**, *19*, 504–507.

[49] B. Silakowski, G. Nordsiek, B. Kunze, H. Blöcker, R. Müller, *Chem. Biol.* **2001**, *8*, 59–69.

[50] G. Hofle, B. Kunzeb, C. Zorzin, H. Reichenbach, *Liebigs Ann. der Chemie* **1998**, *12*, 1883– 1904.

[51] B. Kunze, R. Jansen, G. Höfle, H. Reichenbach, *J. Antibiot. (Tokyo).* **1994**, *47*, 881–886.

[52] F. Sasse, H. Steinmetz, T. Schupp, F. Petersen, K. Memmert, H. Hofmann, C. Heuser, V. Brinkmann, P. von Matt, G. Höfle, et al., *J. Antibiot. (Tokyo).* **2002**, *55*, 543–551.

[53] R. Garcia, K. Gerth, M. Stadler, I. J. Dogma, R. Müller, *Mol. Phylogenet. Evol.* **2010**, *57*, 878–887.

[54] S. Donadio, M. M. Staver, J. J. McAlpine, S. Swanson, L. Katz, *Science (80-.).* **1991**, *252*, 675–679.

[55] S. Borchert, S. S. Patil, M. a Màrahiel, *FEMS Microbiol. Lett.* **1992**, *71*, 175–180.

[56] M. A. Marahiel, *Chem. Biol.* **1997**, *4*, 561–567.

[57] H. von Döhren, U. Keller, J. Vater, R. Zocher, *Chem. Rev.* **1997**, *97*, 2675–2706.

[58] D. E. Cane, C. T. Walsh, *Chem. Biol.* **1999**, *6*, 319–325.

[59] M. Hildebrand, L. E. Waggoner, H. Liu, S. Sudek, S. Allen, C. Anderson, D. H. Sherman, M. Haygood, *Chem. Biol.* **2004**, *11*, 1543–52.

[60] H. D. Mootz, R. Finking, M. a. Marahiel, *J. Biol. Chem.* **2001**, *276*, 37289–37298.

[61] F. Kopp, M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735–749.

[62] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651–2674.

[63] M. A. M. S A Sieber, *Chem. Rev.* **2005**, *105*, 715–738.

[64] D. Mootz, a Marahiel, *Chem. Biol.* **n.d.**

[65] T. Stachelhaus, A. Hüser, M. A. Marahiel, *Chem. Biol.* **1996**, *3*, 913–921.

[66] T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, *J. Biol. Chem.* **1998**, *273*, 22773–22781.

[67] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **2000**, *407*, 215–218.

[68] R. M. Kohli, J. W. Trauger, D. Schwarzer, M. a Marahiel, C. T. Walsh, *Biochemistry* **2001**, *40*, 7099–7108.

[69] D. Krug, R. Müller, *ChemBioChem* **2009**, *10*, 741–750.

[70] C. T. Walsh, R. V. O'Brien, C. Khosla, *Angew. Chemie - Int. Ed.* **2013**, *52*, 7098–7124.

[71] T. Stachelhaus, C. T. Walsh, *Biochemistry* **2000**, *39*, 5775–5787.

[72] C. T. Walsh, H. Chen, T. a Keating, B. K. Hubbard, *Curr. Opin. Chem. Biol.* **2001**, *5*, 525–534.

[73] U. Linne, M. a. Marahiel, *Methods Enzymol.* **2004**, *388*, 293–315.

[74] E. M. Nolan, C. T. Walsh, *ChemBioChem* **2009**, *10*, 34–53.

[75] K. Woithe, N. Geib, K. Zerbe, B. L. Dong, M. Heck, S. Fournier-Rousset, O. Meyer, F. Vitali, N. Matoba, K. Abou-Hadeed, et al., *J. Am. Chem. Soc.* **2007**, *129*, 6887–6895.

[76] H. C. Losey, J. Jiang, J. B. Biggins, M. Oberthür, X. Y. Ye, S. D. Dong, D. Kahne, J. S. Thorson, C. T. Walsh, *Chem. Biol.* **2002**, *9*, 1305–1314.

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

[78] B. J. Rawlings, *Nat. Prod. Rep.* **2001**, *18*, 231–281.

[79] F. E. Ruch, P. R. Vagelos, *J. Biol. Chem.* **1973**, *248*, 8086–8094.

[80] C. Khosla, Y. Tang, A. Y. Chen, N. a Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195–221.

[81] H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, *Mol. Biol. Evol.* **2005**, *22*, 2027–2039.

[82] Y. a Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher, M. G. Thomas, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 14349–14354.

[83] R. J. Heath, C. O. Rock, *Nat. Prod. Rep.* **2002**, *19*, 581–596.

[84] K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry* **1997**, *36*, 13849–13855.

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

[86] P. Caffrey, *ChemBioChem* **2003**, *4*, 654–657.

[87] J. Zheng, A. T. Keatinge-Clay, *Medchemcomm* **2013**, *4*, 34.

[88] A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941–53.

[89] D. H. Kwan, P. F. Leadlay, *ACS Chem. Biol.* **2010**, *5*, 829–838.

[90] C. Bisang, P. F. Long, J. Cortés, J. Westcott, J. Crosby, a L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton, P. F. Leadlay, *Nature* **1999**, *401*, 502–505.

[91] M. E. Horsman, T. P. a. Hari, C. N. Boddy, *Nat. Prod. Rep.* **2015**, *00*, 1–20.

[92] D. E. Cane, *J. Biol. Chem.* **2010**, *285*, 27517–27523.

[93] C. D. Campbell, J. C. Vederas, *Biopolymers* **2010**, *93*, 755–763.

[94] L. Shao, X.-D. Qu, X.-Y. Jia, Q.-F. Zhao, Z.-H. Tian, M. Wang, G.-L. Tang, W. Liu, *Biochem. Biophys. Res. Commun.* **2006**, *345*, 133–139.

[95] Q. Zhang, B. Pang, W. Ding, W. Liu, *ACS Catal.* **2013**, *3*, 1439–1447.

[96] R. J. Cox, *Org. Biomol. Chem.* **2007**, *5*, 2010–2026.

[97] J. Schümann, C. Hertweck, *J. Biotechnol.* **2006**, *124*, 690–703.

[98] Y.-Q. Cheng, G.-L. Tang, B. Shen, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3149–54.

[99] R. a Butcher, F. C. Schroeder, M. a Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1506–1509.

[100] J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, *Angew. Chemie - Int. Ed.* **2007**, *46*, 8195–8197.

[101] J. Piel, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14002–7.

[102] G.-L. Tang, Y.-Q. Cheng, B. Shen, *Chem. Biol.* **2004**, *11*, 33–45.

[103] E. M. Musiol, T. Weber, *Medchemcomm* **2012**, *3*, 871.

[104] K. Jensen, H. Niederkrüger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Pöplau, C. Kohlhaas, C. a Townsend, et al., *Chem. Biol.* **2012**, *19*, 329–39.

[105] E. M. Musiol, A. Greule, T. Härtner, A. Kulik, W. Wohlleben, T. Weber, *Chembiochem* **2013**, *14*, 1343–52.

[106] J. Piel, D. Hui, N. Fusetani, S. Matsunaga, *Environ. Microbiol.* **2004**, *6*, 921–927.

[107] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–33.

[108] M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel, N. J. Oldham, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 1143–7.

[109] C. T. Calderone, *Nat. Prod. Rep.* **2008**, *25*, 845–53.

[110] R. Gurney, C. M. Thomas, *Appl. Microbiol. Biotechnol.* **2011**, *90*, 11–21.

[111] L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis, D. H. Sherman, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 2795–8.

[112] A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson, C. M. Thomas, *J. Biol. Chem.* **2005**, *280*, 6399–6408.

[113] H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen, B. Shen, *J. Am. Chem. Soc.* **2008**, *130*, 6336–7.

[114] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996–1047.

[115] A. Sandmann, J. Dickschat, H. Jenke-Kodama, B. Kunze, E. Dittmann, R. Müller, *Angew. Chemie - Int. Ed.* **2007**, *46*, 2712–2716.

[116] A. O. Brachmann, S. a. Joyce, H. Jenke-Kodama, G. Schwär, D. J. Clarke, H. B. Bode, *ChemBioChem* **2007**, *8*, 1721–1728.

[117] R. H. Blum, S. K. Carter, *Ann. Intern. Med.* **1974**, *80*, 249–259.

[118] I. Chopra, M. Roberts, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 232–260.

[119] B. Shen, in *Biosynth. SE - 1* (Eds.: F. Leeper, J. Vederas), Springer Berlin Heidelberg, **2000**, pp. 1–51.

[120] K. K. Burson, C. Khosla, *Tetrahedron* **2000**, *56*, 9401–9408.

[121] B. S. Moore, C. Hertweck, *Nat. Prod. Rep.* **2002**, *19*, 70–99.

[122] B. S. Moore, J. N. Hopke, *Chembiochem* **2001**, *2*, 35–38.

[123] D. Yu, F. Xu, J. Zeng, J. Zhan, *IUBMB Life* **2012**, *64*, 285–295.

[124] M. B. Austin, J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.

[125] J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416.

[126] S. S. Chirala, S. J. Wakil, *Lipids* **2004**, *39*, 1045–1053.

[127] J. L. Smith, D. H. Sherman, *Science* **2008**, *321*, 1304–1305.

[128] S. Lorente-Cebrián, A. G. V Costa, S. Navas-Carretero, M. Zabala, J. A. Martínez, M. J. Moreno-Aliaga, *J. Physiol. Biochem.* **2013**, *69*, 633–651.

[129] G. Lenihan-Geels, K. S. Bishop, L. R. Ferguson, *Nutrients* **2013**, *5*, 1301–1315.

[130] D. Warude, K. Joshi, A. Harsulkar, *Crit. Rev. Biotechnol.* **2006**, *26*, 83–93.

[131] J. G. Metz, P. Roessler, D. Facciotti, C. Levering, F. Dittrich, M. Lassner, R. Valentine, K. Lardizabal, F. Domergue, a Yamada, et al., *Science* **2001**, *293*, 290–293.

[132] J. a. Napier, *Trends Plant Sci.* **2002**, *7*, 51–54.

[133] J. G. Wallis, J. L. Watts, J. Browse, *Trends Biochem. Sci.* **2002**, *27*, 467–473.

[134] K. Gemperlein, S. Rachid, R. O. Garcia, S. C. Wenzel, R. Müller, *Chem. Sci.* **2014**, *5*, 1733.

[135] J. G. Metz, J. Kuner, B. Rosenzweig, J. C. Lippmeier, P. Roessler, R. Zirkle, *Plant Physiol. Biochem.* **2009**, *47*, 472–478.

[136] H. Jiang, S. R. Rajski, B. Shen, *Tandem Acyl Carrier Protein Domains in Polyunsaturated Fatty Acid Synthases.*, Elsevier Inc., **2009**.

[137] R. Müller, *Chem. Biol.* **2004**, *11*, 4–6.

[138] B. Julien, R. Fehd, *Appl. Environ. Microbiol.* **2003**, *69*, 6299–6301.

[139] S. Jaoua, S. Neff, T. Schupp, *Plasmid* **1992**, *28*, 157–165.

[140] J. Muyrers, *Nucleic Acids Res.* **1999**, *27*, 1555–1557.

[141] S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch, D. V Santi, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15573–15578.

[142] H. G. Menzella, S. J. Reisinger, M. Welch, J. T. Kealey, J. Kennedy, R. Reid, C. Q. Tran, D. V Santi, *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 22–28.

[143] S. C. Mutka, J. R. Carney, Y. Liu, J. Kennedy, *Biochemistry* **2006**, *45*, 1321–1330.

[144] C. Oßwald, G. Zipf, G. Schmidt, J. Maier, H. S. Bernauer, R. Müller, S. C. Wenzel, *ACS Synth. Biol.* **2014**, *3*, 759–772.

[145] K. Gerth, H. Steinmetz, G. Höfle, R. Jansen, *Angew. Chemie Int. Ed.* **2008**, *47*, 600–602.

[146] K. H. Jang, S.-J. Nam, J. B. Locke, C. a. Kauffman, D. S. Beatty, L. a. Paul, W. Fenical, *Angew. Chemie* **2013**, *125*, 7976–7978.

[147] M. E. Hensler, K. H. Jang, W. Thienphrapa, L. Vuong, D. N. Tran, E. Soubih, L. Lin, N. M. Haste, M. L. Cunningham, B. P. Kwan, et al., *J. Antibiot. (Tokyo).* **2014**, *67*, 549–53.

[148] J. Held, T. Gebru, M. Kalesse, R. Jansen, K. Gerth, R. Müller, B. Mordmüller, *Antimicrob. Agents Chemother.* **2014**, *58*, 6378–84.

[149] M. Herrmann, B. Böhlendorf, H. Reichenbach, *Angew. Chemie* **1998**, *110*, 1313–1315.

Chapter 2

Two of a kind
The biosynthetic pathways of chlorotonil and anthracimycin

Previously published in:
Jungmann K.[1], Jansen R.[2], Gerth K.[2], Huch V.[3], Krug D.[1], Fenical W.[4], Müller R.[1]*
ACS Chemical Biology, Article ASAP; Pusblished online 8. September 2015
DOI: 10.1021/acschembio.5b00523

Affiliations
[1] Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Pharmaceutical Biotechnology, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany
[2] Helmholtz-Zentrum für Infektionsforschung GmbH, Inhoffenstraße 7, 38124 Braunschweig
[3] Universität des Saarlandes, Anorganische Chemie, Campus, Gebäude B 2.2, 66123 Saarbrücken
[4] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, United States

Contributions and Acknowledgements

A. Author's efforts

The author designed and performed most of the experiments described and evaluated and interpreted the resulting data. Cultivation of *Sorangium cellulosum* So ce1525 and *Streptomyces* CNH365 as well as analyses of extracts prepared from these strains were performed by the author. The author constructed a cosmid library of So ce1525 and identified, annotated and fully characterized the chlorotonil biosynthetic gene cluster by analyzing the respective sequence data from cosmids. Identification, annotation and *in silico* characterization of the anthracimycin biosynthetic gene cluster was carried out by the author as well. The author conducted all described feeding experiments and evaluated the respective HPLC-MS data. Furthermore, the author constructed the mutants described in the manuscript. The author contributed to conceiving and writing of the manuscript.

B. Contributions by others

Rolf Jansen and Klaus Gerth performed the isolation and structure elucidation of Chlorotonil A and its natural derivatives. Volker Huch did the X-Ray analysis and established a crystal structure for Chlorotonil B. William Fenical provided the anthracimycin producer strain Streptomyces CNH 365 and information about the cultivation.

The project was supervised by Daniel Krug and Rolf Müller, who also contributed to conceiving and writing of the manuscript.

In addition, the authors of the manuscript would like to thank Wolfgang Kessler and Steffen Bernecker and their team for large-scale fermentation, Kerstin Schober for skillful assistance in the isolation of the structural variants of chlorotonil A and Christel Kakoschke for NMR measurements. We are grateful to Michael Hoffmann and Thomas Hoffmann for HPLC-MS measurements and Viktoria Schmitt and Jennifer Herrmann for bioactivity assays.

2 Two of a kind - The biosynthetic pathways of chlorotonil and anthracimycin

2.1 Abstract

Chlorotonil A is a novel polyketide isolated from the myxobacterium *Sorangium cellulosum* So ce1525 that features a unique gem-dichloro-1,3-dione moiety. It exhibits potent bioactivity, most notably against the problematic malaria pathogen *Plasmodium falciparum* in the nanomolar range. In addition, strong antibacterial and moderate anti-fungal activity were determined. The outstanding biological activity of chlorotonil A as well as its unusual chemical structure triggered our interest in elucidating its biosynthesis, a prerequisite for alteration of the scaffold by synthetic biology approaches. This endeavor was facilitated by a recent report describing the strikingly similar structure of anthracimycin from a marine streptomycete, a compound of considerable interest due to its potent antibacterial activity. In this study we report the identification and characterization of the chlorotonil A biosynthetic gene cluster from So ce1525 and compare it with that for anthracimycin biosynthesis. Access to both gene clusters allowed us to highlight commonalities between the two pathways, and revealed striking differences, some of which can plausibly explain the structural differences observed between these intriguing natural products.

2.2 Introduction

In light of the alarming resistance rates for clinical antibiotics, and the concomitant loss of formerly potent therapy options, there is an increasing demand to identify new anti-infective drugs. Natural products such as secondary metabolites isolated from bacteria have proven to be a fruitful source of new lead structures for drug discovery. Evolutionary preselected for potent biological activity, natural products are considered as “privileged structures” in the fight against resistant pathogens as they provide an increased chance to hit previously unknown targets[1]. Over the last two decades, myxobacteria have been established as a rich source of novel bioactive compounds, suggesting great potential for the discovery of new drug candidates with unusual modes of action.[2] In particular, strains of the genus *Sorangium* have been frequently reported as promising producers of novel secondary metabolites with potent biological activity.[3] The

remarkably large genome size of members of *Sorangium* species is thought to reflect their capacity for secondary metabolite biosynthesis.[4] A broadscreening program for bioactive secondary metabolites from *Sorangium* strains led to the isolation of chlorotonil A from *Sorangium cellulosum* So ce1525 in 2008.[5] Structure elucidation identified the highly lipophilic macrolide chlorotonil A which features a unique gem-dichloro-1,3-dione structural moiety. This structural feature is novel amongst natural polyketides and fuelled our interest in elucidating the chlorotonil biosynthetic pathway. Strong antibacterial activity as well as the high potency against the malaria pathogen *Plasmodium falciparum* shown in mouse models fortified our interest, as did the report of the strikingly similar marine metabolite anthracimycin in 2013.[6] Anthracimycin shows good antibacterial activity which could be enhanced via chemical dichlorination of the scaffold, resulting in a derivative structurally mimicking chlorotonil A.

Polyketide metabolites such as chlorotonil A and anthracimycin are produced by modular polyketide synthases (PKS) which catalyze the step-wise connection of simple precursor molecules followed by distinct enzymatic tailoring reactions to form complex natural products. Each extension cycle is catalyzed by a set of enzymatic domains which are grouped into modules to form a biosynthetic assembly line. In the last module the product is most commonly released from the assembly line as either a linear or circular product. Modification of the polyketide core structure by free standing enzyme functionalities can occur during or after assembly on the PKS. A special subtype of polyketide biosynthesis gene clusters, namely the *trans-AT* type clusters have been discovered and investigated in the last couple of years.[7] In these PKSs, the individual modules lack integral acyltransferase (AT) domains for selection of the carboxylic acid substrate required for chain elongation; instead, a discrete AT encoding gene encoded in close proximity to the PKS genes, the product of which is responsible for iterative extender unit selection and transfer can be found. To date, only a small number of *trans*-AT PKS systems have been explored in detail and their mechanism of product assembly is at current not fully understood.

Understanding and elucidating the biosynthesis of chlorotonil A and anthracimycin sets the stage for the bioengineering of novel derivatives using synthetic biology approaches. Here we report the identification of the chlorotonil A biosynthetic pathway in *Sorangium cellulosum* So ce1525 and provide the first insights into its biosynthesis. Gene inactivation experiments and feeding studies were performed to corroborate our findings. A detailed comparison with the biosynthetic pathway of anthracimycin found in

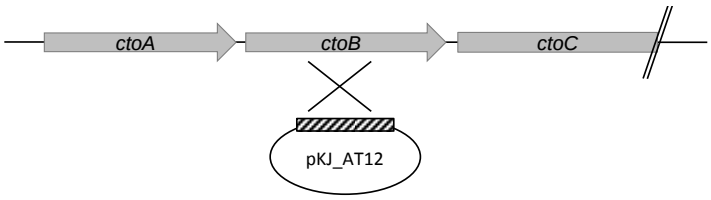
Streptomyces CNH365 sheds light on production of these intriguingly similar compounds by different hosts from different habitats.

2.3 Results and discussion

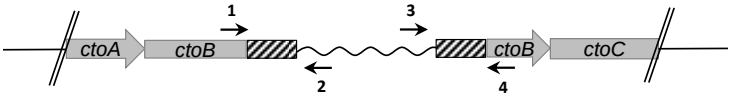
2.3.1 Identification of the chlorotonil biosynthetic gene cluster and new chlorotonil congeners

We sought to identify the gene cluster responsible for chlorotonil biosynthesis in the genome of *Sorangium cellulosum* So ce1525 and initially conducted 454 pyrosequencing for this purpose. The resulting 138 scaffolds were screened for putative PKS pathways matching the retro-biosynthetic proposal for chlorotonil synthesis. As the dichloro-pattern in chlorotonil A suggests involvement of a halogenating enzyme, all scaffolds were additionally searched for genes putatively encoding a halogenase. Using this approach, a concordant gene cluster harboring a putative halogenase gene, as well as a tandem-AT domain encoding gene were found co-localized on one scaffold. In order to probe the involvement of this locus in chlorotonil biosynthesis through gene inactivation, multiple conditions for directed mutagenesis of So ce 1525 were evaluated. After establishing a suitable method for conjugational DNA transfer, targeted disruption of the putative tandem AT domain encoding gene was performed by insertion of a hygromycin cassette using single-crossover homologous recombination (Fig. 1). The resulting So ce1525 mutant phenotype showed no production of chlorotonil A and thus confirmed the identification of the chlorotonil biosynthesis gene cluster. However, insufficient sequence quality precluded in-depth *in silico* analysis of the genomic locus. Gap-closing was therefore performed using PCR and by additional sequencing from a cosmid library (Fig. S10). Moreover, comparative LC-MS analysis of extracts from wildtype and mutant cultures highlighted several signals showing *m/z* values and isotope patterns suggestive of chlorotonil congeners not previously reported (Table. 1, suppl. inf.). Isolation and structural elucidation was pursued for three chlorotonil candidates which became accessible from a large-scale fermentation of the wildtype strain, revealing the structures of new chlorotonils B, C and C2 (suppl. inf.).

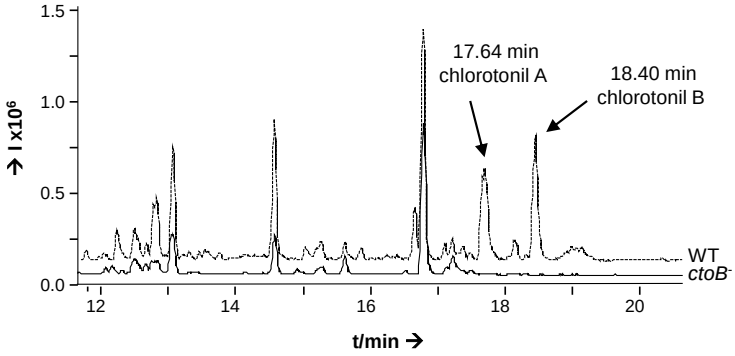
A.



B.



C.



D.

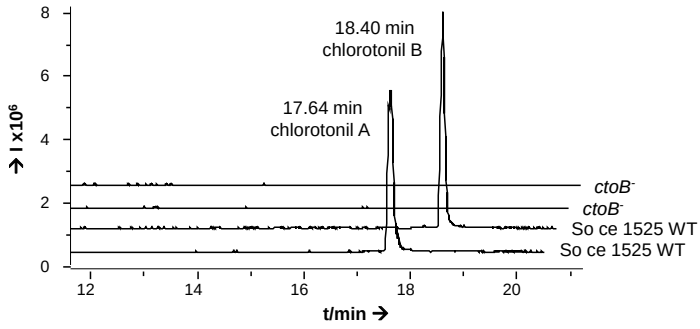


Figure 1: Strategy for the targeted inactivation of the tandem AT domain encoding gene *ctoB* in the chlorotonil A biosynthetic gene cluster A. Schematic overview on the gene organization at the 5' end of the chlorotonil biosynthetic gene cluster showing the target gene *ctoB* and the construct pKJ_AT12 designed for its inactivation via single crossover insertion. B. Genetic organization of the mutant strain So ce1525 *CtoAT::pKJ_AT12* containing the integrated pKJ_AT12. For verification of the genotype, PCR analysis was used.(Fig. S1) The primers used are indicated by the arrows (1: AT_KO_ver_fwd, 2: pSUP_Hyg_rev; 3: pSUPHyg_fwd, 4: AT_KO_ver_rev). C. Analysis of chlorotonil production of So ce1525 wildtype in comparison to the tandem AT knockout mutant by HPLC. Base peak chromatograms (200 – 500 Da) of the wildtype strain So ce 1525 showing production of chlorotonil A and B (upper dashed line) and the mutant strain lacking production of chlorotonils. D. Extracted ion chromatograms of the chlorotonil A target mass $[M+H]^+$ = 479.175 and the chlorotonil B target mass $[M+H]^+$ = 445.214 are shown for the wildtype (lower lines) and the mutant strain.

Besides its outstanding antimalarial activity,[8] chlorotonil A also exhibits remarkable anti-bacterial activity. Since solubility of the compound in polar solvents is very poor, a formulation in THF and cremophor was developed and used for antimicrobial *in vitro* assays (minimum inhibitory concentration, MIC). Overall, the compound strongly inhibited Gram-positive indicator strains such as *Bacillus subtilis* and *Staphylococcus aureus* in the low ng/ml range and displayed moderate antifungal activity while cytotoxicity against established cell lines was found to be relatively low. (IC_{50} for CHO-K1 and HCT-116 cells > 5 µM). It is notable that the MIC values of the dichlorinated chlorotonil A are significantly higher than those determined for the mono- and non-chlorinated congeners. This indicates that the chlorine atoms are crucial for the bioactivity, which is also supported by the results from anti-malaria testing.[8] Other than for anthracimycin, no activity against Gram-negative bacteria could be found.

Table 1: Bioactivity testing of chlorotonil A and its congeners

Table1. Bioactivity testing of chlorotonil A and its congeners

	Chlorotonil A **1**		Chlorotonil B **2**		Chlorotonil C **3**		Chlorotonil C2 **4**	
MIC	**µg/ml**	**µM**	**µg/ml**	**µM**	**µg/ml**	**µM**	**µg/ml**	**µM**
M. luteus DSM-1790	0.0125	0.025	> 3.2	>6.4	> 3.2	>6.4	1.6	3.8
B. subtilis DSM-10	≤ 0.003	≤0.006	> 3.2	>6.4	> 3.2	>6.4	1.6	3.8
S. aureus Newman	0.006	0.012	> 3.2	>6.4	3.2	>6.4	1.6	3.8
E. faecalis ATCC-29212	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4
S. pneumoniae DSM-20566	0.0125	0.025	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4
E. coli (TolC-deficient)	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4
E. coli DSM-1116	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4	> 3.2	>6.4

2.3.2 In silico analysis of the cto biosynthetic gene cluster and a model for chlorotonil biosynthesis

The chlorotonil biosynthetic gene cluster as identified by sequencing and confirmed through gene inactivation spans 80.7 kbp and contains 21 genes, amongst which three large genes *(ctoC, ctoD, ctoE*, table 2*)* encode a ten-modular PKS. As none of the PKS

modules contains an integral AT domain the megaenzyme complex is assigned to the group of *trans*-AT PKSs. In *trans*-AT pathways, the AT functionality is absent from individual modules and is replaced by a tandem-AT domain, encoded by *ctoB* in the case of chlorotonil biosynthesis. Alignment of the two AT domains with other known AT domains from *trans*-AT PKS gene clusters reveals that the AT2 domain falls into the malonyltransferase clade and is consequently responsible for extender unit selection (Fig. S2). This finding is underpinned by the presence of the common GHSxG motif in AT2, which is altered to GASxG in AT1. The AT1 domain groups with domains of the "acyl hydrolase" clade; AT domains belonging to this group exhibit a proof reading functionality by removing stalled acyl units from blocked carrier proteins.[9,10] Thus, the di-domain enzyme CtoB is most likely responsible for loading of the ACP domains with malonyl-CoA extender units and additionally exhibits a proof-reading functionality responsible for off-loading of deadlocked intermediates. As inferred from the chlorotonil structure, acetyl-CoA is serves as a starter unit for the first module. Fused to the C-terminus of the tandem AT domain, an additional ER domain of the 2-nitropropane dioxygenase family can be found. A similar domain arrangement has also already been described for the mupirocin, pederin, chivosazol and kirromycin pathway.[11–14] For bacillaene it could be shown that the ER domain fused to the *trans*-AT domain PksE was able to reduce double bonds in polyketide precursors independent from the action of the AT domain and thus acts as an additional *trans*-ER.[15] In the case of chlorotonil biosynthesis, no such functionality is missing on the assembly line hence the role of the *trans*-ER domain cannot be conclusively defined.

The presence of ten active KS domains in the chlorotonil PKS, all exhibiting the highly conserved HGTGT motif, as well as the DxxCSSLx motif with slight variations (Fig. S3) is in accordance with the retrobiosynthetic analysis that proposes the incorporation of ten acetate-derived extender units.[16] Substrate specificity of all ten KS domains located on *ctoCDE* was predicted via multiple alignment according to Piel et al.,[17] whereupon the prognosis largely matches the biosynthesis model for chlorotonil. (Fig. S16). In the course of chlorotonil biosynthesis the extender units are evidently processed by a reductive loop in eight out of ten cases which leads to either α,β-olefinic or alkylic intermediates. Analysis of the KR domains present in CtoC,D,E revealed an LxD motif in all eight KR domains (KR in modules 2, 3, 4, 6, 7, 8), hence all of them belong to the class of B-type ketoreductases leading to *S*-configured alcohols.[18,19] For KR1 and KR5 no

such LxD motif could be detected which assigns them to the group of A-type ketoreductases forming the respective *R*-configured alcohols.

Furthermore, eight modules of the chlorotonil PKS include DH domains, two of which are found on so-called split modules where the respective module stretches across two genes. Seven out of eight DH domains show the typical motifs needed for dehydratase activity, including highly conserved histidine and aspartate residues (Fig. S3). [20] The deviating dehydratase domain in module one is a truncated variant of a DH domain lacking the essential motifs and conserved residues and is proposed to be inactive, in agreement with the chlorotonil structure. The only ER domain found in module six shows the conserved tyrosine residue correlated with 2S configuration in the polyketide whilst only slight deviations in the LxHxxxGGVGxxAxxxA NADPH binding site consensus motif can be observed (Fig. S5); nevertheless its activity seems to be maintained during biosynthesis, as concluded from the reduction of the double bond introduced by module six. Intriguingly, this ER domain occupies an unusual position in relation to the surrounding domains: contrary to the canonical placement of ER domains in the reductive loop, this domain lies downstream of the ACP domain in module six. Compared to *cis*-AT PKS systems this positioning seems rather unusual, but for *trans*-AT systems this domain arrangement has already been described e.g. in the rhizopodin biosynthetic gene cluster.[21] Whether this positioning influences domain activity or makes the ER available for processing reactions outside its ancestral module remains unclear. Our retro-biosynthetic analysis suggests that two double bond shifts are to be carried out in modules three and seven, to yield the respective α-methylated, β,γ-olefinic intermediate. Deviant from known mechanisms and domain architectures of the earlier described shift modules catalyzing this reaction,as for example known from rhizoxin biosynthesis, no such enzymatic activity can be found in the two modules.[22] It is thus very likely that the double bond shift is executed in the respective elongating module as previously described for bacillaene biosynthesis.[23] Split modules in which one module is partitioned between two proteins are commonly observed in *trans*-AT PKS systems.[7] Two split modules (module 4, module 8) are found in the chlorotonil PKS; this structural peculiarity does not seem to influence the construction of the polyketide chain as both modules yield the respective acrylyl intermediates corresponding to the degree of β-processing expected from the allocated KR and DH domains and stipulated by retro-biosynthetic analysis.

The domain architecture of the chlorotonil PKS diverges notably from the colinearity rule with three apparently superfluous ACP domains. Additional ACP domains have been

proposed to play a role in β-branching events where they are described as a breakpoint for β-branch incorporation, facilitating this modification via retardation of biosynthetic flux through the pathway.[24] Furthermore, it has been speculated that these additional domains could enhance the yield in complex enzymatic reactions as described for curacin biosynthesis.[25] It is tempting to interpret the duplication of ACP domains in module three and six accordingly, as the two modules harbor MT domains and are responsible for incorporation of a α-methyl branch. However, in contrast to all other ten ACPs, the three duplicate ACP domains do not possess the conserved serine in the GIDSxL motif where the phophopantetheinyl arm is attached.[26] Thus, they appear to be non-functional and may represent an inactive flanking subdomain of the upstream KS as previously reported by Keatinge-Clay et. al..[27]

2.3.3 Maturation of the PKS derived chlorotonil core

The α-methyl branches at C8, C10 and C16 in chlorotonil A are plausibly incorporated by the integral methyltransferase domains in modules three, six and seven. Methylation at C2 is proposed as a post-PKS tailoring reaction catalyzed by the free-standing S-adenosylmethionine (SAM)-dependent methyltransferase CtoF. The SAM-origin of all methyl groups was proven by a ^{13}C-methyl-methionine feeding study (Fig. S9). An alignment of all four MTs revealed that the three internal MT domains of modules three, six and seven are highly similar to each other, and distinct from the free-standing MT domain encoded by *ctoF* (Fig. S6-S7.1). We reason that the three integral MT domains are responsible for α-methylation during polyketide chain formation at the appropriate positions, while the external methyltransferase CtoF acts upon the circular pre-mature polyketide (**1** in Fig. 2). This assumption is supported by the finding of a mass of $[M+H]^+$ = 397.23 m/z representing a chlorotonil precursor molecule which is not halogenated at C4 and not methylated at C2, during LC-MS analysis of So ce1525 extracts.(Fig. S13)

A striking feature of chlorotonil is its decalin ring partial structure which likely results from an intramolecular Diels-Alder-like [4+2] cycloaddition reaction. We propose that this reaction takes place during polyketide assembly on the PKS enzyme complex. Whether such [4+2]-cycloadditions can occur spontaneously or involve enzymatic catalysis by so called “Diels-Alderases” remains controversial. For lovastatin biosynthesis in *Aspergillus terreus*, a free standing enoylreductase LovC is thought to assist the ring formation as a partner protein of the lovastatin nonaketide synthase LovB,[28] whereas the methyltransferase homologue SpinF from *Saccharopolyspora spinosa* catalyzes an

intramolecular Diels-Alder reaction leading to a spinosyn precursor molecule.[29] In addition, two distinct enzymes, namely the macrophomate synthase MPS (from *Macrophoma commelinae*),[30] and the solanopyrone synthase Sol5 (from *Alternaria solani*),[31] have been recombinantly produced and associated with Diels-Alder-like biotransformations. Very recently, new enzymes involved in the macrocyclization of spirotetronate containing polyketides have been shown to catalyze [4+2] cycloadditions: in the case of versipelostatin VstJ was found to catalyze ring formation between a highly flexible acyclic diene substructure and an exocyclic olefin.[32] The distance and spatial arrangement between those two moieties seems to be the critical parameter, which necessitates enzymatic catalysis to accomplish ring formation. In-depth *in silico* analysis of the chlorotonil biosynthetic gene cluster and the genomic regions up- and downstream of this locus did not reveal any genes exhibiting significant homology to the afore mentioned enzymes. We thus conclude that the intramolecular Diels-Alder reaction, leading to formation of the decalin moiety occurs spontaneously without dedicated enzymatic catalysis whilst the polyketide chain is still tethered to the assembly line in module eight. In light of the heterogeneity of putative diels-alderase enzymes and their substrates reported to date, it seems likely that there is no generalized enzyme class to promote [4+2] cycloadditions. Especially in the case of versipelostatin it seems that the enzyme identified is highly specific for the versipelostatin precursor. As mentioned by the authors it is very likely that VstJ orientates its substrate into an advantageous conformation for ring formation. In the case of chlorotonil we hypothesize that the diene and the dienophile moiety could arrange in close proximity to each other while the polyketide intermediate is tethered to the ACP domain (Fig. 2, Fig. S11); thus additional enzymatic catalysis of the Diels-Alder reaction may not be necessary.

After ten rounds of chain extension, the polyketide chain is released from the assembly line and cyclized by the TE domain of module ten. This pre-mature macrolactone (**5** in Fig. 2) then undergoes two tailoring steps to form chlorotonil A, and we propose the following timing for these post-PKS modifications: first **5** is halogenated at the C3 position by the flavin dependent halogenase CtoA; second, the methylation at C2 occurs catalyzed by the free-standing methyltransferase CtoF to yield chlorotonil A. Evidence for this order of modification reactions comes from the structures of natural chlorotonil congeners as well as from feeding experiments and is discussed below. The highly abundant chlorotonil B, and less abundant chlorotonil C, differ from chlorotonil A by the lack of a second chlorine atom at C4. Chlorotonil C and C2 also exhibit carbon skeleton

variations: the C10 methyl group and C9-C10 double bond are missing. For the biosynthesis of these derivatives two possible scenarios are conceivable. First, the missing methyl group at C10, which results from inefficient methylation functionality in module seven, could transform the resulting polyketide into a substrate for an external hydrogenase that reduces the C9-C10 double bond. Secondly, we cannot exclude that the oddly positioned ER domain between modules six and seven accessorily acts upon the polyketide chain tethered in module seven when the α-methyl group is absent.

Table 1: Proposed function of the genes involved in chlorotonil biosynthesis and the surrounding open reading frames

Gene/protein	length (bp/aa)	proposed function
ctoA/CtoA	1875/625	FAD-dependent halogenase
ctoB/CtoB	3462/1154	AT_1 (61-993), AT_2 (1063-1884), ER (1920-3156)
ctoC/CtoC	18816/6272	KS_1 (19-1294), DH_1* (2650-3007), KR_1 (3580-4135), ACP_1 (4405-4609), KS_2 (4774-6082), DH_2 (6655-7483), KR_2 (8320-8875), ACP_2 (9211-9397), KS_3 (9625-10930), ACP_{3-1} (11142-11325), DH_3 (11838-12663), KR_3 (13383-13920), MT_3 (14622-15282), ACP_{3-2} (15504-15711), KS_4 (16018-17281), DH_4 17893-18663)
ctoD/CtoD	22806/7602	KR_4 (862-1414), ACP_4 (1699-1900), KS_5 (2158-3457), DH_5 (4015-4839), KR_5 (5584-6139), ACP_5 (6397-6595), KS_6 (6766-8068), ACP_{6-1} (8242-8446), DH_6 (8939-9762), KR_6 (10525-11065), MT_6 (11755-12343), ACP_{6-2} (12580-12790), ER_6 (13096-14023), KS_7 (14104-15379), DH_7 (15928-16752), KR_7 (17428-18022), MT_7 (18721-19384), ACP_7 (19516-19723), KS_8 (19987-21256), DH_8 (21829-22653)
ctoE/CtoEC	7773/2591	KR_8 (844-1393), ACP_8 (1639-1837), KS_9 (2008-3304), ACP_{9-1} (3496-3700), ACP_{9-2} (4324-4531), KS_{10} (4714-6019), ACP_{10} (6637-6838), TE (6943-7723)

		Proposed function of homologous protein	Source of the homologous protein	Identity/ similarity %	Accession number (GenBank)
Orf1	3669/1223	hypothetical protein	*Myxococcus stipitatus*	33/48	WP_015351249
Orf2	4170/1390	hypothetical protein	*Myxococcus stipitatus*	33/48	WP_015351249
CtoF	828/276	SAM-dep. methyltransferase	*Thermincola potens*	41/59	WP_013119426
Orf4	765/255	hypothetical protein	*Mariprofundus ferrooxydans*	48/70	WP_009850450
Orf5	1218/409	ABC transporter - permease	*Enhygromyxa salina*	26/44	KIG16386
Orf6	1233/411	ABC transporter - permease	*Enhygromyxa salina*	31/49	KIG16385
Orf7	696/232	ABC transporter – ATP binding protein	*Myxococcus xanthus*	49/61	WP_011552505
Orf8	1236/412	hypothetical protein	*Syntrophobacter fumaroxidans*	35/48	WP_011699354
Orf9	798/266	ATP binding protein	*Amycolatopsis nirgrescens*	60/74	WP_020673914
Orf10	702/234	putative ABC transporter	*Halococcus hamelinensis*	38/51	WP_007692158
Orf11	708/236	putative ABC transporter	*Methanocella arvoryzae*	28/46	WP_012036131
Orf12	591/197	esterase	*Sorangium cellulosum*	76/84	WP_012236733
Orf13	828/276	LuxR family transcriptional regulator	*Sorangium cellulosum*	83/87	WP_012236732
Orf14	657/219	hypothetical protein	*Chrondomyces apiculatus*	73/82	EYF05039
Orf15	1680/560	hypothetical protein	*Chrondomyces apiculatus*	74/82	EYF05038
Orf16	783/261	hypothetical protein	*Chrondomyces apiculatus*	67/74	EYF05037

1. Chlorotonil PKS

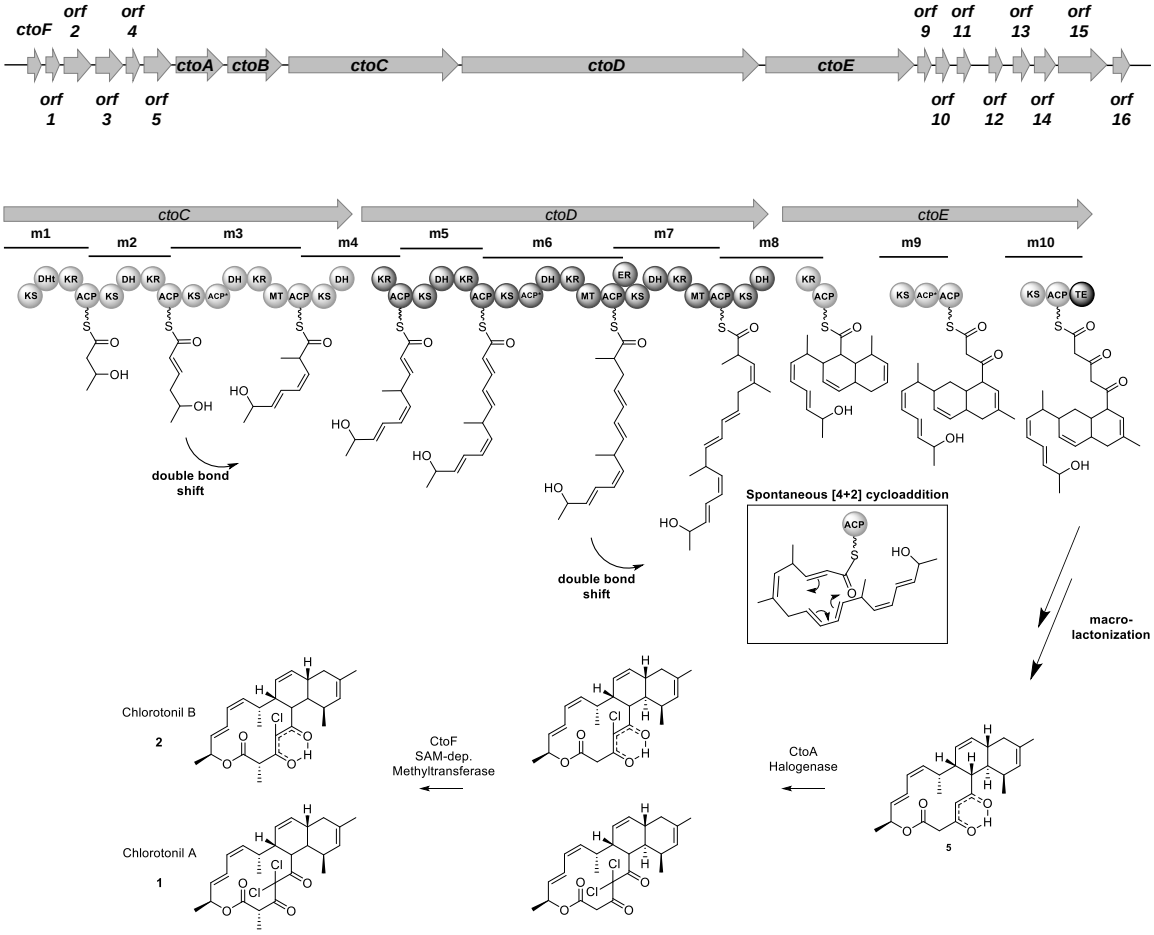


Figure 2: A model for chlorotonil biosynthesis in So ce 1525. The organization of the chlorotonil A biosynthetic gene cluster is schematically shown in A. The proposed biosynthetic pathway for chlorotonil A and B is shown in B. ACP, acyl carrier protein; DH, dehydratase, DHt, inactive truncated dehydratase; ER, enoyl reductase; KR keto reductase; KS, keto synthase; MT, methyl transferase; TE thioesterase.

2.3.4 Geminal di-chlorination of the chlorotonil scaffold by the halogenase CtoA

To the best of our knowledge, the unique gem-dichloro-1,3-dione substructure found in chlorotonil A, has not been reported for any other natural product. This feature appears critical to the bioactivity of chlorotonil as the anti-malarial activity was lost when the non-halogenated derivatives were examined.[8] Similarly, effectiveness of anthracimycin increased when the compound was chemically di-chlorinated.[6] To investigate the role of CtoA, a putative FAD-dependent halogenase, for chlorotonil halogenation we inactivated the *ctoA* gene via the established single crossover protocol: we expected that mutant strain would produce only non-halogenated chlorotonil derivatives. Disappointingly, the resulting mutant produced no chlorotonil or analogues indicating a polar effect on the

downstream genes. Despite numerous attempts a double-crossover excision of the halogenase gene did not succeed.

We then attempted reconstitution of the halogenation reaction *in vitro*, but efforts towards recombinant production of CtoA in *E. coli* were not successful. To corroborate our hypothesis for the timing of tailoring reactions during chlorotonil maturation we instead used the tandem-AT (*ctoB*) knockout mutant in a feeding study to probe halogenation of chlorotonils B (**2** in table 1) and C2 (**4** in table 1). In the *ctoB* mutant, translation of *ctoA* is not hampered by a polar effect as it resides upstream of the disruption (Fig. 2). After incubating the *ctoB* mutant with **2** and **4** for 10 days, extracts were analyzed for the presence of chlorinated compounds using LC-MS but neither extract showed the presence of chlorotonil A. However, further analysis of the MS data from feeding **4** identified two conspicuous peaks with chlorotonil-like retention times and MS patterns consistent with the presence of halogen atoms. The m/z ratios of these, 433.21 and 467.17 respectively, were 14 mass units higher than the anticipated $[M+H]^+$ ions for mono-and di-chlorination of chlorotonil C2, but is consistent with chlorination plus an additional methylation event (Fig. S12). MS fragmentation of the two compounds disclosed the methyl position to be C2 of chlorotonil C2 (data not shown). These results imply that cyclic chlorotonil precursor molecules are chlorinated prior to methylation at C2, and that precursor molecules that are already methylated at C2 are not accepted as substrates by CtoA. The existence of mono- and di-chlorinated C2-methyl derivatives hints that the methylation and the chlorination process are carried out almost in parallel. Further support for this model came from the analysis of wild-type So ce1525 extracts for the corresponding chlorotonil derivatives: in addition to chlorotonil A (479.17 m/z, $[M+H]^+$) and chlorotonil B (445.21 m/z, $[M+H]^+$), the equivalent non-methylated derivatives (465.15 m/z, $[M+H]^+$), (431.19 m/z, $[M+H]^+$) as well as the non-methylated and non-chlorinated precursor (397.23 m/z, $[M+H]^+$) were detected (**8**, **9**, **10** in Fig. S13).

To further exploit the functionality of CtoA a feeding experiment with the natural producer strain So ce1525 was carried out. The chlorine source in the media was depleted and replaced with equimolar amounts of the respective bromide salt, which was achieved by exchanging the media component calciumchloride for calciumbromide. As anticipated, substantial amounts of new brominated compounds were biosynthesized (**11**, **12** in Fig. S14). Consistent with a lack of chloride only trace amounts of chlorotonil A and B were

produced, while the non-methylated, non-halogenated precursor molecule **10** ($[M+H]^+$ = 397.23) accumulated. We thus conclude that CtoA is also capable of conducting bromination reactions.

Given the promising bioactivity seen for the dichloro-derivative of anthracimycin, a biological producer of this compound would be of great interest. We thus elaborated a strategy for integration of *ctoA* under regulation of an *erm*E promoter into the *attB* site of the natural producer *Streptomyces* sp. CNH365, despite legitimate doubts as to the ability of CtoA to transform anthracimycin given its methylation pattern. Despite difficulty in finding conditions for genetic modification of *Streptomyces* sp. CNH365, two mutants harboring the *ctoA* gene were eventually obtained. After verification of the genotype by PCR, the mutants were grown in liquid media and extracts were prepared as described by Jang et al.[33]. However, despite exhaustive LCMS analysis no evidence for the production of halogenated anthracimycin analogues could be found. Given our findings regarding its substrate specificity we conclude that anthracimycin, being methylated at C2, is not a substrate for CtoA.

2.3.5 Comparative analysis of the *atm* biosynthetic gene cluster and a model for anthracimycin biosynthesis

The structure of anthracimycin, a bioactive natural product isolated from a marine streptomycete,[33] closely resembles that of chlorotonil. We therefore set out to identify the anthracimycin biosynthetic gene cluster for comparison of the two pathways. Genome sequencing of the natural producer strain *Streptomyces* sp. CNH365 using the Illumina technique yielded 72 scaffolds which were subjected to automatic secondary metabolite pathway annotation using antiSMASH3.[34] In addition, the genome sequence was searched with BLAST using the chlorotonil PKS genes as the query sequence. Surprisingly, this search revealed the presence of two identical copies of a *trans*-AT PKS gene cluster in the genome of *Streptomyces* sp. CNH365 that were highly likely to encode the anthracimycin biosynthetic machinery (Fig. S15). This peculiarity is rarely seen and the circumstances of its origin are beyond the scope of this work. However, the presence of a duplicated biosynthetic gene cluster suggests that production of anthracimycin may provide a significant selection advantage. The genomic loci each span 70 kb, and consist of three large PKS encoding genes (*atcD, atcE, atcF*), a tandem AT gene with a C-terminal ER domain (*atcC*) and several genes up- and downstream the PKS core region encoding transport- or regulator functions (table 2). The core gene cluster could be

determined as a 53 kb region spanning the genes *atcA-atcI* by comparison with anthracimycin biosynthetic gene cluster of *Streptomyces* sp. T676 which is reported by Alt and Wilkinson contemporaneously with this article (this issue). At first glance the organization of this gene cluster is similar to the chlorotonil biosynthetic gene cluster (Fig. 3). The most obvious difference is the absence of genes encoding homologues for the halogenase CtoA and the free-standing methyltransferase CtoF which agrees well with the chemical structure of anthracimycin in comparison to chlorotonil. In addition to these major differences, the domain distribution amongst the *atcDEF* genes differs from the arrangement seen for *ctoCDE*; this is discussed in section 2.6.8.1 .

Table 2: Proposed function of the genes involved in anthracimycin biosynthesis and the surrounding open reading frames

Gene/protein	length (bp/aa)	proposed function			
atcC/AtcC	3222/1074	AT_1 (45-796), AT_2 (927-1767), ER (1944-2709)			
atcD/AtcD	3462/1154	KS_1 (54-1329), KR_1 (3625-4180), ACP_1 (4459-4660), KS_2 (4813-6124), DH_2 (6658-7482), KR_2 (8269-8824), ACP_2 (9148-9349), KS_3 (9532-10846), DH_3 (11581-12405), KR_3 (13153-13690), MT_3 (14344-14989), ACP_3 (15157-15349), ER_3 (15568-16492), KS_4 (16603-17875), DH_{4-1} (17974-18337), DH_{4-2} (18382-19206)			
atcE/AtcE	18816/6272	ACP_4 (147-355), KS_5 (493-1777), DH_5 (2317-3141), KR_5 (3928-4480), ACP_5 (4759-4960), KS_6 (5128-6412), ACP_{6-1} (6601-6793), ACP_{6-2} (7375-7585), KS_7 (7747-9022), DH_7 (9532-10356), KR_7 (10951-11488), MT_7 (12142-12796), ACP_7 (12883-13087)			
atcF/AtcF	22806/7602	KS_8 (145-1417), DH_8 (1981-2805), KR_8 (3487-4033), ACP_8 (4273-4480), KS_9 (4651-5944), ACP_9 (6892-7066), KS_{10} (7237-8542), ACP_{10} (9178-9364), TE (9508-10312, MT_{10} (10783-11422))			
		Proposed function of homologous protein	Source of the homologous protein	Identity/ similarity, %	Accession number (GenBank)
Orf1	837/279	regulator	*Streptomyces sp.* CNS654	82/88	WP_032767193
Orf2	1533/511	putative amino acid transporter	*Streptomyces ambofaciens*	82/87	CAI78131
Orf3	516/172	hypothetical protein	*Streptomyces violaceusniger*	70/74	WP_014060436
Orf4	702/234	transposase	*Streptomyces sp.*CNQ.431	63/69	WP_033952594
atcA/AtcA	1593/531	putative multidrug efflux protein	*Micromonospora lupini*	36/46	WP_007461885
atcB/AtcB	615/205	TetR family transcriptional regulator	*Streptomyces sp.* CNR698	54/68	WP_027732177
atcG/AtcG	852/284	ATP binding protein	*Methanocella arvoryzae*	55/73	WP_012036133
atcH/AtcH	750/250	ABC transporter permease	*Peanibacillus sp.* OSY-SE	29/44	WP_036705498
atcI/AtcI	708/236	hypothetical protein (partial)	*Streptomyces sp.* NRRL F-5065	80/90	WP_037857030
Orf10	309/103	PadR family transcriptional regulator	*Streptomyces pristinaespiralis*	77/78	WP_005322088
Orf11	1278/426	transporter	*Streptomyces albus*	79/78	WP_030624822

2. Anthracimycin PKS

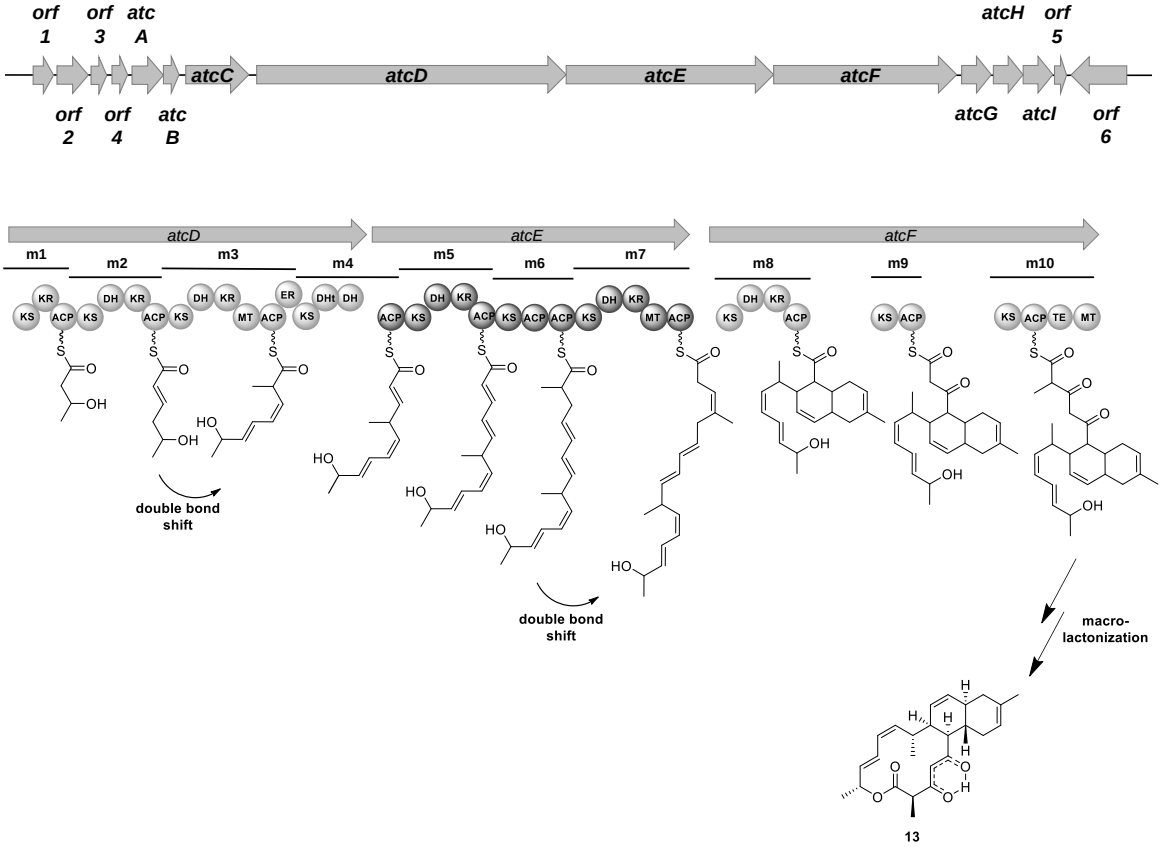


Figure 3: Model for anthracimycin biosynthesis in Streptomyces CNH365. A depicts the organization of the biosynthetic gene cluster. The proposed biosynthetic pathway for anthracimycin is shown in B. ACP, acyl carrier protein; DH, dehydratase, DHt, inactive truncated dehydratase; ER, enoyl reductase; KR keto reductase; KS, keto synthase; MT, methyl transferase; TE thioesterase.

The architectural irregularities of the anthracimycin PKS (versus that anticipated from a retro-biosynthetic analysis) raise questions about how it produces a chemical scaffold so similar to chlorotonil. While the chlorotonil PKS also shows unusual domain arrangements, these peculiarities are not dissimilar to those often seen for *trans*-AT PKS systems. In contrast, the anthracimycin biosynthetic machinery appears to lack several domains whilst others are placed distant from the anticipated point of action. We are nevertheless confident that the *atc* locus represents the anthracimycin biosynthesis gene cluster as careful inspection of the *Streptomyces* sp. CNH365 genome sequence did not afford any other candidate gene cluster encoding an alternative biosynthetic pathway for anthracimycin. A pathway for anthracimycin biosynthesis is proposed in 2.6.8.2.

2.4 Conclusion

In this study we identified the biosynthetic gene clusters for chlorotonil and anthracimycin, two strikingly similar natural product scaffolds produced by a myxobacterium and a marine streptomycete, respectively. Our results indicate that both compounds are formed by *trans*-AT PKS pathways, and in-depth analysis of the *cto* and *atc* gene clusters allowed us to substantiate the reasons for several distinct structural features of chlorotonil A. In particular, the gene set for tailoring reactions readily explains the observed halogenation and methylation pattern of chlorotonil A, and a combination of gene inactivation and feeding experiments enabled us to propose a model for the order and timing of these post-PKS reactions. Both chlorotonil and anthracimycin apparently undergo [4+2] cycloadditions to form their characteristic decalin moiety, and we hypothesize that ring closure takes place during biosynthesis without catalysis by a dedicated “Diels-Alderase”. This notion is supported by the genome-wide absence of plausible candidates exhibiting similarity to any previously reported genes encoding such functionality. The finding that for anthracimycin in comparison to chlorotonil all stereocenters are inverted is very intriguing and we hypothesize that the alternate steric arrangements induced by the C16-methyl group of the PKS intermediates acts to ‘pre-organize’ the PKS-bound intermediates in advance of the [4+2] cyclization reactions (see Fig S11.2). The hydroxyl group introduced during the first round of chain extension is also of opposite stereochemistry in each molecule, and this may also suggest a role in pre-organizing the PKS bound intermediate. It is also conceivable that the orientation of the ACP-bound intermediates is pre-defined by the protein itself.

Despite them being far less conceptual than *cis*-AT PKS, several rules for polyketide assembly by *trans*-AT PKS were deduced from phenomena repetitively observed for these systems. However, urgent questions emerging from our comparative gene cluster analysis presented in this study could not be answered by what is known from *trans*-AT PKS studied to date. This suggests that generally valid rules for polyketide biosynthesis by these tessellated biosynthetic machineries are hard to establish.

Considering the promising bioactivity of chlorotonil A, the production of less lipophilic derivatives based on a system for heterologous expression of the pathway is one of the major goals to achieve in further studies. Analysis of the chlorotonil biosynthetic gene cluster provides valuable insights into chlorotonil biosynthesis and the data gathered in this study is giving direction for the production of chlorotonil derivatives. Molecular

features required for bioactivity of the compound as well as structural prerequisites for the production of chlorotonil derivatives were identified. Consequently starting points for modification of the chlorotonil biosynthetic gene cluster towards production of new bioactive and less lipophilic derivatives could be concluded from this work.

2.5 Experimental section

2.5.1 Bacterial strains and culture conditions

Sorangium cellulosum So ce1525 was grown at 30°C and 180 rpm in liquid H media (soy flour 0.2 %, glucose 0.2 %, starch 0.8 %, yeast extract 0.2 %, CaCl2*2H2O 0.1 %, MgSO4*7H2O 0.1 %, HEPES 1.2 % pH7.4) supplemented with 8 mg/l Fe-EDTA after autoclaving and on solid H agar plates (H media with 1.5 % agar). In liquid culture, the strain was grown for 7 to 12 days. *Escherichia coli* strains were cultured in LB media (lysogeny broth; tryptone 1 %, yeast extract 0.5 %, sodium chloride 0.5 %) at 37 °C until the required cell density was reached. If necessary the appropriate antibiotics were added to the media for selection. *Streptomyces* CNH 365 was cultivated in liquid media as described previously[6]. For maintenance the strain was kept on SM agar plates (soy flour 2 %, mannitol 2 %) and incubated at 30 °C.

2.5.2 Preparation of genomic DNA and PCR

So ce1525 genomic DNA was prepared either by phenol/chloroform/isoamylacohol extraction[36] or with the help of the Gentra Puregene Genomic DNA PurificationKit (Qiagen) according to the manufacturer’s protocol. Plasmid DNAwas prepared by alkaline lysis [Ref. sambrook] or – if higher purity was required - using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). PCR reactions were carried out in an eppendorf mastercycler pro using the following program: 5 min initial denaturation (5 min at 98°C) followed by 30 cycles of denaturation (98°C, 20 sec), annealing (58-68°C, 30 sec), and elongation (0.5 – 1.5 min according to the expected amplicon size). Final extension (5 min, 72 °C) ended the program. PCR amplicons were analyzed by gel electrophoresis and isolated by using the peqGOLD Gel Extraction Kit (Peqlab). PCR fragments were cloned into pJET 1.2 blunt (Thermo scientific) and sequenced by using the pJET1.2For/pJET1.2Rev sequencing primers.

2.5.3 Construction of a So ce1525 cosmid library

Genomic DNA of So ce1525 was partially digested using *Sau*3A1 as described in the Supercos 1 manual (Agilent). A 2 µl fraction of each sample was analyzed by gel electrophoresis on an 0.35 % agarose gel. Samples exhibiting the correct size distribution were combined, extracted and dephosphorylated according to the Supercos1 manual. The supercos 1 vector was prepared as indicated in the respective manual. Complete digestion of the vector – indicated by the appearance aof a 1.1 kb and a 6.5 kb band - was checked by agarose gel electrophoresis. For ligation of the partially digested DNA with the Supercos1 backbone, DNA and Supercos1 were used in a 2.5 : 1 ratio. The ligation preparation was shortly mixed, spun down, and incubated over night at 4 °C. The ligation mix was afterwards precipitated and packaged according to the manufacturer's instructions using the Gigapack III Gold (Agilent) packaging mixture. The resulting packaging mixture was used to infect *E. coli* HS996 cell that were subsequently plated out on LB agar plates supplemented with 50 µg / ml ampicillin. Clones obtained after this selection were transferred into 384 well plates containing 2YT freezing media (tryptone 1.6 %, yeast extract 1.0 %, sodium chloride 0.5 %, $MgSO_4$ x $7H_2O$ 0.068 %, sodiumcitrate x 2 H_2O 0.4 %, $(NH_4)_2SO_4$ 0.81 %, Glycerol 39.6 %, K_2HPO_4 4.23 %, KH_2PO_4 1.62 %) and grown overnight at 30°C. High-density colony filters of the 1865 Cosmid- containing library were generated by spotting cell suspension from every well onto LB soaked nitrocellulose membrane filters that were placed on LB ampicillin agar plates and incubated overnight at 30°C. For spotting a Qpix robot (Genetix, New Milton,Hampshire, UK) was used. The clones gown on the filter membrane were lysed (lysis solution : 0.5 M NaOH, 1.5 M NaCl), cell debris was washed off and the membranes were dried. The dried membranes were then immersed in 2x SSC (NaCl 0.3 M, tri-Sodiumcitrat x 2 H2O 0.03 M) for 5 min, drained, then covered with proteinase K solution and wrapped in aluminum foil. After 2h of incubation at 37°C the membranes were carefully rinsed with water and air-dried on whatman paper. The DNA was thus fixated on the membranes which subsequently underwent a screening procedure.

2.5.4 Screening of the Sorangium cellulosum So ce1525 cosmid library for the chlorotonil biosynthetic gene cluster

The cosmid library was screened using three probes targeting the 5'-end (halo probe) the 3'-end (TE probe) and the core region of the gene cluster. The probes were amplified from genomic DNA of So ce1525. For the halo probe the primers halo_fwd (5'-CTCCTCCAGGTACCTCCGCATCTC -3') and halo_rev (5'-CCGAGGAGCTCATGCAGCAGAC-3') were used, for the core probe core_fwd (5'-GCAGGCGGTGTCCCTGCACGAC-3') and and core_rev (5'-CATCCCGCTCAACAACCGAGGCAG -3') were used and for the TE probe TE_fwd (5'- CTCATCATGAGTCCAGGTACGCTC -3') and TE_rev (5'-CGGCGGGAGATCTGCGCTGTCG -3') were used. Positive cosmids identified in this way were further analyzed by end-sequencing with the T3 and T7 primers. After this the three most promising candidates (cosmid B8, D20, J23) were completely sequenced with the Illumina method. The thus obtained sequencing data was assembled and resulted in a 80 kb genomic region covering the sought biosynthetic gene cluster as well as additional sequence information upstream and downstream this genomic region.

2.5.5 Construction of myxobacterial mutants by insertional mutagenesis

For targeted inactivation of genes within the chlorotonil biosynthetic gene cluster a 1 kb homologous region of the respective gene was amplified from So ce1525 genomic DNA by PCR using the required primers. The thus amplified DNA fragment was ligated to pSUPHyg and subsequently transformed into *E. coli* ET12567 harboring the pUB307 helper plasmid. The construct was then transferred into Soce1525 by conjugation as described previously[37]. Obtained mutants were genotypically verified by PCR. Additionally phenotypic analysis was carried out as described below.

2.5.6 Construction of Streptomyces mutants by insertional mutagenesis

For insertion of the chlorotonil halogenase into the attB site of the Streptomyces genome, the halogenase gene was amplified from the So ce1525 genome using the pErmE-Hal_CNH365_fwd (5' CAGATGATATCGATGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCCAGCGGCGTCAGAGAAGGGAGCGGAAATGGAAGCGAACCCCACTGCA-3') and Hal_CNH365_rev (5'- ACATCTAGATCAGGCCGACACCTCCCGAAG-3') primers. An

ermE promoter sequence was attached to the fwd primer. The thus amplified DNA fragment was cloned into pSET152; the resulting construct was transformed into *E. coli* ET12567 harboring the pUZ8002 helper plasmid. The construct was transferred into Streptomyces by conjugation as described by Hopwood et al.[38] Obtained mutants were genotypically verified by PCR. Additionally phenotypic analysis was carried out as described below.

2.5.7 Phenotypic analysis of myxobacterial wildtype and mutant strains

For phenotypic analysis the respective So ce1525 strains were grown in a 50 ml liquid culture on an orbital shaker (30 °C, 180 rpm) in autolaved H media. After three days of cultivation, 2 % autoclaved absorber resin XAD16 were added to the cultures. Cells and XAD were harvested after 10 days of cultivation by centrifugation (10 min, 8000 rpm); the obtained pellet consisting of cells and XAD was frozen at -20 °C and subsequently freeze dried. Pellets were extracted with 30 ml methanol and three portions of acetone (30 ml). The acetone fraction were combined and evaporated. The remaining residue was dissolved in chloroform. For HPLC-MS measurements, the chloroform sample was diluted 1:100 in methanol.

2.5.8 Phenotypic analysis of streptomycete wildtype and mutant strains

For phenotypic analysis of Streptomyces CNH365, the strains were grown in a 50 ml liquid culture in autoclaved CNH365 media. After 7 days of incubation at 28 °C and 200 rpm, 2 % XAD16 were added and the culture was shaken at 200 rpm and 28°C for another two hours. After 7 days of cultivation, mycelium and XAD were spun down at room temperature (10 min, 8000 rpm) the supernatant was discarded and the pellet was washed with desalted water. The mycelium was resuspended and separated from the XAD by filtration through a XAD sieve. To extract Anthracimycin and its derivatives the XAD was extracted with 3 portions of acetone (30 ml each) for 20 minutes respectively. The combined acetone fractions were evaporated and the resulting residue was dissolved in 1 ml chloroform. For HPLC-MS measurements, the chloroform sample was diluted 1:100 in methanol.

2.5.9 Feeding studies

For feeding of [^{13}C-Methyl] methionine, a sterile aqueous [^{13}C-Methyl] methionine stock was produced (1 mg/ml). Three equal cultures were fed with 6 portions of the stock solution starting with 200 µl on the second day, 100 µl on the fourth day and another 50 µl every second day. The cultures were grown for 10 days whilst after 5 days 2% XAD16 were added. All feeding studies were carried out as triplicates.

The mutant strain So ce1525 CtoAT::pKJAT12 which does not produce Chlorotonil A or any of its derivatives anymore, was grown in 20 ml H-media supplemented with hygromycin at 100 µg/ml. 1 mg chlorotonil C2 were suspended in 500 µl DMSO and added to the liquid culture in four portions. On the third day 250 µl stock solution were added to the culture; on day five 125 µl stock solution were added to the cultures and on the following two days 62.5 µl were fed to the culture respectively. After another 24 hours of incubation, 2 % XAD absorber resin were added to the culture. Cells and XAD were harvested by centrifugation on day 10. Freeze dried cell pellets were extracted with methanol and acetone and subjected to HPLC MS measurements. A non-fed control sample was run in parallel. The feeding study with Chlorotonil B was carried out the same way.

2.5.10 Bioactivity testing of chlorotonil A and its congeners

All microorganisms were handled according to standard procedures and were obtained from the German Collection of Microorganisms and Cell Cultures (*Deutsche Sammlung für Mikroorganismen und Zellkulturen*, DSMZ) or were part of our internal strain collection. For microdilution assays, overnight cultures of *E. coli*, *M. luteus*, *B. subtilis*,*S. aureus* in Müller-Hinton broth (0.2 % beef infusion, 0.15 % corn starch, 1.75 % casein peptone; pH 7.4) and *E. faecalis*, *S. pneumoniae* in Tryptic soy broth (1.7 % tryptone, 0.3 % soytone, 0.5 % NaCl, 0.25 % K_2HPO_4, 0.25 % glucose; pH 7.3; microaerophilic growth conditions) were diluted in the growth medium to achieve a final inoculum of ca. 10^6 cfu/ml. Serial dilutions of chlorotonils in THF/Cremophor EL 1:1 were prepared from $CHCl_3$ stocks (10 mg/ml) in sterile 96-well plates. The cell suspension was added and microorganisms were grown for 16-24 h at 37 °C. Growth inhibition was assessed in two independent experiments by visual inspection and given MIC (minimum inhibitory

concentration) values are the lowest concentration of antibiotic at which no visible growth was observed.

The MTT assay for determination of the cytotoxicity of chlorotonil A against human cell lines was carried out as described by Herrmann et al. 2012[39].

Acknowledgements

The authors would like to thank Wolfgang Kessler and Steffen Bernecker and their team for large-scale fermentation, Kerstin Schober for skillful assistance in the isolation of the structural variants of chlorotonil A and Christel Kakoschke for NMR measurements. We are grateful to Michael Hoffmann and Thomas Hoffmann for HPLC-MS measurements and Viktoria Schmitt and Jennifer Herrmann for bioactivity assays.

Accession numbers

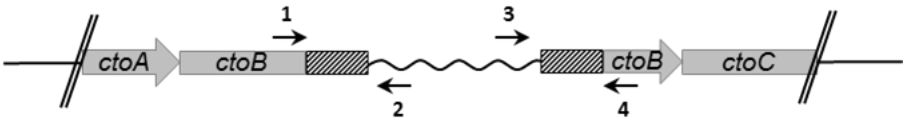
atc_biosynthetic_gene_cluster	KT368179
cto_biosynthetic_gene_cluster	KT368180

Supporting Information

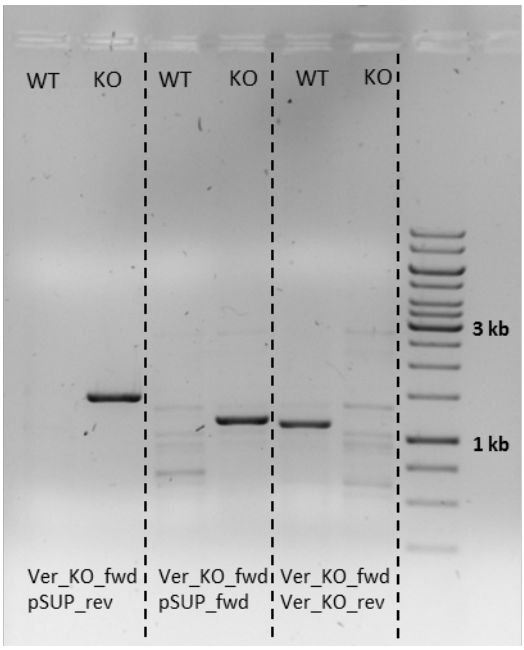
Additional details for analytical and experimental procedures can be found in the supporting information.

2.6 Supporting Information

2.6.1 Genotypic verification of the tandem AT knockout mutant



Primer name	sequence
Ver_KO_fwd(1)	GTCGGGACTGTTGAAGTTGGCGAC
pSUP_rev (2)	GACGAGGTCCTCGAGGCGATC
pSUP_fwd (3)	ATGTAGCACCTGAAGTCAGC
Ver_KO_rev (4)	GGTGATGTCGGCGATATAGG



	expected size	amplicon
	WT	KO
Ver_KO_fwd pSUP_rev	/	~ 1500 bp
pSUP_fwd Ver_KO_rev	/	~ 1200 bp
Ver_KO_fwd Ver_KO_rev	1140 bp	/

Figure S1. Genotypic verification of the AT knockout mutant by PCR. Primers were chosen as indicated in the figure above. Gel electrophoresis of the verification PCR experiments confirmed the expected genotype by showing DNA amplicons of anticipated size.

2.6.2 Alignment tree of AT1 and AT2 of tandem AT domains

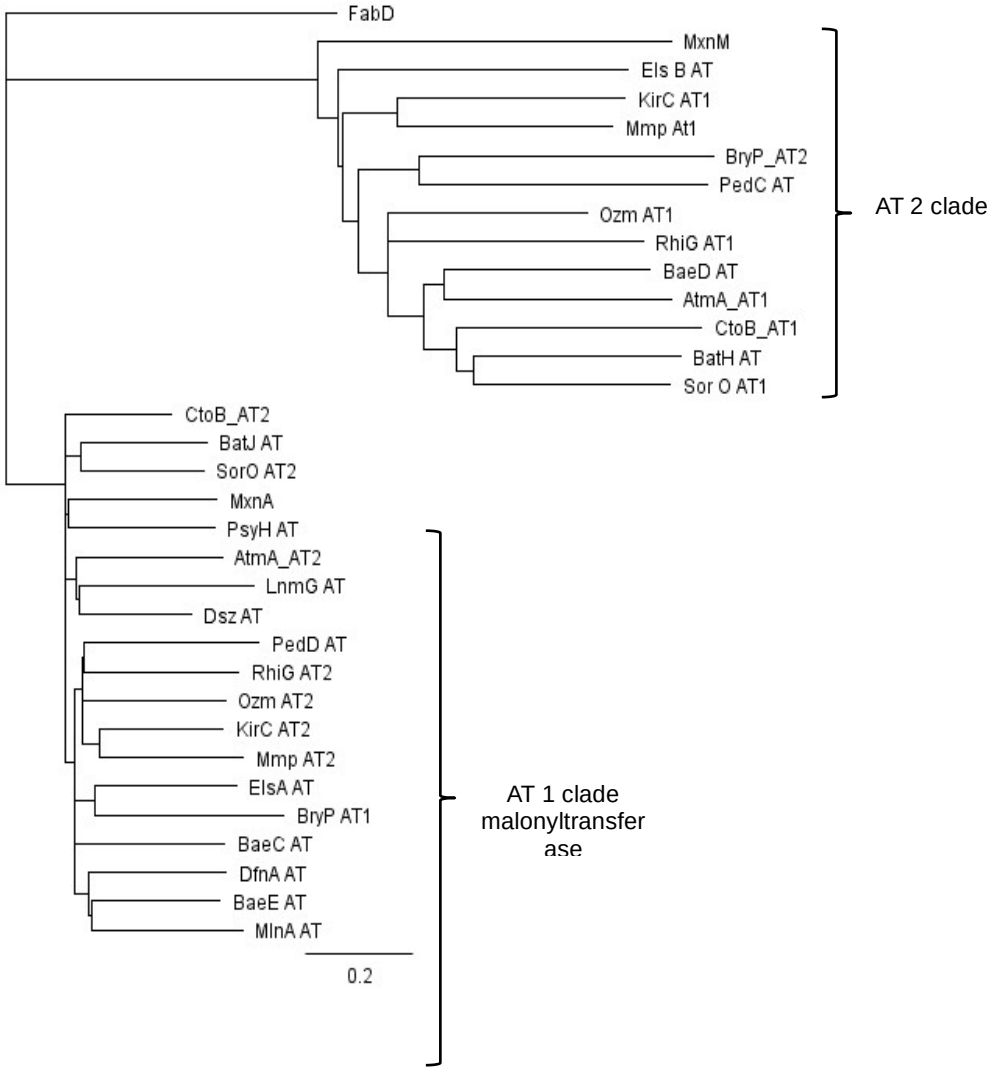


Figure S2. Alignment of the protein sequences was performed by ClustalW; the phylogenetic tree shown was built as Neighbor-Joining tree using the Jukes Cantor substitution model. Branches are labelled with the protein name of the respective AT domain whereas the first AT domain of an AT di-domain is named AT1 and the second AT2. The *E. coli* AT from fatty acid biosynthesis, FabD (YP_489360) was used as an outgroup.

Accession numbers of the AT domains are as follows: BaeC (YP_001421285), BaeD (YP_001421286) and BaeE (YP_001421287) - bacillaene (*Bacillus amyloliquefaciens*); BatH (ADD82949), BatJ (ADD82951) - batumin/kalimantacin (*Pseudomonas fluorescens*); BryP (ABK51299) - bryostatin (*Candidatus*Endobugula sertula); DfnA (ABK51299) - difficidin (*B. amyloliquefaciens*); DszD (AAY32968) - disorazol (*Sorangium cellulosum*); ElsB (YP_003124936) - elansolid (*Chitinophaga pinensis*); KirCI (CAN89639) -kirromycin (*Streptomyces collinus*); LnmG (AAN85520) - leinamycin (*Streptomyces atroolivaceus*); MlnA (YP-001421027) – macrolactin (*Bacillus amyloliquefaciens*); MmpIII (AAM12912) - mupirocin (*Pseudomonas fluorescens*); MxnA (AGS77281) and MxnM (AGS77293) – myxopyronin (*Myxococcus fulvus*), OzmM (ADI12766) - oxazolomycin (*Streptomyces bingchenggensis*); PedC(AAS47559) and PedD (CAE01104) - pederin (*Paederus fuscipes*); PsyH (ADA82589) - psymberin (uncultured symbiont of sponge *Psammocinia* aff. *bulbosa*); RhiG (YP_004029399) - rhizoxin (*Burkholderia rhizoxina*); SorO (ADN68489) - soraphen (*Sorangium cellulosum*). AtmA (this study) – anthracimycin (*Streptomyces* CNH365), CtoB (this study) – chlorotonil (*Sorangium cellulosum*)

2.6.3 Alignment of chlorotonil PKS domains

```
              DxxCSSxL          HGTGT           H
KS_m1       LIDAACASSVVA-----EAHGTGTSL-----NIGHLDA
KS_m2       PVDTACASSLTA-----EAHGTGTAL-----NIGHLES
KS_m3       PVDTACSSSLVA-----EAHGTGTSL-----NIGHLEP
KS_m4       PIDTACSSSLVA-----EAHGTGTKL-----NVGHTLT
KS_m5       ALDTLCSSSLTA-----EAHGTGTAL-----NIGHLES
KS_m6       AVDTMCSSSLTA-----EAHGTGTSL-----NIGHLES
KS_m7       AVNTACSSSLVA-----EAHGTGTKL-----NIGHCVA
KS_m8       PVDTACSSSLVA-----EAHGTGTKL-----NIGHALT
KS_m9       ALDTMCSSSLTA-----EAHGTGTAL-----NIGHLEA
KS_m10      PYNTACSSALVA-----ELHGTGTKL-----NIGHLEP


                HxxxGxxxxP
DH_m2       FLRDHLVGDDLILPGV
DH_m3       IVGDHRVGGQLVFPGV
DH_m4       HVRDHVVEGRPVLPGV
DH_m5       ILADHVVQGHRLLPGM
DH_m6       LLRDHRVNGKTILPGV
DH_m7       TAGHHRVQGNAVVPGV
DH_m8       ALADHIVLGRGTLPGA
DHt         RVALRLWLGG------

              GxGxxGxxxA          LxD
KR_m1       TGGAGELGHAVA-----HAAGVFRPAYF-----SSITAIAAFEGVTDYAYANA
KR_m2       TGGLGGLGLILA-----HTAGVLRDALL-----SSTTALLGNLGQCDYAYANR
KR_m3       VGGAGGIGLEIA-----HSALVLRDRLL-----SSVQSFSGNAGQSNYAAAGT
KR_m4       TGGLGGLGRIFA-----HGAGVLRDALL-----SSMVAVTGNAGQTDYAFACG
KR_m5       GGAGGLGLLFAE-----HAAGVTRDALT-----SSMAAVLGNVGQCDYAFANS
KR_m6       LGGAGGIGLELA-----HSAIVLEDRTI-----SSAQSFSGNPGQSNYAAAST
KR_m7       LGGAGGLGLELG-----HSALVLRDRTV-----SSSQSFLGNAGQSNYAAGCT
KR_m8       TGGAGGLGARLA-----HAAGVLADALL-----SSIAAVVGSAGQCDYGYANA


                GIDSxL
ACP_m1      SFSVYGVDSVLVHHF
ACP_m2      NITDYGVESVGLMEL
ACP_m3-1*   AFTDLGFDAVRLTEL
ACP_m3-2    PYTNLGVDSILSVAI
ACP_m4      PFETYGIDSMMVVGL
ACP_m5      PFERFGIDSVMIMNM
ACP_m6-1*   ALNEIGFDPLSLGAF
ACP_m6-2    QFAEYGIDSILGIDL
ACP_m7      PYSDFGVDSLLAADI
ACP_m8      PFESFGIDSLMSMTL
ACP_m9-1*   DLSACALDRLAAIDL
ACP_m9-2    RFADVGFDSTSFVSL
ACP_m10     EFVEYGIDSILASVI
```

Figure S3 Amino acid alignment (ClustalW) of conserved regions within the chlorotonil PKS domains. Inactive domains are marked with an asterisk. Conserved residues are marked in bold. ACP, acyl carrier protein[40]; KS, ketosynthase[16]; KR, ketoreductase[18,19]; DH, dehydratase[41].

2.6.4 Alignment of anthracimycin PKS domains

```
              DxxCSSxL          HGTGT           H
KS_m1       SIDAACASSLVA-----EAHGTGTSL-----NIGHLDA
KS_m2       PVDTACASSLHA-----EAHGTGTAL-----NIGHLEP
KS_m3       PIDTACSSSLVA-----EIHGTGTAL-----NIGHLES
KS_m4       AVNTACSSSLVA-----EAHGTGTRL-----NIGHLEP
KS_m5       AVDTMCSSSLTA-----EAHGTGTSL-----NVGHTLT
KS_m6       ALDTMCSSSLTA-----EAHGTGTAL-----NIGHLES
KS_m7       AVDTSCSSSLMA-----ECHGTGTRL-----NIGHLES
KS_m8       ALNTACSSSLVA-----EAHGTGTAL-----NIGHCVA
KS_m9       ALDTMCSSSLTA-----EAHGTGTAL-----NIGHALT
KS_m10      PVNTACSSSLVS-----ELHGTGTEL-----NIGHLEA

                HxxxGxxxxP
DH_m2       FLKDHQVDGVPIFPAV
DH_m3       VVAEHRIHGRMVLPGV
DH_m4       YLTDHRVQGYPTLPGV
DH_m5       LVRDHMVDGHALVPAA
DH_m7       LVRDHVVDGKPLLPAA
DH_m8       LVADHVIGGRRVLAGM
DHt         ----------------

              GxGxxGxxxA          LxD
KR_m1       SGGAGGLGRLIT-----HAAGVLRDGLL-----SSVVAVSGNAGQTDYAAANA
KR_m2       SGGAGGIGAHVA-----HCAGVLRDAFL-----SSVAAPLGSSGQAGYAYANS
KR_m3       VGGTGGIGRTLA-----HAAMVLRDSTI-----SSVQSFTGSMGQGNYAAAST
KR_m5       TGALGGLGRRLA-----HLAGVLRDGLV-----SSASGVTGVAGQSDYGFANA
KR_m7       LGGAGGIGAVTA-----HSALVLADSRL-----SSAQSFSGNAGQSNYAAAST
KR_m8       TGGTGHLGLALA-----HAAGVIRDSFL-----SSVVATVGNPGQSDYAYANG

                 GIDSxL
ACP_m1      GFDEYGIDSIRISQF
ACP_m2      SIGDYGFDSLTFNRL
ACP_m3      PLSSYGVDSIVGVEL
ACP_m4      TLDRYGFDSVMALNV
ACP_m5      PFERYGLESVMVMNV
ACP_m6-1*   DLESLGFDTVTLSEL
ACP_m6-2    PHSEFGVDSVLAVEI
ACP_m7      PFTEYGVDSIILVEL
ACP_m8      SFDTYGIDSVMVMRL
ACP_m9      ALGEYGFESVSLKAL
ACP_m10     DLKDFGVDSMLSAMI
```

Figure S4 Amino acid alignment (ClustalW) of conserved regions within the anthracimycin PKS domains. Inactive domains are marked with an asterisk. Conserved residues are marked in bold. ACP, acyl carrier protein[40]; KS, ketosynthase[16]; KR, ketoreductase[18,19]; DH, dehydratase[41].

2.6.5 Alignment of ER domains from the anthracimycin and the chlorotonil PKS in comparison with known ER domains

```
                                                    Y
Fkb_ER6         […]EPRRP---LAPAEVRIGVRAAGLNFRDVLIALGTYPGQ----GVLGGE 53... […]
Fkb_ER7         […]TPRRA---LEAGEVRIDVRAAGLNFRDVLIALGTYTGA----TAMGGE 53... […]
Chm_ER5         […]TDTVA---LAPGQVRIAVRAAGLNFRDTLIALGMYPGE----GVMGAE 53... […]
Myc_ER5         […]TATVS---LLPGQVRIAVRAAGLIFRDTLIALGVYPDQ----AAMGAE 52... […]
Tyl_ER5         […]DAEEP---LAPGQVRIAVRAAGVNFRDALIALGMYPGK----GTMGAE 53... […]
Pik_ER4         […]AETLAPEPLGPGQVRIAIRATGLNFRDVLIALGMYPDP----ALMGTE 56... […]
Ole_ER4         […]DASRP---LGPDEVRIAVRAAGVNFRDVLLALGMYPDE----GLMGAE 53... […]
Ery_ER4         […]DVEQP---LRAGEVRVDVRATGVNFRDVLLALGMYPQK----ADMGTE 53... […]
Meg_ER4         […]DADRP---LAPEEVRVAVRATGVNFRDVLLALGMYPEP----AEMGTE 53... […]
Atc_ER3         […]PVAPG-----PGEVEVQVKAFPVNFSDFLAAKGLYPMMPDFPFTPGVE 49... […]
Cto_ER6         […]PPDPG-----EGEVQVLVRAFSLNFGDLLCVKGLYPTMPDYPFTPGFE 50... […]

                                                LxHxxxGGVGxxAxxxA
Fkb_ER6         AFATAWYGLVDLAGLRPGEKVLIHAATGGVGSAARQIARHLGAEVYATTSAAKRHLVDLD 170
Fkb_ER7         VFATAWYGLVDLGTLRAGEKVLVHAATGGVGMAAAQIARHLGAELYATASTGKQHVLRAA 170
Chm_ER5         VFLSAYYGLRHLAGLRAGQSVLVHAAAGGVGMAAVQLARHFGAEVFGTAGTAKWDALRAQ 170
Myc_ER5         VFLSAYYALKHLARAVPGQSILVHAAAGGVGMAALQLARHLGLEVYGTASPGKWDVLRDQ 169
Tyl_ER5         VLLTSYYALTRLARARTGQTVLVHAAAGGVGMATLQLARHLGLEVYATASTGKWDALQKH 170
Pik_ER4         VFLTAVYALRDLADVKPGERLLVHSAAGGVGMAAVQLARHWGVEVHGTASHGKWDALRAL 173
Ole_ER4         AFLTAYYALHDLAGLRGGESVLVHSAAGGVGMAAVQLARHWDAEVFGTASKGKWDVLAAQ 170
Ery_ER4         AYTTAHYALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAGPAKHGTLRAL 170
Meg_ER4         AFTTAHYALHDLAGLQAGQSVLVHAAAGGVGMAAVALARRAGAEVFATASPAKHPTLRAL 170
Atc_ER3         AFLAMSLAF-ERAAVRPGERVLIRAATGTNGLMAVQLAQLAGAEIIATAGS-RHKVDHLI 167
Cto_ER6         VFLTMHRAF-ELARVTRGERVLIHTAAGGTGLIAVQLAREIGAEIYATAGS-REKLDALA 168
```

Figure S5 Extract from a ClustalW alignment of the ER domains present in the biosynthetic gene clusters of chlorotonil and anthracimycin in comparison with a set of representative ER domains from other PKSs as reported by Kwan et al..[6] Cto_ER6 is the ER domain found in module six of the chlorotonil PKs whilst Atc_ER3 is the ER domain found in module three of the anthracimycin pathway.The conserved tyrosine residues correlated with 2S configuration of the polyketide are marked in bold. The NADPH binding site is also indicated in bold letters.

2.6.6 Alignment of methyltransferase domains from the the chlorotonil biosynthetic pathway

```
MT1             --------------------------------------------------VGRIYKGNTTS 11
MT3             --------------------------------------------------VENIYKGNRLT 11
MT2             --------------------------------------------------VEGVYKGNPVA 11
MT-ctoF         MDLDTIAMLRDPTTHAKLRLASGPEGERLINVDSGASYPIREGVPAFVSRADLTDANQRF 60
                                                                   : ..*

MT1             DYYNGLVASAVRAHLEAAAAALPPG-----RKTRILEIGAGTGGTTALVLDAISSHAGRF 66
MT3             DHFNQVVAWSVRAGLRERLQAPRGG-----EAAAIVEIGAGTGGTTGAVLDGLGAWAKDV 66
MT2             DWCNRLVAEGVQAYVDELISRAAPD-----GSVKILEVGAGTGGTTQAVLPALG-HRDAI 65
MT-ctoF         TKFYDLFAPVYDAMISGYLWVRRLGNDESFRREYLKDLGVKDHDRILEISIGTGRNSRYL 120
                     :.*    * :         .         : ::*.   .    :  . .     .

MT1             EYVFTDVSAGFVKHGQRQLGSSNPGMEFGVLDIERDVEGQGYAPGSVDIVIATNVLHATR 126
MT3             RYVYTDVSAGFTQHGRRRFGAAHPGLEFRVLDIERAPDGQGFAPGEFDVAIATNVLHATK 126
MT2             GYWYTDLSVGFTQSGKKRYGADHPFVRFEILNIEDEVRSQGYTPGAFDVVIAANVLHATR 125
MT-ctoF         PKTCEFFGVDISWGMIKRCKAAMQEQGLRHRLFLANGEYLPFEDDVFDAVFHVGGIKFFN 180
                      ....:     ::  :      :    :        :  . .* .: .. ::  .

MT1             RIADTLARAKALLRTNGWMILNEVTVAQRFTTLTFGLLDGWWLFEDGDVRLPGSPLLDAA 186
MT3             NVARTLSNVARLLRPGGLLVLNEVTAVQRFTTLTFGLLDGWWLFDDEGARLPGSPLLDAP 186
MT2             SLRSTLRNVKSLLKANGVLVLNEVTEVHELTTLTFGLLDGWWLYDDPELRIAGSPLLDPP 185
MT-ctoF         DKARAIREMIRVAKPGAKIMIVDPTEMT-MRDVEGTPIAKSFFKGIKDLTTAPVDLVPKD 239
                    :: .   : :... ::: : *    :  :    :   ::        .   *:

MT1             GWRRLLAEEGFEKSVTLGQVGRDGRGLGQSVIIAES 222
MT3             AWTARLREAGFGAVTSFGLDDGTGRRVGQHVLVAT- 221
MT2             LWDRILAEEGFVSRHHLGGAAADVGSLGQHVIMAES 221
MT-ctoF         MRDVELRDILQGRFYCLQFRKPERAGAAADPTHAMR 275
                     * :        :          .     *
```

Figure S6 ClustalW alignment of the four methyltransferase functionalities present in the chlorotonil biosynthetic gene cluster. Conserved residues are marked by an asterisk.

2.6.7 Alignment of the methyltransferase CtoF and its closest homologue

```
CtoF_           MDLDTIAMLRDPTT—HAKLRLASGPEG---ERLINVDSGASYPIREGVPAFVSRADLTD 55
WP_013119426    MKPENVQFLCSPHTGERLELKVEQGAGGRPVELLVGQKTGAKFPVREGIPLFLDKSQMPG 60
                . :.: : .* *  : :*:: .*. *   * *:. .:**.:*:***:* *:.::::..

CtoF_           ANQRFTKFYDLFAPVYDAMISGYLWVRRLGNDESFRREYLKDLGVKDHDRILEISIGTGR 115
WP_013119426    NNRKQQKFYNLVAPFYDFLHR—MQTARQGGEHLMRMEYIKELGIGDKDKVLEVSIGTGA 118
                 *::  ***:*.**.** :    : . * .:. : **:*:**: :::**:*****

CtoF_           NSRYLPKTCEFFGVDISWGMIKRCKAAMQEQGLRHRLFLANGEYLPFEDDVFDAVFHVGG 175
WP_013119426    NLLYLPRTAQCFGIDISWEMLRRCKRSMQRKGIDVELAMAAAEKLPFVDNVFDVVFNVLG 178
                *  ***:*.: **:**** *::*** :**.:*:  .* :* .* *** *:***.**:* *

CtoF_           IKFFNDKARAIREMIRVAKPGAKIMIVDPTEMTMRDVEGTPIAKSFFKGIKDLTTAPVDL 235
WP_013119426    LRLFNDKGGAIREMLRVAKPGARITIVD----------------------QAKAGAPLHL 216
                :::****. *****:*******:* ***                      :  : **:.*

CtoF_           VPKDMRDVELRDILQGRFYCLQFRKPERAGAAADPTHAMR 275
WP_013119426    LPREIVGPEYKEIYYGHLYCLSFKKAEKK----------- 245
                  :*::: . * ::*   *::***.*:*.*:
```

Figure S7 ClustalW alignment of the the free standing methyltransferase CtoF and its closest congener from *Thermicola potens*. Conserved residues are marked by an asterisk.

2.6.8 Alignment of the methyltransferase CtoF and the methyltransferase domains from the anthracimycin biosynthetic gene cluster

```
atcD_cMT      ----------------------------VASVYRGA—RLVDHCNRLLAEHAVRHIRDTA 30
atcE_cMT      ----------------------------VENVYKGD—PLASGANRAAAREVVARAAGSR 30
atcF_cMT      ----------------------------VAEIYKGS—IQSEFYNRLVAEVVEEHTRHFA 30
CtoF          MDLDTIAMLRDPTTHAKLRLASGPEGERLINVDSGASYPIREGVPAFVSRADLTDANQRF 60
                                         : .:  *      .      :.

atcD_cMT      P---GRRPRILEIGAG-------TGATTAEVLTALRDAGIEADYDYTDVSARFLHDA--- 77
atcE_cMT      NRE-GRPLRILEVGAG-------TGGTTRTVLSALSPYADRVEYVYTDVSTGFLRHG--- 79
atcF_cMT      RRYTRSRPQLFEVGAG-------TGGTSTFVFDALKNLGDRVTYHYTDIGGAFLRAA--- 80
CtoF          TKFYDLFAPVYDAMISGYLWVRRLGNDESFRREYLKDLGVKDHDRILEISIGTGRNSRYL 120
                       : :   .        *         *   . .      ::.    : .

atcD_cMT      ---------ERKWAGGRDVRMSFRPLDIGRPLQAQG-----FGQAEYDVVIAANVLHAVA 123
atcE_cMT      ---------RKEFAS-RYPFVSFKRLDIERSPGDQG-----FEEGGFDVVVAANVLHATQ 124
atcF_cMT      ---------EQQFGT-NHPYIEFDTFDVERSPAVQG-----YEPGTMDVVVASNVLHTTR 125
CtoF          PKTCEFFGVDISWGMIKRCKAAMQEQGLRHRLFLANGEYLPFEDDVFDAVFHVGGIKFFN 180
                        .:.  .     :   .: :     .     :     ..  . ::

atcD_cMT      DLDAALQRTRALLAPSGRLLLTEVTRALAFHTVTFGLLDGWWAYTDDSRRLPSSPLLDTE 183
atcE_cMT      NLDRTLTHVRRLLTADGRLVLCETTAFSAFTTLTFGLLDGWWRFEDGFRRIPASPLAEPA 184
atcF_cMT      SIDVTLAHCRQLLKPGGILVINELTHRLDYNTLTFGLTTGWWLFKDEDKRVQGSPLLDVR 185
CtoF          DKARAIREMIRVAKPGAKIMIVDPTEMTMRDVEGTPIAK---SFFKGIKDLTTAPVDLVP 237
              .   :: .   :  ... ::: : *      .    :      : .  : :  :*:

atcD_cMT      MWRRRLEFTGFTEVDVHG-EADVPG—ELSQRVVTATP 218
atcE_cMT      TWQALLRDKGFEQTVVHDAVAGAPR—SLGLHVLVADL 220
atcF_cMT      GWRDALRRTGFDEVEIRGLVDENENENDQAQCVIVA-- 221
CtoF          KDMRDVELRDILQGRFYCLQFRKPERAGAAADPTHAMR 275
                   :.  .: :  .             .     *
```

Figure S7.2 ClustalW alignment of the the free standing methyltransferase CtoF and the methyltransferase domains from the anthracimycin biosynthetic gene cluster. Conserved residues are marked by an asterisk.

2.6.9 Alignment of the halogenase CtoA and its closest homologue

```
CtoA            MEANPTAG-------TEVVVIGAGIVGVHSAIQFAKRGLKVVLIDNIVGQKKSFKVGESL 53
YP_005628289    MDMSSTAAHQYPPKSNQIVVIGAGIIGLTNALQLAKRGLSVTLIDNLQARRHSFKVGESF 60
                *: ..**.       .::*******:*: .*:*:*****.*.****: .:::*******:

CtoA            LVFSNMFLRTISELDEFN-QKCFPKHGVWFTYGMEGTTSFEEKAEWALEST—LPQAMRD 110
YP_005628289    LVFTSAFLRTIGELDTFITERSFIKLGVWFSAGAEYRHDFGETTEWAVNADPHPPHYLYD 120
                ***:. *****.*** *  ::.* * ****: * *   .* *.:***:::    *: : *

CtoA            AFANKALLRAMADDVQIVRPEAEELMQQTARAHPNITFLDTAKVTNVVIAEGGGPHEVTW 170
YP_005628289    LAPDKKWFRCMFLDMQIVRPEAEAVMLKAVREHPNITLIDNVRIRKVVLAEDDGPHLMHW 180
                  .:*  :*.*  *:******** :* ::.* *****::*..:: :**:**..*** : *

CtoA            ECKATQRTGVVRTTWLIDCSGRNRLLAKKLKHAAEDVELNDGFKTTAVWGQFSGIKDEMF 230
YP_005628289    QGQGCEGTSDAR—WLVDCSGRSRLLARQQGHTAEKHEQRDGFQTTAIWGQFEHAGDAVF 238
                : :. : . .  **:*****.****::  *:**. * .***:***:****.   * :*

CtoA            GENWVNRTSDGARSKRDLNTLHLWGDGYWIWVIRLSEGRISVGATYDQRRPPAGAGYREK 290
YP_005628289    DARWRYLLQDGRSTPRDLYTVHLWGRGYWIWVIRLSQDRISVGVTFDQRTPPEGATPAEK 298
                . .*    .**  : *** *:**** **********:.*****.*:*** ** **   **

CtoA            FWDIIRRYPLFDGMLSDDNMLEFHVFKNCQHITDTFVSEKRYGMIGDAASVIDAYYSQGV 350
YP_005628289    FWNLIRRYPLLDGIINEQTLLELQTFRNVQHWTDTFVSPRRYAMAGDSASIIDAYYSQGM 358
                **::******:**::.::.:**::.*:* ** ****** :**.* **:**:********:

CtoA            SLALVTSWHITNIMERDLRERRLDKEYIARVNRHTRQDWHIMRNMVIEKYTSAMADGRFF 410
YP_005628289    ALGMVTSWHIANVVQRDLLDQHLDRGYIKRINHATRQDWHMLRNMVHQKYSPAIEDSRFF 418
                :*.:******:*:::*** :::**: ** :: ******::**** :**:.*: *.***

CtoA            VMTHLLDMIIFVGAAFPRYLLVRWLVETQGSTARETPVMREMRRYLEENLYYSKIG---S 467
YP_005628289    LLSHMLDMAVFWCMGSTRSALTHWLVQTEGQSSRETAQLRRLRQRLESRLFYSRSPYWLW 478
                :::*:*** :*   . .*  *.:***:*:*.::***. :*.:*: **..*:**:

CtoA            LAPEKVQKVQRGLQASLSERARWRVENGVKVTRLKAIVHAPSGLLKFWKLPLSGQREFED 527
YP_005628289    LSPERVQRLQAHLQRQLARRARWRQDNGIRLPRVTSFLSLTAPLPRVWKLFGGAKAGHID 538
                *:**:**::*  ** .*:.***** :**:::.*:.:::  .: * :.***  ..:  . *

CtoA            ISPKPVKQIPKWLAMTGEETNPRMLKMARPLMASTFFLMYGYDGLSTAVTKVRQRLERLP 587
YP_005628289    VSARDLVQPAEQRPAG---SATWFDRLPISPMTRLGWVIR-----LRPLGLLLQFALCYA 590
                :*.: : * .:  .     : . : ::. . *:   :::        .:  : *     .

CtoA            GAAATAETTAAGRRGEAPEPAMNGAAPVRNVLREVSA 624
YP_005628289    WDAAETRIRRGWRDANRVHAAVNDVTSPQQDAR---- 623
                  ** :.   . * .:  ..*:*..:. ::  *
```

Figure S8 ClustalW alignment of the the free-standing FAD dependent halogenase CtoA and its closest congener from *Xanthomonas oryzae*. Conserved residues are marked by an asterisk.

2.6.10 Comparative analysis of the *atc* and the *cto* biosynthetic gene cluster

The domain distribution amongst the *atcDEF* genes differs from the arrangement seen for *ctoCDE* in the chlorotonil gene cluster. *AtcD* harbors the first four modules (like *ctoC*), while *atcE* is significantly shorter than its counterpart in the chlorotonil biosynthetic gene cluster. This is due to the fact that *atcE* only harbors modules five to seven whereas the remaining three modules eight to ten are encoded on *atcF*. In addition to that, the anthracimycin PKS lacks one reductive loop (module 6) and a KR domain (module 4) and does not exhibit superfluous non-functional ACP domains except in module six. The first three modules of the anthracimycin PKS closely resemble their counterparts in chlorotonil biosynthesis (including predicted KS substrate specificity), except for the ER domain found in module three which is not required to act on the polyketide intermediate at that stage. In contrast to the chlorotonil PKS, no additional ACP domain is found in the methyl-branching module three and the truncated DH domain known from module one of the chlorotonil PKS is missing, whereas a similar truncated DH domain can be found in anthracimycin module four. *AtcE* encodes modules five, six, and seven and the ACP domain of module four. These three modules do not match the anticipated domain arrangement on the basis of retro-biosynthetic analysis: module five (KS-DH-KR-ACP) still exhibits the required domains, but the minimal module six (KS-ACP-ACP) is lacking a full reductive loop and a methyltransferase functionality. This short module contains two ACP domains, but as previously reported for the chlorotonil PKS, the first ACP domain lacks the serine residue for attachment of the phosphopantetheine arm. In contrast, module seven features a methyltransferase domain which is not required in the seventh round of chain extension. The last three modules encoded in *atcF* and the order of encoded domains match the expectation from retrobiosynthetic analysis.

2.6.11 A model for anthracimycin biosynthesis

In the following we propose a model for biosynthesis of anthracimycin under the assumption that the identified polyketide megasynthase, complemented with a few additional functionalities, is sufficient for its production. The starter unit for anthracimycin biosynthesis is acetyl-CoA as inferred from the chemical structure. Notably, in contrast to chlorotonil biosynthesis the KR domain in module one belongs to the B-type clade which results in formation of the respective R-configured alcohol.[18] Assuming that the ER domain in module three does not act upon the growing polyketide chain at that stage, the following two cycles of chain extension resemble chlorotonil biosynthesis: in module two, an α,β-olefinic moiety extends the polyketide chain whilst in module three an α-methylated, β,γ-olefinic intermediate is formed following a double bond shift. For the chain extension in module four, a ketoreduction step is required to generate the proposed double bond. However, no such domain is encoded in this module, leading to the requirement that a KR domain present in another module or an external enzyme must assist

formation of the intermediate. Module five then agrees with the colinearity rule again and catalyzes the production of an α,β-olefinic intermediate. The function of module six is, however, questionable at this step. A module exhibiting a full reductive loop and a methyltransferase domain is expected but only a minimal module comprising a KS domain and two ACP domains (one of which is inactive) is found. Several scenarios are therefore considered: the growing polyketide chain could be shifted back and passed through to module three including the enoylreductase domain encoded there. Iterative use of this module cannot be excluded, albeit at present no bioinformatic tools exist to predict such behaviour. Alternatively, module six could recruit the missing functionalities in *trans,* whereas the ER functionality could be recruited from the *trans*-ER in AtcA. This possibility has been previously proposed for *trans*-AT systems, for example during sorangicin biosynthesis[35]. The situation seen in module two of the sorangicin PKS is comparable, as in this model extension with an alkylic, α-methylated extender is also suggested whereas the required domain set (KR, DH, ER, MT) is missing. Interestingly, phylogenetic analysis of KS7 of the anthracimycin pathway revealed α,β-keto substrate specificity. This finding would be in accordance with the present domain set (KS-ACP) but does not fit the retrobiosynthetic proposal. Furthermore, no β-keto substructure is present in mature anthracimycin. In module seven an α,β-olefinic extender is again incorporated; thus, the available MT domain does evidently not act upon the polyketide chain. The last three extension steps in module eight, nine and ten seem to operate according to textbook PKS logic: the detected domains fit the retro-biosynthetic analysis and the presumed intermediates. Notably, in the anthracimycin PKS module eight is not a split module but is encoded on only one gene, *atcF*. Finally, a rather unusual domain placement is seen in module ten, where the methyltransferase likely catalyzing the α-methylation at C2 is located downstream of the thioesterase domain. Nevertheless, this uncommon positioning does not seem to influence the MTs activity. More detailed analysis of this MT domain, most likely responsible for C2 methylation in anthracimycin, in comparison to CtoF catalyzing C2 methylation in chlorotonil revealed no obvious commonalities (Fig.S7.2). As already shown for the MT domains encoded in the chlorotonil biosynthetic gene cluster, the methyltransferase domains encoded within the PKS differ significantly from the free standing methyltransferase CtoF. As for chlorotonil, we propose a spontaneous [4+2] cycloaddition taking place in module eight, leading to the decalin moiety. In module ten, the polyketide chain is released from the assembly line and cyclized by the TE domain resulting in the final product, anthracimycin.

2.6.12 ^{13}C-methyl-methionine feeding

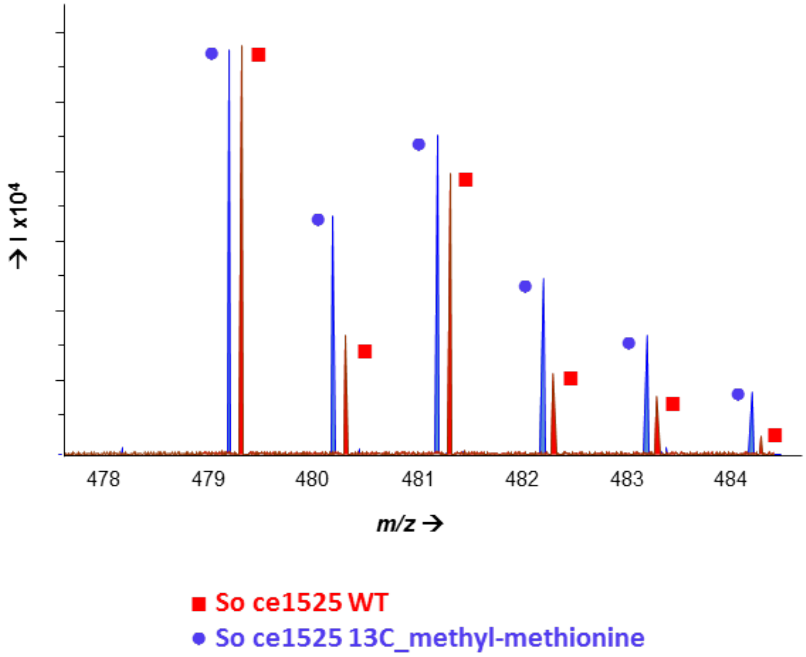


Figure S9 Overlaid isotope patterns of the selected mass of chlorotonil A [M+H]+ = 479.175 found in extracts of So ce 1525 wildtype strain. The isotope pattern marked with squares belongs to the mass found in the non-fed So ce 1525 WT control sample. The isotope pattern marked with dots was found in extracts of the So ce1525 WT culture fed with ^{13}C-methylmethionine as described in the experimental section. To enhance optical comparability the two overlaying patterns were slightly shifted in the spectrum viewer.

2.6.13 Coverage of the chlorotonil biosynthetic gene cluster by cosmids

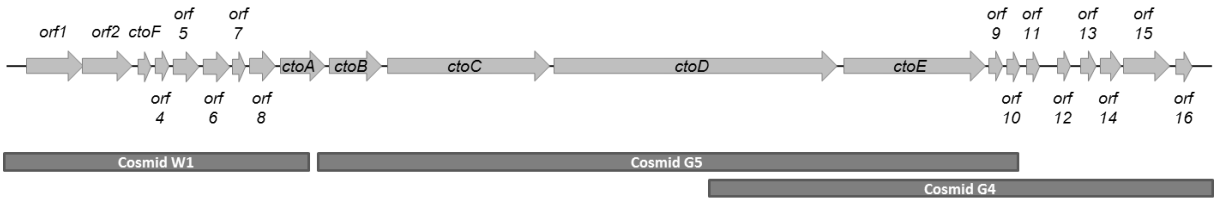
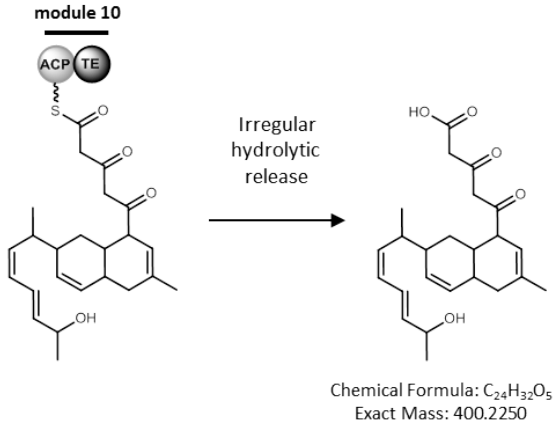


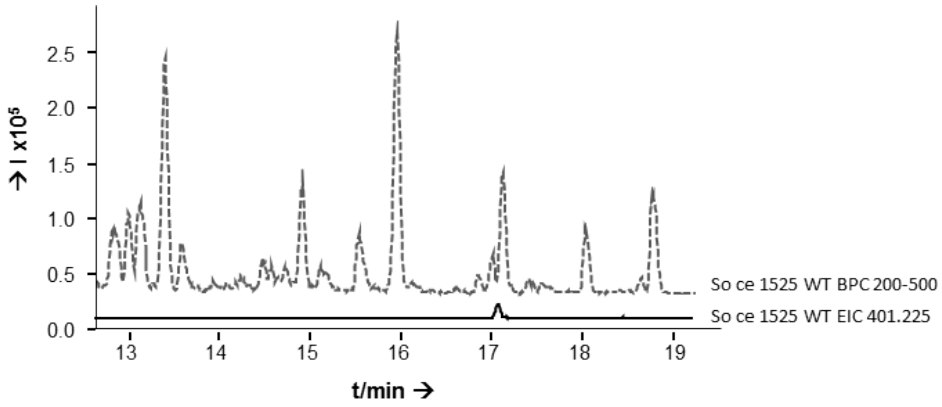
Figure S10 Coverage of the chlorotonil biosynthetic gene cluster through the identified hit cosmids W1, G5 and G4. The 1 kb sequence gap between cosmid W1 and cosmid G5 could be filled with existing data from 454 sequencing.

2.6.14 Diels Alder-like [4+2] cycloaddition reaction during chlorotonil biosynthesis

A.



B.



C.

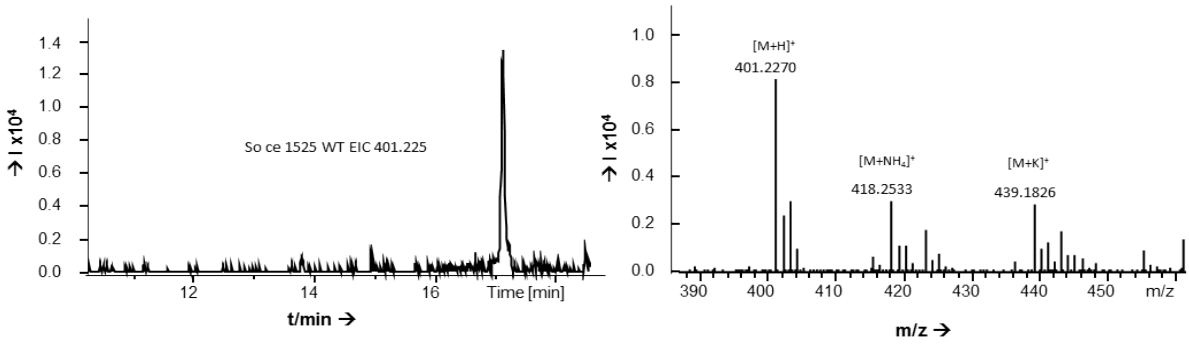
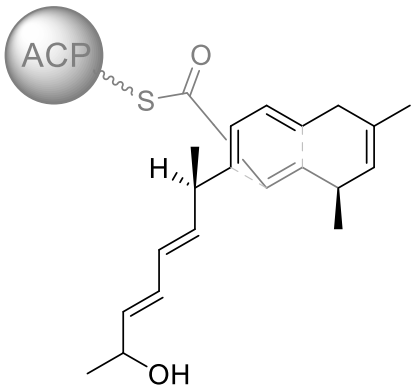


Figure S11 Screening of So ce1525 wildtype strain extracts for the presence of chlorotonil precursor molecules with intact di-ring structures as a result of intramolecular Diel-Alder cyclization, prior to final circularization of the macrolide ring. A.: schematic depiction of possible hydrolytic release of a chlorotonil precursor from the PKS assembly line. B.: HPLC chromatogram showing a BPC of a So ce1525 wildtype extract and an EIC targeting the mass of the wrongly released chlorotonil precursor. C.: EIC of the precursors mass (M+H)+ = 401.225; isotope pattern of the selected mass is shown to the right.

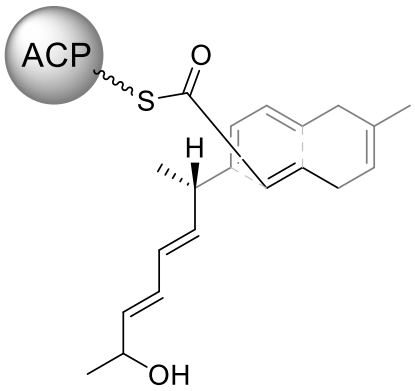
2.6.15 Hypothesis on the diels-alder-like ring formation during chlorotonil and anthracimycin biosynthesis

1. Chlorotonil



Stereochemistry observed after [4+2] cycloaddition of the protein-bound chlorotonil precursor is most likely influenced by the configuration of the methyl-group at C-16 as depicted above. Inverse orientation of the C16- methyl would lead to steric clash with the protein. With the C-16 methyl group directing out of the image plane, and the grey colored substructure of the molecule pointing backwards, the resulting configuration allows for cyclization as depicted above leading to the ring stereochemistry observed in chlorotonil A.

2. Anthracimycin



The opposite orientation of the C-16 methyl group in anthracimycin would consequently lead to inversed stereochemistry after [4+2] cycloaddition.

Figure S11.2 Illustration of the opposite orientation of chlorotonil and anthracimycin prior to [4+2] cycloaddition caused by the differently oriented C16-methyl groups. Black colored parts of the molecule are located in the drawing plane or in front, whilst grey coloured parts point backwards. Dashed lines indicate the ring closure during the cycloaddition.

2.6.16 Feeding of chlorotonil C2 to So ce1525 *ctoB::pKJAT12*

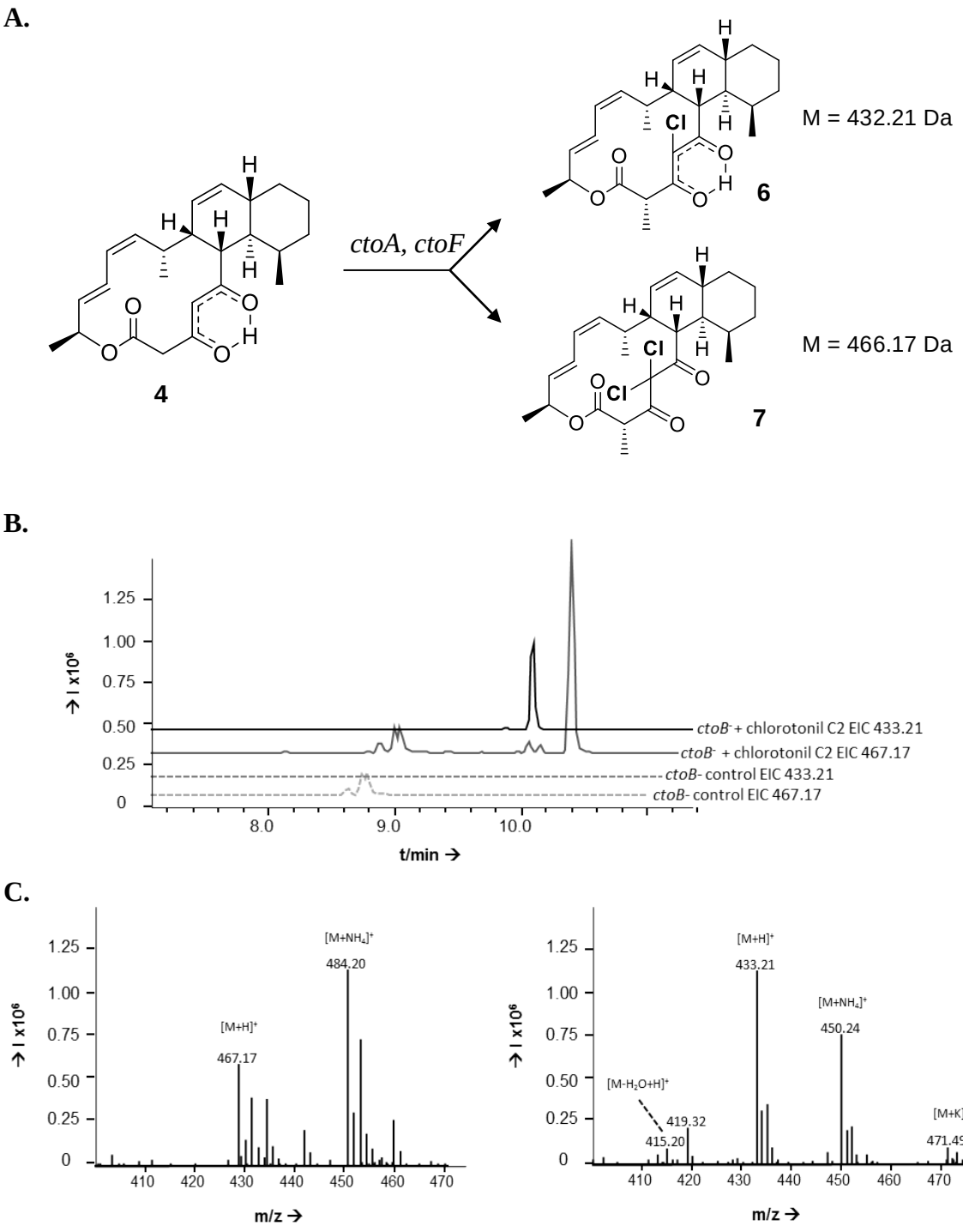


Figure S12 Results from the feeding of chlorotonil C2 to the chlorotonil negative mutant So ce1525 *ctoB::pKJAT12*. A.: conversion of chlorotonil C2 to its mono-methylated, mono-halogenated and the respective di-halogenated from. B.: Extracts of the fed culture showing the sought chlorotonil C2 derivatives in comparison to the control where the respective derivatives are not present. C.: MS spectra showing the isotope pattern of the two chlorotonil derivatives.

2.6.17 HPLC chromatograms of the chlorotonil producer strain So ce1525 showing EIC of chlorotonil precursor molecules

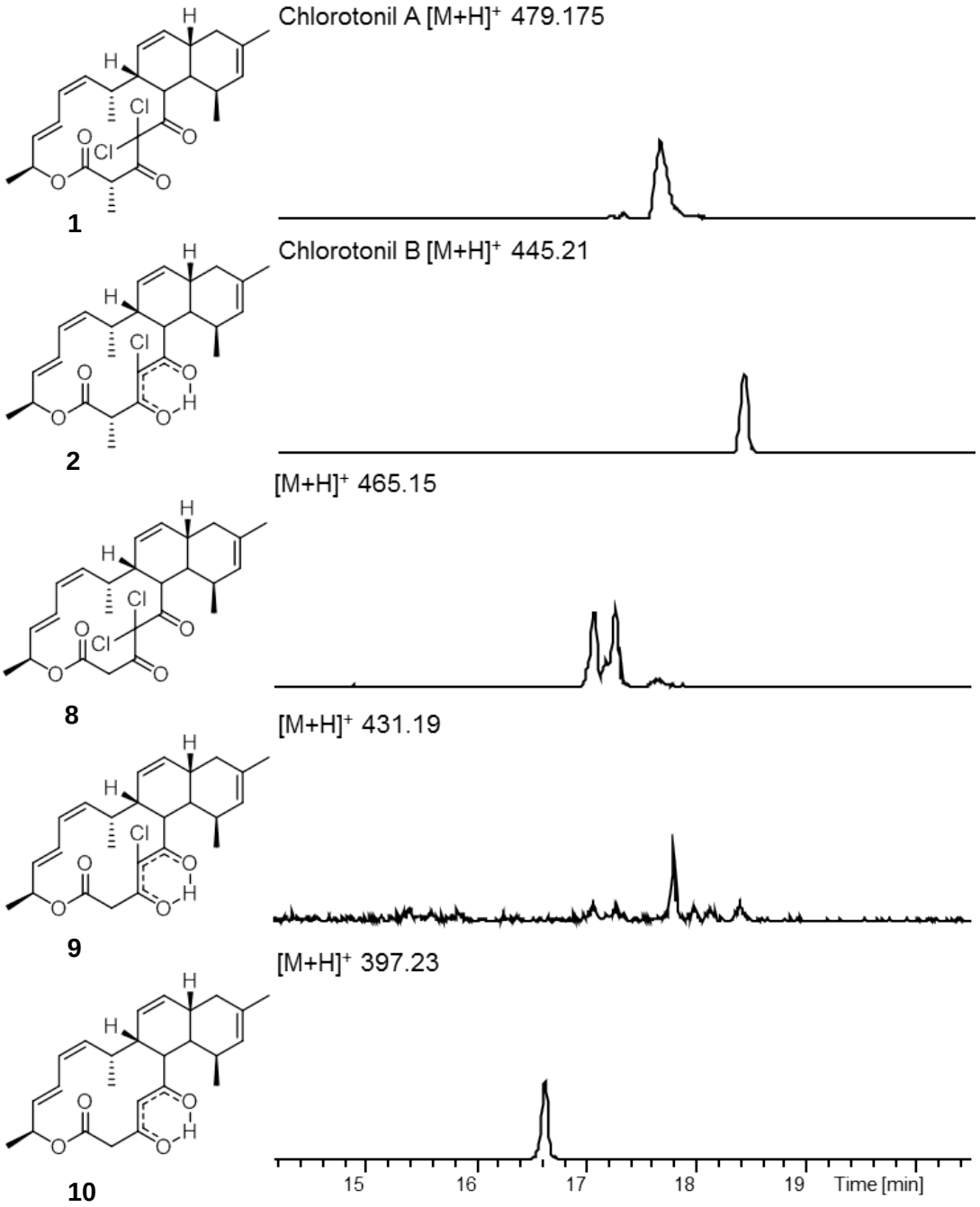
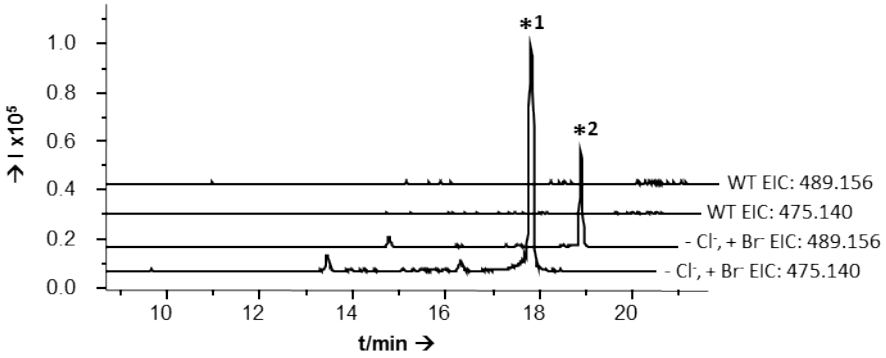


Figure S13 Screening of So ce1525 wildtype strain extracts for the presence of chlorotonil precursor molecules lacking methylation at C2 and having a different degree of halogenation at C4.

2.6.18 Bromine feeding of So ce1525 WT

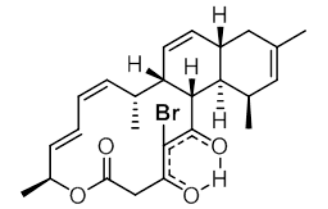
In order to further exploit the functionality of the chlorotonil halogenase and its capability to carry out bromination reactions another feeding experiment with the natural producer strain So ce1525 was carried out. The sodium source in the media was depleted and replaced with equimolar amounts of the respective bromide salt. This media composition significantly deteriorated the growth of So ce1525 which was consequently grown for 14 instead of 10 days. In accordance with deprivation of the chlorine source only trace amounts of chlorotonil A and B were produced, but the non-methylated, non-halogenated precursor molecule $[M+H]^+$ = 397.23 accumulated. In agreement with our expectations substantial amounts of the respective brominated compounds were biosynthesized. A mono- and a C2-methylated, mono-brominated derivative were found, while only trace amounts of the methylated and un-methylated, di-brominated derivative were found.

A.



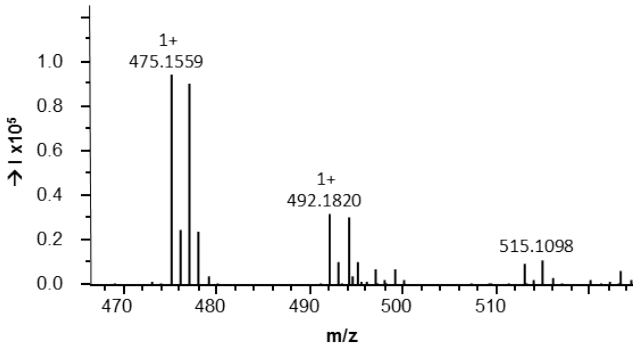
B.

***1**



Chemical Formula: $C_{25}H_{31}BrO_4$
Exact Mass: 474.1406

11



***2**

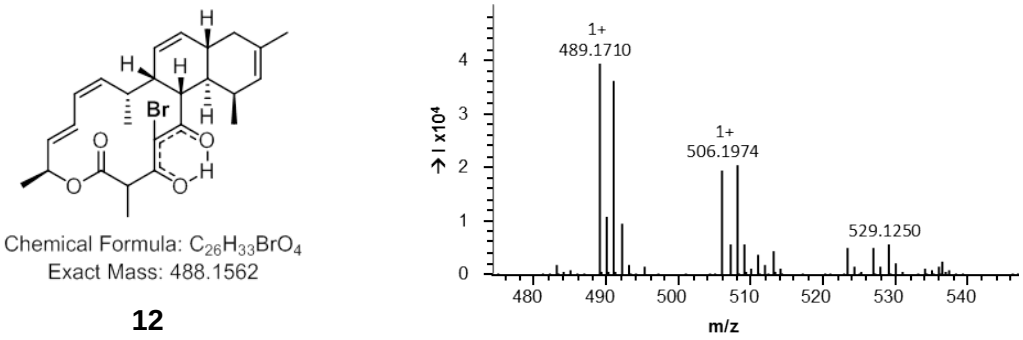
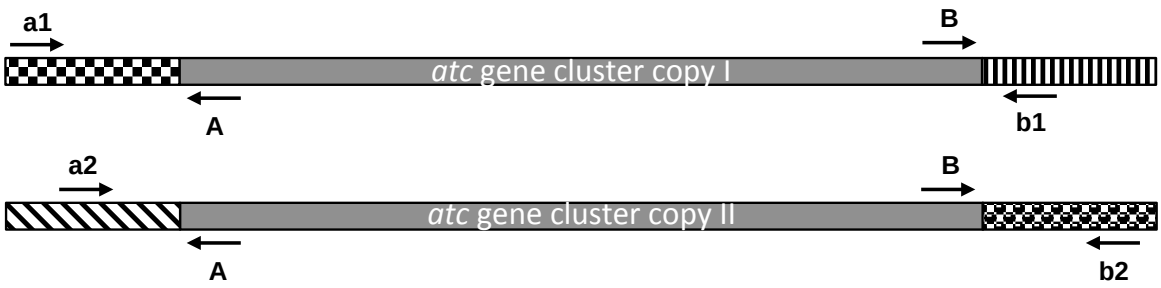


Figure S14 A HPLC chromatogram of Soce1525 wildtype extracts. The chromatograms labelled with "-Cl-, +Br-" represent EIC spectra from extracts of wildtype strain cultivated without chlorine but with addition of a bromine source. The two chromatograms labelled with "WT" represent EIC spectra from extracts of So ce1525 grown in H-media. The two peaks representing the brominated chlorotonil derivatives are marked with an asterisk. B Structural formulas of the respective brominated derivatives and the isotope pattern of the peaks representing them are shown.

2.6.19 PCR proof for the presence of two identical anthracimycin biosynthetic gene clusters in the genome of *Streptomyces* CNH365

In addition to the sequencing data obtained for Streptomyces CNH365 a PCR experiment was carried out in order to corroborate the findings from the sequencing. This was based on the assumption, that if there were two identical gene clusters there should be four differential genomic regions up- and downstream the gene clusters; one upstream and one downstream region for every genomic locus. Thus a set of primers was designed. The A/B primer pair binds at the borders of the identical gene clusters whilst the a1/a2, b1/b2 primers bind up- and downstream in defined distance to the respective anthracimycin gene cluster copy. Indeed, the PCR experiment yielded four different amplicons of the expected size and thus confirms the presence of two identical anthracimycin biosynthetic gene clusters.

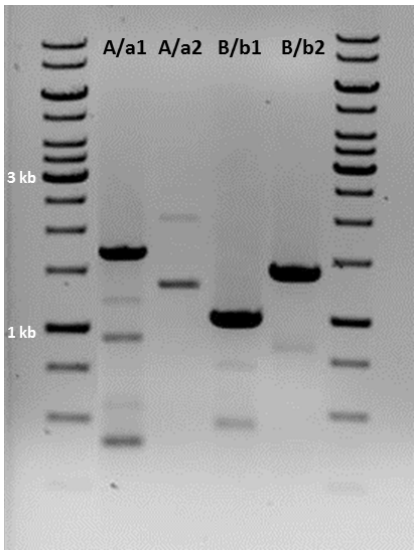
A.



B.

Primer	sequence
A	GAGACGTTCCTGCACTCCCGTGAG
a1	GTACAGAAGGCCGCCGAGTCCATG
a2	AGCACGTACATCGTGCATCTGCGG
B	GCGAGGAAGCAAGGCGATCCGATG
b1	CTGGAACTCGTCGAGCTCGCTCAC
b2	CCGGTAGGTGATGTTGCTCGCGTC

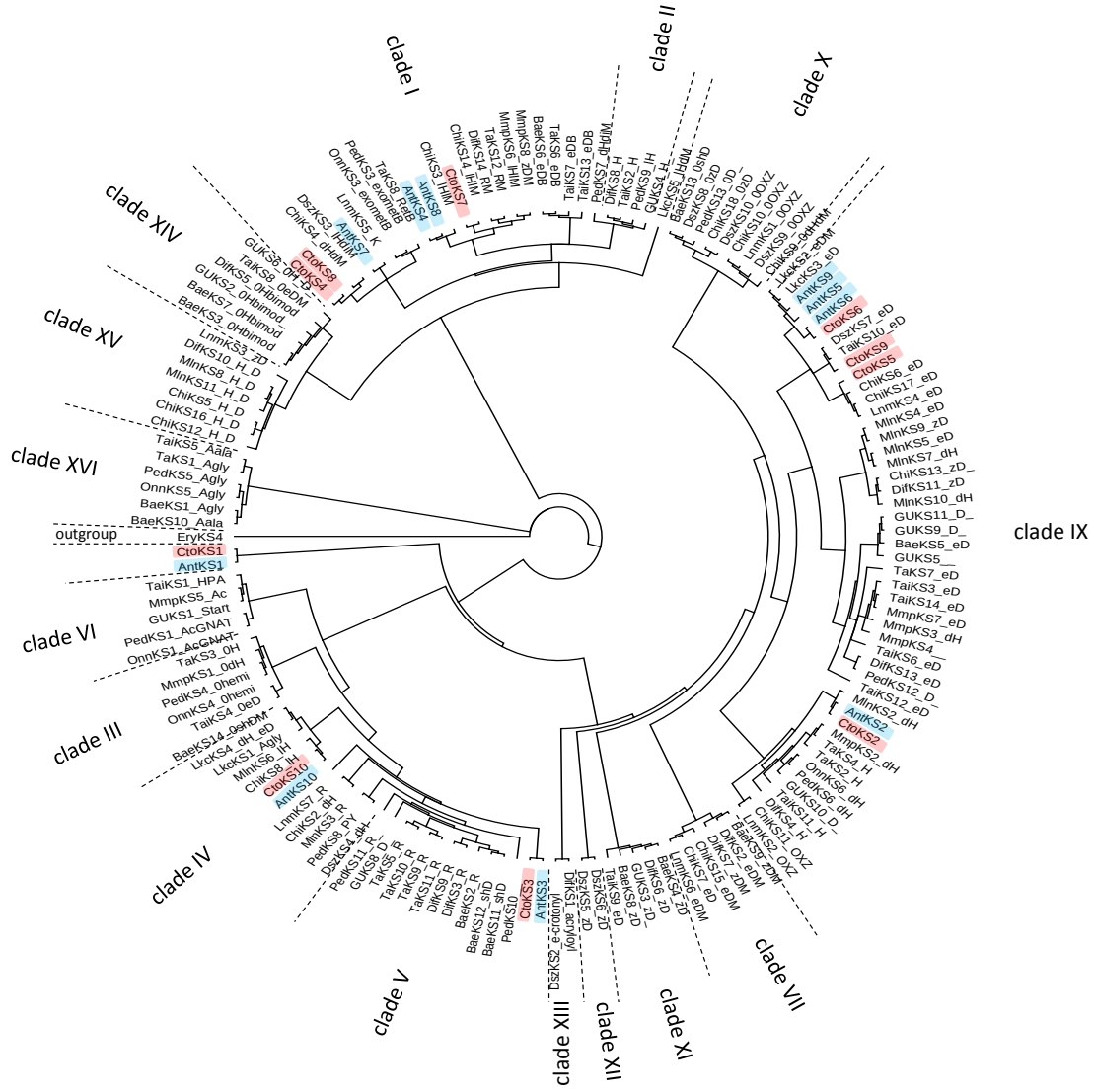
C.



Primer combination	Amplicon length
A / a1	1544 bp
A / a2	1277 bp
B / b1	928 bp
B / b2	1360 bp

Figure S15 Description of the PCR confirming the existence of two copies of the anthracimycin biosynthetic gene cluster. A.: schematic overview on primer design and placement within the cluster copies and their respective up- and downstream regions. B.: table of primers used for the experiment. C.: table with the expected amplicon size for the pecific primer pair; gel electrophoresis of the PCR reactions showing fragments of the expected size.

2.6.20 Phylogenetic analysis of KS domain substrate specificity



	Assignment to clades according to phylogenetic analysis									
	KS1	KS2	KS3	KS4	KS5	KS6	KS7	KS8	KS9	KS10
Anthracimycin	?	VIII	V	I	IX	IX	I	I	IX	IV
Chlorotonil	?	VIII	V	I	IX	IX	I	I	IX	IV

Figure S16 Bayesian cladogram of KS domains from *trans*-AT PKSs - including the KS domains of the anthracimycin and the chlorotonil PKS - adapted from Nguyen et al.[17]; alignment of the protein sequences was performed by ClustalW. Branches are labelled with the protein name of the respective KS domain whereas the number indicates the sequence of the KS domain in the respective pathway. The KS 4 of the erythromycin PKS (*cis*-AT PKS) serves as an outgroup. Abbreviations: Ant: anthracimycin, Bae: bacillaene, Chi: chivosazol, Cto: chlorotonil, Dif: difficidin, Dsz: disorazol, GU: uncharacterized PKS from *Geobacter uraniumreducens*, Lkc: lankacidin, Lnm: leinamycin, Mln: macrolactin, Mmp: mupirocin, Onn: onnamide, Ped: pederin, Ta: myxovirescin, Tai: thailandamide. Amino acid sequences of the 138 KS domains analyzed by Nguyen et al. were retrieved from GenBank.

2.6.21 Structure elucidation of chlorotonil B-C2

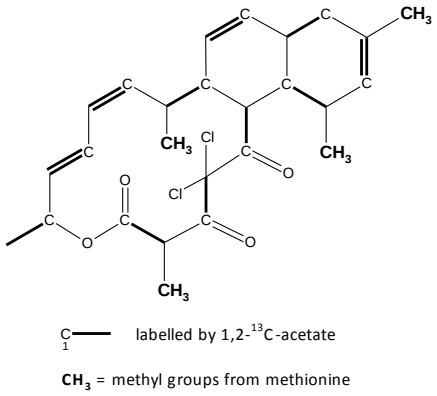


Figure S17.1 Incorporation of labelled precursors in chlorotonil A.

Table 1 ^{13}C NMR Data of Chlorotonil A in $CDCl_3$ after feeding labbeled 1-^{13}C- and 2-^{13}C-acetate and ^{13}C-methyl-methionine.

C	δ_C	m	1-^{13}C acetate	2-^{13}C acetate	^{13}C meth.
1	167.9	s	167.9		
2	47.0	d		47.1[a]	
3	192.0	s	192.0		
4	81.5	s		81.5[b]	
5	196.7	s	196.7		
6	49.6	d		49.7[a]	
7	36.7	d	36.8		
8	30.1	d		30.2	
9	128.0	d	128.0		
10	132.3	s		132.3	
11	38.3	t	38.3		
12	30.3	d		30.3	
13	133.1	d	133.1		
14	123.5	d		123.5	
15	42.7	d	42.7		
16	33.3	d		33.3	
17	139.3	d	139.3		
18	125.4	d		125.4	
19	123.9	d	123.9		
20	130.2	d		130.2	
21	70.3	d	70.3		
22	20.9	q		20.9	
23	17.0	q			17.0
24	14.8	q			14.8
25	23.2	q			23.2
26	15.6	q			15.7

[a] appeared as “triplett” signals, showing a doublet of about 13 Hz besides the singlet of not coupling carbons; [b] the signal of C-4 appeared small, due to two ^{2}J couplings with ^{13}C-2 and ^{13}C-6. In the 1,2-^{13}C-acetate labelled chlorotonil A all acetate units were observed as pairs.

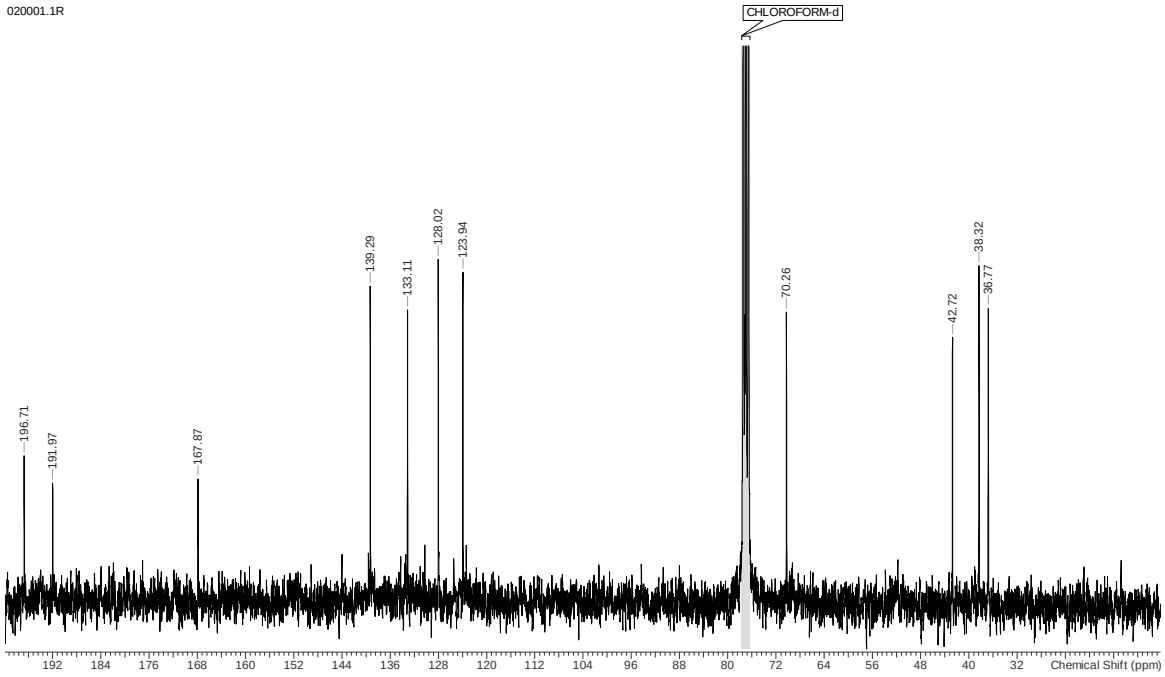


Figure S17.2 ^{13}C NMR spectrum of chlorotonil A after feeding labeled 1-^{13}C-acetate.

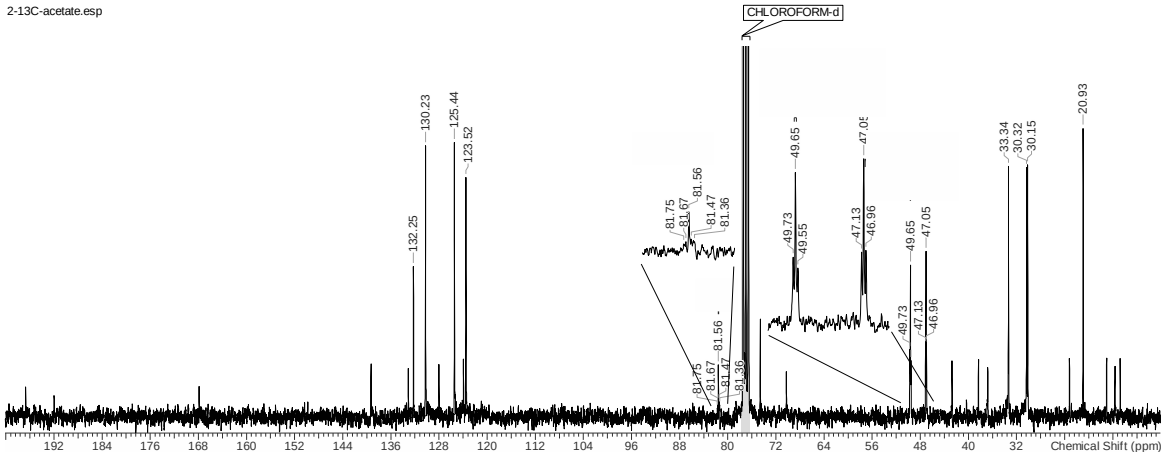


Figure S17.3 ^{13}C NMR spectrum of chlorotonil A after feeding labeled 2-^{13}C-acetate.

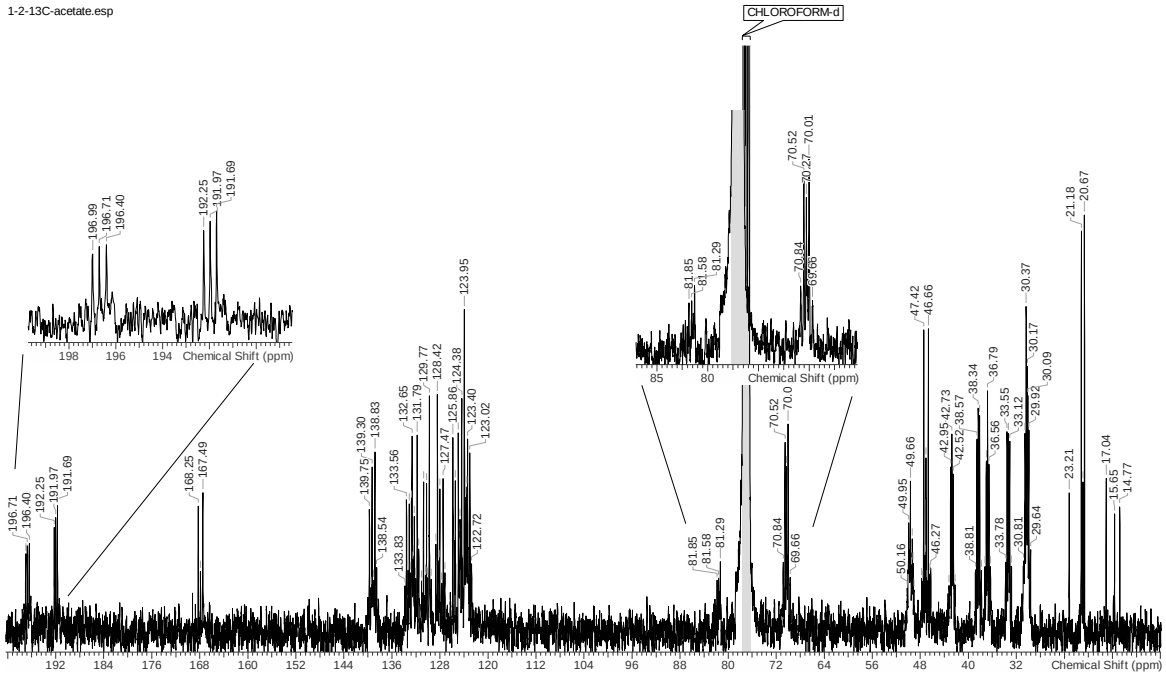


Figure S17.4 ^{13}C NMR spectrum of chlorotonil A after feeding labeled 1,2-^{13}C-acetate.

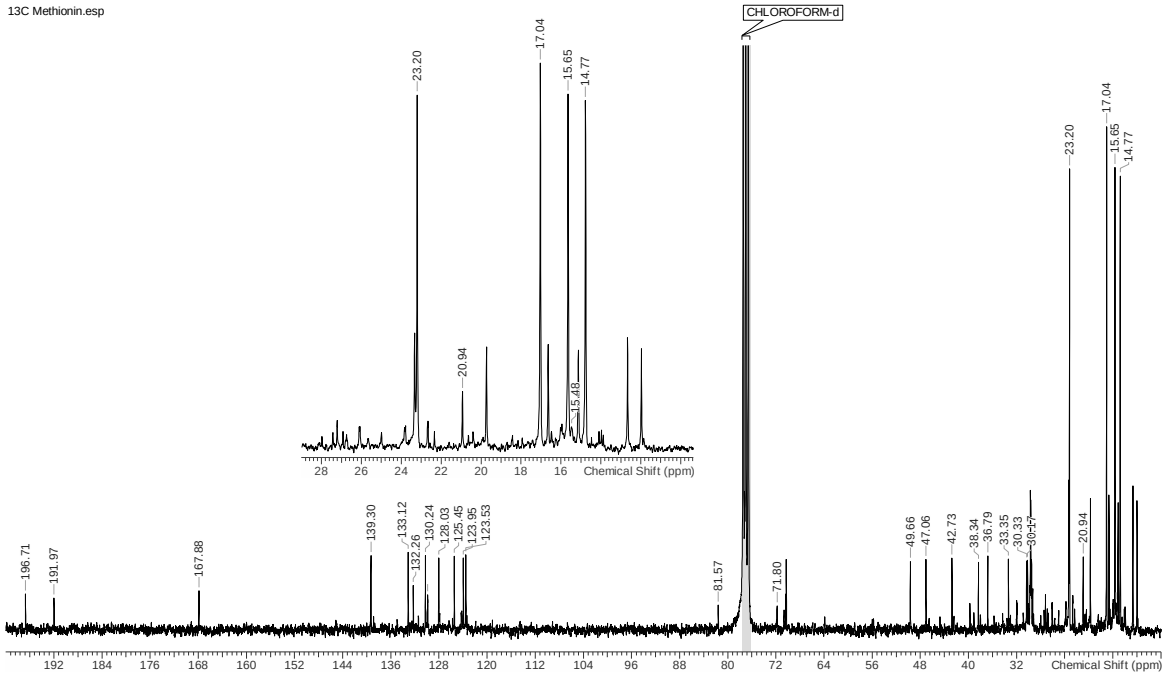


Figure S17.5 ^{13}C NMR spectrum of chlorotonil A after feeding labeled ^{13}C-methylmethionine.

2.6.21.1 General Experimental Procedures

Melting points were measured on a Büchi-510 melting point apparatus; UV data were recorded on a Shimadzu UV/Vis-2450 spectrophotometer in methanol (UVASOL, Merck); IR data were recorded on a Bruker Tensor 27 IR spectrophotometer. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker Ascend 700, Avance DMX 600 or DPX 300 NMR spectrometers, locked to the deuterium signal of the solvent. Data acquisition, processing and spectral analysis were performed with standard Bruker software and ACD/NMRSpectrus. Chemical shifts are given in parts per million (ppm) and coupling constants in Hertz (Hz). HRESIMS data were recorded on a Maxis ESI TOF mass spectrometer (Bruker Daltonics), molecular formulae were calculated including the isotopic pattern (Smart Formula algorithm)

2.6.21.2 Isolation of chlorotonil B isomers

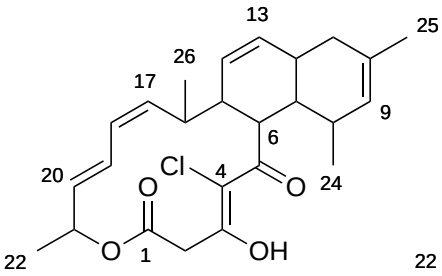
1.63 kg of Amberlite XAD 16 and cell mass were recovered from 70 L of fermentation broth of *Sorangium cellulosum*, strain So ce1525 by centrifugation. RP-HPLC analysis had indicated a production of 51 mg/L (3.57 g) of chlorotonil A at the end of the cultivation. The material was extracted in batches with 4 L of methanol, 3 portions of acetone (5 L each) and 3 portions of chloroform (5 L each). All batches were evaporated and the residues or remaining aqueous mixtures were partitioned between water and dichloromethane (DCM, CH2Cl2). After evaporation of the DCM, the first acetone extract provided 39.6 g, the combined second and third acetone extracts provided 56.4 g and the chloroform extract provided 2.5 g of an oily residue after evaporation of the CH2Cl2. Each was partitioned between 300 mL of methanol and 3 portions of heptane (300 mL). During these partitions an off-white crystalline precipitate was observed in the methanol layer and collected by filtration. The partition was repeated at about 10-fold concentration and further precipitate was collected and finally gave 2.4 g (0.77 g, 1.45 g, and 0.13 g) of chlorotonil A and 0.84 mg of a mixture of chlorotonils from the former chloroform extract.

The contents of the heptane layers of the first acetone extract and the chloroform extract were separated by RP-MPLC in three runs on an ODS-AQ (120 Å 16 µm) column (480×30 mm), using a gradient of 85 to 100 % acetonitrile in 3 hours after 40 min at 85%

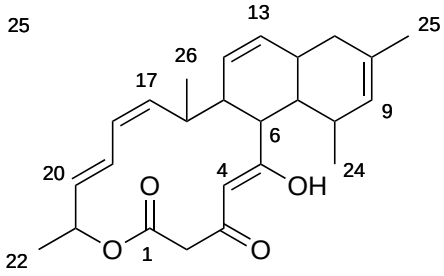
acetonitrile; flow 22 mL/min; detection UV at 226 nm; thus 0.69 g residue from the initial chloroform extract yielded 7 mg of chlorotonil A (R_t = 78 min) and 15 mg of chlorotonil B1 (R_t = 110 min) and 1.54 g residue from the first acetone extract yielded 25 mg of chlorotonil B1. For analytical purposes a portion was further purified by LH-20 chromatography in DCM/MeOH (1:1) and crystalized from $CHCl_3$/DCM/MeOH for x-ray analysis. Although the chromatography yielded pure chlorotonil B1 the NMR spectrum of the mother liquor showed a mixture of B1 and B2 possibly resulting from intermediate warming of the solution before a crystal was obtained.

The mixture of chlorotonils (840 mg) was extracted with DCM and the solution mainly contained chlorotonil A, ca. 50 mg. The residue was dissolved in $CHCl_3$ (150 ml) under warming at 40°C and concentrated to about 75 mL. After addition of petroleum ether (PE)

and standing at 6 °C a precipitate of highly enriched chlorotonil B2 (72 mg) was obtained. LH-20 chromatography in $CHCl_3$ or DCM/MeOH (1:1) and RP-chromatography of this sample caused isomerization to chlorotonil B1 while Si-flash chromatography with $CHCl_3$ yielded another isomer, chlorotonil B3 even after deactivation with trimethylamine.



Chlorotonil B1



Chlorotonil B2

Chlorotonil B3

The structure elucidation of the chlorotonil variants mainly relies on the elemental formulas which were determined by high-resolution ESI-MS (Bruker MaXis) and the 1D- and 2D-NMR data. The reasons for the assignment of chlorotonil B1 as the 3-keto derivative were the high chemical shift of C-3 (188.1 ppm) and of methyl group C-23 (1.45 ppm) and especially the relative intensity of the HMBC signals of the OH proton with slightly decreasing order with C-4, C-5, C-3 while the signals with C-6 and C-2 were very weak. The shift differences in the ^{1}H and ^{13}C NMR data between B2 and B3 are

small. The highest shift difference in the ^{1}H NMR spectra were observed for H-16 which is shifted from 2.38 ppm in B2 to 2.67 ppm in B3 while H-8 was shifted in the opposite direction from 2.65 ppm in B2 to 2.37 ppm in B3. Since chlorotonil B3 was obtained after silica gel chromatography the catalytic action of the silica gel might enable a steric rearrangement in the lactone to keto-enol part of the molecule, which possibly could lessen the ring strain of the lactone without producing conspicuous shifts in the NMR spectra. One argument supporting this rearrangement is the high number of weak ROESY signals observed for methyl group C-23 (i.e. correlations with 19-, 16-, 20-, 21-H, and very weak with 18-H) which were neither observed with chlorotonil B1 nor B2.

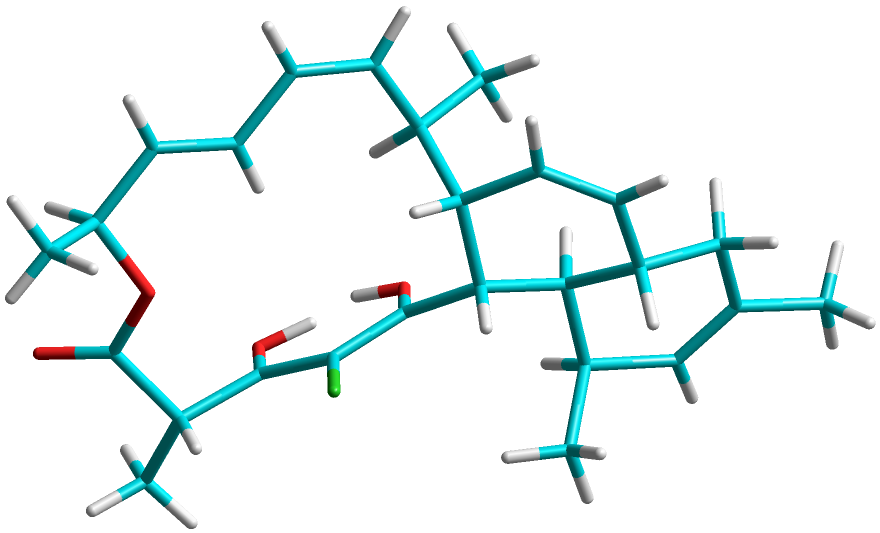


Figure S19.1 X-ray structure of Chlorotonil B1/B2

Chlorotonil B1: C26H33ClO4, M = 445.0; m.p. 185-186 °C; $[\alpha]^{22}_{D}$ = + 48.6 (*c* = 1.1 in $CHCl_3$); ^{1}H and ^{13}C NMR data table; UV (MeOH, ($CHCl_3$): λ_{max} (lg ε) = 237 (4.363), 310 (3.888) nm; HR-ESIMS: $[M+H]^+$ *m/z* 445.2138, calcd 445.2140 for $C_{26}H_{34}ClO_4$; $[M+Na]^+$ *m/z* 467.1957, calcd 467.1960 for $C_{26}H_{33}ClNaO_4$; $[2M+Na]^+$ *m/z* 911.4027, calcd 911.4027 for $C_{52}H_{66}Cl_2NaO_8$.

Table 2 NMR data of chlorotonil B1 in $CDCl_3$.[a]

Pos.	δ_C	type	δ_H	m (*J* [Hz])	COSY	N/ROESY	H HMBC
1	168.10	C					23 >2
2	44.72	CH	4.40	q (6.9)	23 >>OH	23	23, OH[d]
3	188.11	C					OH[d], 23, 2
4	109.12	C					OH[d]
5	193.69	C					6, OH[d]
6	46.55	CH	3.47	dd (12.1, 6.8)	7, 15 >>OH	24, 15	OH[d] >14
7	30.88	CH	2.18	m[b]	12, 6	8, 26	24, 11a, 13, 8, 12 >9[c]
8	30.83	CH	2.51	br s	24, 9	7, 9	24, 11a, 13, 8, 12 >9[c]
9	127.87	CH	5.37	m	25, 8	25, 8 >24	24, 25 >11a
10	132.53	C					25 >11a
11a	38.00	CH_2	2.04	br m (16.9, 3.7)	11b, 12	11b, 12, 13, 25	25 >9, 13
11b		CH_2	1.76	m	11a, 12	11a, 12 >13	
12	35.72	CH	2.18	m[b]	11ab, 11, 7	24, 13, 11ab	24, 11a, 6 >9, 13
13	133.11	CH	5.76	br m (10.1)	15, 14	14, 11a, 12	11a, 14
14	124.22	CH	5.55	ddd (10.0, 4.6, 2.0)	15, 13	13, 15, 17 >26	>13
15	42.41	CH	2.81	m (6.5, 4.3, 2.1)	6, 14 >13	6, 14, 19 >16	26 >6, 14, 13 >16
16	33.02	CH	2.61	br quin (7.2)	26, 17	19, 15, 17	26 >18 >17, 6
17	138.85	CH	5.41	br t (9.9)	16, 18	14, 18 >26	26 >16, 19
18	125.73	CH	5.89	t (10.9)	17, 19	17, 20	20 >19
19	123.34	CH	6.37	dddd (15.1, 11.3, 2.0, 1.0)	18, 20	16, 15, >21	17, 18
20	131.40	CH	5.56	br dd (15.2, 2.3)	19 >21	22, 18	22, 18
21	70.36	CH	5.48	br dt (6.6, 2.1)	22 >20	22	22, 20 >19
22	20.63	CH_3	1.37	d (6.6)	21	20, 21	-
23	11.96	CH_3	1.45	d (7.0)	2	2	2
24	15.12	CH_3	0.78	d (7.2)	8	8, 12, 6 >9	25
25	23.32	CH_3	1.68	s	9	9, 11a	9
26	16.63	CH_3	0.97	m	16	16, 7, 17 >14	17
OH		H	16.11	s	>>2, 6		

[a] (300/75 MHz); [b] H-7, H-12 overlap; [c] C-7, C-8 overlap; [d] relative intensity of the HMBC correlations of the OH proton: C-4, C-5, C-3 >6 >C-2.

Chlorotonil B2: C26H33ClO4, M = 445.0; $[\alpha]^{22}_{D}$ = + 720.6 (*c* = 0.068 in $CHCl_3$); ^{1}H and ^{13}C NMR data table; IR (KBr): ν ~ = 1747 (w), 1713 s, 1597 (s), 1411 (s), 1264 (s), 1183 (s), 769 (m) cm^{-1}; UV (MeOH, ($CHCl_3$): λ_{max} (lg ε) = 232 (4.184), 317 (3.796), 339 (sh) nm; HR-ESIMS: $[M+H]^+$ *m/z* 445.2139, calcd 445.2140 for $C_{26}H_{34}ClO_4$; $[M+Na]^+$ *m/z* 467.1953, calcd 467.1960 for $C_{26}H_{33}ClNaO_4$; $[2M+Na]^+$ *m/z* 911.4028, calcd 911.4027 for $C_{52}H_{66}Cl_2NaO_8$.

Table 3 NMR data of chlorotonil B2 in $CDCl_3$.

Pos.	δ_C	m	δ_H	H Mult (*J* [Hz])	COSY	N/ROESY	H HMBC
1	174.22	C					23, 2
2	45.78	CH	4.56	q (6.6)	23	23	23
3	179.63	C					23, 2
4	107.87	C					
5	194.01	C					6
6	48.16	CH	3.44	dd (11.7, 7.0)	7, 15	24, 7	
7	35.97	CH	1.98	m[b]	6	26, 6	24
8	30.75	CH	2.65	br m	24, 25, 11, 9	24, 11, 9	24
9	130.49	CH	5.46	m[c]	11, 25, 8	25, 8, 24	24, 25
10	130.21	C					25
11a	38.31	CH_2	1.97	m[b]	11, 12, 8	11, 25, 8, 13	25
11b			1.69	m	11, 12, 9	11, 12	25
12	31.14	CH	2.04	br m	11, 11, 14, 13	24, 11, 15, 13	
13	133.41	CH	5.68	br d (9.9)	15, 14, 12	11, 12, 14	
14	124.85	CH	5.48	m[c]	15, 13, 12	15, 17, 13	
15	41.92	CH	2.85	m	16, 6, 14, 13	12, 16, 20, 14, 19	26, 6
16	33.75	CH	2.38	m	26, 17, 15	26, 23, 15, 17, 19	26, 6
17	139.56	CH	5.33	br t (9.7)	16, 18	26, 14, 18, 16	26
18	125.78	CH	5.80	br t (10.6)	17, 19	17, 20	
19	123.32	CH	6.21	br ddd (14.9, 11.6, 1.1)	20, 21, 18	16, 15	
20	131.01	CH	5.47	m[c]	21, 19	22, 15, 21, 18	22, 18
21	69.30	CH	5.56	br q (6.9)	22, 20, 19	22, 20	22
22	20.65	CH_3	1.36	br d (7.0)	21	20, 21	
23	13.80	CH_3	0.91	d (7.0)	2	16, 2	2
24	15.53	CH_3	0.63	d (7.0)	8	12, 8, 6, 9	
25	23.89	CH_3	1.76	s	8, 9	11, 9	11
26	16.54	CH_3	0.77	d (6.6)	16	7, 16, 17	

[a] ^{1}H 600 MHz, ^{13}C 75.5 MHz; [b] overlapping signals H-7/H-11b; [c] overlapping signals H-9/H-14/H-20;

Chlorotonil B3: C26H33ClO4, M = 445.0; $[\alpha]^{22}_{D}$ = + 626.4 (*c* = 0.68 in MeOH); ^{1}H and ^{13}C NMR data table; IR (KBr): ν ~ = 1708 (s), 1598 (s), 1413 (s), 1184 (s), 770 (m) cm^{-1}; UV (MeOH, ($CHCl_3$): λ_{max} (lg ε) = 232 (4.199), 317 (3.860), 339 (sh) nm; (-)DCI-MS (isobutane): *m/z* (%): 444 (60), 408 (55), 400 (100), 363 (90); HR-ESIMS: $[M+H]^+$ *m/z* 445.2139, calcd 445.2140 for $C_{26}H_{34}ClO_4$; $[M+Na]^+$ *m/z* 467.1958, calcd 467.1960 for $C_{26}H_{33}ClNaO_4$; $[2M+Na]^+$ *m/z* 911.4028, calcd 911.4027 for $C_{52}H_{66}Cl_2NaO_8$;

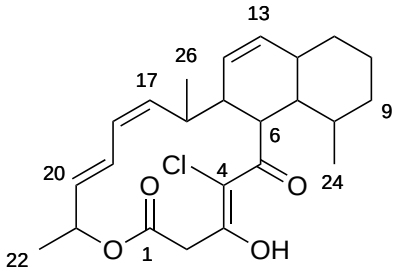
Table 4 NMR data of chlorotonil B3 in $CDCl_3$.[a]

H	δ_C	m	δ_H	m	*J* [Hz]	COSY	ROESY	HMBC
1	175.08	s	-.-			-		23 >> 2, 21
2	45.90	d	4.56	q	6.9	23	19 >21 >22, 26	23 >>
3	180.09	s	-.-			-	--	23 >> 2
4	107.74	s	-.-	-	-	-	--	2, 6
5	194.16	s	-.-			-	--	6 > 7 >> 14
6	48.32	d	3.36	dd	11.9, 7.2	15, 7	24, 7/11a >15 >14	7 > 14 > 16
7	36.34	d	1.98	m	[b]	6, 8	6, 8 >26 >24[b]	24 >> 6, 13, 9, 11a >11b, 12
8	30.71	d	2.37	m		24 > 9, 7, 25 > 11b	7/11a, 9 >12 (24)	24, 9, 7
9	130.25	d	5.39	m	[c]	25, 8	25, 8 >24	25, 24 >> 7/11a > 8, 11b[d]
10	130.30	q	-.-	-	-	--	--	[d]
11a	38.50	t	1.96	m	[b]	12, 11b	13, 25 (11b)[b]	25 >> 13, 9, 7
11b			1.67	dd	15.2, 13.4 br.	12, 11a, 7	12, >13, 7/11a	
12	30.96	d	2.05	m		11a/11b, 7 > 13, 14	24, 11b, 13 >6	6, 8, 7/11a, 14 >11
13	133.59	d	5.67	dd	9.7, 1.1	14 > 15	7/11a, 12 (14)	7/11a > 14, 12, 11b > 15 , 6
14	124.63	d	5.47	ddd	10.0, 4.3, 2.4	13 > 15 > 12	17, 15 (13) >>26	16, 15, 6 > 12
15	41.52	d	2.88	m		6, 14 > 13, 16, 12	16, 14,19, 6	26 >> 6, 13 > 14, 18, 7, 20 >17
16	33.81	d	2.67	m		26 > 17 > 18, 15	19, 15 >23 (26)	26 >> 6, 18, 17 > 13, 15
17	139.56	d	5.34	t	10.1	16, 18	18, 14, 26	26 >> 19, 16 > 18, 21
18	125.35	d	5.78	t	10.6 br.	19, 17	20 (17)	20, 19, 16 > 26
19	124.31	d	6.17	ddd	15.2, 11, 1.5	18, 20 >21	16 >15 >22	22, 18, 17, 21, 20
20	130.84	d	5.41	dd	15.2, 1.8[c]	19 > 21	21, 18, 22	22 >> 18, 21 > 19 > 17
21	69.46	d	5.62	qdd	7, 2, 1.5 br.	22 >> 20, 19	20 >>23 (22)	22 >> 19, 20 > 2
22	20.50	q	1.34	d	7.0	21	20 >2	21, 20 >>19
23	14.51	q	0.97	d	7.0	2	19, 16, 20, 21, 18	2
24	15.53	q	0.59	d	7.0	8	6, 12, 9, >7/11a, (8)	7, 25 > 8 >> 9, 11b
25	23.89	q	1.74	s	br.	9 > 8 > 11b	9, 7/11a	7/11a, 9
26	16.52	q	0.72	d	6.5	16	17, 7/11a, > 2, 14, 12, 8 (16)	17, 16, 15

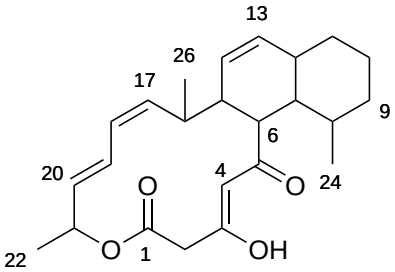
[a] ^{1}H 600 MHz; ^{13}C 100 MHz; [b] overlapping signals H-7/H-11; [c] overlapping signals H-9/H-20; [d] overlapping signals C-9/C-10;

2.6.21.3 Isolation of chlorotonil C

1.2 kg of XAD adsorber resin, which had been recovered from a 70 L fermentation of *Sorangium cellulosum* strain 265, were extracted with 6 L of methanol, 9 L of acetone, and 2.5 L of chloroform to give 9.9 g, 35.1 g and 830 mg of raw material after evaporation of the solvents, respectively. Each portion was partitioned between methanol and heptane to give 1.5 g, 8.14 g, and 290 mg of the heptane layers residues, respectively. The combined residues were partitioned between heptane (110 mL) and aqueous ethanol (110 ml ethanol and 10 mL water) to give a portion of 2.8 g with enriched chlorotonil C in the ethanol layer. This was separated by Sephadex LH-20 using dichloromethane/methanol 80:20 to give 153 mg of highly enriched product. Further product was found in the heptane layer. Repetition of the Sephadex LH-20 chromatography using dichloromethane/methanol 1:1 removed further byproducts. The raw chlorotonil C was separated in 4 portions by preparative RP HPLC [column 250×21 mm Nucleodur 100-10 C18 EC; solvent A: acetonitrile/water 1:1 with 0.1% formic acid, solvent B: acetonitrile with 0.1 % formic acid; isocratic conditions with 78 % solvent B; flow 20 mL/min; UV detection at 235 nm]. Each product fraction was evaporated and the remaining aqueous layer was extracted with dichloromethane to give about 112 mg pure chlorotonil C and a fraction containing 20 mg of the dechloro derivative chlorotonil C2.



Chlorotonil C1



Chlorotonil C2

Chlorotonil C1: $C_{24}H_{31}ClO_4$, M = 418.95; m.p. 160-161 °C; $[\alpha]^{22}_D$ = + 161.0 (*c* = 0.79 in ethanol); ^1H and ^{13}C NMR data table, the characteristic ^1H NMR signals of chlorotonil C1 could be recognized in the NMR spectrum of the raw extract.; IR (KBr): ν~ = 1745 (s), 1596 (m), 1246 (s), 1227 (m), 761 (w) cm^{-1}; UV (MeOH, ($CHCl_3$): λ_{max} (lg ε) = 236 (4.323), 315 (3.897) nm; HR-ESIMS: $[M+H]^+$ *m/z* 419.1985, calcd 419.1983 for $C_{24}H_{32}ClO_4$; $[M+Na]^+$ *m/z* 441.1803, calcd 441.1803 for $C_{24}H_{31}ClNaO_4$; $[2M+Na]^+$ *m/z* 859.3715, calcd 859.3714 for $C_{48}H_{62}Cl_2NaO_8$.

Table 5 NMR Data of Chlorotonil C1 in $CDCl_3$[a]

Pos.	δ_C	type	δ_H	Multiplicity (*J* [Hz])	COSY	NOESY	H in HMBC
1	164.74	C					2ab >>21
2a	43.71	CH_2	4.25	d (13.3)	2a	2b >19, 17/21	OH[e]
2b			3.27	d (13.3)	2b	2a	
3	182.74	C				-	2ab >OH[e]
4	109.67	C					OH[e] >2b >2a, 6
5	195.13	C					6, OH[e] >7
6	47.56	CH	3.47	dd (12.0, 7.0)	7, 15	12 >24, 15 >16, 8	7, 14, OH[e] >16
7	39.57	CH	1.92	ddd (11.9, 10.9, 3.6)	12, 6	8, 26 >9	13, 24 > 9, 11a, 12, 6
8	28.28	CH	2.21	br. qq (7.4, 3.6)	24	24, 9, 7 >>6	24 >10, 9, 7
24	12.33	CH3	0.79	d (7.4)	8	8, 6, 12 >9, 10	7 >9
9	33.19	CH_2	1.57	m		7, 8 >24, 11a, 12	24 >10, 11a[c]
10	20.92	CH_2	1.54	m		11a, 12	9, 11ab
11a	33.75	CH_2	1.79	br. dqd (12.5, 3.1, 1.0)	11b	11b, 10, 12, 13	10, 9, 7, 12, 13[d]
11b			1.08	tdd (12.8, 9.1, 6.8)	11a	11a >10 >7	
12	33.87	CH	2.02	br. ddquin (12.8, 10.5, 2.6)	7	13, 11a, 6 >10	14, 13 >10[d]
13	134.50	CH	5.65	dt (10.1, 1.6)	14	11a, 14 >12	12 >11a, 15
14	123.86	CH	5.47	ddd (10.0, 4.1, 2.8)	17, 13	15, 13	16 >12, 15
15	42.65	CH	2.75	m	6	6, 14, 19 >16	26, 13, 6 >16, 18, 7
16	33.15	CH	2.59	m	26, 17	26, 15, 19	6, 17, 18 >15[c]
26	16.47	CH_3	0.99	d (6.4)	16	16, 7 >17	17 >16 >15
17	138.96	CH	5.43[b]	br. t (10.0)	16, 18, 14	15, 18 >26	26, 16, 19
18	125.67	CH	5.90	t (10.9)	17, 17, 19	17, 20	16, 20 >19
19	123.22	CH	6.41	dddd (15.1, 11.5, 3.2, 1.0)	20, 18	21, 16, 15 **>** 2ab	17, 18, 21
20	131.04	CH	5.57	dd (15.3, 2.3)	19	18, 22	22, 18 >19
21	70.77	CH	5.44[b]	m	22, 18	22, 20, 19	22, 20, 19
22	20.55	CH_3	1.38	d (6.6)	17	20	21, 20
		OH	15.86	s		>9 >19	

[a] ^1H 700.4 MHz, ^{13}C 176.1 MHz; [b] 17/21-H overlap; [c] C-16/C-9 correlations overlap; [d] C-11/C-12 correlations overlap; [e] relative intensity of the HMBC correlations of the OH proton: C-4 > C-5, C-3, C-2 > C-6

Chlorotonil C2: $C_{24}H_{32}ClO_4$, M = 384.51; $[\alpha]^{22}_D$ = + 153.4 (*c* = 1.06 in $CHCl_3$); ^1H and ^{13}C NMR data table; IR (KBr): ν ~ = 1749 (s), 1609 (m), 1252 (s), 1230 (s), 763 (w) cm^{-1};-UV (MeOH, ($CHCl_3$): λ_{max} (lg ε) = 235 (4.381), 287 (3.861) nm; HR-ESIMS: $[M+H]^+$ *m/z* 385.2374, calcd 385.2373 for $C_{24}H_{33}O_4$; $[M+Na]^+$ *m/z* 407.2191, calcd 407.2193 for $C_{24}H_{32}NaO_4$; $[2M+Na]^+$ *m/z* 791.4491, calcd 791.4493 for $C_{48}H_{64}NaO_8$.

Table 6 NMR data of Chlorotonil C2 in $CDCl_3$[a]

Pos.	δ_C	m	δ_H	m (*J* [Hz])	ROESY	H in HMBC
1	165.46	C				2ab, 21 >4
2a	46.43	CH_2	3.50	d (11.8)	2b, 4	4
2b			3.24	d (11.5)	2a	
3	184.71	C				2ab, 4
4	102.79	CH	5.97	s	6, 2a >15/16, 19	6, 2ab
5	196.97	C				6, 4 >2b
6	50.49	CH	2.71	dd (12.3, 6.9)	4, 15/16 >23, 7, 12 >8	7, 14, 4, 16
7	39.30	CH	1.86	ddd (12.3, 11.0, 4.1)	24 > 6, 8 >11b	23, 11ab, 9, 12, 6, 13 >16
8	28.29	CH	2.31	quind (7.1, 4.0)	23 >9, 7	23 >10, 9, 7
9	33.35	CH_2	1.56	m	8, 23, 7 >11b	23 >11ab, 7
10	20.87	CH_2	1.52	m	23 >11ab, 12	9, 11ab
11a	33.83	CH_2	1.07	m	11b > 13, 12, 10	13 >10, 12 [b]
11b			1.77	m	11a >7, 10, 13	
12	33.77	CH	1.95	m	23 >11a >13, 10, 15/16 [d]	9, 7, 14, 11ab [b]
13	134.73	CH	5.63	br d (10.0)	14 >11a, 12	11a, 12, 15
14	124.00	CH	5.45	br dt (10.2, 3.2)	13 >15	12, 15
15	45.77	CH	2.61	m	24 >19, 14, 4, 12, 17[d]	24 >16, 6, 14, 13 >18
16	32.91	CH	2.61	m	24 >19, 14, 4, 12, 17[d]	24 >6, 17, 18 >15
17	139.08	CH	5.41	br t (9.3)	18> 24 >16/15[d]	24 >15, 19
18	125.82	CH	5.88	t (10.7)	17, 20/21[c]	15, 20, 21, 19
19	123.51	CH	6.47	ddt (14.4, 11.4, 1.4)	15/16 [d], 18, 4	17, 18, 21
20	131.25	CH	5.55	m[c]	18, 22[c]	22 >21, 18, 19
21	69.96	CH	5.57	m [c]	22[c]	22 >20, 19 >2b
22	20.70	CH_3	1.35	d (6.6)	20/21	21/20[c]
23	12.40	CH_3	0.79	d (7.4)	6, 8, 12 >9, 10	9, 7
24	16.20	CH_3	0.97	d (6.6)	7, 15/16 [d] >17	15, 17
OH		OH	15.44	br s		

[a] $^1H/^{13}C$ 700.4/176.1 MHz; [b] C-11/C-12 correlations overlap; [c] H-20/H-21 correlations overlap; [d] H-15 and H-16 overlap;

2.6.21.4 X-Ray analysis of chlorotonil B

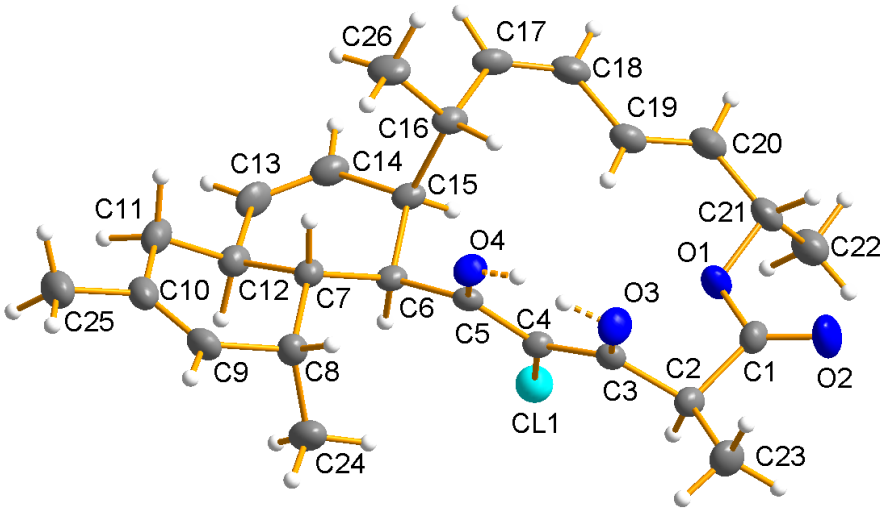


Figure S21 Result of the X ray analysis of chlorotonil B1 and B2 giving the atom numbering used.

2.6.23 X-Ray Structure Determination of Chlorotonil B

Crystals suitable for single-crystal x-ray analysis were obtained from $CHCl_3$/DCM/MeOH. The data were collected at 152 K on a BrukerAXS X8Apex CCD diffractometer operating with graphite-monochromatized Mo Kα radiation. Frames of 0.5° oscillation were exposed; deriving 5709 independent reflections (R_{int} = 0.04) in the θ range of 2 to 28° with a completeness of ~99%. Structure solution and full least-squares refinement with anisotropic thermal parameters of all non-hydrogen atoms and free refinement of the hydrogen were performed using SHELX.[42] The final refinement result in: R1= 0.036, wR2= 0.071. The absolute structure could be established by anomalous-dispersion effects (Flack: 0.03(3))

Crystallographic data for the structure have been deposited with the Cambridge Crystallographic Data Centre, CCDC, 12 Union Road, Cambridge CB21EZ, UK. Copies of the data can be obtained free of charge on quoting the depository number CCDC 1411222 www.ccdc.cam.ac.uk/data_request/cif

The keto-enol tautomerism in chlorotonil B shows up in the structure presentation by a 50 % presence of the enol protons at C-3 and C-5 simultaneously (dashed bonds) while the C-O and C-C distances are nearly the same, thus suggesting that about 50 % of the molecules have the OH group at C-3 and C-5, respectively.

Table 7: Crystal data and structure refinement for chlorotonil B1 sh3478 = So 2783 isomers

Identification code	sh3478
Empirical formula	C26 H33 Cl O4
Formula weight	444.97
Temperature	152(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 8.9278(7) Å; $\alpha = 90°$
	b = 10.7681(8) Å; $\beta = 90°$
	c = 24.8690(17) Å; $\gamma = 90°$
Volume	2390.8(3) $Å^3$
Z	4
Density (calculated)	1.236 Mg/m^3
Absorption coefficient	0.189 mm^{-1}
F(000)	952
Crystal size	0.674 x 0.546 x 0.028 mm^3
Theta range for data collection	1.638 to 27.898°.
Index ranges	-11<=h<=8, -13<=k<=14, -32<=l<=28
Reflections collected	20355
Independent reflections	5709 [R(int) = 0.0405]
Completeness to theta = 25.242°	100.0 %
Absorption correction	multi-scan
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5709 / 81 / 416
Goodness-of-fit on F^2	1.011
Final R indices [I>2sigma(I)]	R1 = 0.0357, wR2 = 0.0706
R indices (all data)	R1 = 0.0512, wR2 = 0.0762
Absolute structure parameter	0.03(3)
Extinction coefficient	n/a
Largest diff. peak and hole	0.192 and -0.167 $e.Å^{-3}$

Table 8: Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) For sh3478. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U (eq)
Cl(1)	4345(1)	9161(1)	7850(1)	29(1)
O(1)	3072(2)	8272(2)	9059(1)	31(1)
O(2)	4436(2)	7300(2)	9688(1)	48(1)
O(3)	4071(2)	5687(2)	8311(1)	31(1)
O(4)	2726(2)	5921(2)	7450(1)	25(1)
C(1)	4241(3)	7614(2)	9232(1)	27(1)
C(2)	5299(3)	7336(2)	8766(1)	25(1)
C(3)	4402(3)	6860(2)	8293(1)	22(1)
C(4)	3915(2)	7584(2)	7862(1)	20(1)
C(5)	3040(2)	7085(2)	7443(1)	20(1)
C(6)	2375(3)	7852(2)	6999(1)	21(1)
C(7)	2228(3)	7156(2)	6462(1)	23(1)
C(8)	3661(3)	6491(2)	6276(1)	28(1)
C(9)	3451(3)	6019(3)	5710(1)	32(1)
C(10)	2364(3)	6346(2)	5378(1)	32(1)
C(11)	1219(3)	7309(2)	5533(1)	32(1)
C(12)	1695(3)	8062(2)	6025(1)	28(1)
C(13)	502(3)	8925(2)	6222(1)	31(1)
C(14)	122(3)	9081(2)	6729(1)	29(1)
C(15)	861(3)	8416(2)	7190(1)	22(1)
C(16)	-252(3)	7520(2)	7488(1)	25(1)
C(17)	-1328(3)	8268(2)	7824(1)	29(1)
C(18)	-1074(3)	8657(2)	8324(1)	30(1)
C(19)	293(3)	8467(2)	8624(1)	30(1)
C(20)	504(3)	8782(2)	9135(1)	32(1)
C(21)	1917(3)	8634(2)	9445(1)	32(1)
C(22)	2393(4)	9827(3)	9723(1)	39(1)
C(23)	6530(3)	6442(3)	8931(1)	36(1)
C(24)	5090(3)	7272(3)	6324(1)	40(1)
C(25)	2200(4)	5794(4)	4824(1)	44(1)
C(26)	-1126(3)	6637(3)	7120(1)	35(1)

Table 9: Bond lengths [Å] and angles [°] for sh3478

Cl(1)-C(4)	1.741(2)
O(1)-C(1)	1.332(3)
O(1)-C(21)	1.461(3)
O(2)-C(1)	1.195(3)
O(3)-C(3)	1.298(3)
O(3)-H(0A)	0.89(8)
O(4)-C(5)	1.285(3)
O(4)-H(0B)	0.83(8)
C(1)-C(2)	1.526(3)
C(2)-C(3)	1.511(3)
C(2)-C(23)	1.518(4)
C(2)-H(1)	0.96(3)
C(3)-C(4)	1.396(3)
C(4)-C(5)	1.408(3)
C(5)-C(6)	1.502(3)
C(6)-C(7)	1.537(3)
C(6)-C(15)	1.556(3)
C(6)-H(2)	0.97(2)
C(7)-C(12)	1.535(3)
C(7)-C(8)	1.537(3)
C(7)-H(3)	1.01(3)
C(8)-C(9)	1.510(3)
C(8)-C(24)	1.532(4)
C(8)-H(4)	0.98(2)
C(9)-C(10)	1.322(4)
C(9)-H(5)	0.92(3)
C(10)-C(11)	1.507(4)
C(10)-C(25)	1.507(4)
C(11)-C(12)	1.528(3)
C(11)-H(6)	1.01(3)
C(11)-H(7)	1.01(3)
C(12)-C(13)	1.495(4)
C(12)-H(8)	1.02(3)
C(13)-C(14)	1.317(4)
C(13)-H(9)	0.90(3)
C(14)-C(15)	1.505(3)
C(14)-H(10)	1.00(3)
C(15)-C(16)	1.570(3)
C(15)-H(11)	0.96(2)
C(16)-C(17)	1.508(3)
C(16)-C(26)	1.532(4)
C(16)-H(12)	0.97(2)
C(17)-C(18)	1.330(4)
C(17)-H(13)	0.96(2)
C(18)-C(19)	1.446(4)
C(18)-H(14)	0.98(3)
C(19)-C(20)	1.329(4)
C(19)-H(15)	0.89(3)
C(20)-C(21)	1.487(4)
C(20)-H(16)	0.97(3)
C(21)-C(22)	1.521(4)
C(21)-H(17)	1.02(2)
C(22)-H(18)	1.02(3)
C(22)-H(19)	0.96(4)
C(22)-H(20)	1.00(3)
C(23)-H(21)	1.00(3)
C(23)-H(22)	1.02(3)
C(23)-H(23)	1.01(3)
C(24)-H(24)	0.99(3)
C(24)-H(25)	0.98(4)
C(24)-H(26)	0.96(3)
C(25)-H(27)	1.01(4)
C(25)-H(28)	0.98(3)
C(25)-H(29)	0.95(3)
C(26)-H(30)	1.02(3)
C(26)-H(31)	0.98(3)
C(26)-H(32)	0.97(3)

Table 9.2: Bond lengths [Å] and angles [°] for sh3478 (part II)

C(1)-O(1)-C(21)	118.90(18)
C(3)-O(3)-H(0A)	100(3)
C(5)-O(4)-H(0B)	112(5)
O(2)-C(1)-O(1)	124.8(2)
O(2)-C(1)-C(2)	125.1(2)
O(1)-C(1)-C(2)	110.13(18)
C(3)-C(2)-C(23)	112.3(2)
C(3)-C(2)-C(1)	109.24(19)
C(23)-C(2)-C(1)	111.5(2)
C(3)-C(2)-H(1)	106.3(15)
C(23)-C(2)-H(1)	111.4(16)
C(1)-C(2)-H(1)	105.8(15)
O(3)-C(3)-C(4)	119.9(2)
O(3)-C(3)-C(2)	115.14(19)
C(4)-C(3)-C(2)	124.95(19)
C(3)-C(4)-C(5)	121.9(2)
C(3)-C(4)-Cl(1)	119.30(17)
C(5)-C(4)-Cl(1)	118.80(17)
O(4)-C(5)-C(4)	118.9(2)
O(4)-C(5)-C(6)	117.4(2)
C(4)-C(5)-C(6)	123.61(19)
C(5)-C(6)-C(7)	113.90(19)
C(5)-C(6)-C(15)	109.42(18)
C(7)-C(6)-C(15)	112.45(19)
C(5)-C(6)-H(2)	108.2(13)
C(7)-C(6)-H(2)	108.3(13)
C(15)-C(6)-H(2)	104.0(14)
C(12)-C(7)-C(8)	109.99(19)
C(12)-C(7)-C(6)	109.33(18)
C(8)-C(7)-C(6)	114.60(19)
C(12)-C(7)-H(3)	104.7(15)
C(8)-C(7)-H(3)	108.3(15)
C(6)-C(7)-H(3)	109.4(14)
C(9)-C(8)-C(24)	111.2(2)
C(9)-C(8)-C(7)	109.5(2)
C(24)-C(8)-C(7)	114.5(2)
C(9)-C(8)-H(4)	107.0(13)
C(24)-C(8)-H(4)	107.5(14)
C(7)-C(8)-H(4)	106.8(13)
C(10)-C(9)-C(8)	125.8(3)
C(10)-C(9)-H(5)	116.0(16)
C(8)-C(9)-H(5)	118.2(16)
C(9)-C(10)-C(11)	121.4(2)
C(9)-C(10)-C(25)	122.5(3)
C(11)-C(10)-C(25)	116.1(3)
C(10)-C(11)-C(12)	112.5(2)
C(10)-C(11)-H(6)	108.5(17)
C(12)-C(11)-H(6)	110.4(16)
C(10)-C(11)-H(7)	108.1(14)
C(12)-C(11)-H(7)	111.2(14)
H(6)-C(11)-H(7)	106(2)
C(13)-C(12)-C(11)	113.2(2)
C(13)-C(12)-C(7)	112.6(2)
C(11)-C(12)-C(7)	108.37(19)
C(13)-C(12)-H(8)	110.5(13)
C(11)-C(12)-H(8)	105.6(13)
C(7)-C(12)-H(8)	106.2(13)
C(14)-C(13)-C(12)	125.2(2)
C(14)-C(13)-H(9)	120.7(18)
C(12)-C(13)-H(9)	114.0(18)
C(13)-C(14)-C(15)	123.8(2)
C(13)-C(14)-H(10)	119.2(14)
C(15)-C(14)-H(10)	116.9(14)
C(14)-C(15)-C(6)	109.47(19)
C(14)-C(15)-C(16)	111.95(19)
C(6)-C(15)-C(16)	116.99(17)
C(14)-C(15)-H(11)	109.2(13)
C(6)-C(15)-H(11)	106.5(13)
C(16)-C(15)-H(11)	102.1(13)
C(17)-C(16)-C(26)	109.7(2)
C(17)-C(16)-C(15)	109.68(18)
C(26)-C(16)-C(15)	115.1(2)
C(17)-C(16)-H(12)	106.6(13)
C(26)-C(16)-H(12)	107.3(13)
C(15)-C(16)-H(12)	108.2(14)
C(18)-C(17)-C(16)	125.3(2)

C(18)-C(17)-H(13)	118.6(14)
C(16)-C(17)-H(13)	116.0(15)
C(17)-C(18)-C(19)	125.6(2)
C(17)-C(18)-H(14)	119.9(14)
C(19)-C(18)-H(14)	114.5(14)
C(20)-C(19)-C(18)	125.2(3)
C(20)-C(19)-H(15)	117.4(17)
C(18)-C(19)-H(15)	117.4(17)
C(19)-C(20)-C(21)	126.0(2)
C(19)-C(20)-H(16)	118.8(16)
C(21)-C(20)-H(16)	115.1(16)
O(1)-C(21)-C(20)	106.7(2)
O(1)-C(21)-C(22)	109.1(2)
C(20)-C(21)-C(22)	112.5(2)
O(1)-C(21)-H(17)	108.7(14)
C(20)-C(21)-H(17)	111.3(14)
C(22)-C(21)-H(17)	108.5(13)
C(21)-C(22)-H(18)	108.7(17)
C(21)-C(22)-H(19)	110(2)
H(18)-C(22)-H(19)	107(2)
C(21)-C(22)-H(20)	110.6(17)
H(18)-C(22)-H(20)	110(2)
H(19)-C(22)-H(20)	110(3)
C(2)-C(23)-H(21)	108.3(17)
C(2)-C(23)-H(22)	112.2(18)
H(21)-C(23)-H(22)	108(3)
C(2)-C(23)-H(23)	109.5(17)
H(21)-C(23)-H(23)	112(3)
H(22)-C(23)-H(23)	107(2)
C(8)-C(24)-H(24)	110.9(17)
C(8)-C(24)-H(25)	109(2)
H(24)-C(24)-H(25)	108(3)
C(8)-C(24)-H(26)	113(2)
H(24)-C(24)-H(26)	107(2)
H(25)-C(24)-H(26)	109(3)
C(10)-C(25)-H(27)	114(2)
C(10)-C(25)-H(28)	111.9(19)
H(27)-C(25)-H(28)	103(3)
C(10)-C(25)-H(29)	110.4(17)
H(27)-C(25)-H(29)	107(3)
H(28)-C(25)-H(29)	110(3)

C(16)-C(26)-H(30)	112.4(16)
C(16)-C(26)-H(31)	111.1(17)
H(30)-C(26)-H(31)	103(2)
C(16)-C(26)-H(32)	108.8(16)
H(30)-C(26)-H(32)	110(2)
H(31)-C(26)-H(32)	111(2)

Table 10 Anisotropic displacement parameters ($Å^2 x\ 10^3$) for sh3478. The anisotropic displacement factor exponent takes the form: $-2p^2[\ h^2\ a^{*2}U^{11} + ... + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
Cl(1)	34(1)	18(1)	34(1)	4(1)	-10(1)	-7(1)
O(1)	30(1)	40(1)	25(1)	4(1)	5(1)	9(1)
O(2)	48(1)	69(1)	25(1)	12(1)	-1(1)	14(1)
O(3)	39(1)	20(1)	33(1)	5(1)	-5(1)	-1(1)
O(4)	29(1)	15(1)	32(1)	1(1)	-3(1)	-2(1)
C(1)	29(1)	29(1)	24(1)	3(1)	-4(1)	-2(1)
C(2)	24(1)	26(1)	26(1)	5(1)	-2(1)	1(1)
C(3)	19(1)	22(1)	24(1)	1(1)	2(1)	2(1)
C(4)	20(1)	15(1)	26(1)	1(1)	1(1)	0(1)
C(5)	16(1)	20(1)	25(1)	0(1)	4(1)	0(1)
C(6)	18(1)	18(1)	28(1)	2(1)	-1(1)	-2(1)
C(7)	22(1)	21(1)	25(1)	2(1)	-2(1)	-1(1)
C(8)	27(1)	31(1)	26(1)	1(1)	0(1)	3(1)
C(9)	30(1)	36(2)	30(1)	-2(1)	9(1)	1(1)
C(10)	33(1)	37(1)	25(1)	1(1)	2(1)	-9(1)
C(11)	34(2)	33(1)	28(1)	7(1)	-9(1)	-4(1)
C(12)	28(1)	25(1)	30(1)	5(1)	-3(1)	-2(1)
C(13)	32(1)	22(1)	40(1)	11(1)	-10(1)	3(1)
C(14)	24(1)	20(1)	43(2)	4(1)	-4(1)	2(1)
C(15)	20(1)	16(1)	32(1)	0(1)	-2(1)	0(1)
C(16)	20(1)	18(1)	36(1)	4(1)	0(1)	1(1)
C(17)	19(1)	22(1)	46(2)	7(1)	2(1)	2(1)
C(18)	23(1)	24(1)	41(2)	3(1)	9(1)	2(1)
C(19)	24(1)	29(1)	35(1)	4(1)	8(1)	2(1)
C(20)	29(1)	32(1)	35(1)	1(1)	11(1)	3(1)
C(21)	33(1)	35(1)	26(1)	3(1)	10(1)	1(1)
C(22)	39(2)	43(2)	36(2)	-4(1)	7(2)	-4(1)
C(23)	32(2)	42(2)	35(2)	7(1)	-4(1)	8(1)
C(24)	26(1)	56(2)	39(2)	-7(2)	2(1)	-1(2)
C(25)	44(2)	60(2)	30(1)	-6(2)	4(1)	-10(2)
C(26)	29(1)	28(1)	49(2)	-4(1)	3(1)	-7(1)

Table 11: Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for sh3478.

	x	y	z	U (eq)
H(0A)	3560(70)	5610(50)	8000(30)	28(12)
H(0B)	3000(80)	5590(60)	7730(30)	28(12)
H(1)	5700(30)	8130(20)	8657(10)	32(7)
H(2)	3010(30)	8570(20)	6942(8)	18(6)
H(3)	1400(30)	6520(20)	6490(10)	32(7)
H(4)	3780(30)	5760(20)	6508(9)	18(6)
H(5)	4120(30)	5450(20)	5580(10)	29(7)
H(6)	1050(30)	7870(30)	5216(11)	49(8)
H(7)	230(30)	6870(20)	5598(10)	31(7)
H(8)	2620(30)	8550(20)	5907(9)	26(6)
H(9)	40(30)	9350(30)	5960(11)	42(8)
H(10)	-760(30)	9620(30)	6818(10)	37(7)
H(11)	1100(20)	9010(20)	7468(9)	21(6)
H(12)	320(30)	7010(20)	7736(9)	21(6)
H(13)	-2290(30)	8420(20)	7669(9)	28(7)
H(14)	-1840(30)	9130(20)	8515(10)	31(7)
H(15)	1050(30)	8110(20)	8453(10)	35(8)
H(16)	-310(30)	9180(30)	9327(11)	44(8)
H(17)	1820(30)	7950(20)	9728(9)	24(7)
H(18)	1580(40)	10080(30)	9989(11)	44(8)
H(19)	3290(40)	9690(30)	9927(12)	54(10)
H(20)	2560(30)	10500(30)	9455(11)	47(8)
H(21)	7180(40)	6280(30)	8613(12)	52(9)
H(22)	6110(40)	5610(30)	9059(12)	55(9)
H(23)	7110(40)	6810(30)	9242(12)	53(9)
H(24)	5320(30)	7450(30)	6706(11)	41(8)
H(25)	5930(40)	6800(30)	6175(13)	68(11)
H(26)	5020(40)	8060(30)	6140(12)	52(9)
H(27)	2950(40)	5120(30)	4741(14)	68(11)
H(28)	2400(40)	6400(30)	4541(13)	60(10)
H(29)	1220(40)	5450(30)	4780(11)	46(9)
H(30)	-450(30)	6000(30)	6937(10)	45(8)
H(31)	-1580(30)	7090(30)	6819(11)	41(8)
H(32)	-1890(30)	6220(30)	7331(10)	41(8)

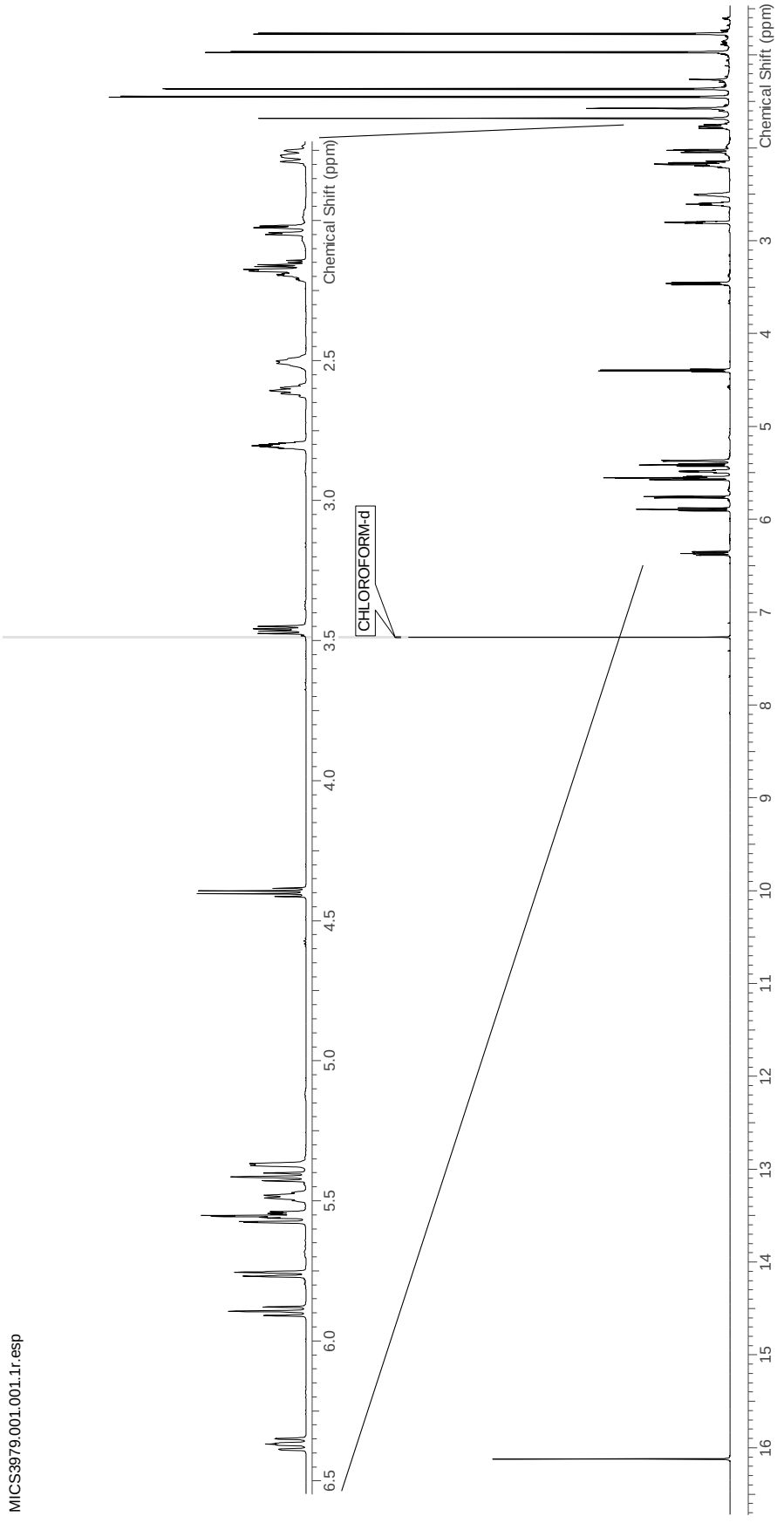


Figure S22.1 ^{1}H NMR spectrum of chlorotonil B1 in $CDCl_3$ at 700 MHz.

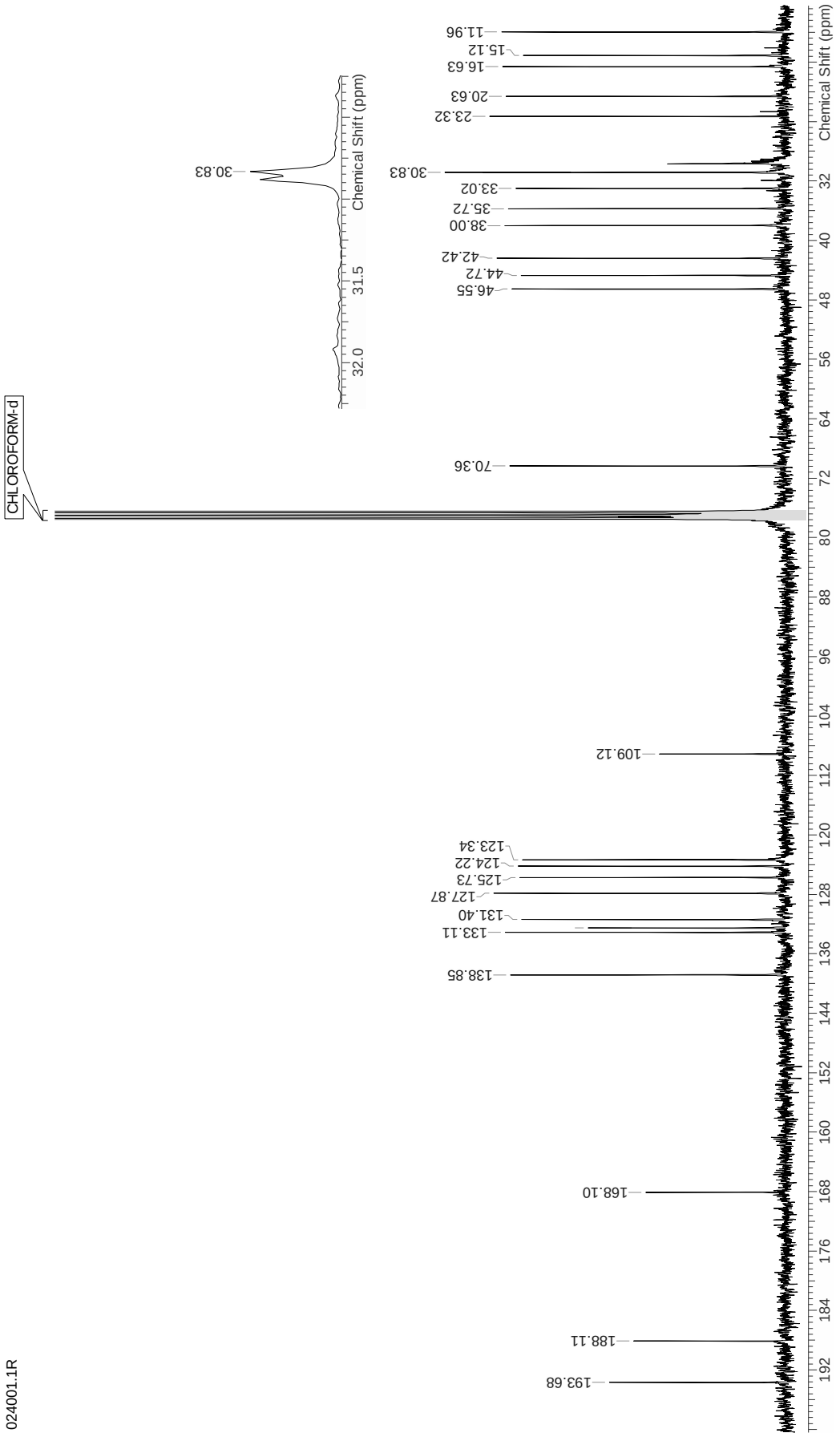


Figure S22.2 ^{13}C NMR spectrum of chlorotonil B1 in $CDCl_3$ at 75.5 MHz.

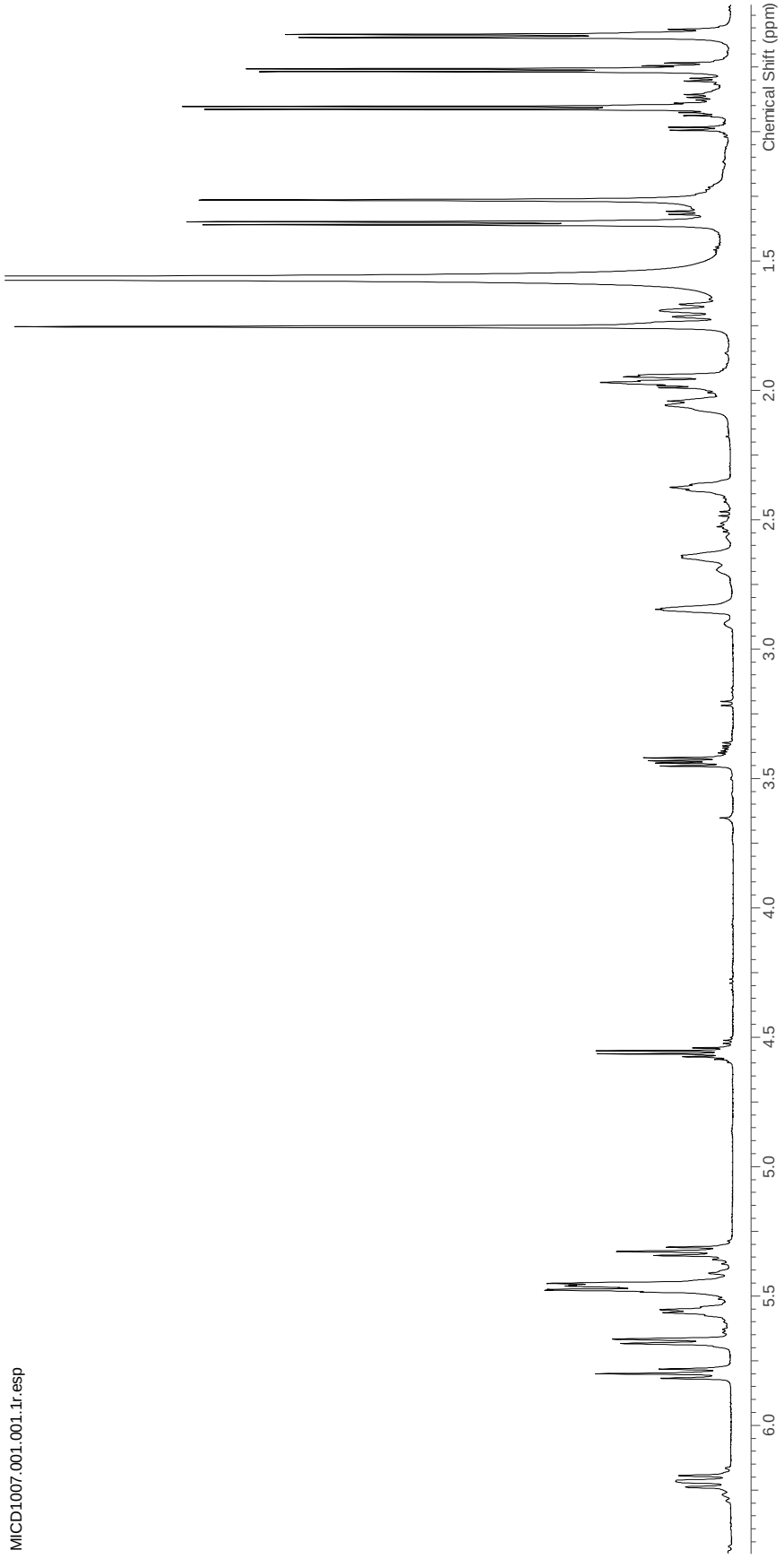


Figure S22.3 ^{1}H NMR spectrum of chlorotonil B2 in $CDCl_3$ at 600 MHz.

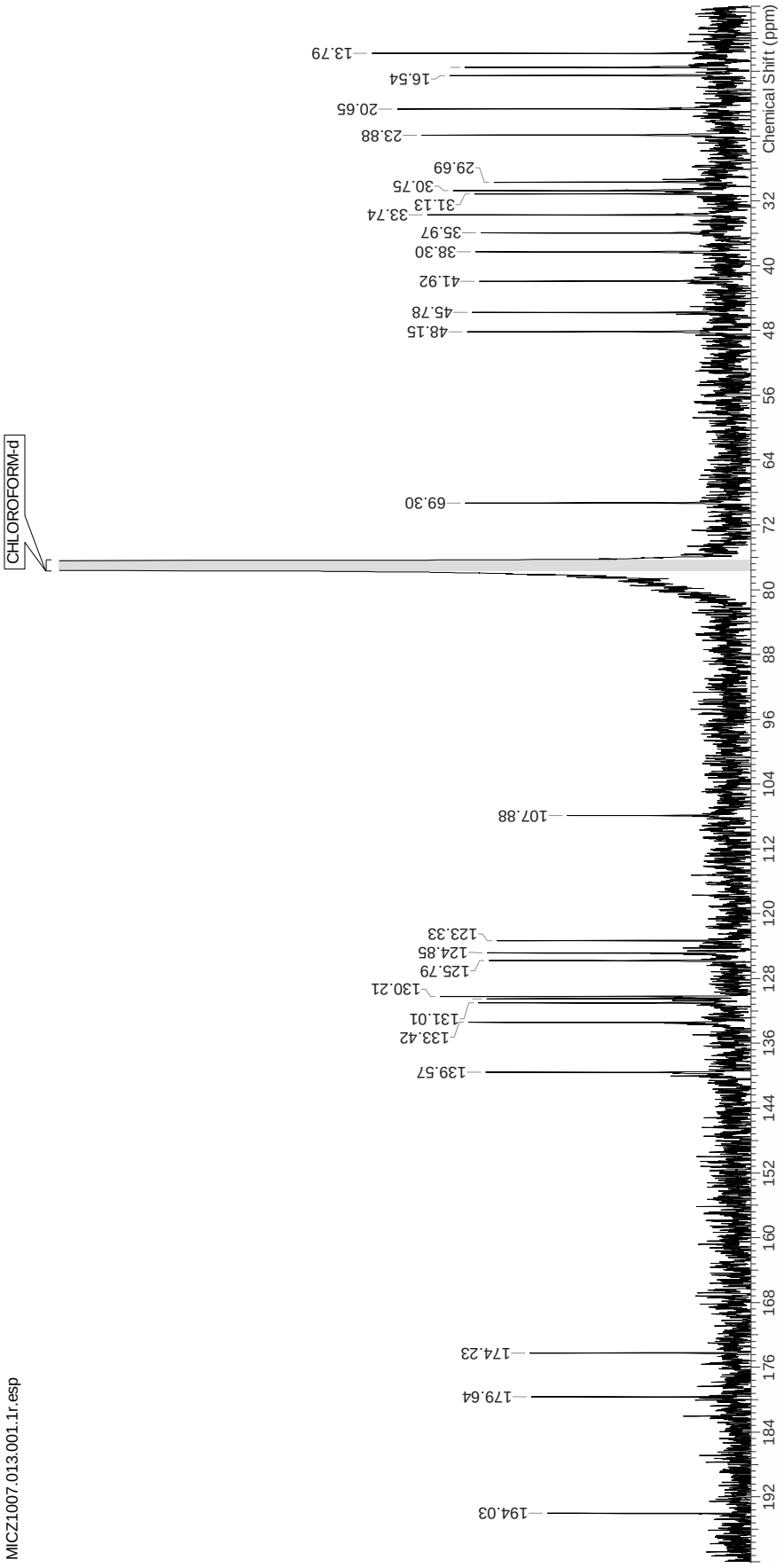


Figure S22.4 ^{13}C NMR spectrum of chlorotonil B2 in $CDCl_3$ at 75.5 MHz.

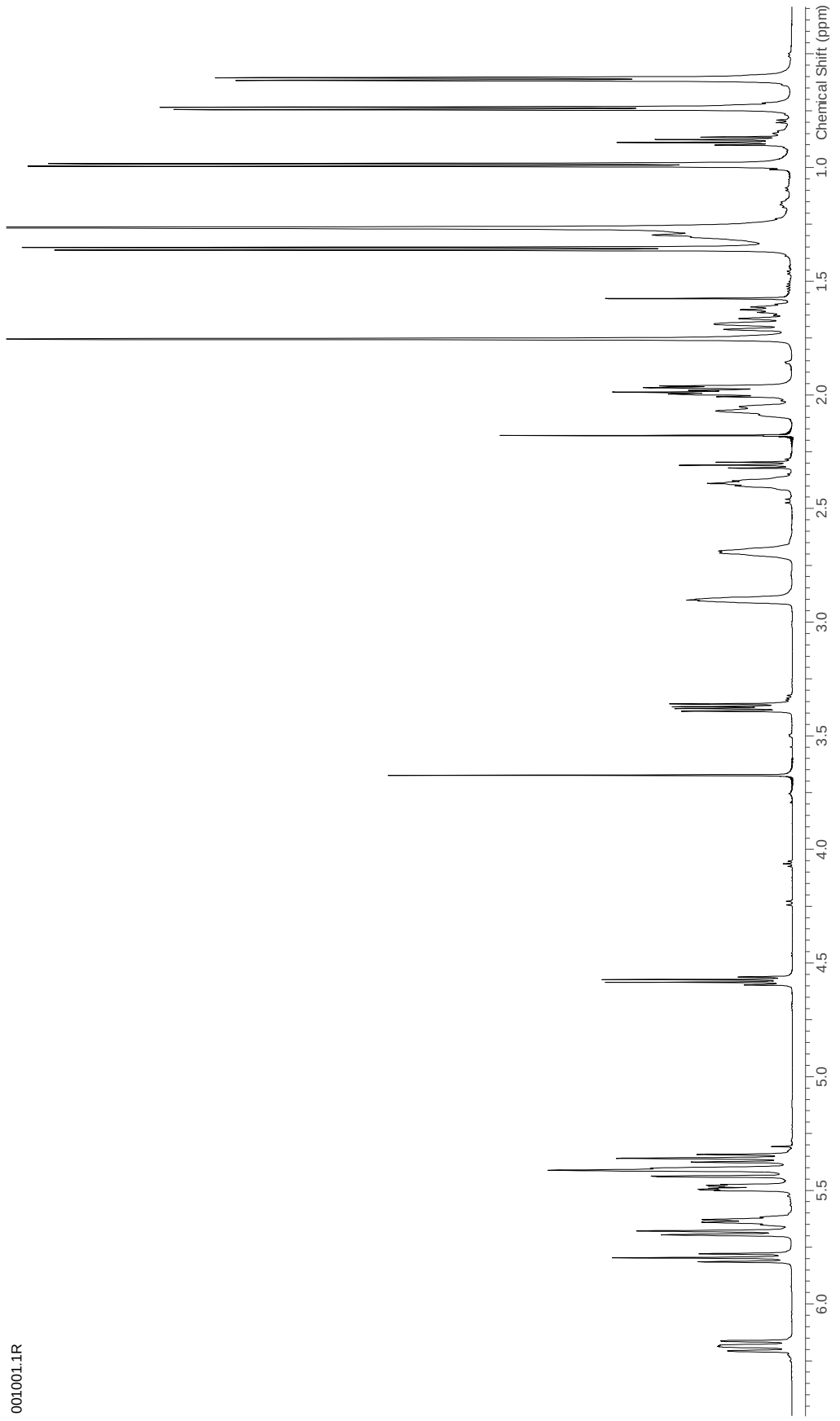


Figure S22.5 ^{1}H NMR spectrum of chlorotonil B3 in $CDCl_3$ at 600 MHz.

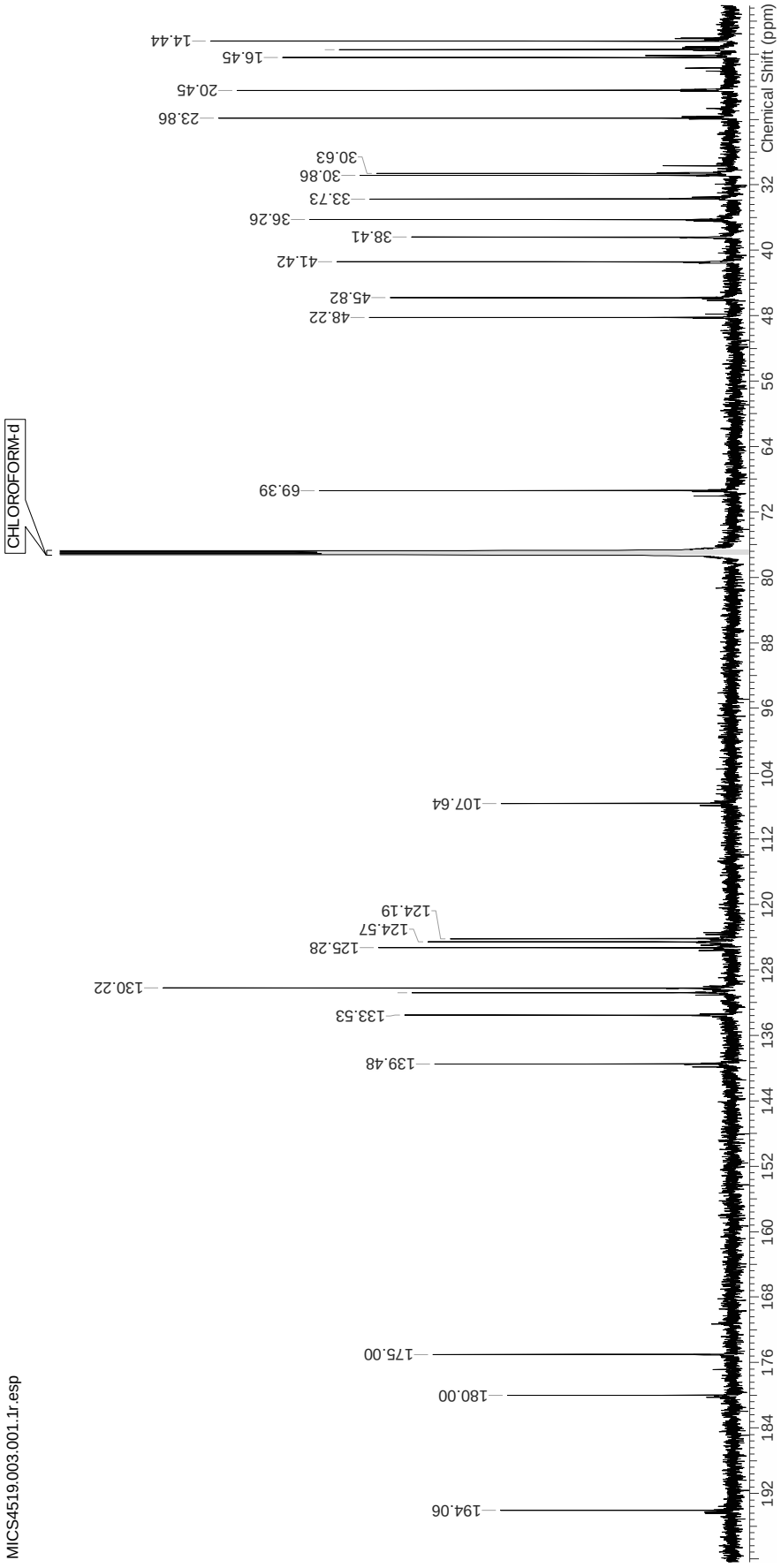


Figure S22.6 ^{13}C NMR spectrum of chlorotonil B3 in $CDCl_3$ at 100.6 MHz.

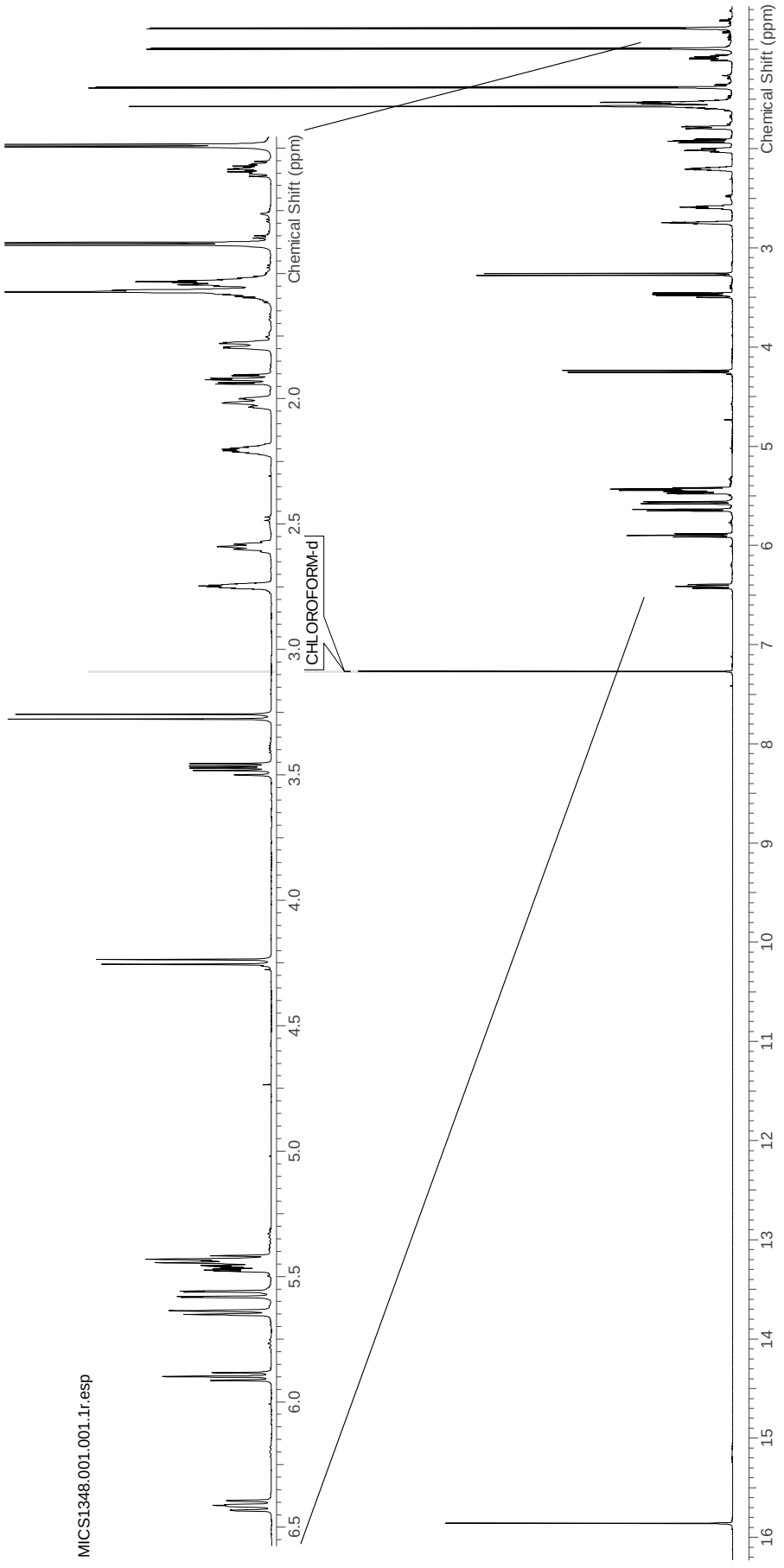


Figure S22.7 ^{1}H NMR spectrum of chlorotonil C1 in $CDCl_3$ at 700 MHz.

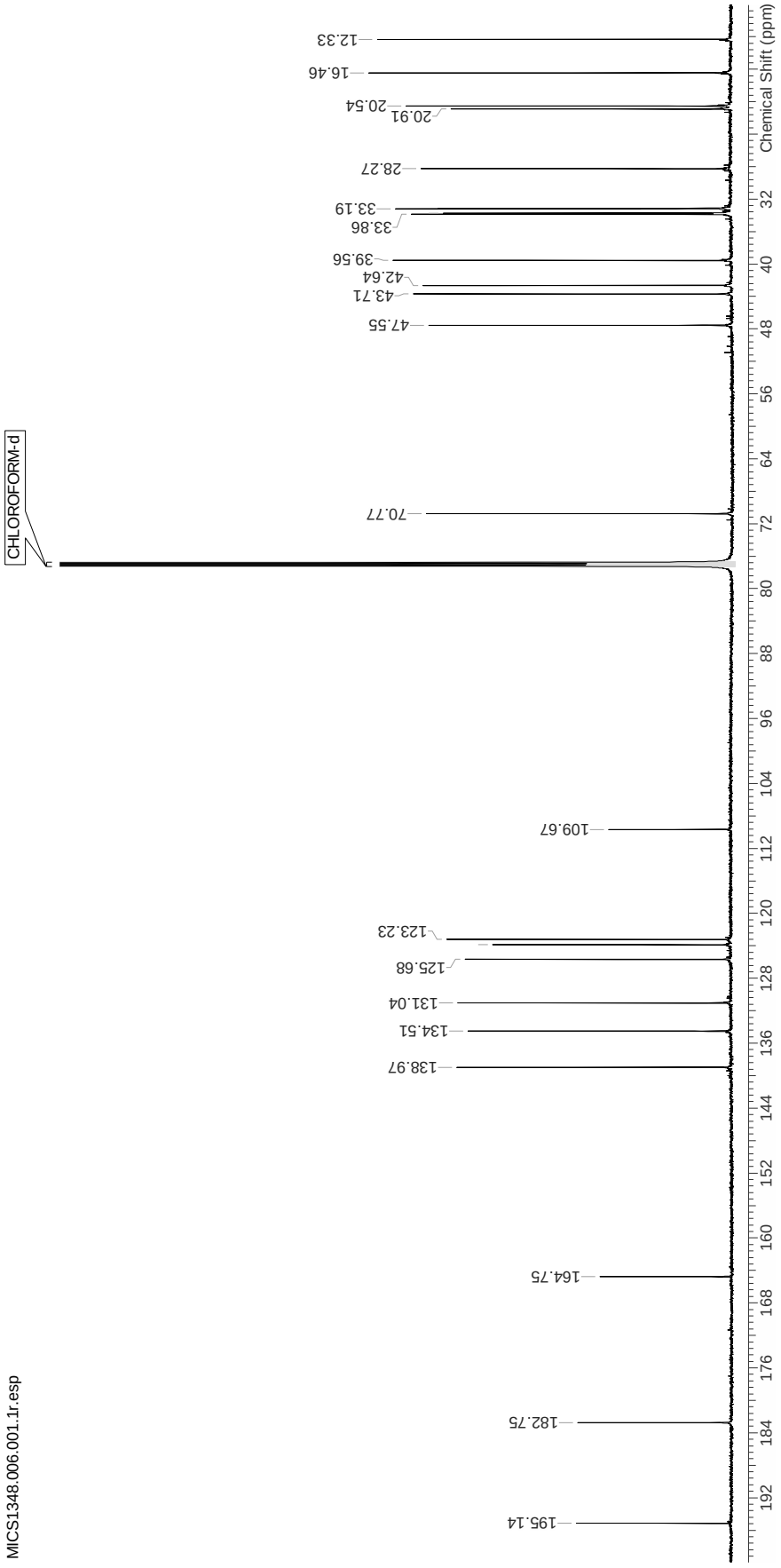


Figure S22.8 ^{13}C NMR spectrum of chlorotonil C1 in $CDCl_3$ at 176.12 MHz.

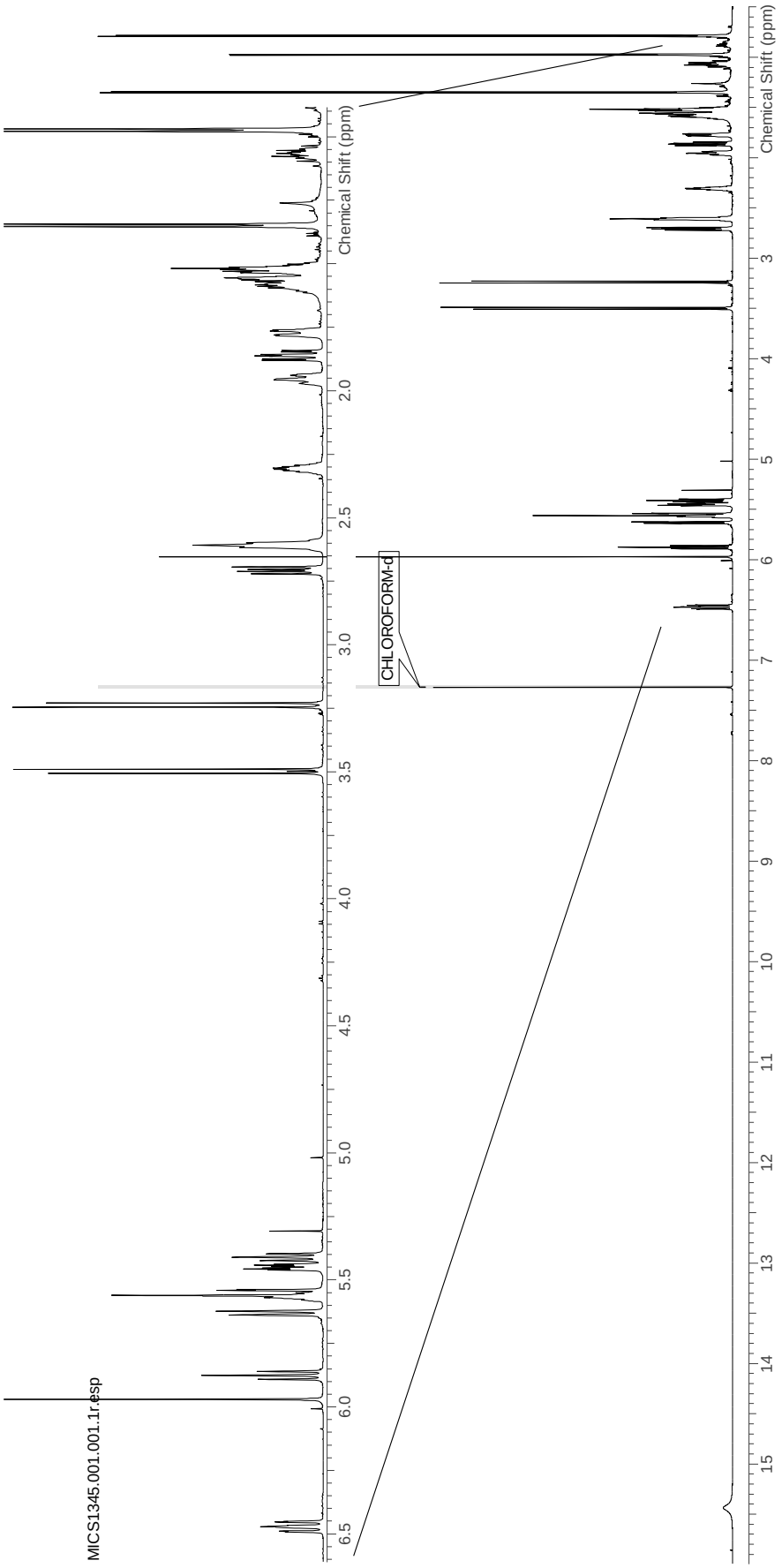


Figure S 22.9 ^{1}H NMR spectrum of chlorotonil C2 in $CDCl_3$ at 700 MHz.

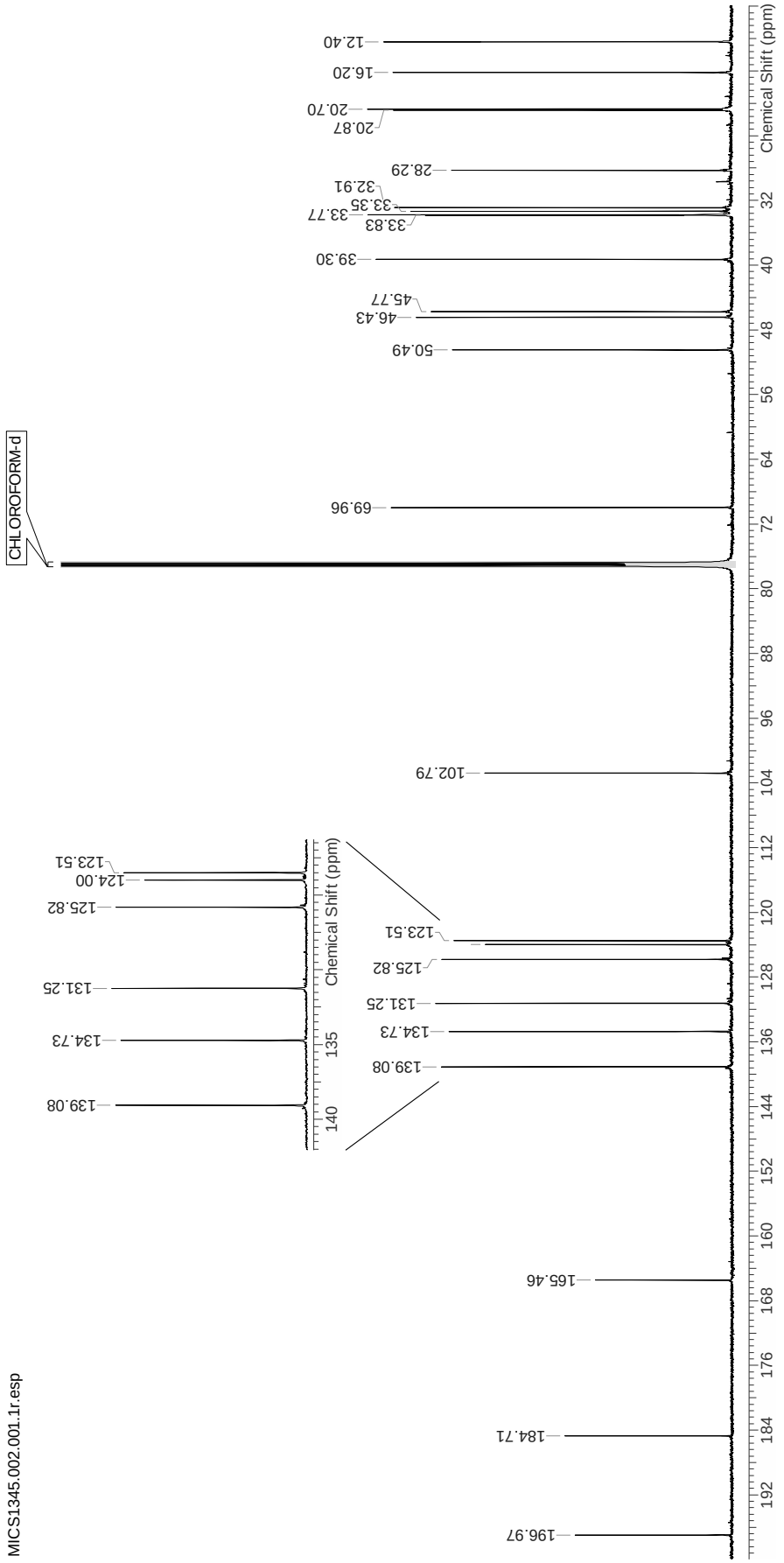


Figure S22.10 ^{13}C NMR spectrum of chlorotonil C2 in $CDCl_3$ at 176.12 MHz

2.7 References

[1] F. E. Koehn, G. T. Carter, *Nat. Rev. Drug Discov.* **2005**, *4*, 206–220.

[2] S. C. Wenzel, R. Müller, *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606.

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

[4] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, et al., *Nat. Biotechnol.* **2007**, *25*, 1281–1289.

[5] K. Gerth, H. Steinmetz, G. Höfle, R. Jansen, *Angew. Chemie Int. Ed.* **2008**, *47*, 600–602.

[6] K. H. Jang, S.-J. Nam, J. B. Locke, C. a Kauffman, D. S. Beatty, L. a Paul, W. Fenical, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7822–4.

[7] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996–1047.

[8] J. Held, T. Gebru, M. Kalesse, R. Jansen, K. Gerth, R. Müller, B. Mordmüller, *Antimicrob. Agents Chemother.* **2014**, *58*, 6378–84.

[9] E. M. Musiol, A. Greule, T. Härtner, A. Kulik, W. Wohlleben, T. Weber, *Chembiochem* **2013**, *14*, 1343–52.

[10] E. E. M. Musiol, T. Weber, *Medchemcomm* **2012**, *3*, 871.

[11] J. Piel, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14002–7.

[12] R. Gurney, C. M. Thomas, *Appl. Microbiol. Biotechnol.* **2011**, *90*, 11–21.

[13] O. Perlova, K. Gerth, O. Kaiser, A. Hans, R. Müller, *J. Biotechnol.* **2006**, *121*, 174–191.

[14] T. Weber, K. J. Laiple, E. K. Pross, A. Textor, S. Grond, K. Welzel, S. Pelzer, A. Vente, W. Wohlleben, *Chem. Biol.* **2008**, *15*, 175–188.

[15] S. B. Bumpus, N. a Magarvey, N. L. Kelleher, C. T. Walsh, C. T. Calderone, **2009**, *130*, 11614–11616.

[16] A. Witkowski, A. K. Joshi, Y. Lindqvist, S. Smith, *Biochemistry* **1999**, *38*, 11643–11650.

[17] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–33.

[18] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[19] P. Caffrey, *ChemBioChem* **2003**, *4*, 654–657.

[20] A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941–53.

[21] D. Pistorius, R. Müller, *Chembiochem* **2012**, *13*, 416–26.

[22] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chemie* **2010**, *49*, 1502–1506.

[23] J. Moldenhauer, D. C. G. Götz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Süssmuth, M. Engeser, H. Gross, G. Bringmann, J. Piel, *Angew. Chemie* **2010**, *122*, 1507–1509.

[24] C. T. Calderone, *Nat. Prod. Rep.* **2008**, *25*, 845–53.

[25] L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis, D. H. Sherman, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 2795–8.

[26] J. Crosby, M. P. Crump, *Nat. Prod. Rep.* **2012**, *29*, 1111–37.

[27] D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel, A. T. Keatinge-Clay, *Structure* **2014**, *22*, 444–451.

[28] K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van Den Heever, C. R. Hutchinson, J. C. Vederas, C. Tg, V. Uni, R. V August, et al., **2000**, *1*, 11519–11520.

[29] H. J. Kim, M. W. Ruszczycky, S. Choi, Y. Liu, H. Liu, *Nature* **2011**, *473*, 109–12.

[30] M. Serafimov, D. Gillingham, S. Kuster, D. Hilvert, **2008**, *1*, 7798–7799.

[31] K. Katayama, T. Kobayashi, M. Chijimatsu, A. Ichihara, H. Oikawa, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 604–7.

[32] T. Hashimoto, J. Hashimoto, K. Teruya, T. Hirano, K. Shin-Ya, H. Ikeda, H.-W. Liu, M. Nishiyama, T. Kuzuyama, *J. Am. Chem. Soc.* **2015**, 2–5.

[33] K. H. Jang, S.-J. Nam, J. B. Locke, C. a. Kauffman, D. S. Beatty, L. a. Paul, W. Fenical, *Angew. Chemie* **2013**, *125*, 7976–7978.

[34] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. a. Fischbach, R. Muller, W. Wohlleben, et al., *Nucleic Acids Res.* **2015**, 1–7.

[35] H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel, R. Müller, *Chembiochem* **2010**, *11*, 1840–9.

[36] M. R. Green, J. Sambrook, *Molecular Cloning : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, **n.d.**

[37] M. Kopp, H. Irschik, F. Gross, O. Perlova, A. Sandmann, K. Gerth, R. Müller, *J. Biotechnol.* **2004**, *107*, 29–40.

[38] D. A. H. Tobias Kieser, Mervyn J. Bibb, Mark J. Buttner, Keith F. Chater, *Practical Streptomyces Genetics*, Norwich, **2000**.

[39] J. Herrmann, Y. a. Elnakady, R. M. Wiedmann, A. Ullrich, M. Rohde, U. Kazmaier, A. M. Vollmar, R. Müller, *PLoS One* **2012**, *7*, 1–12.

[40] S. C. Findlow, C. Winsor, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **2003**, *42*, 8423–8433.

[41] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J L. Smith, *Structure* **2010**, *18*, 94–105.

[42] G. M. Sheldrick, *Acta Crystallogr. Sect. A Found. Crystallogr.* **2007**, *64*, 112–122.

Chapter 3

Towards heterologous expression of the chlorotonil biosynthetic pathway in *Myxococcus xanthus* DK1622

Contributions and Acknowledgements

A. Author's efforts

The author designed and performed the experiments described and evaluated and interpreted the resulting data. The author constructed a cosmid library of *Sorangium cellulosum* 1525 and screened it for cosmids harboring the chlorotonil biosynthetic gene cluster. After analysis of hit cosmids, the author developed a cloning strategy for assembly of the chlorotonil biosynthetic gene cluster with the help of Silke Wenzel. All constructs described were prepared by the author; likewise the author conducted all cloning and transformation steps involved in the project. The author cultivated and extracted all mutants described and analayzed the respective HPLC-MS data. Furthermore the author conceived and wrote the chapter.

B. Contributions by others

The author would like to thank Silke Wenzel for her help and assistance in planning multi-step cloning procedures and for her advice throughout the project. Scientific supervision of the project as well as proofreading of the chapter was done by Daniel Krug and Rolf Müller. The author would also like to thank Thomas Hoffmann and Michael Hoffmann for HPLC-MS measurements and Katja Gemperlein and Anja Schwarz for advice concerning cosmid libraray screening.

3 Towards heterologous expression of the chlorotonil biosynthetic pathway in *Myxococcus xanthus* DK1622

3.1 Abstract

Chlorotonil A is a novel polyketide that features an unusual gem-dichloro-1.3-dione substructure. The compound which was isolated from the myxobacterial producer strain *Sorangium cellulosum* So ce1525 shows outstanding bioactivity against the malaria parasite *Plasmodium falciparum* in a nano-molar range. In addition, good anti-bacterial and a moderate anti-fungal activity could be observed for chlorotonil A while at the same time the cytotoxicity level is relatively low. The chlorotonil A biosynthetic gene cluster which was identified previously contains 21 genes, three of which encode a ten-modular *trans*-AT polyketide synthase, whilst the remaining genes encode transporter and regulator functionalities as well as tailoring enzymes involved in chlorotonil biosynthesis. The gene cluster spanning 80.7 kbp was cloned from individual cosmids and two separate constructs for stepwise integration into the host genome of *Myxococcus xanthus* DK1622 were prepared. *M. xanthus* is more amenable to mutagenesis than the original *Sorangium* producer strain and at the same time, has significantly faster doubling rate than So ce1525. Despite the good chlorotonil A production level of 77 mg/L observed in the *Sorangium* producer strain, heterologous expression of the chlorotonil biosynthetic gene cluster is expected to facilitate alterations of gene cluster aiming for production of derivatives. Furthermore this platform should enable profound investigations of the trans-AT PKS responsible for chlorotonil biosynthesis.

3.2 Introduction

With respect to growing resistance rates against known therapeutics, in particular also in the case of tropical diseases such as malaria, there is an increasing need for new anti-infective drugs with new modes of action. Chlorotonil has shown convincing activity against *Plasmodium falciparum* strains *in vitro* in and *in vivo* and exhibits low toxicity compared to the standard treatment chloroquin.[1] One major drawback of the compound is its high lipophilicity which makes it practically insoluble in all sorts of polar solvents. One option to address this problem is the synthesis of less lipophilic derivatives. But as frequently seen for natural products, also in the case of chlorotonil A, chemical synthesis

was established but the complexity of the route to the stereochemically challenging compound renders this approach rather inconvenient. As reported by Rahn and Kalesse, total synthesis of chlorotonil A comprised 21 linear steps with an overall yield of 1.5 %; for large scale production of less lipophilic analogues a more efficient process would be favoured.[2] The establishment of a platform for heterologous expression of chlorotonil A in a host with good growth properties, which at the same time is amenable to mutagenesis, became the focal point of our research interest. The biosynthetic pathway for chlorotonil A, which was reported previously, consists of a ten-modular polyketide synthase of *trans*-AT type and several additional functionalities encoded up- and downstream of the respective locus. Polyketide synthases are large multifunctional megaenzyme complexes that catalyse the stepwise assembly of a polyketide from simple precursor molecules such as malonyl-CoA or methylmalonyl-CoA. The most basic representative of this class of biosynthetic machinery is a type I PKS following the so-called colinearity rule.[3,4] This means that the number and order of domains grouped into modules reflects the sequence and the structural variations seen in the final polyketide product. A minimal module in a type I PKS pathway consequently consists of an AT (acyltransferase) domain selecting the activated extender unit, an ACP (acyl-carrier protein) domain whose ppant arm is loaded with the extender unit and a KS (ketosynthase) domain that links the extender unit to the growing polyketide chain. Optional reducing domains (ketoreductase [KR], dehydratase [DH], enoylreductase [ER]) are involved in β-processing of the accruing polyketide.[5,6] In derogation thereof, *trans*-AT PKS systems are lacking integral AT domains in their modules; thus the selection of extender units takes places via free-standing AT functionalities which are in most cases encoded in close proximity to the PKS gene cluster.[7] This could also be observed for the chlorotonil PKS which is lacking cognate ATs in its modules but is equipped with a free standing tandem AT domain fused to an ER domain encoded by *ctoB*. Two more striking peculiarities were reported for the chlorotonil biosynthetic pathway: an FAD-dependent halogenase responsible for di-chlorination at the C4 position of chlorotonil and an SAM-dependent methyltransferase attaching a methyl-group at the C2 position of chlorotonil. In-depth analysis revealed the interdependence of these two processes which also has been previously described. Here we report the cloning of the 80.7 kbp biosynthetic gene cluster from cosmids and the construction of two constructs harbouring the 5’ and the 3 ‘ part the megaenzyme complex including several genes encoded up- and downstream of the core PKS part. For cloning and assembly of the chlorotonil biosynthetic gene cluster,

the advanced RedET recombineering technique was applied.[8] The technique allows for rapid and targeted alterations of plasmids in suitable *E. coli* host strains adapted for this purpose. DNA sequences can be introduced to or removed from plasmid constructs with the help of homologous regions while selection of the resulting construct is achieved by introduction of antibiotic resistance cassettes in the same step. Comprehensive and systematic planning of the desired cloning steps performed by RedET recombineering enables straightforward assembly of larger biosynthetic pathways as described in the following. Furthermore, endeavours towards stepwise integration of these constructs into the host genome of *Myxococcus xanthus* DK1622 are reported[9].

3.3 Results and discussion

3.3.1 Identification of cosmids harbouring the chlorotonil biosynthetic gene cluster

In order to obtain DNA fragments harbouring the chlorotonil biosynthetic gene cluster, two cosmid libraries from genomic DNA of *Sorangium cellulosum* So ce1525 were constructed and spotted on membranes. Three labelled probes - one targeting the halogenase-encoding gene *ctoA*, one targeting module five encoded in *ctoD* and one targeting the 3’ end of *ctoE* encoding the TE domain - were amplified via PCR. Hybridization of the membranes resulted in 10 putative hit cosmids which were end-sequenced at their 3’ and 5’ ends in order to evaluate the coverage of the chlorotonil biosynthetic gene cluster through the respective cosmid. Four candidate cosmids could thus be identified; three of them, - W1, G5 and G4 – were sequenced using the Illumina technique. The 32 kbp cosmid W1 covers the first PKS module, as well as the upstream region of the chlorotonil biosynthetic gene cluster plus an additional 9 kbp sequence upstream the presumed 5’ end of the gene cluster. The 36 kbp cosmid G5 contains the rest of the PKS core genes *ctoCDE* as well as two transporter genes encoded downstream the PKS genes. Cosmid G4 which spans 30 kbp covers the modules 7-9 and another 8 kbp of sequence downstream the PKS core genes. Notably a 1 kb fragment between cosmid W1 and G5, which encodes the ACP domain of module three is not covered by the three cosmids, but the fourth candidate – cosmid G3 – which mainly covers the same genomic region as cosmid W1 also contains this sequence part and was consequently used for cloning of the biosynthetic pathway. The strategy for stitching the chlorotonil gene cluster involved five RedET recombineering steps in which the constructs were modified according to the needs as described below.[8]

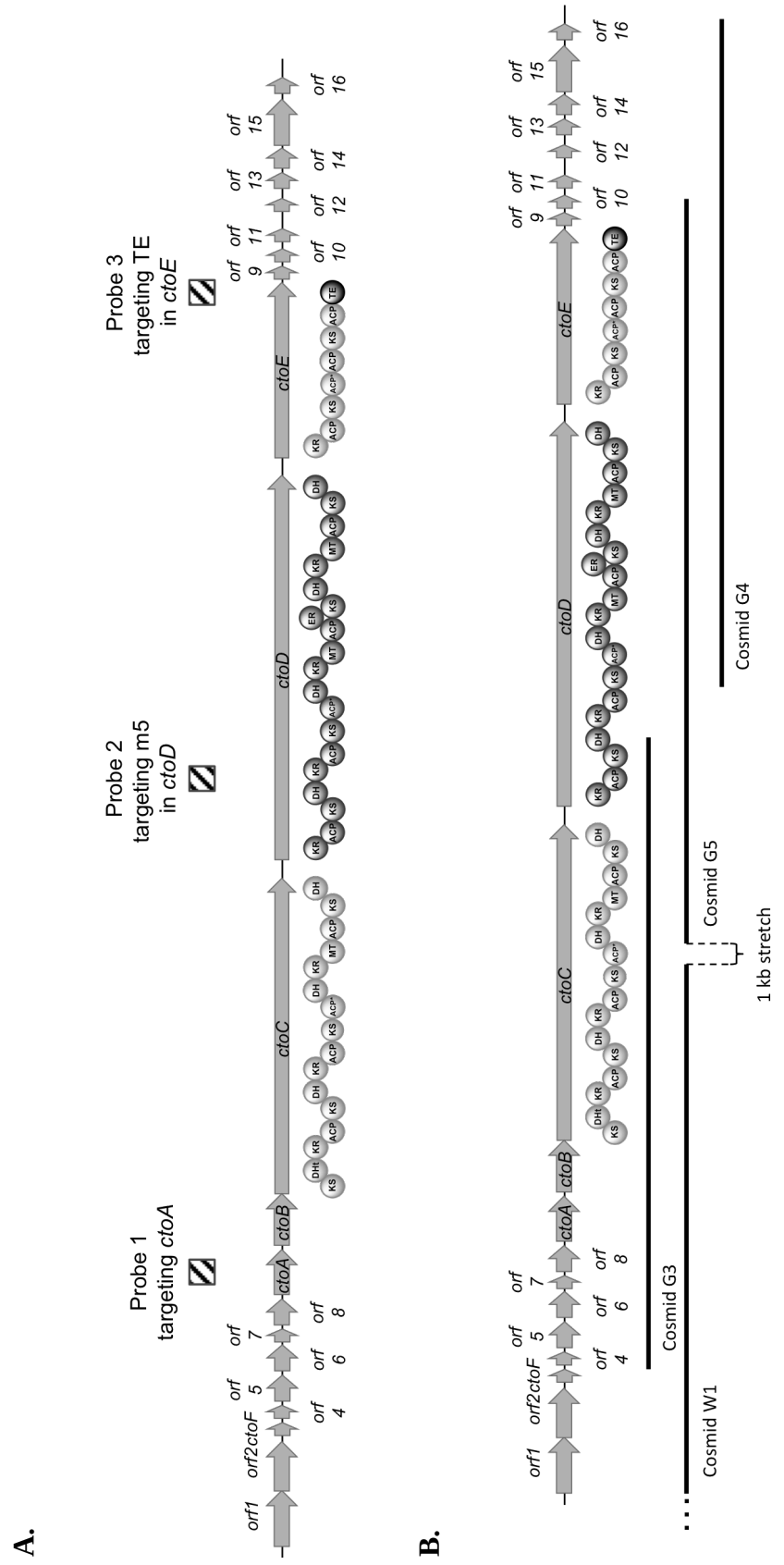


Figure 1: A.: schematic depiction of the localization of hybridization probes used for screening of the So ce1525 cosmid libraries to identify cosmids harbouring parts of the chlorotonil biosynthetic gene cluster. B.: coverage of the chlorotonil biosynthetic gene cluster by the four identified hit cosmids.

3.3.2 Design strategy for the construct W1_MChr harbouring the 5’ part of the chlorotonil biosynthetic gene cluster

A comparison of the chlorotonil biosynthetic gene cluster found in So ce1525 with the version that what was found in the alternative producer So ce1128 allowed to determine the approximate borders of the gene cluster. Furthermore, a functional analysis of the genes encoded in the respective locus by BLAST analysis gave additional hints about their involvement in chlorotonil biosynthesis. In accordance to that, cosmid W1 covering the upstream region and the first two modules of the chlorotonil PKS was shortened in a way, that the sequence starts at the outlying methyltransferase encoding gene *ctoF*. Homology arms were designed such, that 9.3 kb upstream the MT encoding gene were removed from the cosmid together with the ampicillin resistance cassette of the Supercos1 backbone. Instead, a hygromycin resistance cassette, amplified from pSUP_Hyg[10], flanked by *Pac*I restriction sites was inserted at that position and served as selection marker. After successful shortening of W1, the hygromycin resistance cassette was removed again from the construct via *Pac*I digest and replaced by a 1 kb fragment homologous to the *mchC* gene from the myxochromide gene cluster, resulting in construct of 31.2 kbp. The *mchC* fragment serves as homology region for integration of the W1 construct into the host genome of *M. xanthus* DK1622. Alternatively also an mx8 or mx9 attB integration cassette could be used for this purpose. The 3’ end of the construct contains 1.2 kb sequence encoding for the KS domain of module three; this region serves as integration site for the second construct once the construct W1 was successfully integrated into the *mchr* site of DK1622.

3.3.3 Design strategy for the construct G5_BAC_G4 harbouring the 3’ part of the chlorotonil biosynthetic gene cluster

The second construct which spans 54 kbp, covers most of the PKS core genes as well as the downstream region of the gene cluster. Furthermore this construct contains the 1 kb region not covered by any of the three cosmids W1, G5 and G5. To facilitate the maintenance and stabilization of the large construct, a BAC backbone was used instead of the Supercos1 backbone.

The assembly of this construct was made in three consecutive cloning steps involving conventional ligation as well as RedET recombineering. Initially, the backbone exchange for cosmid G5, combined with attachement of the 1 kb fragment linking G5 to W1 was

envisaged. For this purpose a tetracycline resistance cassette, amplified from pSWU22 (Wu, Kaiser, personal communication), was introduced into pBeloBAC11 via RedET. In the same step a restriction site for *Eco*72I was added to the vector. The 1 kb gap between cosmid W1 and G5 was bridged with the respective DNA fragment amplified from cosmid G3 using a stitching PCR approach. At the same time, a 1 kb homology region to cosmid W1 as well as a tandem homology arm for combination with cosmid G5 in a later RedET step was added. The PCR fragment was ligated to the multi-purpose cloning vector pJET which allowed for preparation of larger DNA amounts from *E. coli*. The resulting pJet construct as well as the modified pBeloBAC vector pBeloBAC_tet were incubated with the restriction enzymes *Sca*I and *Eco*72I and subsequently the thus released PCR product was ligated to pBeloBAC_tet. During restriction and ligation, the chloramphenicol resistance gene was removed from the pBeloBAC_tet backbone and replaced with the PCR fragment, yielding the construct pBeloBAC_stretch.

In the following step the resulting pBeloBAC_stretch was linearized with *Xma*JI. The vector then served as linear fragment in a LHRC RedET step with cosmid G5. As a result the Supercos1 backbone was to be removed from cosmid G5 and replaced with the BAC backbone of pBeloBAC to form the construct G5_BAC. The RedET experiment for the backbone exchange in G5 was not straight forward. Several trials using pBeloBAC_stretch linearized with *Xma*JI - and G5 as the circular construct in a LCHR RedET experiment failed. Thus in-depth *in silico* analysis of the two constructs was carried out. Alignment of the G5 cosmid and the pBeloBAC fragment revealed the presence of a 650 bp span on the cosmid backbone representing a truncated tetracycline resistance gene. This 650 bp fragment hindered the RedET experiment with pBeloBAC_stretch. Since the linear BAC fragment also contains a tetracycline resistance gene, the truncated version on the supercos1 backbone hampered the RedET cloning as it acted as concurring third homology arm. To solve this problem a set of restriction sites was sought to remove the unwanted part from the supercos1 backbone without affecting the integrity of the remaining construct. Two *Mun*I restriction sites framing the respective region as depicted below were identified and used for removal of the truncated tetR gene whilst maintaining the pUC ori and the ampicillin resistance cassette for selection.

The *Mun*I-digested G5 cosmid was re-ligated and recombination with pBeloBAC_stretch could then be successfully conducted to yield G5_BAC.

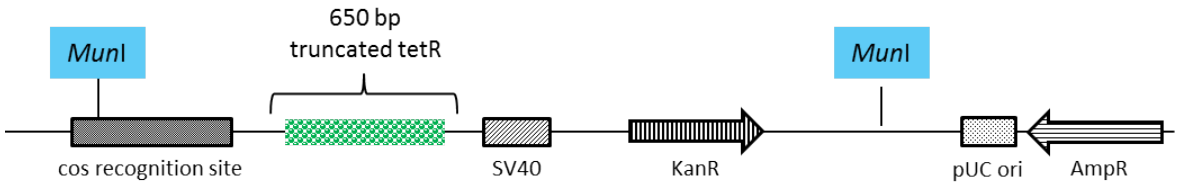


Figure 2 Drawing showing the location of the truncated tetR gene on the supercos1 backbone and the location of the two MunI restriction sites.

Prior to conjunction of the cosmid G4 with G5_BAC, redundant DNA sequence was removed from the cosmid. The G4 cosmid harbouring the 3' end of the chlorotonil A biosynthetic gene cluster plus the genomic region downstream of the gene cluster was shortened in an approach similar to that for cosmid W1. An apramycin resistance cassette was amplified and suitable homology arms were attached via the respective primers. In a LCHR experiment the cosmid was shortened whilst at the same time, the kanamycin resistance cassette was removed from the Supercos1 backbone and replaced by the apramycin cassette. The final construct, G4_Apra, was linearized using *Nsi*I to serve as linear fragment for the following steps. As the resulting 5.4 kb fragment of cosmid G4 and the construct G5_BAC share approximately 1 kb of the chlorotonil gene cluster, this region served as a "natural" homology arm. After this RedET step, the final construct G5_BAC_G4, spanning 54.1 kbp was obtained.

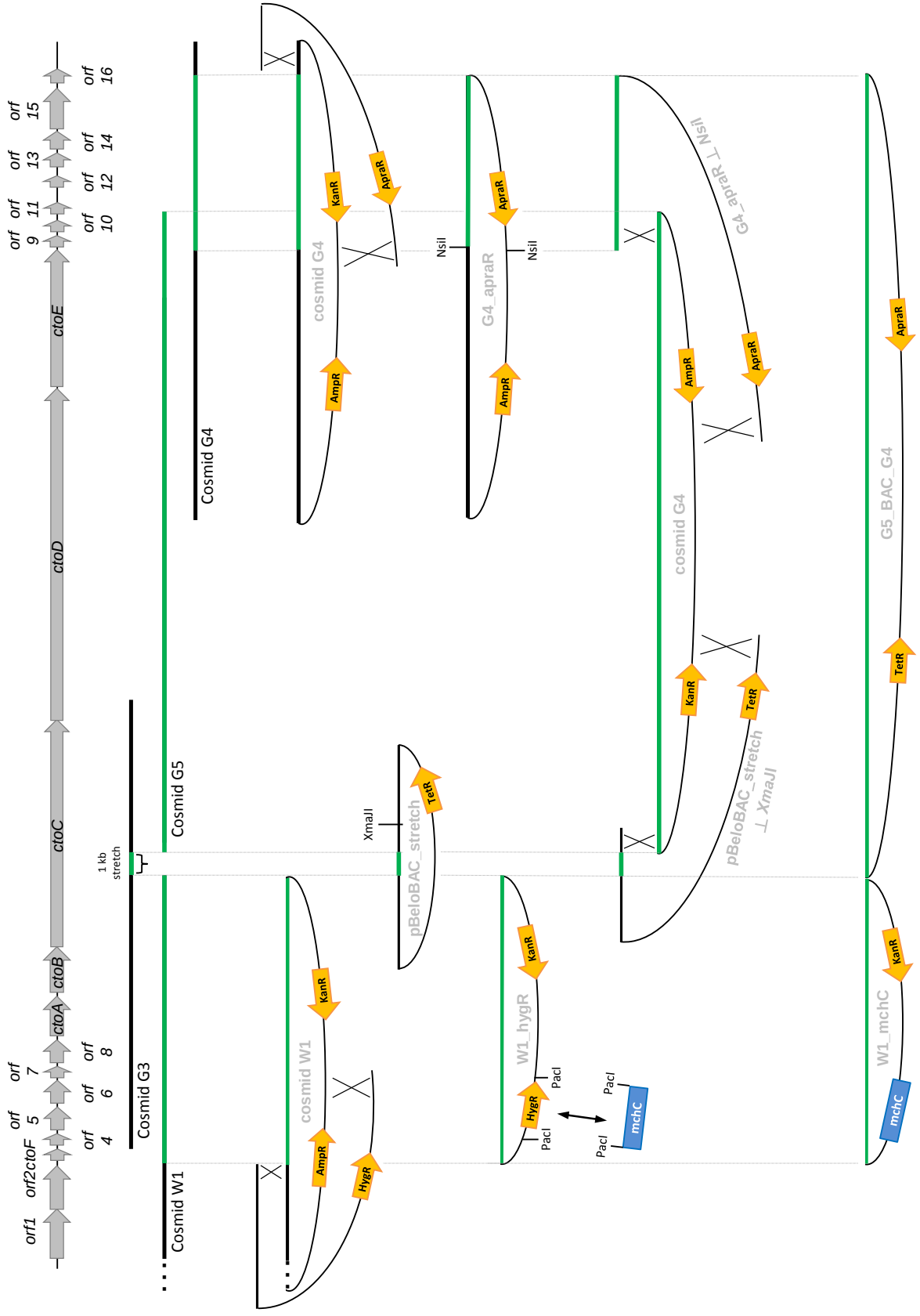


Figure 3 Overview of the stitching strategy applied for assembly odf the chlorotonil biosynthetic gene cluster from the cosmids W1, G3, G4, G5

3.3.4 Consecutive integration of the constructs W1_Mchr and G5_BAC_G4 into *Myxococcous xanthus* DK1622

Upon finalization of the two constructs W1_mchC and G5_BAC_G4, together covering the chlorotonil biosynthetic gene cluster, Illumina sequencing of both constructs was conducted in order to check for putative mutations or deletions acquired during the assembly process. As no such alterations could be detected after sequencing, consecutive integration of the two constructs into the host genome of *Myxococcus xanthus* DK1622 was pursued. As already mentioned the myxochromide gene cluster present in *M. xanthus* DK1622 was chosen as integration locus; this strategy was based on the knowledge that the myxochromide gene cluster is well expressed in the host strain, which results in a reasonable expression level of myxochromides.[11] Thus it was expected to benefit from this position effect when choosing the myxochromide gene cluster locus as integration site for the chlorotonil biosynthetic gene cluster. For this, stepwise transformation of the two constructs was conducted. Initially the construct W1_mchC was transformed into *M. xanthus* via electroporation yielding the kanamycin resistant mutant *Myxococcus xanthus* DK1622::W1_mchC. The obtained mutant was then transformed with the second construct G5_BAC_G4, whereas the tetracycline resistance cassette served as selection marker. Several mutants, resistant to kanamycin and tetracycline were obtained and pre-screened via PCR. Whole genome Illumina sequencing of one of the *Myxococcus xanthus* DK1622::W1_mchC::G5_BAC_G4 mutants was carried out to monitor correct and complete integration of the two constructs. Results of the genome sequencing are still pending.

3.3.5 Phenotypic analysis of the DK1622::W1_mchC::G5_BAC_G4 mutant

In order to check whether the produced mutant was capable of expressing the chlorotonil pathway and thus produces chlorotonil or derivatives of the compound, a phenotypic analysis of DK1622::W1_mchC::G5_BAC_G4 was carried out. For this purpose the mutant strain was grown in liquid culture in triplicates with addition of the adsorber resin XAD 16. As a control the mutant *Myxococcus xanthus* DK1622::W1_mchC was cultivated in parallel. After three days cultivation time, cells and XAD were harvested and extracts which subsequently underwent HPLC-MS analysis were prepared. HPLC_MS analysis of the extracts in comparison to the wildtype strain showed no signs of

production of chlorotonil or any of its congeners or precursors reported previously (comp. chapter two).

3.3.6 Plasmid recovery from DK1622::W1_mchC::G5_BAC_G4

The strategy for integration of the two constructs was designed in a way that upon successful integration of the complete biosynthetic gene cluster into *M. xanthus* DK1622, recovery of a single plasmid harbouring the complete gene cluster was enabled. For this two *Pac*I sites flanking the chlorotonil biosynthetic gene cluster together with the pBeloBAC backbone were introduced during the assembly. For selection of this construct an apramycin resistance cassette and a tetracycline resistance cassette would be available. *Pac*I restriction digest of genomic DNA from DK1622::W1_mchC::G5_BAC_G4 followed by re-ligation of the mixture would lead to formation of this construct. Transformation and subsequent selection on apramycin and tetracycline allows for selection of the desired plasmid. Upon verification of complete and correct integration of the chlorotonil gene cluster into the genome of *Myxococcus xanthus* DK1622, plasmid recovery would be pursued.

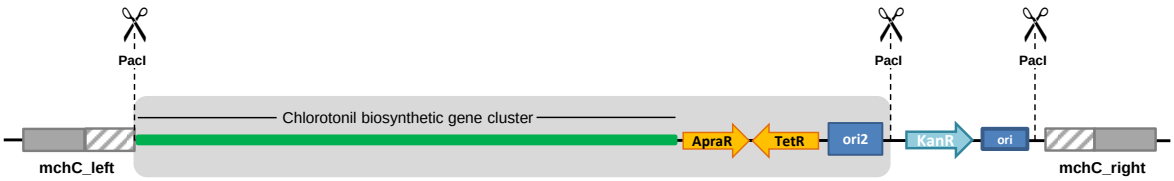


Figure 4: Schematic depiction of the plasmid recovery strategy of the complete biosynthetic gene cluster for chlorotonil. Restriction of genomic DNA from DK1622::W1_mchC::G5_BAC_G4 isolated the pathway encoding genes including the pBeloBAC_tet backbone. The plasmid resulting from relegation of the obtained mixture can be selected for using apramycin and tetracycline.

3.4 Conclusion

Chlorotonil A is a promising new lead structure with outstanding activity *in vitro* and *in vivo* against the problematic malaria pathogen *Plasmodium falciparum*. With a very early onset, chlorotonil A was shown to be active against all blood stages of the disease whilst at the same time only low cytotoxicity could be observed. Moreover, also *Plasmodium* species resistant to the standard treatment with chloroquine were sensitive against chlorotonil, indicating that chlorotonil hits target different to that of chloroquine.[1] In addition to these very promising results, chlorotonil also shows good antibacterial activity against Gram-positive indicator strains (see chapter 2). The simultaneous observation of antiparasitic and antibacterial activity hints that the target of chlorotonil A in *Plasmodium* might be its apicoplast, but this presumption is still highly speculative as identification of the chlorotonil target is still work in progress.[12] Nevertheless, chlorotonil A is a good candidate for further development as antimalaria-drug. To address the issues arising from the high lipophilicity of the compound, two different approaches were carried out. On the one hand semi-synthetic derivatisation of chlorotonil A is pursued (see chapter 4) while synchronous endeavors towards the production of derivatives via alterations in the chlorotonil biosynthetic machinery were initiated. For the later approach the establishment of a platform for heterologous expression of the chlorotonil pathway is an indispensable prerequisite as the producer strain, *Sorangium cellulosum* So ce1525, is very slow growing and notoriously hard to manipulate with genetic tools.

With the above described cloning of the biosynthetic pathway, we set the stage for the design and construction of gene cluster variants enabling the production of chlorotonil derivatives. The fact that expression of the chlorotonil pathway in *Mycococcus xanthus* was not met with success so far, raises several questions which have to be looked into in further detail. In case the whole genome sequencing of the obtained double mutant *Myxococcus xanthus* DK1622::W1_mchC::G5_BAC_G4 reveals complete and faultless integration of the constructs into the host genome, several scenarios responsible for absence of chlorotonil production are conceivable. It might be that the native promoters are not functional or not sufficiently strong to drive expression of the gene cluster in *Myxococcus*. This could be improved by furnishing the construct W1_mchC with an inducible promoter system, such as the vanillate promoter established for *Myxococcus*, prior to integration into the host genome.[13] However it is also possible, that in the case of

the chlorotonil pathway, *Myxococcus* is not a suitable host strain which could result from difficulties with the codon usage of the *Sorangium*-derived gene cluster. Consequently another host, most favourable from the same genus *Sorangium* would have to be established for heterologous expression; after plasmid recovery from *Mycococcus*, the resulting construct – adapted to the requirements for integration into the respective *Sorangium* genome – could be used to achieve this goal. Ultimately it might as well be the case that some functionalities are missing in the gene cluster isolated so far, hence in-depth analysis of up- and downstream regions of the biosynthetic gene cluster in the producer strain and if necessary complementation of the existing construct with additionaly required functionalities could facilitate functional expression of the chlorotonil biosynthetic gene cluster in a heterologous host.

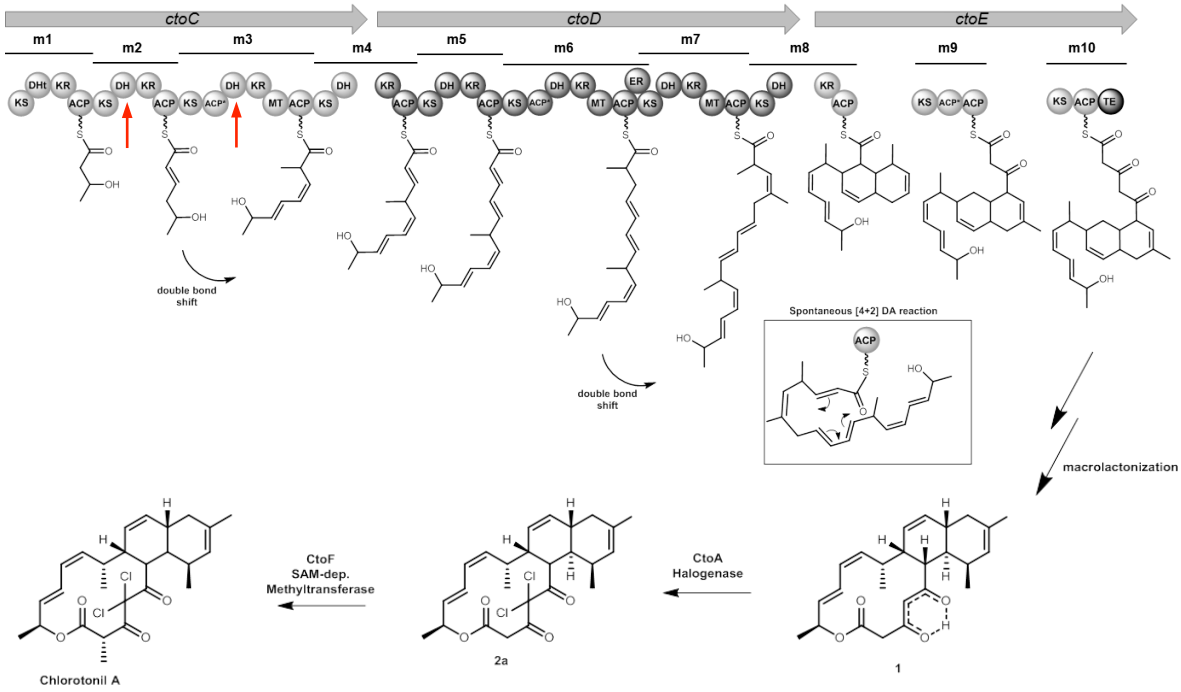


Figure 5 A model for chlorotonil biosynthesis in So ce 1525. The organization of the chlorotonil A biosynthetic gene cluster is schematically shown in A. The proposed biosynthetic pathway for chlorotonil A and B is shown in B. ACP, acyl carrier protein; DH, dehydratase, DHt, inactive truncated dehydratase; ER, enoyl reductase; KR keto reductase; KS, keto synthase; MT, methyl transferase; TE thioesterase. The candidate DH domains for inactivation in modules two and three are marked with red arrows.

As soon as heterologous expression of the chlorotonil biosynthetic pathway can be achieved, the question arises how to modify the biosynthetic gene cluster in order to obtain viable derivatives of the compound. In this context there are two aspects that need to be considered. Initially the *gem*-dichloro-1,3-dione structure which is considered the pharmacophore of chlorotonil A (see chapter 2 and 4) is not be touched as it could already

be shown that such modification renders the substance inactive. Furthermore, the potential target regions for modifications in the biosynthetic gene cluster have to be chosen carefully, as several reducing domains are crucial to produce a precursor molecule fulfilling the conditions for spontaneous Diels-Alder cyclization (see chapter 2). Consequently there are dehydratase domains that may not be inactivated as the resulting spatial arrangement would impede the [4+2] cycloadditon and thus strongly affect the basic framework of chlorotonil A. Following the biosynthesis scheme only two target points remain after taking the above mentioned aspects into account. Free hydroxyl groups would clearly contribute to the hydrophilicity of chlorotonil, hence inactivation of dehydratase domains responsible for dehydration and installation of a double bond in the respective β-keto position is the consistent approach. Considering the prerequisites for formation of the chlorotonil basic framework as explained above, only the dehydratase domains of module two and three (see figure 5) are suitable candidates for inactivation. Simply thought, inactivation of one or both of the DH domains would lead to one or two free hydroxyl groups at the positions C19/C20.

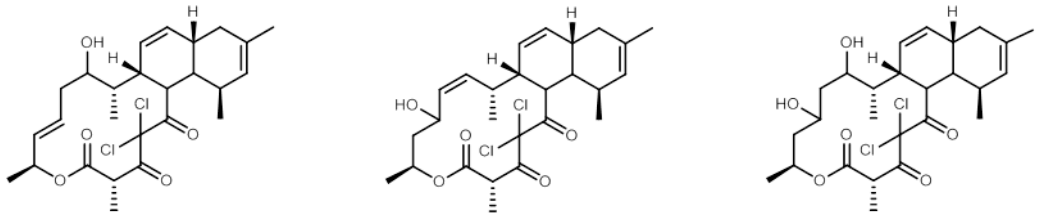


Figure 6: putative hydroxyl-chlorotonil derivatives originating from individual inactivation of the two DH domains in modules two and three or simultaneous inactivation of both DH domains.

But thinking beyond this obvious aim, the substrate specificity of ketosynthase domains which is commonly seen in *trans*-AT PKS systems such as the chlorotonil PKS could obstruct this plan.[14] The previously planar double bond moiety in the molecule would thereby change into a sterically more demanding branched substructure in both cases, which might lead to problems in further processing of the resulting intermediates and could result in offloading of the polyketide intermediate. Therefore careful *in silico* analysis of the respective ketosynthase binding pockets and presumably also loading assays with different substrates to probe the substrate tolerance of the two succeeding KS domains would probably be required.[15]

Summing up, through isolation and cloning of the biosynthetic gene cluster responsible for chlorotonil A biosynthesis, we paved the way to establish a platform for heterologous expression of this biosynthetic machinery. Once a system of such kind is set up, further investigations of biosynthetic processes can be carried out as well as yield optimization in the heterologous host. The main goal of our work on chlorotonil is however to find suitable routes to access derivatives of the compound which retain their bioactivity. Besides the production of semi-synthetic derivatives also genetic engineering of the biosynthetic pathway is one option to achieve this aim. In comparison to chemical derivatisation, the genetic engineering approach has the advantage that upon implementation of an adequate biological production process, derivatives can be obtained in a continuous process without prior isolation of starting material and several intermediate steps of workup.

3.5 Experimental section

3.5.1 Bacterial strains and culture conditions

Sorangium cellulosum So ce1525 as well as *Escherichia coli* strains were grown as reported previously. *Myxococcus xanthus* DK1622 was cultivated at 30°C and 180 rpm in liquid CTT media (casitone, Bacto (10 gL_1), MgSO4 (8 mM), Tris-HCl (10 mM) pH 8.0, K2HPO4 (1 mM) pH 7.6, final pH adjusted to 7.6) and maintained on CTT agar plates (CTT medium with 1.5 % bacto agar); CTT soft agar was prepared with 0.75 % bacto agar. DK1622 mutants strains were grown in the presence of kanamycin 50 μg/ml or kanamycin 50 μg/ml and oxytetracycline 5 μg/ml. For liquid cultures, 20 ml of CTT media were inoculated with DK1622 cells from a CTT agar plate and incubated for 2-3 days at 30 °C. 0.5 - 2ml of thus set up pre-cultures were used to inculate 25 ml of fresh CTT media in 100 ml Erlenmeyer flasks. These cultures were incubated at 30 °C and 180 rpm shaking for 2 – 4 days. For permanent stocks, agar pieces grown with the strain were frozen at –80°C in cryovials.

3.5.2 Preparation of genomic DNA for construction of a cosmid library

So ce1525 genomic DNA was prepared either by phenol/chloroform/isoamylacohol extraction. For this a 50 ml liquid culture of So ce1525, grown for seven days, was spun down at 13.000 rpm for 10 min; the supernatant was discarded. The cell pellet was washed twice with SET buffer (75 mM NaCl, 25 mM EDTA pH8.0, 20 mM Tris-HCl pH

7.5) and were finally resuspended in 5 ml SET buffer. 50 µl RNAse A stock solution (10 mg/ml in water) were added to the cells and the sample was incubated at 37 °C for 45 min.. Upon addition of 300 µl proteinase K solution (10 mg/ml in 50 mM Tris-HCl pH 8.0; 1 mM $CaCl_2$) the sample was inverted several times; subsequently 600 µl of an 10 % SDS solution were added and the sample was incubated at 55 °C for two hours whereby the sample was inverted every 15 minutes. Thereafter one volume of Phenol:Chloroform:Isoamylalcohol mixture (25:24:1) was added and the sample was swung at 5 rpm for 60 minutes. Upon centrifugation of the mixture at 8000 rpm for 10 min at 4°C, the upper phase of the sample was transferred to a new tube using a wide – end 1ml tip; special care was taken no to contaminate the sample with the white interphase layer. This procedure was repeated until the white interphase was significantly reduced. As last extraction step one volume of Chloroform:Isoamylalcohol (24:1) was added to the sample and the above described procedure was repeated once. The upper phase was again transferred to a new tube; 1/10 volume of 3M Na-acetate pH 5.5 was added and the sample was inverted several times. Thereupon 2.5 volumes of ice-cold absolute ethanol were added to the tube which was then inverted several times until the appearance of white cotton-like genomic DNA. The DNA was transferred to a new tube containing ice-cold ethanol 70 % and washed by inverting the tube several times. The sample was carefully spun down and air dried overnight in the open tube. Finally the genomic DNA was dissolved in 5.5- 1 ml sterile water.

3.5.3 Construction and screening of a So ce1525 cosmid library

A cosmid library of the strain *Sorangium cellulosum* So ce1525 was constructed and screened as previously described. Thus identified hit cosmids were sent for end-sequencing whereas the candidates found in this pre-screening were sent for Illumina sequencing.

3.5.4 Protocol for RedET recombineering – linear to circular homologous recombination (LCHR)

An overnight culture of *E. coli* HS996tet harboring the respective circular construct was used to inoculate 1.4 ml LB media containing 5 µg/ml tetracycline. Different inoculum sizes were chosen (e.g. 5, 15 and 30 µl) and every dilution was setup three times. One out of three identical cultures of every dilution was marked and used of OD_{600} determination.

The cultures were incubated at 30 °C and 800 rpm shaking until an OD600 of 0.2 was reached but at least for two hours. 21 µl of arabinose 10% were added to the respective tubes which were then shifted to 37 °C and incubated for another 45 min until an OD_{600} of 0.4 was reached. The cells were spun down at 13000 rpm, 4 °C for 2 minutes and washed three times with 500 µl of ice-cold sterile water. After the last washing step, ca. 50 µl of water were in left in the tube to resuspend the cells. 1 µl of the respective linear construct was added to the cells, the mixture was transferred to an electroporation cuvette and electroporated at standard *E. coli* conditions (1250 V, 200Ω, 25µF). 1 ml of fresh LB media was added to the cells that were subsequently incubated at 37 °C for one hour. The cells were plated out on LB agar plates containing the required antibiotics and incubated overnight at 30 °C.

3.5.5 Transformation of Myxococcus xanthus DK1622

For transformation *M. xanthus* DK1622 was grown in CTT media until an OD600 of 0.8 – 1.0 was reached. 1.5 ml of the liquid culture were then transferred to an eppendorf tube and spun down. After two washing steps with 1 ml H2O at room temperature, competent cells were resuspended in 30 µl H_2O. To 30 µl competent cells, 0.3 – 0.6 µg of DNA were added; the mixture was transferred to an electroporation cuvette and transformed under the given conditions:

Resistance	400Ω
Voltage	0.65kV
Capacity	25 µF

The time constants to be reached were around 8-10 ms. Directly after electroporation 1 ml of CTT media was added to the cells and the samples were shaken at 30°C for 6 hours. Afterwards three different aliquots (e.g. 100, 300 and 600 µl) of the cell suspension were added to 3.5 ml CTT softagar containing the appropriate antibiotic, each. This soft agar mixture was then poored onto CTT agar plates also containing the respective antibiotics. Plates were incubated at 30 °C until clones were visible.

3.5.6 Phenotypic analysis of DK1622::W1_mchC::G5_BAC_G4

For analyzing the metabolic profile of DK1622::W1_mchC::G5_BAC_G4, cells of the mutant strain maintained on a CTT agar plate supplemented with kanamycin and

oxytetracycline were used to inoculate 10 ml of CTT liquid media equally supplemented with kanamycin and tetracycline. This preculture, grown at 30°C and 180 rpm for two days, was then used to inoculate 3 300 ml flasks each containing 50 ml CTT media and the appropriate antibiotics. After cultivation overnight under the described conditions, 2 % XAD16 were added to the flasks and the cultures were grown for another one to two days. Subsequently cells and XAD were harvested by centrifugation and the resulting pellets were frozen at -20 °C overnight.

The pellets were first extracted with 30 ml methanol to remove polar substances and then three times with 30 ml acetone for 20 min each. The combined acetone fractions were evaporated and the resulting residue was dissolved in 1 ml chloroform. For further analysis the samples were centrifuged; aliquots were taken and diluted 1:10 with methanol for HPLC analysis; the remaining samples were stored at -20 °C.

3.6 Supporting Information

3.6.1 Images of nitrocellulose membranes spotted with the So ce1525 cosmid library after membrane hybridization

A. Spotting scheme

Obtained cosmid libraries containing up to 1920 clones were spotted on nitrocellulose membranes using the Q-Pix spot picker robot system. The below shown scheme (“4x4 array”) was applied for arranging the *E. coli* clones from the five 384 well plates containing the cosmid clones. This array allows obtaining unambiguous signals which can be backtracked to the specific plate and well they originate from. Following this approach six identical membranes of the cosmid library were produced; this opened the opportunity for memebrane hybridization experiments with different probes in parallel.

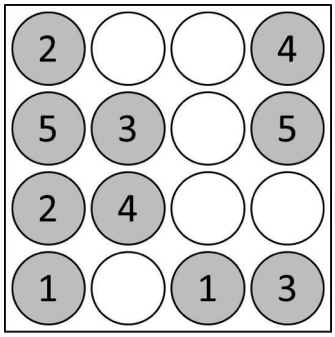
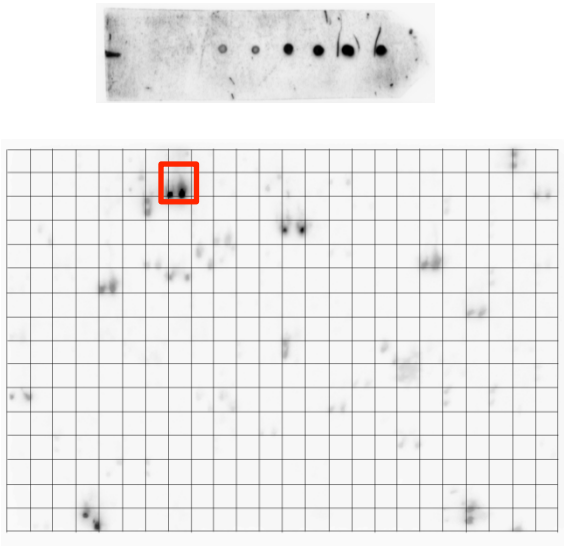


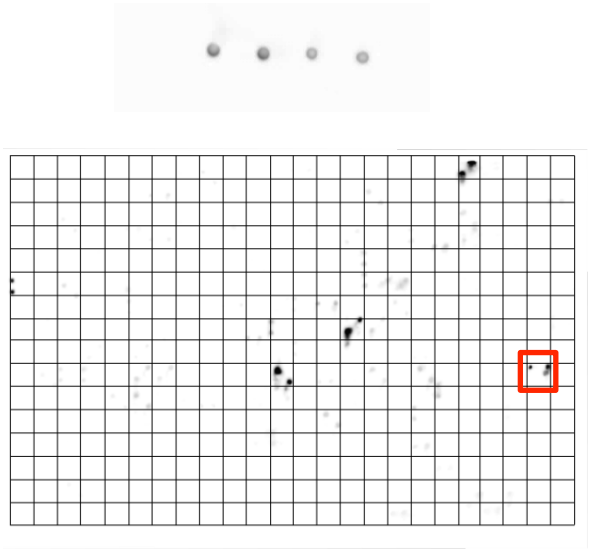
Figure 3 Spotting scheme for cosmid libraries on nitrocellulose membranes. This diagram is exemplary for spotting five 384 well plates on the membranes. Every well is spotted twice following a distinct pattern to ensure clear signals. A given signal can only be considered positive, if it shows the respective pattern.

Prior to membrane hybridization the respective A1 position of every membrane was characteristically marked to allow for correct orientation of the membranes during evalution of the obtained results. A dot-blot positive control was included during every round of memebraen hybridization. For this, the probe itself was spotted on a piece of nitrocellulose membrane in several dilutions and crosslinked via UV radiation. Thus produced positive control samples were carried along during membrane hybridization.

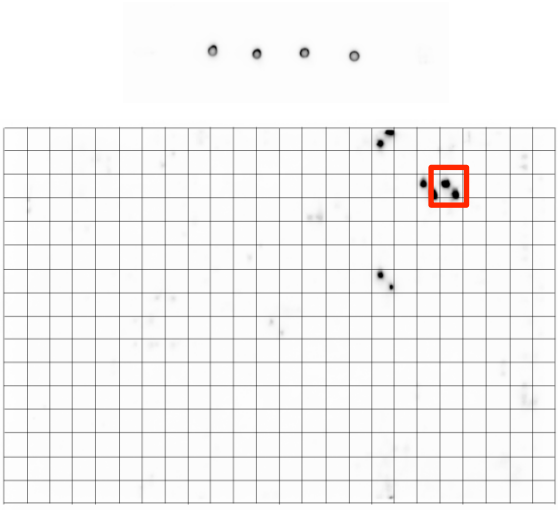
A. Cosmid W1 – Probe 1 (*ctoA*) position: B8



C. Cosmid G5 – Probe 2 (*ctoD*) position: J23



B. Cosmid G3 –Probe 1 (*ctoA*) position: C20



D. Cosmid G4 – Probe 3 (TE *ctoE*) position: D20

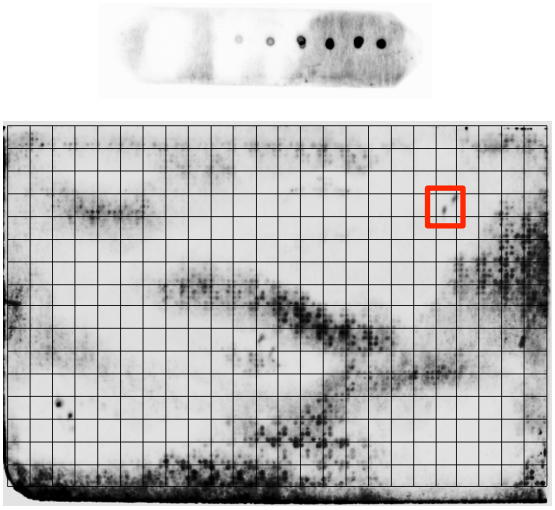


Figure 4 Recording of nitrocellulose membrane images after memebrane hybridization. After detection, the membranes were superimposed with a 24x16 (384) grid The highlighted signals were back traced to the respective 384 well plate position.

3.6.2 Oligonucleotides

Table 3: Oligonucleotide primers used during assembly of the chlorotonil biosynthetic gene cluster from cosmids.

Primer	5'-Sequence-3'	Restriction site
W1_HygR_fwd	GCTGGGCCGTGCGCGGGCCGACCGCGAGAATAGATCGCCGGTGCGC**TTAATTAA**CAGCTTCACGCTGCCGCAAGC	*Pac*I
W1_HygR_rev	TAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAG**TCTAGA**CCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAG**TTAATTAA**TCAGGCGCCGGGGGCGGTGTC	*Xba*I, *Pac*I
Apra_fwd(G4)	ATTTGTCCGCCTCCGCCGGCCTCGGATAGGCTTTCTTGTCGGGCGCGGTTCATGTGCAGCTCCATCA	
Apra_rev(BAC NsiI)	AATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCTTGGTCGGTCAT**ATGCAT**AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGAC	*Nsi*I
TetR_fwd (HA_BAC)	TGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTCGTAATTCTCATGTTTGACAGC	
TetR_rev (HA_BAC)	TTATCGCGAATAAATACCTGTGACGGAAGATCACTTCGCAGAATAA**CACGTG**CAAGGGTTGGTTTGCGCATTC	*Eco*72I
P1 (W1 HA_fwd)	GAT**GGCCGCCCGGGCC**T**TAATTAA**CCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGTTAATTAACGACAAGTTCGACGCCAGGTT	*Sfi*I, *Pac*I
P2	AGCCGCTCCCAGGGCCTCGAC	
P3	GTCACCAGGTCGAGGCCCTGGGAGCGGCTGTGCGACGACGGAGGGCGGG	
P4 (2xG5HA_rev)	GAT**CACGTG*ATGCAT***AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACA**CCTAGG**CTCGCCGTCGACGACGAGGGGGCGCAGC	*Eco*72I, *Nsi*I, *Xma*JI
Apra_rev (BAC-NsiI)	AATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCTTGGTCGGTCAT**ATGCAT**AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACATCAGCCAATCGACTGGCG	*Nsi*I
MChr16_fwd	GATT**TAATTAA**GTCAGGCCCAGAGGACGCC	*Pac*I
MChr19_rev	ACAT**TAATTAA**CACCCGGCTCAAGGAGGCG	*Pac*I

3.6.3 Results from Illumina sequencing of the two constructs W1_mchC and G5_BAC_G4

The two constructs W1_mchC and G5_BAC_G4 were isolated from *E. coli* GB2005 and subsequently sent for Illumina sequencing. The resulting data was assembled in-house with the help of the in-silico sequence data compiled from cosmid sequence data and plasmid maps. The thus obtained sequence data of the two constructs was mapped to the *in-silico* data using MAUVE alignment to detect deletion, insertions and mutations. In W1_mchC several deletions, insertions and mutations could be observed in the cos-recognition site stemming from the supercos1 backbone; furthermore two smaller insertions in non-coding regions and a 11 bp deletion in the pUC ori could be detected. As the cos recognition site is not needed and the pUC ori was only used to initiate replication in *E. coli*, neither the transformation of the construct into *M. xanthus* DK1622 nor the expression of the gene cluster in the host strain should be affected by these alterations.

Table 4: Alterations in W1_mchC acquired during cloning

Alterations in W1_mchC acquired during cloning		
Position	**type of modification**	**effect**
cos recognition site (24074 – 25127)	numerous deletions and point mutations	none observed
pUC ori (28944 -29622)	11 bp deletion at the 5’ end	none observed
29769 (non-coding)	insertion G	none observed
29813 (non-coding)	insertion G	none observed

For the construct G5_BAC_G4 also several alterations could be detected, most of which were located in the BAC backbone portion of the construct. Here, a 6 bp deletion upstream the introduced tetracycline resistance cassette was observed; a silent mutation in the sopB gene encoded on the pBeloBAC backbone was found as well as the insertion of a 2 bp GG in a non-coding region of pBeloBAC. The sopB gene, together with sopA and sopC is responsible for partitioning, meaning the sopAB act at sopB to make sure that every *E. coli* daughter cell is equipped with one copy of the plasmid. As the observed mutation is silent it does not influence this process, neither is it relevant for integration of

the construct into the host genome or expression of the chlorotonil gene cluster in DK1622. Summing up, the number of alterations especially seen in the region of the pBeloBAC backbone hints that the plasmid map provided in the database might be inaccurate. The G → A point mutation downstream the ACP3 encoding region of ctoC should also not influence correct expression of the chlorotonil biosynthetic gene cluster.

Table 5: Alterations in G5_BAC_G4 acquired during cloning

Alterations in G5_BAC_G4 acquired during cloning		
Position	**type of modification**	**effect**
5043 (non-coding)	insertion C	none observed
pBeloBAC backbone (6439-6444)	6 bp deletion, region of tetR_fwd primer	none observed
sopB pBeloBAC backbone (10365-10368)	silent 4 bp mutation TAGT → ATCG	none observed
pBeloBAC backbone (13452-13453)	insertion of GG	none observed
downstream ACP; non-coding (15070)	point mutation G → A	none observed

3.7 References

[1] J. Held, T. Gebru, M. Kalesse, R. Jansen, K. Gerth, R. Müller, B. Mordmüller, *Antimicrob. Agents Chemother.* **2014**, *58*, 6378–84.

[2] N. Rahn, M. Kalesse, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 597–9.

[3] M. A. Marahiel, *Chem. Biol.* **1997**, *4*, 561–567.

[4] H. von Döhren, U. Keller, J. Vater, R. Zocher, *Chem. Rev.* **1997**, *97*, 2675–2706.

[5] S. Donadio, M. M. Staver, J. J. McAlpine, S. Swanson, L. Katz, *Science (80-.).* **1991**, *252*, 675–679.

[6] C. Hertweck, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688–716.

[7] Y.-Q. Cheng, G.-L. Tang, B. Shen, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3149–54.

[8] J. Muyrers, *Nucleic Acids Res.* **1999**, *27*, 1555–1557.

[9] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, a S. Durkin, J. a Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15200–15205.

[10] S. Pradella, A. Hans, C. Spröer, H. Reichenbach, K. Gerth, S. Beyer, *Arch. Microbiol.* **2002**, *178*, 484–492.

[11] K. Gerth, H. Reichenbach, G. Hofle, K. Gerth, V. Wray, M. Weg, A. From, G. Bacteria, D. Stamm, **1993**, 1233–1237.

[12] M. Kalanon, G. I. McFadden, *Biochem. Soc. Trans.* **2010**, *38*, 775–782.

[13] A. a. Iniesta, F. García-Heras, J. Abellón-Ruiz, A. Gallego-García, M. Elías-Arnanz, *J. Bacteriol.* **2012**, *194*, 5875–5885.

[14] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–33.

[15] M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel, N. J. Oldham, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 1143–7.

Chapter 4

Epochlorotonil A and B, two new bioactive semi-synthetic chlorotonil derivatives with advanced pharmacokinetics

Contributions and Acknowledgements

A. Author's efforts

The author designed and performed most of the experiments described and evaluated and interpreted the resulting data. The author cultivated the chlorotonil producer strain So ce 1525 and isolated chlorotonil A. The author performed, adapted and monitored the described derivatisation reactions yielding epochlorotonil A and B. Furthermore the author carried out the solubility testing for all described chlorotonil derivatives.
The author conceived and wrote the chapter.

B. Contribution of others

The author would like to thank Kirsten Harmrolfs for her help and assistance in planning semi-synthesis routes and purification protocols and for her advice throughout the project. Moreover the author would like to thank Rolf Jansen for providing dihydro-chlorotonil and the related NMR data and Lena Keller for structure elucidation of epochlorotonil derivatives. Additionally the author would like to thank Jennifer Herrmann and Viktoria Schmitt as well as Jana Held (University of Tübingen; anti-plasmodial activity testing) for bioactivity testing of the produced derivatives. Scientific supervision of the project as well as proofreading of the chapter was done by Daniel Krug and Rolf Müller.

The author would like to thank David Auerbach for his help in purification of semi-synthetic chlorotonil derivatives and Thomas Hoffmann and Michael Hoffmann for HPLC-MS measurements.

4 Epochlorotonil A and B, two new bioactive semi-synthetic chlorotonil derivatives with advanced pharmacokinetics

4.1 Abstract

Epochlorotonil A and B are two new semi-synthetic derivatives of chlorotonil A, a structurally peculiar polyketide produced by the myxobacterium *Sorangium cellulosum* So ce1525. Chlorotonil A has received particular attention for its outstanding *in vitro* and *in vivo* activity against the problematic malaria pathogen *Plasmodium falciparum*. It could be shown that with a very early onset the compound is active against all blood stages of the disease. Chlorotonil A also exhibits good antibacterial activity against Gram-positive indicator strains. In comparison to chlorotonil A, the two epochlorotonils still show reasonable antibacterial and good anti-plasmodial activity with epochlorotonil B being the more active derivative. Furthermore it could be shown that in contrast to another semi-synthetic derivative, the position of the modification in dihydro-chlorotonil A did not lead to loss of activity.

4.2 Introduction

In the last decades the resistance situation in the field of infectious diseases has steadily deteriorate. This is reflected by the growing number of nosocomial infections as well as by the re-emergence of formerly eradicated diseases. At the same time there is a lack of therapy options for tropical diseases such as malaria; although there are some well-established remedies and several therapy regimens, severe malaria caused by resistant *Plasmodium* strains is still incurable in most cases. Seen against this backdrop, there is a growing need for new anti-infectives with new modes of action, which are able to bypass existing resistance mechanisms. In this context, natural sources like bacteria, fungi and sponges have proven to be rich sources for new bioactive substances. After decade-long investigation of the streptomycetes metabolome, now the myxobacteria – soil-dwelling Gram-negative δ-proteobacteria – are increasingly gaining significance. With their typically large genomes of 9-15 Mbp myxobacteria are bearing the potential for biosynthesis of a plethora of secondary metabolites. This applies especially to myxobacteria of the genus *Sorangium*, which in several cases have proven to be prolific producers of structurally diverse, bioactive secondary metabolites. Likewise, chlorotonil

A, produced by the myxobacterium *Sorangium cellulosum* So ce 1525, showed promising bioactivity against Gram-positive bacteria and the malaria pathogen *Plasmodium falciparum.* One major drawback of the compound is its high lipophilicity which leads to insolubility in practically all polar solvents. This property of chlorotonil A hampers its bioavailability and makes formulation for oral administration impossible. Thus the aim of the work presented, was to semi-synthetically generate chlorotonil derivatives with decreased lipophilicity and enhanced bioavailability. As the target of the compound and its pharmacophore are still unknown, sites in the molecule which were accessible and at the same time not crucial for the bioactivity had to be addressed. However, initial bioactivity testing of chlorotonil A and several mono- and non-chlorinated congeners already provided first indications for the structural features of chlorotonil A required for bioactivity. It could be shown that the *gem*-dichloro-1,3-dione substructure and especially the dichloro-pattern, is crucial for bioactivity, as mono- and non-chlorinated derivatives were less or even not active anymore (Jungmann et al., *ACS Chemical Biology,* DOI: 10.1021/acschembio.5b00523).

4.3 Materials and Methods

4.3.1 Production and purification of semi-synthetic chlorotonil derivatives

The two epoxide derivatives epochlorotonil A and B were produced by a Prilezhaev epoxidation of chlorotonil A. Chlorotonil A was isolated and purified as described previously. For the production of the epochlorotonils 100 mg chlorotonil A (2.1×10^{-4} mol) were dissolved in chloroform to a final concentration of 5 mg/ml. To the stirred solution of chlorotonil A, 2 equivalents (4.2×10^{-4} mol) of meta-chloroperoxybenzoic acid (MCPBA) in chloroform (0.05 mg/ml) were added. The mixture was stirred at 200 rpm for 2 hours at room temperature. To quench the reaction, three volumes of a saturated sodiumhydrogencarbonate solution were added to the reaction mixture which was subsequently shaken thoroughly. Upon quenching, the water fraction and the organic fraction were separated. The water fraction was extracted once again with chloroform in a 1:1 ratio. The organic phases were combined and extracted with distilled water in a 1:1 ratio twice to remove salts. Chloroform was removed from the sample in a rotary evaporator under reduced pressure at 40 °C. An aliquot (1 mg) of the resulting white product was dissolved in chloroform to a final concentration of 5 mg/ml and diluted in methanol 1:10. The sample was subsequently analyzed by high resolution HPLC-MS. Purification of the product mixture on a semi-preparative Agilent Zorbax XDB C8 (250x9.4 mm, 5 µm) column with a 95 min H_2O/MeOH 5-95 % gradient provided 9.5 mg epochlorotonil A and 29 mg epochlorotonil B.

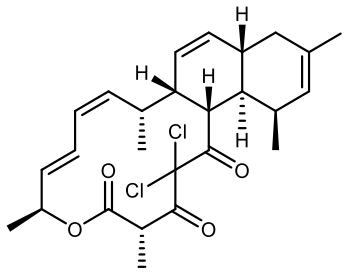


Figure 1: Structure of chlorotonil A from Sorangium cellulosum So ce1525

For reduction of chlorotonil A using dimethylaminoboran, 100 mg chlorotonil A (2.1×10^{-4} mol) were dispersed in 1 ml dry dichloromethane. 124 mg dimethylaminoborane (2.1×10^{-3} mol) in 0.5 mL dry dichloromethane were added slowly under stirring at room temperature. After 1 h another portion of 124 mg pure dimethylaminoboran (2.1×10^{-3}

mol) was added and the solution was stirred for another 2 h before 300 µL of acetone were added slowly. After 30 min dichloromethane was added to give 10 mL solution, which was extracted three times with 10 mL of 0.1 N HCl and three times with 10 mL water. The dichloromethane solution was dried with sodium sulfate and evaporated to give 90 mg raw product. Si-Flash chromatography on 12 g silica gel cartridge (Reveleris) with dichloromethane provided 29 mg of pure product #.

4.3.2 NMR and MS measurements

NMR spectra were recorded in chloroform-*d* on a Bruker Ascend 700 spectrometer with a 5 mm TXI cryoprobe (^{1}H at 700 MHz, ^{13}C at 175 MHz) DQFCOSY, HSQC, HMBC, and ROESY experiments were recorded using standard pulse programs. HSQC experiments were optimized for $^{1}J_{C\text{-}H}$ = 145 Hz, and HMBC spectra were optimized for $^{2,3}J_{C\text{-}H}$ = 6 Hz. The samples were dissolved in chloroform-*d* and the chemical shifts of the solvent signals at 7.24 ppm (δ_H) and 77.23 ppm (δ_C) were considered as internal standard (reference signal). The observed chemical shift (δ) values were given in ppm and the coupling constants (*J*) in Hz. 5mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA) were used to increase sensitivity.

All HPLC-MS measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 50 x 2.1 mm, 1.7 µm dp column. Separation of 1 µl sample was achieved by a linear gradient with (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 1 min isocratic step at 5 % B, followed by an increase to 95 % B in 6 min to end up with a 1.5 min step at 95 % B before reequilibration under the initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 2000 m/z at a 2 Hz scan speed.

4.3.3 MIC activity of the semi-synthetic chlorotonil derivatives against Gram-positive bacteria

The microorganisms used for MIC determination, were obtained from the German Collection of Microorganisms and Cell Cultures (*Deutsche Sammlung für*

Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection. They were were handled according to standard procedures. For microdilution assays, overnight cultures of *E. coli*, *M. luteus, B. subtilis,S. aureus* in Müller-Hinton broth (0.2 % beef infusion, 0.15 % corn starch, 1.75 % casein peptone; pH 7.4) and *E. faecalis, S. pneumoniae* in Tryptic soy broth (1.7 % tryptone, 0.3 % soytone, 0.5 % NaCl, 0.25 % K_2HPO_4, 0.25 % glucose; pH 7.3; microaerophilic growth conditions) were diluted in the growth medium to achieve a final inoculum of ca. 10^6 cfu/ml. Chlorotonil and its derivatives were prepared as $CHCl_3$ stocks (10 mg/ml). Serial dilutions of the compounds in THF/Cremophor EL 1:1 were prepared from the 10 mg/ml stock solutions in sterile 96-well plates. The cell suspension was added and microorganisms were grown for 16-24 h at 37 °C. Growth inhibition was assessed in two independent experiments by visual inspection and given MIC (minimum inhibitory concentration) values are the lowest concentration of antibiotic at which no visible growth was observed.

4.3.4 MIC activity of semi-synthetic chlorotonil derivatives against *Plasmodium falciparum* strains

The MIC values for the two epochlorotonil derivatives and the dihydro-chlorotonil derivative against *Plasmodium falciparum* were determined as previously described by Held et al..

4.3.5 Solubility testing of epochlorotonil A and B

To assay the solubility of epochlorotonil A, B and dihydrochlorotnil in comparison to chlorotonil A, 3.3 mg of each compound were dissolved in 333 µl ml methanol at room temperature with ultrasonication in an ultrasonic bath for 10 minutes. The samples were subsequently centrifuged at maximum speed (15 000 rpm) from 20 minutes at room temperature. A 300 µl aliquot of each sample was taken and transferred to a pre-weighed glass vial. Methanol was removed under high vacuum and the vials were subsequently weighed out. For better comparability the obtained results were calculated back to a total volume of 1 ml by multiplication with the factor 3.3.

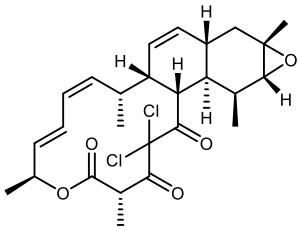
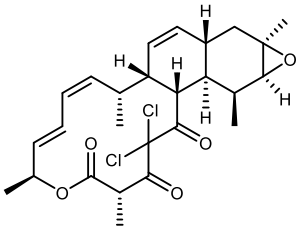
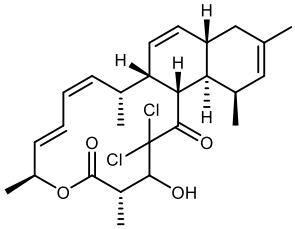
4.4 Results

4.4.1 Production and structure elucidation of semi-synthetic chlorotonil derivatives

Looking at the structure of chlorotonil A and keeping in mind that the *gem*-dichloro-1,3-dione substructure might be the pharmacophore of the compound, the isolated double bond at C-9/C-10 seemed to be an appropriate target position for derivatisation. In a very straightforward approach we thus succeeded in derivatising chlorotonil A with MCPBA and obtained a reasonable amount of epoxide derivatives. HPLC-MS analysis of the epoxidation reaction product showed a double-peak with a mass of $[M+H]^+$ = 495.162 m/z in agreement with the predicted mass for mono-epoxide derivatives of chlorotonil A. Furthermore a smaller peak groups, most likely presenting di-epoxide derivatives of chlorotonil A, exhibiting a mass of $[M+H]^+$ = 511.175 m/z could be observed. During semi-preparative purification of the reaction mixture, the double-peak with $[M+H]^+$ = 495.162 m/z split into two peaks at retention times of 66.5 min and 68.0 min in the gradient, which were separately collected. After solvent evaporation and drying the two compounds under high vacuum conditions, NMR analysis of the compounds was carried out. The absolute configuration of the epoxide residues was deduced from ROE correlations. 2-D NMR analysis revealed the two compounds as diastereomers, whereas epochlorotonil A (retention time: 66.5 min), represents the minor product and epochlorotonil B (retention time: 68.0 min) represents the major product. Looking at the 3-D models of the two structures, it seems obvious that sterical hindrance caused by the methyl group at C-8 of epochlorotonil A leads to the fact that epochlorotonil A is the minor product resulting from the Prileschajew epoxidation. The overall yield of the reaction was found to be 38.5 % for the mono-epoxides whereas the minor product's yield was 9.5 % and the major product's yield was determined as 29.0 %.

In order to corroborate or disprove our pharmacophore hypothesis, also a derivative altered in the alleged pharmacophore substructure was produced. Using dimethylaminoboran, the carbonyl group at C3 of chlorotonil A was selectively reduced to the respective alcohol. The overall yield of the reaction was 29 % of dihydro-chlorotonil.

Table 6: NMR chemical shifts of chlorotonil derivatives in in $CDCl_3$

	epochlorotonil A			**epochlorotonil A**			**dihydro-chlorotonil**		
									
s.	δ_C	M	δ_H	δ_C	m	δ_H	δ_C	m	δ_H
1	167.6	C		168.0	C		171.16	C	
2	47.2	CH	4.51	47.3	CH	4.52	41.63	CH	3.51
3	191.6	C		192.3	C		81.55	CH	4.69
4							88.73	C	
5	195.7	C		197.3	C		202.79	C	
6	49.4	CH	3.69	48.8	CH	3.65	49.53	CH	3.64
7	31.6	CH	2.16	37.0	CH	1.95	37.31	CH	2.17
8	29.1	CH	2.43	28.1	CH	2.33	30.47	CH	2.26[b]
9	65.6	CH	2.82	64.8	CH	3.04	128.58	CH	5.43
10	57.0	C		59.5	C		131.86	C	
11a	36.6	CH_2	1.93	39.3	CH_2	2.09	38.76	CH_2	2.09
11b			1.58			1.39			1.76
12	28.6	CH	1.98	28.0	CH	2.04	29.43	CH	2.27[b]
13	132.3	CH	5.55	132.7	CH	5.59	132.56	CH	5.75[c]
14	124.3	CH	5.47	123.9	CH	5.45	123.41	CH	5.53
15	42.7	CH	2.98	42.4	CH	2.98	44.82	CH	3.34
16	33.4	CH	2.74	33.3	CH	2.75	33.63	CH	2.84
17	139.3	CH	5.26	139.2	CH	5.24	138.06	CH	5.37
18	125.7	CH	5.85	125.7	CH	5.86	125.93	CH	5.92
19	124.2	CH	6.02	124.0	CH	5.99	127.73	CH	7.24
20	130.6	CH	5.49	130.6	CH	5.49	130.15	CH	5.51
21	70.5	CH	5.57	70.5	CH	5.59	69.63	CH	5.75[c]
22	21.1	CH_3	1.30	21.2	CH_3	1.30	21.21	CH_3	1.34
23	17.1	CH_3	1.64	17.3	CH_3	1.64	18.03	CH_3	1.47
24	10.5	CH_3	0.90	9.8	CH_3	0.92	14.94	CH_3	0.94
25	24.9	CH_3	1.33	23.1	CH_3	1.34	23.16	CH_3	1.67
26	15.9	CH_3	0.91	15.8	CH_3	0.92	15.36	CH_3	0.86

4.4.2 Antimicrobial activity of semi-synthetic chlorotonil derivatives

Table 7: Antibacterial activity of chlorotonil derivatives in comparison to chlorotonil A

MIC [µg/ml]	Chlorotonil A		Epochlorotonil A		Epochlorotonil B		Dihydro-chlorotonil	
	(µg/ml)	µM	(µg/ml)	µM	(µg/ml)	µM	(µg/ml)	µM
E. coli DSM-1116	> 32	> 64	> 32	> 64	> 32	> 64	> 32	> 64
E. coli TolC	> 32	> 64	> 32	> 64	> 32	> 64	> 32	> 64
B. subtilis DSM-10	0.03	0.06	0.125	0.25	0.06	0.12	> 32	> 64
C. glutamicum DSM-20300	0.06	0.12	0.5	1.0	0.25	0.5	> 32	> 64
E. faecalis ATCC29212	8	16	4	8	> 32	> 64	> 32	> 64
M. luteus DSM-1790	0.06-0.125	0.12-0.25	0.5-1	1.0-2.0	0.25	0.5	> 32	> 64
S. aureus Newman	0.03	0.06	0.125	0.25	0.06	0.12	> 32	> 64
S. pneumoniae DSM-20566	0.06	0.12	0.5	1.0	0.25-0.5	0.5 – 1.0	> 32	> 64

The antimicrobial activity of chlorotonil derivatives was assayed in comparison to chlorotonil A using the same indicator strain panel as previously described. It could be shown that the major product epochlorotonil B is half as active as the chlorotonil reference whilst the minor product epochlorotonil A was found to be half as active as epochlorotnil B and thus exhibits only about 25 % of the activity level observed for chlorotonil A. The only outlier in this test series is the activity against *Enterococci* which was found to be the highest for epochlorotonil A. The reduced chlorotonil variant dihydro-chlorotonil was found to be inactive in all cases.

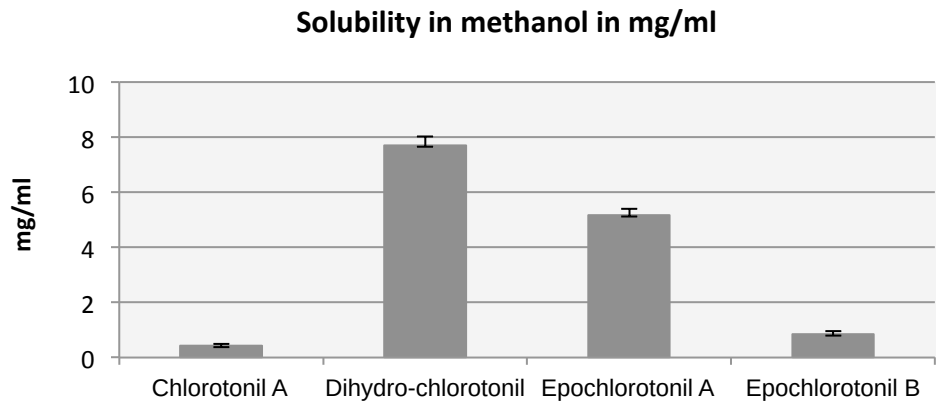
Table 8: Antiplasmodial activity of chlorotonil derivatives in comparison to chlorotonil A

	Chlorotonil A IC50		Epochlorotonil A IC50		Epochlorotonil B IC50		Dihydro-chlorotonil IC50	
	(µg/ml)	µM	(µg/ml)	µM	(µg/ml)	µM	(µg/ml)	µM
P. falciparum 3D7	0.011	0.023	1.0/2.8	3.8/5.8	0.05/0.08	0.1/0.17	10-50	20-100

Similarly all three derivatives were tested in an anti-plasmodial screening in the institute of tropical medicine in Tübingen. The results obtained in this assay were consistent with those from the antibacterial testing. The major derivative epochlorotonil B, with an MIC of 0.2 µM showed a 5- to 10-fold reduced activity compared to chlorotonil A whereas the activity of epochlorotonil A was 200-fold reduced. With an MIC greater than 20 µM, the dihydro-derivative of chlorotonil A practically did not show activity against *Plasmodium falciparum* anymore.

4.4.3 Effects of semi-synthetic derivatisation of chlorotonil A on the solubility of the compound

Solubility testing of the chlorotonil derivatives in comparison to the parent compound chlorotonil A provided conclusive results. First of all it could be shown, that derivatistion led to an enhanced solubility in polar solvents in all three cases. The biggest contribution to the hydrophilicity of chlorotonil was provided by reduction of the C2-carbonyl to the respective alcohol. This derivative could be dissolved up to a final concentration of 7.7 mg/ml in methanol.



For the two epoxide derivatives epochlorotonil A and B, the obtained results are quite different. Epochlorotonil A was found to be better soluble in methanol with a final concentration of up to 5.2 mg/ml whereas epochlorotonil B could only be dissolved to a final concentration of 0.85 mg/ml. These results indicate that in the two epoxide diastereomers, the orientation of the epoxide moiety highly impacts the accessibility for polar solvents.

The apparent interdependency of the hydrophobicity of the various derivatives and their observed bioactivity is particularly striking. In this small test setting, the best soluble compound is the least active. The results indicate that apart from the *gem*-dichloro-1,3-dione pharmcophore, also the hydrophobicity of the molecules plays a role for their bioactivity.

4.5 Discussion

Chlorotonil A is a promising new lead structure with outstanding activity against the malaria parasite *Plasmodium falciparum* as well as good activity against Gram-negative bacteria. In the presented study we addressed the solubility issues of chlorotonil A in polar solvents and also gained further insights concerning the pharmacophore of the molecule. Along these lines we produced three new semi-synthetic chlorotonil derivatives, two diastereomeric epoxide derivatives and a reduced chlorotonil derivative. These alterations in the molecule contributed to the hydrophilicity of the compound as in comparison to chlorotonil A (0.43 mg/ml), the methanol solubility of epochlorotonil A, epochlorotonil B and dihydro-chlorotonil could be improved.

As anticipated from earlier studies in which the bioactivity of natural chlorotonil congeners was assayed as well, the *gem*-dichloro-1,3-dione substructure was considered as pharmacophore of the molecule. This is due to the dramatic loss of bioactivity observed for natural chlorotonil derivatives with alterations in the *gem*-dichloro-1,3-dione pattern. To corroborate this assumption, the C3- carbonyl group in chlorotonil was reduced to the respective alcohol yielding dihydro-chlorotonil. Bioactivity assays were conducted with all three derivatives and chlorotonil A as a reference. By this means we could show that derivatisation of chlorotonil at the alleged pharmacophore substructure – as realized in dihydro-chlorotonil – indeed leads to complete abolishment of bioactivity. At the same time the epoxide derivatives retained their bioactivity even though it was reduced to a 50 – 25 % of the values determined for chlorotonil A. Moreover, the solubility issues in polar solvents could be successfully addressed as all derivatives showed enhanced solubility in methanol compared to chlorotonil A. Epochlorotonil B was found to be twice as soluble as chlorotonil A, whereas epochlorotonil A and dihydro-chlorotonil showed enhanced solubility by factor 12 and 18, respectively. Additionally the results from the solubility testing also provided further insights into the molecular features responsible for the bioactivity of chlorotonil A. The collected data provides evidence for an interdependence of hydrophobicity and bioactivity of the compounds, suggesting that part of the bioactivity is based on hydrophobic interactions.

In this study we could demonstrate, that chlorotonil A is accessible to semi-synthetic approaches, which opens up the opportunity to modulate pharmacokinetics of the compound by chemical optimization. Furthermore our study supports the pharmacophore

hypothesis for chlorotonil as we observed that the position of modification has a major impact on the bioactivity of the derivatives. Alterations in the *gem*-dichloro-1,3-dione substructure of chlorotonil A have so far always lead to either a significant reduction or total abolishment of bioactivity. Moreover we could verify the effects of the derivatisation on the hydrophilicity of the respective derivatives by probing their methanol solubility. In addition, this experiment provided more details about the molecular features needed for bioactivity as a connection between hydrophobicity and bioactivity could be observed.

Summarizing the above, this work paved the way for further focused medicinal-chemistry studies on chlorotonil A. Two important parameters which are most likely crucial for the bioactivity could be identified: On the one hand the *gem*-dichloro-1,3-dione substructure was – as already anticipated previously (see chapter 2) – found to play a major role. On the other hand also the hydrophobicity seems to influence the bioactivity. Once the molecular target of chlorotonil can be identified (work in progress), a more targeted approach can be used to produce further variants of chlorotonil A exhibiting better solubility whilst maintaining their bioactivity.

4.6 Supporting Information

4.6.1 HPLC-MS analysis of the crude reaction mixture

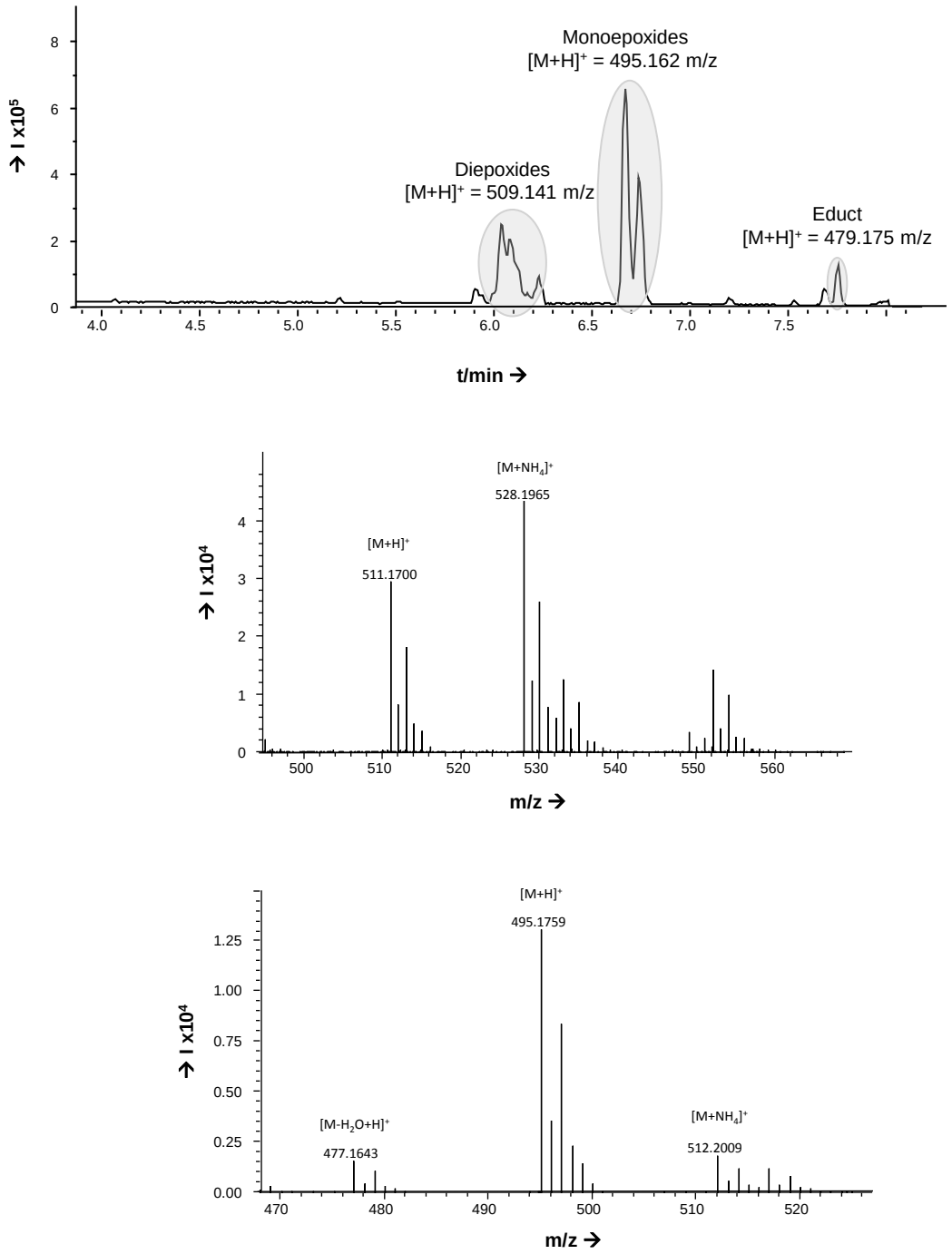


Figure 1 HPLC-MS traces of the reaction mixture after Prilezhaev epoxidation. Isotope patterns for mono- and di-epoxide derivatives of chlorotonil A are shown.

4.6.2 UV traces recorded during separation and purification of the two epoxide derivatives epochlorotonil A and B

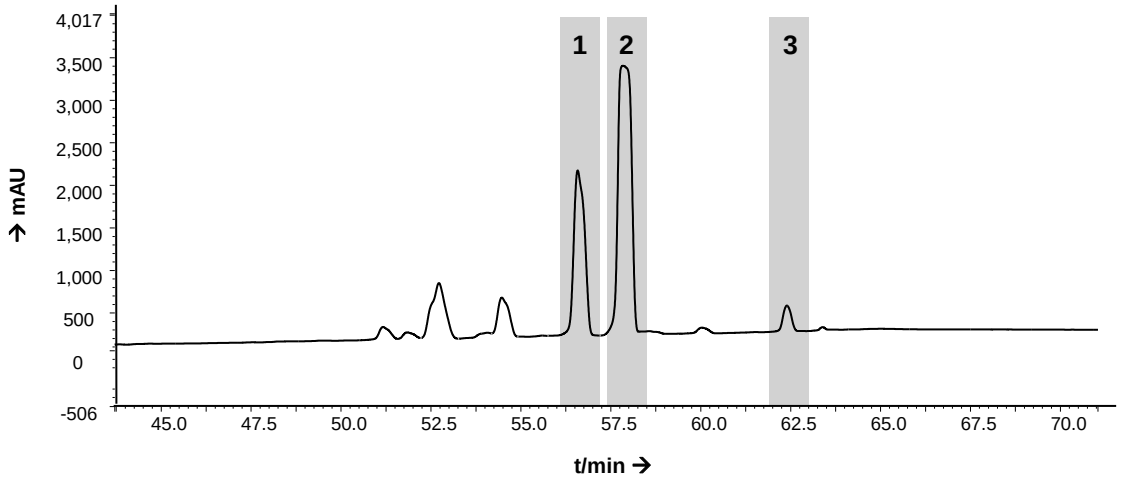


Figure 2: UV trace rcorded at 210 nm during semi-preparative purification of epochlorotonil A and B. Peak 1 represents epochlorotonil A, peak 2 represents epochlorotonil B and peak three is the residual educt chlorotonil A. all three peaks were collected throughout several runs.

4.6.3 Solubility testing of chlorotonil derivatives

Table 9: Determination of the methanol solubility of chlorotonil A and the semi-synthetic derivatives.

Vial No.	compound	TARA vial	Weighing result	Difference (weighing result – TARA)	Result * 3.3	Mean	Standard deviation
1	chlorotonil A	2592.712	2592.855	0.143	0.4719	0.4323	0.05368631
2	chlorotonil A	2606.751	2606.893	0.142	0.4686		
3	chlorotonil A	2581.707	2581.815	0.108	0.3564		
4	dihydro-chlorotonil	2588.881	2591.098	2.217	7.3161	7.7055	0.32067217
5	dihydro-chlorotonil	2553.399	2555.732	2.333	7.6989		
6	dihydro-chlorotonil	2586.383	2588.838	2.455	8.1015		
7	epochlorotonil A	2574.498	2576.027	1.529	5.0457	5.1733	0.21395495
8	epochlorotonil A	2559.208	2560.867	1.659	5.4747		
9	epochlorotonil A	2647.481	2648.996	1.515	4.9995		
10	epochlorotonil B	2633.784	2634.003	0.219	0.7227	0.8459	0.10387868
11	epochlorotonil B	2620.979	2621.275	0.296	0.9768		
12	epochlorotonil B	2631.995	2632.249	0.254	0.8382		

4.6.4 Structure elucidation of chlorotonil A and B

A. Epochlorotonil A

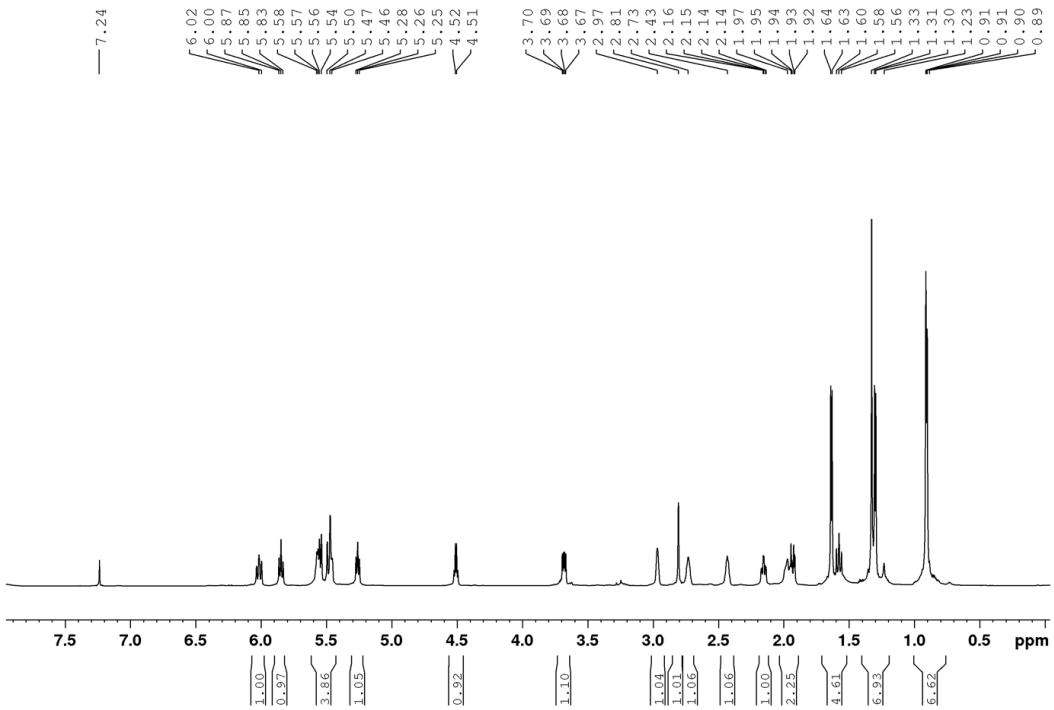


Figure 3: ^{1}H NMR spectrum of epochlorotonil A in $CDCl_3$ at 700 MHz

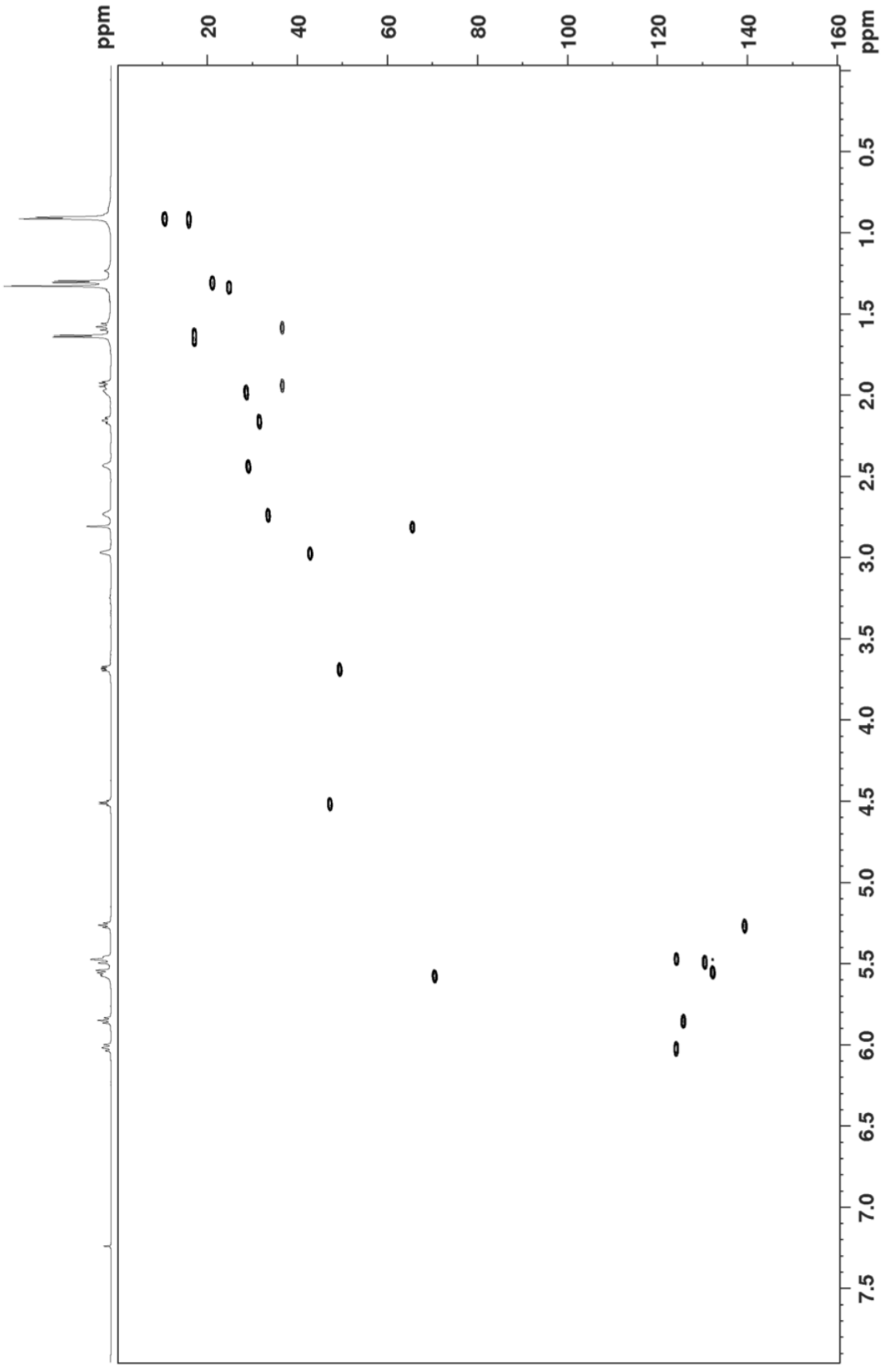


Figure 4: HSQC spectrum of epochlorotonil A in in $CDCl_3$ at 700 MHz

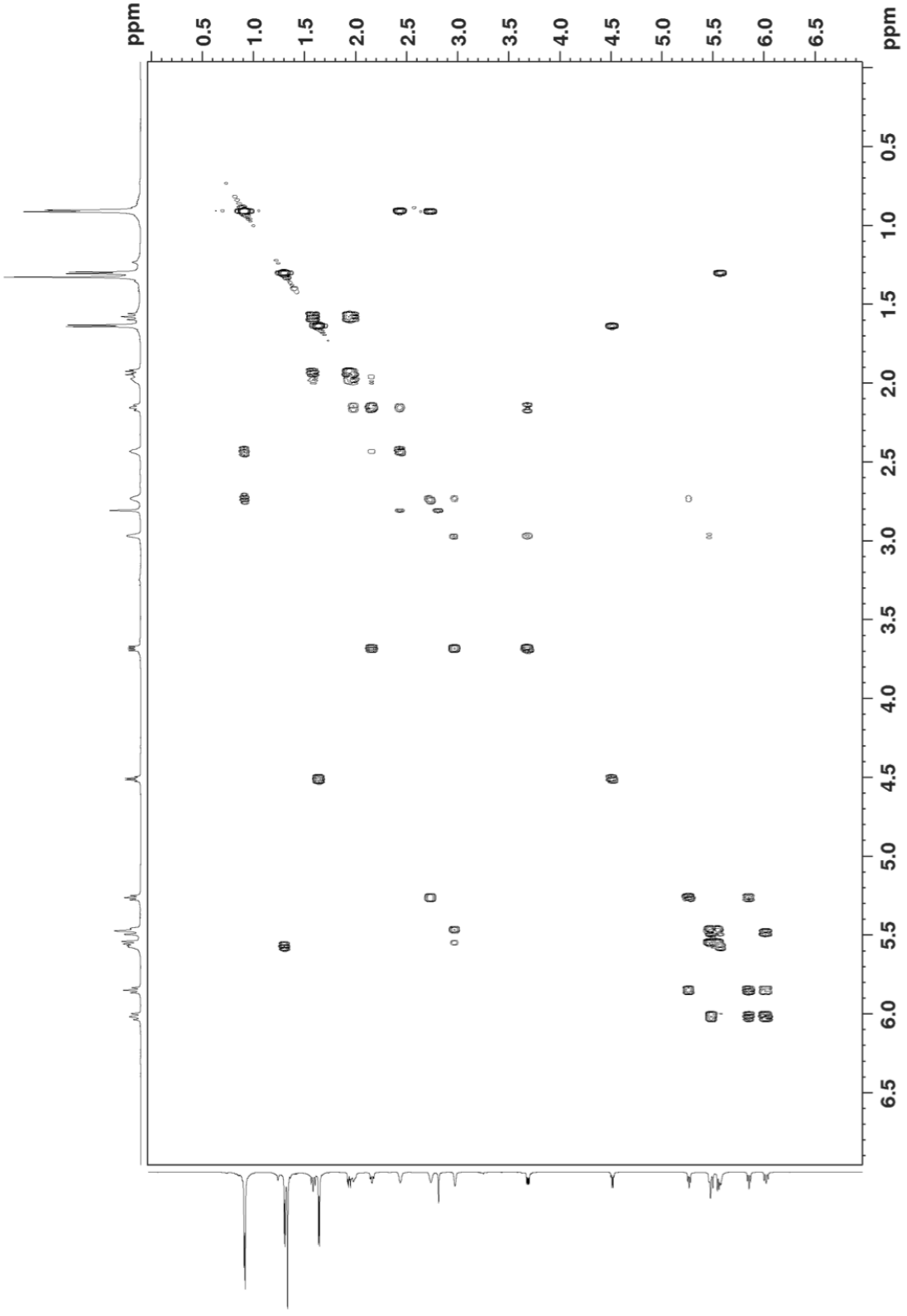


Figure 5: COSY spectrum of epochlorotonil A in in $CDCl_3$ at 700 MHz

B. Epochlorotonil B

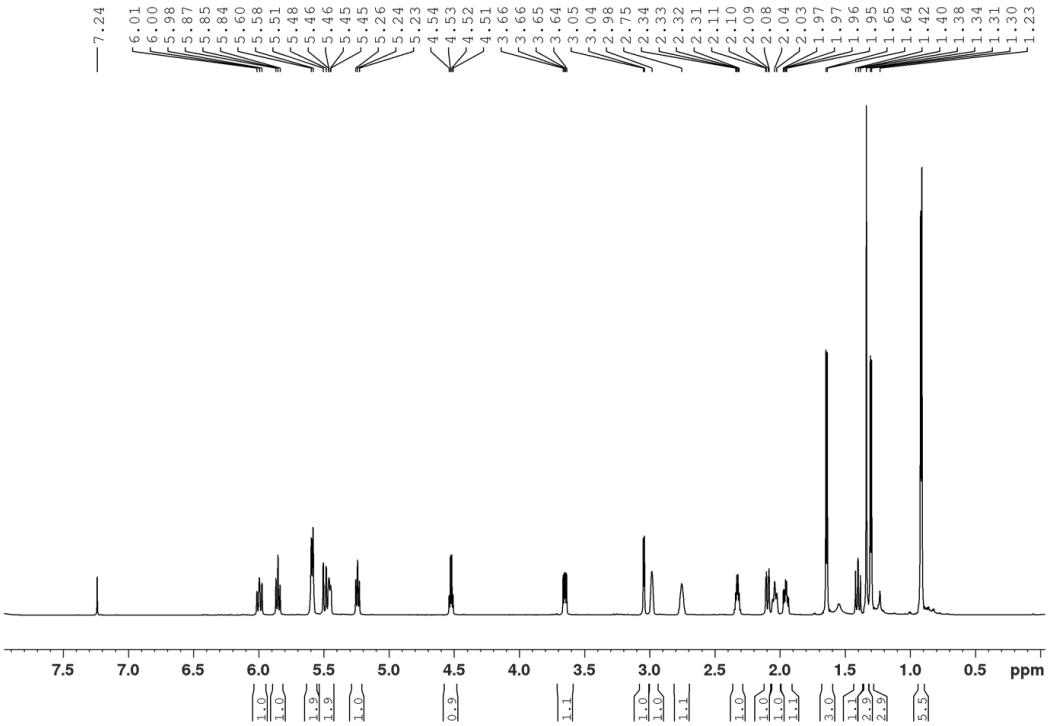


Figure 6: ^{1}H NMR spectrum of epochlorotonil B in $CDCl_3$ at 700 MHz

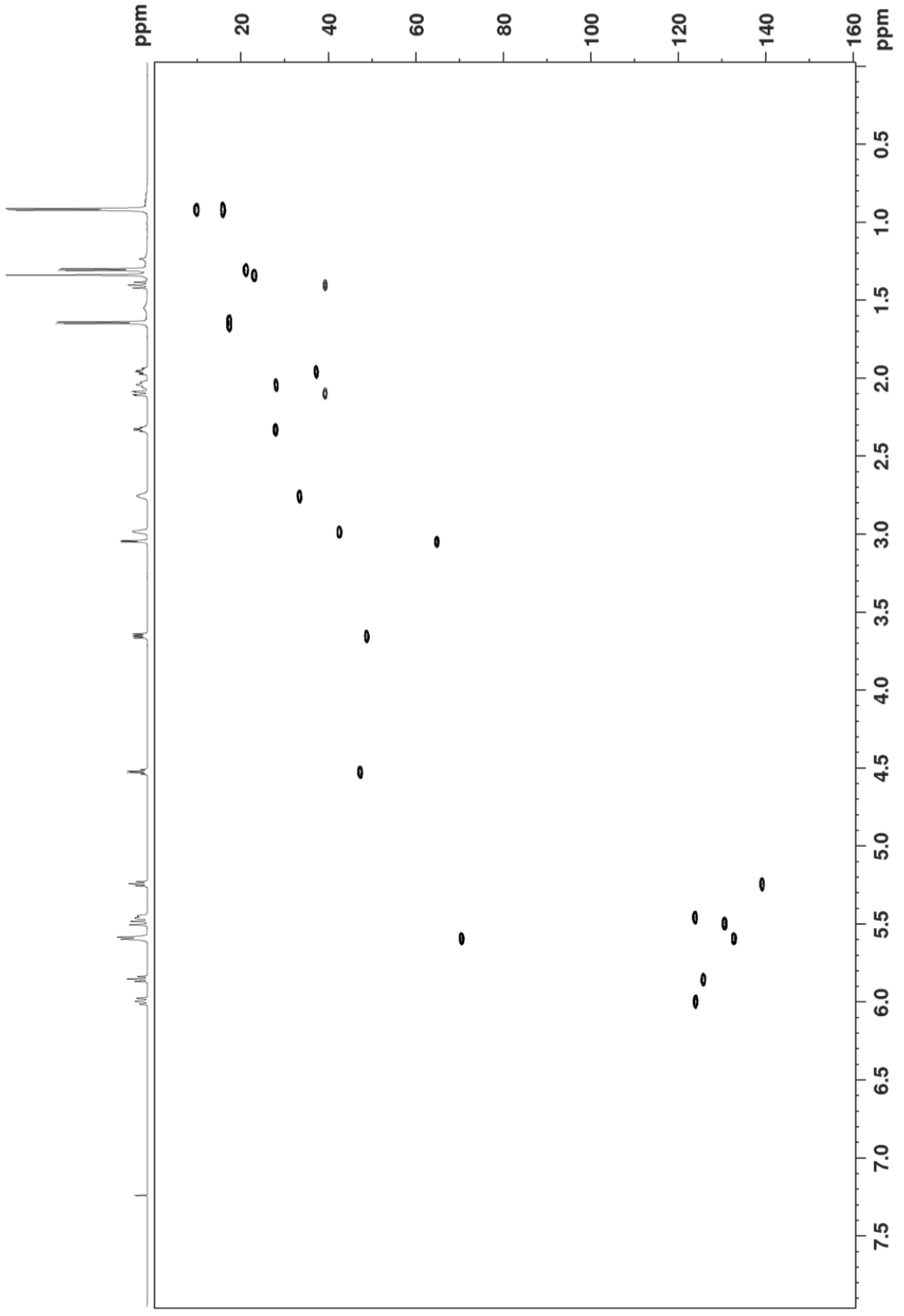


Figure 7: HSQC spectrum of epochlorotonil B in in $CDCl_3$ at 700 MHz

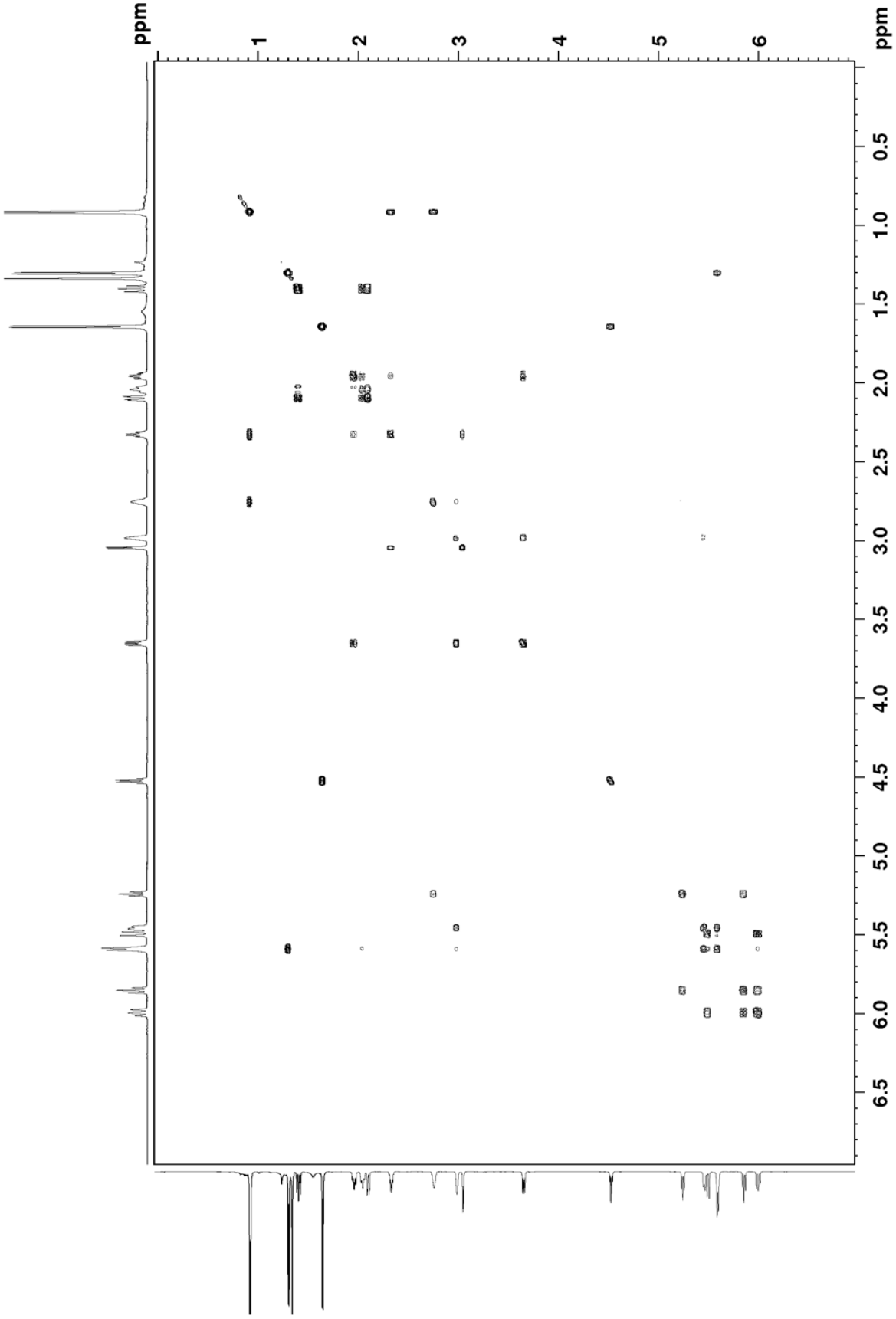


Figure 8: COSY spectrum of epochlorotonil B in in $CDCl_3$ at 700 MHz

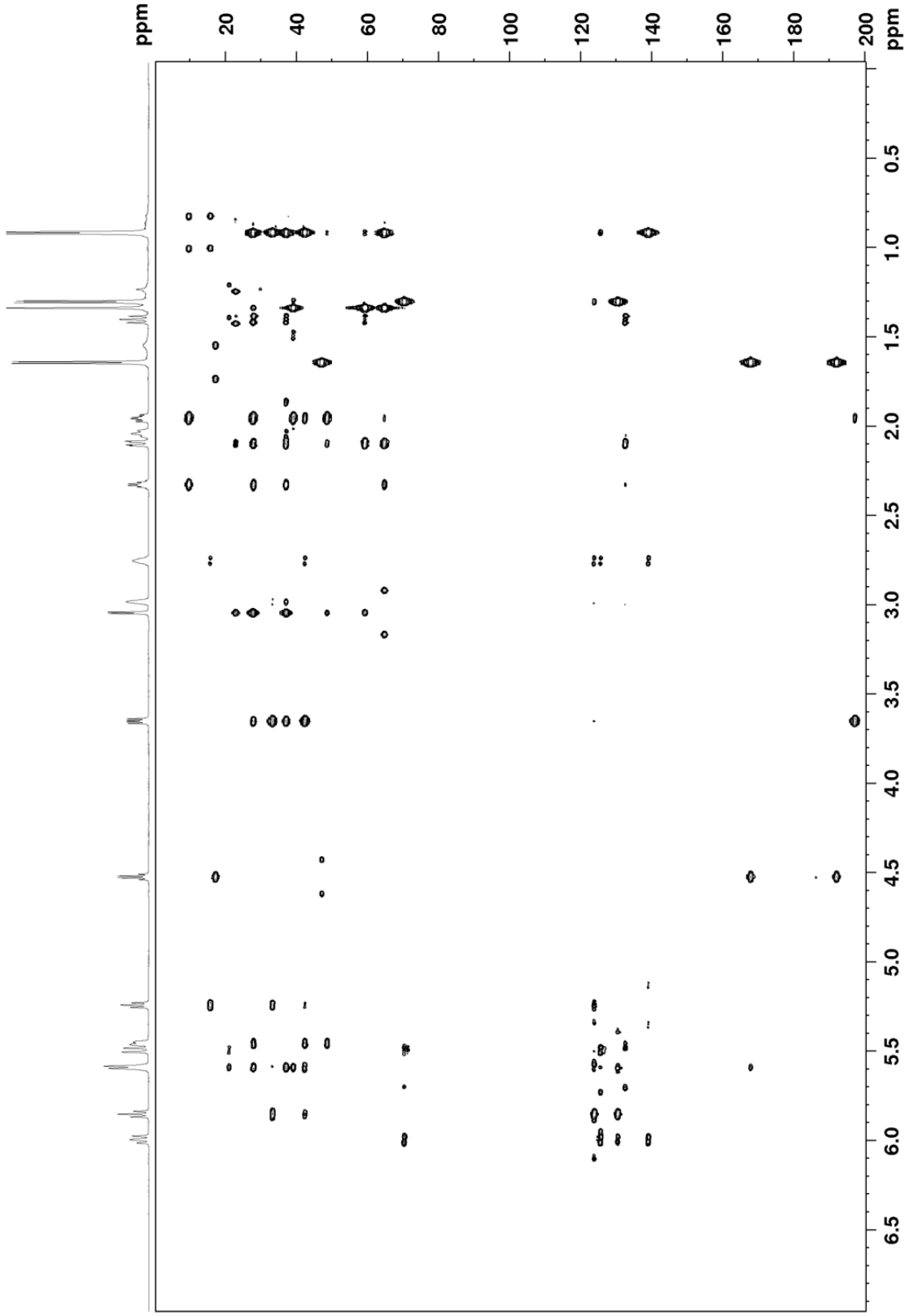


Figure 9: HMBC spectrum of epochlorotonil B in in $CDCl_3$ at 700 MHz

C. Dihydro-chlorotonil A

Dihydro-chlorotonil A: $C_{26}H_{34}Cl_2O_4$, M = 481.45; $[\alpha]^{22}_D$ = + 146.2 (*c* = 0.4 in ethanol); ^1H and ^{13}C NMR data table #; UV (EtOH): λ_{max} (lg ε) = 235 (4.379) nm; HR-ESIMS: $[M+H]^+$ *m/z* 481.1918, calcd 481.1907 for $C_{26}H_{35}Cl_2O_4$; $[M+Na]^+$ *m/z* 503.1736, calcd 503.1726 for $C_{26}H_{34}Cl_2NaO_4$; $[2M+Na]^+$ *m/z* 961.3752, calcd 961.3741 for $C_{52}H_{69}Cl_4NaO_8$.

Table 10. NMR Data of Dihydro-chlorotonil A (#) in CDCl3a

Pos.	δ_C	m	δ_H	H Multiplicity (*J* [Hz])	COSY	ROESY	H in HMBC
1	171.16	C					23 >2, 3 >21
2	41.63	CH	3.51	qd (7.2, 1.9)	23, 3	(23) >3	23 >OH, 3
3	81.55	CH	4.69	dd (6.2, 1.9)	OH, 2	OH, 23 >2	23 >OH, 2
4	88.73	C					2 > 3
5	202.79	C					6 >>7
6	49.53	CH	3.64	dd (12.4, 7.2)	7, 15	12/8, 16, 24 >7	7, 14 >16
7	37.31	CH	2.17	ddd (12.1, 10.9, 4.1)	12, 6	26 >11b, >(6)	24, >11ab, 13, 9 >8
8	30.47	CH	2.26[b]	dd (4.8, 2.5)	24, 25, 9	24, 9	24 >7, 9 >6
9	128.58	CH	5.43	d (5.4)	25, 11b, 8	25, 8 >24	24, 25 >11ab, 8
10	131.86	C					25 >11ab >12
11a	38.76	CH_2	2.09	dd (17.1, 5.1)	25, 11b, 12	(11b) >25, 12/8, 13	25 >7, 9, 13
11b			1.76	m (16.3, 12.5)	25, 11a, 12, 9	(11a) >7 >13	---
12	29.43	CH	2.27[b]	m	11ab, 7, 24,	24, 11a, 13, 6	11ab, 14, 13,7, 6
13	132.56	CH	5.75[c]	dt (10.1, 1.9)	15, 14	12, 11a	11ab
14	123.41	CH	5.53	ddd (10.1, 4.5, 2.4)	12, 15, 13/21	17, 15 >26	16
15	44.82	CH	3.34	m (7.1, 4.7, 2.4)	16, 6, 13/21, 14	6, 16, 14 >19 >17	26 >6, 13 >14, 16, 7, 18
16	33.63	CH	2.84	m	26, 15, 17	19 >OH, (15, 26) >17	26, 6, 18 >17
17	138.06	CH	5.37	m (10.2, 9.6, 0.6)	16, 18	(18), 14, 26 >> 15, 16	26 >16, 19
18	125.93	CH	5.92	t (11.0)	17, 19	17, 20	20, 16, 19
19	127.73	CH	7.24	dddd (15.1, 11.8, 2.2, 0.9)	20, 18	16 >15, OH, 21	17 >18, 21
20	130.15	CH	5.51	m (15.3, 2.2)	19	18, (21), 22	22 >18, 21 >19
21	69.63	CH	5.75[c]	m	22	(20, 22) >19	22 >20, 19
22	21.21	CH_3	1.34	d (6.9)	13/21	(21), 20	20, 21
23	18.03	CH_3	1.47	d (7.3)	2	(2) >3 >OH	2, 3
24	14.94	CH_3	0.94	d (7.1)	8, 12	12/(8) 6, 9 >7	7, 8, 11ab, 25
25	23.16	CH_3	1.67	s	11ab, 8, 9	9, 11a	11ab, 9
26	15.36	CH_3	0.86	d (6.5)	16	7 >16, 17 >8/12, 14, OH	17 >16, 15
OH			3.42	d (6.0)	3	(3), 16 >19 >23, 26	

[a] ^1H 700.3, ^{13}C 176.1 MHz; [b, c] overlapping signals;

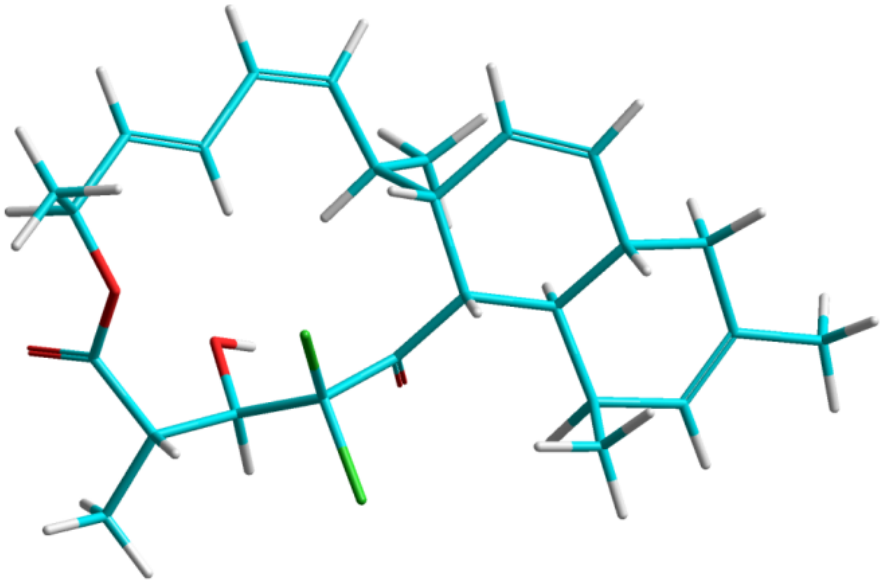


Figure 10: Calculated 3D structure of dihydro-chlorotonil A (#) with *R* configuration at C-3

4.7 References

[1] K. H. Jang, S.-J. Nam, J. B. Locke, C. a Kauffman, D. S. Beatty, L. a Paul, W. Fenical, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7822–4.

[2] M. R. Green, J. Sambrook, *Molecular Cloning : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, **n.d.**

[3] M. Kopp, H. Irschik, F. Gross, O. Perlova, A. Sandmann, K. Gerth, R. Müller, *J. Biotechnol.* **2004**, *107*, 29–40.

[4] D. A. H. Tobias Kieser, Mervyn J. Bibb, Mark J. Buttner, Keith F. Chater, *Practical Streptomyces Genetics*, Norwich, **2000**.

[5] J. Herrmann, Y. a. Elnakady, R. M. Wiedmann, A. Ullrich, M. Rohde, U. Kazmaier, A. M. Vollmar, R. Müller, *PLoS One* **2012**, *7*, 1–12.

[6] S. C. Findlow, C. Winsor, T. J. Simpson, J. Crosby, M. P. Crump, **2003**, 8423–8433.

[7] A. Witkowski, A. K. Joshi, Y. Lindqvist, S. Smith, *Biochemistry* **1999**, *38*, 11643–11650.

[8] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[9] P. Caffrey, *ChemBioChem* **2003**, *4*, 654–657.

[10] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure* **2010**, *18*, 94–105.

[11] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–33.

Chapter 5

Towards heterologous expression of the maracen biosynthetic pathway assembled from artificial DNA segments

Contributions and Acknowledgements

A. Author's efforts

The author designed and performed most of the experiments described and evaluated and interpreted the resulting data. Analysis and annotation of the maracen/maracin biosynthetic gene cluster as well as completion of the sequence information was done by the author. The author designed an artificial maracen/maracin biosynthetic gene cluster and the required cloning vectors together with Silke Wenzel and in collaboration with ATG:Biosynthetics GmbH. All cloning and transformation steps involved in assembly of the artificial gene cluster were conducted by the author. The author created the mutants described in the chapter; similarly the author performed genotypic and phenotypic analysis thereof. The author conceived and wrote the chapter.

B. Contributions by others

Katja Gemperlein provided a first biosynthesis hypothesis for maracen and maracin, which served as working hypothesis throughout the project and provided a construct harbouring the *pfa2* PUFA gene cluster. Thomas Hoffmann initially evaluated the myxobacterial strain collection for additional maracen/maracin producers. The design of an artificial pathway was performed in collaboration with ATG:Biosynthetics GmbH. Scientific supervision of the project as well as proofreading of the chapter was done by Silke Wenzel and Rolf Müller.

Furthermore, the author would like to thank Thomas Hoffmann and Michael Hoffmann for HPLC-MS measurements and Isabell Hofmann for her support during assembly of the artificial maracen/maracin biosynthetic gene cluster.

5 Towards heterologous expression of the maracen biosynthetic pathway assembled from artificial DNA segments

5.1 Abstract

Natural products isolated from plants, fungi and bacteria have increasingly gained importance in the last decades and continue to provide a valuable source for new pharmaceutical lead structures, as they often show new structural features and exhibit unknown modes of action. As total synthesis of natural products is often impaired by stereochemically demanding substructures and the overall complexity of the compounds, fermentation of the producer strain often is the only viable production route. Moreover the physicochemical properties of natural products often have to be optimized prior to their application. Thus the production of natural product derivatives with enhanced pharmacokinetics by medicinal chemistry or biosynthetic engineering is a major task in the development of natural products towards clinical use. For the latter approach a detailed understanding of the biosynthetic machinery in the producing host is of great importance. But knowledge of the biosynthetic gene cluster responsible for production of the respective metabolite is often not sufficient as regulation thereof in the producer strain might lead to low production levels. Furthermore, the production of derivatives by alterations of the gene cluster is most of the time not straightforward since genetic tools for new genera and species often have to be established arduously; the same is true for in-depth and detailed investigation of peculiar steps during biosynthesis. Consequently heterologous expression of intriguing gene clusters in more accessible host strains represents an approved strategy in natural product research. However, this approach has its limitations since assembly of gene clusters often depends on previous isolation of the sought DNA sequence from the host strain which also requires the establishment of various micro- and molecular biology techniques as well as a reasonable cultivability of the producer strains. These issues can be addressed by synthetic biology which provides the opportunity to piece together complete biosynthetic pathways independent of the producer strain's manageability. In this study we report design and reassembly of the putative maracen/maracin biosynthetic pathway from a *Sorangium sp.* for expression in the well-established heterologous host strain *Myxococcus xanthus* DK1622. Maracen and

maracin represent two unusual myxobacterial secondary metabolites with structural similarity to polyunsaturated fatty acids. Following an adept cloning strategy the gene cluster was assembled from synthetic DNA fragments and subsequently adapted for integration into the *Myxococcus* genome. By this approach we expect to prove the responsibility of the assembled genes for maracen/maracin biosynthesis, whilst the flexible cloning strategy allows for deeper analysis of the biosynthetic pathway by selective removal of individual genes from the gene cluster.

5.2 Introduction

As reflected from the number of newly approved drugs in the last years and their origin, natural products are still highly important in the search for new pharmaceuticals.[1] Secondary metabolites from plants, fungi and bacteria are commonly considered privileged structures as they were optimized by evolutionary processes and often specifically bind to yet under-exploited targets.[2] In this context, the relatively young research field of myxobacteria provides a rich source for the discovery of new natural products. Myxobacteria are soil-dwelling Gram-negative δ-proteobacteria which are similarly to actinomycetes very often isolated from competitive habitats such as soil or dung. With their typically large genomes these bacteria bear the potential for production of numerous secondary metabolites. Especially strains of the genus *Sorangium* are frequently revealed as prolific producers of bioactive natural products. The anti-malaria agent chlorotonil as well as the anticancer drug epothilone for example are produced by two *Sorangium cellulosum* strains[3,4]. The vast majority of natural products from myxobacteria characterized so far is produced by polyketide synthases (PKS), non-robosomal peptide synthetases (NRPS) and hybrids thereof. Another interesting substance class exhibiting a PKS/FAS (fatty acid synthase)-like structure are the maracenes and maracines whose structural peculiarities triggered the interest in their biosynthesis in.[5] However, the drawback of myxobacteria and especially strains of the genus *Sorangium* is that they are notoriously hard to cultivate with a low doubling rate and several special demands regarding the nutrient sources. Furthermore genetic manipulations can hardly be achieved which is also due to the fact that *Sorangium* strains do have natural resistance against almost all antibiotics in laboratory use. Considering all these points, biosynthetic gene clusters from *Sorangium* strains appear to be the perfect candidates for synthetic biology approaches towards establishment of heterologous expression platforms. It is only in the recent past that synthetic biology offered the possibility to bypass some of the

tedious steps towards establishment of platforms for heterologous expression. Synthesis of the DNA encoding the desired biosynthetic pathway allows for modifications and adaptions of the pathway and also enables more detailed investigation of biosynthetic processes. This has already exemplarily been achieved for the previously mentioned epothilone biosynthetic pathway which has been assembled from synthetic DNA fragments and successfully expressed in the host strain *Myxococcus xanthus*, as well as in an *E. coli* host system.[6,7] Another example just recently accomplished was the expression of the biosynthetic pathway for argyrin A[8], a cyclic peptide with promising anti-tumor activity produced by the myxobacterial strain *Archangium gephyra*. (S. Wenzel et al., unpublished) In the light of these results, also for investigations of maracen/maracin biosynthesis in *Sorangium cellulosum*, a synthetic biology approach was chosen. Maracen and maracin represent two secondary metabolites produced by several *Sorangium* strains, which structurally resemble polyunsaturated fatty acids.[5] Both compounds show a good *in vitro* activity against *Mycobacterium tuberculosis*, whilst maracen also exhibits additional antibacterial activity against a panel of Gram-positive indicator strains. Comparison of all maracen producer strains on the gene level allowed for rapid identification of a candidate gene cluster *in vitro*. The found pathway was then synthesized in smaller pieces and reassembled on a construct designed for integration into the *Myxococcus* host genome. Using this approach we aimed to prove our biosynthesis hypothesis and at the same time establish a heterologous expression platform for maracen/maracin. Moreover the whole construct was designed in a way that selective removal of individual genes from the gene cluster was possible. This was to enable studies on necessity of individual genes in the gene cluster for production of the secondary metabolites.

5.3 Results and discussion

5.3.1 Identification of the maracen biosynthetic gene cluster by combined analysis of genome and HPLC data

The ethinyl-vinylether maracin and the α-chloro-divinylether maracen are both reportedly produced by two representatives of the genus *Sorangium*, So ce1128 and So ce880, for one of which, So ce1128 genome data was available.[5] In order to identify a biosynthetic gene cluster responsible for production of the two compounds, the sequenced genome of So ce1128 was screened for putative candidate gene clusters of PKS type but no match could be found. A database search within our in-house database “myxobase” comprising among other things the metabolome data of several thousand myxobacterial strains revealed several other maracen/maracin producer strains for some of which genome data was available as well. Inferred from the structural peculiarities of the two compounds, a linkage to fatty acid biosynthesis seemed obvious and became the working hypothesis. By the time a new type of PUFA gene cluster has been discovered exclusively in *Sorangium* strains that also produce maracen and maracin, this was considered as further confirmation of our hypothesis (K. Gemperlein, unpublished). Based on these findings the available genome or cosmid data of five myxobacterial producer strains was examined carefully around the genomic locus of the newly identified EPA-type PUFA gene cluster. A genomic region spanning about 45 kb consistently composed of the same set of genes – including the PUFA synthase encoding genes - could be identified in the genome data of all five strains. Through heterologous expression of the two PUFA genes from *Sorangium sp.* So026 (MSr9366) complemented with the missing enoyl-reductase functionality in *Myxococcus* it could already be proven that the other surrounding genes were not required for PUFA biosynthesis (K. Gemperlein, unpublished). Hence a hypothesis for the production of maracen and maracin starting from eicosapentaenoic acid with the help of the functionalities encoded in close proximity to the PUFA gene cluster was elaborated.

5.3.2 Biosynthesis hypothesis for maracen/maracin

Based on *in silico* analysis of the 19 genes encoded up- and downstream of the EPA-producing *pfa* gene cluster found in *Sorangium sp.* So026 (MSr9366), a biosynthesis hypothesis for maracen and maracin was elaborated, which is summarized in Figure 4.

Table 2: Results of the BLAST analysis of the 19 genes flanking the *pfa* gene cluster in *Sorangium sp.* So026 (MSr9366)

Gene/protein	Length (bp/aa)	Proposed function			
		Proposed function of homologous protein	Source of the homologous protein	Identity/ similarity, %	Accession number (GenBank)
mrc0/Mrc0	1656/552	protein kinase	*Sorangium cellulosum*	83/85	WP_015351249
mrc1/Mrc1	960/320	delta(12)-fatty acid dehydrogenase	*Sorangium cellulosum*	39/60	WP_044969276
mrc2/Mrc2	678/226	hypothetical protein			
mrc3/Mrc3	903/301	ferritin, fatty acid desaturase	*Burkholderia sp.*	52/52	WP_007180466
mrc4/Mrc4	1056/352	hypothetical protein, α,β-hydrolase (DUF 2048)	*Myxococcus xanthus*	46/47	AAO22902
mrc5/Mrc5	1041/347	zinc-dependent alcohol dehydrogenase	*Streptomyces scabiei*	41/42	WP_037705470
mrc6/Mrc6	816/272	phospholipid/glycerol acyltransferase	*Anaeromyxobacter sp.*	43/43	ABS25954
mrc7/Mrc7	768/256	patatin-like phospholipase	*Stigmatellla aurantiaca*	32/33	WP_002618026
mrc8/Mrc8	801/267	enoyl-CoA hydratase	*Myxococcus fulvus*	56/57	WP_046711249
mrc9/Mrc9	2250/750	patatin-like phospholipase	*Stigmatellla aurantiaca*	27/28	EAU67493
mrc10/Mrc10	417/139	hypothetical protein	*Chondromyces apiculatus*	58/58	WP_044234628
mrc11/Mrc11	1023/341	hypothetical protein, α,β-hydrolase (DUF 2048)	*Myxococcus xanthus*	46/47	AAO22902
mrc12/Mrc12	2250/750	patatin-like phospholipase	*Stigmatellla aurantiaca*	29/29	EAU67493
mrc13/Mrc13	1431/477	cytochrome P450	*Cyanotheca sp.*	41/41	WP_012596349
mrc14/Mrc14	1644/548	lipoxygenase	*Cystobacter violaceus*	46/46	WP_043397944
mrc15/Mrc15	2538/846	lipase maturation factor 1	*Parachlamydia acanthamoebae*	26/26	KIA77111
mrc16/Mrc16	2250/750	cyclic nucleotide-binding/patatin-like phospholipase domain-containing protein	*Corallococcus coralloides*	34/34	AFE10789
mrc17/Mrc17	1341/447	aminooxidase, PUFA isomerase	*Propionibacterium acnes*	22/22	WP_002513679
mrc18/Mrc18	1392/464	aminooxidase, PUFA isomerase	*Propionibacterium acnes*	26/26	WP_002513679

Initially EPA is released from phospholipids catalysed by phospholipases. Interestingly, this functionality is encoded four times in the putative maracen biosynthetic gene cluster; the genes *mrc7, 9, 12 and 16* all code for patatin-like phospholipases, whereas *mrc9* and *mrc12* are almost identical (92.8 % identical sites). Patatin denominates a group of glycoproteins which also show phospholipase activity and were first identified in *Solanum tuberosum* species; up to now, several related patatin-like phospholipases have also been reported from eukarya and prokarya.[9,10] Overall, patatin-like phospholipases are characterized by a distinct architecture in their active site that constitutes a catalytic Ser-Asp dyad instead of the classical triad seen in other phospholipases. The observed

folding and the 3D structure in the catalytic site of those enzymes differs considerably from the topology of other types of lipases.[11] For some of the mammalian isoenzymes, a role in lipometabolism could be demonstrated, but the physiological function of other members still remains to be elucidated.[12] Similarly the lipase maturation factor encoded by *mrc15* might be involved in the EPA release from phospholipids as well, as this chaperon protein is known to play a key role in the post-translational activation of several lipases.[13]

The available EPA is then assumed to undergo an isomerization process to convert the 8-*cis* double bond to a 7-*trans* double bond. The maracen biosynthetic gene cluster encodes two putative isomerases, Mrc17 and Mrc18, that might be involved in the proposed double bond shift.[14,15] It is striking that also in this case the two genes are highly similar showing 80.3 % identity. In the next step the isomerization product is oxidized to yield a 6-hydroperoxide derivative most likely catalysed by Mrc14, which shows homology to lipoxygenases (LOX).[16] Representatives of this enzyme class are wide spread in plants and also the human isoenzymes are well studied as the play a major role in second messenger formation for example in the generation of arachidonic acid. Bacterial LOX are rarely reported, therefore it has been assumed that horizontal gene transfer is responsible for the presence of lipoxygenases in bacteria.[17] In the same transcription unit with *mrc14*, a putative cytochrome P450 enzyme (Mrc13) is encoded, which is most likely involved in the conversion of the 6-hydroperoxide fatty acid into the respective epoxy fatty acid as shown in Figure 1. This intermediate is highly unstable and assumed to be transformed into the corresponding divinyl-ether by the same P450 functionality, which would then act as a divinyl-ether-synthase also known from the plant lipoxygenase pathway.[18]

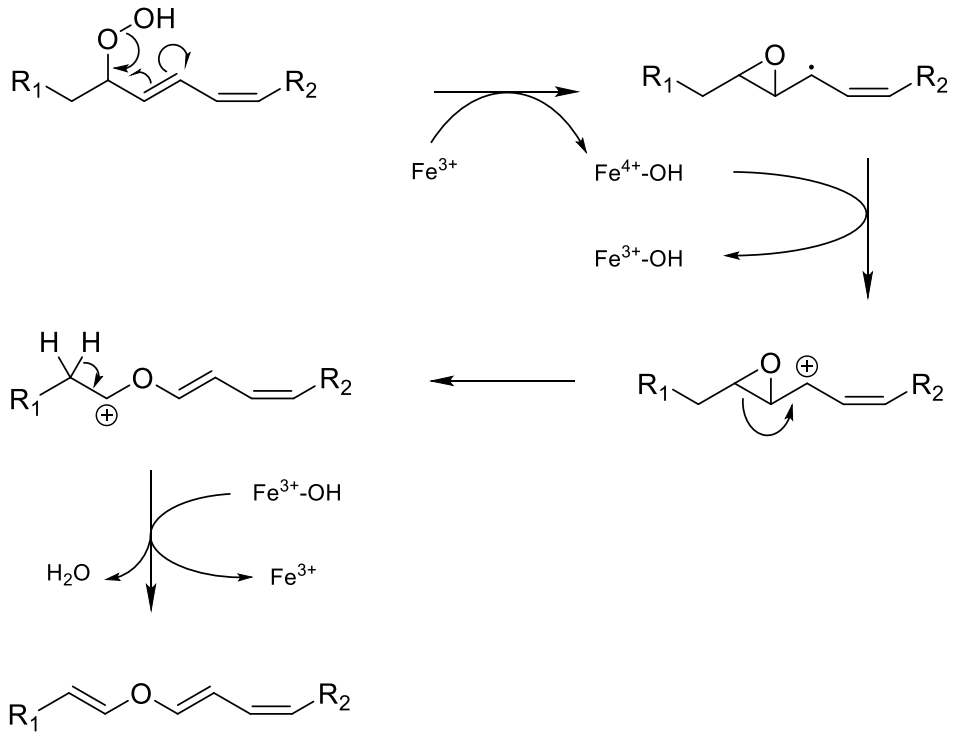


Figure 11: Proposed reaction mechanism for introduction of the divinyl-ether moiety in the maracen precursor molecule. Adapted from Grechkin, Prostaglandins & other Lipid Mediators 68–69 (2002) 457–470

In the next step, the thus generated C3-C4-*trans* double bond is shifted to C2 resulting in a C2-C3-*trans* double bond. The putative PUFA isomerases Mrc17 and Mrc18 represent possible candidates for catalyzing this reaction. Subsequently the C2-C3-*trans* double bond is oxidized resulting to the respective 2-hydroxy-fatty acid. This reaction is commonly seen for branched chain fatty acids, where a β-methyl-branch hinders access of the enoyl-CoA hydratase responsible for β-oxidation. In the present case, the fatty acid substrate is not methylated and an enoyl-CoA hydratase encoded by *mrc8* would be available to carry out β-oxidation. However, β-oxidation of the fatty acid substrate would lead to the removal of two carbons yielding a C18 chain, whilst maracen and maracin consist of a C19 carbon chain. Thus α-oxidation of the fatty-acid intermediate followed by elimination of one carbon is the most apparent scenario carried out in the subsequent two steps. According to this, the 2-hydroxy-group is reduced to the respective carbonyl moiety, most likely catalysed by the zinc-dependent alcohol dehydrogenase-homolog Mrc5, followed by elimination of the terminal $–CO_2$ moiety. The enzyme catalysing this cleavage step is still unknown, but two candidate genes, *mrc4* and *mrc11* encode hypothetical proteins that might be involved in this reaction. Nevertheless their function in maracen biosynthesis still remains to be elucidated. Next another isomerization step

takes place in which the two *cis* double bonds at C10-C11 and C13-C14 are converted into *trans* double bonds. Also for this isomerization steps the two PUFA isomerase encoding genes *mrc17* and *mrc18* come into consideration. Subsequently the terminal C19-C20 bond is desaturated yielding the respective double bond. For this reaction, taking place at the ω-1 position of the fatty acid intermediate, no evident candidate gene encodes the required enzymatic functionality. ω-3 and ω-6 fatty acid desaturase, introducing a double bond at the respective ω-3 and ω-6 position of fatty acids are well known from higher plants, but to the best of our knowledge, no description of an ω-1-fatty acid desaturase has been made.[19–21] However *mrc3* encodes a protein belonging to the ferritin-like superfamily in which also several fatty acid desaturases can be found.[22,23] Consequently, it might be, that the product of *mrc3* is involved in the double bond formation of the fatty acid terminus. In the next step the triple bond is to be introduced into the molecule at the C6-C7 position, leading to the formation of maracen. A fairly similar reaction is known from the crepenynate synthesis from linoleate, where the C12-C13 double bond is conversed into a triple bond via catalysis of a delta-(12)-fatty acid desaturase.[24–26] The proposed reaction mechanism is shown in Figure 2. Also in the alleged maracen biosynthetic gene cluster this functionality can be found, encoded by *mrc1*.

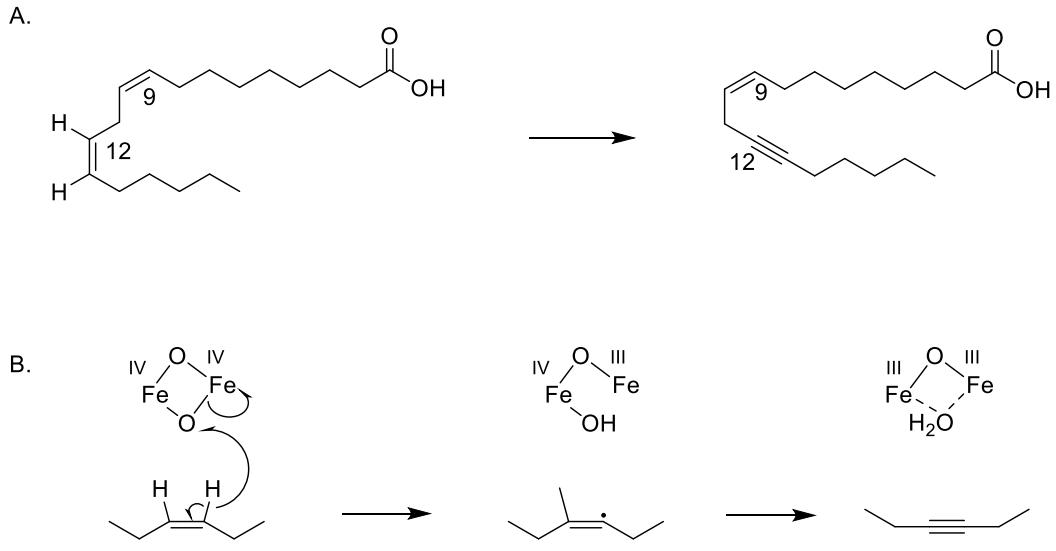


Figure 12: Proposed mechanism for fatty acid desaturase reaction.[27] Scheme of the reaction mechanism adapted from Reed et al., *J. Am. Chem. Soc*. 2003,125, 10635-10640.

To obtain maracen, the last step of our biosynthesis hypothesis proposes the addition of chlorine at the triple bond. Such a reaction is commonly applied in organic synthesis,

where the π-electrons of the triple bond attack the hydrogen of the hydrogen halide, which is then covalently bound to the carbon forming the respective carbocation intermediate. The resulting nucleophilic halide subsequently attacks the carbocation and is then placed at the highest substituted carbon. This reaction as depicted in Figure 3, follows the Markovnikov rules, meaning that addition of the hydrogen occurs at the carbon which has most hydrogen substituents whereas addition of the halide occurs at the highest substituted carbon with the least hydrogen substituents.

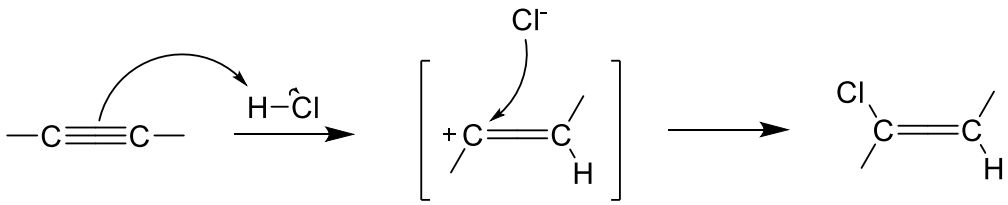


Figure 13: Electrophilic addition of hydrogen chloride to an internal alkyne moiety.

Also in nature, halogenated fatty acids are frequently reported, especially from organisms that were collected in marine habitats.[28] The chloroperoxidases were found to be the enzymes responsible for this halogenation as they catalyse the oxidation of halides to hypohalous acids, that can subsequently react with the unsaturated carbon-carbon bonds of unsaturated fatty acids to yield the respective halohydroxy-fatty acids.[29] But in the present case no additional hydroxyl group exists. Consequently a mechanism differing from the chloroperoxidase pathway is carried out in this step, as elimination of water from the chloro-hydroxy-alkene moiety is not possible. Several studies on unusual 6-brominated fatty acid resulted in a new hypothesis for biosynthesis of these metabolites.[30–32] Thus Lam et al. propose the addition of Br_2 across a double or a triple bond followed by the elimination of HBr to yield a mono-6-brominated fatty acid.[30] Also in maracen, the 6-position is mono-chlorinated, which indicates a similar mechanism. Two more candidate genes from the maracen biosynthetic gene cluster, whose function is still unclear (*mrc2* and *mrc10*) might be involved in this halogenation reaction. However, no further studies are reported to prove this biosynthesis hypothesis.

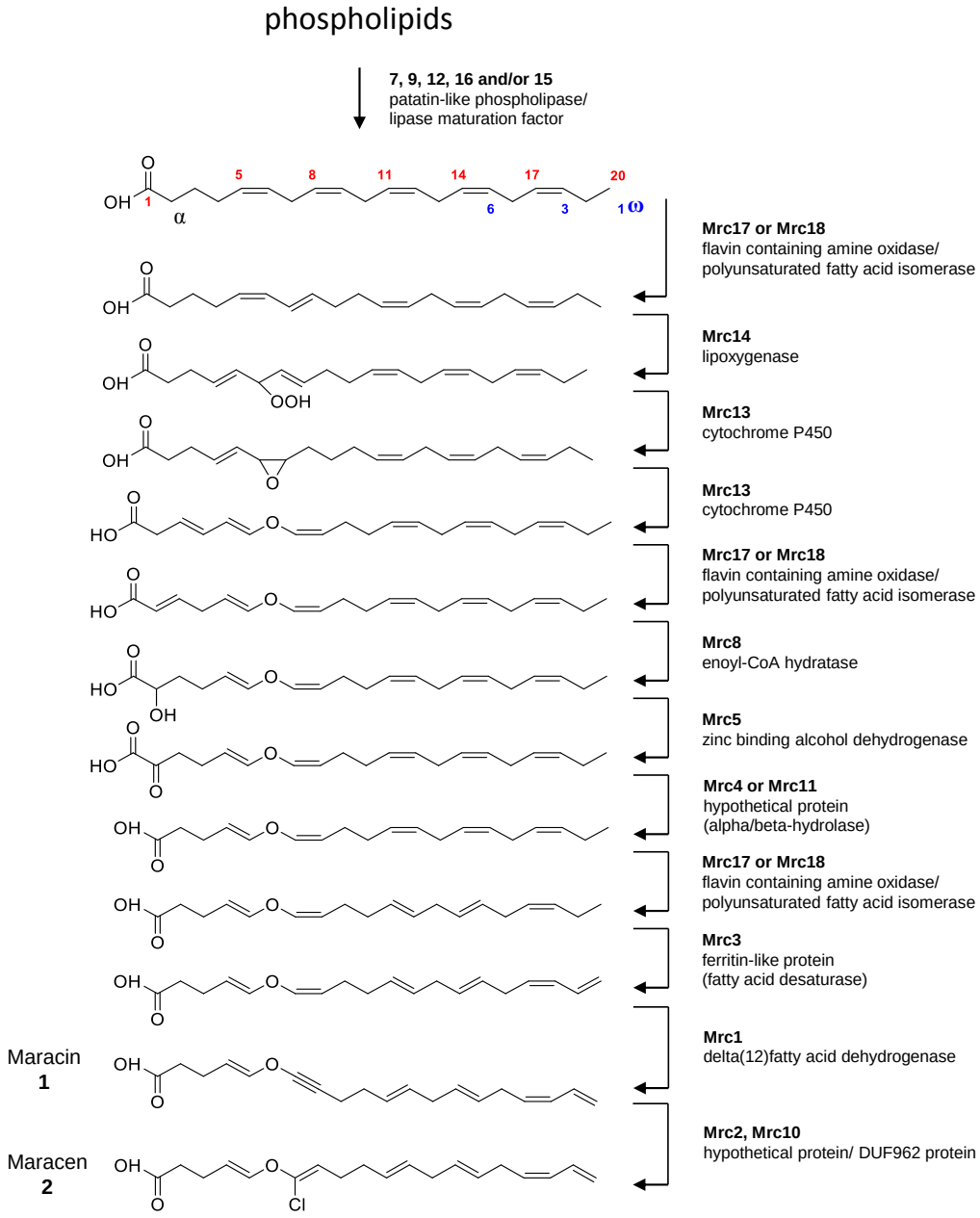


Figure 14: Putative maracen biosynthesis route starting from EPA: EPA is stepwise transformed via the enzymes encoded in the putative maracen gene cluster.

The above shown biosynthesis hypothesis appears conclusive considering the postulated connection between EPA and maracen/maracin production. Yet several inconsistencies between the predicted function of some of the genes and the reaction that is supposed to be catalysed by them remain. This especially holds true for the α-hydroxylation step in our hypothesis. Enzymes from the enoyl-CoA hydratase family such as Mrc8 indeed carry out hydroxylation reactions, but in general this takes place at the β-position of fatty acids.

On the other hand no other enzyme catalysing hydroxylation reactions is encoded in the gene cluster. Whether a hydroxyl-migration is involved in formation of the α-hydroxy fatty acids and what the underlying mechanisms would be, can at the moment not be stated with certainty. Apart from this, the question arises why several functionalities – such as the PUFA isomerase and the patatin-like phospholipase – are encoded repeatedly and whether they can replace each other or whether every single enzyme is needed to carry out one distinct reaction. From the biochemists point of view the proposed reactions catalysed by the cytochrome P450 enzyme, the lipoxygenase and the delta-(12)-fatty-acid isomerase are the interesting key steps in the proposed biosynthetic model. Further investigation of these reactions with the help of synthesized precursor molecules for instance could hence provide a more detailed insight into the role of these enzymes. Ultimately, we were able to set up a largely coherent hypothesis for maracen and maracin production considering all the functionalities encoded in the described genomic locus which exists in all maracen/maracin producer strains. In the following endeavours towards the establishment of a system for heterologous expression of the maracen/maracin biosynthetic pathway are described. With such a system implemented successfully, we aim to shed more light on the enzymatic processes involved in maracen and maracin biosynthesis.

5.3.3 Preparatory work towards design of an artificial maracen biosynthetic pathway

As genetic manipulations in the main producer strain of maracen and maracin – *Sorangium cellulosum* So ce1128[5] – are notoriously hard to achieve, we aimed to establish a heterologous production platform using modern technologies based on artificial DNA. According to the proposed biosynthetic pathway (Figure 4) we intended to construct a synthetic gene cluster comprising artificial versions of *mrc0-mrc18* (Table 1), which should allow for maracen/maracin production based on EPA. For EPA biosynthesis, a functional expression construct with native genes from *Sorangium sp.* SBSo026 (MSr9366) was already available (K. Gemperlein, unpublished), which could be easily adapted to the requirements for this project. To set the basis for the design of an artificial *mrc0-mrc18* gene cluster, the native gene sequences from *Sorangium sp.* SBSo026 had to be completely deciphered and verified. The presence of multiple copies of genes encoding identical functionalities as mentioned above impeded whole and accurate sequencing of

this region via the applied shotgun Illumina approach. In order to complete the gene cluster sequence PCR amplicons spanning the respective loci were generated and sequence verified.(see 5.6.1) By integration of the additional sequencing results, a continuous, gap-free sequence spanning 45 kbp was established facilitating all subsequent pathway design and engineering steps.

5.3.4 Design of an artificial maracen biosynthetic pathway

In collaboration with the bioinformatics company ATG:biosynthetics GmbH detailed analysis of the 45 kbp maracen biosynthetic gene cluster was carried out with particular emphasis on the transcriptional and translational initiation regions. This also included analysis of promoter and terminator structures in order to determine the native transcription units. The identified putative promoters were categorized as weak (w), medium strong (m) and strong (s), whereas only the latter two groups were considered as relevant parameters for constructional pathway design (Figure 5). In the constructional design, the present pathway was modularized enabling straightforward adaptions and alterations. This included a rational subdivision of the biosynthetic pathway into smaller fragments, which can be obtained via gene synthesis, as well as the engineering of unique restriction sites for pathway assembly and manipulation. A set of 29 restriction enzymes suitable for the envisaged cloning strategy was defined and the corresponding recognition sequences were eliminated from the native gene cluster sequence. Unique restriction sites (R-sites) for nine of these enzymes were re-introduced at the 5’ends of selected genes located at the beginning of proposed transcription units (Figure 5, Table 2). The remaining 21 enzymes from the list were required for vector backbone and additional pathway engineering as described below. In general, restriction sites engineering in coding sequences (CDS) was performed via silent point mutations and relevant genetic features (e.g. promoter structures, Shine-Dalgarno sites) were conserved revealing an artificial maracen pathway expressing the native biosynthesis machinery driven by native regulatory elements. As illustrated in Figure 5 the overall gene cluster arrangement was almost retained except the omission of the *pfa* biosynthetic genes and the inversion of the *mrc18-17-16* transcription unit, which later on allows for insertion of artificial bidirectional promoter boxes between *mrc15*/*mrc18*.

Table 3: List of unique restriction sites engineered in the maracen/maracin biosynthetic gene cluster.

No.	Location	Unique R-site
R1	5'-orf3	*Sal*I
R2	5'-orf4	*Pvu*II
R3	5'-orf6	*Pci*I
R4	5'-orf7	*Nco*I
R5	5'-orf9	*Bsp*EI
R6	5'-orf12	*Pst*I
R7	5'-orf15	*Bsr*GI
R8	5'-orf18	*Fse*I
R9	5'-orf16	*Nru*I

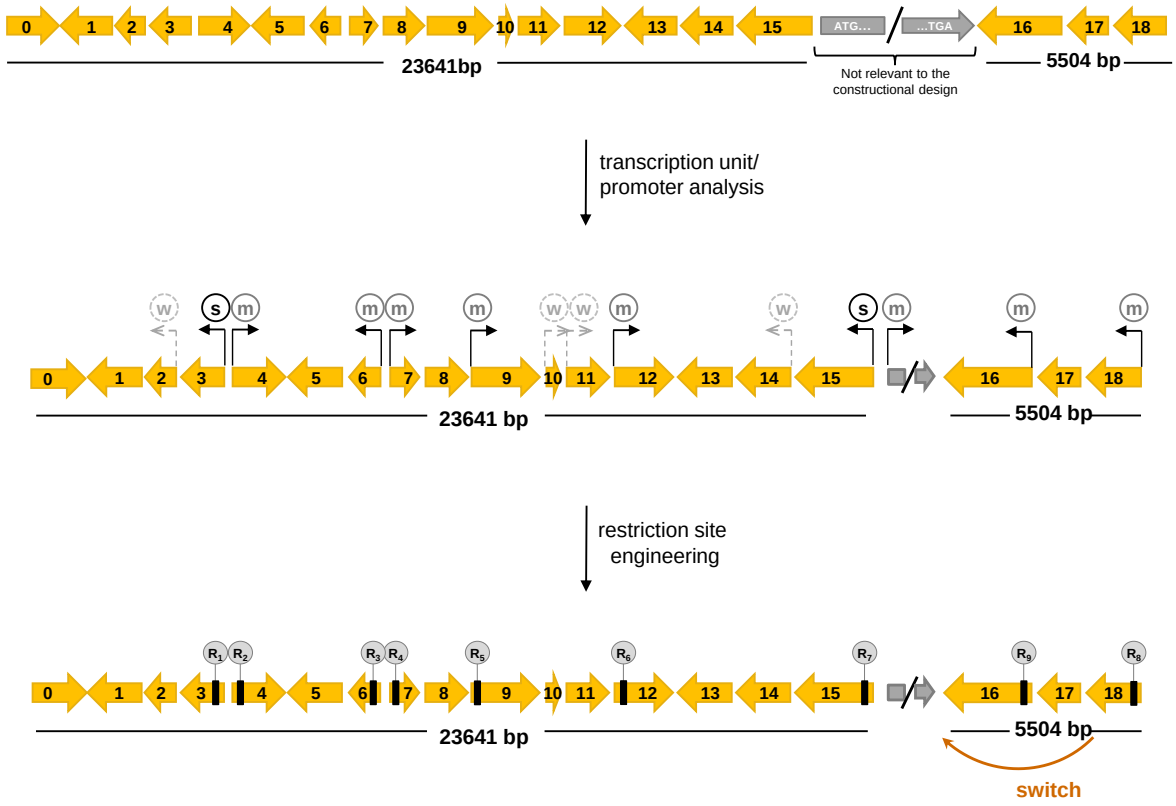


Figure 15: Initial analysis and design steps carried out for the maracen/maracin biosynthetic gene cluster. After transcription unit and promoter analysis within the gene cluster, constructs were planned for gene synthesis and the first set of restriction enzymes listed in table 2 were engineered in the gene cluster.

The constructional design process also involved the introduction of so-called splitter elements at the 5' and 3' ends of all genes except *mrc0*, *mrc10* and the nine positions where unique restriction sites have been already engineered (Figure 9). These splitter elements represent small DNA stretches (28 bp), which are temporarily integrated into the

CDS to provide additional unique restriction sites for pathway engineering, e.g. for targeted deletion of selected genes to investigate their role during maracen biosynthesis. Such modifications can be performed at an early stage of the assembly process based on constructs harboring subsets of the biosynthetic pathway (constructs pUCMrc0+1-5; Figure 9). Prior to assembly of the gene cluster for heterologous expression, splitter elements are removed in a process called “desplitting”, which can be performed as one-pot- reaction. Scarless reconstitution of the CDS is achieved by using a type IIS restriction enzyme (*Bsa*I), which can release the splitter elements by generating cluster fragments with compatible overhangs for religation. Further details on the splitter elements and the design of an artificial maracen biosynthetic gene cluster from ‘de-split’ constructs are given in the supporting information (5.6.5 – 5.6.9).

5.3.5 Design of the cloning vector pUCmrc

The envisaged construction of an artificial maracen biosynthetic pathway based on synthetic DNA building blocks also required the design of a suitable cloning vector. As early stage pathway assembly and desplitting processes are based on fragments which are less than 8 kb in size, a high copy vector backbone harboring a minimal set of genetic features (origin of replication, resistance gene, multiple cloning site) appeared applicable. To generate an appropriate plasmid, the sequence of the commercially available and well-established pUC18 vector was redesigned to meet the special requirements of the maracen project. Restriction sites of enzymes implemented in the cloning strategy were removed from the vector backbone sequence by appropriate point mutations, which were “silent” within CDS regions. In addition, a customized multiple cloning site (MCS) was generated, which is consistent to the MCS of the expression vector pEXPmrc. A map of the designed cloning vector named “pUCmrc” is illustrated in Figure 6, details on the sequence design are given in the supporting information (see chapter 5.6.3).

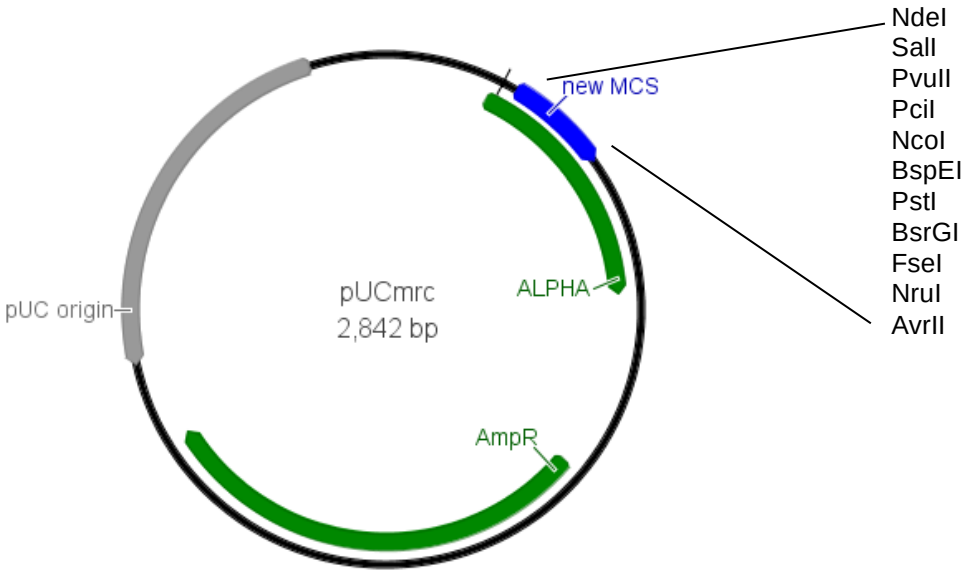


Figure 16: Depiction of the multi-purpose cloning vector pUCmrc; close-up on its multiple cloning site.

5.3.6 Design of the expression vector pEXPmrc

The first choice as expression host for the artificial maracen biosynthetic pathway was the myxobacterial model strain *M. xanthus* DK1622, in particular a mutant in which the well-expressed myxochromide A biosynthetic pathway (*mchA*) was deleted from the chromosome by replacement with an oxytetracycline resistance gene (*M. xanthus* DK1622 *ΔmchA-tet*). This mutant strain was recently established as expression host for native and artificial myxobacterial natural product pathways and a suitable minimal expression vector backbone, pSynBio2, is available (Burgard et al, unpublished). The vector features a p15A low copy origin of replication, an oxytetracycline resistance gene for homologous recombination into the host chromosome and a zeocin resistance gene that can be used for selection. Several adaptions of the expression vector for the maracen/maracin project were made; this includes the implementation of a suitable MCS (consistent with the one from the pUCmrc vector) plus engineering/conservation of unique restriction sites between functional vector elements. The resulting 'modular' expression plasmid pEXPmrc can be modified in various ways, e.g. by exchange of selection markers and transfer cassettes or the insertion of additional recombinant fragments. Thereby it is possible to add the EPA biosynthetic genes (*pfa* genes) to have the entire pathway on the same physical entity alternatively to the planned multiplasmid approach. In order to extend the expression toolbox an artificial Mx8 phage integrase cassette[33] was designed by modifying the respective region from plasmid pCK_T7A1_att (C. Kegler et al., unpublished) according to the given restriction sites requirements. The generated fragment can replace the oxytetracycline cassette to enable integration via a phage-derived system, which can then not only be applied for the *M. xanthus ΔmchA-tet* mutant, but also for the corresponding wildtype and other strains harboring an Mx8 attachment site. A map of the generated pEXPmrc expression vector is illustrated in Figure 7, details on the sequence design including the *mx8* cassette are given in the supporting information (see chapter 5.6.4).

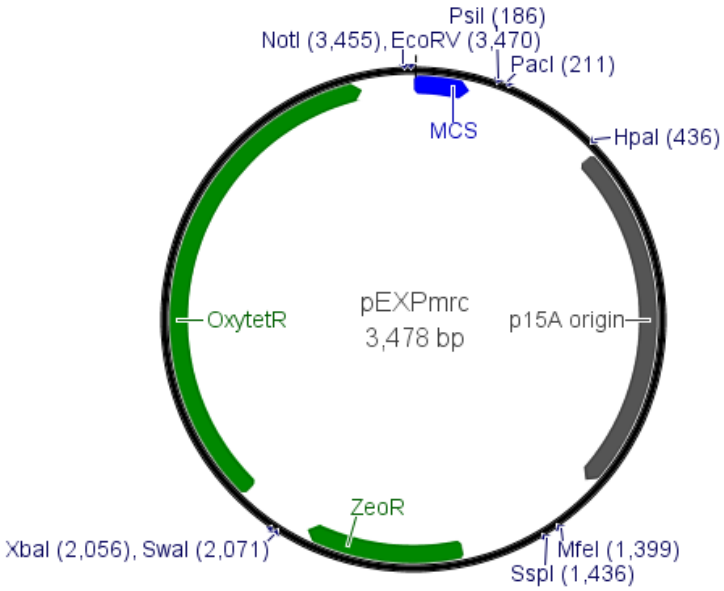


Figure 17: Depiction of the expression vector pEXPmrc; all relevant restriction sites are shown.

5.3.7 Gene synthesis to physically generate the artificial maracen pathway and the synthetic vectors

After completing the sequence design described in chapters 5.3.4 - 5.3.6 the artificial DNA sequences were generated via gene synthesis commissioned by ATG:biosynthetics GmbH. Due to its large size the maracen biosynthetic pathway was split into eight synthetic fragments, which span between 2.0 and 4.3 kb (see Table 3). For the upstream part of the gene cluster two alternative fragments, with and without gene *mrc0*, were generated (Mrc0 and Mrc1). In general, all fragments were flanked by unique restriction sites that were used for subsequent cloning procedures as described in chapter 5.3.8.

In addition to the gene cluster fragments the two synthetic vector backbones, pUCmrc and pEXPmrc, plus the above mentioned mx8 integrase cassette were synthesized (Table 3). In total, 12 synthetic DNA fragments were generated via gene synthesis and delivered in the standard cloning vector "pGH" (pUC18 derivative harboring an ampicillin resistance gene). The synthesized vector backbones were subsequently released and circularized by applying conventional restriction/ligation techniques to generate pUCmrc and pEXPmrc.

Table 3: Constructs generated via gene synthesis. The fragment size and the flanking restriction sites used for later assembly are listed respectively.

Construct	Size	included genes / gene fragments	flanking restriction sites
pGH-Mrc0	2016 bp	0, 1, 2, 3 & 5'-end gene 4	*Nde*I, *Pvu*II
pGH-Mrc1	3686 bp	1, 2, 3 & 5'-end gene 4	*Nde*I, *Pvu*II
pGH-Mrc2	3569 bp	3'-end gene to 5''-end gene 7	*Pvu*II, *Nco*I
pGH-Mrc3-1	3893 bp	3'-end gene 7 to 5'-end gene 9	*Nco*I, *Kpn*I
pGH-Mrc3-2	2430 bp	3'-end gene 9 to 5'-end gene 12	*Kpn*I, *Pst*I
pGH-Mrc4-1	3902 bp	3'-end gene 12, to 3'-end gene 14	*Pst*I, *Kpn*I
pGH-Mrc4-2	4261 bp	5'-end gene 14 to 3'-end gene 15	*Kpn*I, *BsrG*I
pGH-Mrc5-1	3671 bp	5'-end gene 15 to 5'-end gene 17	*BsrG*I, *Kpn*I
pGH-Mrc5-2	2872 bp	3'-end gene 17 to 3'-end gene 16	*Kpn*I, *Avr*II
pGH-mx8	2960 bp	-	*EcoRV,NotI//SwaI,XbaI,*
pGH-pUCmrc	2842 bp	-	-
pGH-pEXPmrc	3478 bp	-	-

5.3.8 In vitro reconstitution of the maracen biosynthetic pathway from synthetic DNA fragments

To reconstitute the maracen biosynthetic pathway the synthetic gene cluster fragments obtained from ATG:biosynthetics GmbH (Table 3) were first subcloned into the pUCmrc vector, which has an engineered restriction sites pattern compatible with downstream cloning steps. In this process some of the fragments were stitched (Mrc3-1 with Mrc3-2, Mrc4-1 with Mrc4-2, Mrc5-1 with Mrc5-2) to cover the artificial pathway with five constructs (constructs pUCMrc0+1-5; Figure 9). For construct 1, two versions (with and without gene *mrc0*) have been generated. The five complementary constructs built the basis for gene cluster reconstitution, which involves three major steps: (1) Definition of the gene set: Making use of unique restriction sites (including such in splitter elements) selected genes can be deleted or replaced, e.g. for functional studies. (2) 'De-splitting' process: Splitter elements are subsequently removed in a one-pot reaction using the type IIS enzyme *Bsa*I to reconstitute the CDS. (3) Gene cluster assembly: The five gene cluster parts are stitched into the expression vector pEXPmrc via unique restriction sites. Following this strategy, we first aimed to construct a full-length version of the *mrc* biosynthetic gene cluster harboring genes *mrc0-18*. Starting from constructs pUCMrc0+1, pUCMrc2, pUCMrc3, pUCMrc4, pUCMrc5 (Table 4), splitter elements were removed via hydrolysis with the respective type IIS enzyme. Prior to rejoining of cluster fragments the significantly smaller splitter elements (28 bp) were removed from the fragment mixture using spin columns with a 100 bp cut-off. The eluted fragment mixture was then re-ligated via unique and compatible fusion sites. The expected products, completely 'de-split' constructs, were verified by restriction analysis to confirm correct reassembly and to exclude remains of splitter elements. For this, a combination of two restriction sites was chosen, one located in the splitter elements and one in the cloning vector backbone. (see supporting information chapter 5.6.8) To further confirm these results, all 'de-split' constructs (pMrc0+1 – pMrc5, Table 4) were additionally verified by Sanger sequencing using primers flanking the 'splitter regions' (see supporting information, chapter 5.6.9). Based on the obtained 'de-split' constructs the entire *mrc* gene cluster was reconstituted on the expression vector pEXPmrc by applying conventional restriction and ligation methods. In general, after subcloning the first fragment (Mrc0+1) the downstream gene cluster parts were introduced via the following strategy: the precursor construct was hydrolyzed using two enzymes, one having a unique restriction site at the 3' end of the

insert (that corresponds to 5’ end of the next incoming fragment) whereas the other enzyme cuts in the MCS of pEXPmrc, which is illustrated in Figure 8. In this manner, stepwise reconstitution of the artificial pathway starting from the 5’ end of the biosynthetic gene cluster was achieved.

The first cluster fragment, the 6.1 kb Mrc0+1 fragment from pMrc0+1, was subcloned into the pEXPmrc vector via *Nde*I/*Pvu*II to yield pEXPMrc0+1. The connecting building block, the 3.4 kb Mrc2 fragment from pMrc2, was then introduced via *Pvu*II/*Nco*I restriction sites resulting in construct pEXPMrc0-2. Subsequently, the third building block, the 6.2 kb Mrc3 fragment from pMrc3, was added via *Nco*I/*Pst*I restriction sites resulting in construct pEXPMrc0-3, which was then extended with the 7.9 kb Mrc4 fragment from pMrc4 using the *Pst*I/*BsrG*I restriction sites. Ultimately, the last gene cluster fragment, the 6.5 kb Mrc5 fragment from pMrc5, was added via *BsrG*I/*Avr*II restriction sites leading to the complete expression construct pEXPMrc0-5.

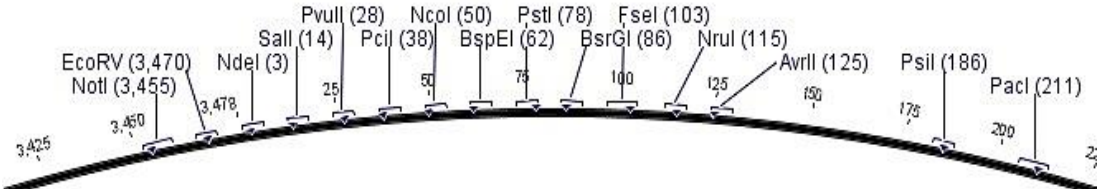


Figure 18: Section of the expression vector pEXPmrc showing the multiple cloning site (MCS) used for stepwise assembly of the maracen biosynthetic gene cluster.

Table 4: List of constructs generated in this study.

Constructs generated in this study		
Construct name	formation	size (bp)
pUCMrc	self-circularization from pGH-pUCMrc	2842
pEXPMrc	self-circularization from pGH-pEXPMrc	3478
pUCMrc1	cloning of fragment Mrc1 from pGH-Mrc1 into pUCMrc	6373
pUCMrc0+1	addition of fragment Mrc0 from pGH-Mrc0 to pUCMrc1	8149
pUCMrc2	cloning of fragment Mrc2 from pGH-Mrc2 into pUCMrc	6282
pUCMrc3-1	cloning of fragment Mrc3-1 from pGH-Mrc3-1 into pUCMrc	6606
pUCMrc3	addition of fragment Mrc3-2 from pGH-Mrc3-2 to pUCMrc3-1	9006
pUCMrc4-1	cloning of fragment Mrc4-1 from pGH-Mrc4-1 into pUCMrc	6615
pUCMrc4	addition of fragment Mrc4-2 from pGH-Mrc3-2 to pUCMrc4-1	10846
pUCMrc5-1	cloning of fragment Mrc5-1 from pGH-Mrc5-1 into pUCMrc	6384
pUCMrc5	addition of fragment Mrc5-2 from pGH-Mrc5-2 to pUCMrc5-1	9226
pMrc0+1	by desplitting and rejoining of pUCMrc0+1	8029
pMrc2	by desplitting and rejoining of pUCMrc2	6186
pMrc3	by desplitting and rejoining of pUCMrc3	8862
pMrc4	by desplitting and rejoining of pUCMrc4	10702
pMrc5	by desplitting and rejoining of pUCMrc5	9130
pEXPMrc0+1	cloning of the desplit fragment Mrc0+1 from pMrc0+1 into pEXPMrc	8752
pEXPMrc0-2	cloning of the desplit fragment Mrc2 from pMrc2 into pEXPMrc0+1	12171
pEXPMrc0-3	cloning of the desplit fragment Mrc3 from pMrc3 into pEXPMrc0-2	18266
pEXPMrc0-4	cloning of the desplit fragment Mrc4 from pMrc4 into pEXPMrc0-3	26213
pEXPMrc0-5	cloning of the desplit fragment Mrc5 from pMrc5 into pEXPMrc0-4	32573
pEXPMrc0-5_pfa	cloning of the *pfa* gene cluster into pEXPMrc0-5	51240
pCm15A_H-right_pfa	derivative of pHybpfa1-mx9.2; details see chapter 5.6.10	21938

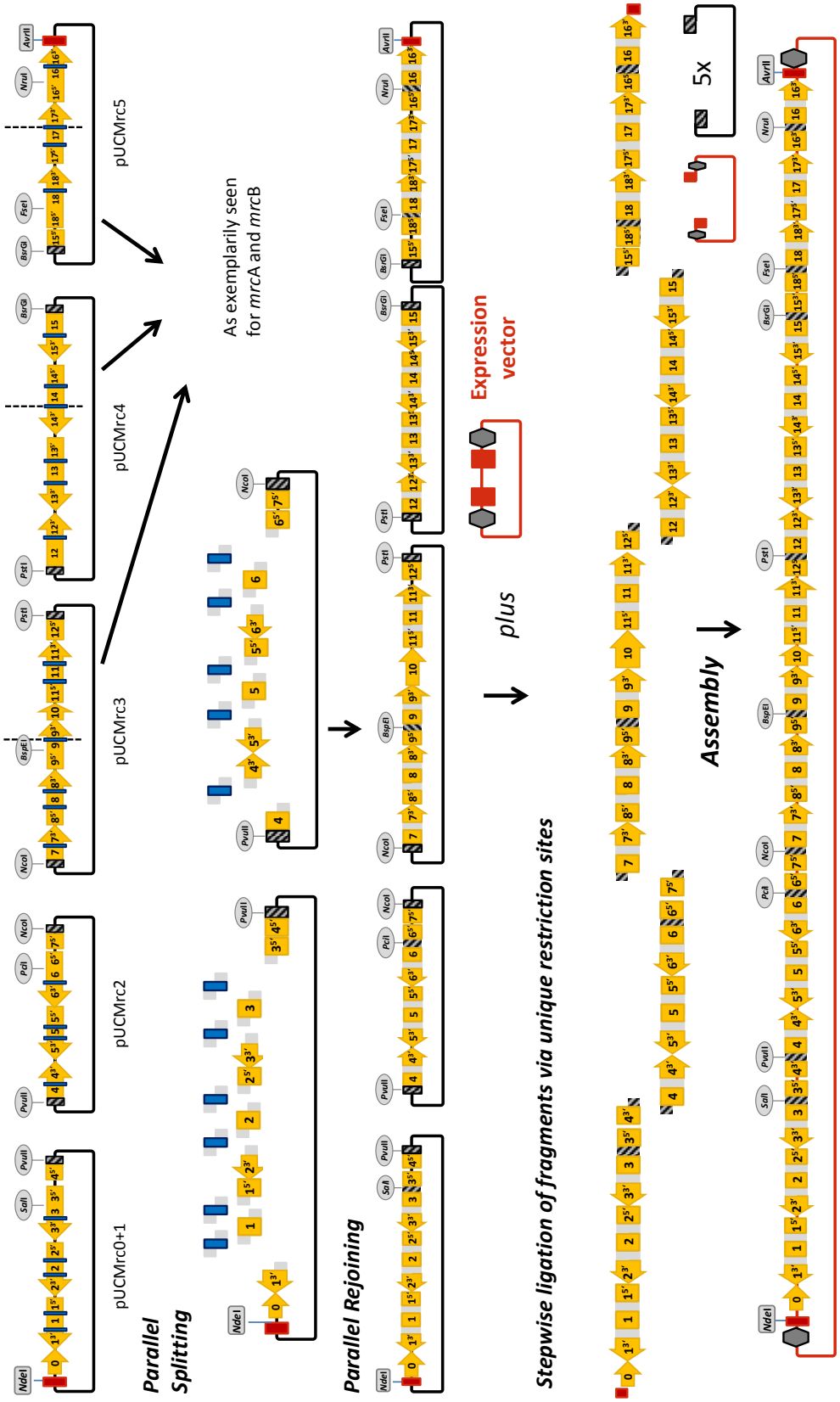


Figure 19: Overview of the two major steps performed towards assembly of the maracen/maracin biosynthetic gene cluster. Initially, splitter elements were removed from synthetic constructs which were subsequently rejoined. To build the final construct, synthetic fragments released from their backbone vector were combined sequentially via distinct compatible ends generated by DNA restriction using defined enzyme combinations.

5.3.9 Strategies for integration of the maracen biosynthetic pathway into various host genomes

For integration of the maracen biosynthetic pathway into the heterologous host's genome several strategies were envisaged. Initially, two *Myxococcus xanthus* host systems, the DK1622 wildtype and the DK1622 *ΔmchA* deletion mutant were available and as described in chapter 5.3.6 tools for chromosomal integration for both strains were designed. Due to the anticipated interdependence of PUFA and maracen biosynthesis (maracen is assumed to originate from EPA, which is not provided by the native *M. xanthus* metabolism), co-expression of the *pfa* biosynthetic gene cluster was pursued.

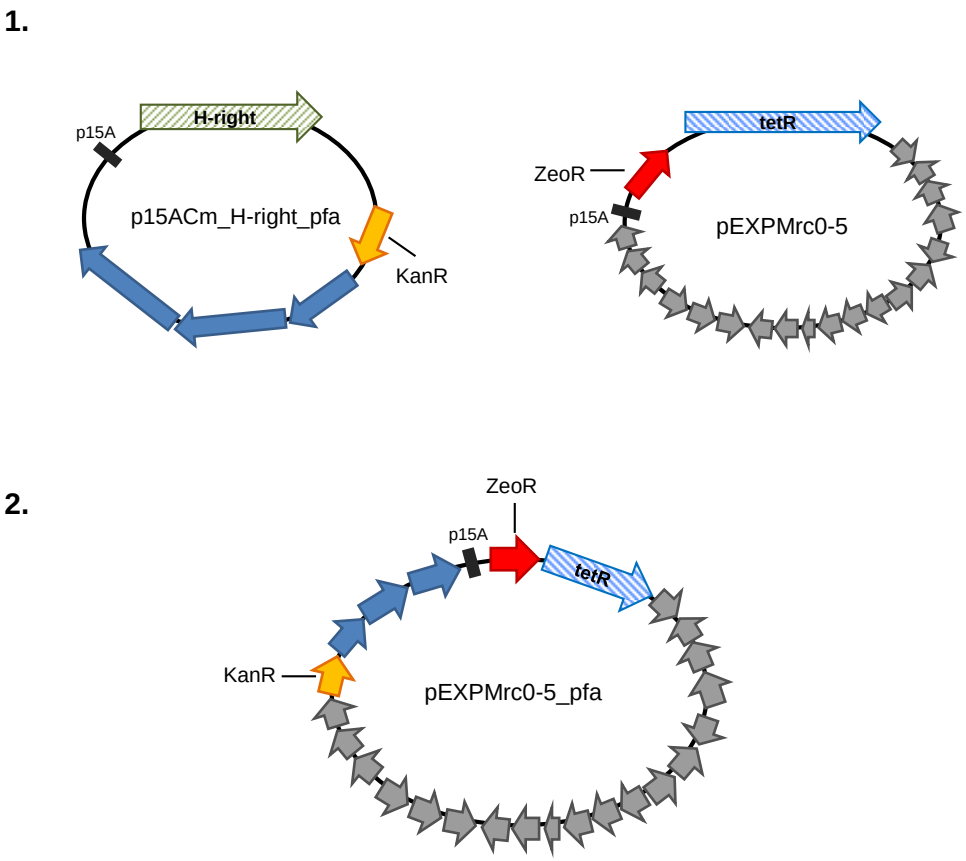


Figure 20: Constructs designed for integration of the maracen biosynthetic pathway into the *Myxococcus* host genome. In 1. the two constructs for the multi-plasmid approach for *Myxococcus xanthus* DK1622 *ΔmchA-tet* are shown. p15ACm_H.right_pfa contains the *pfa* gene cluster whilst pEXPMrc0-5 harbors all maracen biosynthesis genes. In 2. the construct for the single plasmid approach harboring the *pfa* and the *mrc* gene cluster is shown.

In general, two options were available: a multi-plasmid or a single plasmid approach. For the multi-plasmid approach, the *pfa* gene cluster is transformed as an autonomous expression construct and upon successful integration and heterologous EPA production, the respective mutant is additionally transformed with the construct containing the maracen biosynthesis genes. Alternatively, in the single plasmid approach, the *pfa* gene cluster is subcloned onto the same expression construct with the *mrc* genes (construct pEXPmrc0-5 via *Psi*I/*Pac*I restriction sites) to transfer and express the entire set of biosynthetic genes on the same physical entity. For integration of the expression constructs into the respective host genome different strategies were elaborated including the use of phage-derived systems (deriving from the Mx8 and Mx9 phages[33,34]) or homologous chromosomal fragments to allow for recombination events. Following the multi-plasmid approach, the overall strategy was to initially establish a mutant which contains the *pfa* gene cluster and thus produces the required fatty acid eicosapentaenoic acid. Thereupon integration of the maracen biosynthesis genes was pursued. Initially only *Myxococcus xanthus* DK1622 *ΔmchA-tet* was used as host strain whereas the former locus of the myxochromide A gene cluster was chosen as integration site. The PUFA construct p15ACm_H-right_pfa (see chapter 5.6.10) was thus to be introduced downstream the tetracycline resistance gene into the right homology arm whereas integration of the maracen biosynthesis genes was intended via homologous recombination at the tetracycline resistance gene as shown in Figure 11.

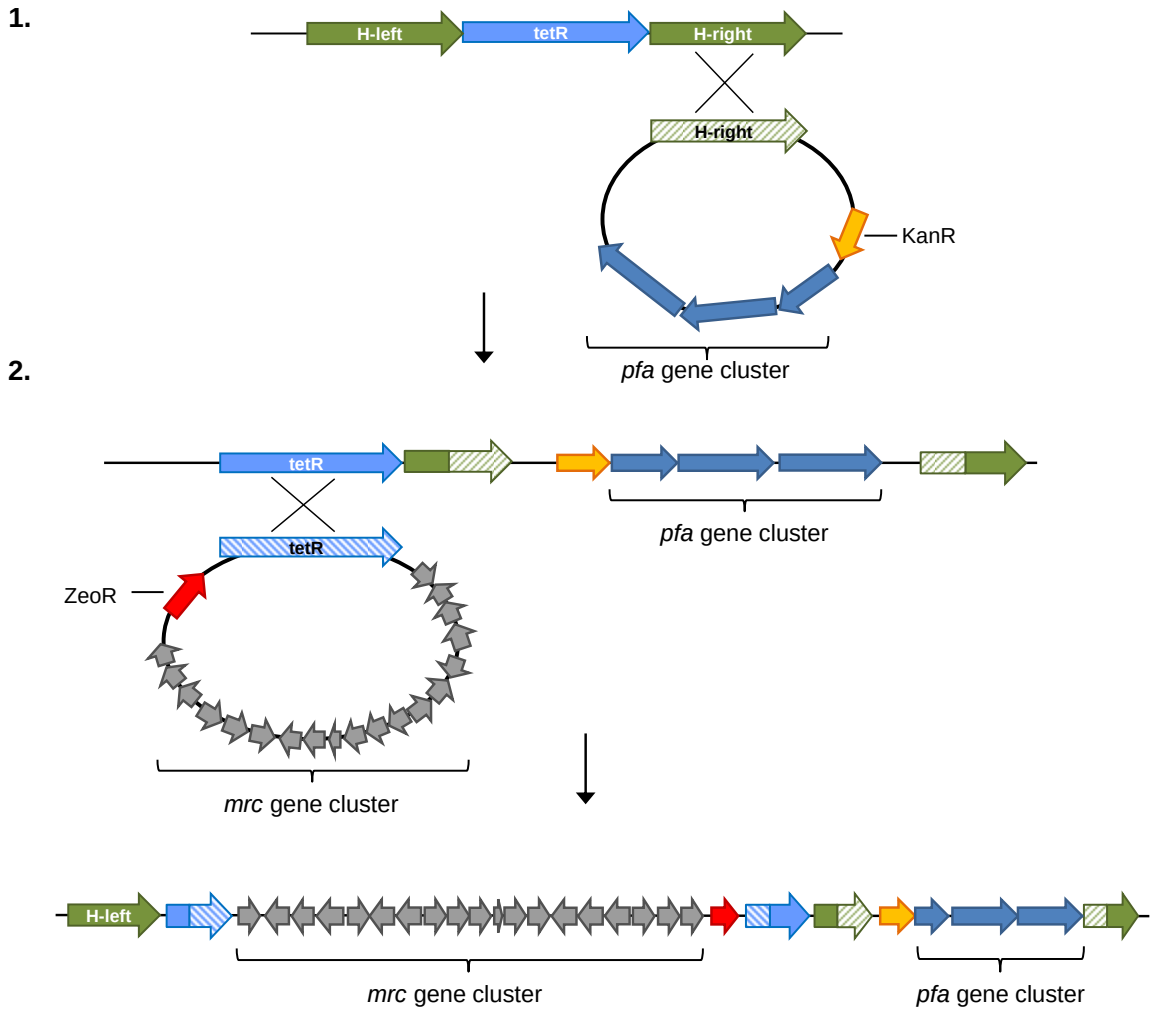


Figure 21: Genetic organisation of *Myxococcus xanthus* DK1622 *ΔmchA-tet* in the former locus of the myxochromide biosynthetic pathway, which was deleted from *M. xanthus* DK1622 and replaced with the H-left/tet/H-right cassette shown above. **1.** Integration of the *pfa* gene cluster takes places via homologous recombination at the right homology arm H-right leading to a kanamycin resistant mutant with the genotype shown. 2. Integration of the maracen biosynthesis genes is achieved via homologous recombination via the tetracycline resistance gene leading to a kanamycin and zeocin resistant double mutant.

5.3.10 Towards expression of the maracen biosynthetic pathway

Upon finalization of the maracen expression construct pEXPMrc0-5 (as described in chapter 5.3.8), the *pfa* gene cluster was added to the construct via the designated *Psi*I/*Pac*I restriction sites, yielding pEXPMrc0-5_pfa. The thus resulting construct was re-sequenced using the Illumina technique to verify correct assembly and error-free sequence. In a first approach, pEXPMrc0-5_pfa (see Figure 10) was transformed to *Myxococcus xanthus* DK1622 *ΔmchA-tet*, but upon selection with zeocin, no mutants

could be obtained. Also numerous repetitions of the transformation protocol did not result in zeocin resistant clones. Consequently, in a second series of experiments the multi-plasmid approach was pursued. After transformation of the PUFA construct, equipped with a 1 kb homology cassette to H-right, a kanamycin resistant mutant was obtained. PCR analysis confirmed the presence of the gene cluster at the intended locus (see supporting information 5.6.11). To check for production of PUFAs, the mutant *M. xanthus* DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa was grown in liquid media for three days and the PUFA production was assayed via GC-MS analysis (see supporting information 5.6.12). The analysis showed that the generated mutant produced eicosapentaenoic acid as well as arachidonic acid which substantiates the functional expression of the PUFA biosynthetic gene cluster in the heterologous host. In the following this mutant was transformed with expression construct pEXPMrc0-5 to completely reconstitute the proposed maracen biosynthetic pathway in the host. However, despite numerous trials a corresponding mutant strain harboring the additional *mrc* gene set could not be obtained so far. For this, several reasons are conceivable. The fact that integration of the maracen/maracin biosynthesis genes could not be achieved can be attributed to a very low transformation- and recombination efficiency but it might as well be the case that several Mrc proteins are toxic to the selected host strain. The latter issue could be addressed by the replacement of native promoters for inducible promoter systems such as the vanillate promoter.[35] Another conceivable scenario would be that maracen and maracin itself are toxic to the host strain and therefore the construct is not integrated into the host's genome. Yet no indications for that could be found in previous activity testings of the compounds, but it might as well be the case that genes encoding important functionalities for self-resistance mechanisms are missing on the synthetic construct.

5.4 Conclusion & Outlook

In this work we described a new type of biosynthetic pathway most likely responsible for the production of maracen and maracin in several myxobacterial producer strains. The anticipated interdependence between polyunsaturated fatty acid biosynthesis by a new type of *pfa* gene cluster and the biosynthesis of maracen and maracin, respectively, is one of the most striking aspects of this work. By applying a synthetic biology approach we thus set out to investigate the underlying enzymatic steps for transformation of eicosapentaenoic acid (EPA) into maracen/maracin. Other than the classical gene inactivation and metabolite profiling approach in the native producer strain we aimed to construct a heterologous expression platform in a suitable host (*M. xanthus*) based on synthetic DNA. The developed strategy allows for highly flexible and rapid pathway engineering, e.g. to perform targeted modifications within the gene cluster for gene-to-function studies. The applied splitter technology represents a unique feature facilitating gene cluster engineering in a very straightforward fashion as individual genes including their regulatory elements can be easily removed or replaced. This concept not only provides a valuable platform for biosynthesis studies, but also enables the rapid construction of gene cluster variants for production of novel derivatives. In the proposed model for maracen biosynthesis, the reactions catalysed by the lipoxygenase and the delta-(12)-fatty acid dehydrogenase are of major interest as especially bacterial lipoxygenases are reported rarely and their function is not well understood. To study the underlying biochemistry in more detail, *in-vitro* assays using recombinant enzymes and synthetic maracen precursor molecules might be conceivable. Another interesting aspect is the presence of multiple copies of genes encoding identical functionalities such as the PUFA isomerases Mrc17 and 18, and the patatin-like phospholipases Mrc7, 9, 12 & 16. Finding out whether these enzymes are interchangeable or whether each of them catalyzes a distinct reaction is therefore another important point to inspect.

In our study aiming at the establishment of a heterologous expression platform for the putative maracen/maracin biosynthetic gene cluster, transformation of *M. xanthus* with the *mrc0-18* expression construct could not be achieved for which possible reasons were discussed. A future approach towards a heterologous maracen/marcin production system is the evaluation of alternative hosts closer related to the native producer strains. Promising candidates are strains belonging to the genus *Sorangium*, which do not

naturally produce maracen or maracin and for which genetic tools have been established. Taking these aspects into account, *Sorangium cellulosum* So ce1525 was considered a potential host strain complying with all the above mentioned requirements. The strain is well-investigated as chlorotonil producer and the identified biosynthetic gene locus probably provides a suitable integration site for the maracen biosynthetic pathway.[3] We thus planned for a synthetic DNA cassette containing all features necessary for transfer of the construct and later selection (Figure 12).

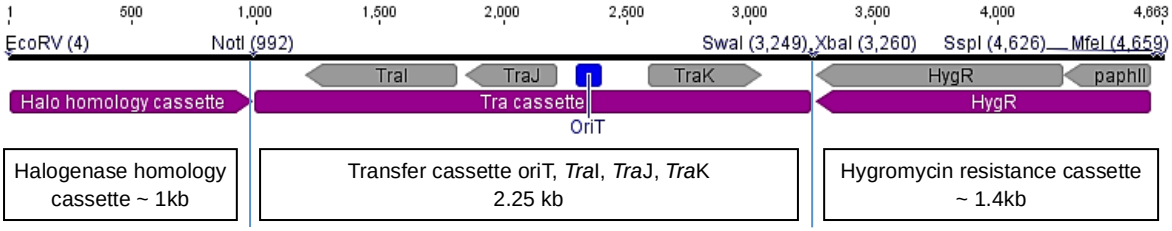


Figure 22: Structure of the transfer cassette designed for transformation and integration of the maracen/maracin biosynthetic gene cluster into the alternative host *Sorangium cellulosum* So ce1525.

The genes for transfer and selection were derived from the pSUP_Hyg vector,[36] whereas the “halo cassette” represents a 1 kb homology region to the chlorotonil halogenase encoding gene. The 4.7 kb transfer cassette was obtained via gene synthesis and can be used to modify expression construct pEXPMrc0-5_pfa by replacing the tetracycline and zeocin resistance genes Following the “single plasmid approach” the modified expression plasmid harbouring both, *pfa* and *mrc* genes, can be subsequently transferred into the *Sorangium* host via conjugation. It remains to be evaluated if the 55.2 kb expression construct can be efficiently transferred, but the usage of *S. cellulosum* as closer related host represents a promising future approach.

5.5 Experimental section

5.5.1 Bacterial strains and culture conditions

Myxococcus xanthus DK1622 was cultivated at 30 °C in CTT media [Casitone, Difco (10 g/l), $MgSO_4$ (8 mM), Tris-HCl (10 mM) pH 8.0, K_2HPO_4 (1 mM) pH 7.6, final pH adjusted to 7.6]. For liquid cultures, 20 ml of CTT media were inoculated with DK1622 cells from a CTT agar plate and incubated for 2 - 3 days at 30 °C. 0.5 - 2ml of thus set up pre-cultures were used to inoculate 25 ml of fresh CTT media in 100 ml Erlenmeyer flasks. These cultures were incubated at 30 °C and 180 rpm shaking for 2 - 4 days. For permanent stocks, agar pieces grown with the strain were frozen at –80°C in cryo vials.

The mutant DK1622 Δ*mch*A-*tet* was grown in the same way but with addition of oxytetracycline at 6.25 µg /ml.

Escherichia coli strains were cultured in LB media (lysogeny broth; tryptone 1 %, yeast extract 0.5 %, sodium chloride 0.5 %) at 37 °C until the required cell density was reached. If necessary the appropriate antibiotics were added to the media for selection.

5.5.2 DNA manipulation, analysis and PCR

For site directed digest of DNA type II restriction endonucleases (Thermo scientific) were used. The enzymes were applied according to the manufacturer's recommendations using the required buffer and temperature conditions. To perform a double digest for subsequent ligation experiments the two required enzymes were applied one after another. The reaction was set up in a total volume of 100 µl consisting of 40 µl plasmid DNA (1-5 µg), 10 µl enzyme buffer, 2.5 µl enzyme I and 47.5 µl sterile water. After incubation for 2h at the needed temperature the DNA was precipitated adding 1/10 vol 3M NaOAc, 2.5 vol ethanol p.a. and incubation at -20°C for at least 30 min. The DNA pellet formed by centrifugation at 15000 rpm and 4°C was subsequently washed, air dried and dissolved in 90 µl sterile water. The second part of the digestion was performed by adding 10 µl of enzyme buffer II and 2.5 µl of enzyme II. In case of a following ligation the vector was dephosphorylated adding 1.0 µl SAP to the mixture. After another two hours of incubation the DNA was separated on an agarose gel to check for complete digest.

PCR reactions were carried out in an eppendorf mastercycler pro using the following program: 5 min initial denaturation (5 min at 98°C) followed by 30 cycles of denaturation (98°C, 20 sec), annealing (58-68°C, 30 sec), and elongation (0.5 – 1.5 min according to the expected amplicon size). A final extension step (5 min at 72°C) was carried out at the end of the program. PCR amplicons were analyzed by gel electrophoresis and isolated by using the peqGOLD Gel Extraction Kit (Peqlab).

5.5.3 Ligation

To ligate insert and vector DNA, a T4 DNA polymerase was used. For efficient ligation a vector-insert-ratio of 2:1 was aimed for. After purification of digested DNA via gel electrophoresis and subsequent gel extraction of the DNA samples the DNA concentration was measured using a NanoDrop 2000 (thermo scientific). The ligation mixture was set up in a total volume of 10 µl consisting of vector and insert (1:2), 1.0 µl T4 ligase, 1 µl T4 ligase buffer and sterile water. The mixture was incubated for either 2 hours at room temperature or if not applicable overnight at 16°C. If necessary the ligated DNA was purified using phenol-chloroform-isoamylalcohol extraction.

5.5.4 Desplitting and re-ligation

For desplitting and re-ligation constructs 1-5 (Figure 8) were incubated with the type IIS restriction enzyme *Bsa*I according to the manufacturer's instructions for 2 hours at 37°C. After that, an aliquot (5 µl) of the mixture was taken for agarose gel analysis whilst the rest of the sample was heat-inactivated for 10 min at 65°C. To remove the splitter elements the sample was then purified via a Nucleo Spin PCR Clean Up column (Macherey-Nagel), whereas the flow-through was removed and the eluted fraction was subsequently used for ligation. In a 10 µl volume, the DNA fragments were re-ligated at 16 °C overnight using a T4 ligase. After heat inactivation of the ligase at 65°C for 10 min, an aliquot of the ligation mixture was electroporated to *E. coli* DH 10b. Clones were selected on LB agar plates supplemented with 6.25 µg/ml tetracycline. To verify complete removal of all splitter elements, a sophisticated restriction digest scheme was developed (see supporting information chapter 5.6.8). Correct constructs were further verified by Sanger sequencing covering the former locus of the respective splitter elements. The sequencing primers used are listed in Table S9

5.5.5 Extraction of cellular fatty acids

The cellular fatty acids were extracted using the FAME method.[37] For this purpose, 50 ml of a culture were harvested at 8000 rpm for 10 min at room temperature. The cell pellet was transferred to a glass vial and dried in a vacuum concentrator. 500 µl of a mixture of methanol, toluene, and sulphuric acid (50:50:2, v/v/v) were added. The vial was sealed with a teflon-lined screw cap and incubated at 80°C for 18 h. After the mixture was cooled down to room temperature, 400 µl of an aqueous solution consisting of 0.5 M NH_4HCO_3 and 2 M KCl were added, and the sample was vortexed for 5 s. Phase separation was achieved by centrifugation at 4000 rpm for 5 min at room temperature. 75 µl of the upper phase were used for GC-MS analysis.

5.5.6 Analysis of fatty acid methyl esters (FAME) by GC-MS

GC-MS was carried out on an Agilent 6890N gas chromatograph (Agilent Technologies) equipped with a 7683B split/splitless injector with autosampler (Agilent Technologies) and coupled to a 5973 electron impact mass selective detector (Agilent Technologies). Helium was used as the carrier gas at a flow rate of 1 ml/min. 1 µl of the sample was injected in split mode (split ratio, 10:1). The analytical column was a (5% phenyl)-methylpolysiloxane capillary column (Agilent J&W DB-5ht; 30 m x 0.25 mm i.d. x 0.1 µm film thickness, maximum temperature 400 °C; Agilent Technologies). The column temperature was kept at 130°C for 2.5 min, increased to 240°C at a rate of 5°C/min, then ramped to 300°C at 30°C/min, and held at 300°C for 5 min. Other temperatures were as follows: inlet, 275°C; GC-MS transfer line, 280°C; ion source, 230°C; and quadrupole, 150°C. The mass selective detector was operated in scan mode, scanning the mass range from *m*/*z* 40 to 700. Scan control, data acquisition, and processing were performed by MSD ChemStation and AMDIS software, version 2.68, based on the fragmentation patterns and retention times, in comparison with Supelco 37 Component FAME Mix and LC-PUFAs (all Sigma-Aldrich) as reference standards, and NIST 08 library.

5.6 Supporting Information

5.6.1 Decipherment of the complete maracen and maracin biosynthetic gene cluster sequence

For sequence completion, PCR amplicons were generated using primer pairs flanking problematic regions and sequence gaps (Table S1). A fosmid from a gene library of *Sorangium sp.* SBSo026 (MSr9366) served as a template (pCC1FOS C:C03, K. Gemperlein, personal communication) After amplification the fragments were purified via gel electrophoresis and upon gel extraction, samples were send for sequencing using the same set of primers as sequencing primers. The obtained sequence data were subsequently assembled with the available sequence information using Geneious v8.1.4 in order to completely decipher the *mrc* gene cluster sequence.

Table S1: Primers used for amplification of short DNA fragments within the maracen biosynthetic gene cluster

Name	Sequence (5‘ → 3‘)
So026_0fwd	ACCAGCACGGGCTTGGTGCTC
So026_0rev	CTCGCGCTCTACGCCGTGCAG
So026_1fwd	GGAGCTGTCGGAGGTGACCTC
So026_1rev	CCCCACATCGCGGCGACCATC
So026_2fwd	CCGTGCCCGTCCTGTTCCAGC
So026_2rev	GAGCTCCTGCGCGACGCTGAG
So026_3fwd	CAGCTGATGGCCACCGAGGAC
So026_3rev	AACCTCATGCGGGCCGATCAC
So026_3rev	GTGATCGGCCCGCATGAGGTT
So026_4fwd	GAGACGGCCAATTCCACCGCGGAC
So026_4rev	GACCCCGAGCTTGATCATCGCGTCCAG
So026_5fwd	GCGGTGGTGCGGCACACGGAG
So026_5fwd	GCGGTGGTGCGGCACACGGAG
So026_5rev	AGGTCTCCTGTCGCCATCCGATGC
So026_5rev	GCATCGGATGGCGACAGGAGACCT
So026_6fwd	GCCGCGTTCGGGGACGCCGGGGAG
So026_6rev	CTTGTGCGCGATGAGCTTCTTCTCGACCAG
So026_7fwd	AGCATGAGCCGGTGCACCGCG
So026_7rev	CGCCAGCACGTCGGGTATCAC
So026_8fwd	GGTCTGGATGTCCACCGCCAC
So026_8rev	ACGCACCAGTCGGTCGGGATC
So026_9fwd	CGCGATCGGCACGATCCAGTG
So026_9rev	GAGCTCGGCATCGCCACGAAG

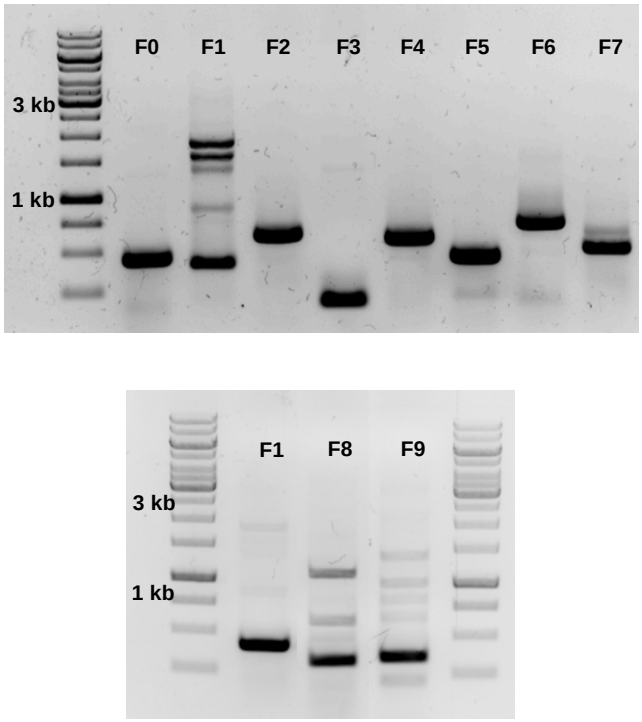


Figure S1: Agarose gel images showing the PCR amplicons obtained in a polymerase chain reaction using the respective primer pairs shown in Table S1

Table S2: Expected fragment size for amplicons produced in the PCR

fragment	size (bp)
F0	461
F1	440
F2	647
F3	273
F4	613
F5	490
F6	753
F7	415
F8	393
F9	362

5.6.2 Restriction sites used in the maracen project

Table S3: Restriction sites and their intended use in the project. In the last two columns the abundance in the native PUFA and *mrc* gene cluster is listed. All restriction sites used were removed from the maracen biosynthetic gene cluster by appropriate point mutations using computational methods.

R-site	use	Enzyme	Recognition sequence	overhang		methylation	Close to end cleavage					No. of sites in mrc	
							1	2	3	4	5	with pfa	w/o pfa
1	R3_unique RS mrc	*Apa*I	GGGCC^C	sticky (4)	37°C	Dcm (impaired)	xxx	xxx	xxx	xxx	xxx	17	10
2	R_{rechts} unique	*Avr*II	C^CTAGG	sticky(4)	37°C	/	xx	xx	xxx	xxx	xxx	0	0
3	ori exchange	*Bgl*II	A^GATCT	sticky(4)	37°C	EcoBI (maybe, impaired)	/	xxx	xxx	xxx	xxx	10	9
4	6x R-sites (backup)	*Bsi*WI	C^GTACG	sticky(4)	37°C	/	x	xx	xxx	xxx	xxx	13	5
5	R5_unique RS mrc	*Bsp*EI	T^CCGGA	sticky(4)	37°C	/	/	xxx	xxx	xxx	xxx	13	10
6	6x R-sites	*Eco*RI	G^AATTC	sticky(4)	37°C	/	x	x	xx	xx	xxx	2	2
7	vector degradation	*Hind*III	A^AGCTT	sticky(4)	37°C	EcoBI (maybe, impaired)	/	x	xxx	xxx	xxx	1	1
8	ori exchange	*Hpa*I	GTT^AAC	blunt	37°C	EcoBI (maybe, blocked)	xx	xxx	xxx	xxx	xxx	0	0
9	6x R-sites	*Kpn*I	GGTAC^C	sticky(4)	37°C	/	x	xxx	xxx	xxx	xxx	7	6
10	R2_unique RS mrc	*Mlu*I	A^CGCGT	sticky(4)	37°C	/	xx	xx	xxx	xxx	xxx	16	10
11	R1_unique RS mrc	*Msc*I	TGG^CCA	blunt	37°C	Dcm (maybe, blocked)	x	xxx	xxx	xxx	xxx	11	7
12	R4_unique RS mrc	*Nco*I	C^CATGG	sticky(4)	37°C	/	/	xxx	xxx	xxx	xxx	19	14
13	R_{links} unique	*Nde*I	CA^TATG	sticky(2)	37°C	/	x	xx	xxx	xxx	xxx	0	0
14	transfer cassette exchange	*Not*I	GC^GGCCGC	sticky(2)	37°C	/	xx	xxx	xxx	xxx	xxx	12	3
15	6x R-sites	*Nsi*I	ATGCA^T	sticky(4)	37°C	/	xx	xxx	xxx	xxx	xxx	0	0
16	pfa cloning	*Pac*I	TTAAT^TAA	sticky(2)	37°C	/	xx	xxx	xxx	xxx	xxx	0	0
17	6x R-sites	*Pci*I	A^CATGT	sticky(4)	37°C	/	/	xxx	xxx	xxx	xxx	4	3
18	pfa cloning	*Psi*I	TTA^TAA	blunt	37°C	EcoKI (maybe, not det.)	/	x	xx	xx	xxx	0	0
19	R9_unique RS mrc	*Pvu*I	CGAT^CG	sticky(2)	37°C	/	xx	xx	xxx	xxx	xxx	70	37
20	6x R-sites	*Pvu*II	CAG^CTG	blunt	37°C	/	xxx	xxx	xxx	xxx	xxx	10	7
21	R6_unique RS mrc	*Sal*I	G^TCGAC	sticky(4)	37°C	/	xx	xx	xxx	xxx	xxx	40	20
22	transfer cassette exchange	*Sca*I	AGT^ACT	blunt	37°C	EcoBI (maybe, blocked)	x	xxx	xxx	xxx	xxx	2	2
23	6x R-sites	*Spe*I	A^CTAGT	sticky(4)	37°C	EcoKI (maybe, not det.)	xxx	xxx	xxx	xxx	xxx	0	0
24	ori exchange	*Ssp*I	AAT^ATT	blunt	37°C	/	x	xxx	xxx	xxx	xxx	0	0
25	transfer cassette exchange	*Swa*I	ATTT^AAAT	blunt	30°C	/	x	xxx	xxx	xxx	xxx	0	0
26	transfer cassette exchange	*Xba*I	T^CTAGA	sticky(4)	37°C	Dam (maybe, blocked)	xx	xxx	xxx	xxx	xxx	0	0
27	R8_unique RS mrc	*Xho*I	C^TCGAG	sticky(4)	37°C	/	x	xxx	xxx	xxx	xxx	59	35
28	R7_unique RS mrc	*Xma*I	C^CCGGG	sticky(4)	37°C	/	xx	xxx	xxx	xxx	xxx	26	19
Typ IIs cutter													
29	splitter removal	*Bsa*I	GGTCTC(1/5)	5' (4 nucleotides)								28	18

5.6.3 Design of the multi-purpose cloning vector pUCmrc

For design of the multi-purpose cloning vector pUCmrc, several restriction sites had to be removed from the sequence, which in coding sequences were removed via silent point mutations. Table S4 shows the positon of the respective restriction site and its sequence within the respective coding sequence (bold and underlined), as well as the resulting peptide sequence. The last column explains how the relevant codon was altered in order to remove the recognition site for the respective restriction enzyme.

Table S4: Restriction sites removed in the pEXPmrc vector by silent mutations. The table shows the respective restriction site, the DNA sequence surrounding it, the peptide built from that sequence fragment and how the bases were mutated in order to maintain the coding sequence and remove the restriction site.

pUC18

	R-site	**Sequence**	**Peptide**	**Mutation**
Alpha				
Pos 14-20	*Eco*RI	AC**G AAT TC**G	TNS	Asp AAT >Asp AAC
Pos 26-32	*Kpn*I	TC**G GTA CC**C	SVP	Val GTA > Val GTC
Pos 41-47	*Xba*I	CC**T CTA GA**G	PLE	Leu CTG > Leu CTC
Pos 47-53	*Sal*I	GA**G TCG AC**C	EST	Ser TCG > Ser TCC
Pos 53-59	*Pst*I	AC**C TGC AG**G	TCR	Cys TGC > Cys TGT
Pos 65-71	*Hind*III	GC**A AGC TT**G	ASL	Leu TTG > Leu CTG
Pos 158-164	*Pvu*II	G**CC AGC TG**G	ASW	Ala GCC > Ala GCA
Pos 281-284	*Nde*I	CG**C ATA TG**G	RIW	Ile ATA > Ile ATC
AmpR				
Pos 304-310	*Sca*I	G**AG TAC T**CA	EYS	Glu GAG > Glu GAA
Pos714-720	*Bsa*I	G**GG TCT C**GC	GSR	Ser TCT > Ser TCC
pUC ori				
5’ end	*Pci*I	**ACATGT**		ACATGT > ACATGA

5.6.4 Design of the expression vector pEXPmrc

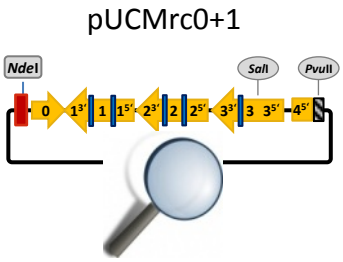
For design of the multi-purpose cloning vector pEXPmrc, several restriction sites had to be removed from the sequence, which in coding sequences were removed via silent point mutations. Table S5 shows the positon of the respective restriction site and its sequence within the respective coding sequence (bold and underlined), as well as the resulting peptide sequence. The last column explains how the relevant codon was altered in order to remove the recognition site for the respective restriction enzyme.

Table S5: Restriction sites removed in the pUCmrc vector by silent mutations. The table shows the respective restriction site, the DNA sequence surrounding it, the peptide built from that sequence fragment and how the bases were mutated in order to maintain the coding sequence and remove the restriction site. In the case of the zeocin resistance cassette no silent mutation was possible thus the threonine upstream the start codon was changed into a serine.

pEXPmrc

	R-site	**Sequence**	**Peptide**	**Mutation**
ZeoR				
Pos. -2-6	*Nco*I	A**CC ATG G**CC	TMA	Thr ACC > Ser AGC
Pos. 2-6	*Msc*I	**ATG GCC** AAG	MAK	Ala GCC > Ala GCG
Pos 88-94	*Xma*I	T**CC CGG G**AC	SRD	R: CGG > CGC
TD1				
Pos 11-17	*Apa*I	AA**GGGCCC**G		
Pos 33-39	*Apa*I	TCC**GGGCCC**		
Mx8				
Pos 79-85	*Mlu*I	G**AC GCG T**TC	DAF	Ala GCG > Ala GCA
Pos 248-254	*BspE*I	CG**T CCG GA**C	RPD	Pro CCG > CCC
Pos 359 -365	*Sal*I	GA**G TCG AC**T	EST	Ser TCG > Ser TCC
Integrase				
Pos 312-318	*Mlu*I	**ACG CGT** CGG	TRR	Thr ACG > Thr ACC
Pos 1050 - 1056	*Sal*I	**GTC GAC** GCC	VDA	Asp GAC > Asp GAT
Pos 1138-1144	*Nco*I	A**CC ATG G**CC	TMA	Thr ACC > Thr ACG
Pos 1503 - 1511	*Not*I	**GCG GCC GC**C	AAA	Ala 2 GCC > Ala GCG

5.6.5 Detailed description of constructional design including splitter elements exemplarily described for construct pUCMrc0+1



Construct pUCMrc0+1 consists of the *mrc0,1,2* and the 3' end plus the core sequence of gene *mrc3*. Five splitter elements with five unique restriction sites are engineered in the sequence. The DNA fragment is flanked by the unique restriction sites R_{uL} (*Nde*I) and R_2 (*Pvu*II). There are two types of splitter elements used in the sequence: type A (conventional SEs) has two recognition sites for the type IIS cutter *Bsa*I which frame a unique restriction site. Splitting and re-ligation of the construct leads to the native sequence without any alterations. In case B, there is a restriction enzyme recognition site encoded in the *Bsa*I overhang sequence; this restriction site will remain even after de-splitting and-re-ligation. Type B splitter elements are mainly used at the borders of synthetic fragments. To obtain the desired specific fusion site for every DNA fragment after *Bsa*I digest, a 1bp spacer had to be introduced in the splitter element, too, as *Bsa*I cuts 1bp distant from its recognition site on the leading strand and in 5 bp distance on the lagging strand resulting in a 4 bases overhang on the lagging strand.

Eco31I (*Bsa*I) 5'...G G T C T C $(N)_1$↓...3'

3'...C C A G A G $(N)_5$↑..

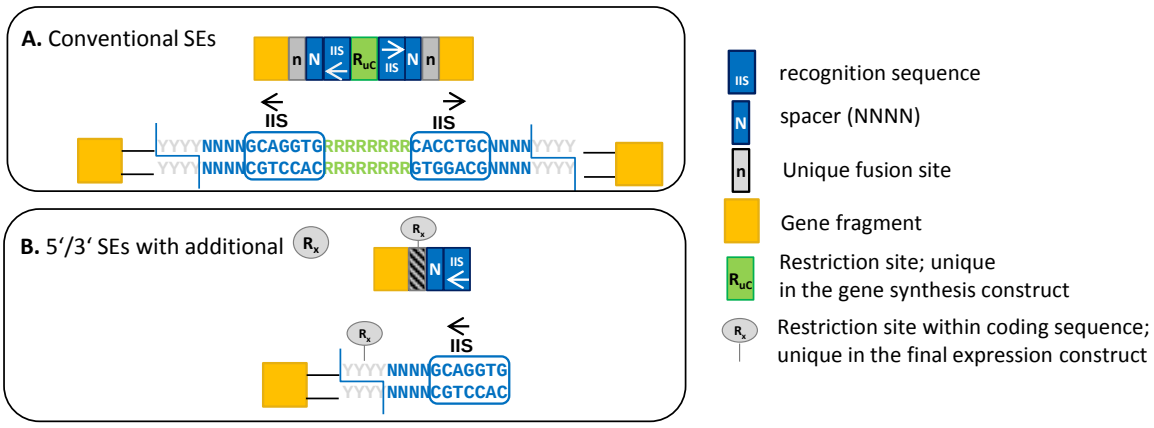


Figure S2: Composition of a splitter element with a type IIS cutter, cleaving in N base pairs distance to its recognition site. B If a unique restriction site engineered in the coding sequence was to remain present after desplitting the respective recognition sequence was introduced in the unique fusion site.

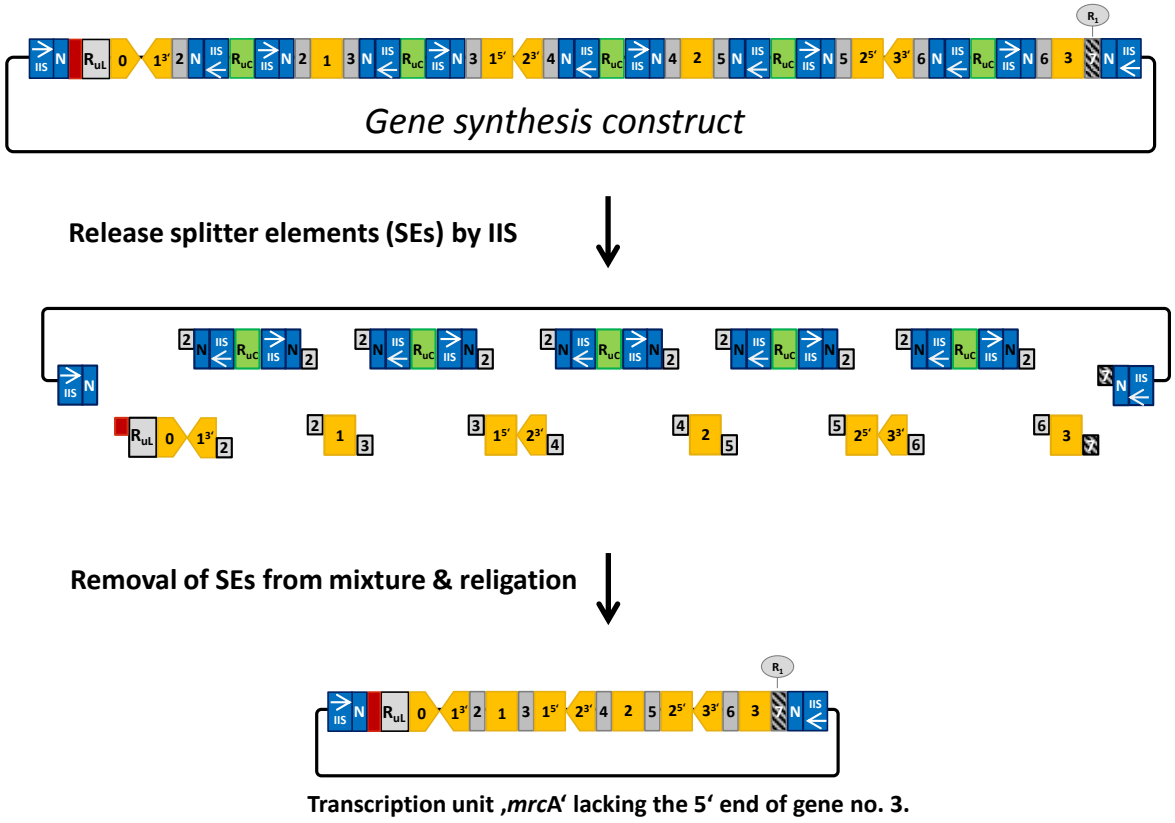


Figure S3: Detailed assembly strategy exemplified for the construct pUCmrc0+1. The construct is de-split in pUCmrc; splitter elements are removed from the mixture prior to re-ligation of the fragment pieces into pUCmrc.

Position and overhang sequences of the typeIIS cutters' restriction site were chosen in way that unambiguous fusion sites for later re-ligation resulted. Additionally, for every splitter element, a spacer sequence of 4 bp for every *Bsa*I restriction site had to be added. Details on the splitter elements' sequences are listed in Table S7 (chapter 5.6.7).

5.6.6 Overhang sequences within the coding sequence generated after *Bsa*I digest for all splitter elements

Table S6: Unique fusion sites generated after 'de-splitting' with *Bsa*I. The splitter elements' positions were chosen in a way that the resulting overhang provides a unique fusion site for re-ligation of the fragments. (Numbering from 1-25 according to the order of splitter elements in genes 1-18)

No.	Sequence	Position
1	ACGG	3‘ orf1
2	CGAG	5‘ orf1
3	CACG	3‘ orf2
4	GCTC	5‘ orf2
5	GATA	3‘ orf3
6	CTGC	3‘ orf4
7	CACG	3‘ orf5
8	GGAC	5‘ orf5
9	AGGT	3‘ orf6
10	AGCC	3‘ orf7
11	TGCT	5‘ orf8
12	CTGC	3‘ orf8
13	ACAG	3‘ orf9
14	GTCG	5‘ orf11
15	GGTG	3‘ orf11
16	ATGG	3‘ orf12
17	TGGC	3‘ orf13
18	GAGC	5‘ orf13
19	CGTC	3‘ orf14
20	GCAG	5‘ orf14
21	CTGC	3‘ orf15
22	CTGG	3‘ orf16
23	CCGT	3‘ orf17
24	GAAC	5‘ orf17
25	CGTA	3‘ orf18

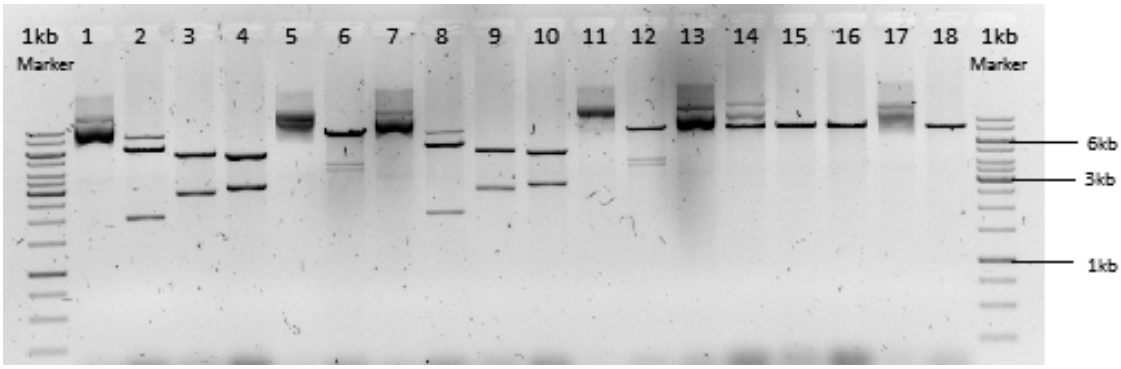
5.6.7 Sequence features of splitter elements

Table S7: Table with detailed description of all splitter elements engineered in the maracen biosynthetic gene cluster. The relative position in the gene cluster, the respective restriction site engineered in the splitter element as well as the "filling-nucleotide N" are specified.

No.	Fusion site	Position	R-site	Recognition sequence	Splitter element **FS**-N-BsaI-*R-site*-BsaI-N-**FS**	Splitter element **FS**-N-BsaI-*R-site*-BsaI-N-**FS**
1	ACGG	3‘ orf1	*Afl*II	C*TTAAG	**ACGG**N GAGACC *CTTAAG* GGTCTC N **ACGG**	**ACGG**T GAGACC *CTTAAG* GGTCTC G **ACGG**
2	CGAG	5‘ orf1	*Bsi*WI	C*GTACG	**CGAG**N GAGACC *CGTACG* GGTCTC N **CGAG**	**CGAG**C GAGACC *CGTACG* GGTCTC A **CGA G**
3	CACG	3‘ orf2	*Eco*RI	G*AATTC	**CACG**N GAGACC *GAATTC* GGTCTC N **CACG**	**CACG**T GAGACC *GAATTC* GGTCTC G **CACG**
4	GCTC	5‘ orf2	*Kpn*I	GGTAC*C	**GCTC**N GAGACC *GGTACC* GGTCTC N **GCTC**	**GCTC**T GAGACC *GGTACC* GGTCTC T **GCTC**
5	GATA	3‘ orf3	*Nsi*I	ATGCA*T	**GATA**N GAGACC *ATGCAT* GGTCTC N **GATA**	**GATA**C GAGACC *ATGCAT* GGTCTC A **GATA**
6	CTGC	3‘ orf4	*Afl*II	C*TTAAG	**CTGC**N GAGACC *CTTAAG* GGTCTC N **CTGC**	**CTGC**T GAGACC *CTTAAG* GGTCTC A **CTGC**
7	CACG	3‘ orf5	*Bsi*WI	C*GTACG	**CACG**N GAGACC *CGTACG* GGTCTC N **CACG**	**CACG**C GAGACC *CGTACG* GGTCTC G **CACG**
8	GGAC	5‘ orf5	*Eco*RI	G*AATTC	**GGAC**N GAGACC *GAATTC* GGTCTC N **GGAC**	**GGAC**T GAGACC *GAATTC* GGTCTC A **GGAC**
9	AGGT	3‘ orf6	*Kpn*I	GGTAC*C	**AGGT**N GAGACC *GGTACC* GGTCTC N **AGGT**	**AGGT**C GAGACC *GGTACC* GGTCTC G **AGGT**
10	AGCC	3‘ orf7	*Afl*II	C*TTAAG	**AGCC**N GAGACC *CTTAAG* GGTCTC N **AGCC**	**AGCC**A GAGACC *CTTAAG* GGTCTC G **AGCC**
11	TGCT	5‘ orf8	*Bsi*WI	C*GTACG	**TGCT**N GAGACC *CGTACG* GGTCTC N **TGCT**	**TGCT**C GAGACC *CGTACG* GGTCTC G **TGCT**
12	CTGC	3‘ orf8	*Eco*RI	G*AATTC	**CTGC**N GAGACC *GAATTC* GGTCTC N **CTGC**	**CTGC**T GAGACC *GAATTC* GGTCTC G **CTGC**
13	ACAG	3‘ orf9	*Kpn*I	GGTAC*C	**ACAG**N GAGACC *GGTACC* GGTCTC N **ACAG**	**ACAG**C GAGACC *GGTACC* GGTCTC G **ACAG**
14	GTCG	5‘ orf11	*Nsi*I	ATGCA*T	**GTCG**N GAGACC *ATGCAT* GGTCTC N **GTCG**	**GTCG**T GAGACC *ATGCAT* GGTCTC A **GTCG**
15	GGTG	3‘ orf11	*Spe*I	A*CTAGT	**GGTG**N GAGACC *ACTAGT* GGTCTC N **GGTG**	**GGTG**C GAGACC *ACTAGT* GGTCTC A **GGTG**
16	ATGG	3‘ orf12	*Afl*II	C*TTAAG	**ATGG**N GAGACC *CTTAAG* GGTCTC N **ATGG**	**ATGG**C GAGACC *CTTAAG* GGTCTC G **ATGG**
17	TGGC	3‘ orf13	*Bsi*WI	C*GTACG	**TGGC**N GAGACC *CGTACG* GGTCTC N **TGGC**	**TGGC**T GAGACC *CGTACG* GGTCTC A **TGGC**
18	GAGC	5‘ orf13	*Eco*RI	G*AATTC	**GAGC**N GAGACC *GAATTC* GGTCTC N **GAGC**	**GAGC**A GAGACC *GAATTC* GGTCTC A **GAGC**
19	CGTC	3‘ orf14	*Kpn*I	GGTAC*C	**CGTC**N GAGACC *GGTACC* GGTCTC N **CGTC**	**CGTC**T GAGACC *GGTACC* GGTCTC G **CGTC**
20	GCAG	5‘ orf14	*Nsi*I	ATGCA*T	**GCAG**N GAGACC *ATGCAT* GGTCTC N **GCAG**	**GCAG**C GAGACC *ATGCAT* GGTCTC A **GCAG**
21	CTGC	3‘ orf15	*Spe*I	A*CTAGT	**CTGC**N GAGACC *ACTAGT* GGTCTC N **CTGC**	**CTGC**T GAGACC *ACTAGT* GGTCTC G **CTGC**
22	CTGG	3‘ orf16	*Eco*RI	G*AATTC	**CTGG**N GAGACC *GAATTC* GGTCTC N **CTGG**	**CTGG**C GAGACC *GAATTC* GGTCTC A **CTGG**
23	CCGT	3‘ orf17	*Kpn*I	GGTAC*C	**CCGT**N GAGACC *GGTACC* GGTCTC N **CCGT**	**CCGT**C GAGACC *GGTACC* GGTCTC A **CCGT**
24	GAAC	5‘ orf17	*Nsi*I	ATGCA*T	**GAAC**N GAGACC *ATGCAT* GGTCTC N **GAAC**	**GAAC**T GAGACC *ATGCAT* GGTCTC A **GAAC**
25	CGTA	3‘ orf18	*Spe*I	A*CTAGT	**CGTA**N GAGACC *ACTAGT* GGTCTC N **CGTA**	**CGTA**G GAGACC *ACTAGT* GGTCTC A **CGTA**

5.6.8 Analysis of results from control digest after splitter removal exemplarily for construct pMrc0+1

The control digests of ‘de-split’ and re-ligated constructs was performed using the enzyme combinations described in Table S8; one targeted restriction site was present in the construct backbone whilst the other restriction site was located in the splitter element that was to be removed. Subsequently samples were analyzed by agarose gel electrophoresis. With the samples also a non-treated control was run on the gel. In case of complete removal of all splitter elements the band pattern of the loaded sample should not differ from the reference, as only one distinct band representing the linearized construct should be seen. Below this is exemplarily shown for three samples of the construct Mrc0+1. In the first and in the second sample, splitter elements were not removed successfully thus a distinct band pattern described in Table S8 could be observed. In the case of sample three, all splitter elements were removed and consequently only bands representing the linearized construct can be observed. As not for all constructs adequate conditions for such a double digest could be found, promising candidate constructs were also send for sequencing to ensure complete removal of all splitter elements.



1 = Mrc 0+1 reference, # 1
2 = Mrc 0+1 (*Afl*II, *Not*I), # 1
3 = Mrc 0+1 (*BsiW*I, *Pae*I), # 1
4 = Mrc 0+1 (*EcoR*I, *Not*I), # 1
5 = Mrc 0+1 (*Kpn*I, *Pae*I), # 1
6 = Mrc 0+1 (*Nsi*I, *EcoR*V), # 1
7 = Mrc 0+1 reference, # 2
8 = Mrc 0+1 (*Afl*II, *Not*I), # 2
9 = Mrc 0+1 (*BsiW*I, *Pae*I), # 2
10 = Mrc 0+1 (*EcoR*I, *Not*I), # 2
11 = Mrc 0+1 (*Kpn*I, *Pae*I), # 2
12 = Mrc 0+1 (*Nsi*I, *EcoR*V), # 2
13 = Mrc 0+1 reference, # 3
14 = Mrc 0+1 (*Afl*II, *Not*I), # 3
15 = Mrc 0+1 (*BsiW*I, *Pae*I), # 3
16 = Mrc 0+1 (*EcoR*I, *Not*I), # 3
17 = Mrc 0+1 (*Kpn*I, *Pae*I), # 3
18 = Mrc 0+1 (*Nsi*I, *EcoR*V), # 3

Figure S4: Agarose gel showing the analysis of three samples of construct pMrc0+1 after control digest

Control digest scheme after splitter removal

Construct	*R-site*	*NdeI, AvrII*	*Control digest*	*band pattern (bp)*
mrc0,11	*Afl*II(BspTI)	*5313 2722*	*No1*	*6155,2002*
(8149)	*Bsi*WI(Pfl23II)		*No.6*	*5364, 2793*
	*Eco*RI		*No.2*	*5178, 2979*
	*Kpn*I		*No.7*	*6108, 2049*
	*Nsi*I(Mph1103I)		*No.3*	*4280, 3873*
mrc1	*Afl*II(BspTI)	*3537 2722*	*No1*	*6155, 226*
(6373)	*Bsi*WI(Pfl23II)		*No.6*	*3588, 2793*
	*Eco*RI		*No.2*	*5178, 1203*
	*Kpn*I		*No.7*	*4332, 2049*
	*Nsi*I(Mph1103I)		*No.3*	*4280, 2097*
mrc2	*Afl*II(BspTI)	*3470 2722*	*No1*	*5438, 852*
(6282)	*Bsi*WI(Pfl23II)		*No.8*	*3817, 2473*
	*Eco*RI		*No.2*	*4266, 2024*
	*Kpn*I		*No.9*	*4997, 1293*
mrc3	*Afl*II(BspTI)	*6146 2722*	*No1*	*8484, 530*
(9006)	*Bsi*WI(Pfl23II)		*No.10*	*6872, 2139*
	*Eco*RI		*No.2*	*7505, 1509*
	*Kpn*I		*No.*	-
	*Nsi*I(Mph1103I)		*No.3*	*4868, 4142*
	*Spe*I(BcuI)		*No.4*	*5680, 3330*
mrc4	*Afl*II(BspTI)	*7986 2722*	*No1*	*8749, 2105*
(10846)	*Bsi*WI(Pfl23II)		*No.11*	*8441, 2409*
	*Eco*RI		*No.2*	*7169, 3685*
	*Kpn*I		*No.*	-
	*Nsi*I(Mph1103I)		*No.13*	*8077,2777*
	*Spe*I(BcuI)		*No.4*	*5784, 5066*
mrc5	*Eco*RI	*6414 2722*	*No.2*	*6288, 2946*
(9226)	*Kpn*I		*No.*	-
	*Nsi*I(Mph1103I)		*No.3*	*6677, 2553*
	*Spe*I(BcuI)		*No.4*	*6942, 2288*

Table S8: Restriction enzymes combinations for control digest. Instructions for enzyme reactions are given above.

No.	Instructions
1	Buffer O BspTI (AflII); NotI Incubate at 37°C
2	Buffer O EcoRI; NotI Incubate at 37°C
3	Buffer R Eco32I (EcoRV); Mph1103I (NsiI) Incubate at 37°C
4	Buffer G 2-fold excess of BcuI (SpeI) 2-fold excess of Eco32I (EcoRV) Incubate at 37°C
5	Buffer 1X Tango™ BcuI (SpeI); 2-fold excess of PaeI (SphI) Incubate at 37°C
6	Buffer 1X Tango™ 2-fold excess of PaeI (SphI); Pfl23II (BsiWI) Incubate at 37°C
7	Buffer KpnI KpnI;2-fold excess of PaeI (SphI) Incubate at 37°C
8	Buffer G BamHI; 2-fold excess of Pfl23II (BsiWI) Incubate at 37°C
9	Buffer BamHI BamHI; 2-fold excess of KpnI Incubate at 37°C
10	Buffer 1X Tango™ 2-fold excess of Eam1105I (AhdI); Pfl23II (BsiWI) Incubate at 37°C
11	Buffer G 2-fold excess of Eco32I (EcoRV); 2-fold excess of Pfl23II (BsiWI) Incubate at 37°C
12	Buffer R Mph1103I (NsiI); 2-fold excess of NdeI Incubate at 37°C
13	Buffer R Mph1103I (NsiI); 2-fold excess of XmaJI (AvrII) Incubate at 37°C

5.6.9 Sequencing of ‘de-split’ constructs

After verification of ‘de-split’ and re-ligated constructs via restriction analysis as described in chapter 5.6.8, the absence of the splitter elements and correct re-assembly of the cluster fragments was verified by sequencing. For this, primers located 300-400 bp up- and downstream of the former splitter element position were designed (Table S9). Analysis of the obtained Sanger sequencing results confirmed the complete removal of all splitter elements and reconstitution of intact CDS for all constructs.

Table S9: List of primers used for sequencing particular stretches of ‘de-split’ and re-ligated pMrc constructs

Primer	5’-Sequence-3’
mrc1_1rev	CCGCGTCCGCTGGCTGAACTC
mrc1_2fwd	CGTAGGGGCAGCATCCACGAG
mrc1_2rev	CGGTGACCGGGACCGTGATCG
mrc1_3fwd	GCGGCTGTGCACCGTCAAGAG
mrc1_3rev	GCCATGATCCGCAAGGACGACG
mrc1_4fwd	GTGAACGGCGTCTCCGGGAAG
mrc1_4rev	GCGGTGAGCTCAGGTCGATTGC
mrc1_5fwd	CTTGACCCGCCAGGATGGCAC
mrc2_1rev	GAGGGATCAGACCCCTTGCTGC
mrc2_2fwd	CCGCACGTCAGCCGGAGCTG
mrc2_2rev	AGGGGCGCGATGTTCGGCTCG
mrc2_3fwd	CACGCGTCGCTGAGATTGTCGC
mrc2_3rev	AGCCGGTCGTGGGGACGATCC
mrc3_1rev	GCACCTCGCGCGTGATGTAGC
mrc3_2fwd	GACCGGGCGTACCCATGTCTG
mrc3_2rev	CGCGGCGCTCGATGAAGGACG
mrc3_5fwd	GGATCCCGATCGACGTGGTCG
mrc3_5rev	CCTTCATCGGGGCGACCTCGC
mrc3_6fwd	CGGAGCAGGCCAGGCTTGTCG
mrc3_6rev	GGCAGGAGCTCGCGCACATGG
mrc3_7fwd	CGCTCGCCGGTCGCAGATCTC
mrc3_7rev	TCGACAGGGGGACCACCGCTC
mrc4_3fwd	CAGGACATGCCGTTGCCGACG
mrc4_3rev	CGGGCCACTACATCCCGTTCG
mrc4_5fwd	CGCGCTTGCGCTGCCAGTCC
mrc4_5rev	CCGTCGAGGACCTCTGCGACC
mrc4_7fwd	GCCATGCGCCGGATGATCATGG
mrc4_7rev	TCGAGGGCGCCGTCAACTACG
mrc5_3fwd	AGCAGGGGATCGACCTGACCG
mrc5_3rev	CCTCGTAGTCGCGGAGGATCC
mrc5_5fwd	CTCGATCGGGTCCGGCTACG
mrc5_5rev	CCAGATAGGCGCACACCTCGC
mrc5_8fwd	GCGATCGTCCCGGAGATCGTG

5.6.10 Generation of the mutant *M. xanthus* DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa from *M. xanthus* DK1622 *ΔmchA*

To establish PUFA production in the heterologous host *Myxococcus xanthus* DK1622 *ΔmchA-tet*, the *pfa* gene cluster previously described by Gemperlein et al.(unpublished) was to be introduced into the host strain genome. For this, minor alterations of the construct were required as the selection marker zeocin present on the construct interfered with zeocin cassette on the maracen expression construct. Furthermore the right homology arm "H-right" was chosen as integration locus in DK1622 *ΔmchA-tet* (see section 5.3.6), whilst the original PUFA construct pHybpfa1-mx9.2 was designed for integration into the mx9 attachment site.

Thus a new backbone was designed. For this the commercially available cloning vector pACYC184[38] was used as foundation. In a PCR reaction using the primers H-right_fwd (5'-GATACATGATCAGTTTAAACACTAGTCGATGTCGCGCAGGTACG-3') and H-right_rev (5'-ATCTGTAAGCTTTTAATTAATGTATCTTATAACATCCAGCAGAAGGCCATC-3') were used to amplify the 1 kb H-right homology arm, which was subsequently cloned into pACYC184 to replace its tetracycline resistance gene yielding in p15ACm_H-right. After linearization of p15ACm_H-right with *Psi*I, *Pac*I, exchange of the vector backbone of pHybpfa1-mx9.2 was simple to execute. The resulting construct p15ACm_H-right_pfa could then be transformed into DK1622 *ΔmchA-tet* for integration of the *pfa* gene cluster into the H-right locus; the kanamycin resistance gene *pfa* gene cluster construct allowed for selection of positive mutants. The obtained mutants were subsequently verified by genotypic and phenotypic analysis as described in chapter 5.6.12 .

5.6.11 Genotypic verification of DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa

For genotypic verification of the mutant DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa PCR was carried out. The primers used are indicated in figure S4 and their sequence is shown in Table S10. The expected band pattern resulting from the primer combination listed in Table S11 confirmed the correct integration of the *pfa* gene cluster at the target locus.

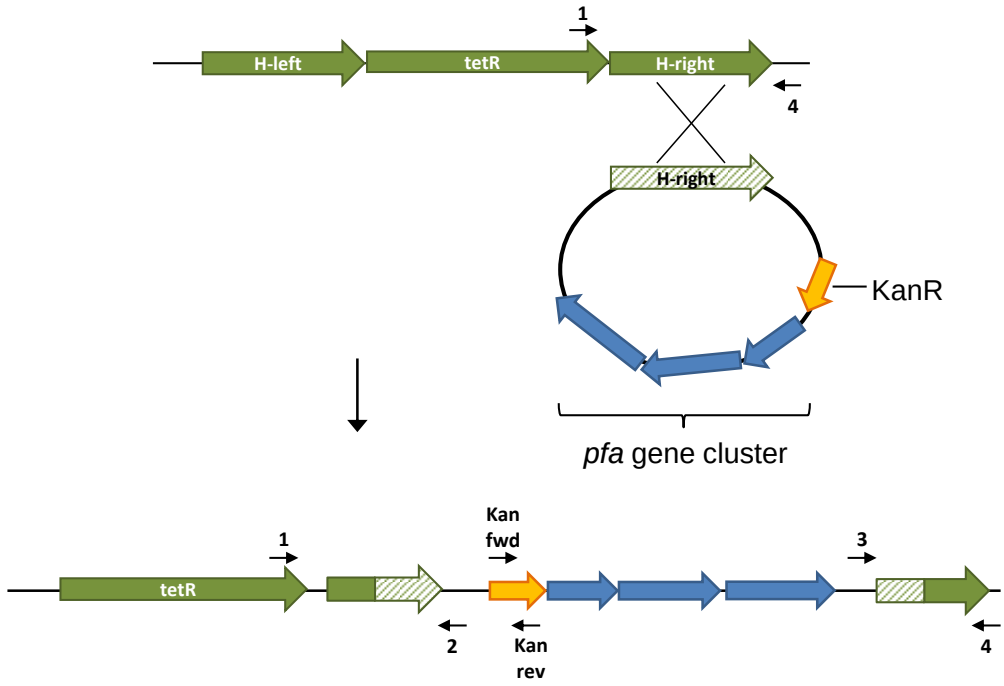


Figure S5: .Schematic overview on the gene organization at the former myxochromide gene cluster locus showing the target locus *H-right* and the construct pCm15A_H-right_pfa designed for its integration of the *pfa* gene cluster via single crossover insertion. B. Genetic organization of the mutant strain DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa containing the integrated pCm15A_H-right_pfa. For verification of the genotype, PCR analysis was used.

Table S10: Primers used for verification of the DK1622 *ΔmchA-tet*::H-right pCm15A_H-right_pfa mutant

Primer	sequence
1	TCTTGCGGAGAACTGTGAATGCGC
2	CCTGAGGCCAGTTTGCTCAGG
3	GCTCAGTATTGCCCGCTCCAC
4	GACGATCTTCAGCTTCACGACACC
Kan_fwd	ATTCCGGAATTGCCAGCTGGG
Kan_rev	TCAGAAGAACTCGTCAAGAAG

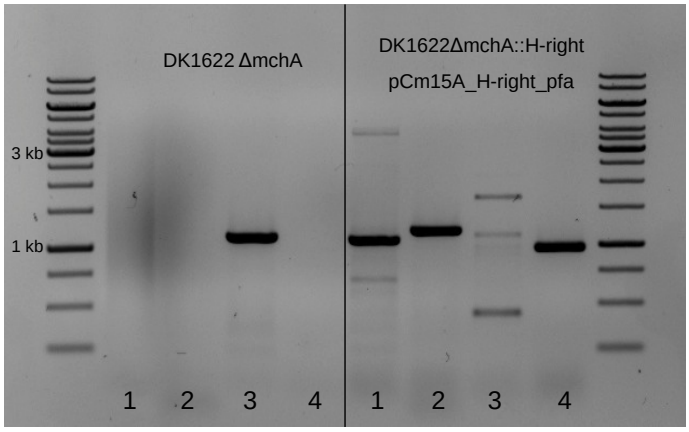


Figure S6: Genotypic verification of the DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa mutant by PCR. Primers were chosen as indicated in the figure above. Gel electrophoresis of the verification PCR experiments confirmed the expected genotype by showing DNA amplicons of anticipated size.

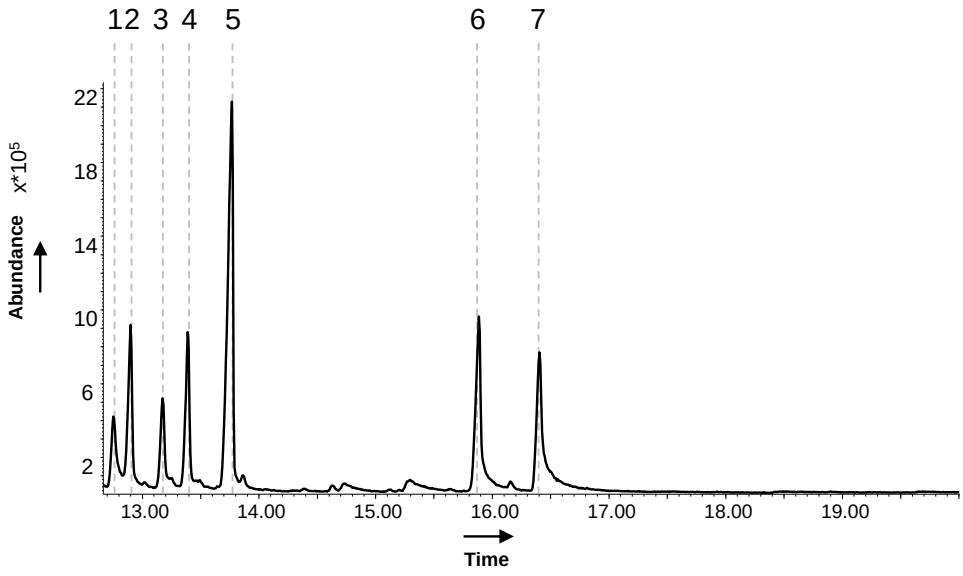
Table S111: Primer combinations used in the verification PCR and amplicon size expected from this experiment

	expected amplicon size (bp)	
	ΔmchA	**mutant**
1 & 2	/	~ 1075
3 & 4	/	~ 1130
1 & 4	1112	/
Kan_fwd&rev		933

5.6.12 Phenotypic analysis of DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa

Upon verification of the genotype of DK1622 *ΔmchA-tet::*H-right pCm15A_H-right_pfa, the mutant strain was grown in liquid media as described in section 5.1.8. Subsequently the production of fatty acids was assayed via GC-MS analysis of the respective fatty acid methyl-esters (FAME) as described in sections 5.1.13 and 5.1.14. As negative control, the host strain without the *pfa* gene cluster DK1622 *ΔmchA-tet* was grown and analysed in the same way.

A.



B.

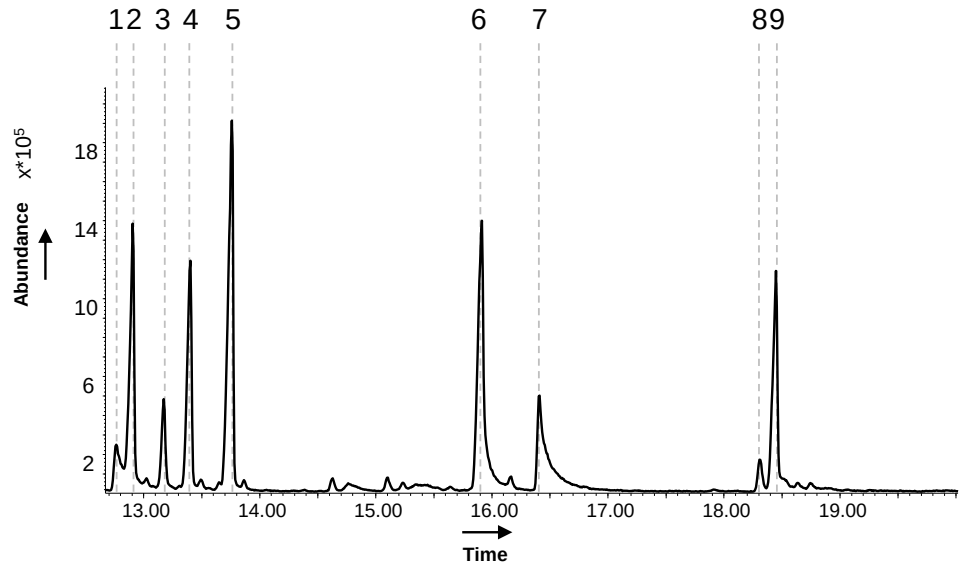


Figure S6: GC-MS traces of the extracts prepared from DK1622 *ΔmchA-tet* and DK1622 *ΔmchA-tet ::*H-right pCm15A_H-right_pfa. **A.** Reference chromatogram of an extract of DK1622 *ΔmchA-tet* producing the fatty acids 1-7 **B.** Chromatogram of an extract produced from the mutant strain DK1622 *ΔmchA-tet ::*H-right pCm15A_H-right_pfa additionally producing arachidonic acid (8) and EPA (9).

Table S12: Fatty acids detected via GC-MS analysis as FAMEs (listed according to the sequence of their occurrence in the chromatogram).

No.	fatty acid
1	16:0 3-OH
2	iso-17:2ω5,11 all cis
3	iso-17:1ω11cis
4	iso-17:1ω5cis
5	iso-17:0
6	iso-17:0 2-OH
7	iso-17:0 3-OH
8	20:4ω6,9,12,15 all cis arachidonic acid
9	20:5ω3,6,9,12,15 all cis eicosapentaenoic acid

5.7 References

[1] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–35.

[2] S. Renner, W. A. L. van Otterlo, M. Dominguez Seoane, S. Möcklinghoff, B. Hofmann, S. Wetzel, A. Schuffenhauer, P. Ertl, T. I. Oprea, D. Steinhilber, et al., *Nature* **2010**, *87131*, 585–592.

[3] K. Gerth, H. Steinmetz, G. Höfle, R. Jansen, *Angew. Chemie Int. Ed.* **2008**, *47*, 600–602.

[4] D. M. Bollag, P. a McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, C. M. Woods, *Cancer Res.* **1995**, *55*, 2325–2333.

[5] M. Herrmann, B. Böhlendorf, H. Reichenbach, *Angew. Chemie* **1998**, *110*, 1313–1315.

[6] C. Oßwald, G. Zipf, G. Schmidt, J. Maier, H. S. Bernauer, R. Müller, S. C. Wenzel, *ACS Synth. Biol.* **2014**, *3*, 759–772.

[7] S. C. Mutka, J. R. Carney, Y. Liu, J. Kennedy, *Biochemistry* **2006**, *45*, 1321–1330.

[8] I. Nickeleit, S. Zender, F. Sasse, R. Geffers, G. Brandes, I. Sörensen, H. Steinmetz, S. Kubicka, T. Carlomagno, D. Menche, et al., *Cancer Cell* **2008**, *14*, 23–35.

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

[10] P. C. Kienesberger, M. Oberer, A. Lass, R. Zechner, *J. Lipid Res.* **2009**, *50 Suppl*, S63–S68.

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

[12] P. a Wilson, S. D. Gardner, N. M. Lambie, S. a Commans, D. J. Crowther, *J. Lipid Res.* **2006**, *47*, 1940–1949.

[13] M. Péterfy, A. Manuscript, *Biochim. Biophys. Acta* **2012**, *1821*, 790–794.

[14] A. Liavonchanka, E. Hornung, I. Feussner, M. G. Rudolph, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 2576–2581.

[15] A. Liavonchanka, M. G. Rudolph, K. Tittmann, M. Hamberg, I. Feussner, *J. Biol. Chem.* **2009**, *284*, 8005–8012.

[16] A. Andreou, I. Feussner, *Phytochemistry* **2009**, *70*, 1504–1510.

[17] H. Porta, M. Rocha-Sosa, *Microbiology* **2001**, *147*, 3199–200.

[18] A. N. Grechkin, *Prostaglandins Other Lipid Mediat.* **2002**, *68-69*, 457–470.

[19] N. S. Yadav, a Wierzbicki, M. Aegerter, C. S. Caster, L. Pérez-Grau, a J. Kinney, W. D. Hitz, J. R. Booth, B. Schweiger, K. L. Stecca, *Plant Physiol.* **1993**, *103*, 467–476.

[20] C. Somerville, J. Browse, C. Somerville, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1991**, *42*, 467–506.

[21] B. Lemieux, M. Miquel, C. Somerville, J. Browse, *Theor. Appl. Genet.* **1990**, *80*, 234–240.

[22] J. Shanklin, J. E. Guy, G. Mishra, Y. Lindqvist, *J. Biol. Chem.* **2009**, *284*, 18559–18563.

[23] S. C. Andrews, *Biochim. Biophys. Acta - Gen. Subj.* **2010**, *1800*, 691–705.

[24] R. E. Minto, B. J. Blacklock, *Prog. Lipid Res.* **2008**, *47*, 233–306.

[25] P. H. Buist, *Nat. Prod. Rep.* **2004**, *21*, 249–262.

[26] M. Lee, M. Lenman, a Banaś, M. Bafor, S. Singh, M. Schweizer, R. Nilsson, C. Liljenberg, a Dahlqvist, P. O. Gummeson, et al., *Science* **1998**, *280*, 915–918.

[27] D. W. Reed, D. R. Polichuk, P. H. Buist, S. J. Ambrose, R. J. Sasata, C. K. Savile, A. R. S. Ross, P. S. Covello, *J. Am. Chem. Soc.* **2003**, *125*, 10635–10640.

[28] V. M. Dembitsky, M. Srebnik, *Prog. Lipid Res.* **2002**, *41*, 315–367.

[29] H. Mu, P. Sundin, C. Wesén, *TrAC Trends Anal. Chem.* **1997**, *16*, 274–286.

[30] W. Lam, S. Hahn, E. Ayanoglu, C. Djerassi, *J. Org. Chem.* **1989**, *54*, 3428–3432.

[31] N. M. Carballeira, a Emiliano, *Lipids* **1993**, *28*, 763–766.

[32] M. J. Garson, M. P. Zimmermann, M. Hoberg, R. M. Larsen, C. N. Battershill, P. T. Murphy, *Lipids* **1993**, *28*, 1011–1014.

[33] V. Magrini, M. L. Storms, P. Youderian, **1999**, *181*, 4062–4070.

[34] B. Julien, *J. Bacteriol.* **2003**, *185*, 6325–6330.

[35] A. a. Iniesta, F. García-Heras, J. Abellón-Ruiz, A. Gallego-García, M. Elías-Arnanz, *J. Bacteriol.* **2012**, *194*, 5875–5885.

[36] G. Bacteria, R. Simon, M. O. Connell, M. Labes, M. O'Connell, M. Labes, A. Pühler, *METHODS Enzymol.* **1986**, *118*, 640–659.

[37] H. B. Bode, M. W. Ring, G. Schwär, R. M. Kroppenstedt, D. Kaiser, R. Müller, *J. Bacteriol.* **2006**, *188*, 6524–6528.

[38] M. R. Green, J. Sambrook, *Molecular Cloning : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, **n.d.**

6. Discussion & Outlook

The work described in this thesis covers several aspects of natural product research, whereas the main focus is on the biosynthesis of myxobacterial secondary metabolites exhibiting promising bioactivity *in vitro* and *in vivo*. The main achievement of the endeavors reported in detail in chapters 2 to 4 are the elucidation of the biosynthetic pathway responsible for chlorotonil biosynthesis as well as further insights into the underlying biochemical processes. Moreover, initial efforts towards development of chlorotonil derivatives with advanced pharmacokinetic properties were carried out and met with success. Apart from these translational research aspects, the work on chlorotonil and anthracimycin also contributes significantly to fundamental research as the underlying biosynthetic machineries show several peculiarities that rise questions about the rules established for *trans*-AT PKS systems so far. Comparison of the two biosynthetic pathways provided the unique opportunity to investigate the architectural differences responsible for the production of two highly similar scaffolds, showing small but significant differences which are also reflected in the respective bioactivity. The outcome of this comparison challenges the currently valid model for polyketide biosynthesis by *trans*-AT PKS systems and thus suggests that the operational differences between *cis*- and *trans*-AT PKSs are bigger than anticipated.

Chapter 5, additionally deals with investigations of the biosynthetic machinery of the natural products maracen and maracin, following an innovative approach. Implementing a comprehensive concept linking metabolome and genome data, a model for maracen and maracin biosynthesis was elaborated based on the combined data. Making use of modern tools available from the steadily growing field of synthetic biology, the candidate biosynthetic pathway was analyzed, adapted and ultimately synthesized in defined fragments. This strategy allowed the rapid assembly of the alleged maracen/maracin biosynthetic machinery and also facilitates rapid straightforward alterations in the gene cluster. We thus paved the way for establishment of a platform for heterologous expression of the pathway, which at the same time enables detailed investigation of the involved enzymatic steps and also should allow for the production of derivatives via targeted genetic modifications.

6.1 Chlorotonil and anthracimycin – Biosynthesis and generation of derivatives

6.1.1 Biosynthesis of chlorotonil and anthracimycin

The polyketides chlorotonil and anthracimycin, both produced by a *trans*-AT polyketide synthase basically share the same carbon skeleton, whereas they show slight variations in stereochemistry. The major difference, however, is the gem-dichloro-1,3-dione structure uniquely seen in chlorotonil. Nevertheless, these two highly similar natural products originate from two different organisms from two different habitats. Whilst chlorotonil is produced by the terrestrial myxobacterium *Sorangium cellulosum* So ce1525, anthracimycin was initially isolated from a streptomycete living in a marine environment,[1,2] although it has been brought to our attention recently that this compound has also been reported from terrestrial streptomycetes (Wilkinson et al., *ACS Chem. Biol.* DOI: 10.1021/acschembio.5b00525). Looking through the variety of myxobacterial compounds isolated so far, secondary metabolites from marine sources showing strikingly similar structures have been reported occasionally. As an example, in the very recent case of the unusual peptide natural products microsclerodermins the compound class which was already known from a marine sponge has also been isolated form several myxobacteria of the genus *Jahnella* and *Sorangium*.[3,4] The finding of this compound class in terrestrial myxobacteria and marine sponges thus led to a “symbiont hypothesis” suggesting that a symbiont bacterium of the sponge might be responsible for the production of the microsclerodermins.[5] But this is by far not the only case. The cyclodepsipeptide chondramide produced by the myxobacterium *Chondromyces crocatus* Cmc5 highly resembles jaspamide which was isolated from a marine *Jaspis* sponge, suggesting that the underlying biosynthetic machinery should be highly similar as well.[6–8] Similarly, the bengamides were isolated from a *Jaspis* sponge as well as from the myxobacterium *Myxococcus virescens*.[9,10] Another example is the macrolides salicylihalamide and apicularen, found in a sponge of *Halicona species* in the first case and in a *Chrondomyces species* in the second case.[11,12] And also in case of renieramycin and saframycin the same observation was made: saframycin MX1 is produced by the myxobacterium *Myxococcus xanthus* Mx x48, whereas the related renieramycin was isolated from a *Reniera* sponge.[13,14] Moreover, saframycin was also isolated from a terrestrial *Streptomyces lanvandulae*, which indicates that different habitats are not necessarily the point in this contemplation but that several biosynthetic gene clusters

spread beyond the species barrier.[15] In the present case, chlorotonil is uniquely produced by terrestrial myxobacteria whilst anthracimycin has been isolated from marine and terrestrial streptomycete species. The two gene clusters encoding the two PKS systems producing the compounds, overall show several commonalities. Both biosynthetic machineries are polyketide synthases of *trans*-AT type consisting of 10 modules, which are distributed on three large genes respectively.[16] A tandem AT domain composed of an acyl hydrolase domain, an acyl transferase domain and an enoyl reductase domain is encoded upstream to the core PKS genes in both cases.[17] The predicted ketsoynthase substrate specificity is also consistent amongst the set of 10 KS domains present in the respective biosynthetic pathway.[18,19] Moreover, the "odd" placement of domains in single modules is similarly seen in both PKSs. The structure of both compounds suggests that an intramolecular [4+2] cycloadditon takes place during biosynthesis; the topic of such Diels-Alder like ring formations in a biological context has been debated over and over again, but so far, no Diels-Alderase enzyme has been reported. Only very specific enzymes adjusting the substrate molecules in favorable orientation supporting the cycloadditon were found.[20] In the case of chlorotonil and anthracimycin no such functionality is encoded in the respective gene cluster; importantly, Wilkinson et al. (*ACS Chem. Biol.* DOI: 10.1021/acschembio.5b00525) verified via heterologous production of anthracimycin that the set of genes reported in chapter 2 is sufficient for production of the compound. Considering these data, an enzyme-catalyzed intramolecular [4+2] cycloadditon seems unlikely. we thus propose spontaneous formation of the dekalin moiety in both cases. The same holds true for the double bond shifts in modules two and six: None of the previously reported architectural prerequisites ("shift modules") for catalysis of a double bond shift can be found in either of the pathways, thus a spontaneous isomerization is conceivable.[21] However, to address these aspects of the biosynthesis in a more detailed way, further experiments would be necessary. To gain further insights in the timing and procedure of double bond shifts and cycloaddition several approaches are conceivable. In order to obtain intermediates of the biosynthesis, which could be analyzed by HPLC-MS, the introduction of point mutations into ACP domains would be an option. If the active-site serine of an ACP domain was exchanged for another amino acid, the growing polyketide chain could not be passed on by this domain leading to off-loading of the respective intermediate. Though this approach requires adequate genetic tools and, in the best case, a well-established platform for heterologous expression with an easy-to-manipulate host system. An alternative method, also based on mass spectrometry would

be the off-loading of polyketide intermediates through the incorporation of non-hydrolysable synthetic mimics of extender units. This method takes advantage of the fact that polyketides are built sequentially by repeated decarboxylation and condensation of malonyl-CoA units. The chemical mimics developed by Tosin and coworkers interfere in this process as they compete with respective ACP domain for the growing polyketide chain, leading to off-loading of unfinished intermediates which can be analyzed by HPLC-MS.[22–24] This method is more advantageous as it can be carried out *in vivo* by feeding the carba(dethia)cysteamine analogues to the respective producer strain. Furthermore the only requirements are the culturability of the producer strain and the synthesis of the synthetic mimics. Consequently this straightforward approach can provide valuable insights into key –steps during biosynthesis.

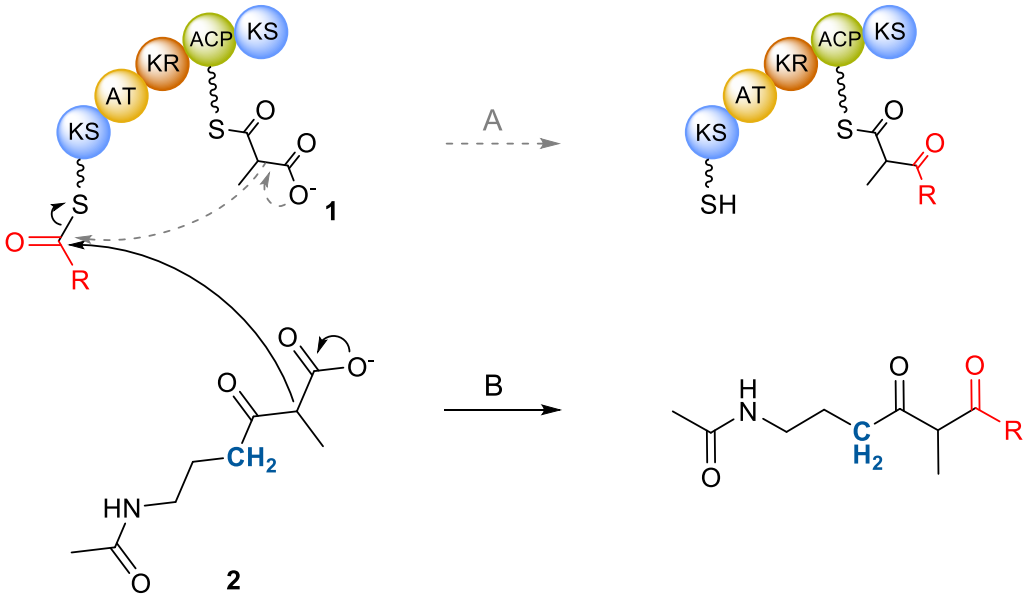


Figure 23: Decarboxylative condensation of methylmalonyl-ACP (**1**) leading to polyketide formation in type I modular PKS via route A. Route B shows how the feeding of the respective nonhydrolysable mimic of methylmalonyl-ACP, methyl-malonyl-carba(dethia)-*N*-acetyl cysteamine which intervenes in polyketide formation and leads to off-loading of intermediates. Adapted from Tosin et al. *Chem.Commun.*, 2011, 47, 3460-3462

The in-depth *in silico* analysis of both pathways led to the assumption that regarding the stereochemistry of the two compounds, the ketoreductase domain in module one is likely the critical determinant. In the chlorotonil PKS this KR domain belongs to the A-type group of KRs, responsible for formation of an S-configured alcohol whilst in the anthracimycin PKS the counterpart KR, belonging to the B-type group, forms an R-

configured alcohol.[25,26] Looking at the differences between the two pathways it is striking that the extent to which the anthracimycin PKS deviates from textbook PKS logic surpasses what is seen in the chlorotonil pathway by far. In the chlorotonil PKS, the lack of cognate AT domains within all ten modules is the most obvious divergence from cis-AT PKS. Apart from that, all functionalities predicted from a retro-biosynthetic analysis can be found. This is not the case for the anthracimycin PKS where a full reductive loop and a methyltransferase domain are missing in module 6 and the ER domain is located far from its anticipated point of action. Since it could be demonstrated by Wilkinson and coworkers that the present gene cluster is indeed responsible for anthracimycin biosynthesis, the enzymatic functionalities missing at a certain stage during biosynthesis must be replaced by domains encoded in trans – such as the ER domain attached to the tandem AT domain – or the growing polyketide chain is iteratively passed through a module exhibiting all domains required. A similar scenario is described during sorangicin biosynthesis, where module two (KS-ACP-ACP) is also missing the reductive loop and a methyltransferase domain. Also in this case the authors suggest that the functionalities are provided in *trans* but no mention of candidate enzymes is made.[27]

Looking at the big picture, the analyzed data suggest that the anthracimycin gene cluster and the chlorotonil gene cluster have a common origin. The alternative explanation, convergent evolution of these two pathways, seems very unlikely. Comparing of the codon usage in both gene clusters reveals identical preferences. Following this hypothesis, there are indications that the anthracimycin gene cluster evolved from the chlorotonil pathway. As for example the superfluous inactive ACP domains observed within the chlorotonil PKS – with one exception - cannot be found in the anthracimycin PKS anymore. The absence of a complete set of domains in module six also corroborates this hypothesis. Consequently it might be, that the anthracimycin gene cluster is a version of the chlorotonil gene cluster in which all redundant functionalities were removed. The simultaneous presence of two identical copies of the anthracimycin PKS in *Streptomyces* CNH365 and their relative position in the linear genome also suggests that a duplication of these genes happened in the more recent past.

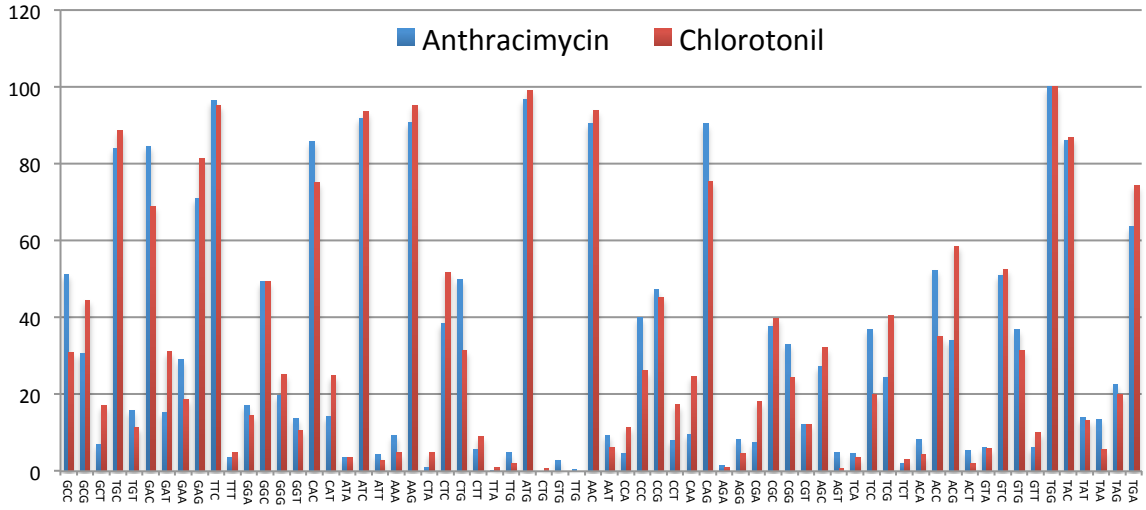


Figure 24: Comparison of codon usage in the anthracimycin and the chlorotonil biosynthetic gene cluster. Similar to identical preferences can be observed. The analysis was carried out across the genomic loci extending the respective gene cluster.

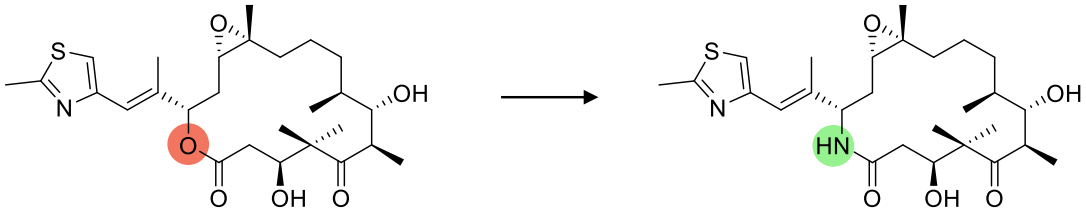
Beyond the question of evolutionary linkage of the two PKSs, the comparative analysis of both pathways revealed, that the relatively new field of *trans*-AT PKS is not as traceable as would be desired for productive utilization by natural product chemists. Many tools such as KS substrate specificity prediction through phylogenetic analysis or the assignment of KR types via fingerprint analysis provide useful support during *in silico* analysis of biosynthetic pathways, but especially in the case of *trans*-AT PKS, many questions still remain open. In particular, despite in-depth analysis we cannot answer the question how these two intriguingly similar pathways, at the same time showing several striking disparities, can produce these two highly similar natural products. The genetic differences observed amongst the gene clusters do not adequately explain all structural peculiarities.

6.1.2 Production of chlorotonil and anthracimycin derivatives

Besides the fundamental research on the biosynthesis of chlorotonil and anthracimycin, the bioactivity observed for both compounds is a major aspect that mainly triggered our interest in these two natural products. Anthracimycin initially showed promising activity against the spore-forming Gram-positive anthrax pathogen *Bacillus anthracis* as well as against methicillin resistant *Staphylococcus aureus*.[2,28] In addition, bioactivity against other Gram-positive indicator strains such as enterococci and streptococci could be detected but no activity towards Gram-negative pathogens was found. Thus, inspired by the unique *gem*-dichloro-1,3-dione structure of chlorotonil, Fenical and coworkers chemically modified anthracimycin in analogy to the last step of the chlorotonil total synthesis to yield the respective *gem*-dichloro-1,3-dione derivative.[29] The obtained derivative lost half of its potency against the previously tested indicator pathogens but the dichlorinated compound additionally exhibited bioactivity against a panel of Gram-negative pathogens. Thus it could be shown that – as often observed for natural products – chlorination plays a major role for bioactivity.[30] The same holds true in the case of chlorotonil, where only the di-chlorinated compound shows outstanding bioactivity whilst the bioactivity of mono- and non-chlorinated congeners is significantly lower or cannot even be observed anymore. We thus conclude that the di-chloro pattern is crucial for the biological activity of chlorotonil A. One major drawback of chlorotonil is its high lipophilicity and thus the low bioavailabilty. Although the in vivo testing of the compound in malaria infected mice revealed excellent IC_{50} values for chlorotonil A, the effective dose of the compound and thus the systemic uptake and the resulting plasma values are not known. Given the hydrophobicity of chlorotonil A one could presume that the effective dose is only a small fraction of the amount that was administered to the mice. Consequently, the drug concentration at the site of action would only be reached via a concentration gradient which means that chlorotonil is excreted un-metabolized to a great extent. Considering the efforts spent towards isolation and purification of natural products, such a tremendous loss of compound is undesirable and requires countermeasures. Two key disciplines of pharmaceutical sciences, pharmaceutical technology and medicinal chemistry, provide tools and approaches to address such issues. Pharmaceutical technology aims to find a suitable vehicle for a targeted transport of the respective drug to its site of action. As the available carrier systems are highly diverse, an iterative process is frequently employed to find individual solutions for the respective

drug. The spectrum ranges from solubility enhancers such as polyethylenglycol to more advanced systems such as liposomes enclosing the compound or nanoparticles coated with the drug. Already for the bioactivity testing of chlorotonil A, a simple formulation of the compound was needed due to its poor solubility in the aqueous growth media of test strains. Therefore, a 1:1 cremophor/THF formulation based on the paclitaxel formulation currently on the market was developed for chlorotonil A.[31] In parallel also endeavors towards establishment of formulation for peroral application is developed in the Drug Delivery department of the Helmholtz Institute for Infection Research Saarland, but the outcome of this study is not the subject of this work.

Medicinal chemistry on the contrary follows a different approach, aimed at optimizing hit structures to lead compounds. In the context of natural products this means that the compound itself is modified towards advanced pharmacokinetic properties. The overall objective of these endeavors is to generate an optimized drug with increased activity which selectively binds to the target and at the same time exhibits favourable pharmokokinetic properties such as plasma stability. Also in the field of natural product research, medicinal chemistry plays an important role as very often, the unmodified natural product shows good bioactivity but its poor physicochemical properties are a hindrance for direct use in the clinics. The anticancer drug epothilone, isolated from the myxobacterium *Sorangium cellulosum* is a good example for this. The epothilones show good cytotoxic activity and it could be shown that they induce tubulin polymerization, cell cycle arrest and apoptosis. The fact that the lactone ring of the epothilones was prone to ester cleavage, however, impeded *in vivo* usage as along with this also the potency of the epothilones was lost due to degradation. By means of semi-synthesis Bristol-Myers Squibb set out to address this issue and in extended structure-activity studies and introduced numerous alterations to the epothilone B backbone. Finally, replacing the lactone for a lactame leading to the epothilone B derivative ixabepilone which was approved by the FDA for cancer treatment, was the breakthrough.[32,33]



Epothilon B Ixabepilone (Ixempra®)

Figure 25: Semi-synthetic alteration of epothilone B leading to ixabepilone with increased plasma stability.

In the case of chlorotonil A, the molecular target of the compound is still unknown but endeavors towards identification thereof are in progress. Nevertheless, the data obtained during activity testing of chlorotonil provided us with first insights into the structural properties responsible for the bioactivity of the macrolide. It was striking that only chlorotonil A but none of its mono-chlorinated congeners showed outstanding bioactivity. This finding was consistent in the anti-bacterial as well as in the anti-malaria testing and thus indicated that the dichloro-pattern uniquely seen in chlorotonil A is crucial for the bioactivity. In a first attempt to increase the solubility of the compound in polar solvents the dihydro-chlorotonil derivative was produced via reduction of the C3-carbonyl group to the respective alcohol. This modification led to a complete loss of bioactivity in all assays suggesting that the *gem*-dichloro-1,3-dione structure in chlorotonil A is required for bioactivity and thus represents an important part of the pharmacophore. Keeping this in mind we elaborated on a strategy for the production of chlorotonil derivatives with improved solubility. For this objective a biotechnological and a semi-synthetic approach were chosen. The biotechnological approach which is described in detail in chapter three relies on the establishment of a platform for heterologous expression of the chlorotonil biosynthetic pathway. Upon successful completion of this task the manipulation of the biosynthetic pathway at two distinct positions is assumed to lead to the production of mono-or di-hydroxylated chlorotonil derivatives. Still, heterologous expression of the biosynthetic pathway is a task to fulfill.

In the semi-synthetic approach, the isolated C9-C10 double bond was targeted for the introduction of modifications, whereas an epoxidation at this position was envisaged. The derivatisation reaction yielded two diastereomeric epoxide derivatives named epochlorotonil A and B named after their occurrence in the chromatogram. Upon scale up

of the reaction, separation and purification of the two derivatives was performed as well as activity testing *in vitro*. Interestingly the two diastereomeric epoxide derivatives differed considerably in their bioactivity as epochlorotonil B, the major product showed half the potency of chlorotonil A, whilst epochlorotonil A was found to be only half as potent as its diastereomeric counterpart. Notwithstanding, both derivatives still showed considerable bioactivity unlike the dihydro-derivative. This finding corroborated our pharmacophore hypothesis and thus set the stage for further modifications of chlorotonil A.

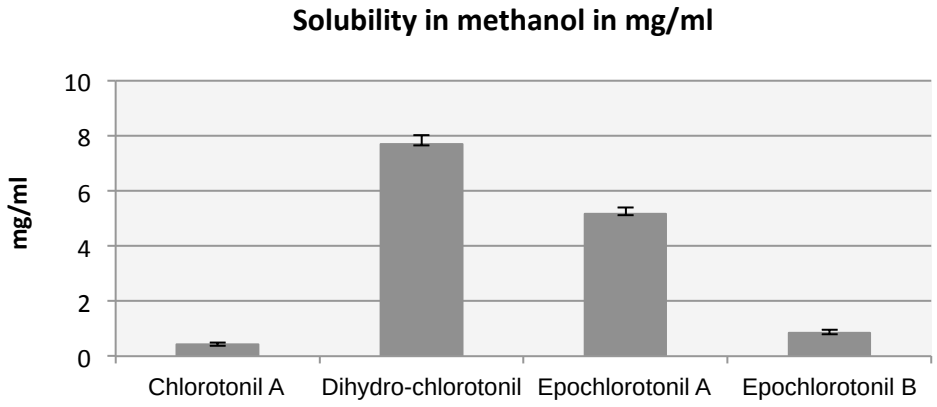


Figure 26: Methanol solubility of semi-synthetic chlorotonil derivatives in comparison to chlorotonil A.

To quantify the influence of the respective modification on the solubility of the compounds the methanol solubility of the three derivatives in comparison to chlorotonil A was assayed. This led to the surprising observation that polarity of the compounds is seemingly related to their biological activity, thereby providing more insights into the physicochemical properties required for bioactivity.

The highest solubility with up to 7.7 mg/ml was observed for the inactive dihydro-chlorotonil; the epochlorotonil derivatives however differ significantly in their methanol solubility with 5 mg/ml for the less active epochlorotonil A and 0.8 mg/ml for epochlorotonil B. For chlorotonil A, a value of 0.4 mg/ml was determined. The fact that the two epoxide derivatives differ tremendously in their solubility implies that the orientation of the epoxide residue highly influences how well the solvent can access the molecule. It is striking that there seems to be a correlation between hydrophilicity and bioactivity as our small test setting indicates that, the poorer the solubility in polar solvents, the more active the compounds. These data suggests that not only the supposed

gem-dichloro-1,3-dione pharmacophore is essential for bioactivity, but also a hydrophobic interaction between the molecule and its target might be of importance. Therefore in further derivatisation attempts a reasonable balance between hydrophilicity and bioactivity has to be taken care of. The two epoxide derivatives in this context represent a step in the right direction as solubility could be improved whilst bioactivity was maintained.

Following this approach we were able to show that chlorotonil A can be chemically modified and moreover several prerequisites for further studies were defined. The derivatives produced so far and their bioactivity provided valuable information contributing to the pharmacophore hypothesis proposed for chlorotonil A. Given its outstanding bioactivity especially against malaria, chlorotonil A indeed represents a new lead structure that might find its way into the clinics once the bioavailability issue can be addressed satisfactorily. Overall the semi-synthetic approach might as well be helpful for identification of the molecular target of chlorotonil. Functionalization of the compound with a biotin linker for instance enables target fishing via affinity chromatography or proteome chip based methods.[34] The preparation of further chlorotonil derivatives is work in progress.

6.2 Maracen and maracin – synthetic-biology-assisted investigation of biosynthesis

The discovery and structure elucidation of the two myxobacterial compounds maracen and maracin from a *Sorangium cellulosum* strain raised the question how these two secondary metabolites are produced in their host strain. With genome data in hands, a candidate biosynthetic gene cluster of PKS type was sought for but no match could be found. However, looking at the structure of the two compounds also a connection to fatty acid biosynthesis could not be excluded. A project especially dedicated to the production of polyunsaturated fatty acids in myxobacteria finally provided crucial indications for this connection as described in chapter 5. With the working hypothesis elaborated for maracen and maracin biosynthesis, two options for verification or falsification of the hypothesis are conceivable. In a rather traditional approach, well-established for investigation of biosynthetic models or pathways, the inactivation of several candidate genes via single crossover experiments accompanied by the analysis of the metabolite profiles of the produced mutants is the first way to go. If the involvement of the targeted gene in biosynthesis of the respective compound could be proven that way, several other approaches for in-depth investigation of the biosynthetic pathway are conceivable. *In vivo* assays aiming to determine the functionality of enzymatic key steps during biosynthesis are one possibility, whereas the enzyme of interest has to be expressed recombinantly as a soluble protein, suitable assay conditions including pH and cofactors have to be established and the substrate to be transformed by the enzyme has to be available either from total synthesis or a biological source. In case all requirements can be met, in most cases the read-out of such assays provides valuable information about the enzyme and its kinetics. Furthermore the deletion or point mutation of single genes or domains within the respective gene cluster is another approach to investigate biosynthetic processes. This approach, however, requires well-established protocols for mutagenesis of the producer strain which is very often not the case especially for myxobacteria from the genus *Sorangium*. However, if the respective gene cluster can be modified directly in the host, the feeding of mutasynthons allows for the production of derivatives and can additionally provide insights into the substrate specificity of specific enzymes.[35] In case the producer strain is not amenable to mutagenesis, at least PKS biosynthetic pathways can be analyzed using the previously mentioned chain-terminating analogous developed by Tosin et al..[22] The heterologous expression of a gene cluster is also a good way to investigate biosynthesis and at the same time establish a heterologous host more amenable to

mutagenesis. This strategy is quite labour-intensive as it requires the mobilization of the genes from the producer strain. In the past, the construction of genomic BAC (bacterial artificial chromosome) or cosmid libraries was the method of choice to reach this goal. However, as soon as the target gene cluster exceeds a certain size, the complete pathway has to be assembled from several cosmids, which depending on the number of cosmids can be very time consuming. Only recently more straightforward methods such as direct cloning of complete gene clusters via TAR cloning (transformation assisted recombination) were established for bigger gene clusters as well and provide an alternative for mobilization of large pathways in one step.[36,37] If heterologous expression of the desired pathway is met with success, the door is open for further investigations of biosynthetic processes via modifications in the cluster sequence for instance. An established platform for heterologous expression moreover allows for production of derivatives via targeted mutations. In summary, it can be said that the classical approach to investigate biosynthetic pathways comprises many tools but all of them require that the respective strain can be cultivated under laboratory conditions and that it can be mutagenized to a greater or lesser extent. It would thus be favourable to have an alternative approach for in-depth investigation of biosynthetic pathways and heterologous expression, especially in the case of difficult to access bacterial strains producing compounds of major interest.

This alternative approach might be found in the emerging field of synthetic biology. Synthetic biology, understood as an interdisciplinary field involving many research areas such as molecular biology, bioinformatics, physics and engineering for instance, is based on the concept that biologic processes and regulatory networks are predictable, calculable and characterizable using mathematical models.[38] As a result synthetic biology allows for modification and reprogramming of biological processes to an extent beyond conventional biotechnology. The great promise of synthetic biology is that tuning of existing metabolic pathways by genetic methods is no longer the only option, but the redesign or even *de novo* construction of biosynthetic pathways featuring non-natural functionalities can be realized.[39] The fact that biosynthesis genes are often clustered, as observed for PKS and NRPS pathways is a factor that is highly beneficial for this approach. The steadily decreasing costs for genome sequencing enabled genome mining in rich bacterial sources such as streptomyces and myxobacteria, leading to the discovery of a plethora of metabolic pathways *in silico*. However, in many cases no candidate

metabolite can be detected in the respective metabolome, so the respective gene cluster is considered a silent or cryptic pathway. The awakening of such silent pathways is in the focus of natural product research for quite a while, but synthetic biology now offers the unique opportunity to exchange the native regulation system of such pathways for controllable synthetic regulatory network.[40,41] Looking at recent developments in the field, synthetic biology plays an increasing role in natural product research and provides valuable tools for investigation of biosynthetic processes as well as for modulation and manipulation of biosynthetic pathways *ex vivo*. Several application areas are conceivable and first examples prove the feasibility of this idea.[42–45]

Consequently, we decided to rely on synthetic biology methods to investigate the biosynthesis of maracen and maracin. The natural producer strains of the two compounds were notoriously hard to mutagenize, thus we were not able to prove involvement of the candidate biosynthetic pathway in metabolite production via conventional methods. Instead, an artificial version of the gene cluster based on the native sequence was designed with initially no profound changes concerning regulation of the gene cluster. In order to probe the functionality of specific enzymes encoded in the alleged biosynthetic pathway, a sophisticated system for selective removal of single genes from the gene cluster was employed. With this tool we aimed to remove selected functionalities from the gene cluster and monitor the effects on metabolite production. Especially the duplicated isomerase encoding genes (*mrc17, mrc18*) and the multiple copies of the phospholipase encoding genes (*mrc7,9,12,16*) were in the focus of these considerations. Nevertheless the applied system also facilitates selective removal of all other genes. Upon completed design and successful synthesis of all required DNA fragments the first version of the gene cluster was assembled in a two-step process. For production of a native gene cluster version, first all splitter elements were removed from the DNA fragments which were subsequently combined stepwise. For overall improvement of the assembly strategy this step requiring several rounds of DNA restriction, ligation and transformation could also be replaced by a more advanced technique. In order to build a complete pathway from DNA fragments, also one pot assembly strategies like Gibson assembly would be an option. Alternatively also in this case TAR cloning is applicable, whereby Gibson assembly is preferable as it does not depend on host systems such as yeast or *E. coli* .[46–48] Nevertheless stepwise assembly of the gene cluster by ligation was carried out with success to give an expression construct of 32.5 kbp. Having our biosynthesis model in

mind, maracen and maracin production relies on the fatty acid precursor eicosapentaenoic acid, thus the PUFA gene cluster described previously was also part of the biosynthetic pathway. To be more flexible the PUFA pathway could either be cloned on the expression construct harbouring the maracen genes, or it was transferred to the heterologous host on a separate construct.[44]

During the planning of the maracen/maracin project, the selection of a suitable host system for heterologous expression was one of the key issues to address. In the last years, the myxobacterial strain *Myxococcus xanthus* DK1622 has proven worthwile as host strain for heterologous expression of different metabolic pathways.[44,49,50] Other than myxobacteria from the genus *Sorangium, M. xanthus* DK1622 exhibits higher growth rates leading to dense liquid cultures growing in suspension. Furthermore, the strain is easier to manipulate and several well-established robust protocols for introduction of foreign DNA into the genome of DK1622 are available.[51,52] Since also the PUFA gene cluster, which is considered an integral component of the maracen/maracin pathway has already been successfully expressed in DK1622, this host became the first choice for this project as well. However, considering the size of the construct to be introduced (~ 46 kbp) and the differences between the *sorangium* producer and the *myxococcus* host strain concerning codon preferences for instance, we also had concerns about this choice. Therefore we planned for an alternative host system as well. We were consequently looking for a suitable strain of the genus *Sorangium* for which established protocols for mutagenesis are available and which does not produce maracen and maracin. *Sorangium cellulosum* So ce 1525, the chlorotonil producer strain fulfills all these conditions and was hence chosen as an alternative host. For integration of the gene cluster in So ce1525 a synthetic transfer cassette was designed and synthesized. The genetic elements of this cassette were derived from the pSUP_Hyg vector and thus allowed for integration and selection in *Sorangium cellulosum*.[53]

In our first series of experiments aiming to establish heterologous expression of the maracen and maracin biosynthetic gene cluster in *Myxococcus xanthus* DK1622, we followed the multi-plasmid approach. First we generated a DK1622 mutant harbouring the PUFA gene cluster which was shown to be expressed successfully as production of eicosapentaenoic acid could be observed. Subsequently, we transformed the expression construct harboring the remaining maracen/maracin biosynthetic gene cluster to this

mutant but despite several series of transformation our endeavors were not met with success. The finding that the maracen/maracin gene cluster cannot readily be integrated in the myxococcus host genome is open to different possible interpretations. On the one hand it might be that (as anticipated earlier), a gene cluster of such size stemming from a related but still rather distant myxobacterium is not integrated in the host genome but just recognized as foreign DNA and thus degraded. If that was the case the alternative host system could facilitate heterologous expression, since the alternative producer is a *Sorangium cellulosum* species. As the required tools for this approach are already available, trying to achieve heterologous expression in the alternative host strain would be the next step in this project. However, it is also conceivable that the maracen/maracin gene cluster is possibly not integrated in the host genome since the two metabolites are toxic to the heterologous host. In this case also the alternative host system might be an option but the genetic locus of the biosynthetic pathway in the producer strain has to be carefully re-checked for functionalities conveying self-resistance to the producer strain as well.

Once the current difficulties are overcome and a system for heterologous expression of the maracen/maracin pathway can be established, this prepares the way for in-depth analysis of biosynthetic processes using the splitter system. As mentioned previously the intriguing biosynthetic route proposed for maracen/maracin with regard in particular to the involvement of a bacterial lipoxygenase and a delta-(12)-fatty acid dehydrogenase is of high interest. A platform for heterologous expression of the biosynthetic pathway would thus allow for straightforward, detailed investigations by selective removal of enzyme functionalities from the biosynthetic route using the splitter system. Upon verification of our biosynthetic hypothesis, the maracen pathway would clearly represent a novel type of secondary metabolite production route depending on the just recently discovered new type of polyunsaturated fatty acid synthase *pfa*. Moreover, the platform can be used to achieve enhanced production of the metabolites in the heterologous host and also allows for the generation of analogues.

6.3 Final thoughts

Natural products continue to play a major role especially in the field of anti-infective research. Thanks to steadily improved techniques and advances particularly in the field of genome sequencing, the access to natural products and the respective biosynthetic gene clusters becomes more and more straightforward. However, modifications of the molecules are often required to adapt their pharmacokinetic properties. To accelerate this process, heterologous expression of biosynthetic pathways is a valuable tool as once a production platform is established, the generation of analogues via modification of the gene cluster can be envisaged. The idea of synthetic biology takes this concept one step further, as this discipline aims to create tailored natural product factories using a comprehensive approach which is based on the gathered knowledge from various biosynthetic pathways. The synthetic biology toolbox is thus likely to contribute to the efficiency of natural product research in the future, as it potentially uncouples the research process from the obstacles and difficulties encountered when working with scarcely manageable microorganisms.

Summing up, natural products continue to be a rich source for new lead structures, reflected by the number of approved drugs with natural product background. Recent advances in the filed especially in the technologies applied have had a big impact on natural product research. However, despite the advantages compared to synthetic drugs, the discovery of natural products and their establishment as pharmaceuticals still is a laborious process with many incalculable parameters. Elucidating biosynthetic pathways supports these endeavors and has the potential to facilitate this process. Once a wider background knowledge of the involved biosynthetic processes is obtained, the door is open for a more streamlined exploitation of microbial natural products, including the implementation of advanced synthetic biology tools.

6.4 References

[1] K. Gerth, H. Steinmetz, G. Höfle, R. Jansen, *Angew. Chemie Int. Ed.* **2008**, *47*, 600–602.

[2] K. H. Jang, S.-J. Nam, J. B. Locke, C. a Kauffman, D. S. Beatty, L. a Paul, W. Fenical, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7822–4.

[3] T. Hoffmann, S. Müller, S. Nadmid, R. Garcia, R. Müller, *J. Am. Chem. Soc.* **2013**, *135*, 16904–16911.

[4] E. a. Guzmán, K. Maers, J. Roberts, H. V. Kemami-Wangun, D. Harmody, A. E. Wright, *Invest. New Drugs* **2015**, *33*, 86–94.

[5] B. Kunze, B. Böhlendorf, H. Reichenbach, G. Höfle, *J. Antibiot. (Tokyo).* **2008**, *61*, 18–26.

[6] P. Crews, L. V Manes, M. Boehler, *Tetrahedron Lett.* **1986**, *27*, 2797–2800.

[7] T. M. Zabriskie, J. A. Klocke, C. M. Ireland, A. H. Marcus, T. F. Molinski, D. J. Faulkner, C. Xu, J. Clardy, *J. Am. Chem. Soc.* **1986**, *108*, 3123–3124.

[8] F. Sasse, B. Kunze, T. M. Gronewold, H. Reichenbach, *J. Natl. Cancer Inst.* **1998**, *90*, 1559–1563.

[9] S. Haag-Richter, H. Hoffmann, M. Kurz, H. Tietgen, **2005**.

[10] T. A. Johnson, J. Sohn, Y. M. Vaske, K. N. White, T. L. Cohen, H. C. Vervoort, K. Tenney, F. A. Valeriote, L. F. Bjeldanes, P. Crews, *Bioorg. Med. Chem.* **2012**, *20*, 4348–55.

[11] K. L. Erickson, J. A. Beutler, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8188–8192.

[12] B. Kunze, R. Jansen, F. Sasse, G. Höfle, H. Reichenbach, *J. Antibiot. (Tokyo).* **1998**, *51*, 1075–1080.

[13] H. Irschik, W. Trotzwitzsch-Kienast, K. Gerth, G. Höfle, H. Reichenbach, *J. Antibiot. (Tokyo).* **1988**, *XLI*, 993–998.

[14] Y. Nakao, T. Shiroiwa, S. Murayama, S. Matsunaga, Y. Goto, Y. Matsumoto, N. Fusetani, *Mar. Drugs* **2004**, *2*, 55–62.

[15] T. Arai, K. Takahashi, A. Kubo, *J. Antibiot. (Tokyo).* **1988**, *XLI*, 993–998.

[16] J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996–1047.

[17] E. M. Musiol, T. Weber, *Medchemcomm* **2012**, *3*, 871.

[18] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–33.

[19] M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel, N. J. Oldham, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 1143–7.

[20] T. Hashimoto, J. Hashimoto, K. Teruya, T. Hirano, K. Shin-Ya, H. Ikeda, H.-W. Liu, M. Nishiyama, T. Kuzuyama, *J. Am. Chem. Soc.* **2015**, 2–5.

[21] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chemie* **2010**, *49*, 1502–1506.

[22] M. Tosin, D. Spiteller, J. B. Spencer, *Chembiochem* **2009**, *10*, 1714–23.

[23] M. Tosin, L. Betancor, E. Stephens, W. M. A. Li, J. B. Spencer, P. F. Leadlay, *Chembiochem* **2010**, *11*, 539–46.

[24] M. Tosin, Y. Demydchuk, J. S. Parascandolo, C. B. Per, F. J. Leeper, P. F. Leadlay, *Chem. Commun. (Camb).* **2011**, *47*, 3460–2.

[25] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[26] A. Kitsche, M. Kalesse, *Chembiochem* **2013**, *14*, 851–61.

[27] H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel, R. Müller, *Chembiochem* **2010**, *11*, 1840–9.

[28] M. E. Hensler, K. H. Jang, W. Thienphrapa, L. Vuong, D. N. Tran, E. Soubih, L. Lin, N. M. Haste, M. L. Cunningham, B. P. Kwan, et al., *J. Antibiot. (Tokyo).* **2014**, *67*, 549–53.

[29] N. Rahn, M. Kalesse, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 597–9.

[30] K. Naumann, *Euro-Chlor Sci. Doss.* **2003**.

[31] A. K. Singla, A. Garg, D. Aggarwal, *Int. J. Pharm.* **2002**, *235*, 179–192.

[32] K.-H. Altmann, F. Z. Gaugaz, R. Schiess, *Mol. Divers.* **2011**, *15*, 383–99.

[33] J. T. Hunt, *Mol. Cancer Ther.* **2009**, *8*, 275–281.

[34] B. Lomenick, R. W. Olsen, J. Huang, *ACS Chem. Biol.* **2011**, *6*, 34–46.

[35] J. H. Sahner, H. Sucipto, S. C. Wenzel, M. Groh, R. W. Hartmann, R. Müller, *ChemBioChem* **2015**, n/a–n/a.

[36] X. Bian, F. Huang, F. a. Stewart, L. Xia, Y. Zhang, R. Müller, *ChemBioChem* **2012**, *13*, 1946–1952.

[37] K. Yamanaka, K. a Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein, B. S. Moore, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1957–62.

[38] A. S. Khalil, J. J. Collins, *Nat. Rev. Genet.* **2010**, *11*, 367–379.

[39] S. Rollié, M. Mangold, K. Sundmacher, *Chem. Eng. Sci.* **2012**, *69*, 1–29.

[40] M. H. Medema, R. Breitling, E. Takano, *Methods Enzymol.* **2011**, *497*, 485–502.

[41] A. A. Brakhage, V. Schroeckh, *Fungal Genet. Biol.* **2011**, *48*, 15–22.

[42] W. Wohlleben, Y. Mast, G. Muth, M. Röttgen, E. Stegmann, T. Weber, *FEBS Lett.* **2012**, *586*, 2171–2176.

[43] M. H. Medema, M. T. Alam, R. Breitling, E. Takano, *Bioeng. Bugs* **2011**, *2*, 230–233.

[44] K. Gemperlein, S. Rachid, R. O. Garcia, S. C. Wenzel, R. Müller, *Chem. Sci.* **2014**, *5*, 1733.

[45] R. P. D. Müller, S. D. Wenzel, R. Garcia, *Synthetic Pathway Enzymes for the Production of Argyrins*, **2010**.

[46] D. G. Gibson, G. a Benders, K. C. Axelrod, J. Zaveri, M. a Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith, C. a Hutchison, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20404–9.

[47] D. G. Gibson, *Nucleic Acids Res.* **2009**, *37*, 6984–90.

[48] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. a Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–5.

[49] C. Oßwald, G. Zipf, G. Schmidt, J. Maier, H. S. Bernauer, R. Müller, S. C. Wenzel, *ACS Synth. Biol.* **2014**, *3*, 759–772.

[50] Y. Chai, S. Shan, K. J. Weissman, S. Hu, Y. Zhang, R. Müller, *Chem. Biol.* **2012**, *19*, 361–371.

[51] V. Magrini, M. L. Storms, P. Youderian, **1999**, *181*, 4062–4070.

[52] B. Julien, *J. Bacteriol.* **2003**, *185*, 6325–6330.

[53] S. Jaoua, S. Neff, T. Schupp, *Plasmid* **1992**, *28*, 157–165.

7. Appendix

7.1 Curriculum Vitae

Persönliche Daten

Name	Katrin Jungmann
Geburtsdaten	19.06.1986 in Völklingen
Familienstand	ledig

Wissenschaftlicher Werdegang

09/2015 - 10/2015	Forschungsaufenthalt bei Prof. Valerie Mizrahi, Institute of Infectious Diseases and Molecular Medicine (IDM), Molecular Mycobacteriology Research Unit, University of Cape Town, South Africa
ab 01/2012	Promotion unter Anleitung von Prof. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS),Saarbrücken, Abteilung Mikrobielle Naturstoffe Titel der Dissertation: “Investigation of bacterial secondary metabolite pathways from *Sorangium cellulosum*.”
10/2006 – 09/2010	Studium der der Pharmazie, Universität des Saarlandes Abschluss: 2. Staatsexamen
03/2009 - 06/2009	Forschungsaufenthalt bei Prof. Jennifer Littlechild, Exeter Biocatalyis Center, Department of Cellular and Chemical Biology, University of Exeter, Exeter United Kingdom
10/2005 - 10/2006	Studium der Chemie an der Universität des Saarlandes

Schuldbildung

08/1996 - 06/2005	Allgemeine Hochschulreife, Marie Luise Kaschnitz Gymnasium Völklingen, Abschluss: Abitur

Mitgliedschaften

Alumni und Freunde der Fachrichtung Pharmazie an der Universität des Saarlandes e.V.

Deutsche Pharmazeutische Gesellschaft (DPhG)