

Exploitation and Engineering of Lipopeptide Biosynthesis in Myxobacteria

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Christian Burgard

Saarbrücken

2017

Tag des Kolloquiums: 14.12.2017

Dekan: Prof. Dr. Guido Kickelbick

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Elmar Heinzle

Vorsitz: Prof. Dr. Uli Kazmaier

Akad. Mitarbeiter: Dr. Christian Völzing

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller in der Fachrichtung Pharmazie, am Lehrstuhl für Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes und am Helmholtz-Institut für Pharmazeutische Forschung Saarland im Zeitraum von April 2013 bis Februar 2017.

Danksagung

Der größte Dank soll an dieser Stelle meiner Familie zuteilwerden, insbesondere meiner Mutter Hanne, meiner Schwester Janine, meiner Lebensgefährtin Annika sowie ihren Eltern Uwe und Ulrike. Ohne euch würde diese Arbeit nicht existieren!

Prof. Dr. Rolf Müller danke ich herzlich für die Aufnahme in seine Arbeitsgruppe, für die fortwährende Unterstützung, für seine Ehrlichkeit sowie das mir entgegengebrachte Vertrauen während der kompletten Promotionszeit. Durch dich bot sich mir eine einmalige Chance!

Bei Prof. Dr. Elmar Heinzle möchte ich mich dafür bedanken, dass er trotz Ruhestand diese Arbeit wissenschaftlich begleitet und als Berichterstatter ebendieser fungiert hat. Respekt!

Ein ganz besonderer Dank gilt Dr. Silke C. Wenzel für die grandiose wissenschaftliche Betreuung, ihre nicht enden wollende Unterstützung und für die „Überlassung“ ihres enormen Fundus an wissenschaftlicher Expertise. Durch dich habe ich gelernt „anders“ zu denken!

Mein Dank gilt auch der gesamten Arbeitsgruppe sowie besonders den vielen internen und externen Kollegen, die direkt oder indirekt zu dieser Arbeit beigetragen haben und mir stets mit Rat und Tat zur Seite standen. Ihr habt mich nicht nur fachlich, sondern auch persönlich weitergebracht!

Großen Anteil daran, dass der Laden überhaupt läuft und damit ein riesiges Dankeschön haben insbesondere unsere technischen Assistenten und Laborhelfer verdient: Karsten Mayr, Claudia Helbig, Viktoria Schmitt, Daniel Sauer, Verena Qallaku, Eva Luxenburger und Irene Kochems. Ihr habt mich nicht nur stets mit allen Materialien versorgt, die es für erfolgreiche Forschung braucht, sondern ihr ward auch immer beste Stimmungsaufheller, falls die Forschung mal doch nicht so erfolgreich lief!

Ich danke allen ehemaligen Mitarbeitern, die mich auf meinem Weg begleitet haben, insbesondere Dr. Khai Bui, Dr. Ram Prasad Awal, Dr. Angela Kling, Anja Schwarz und Simon Frewert. Euer Humor war ebenso essentiell wie eure Unterstützung jedweder Art!

Das Schönste ist zu wissen, dass einige über die Laborbank hinaus zu Freunden wurden!

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:

Publikationen

Burgard, C., Zaburannyi, N., Nadmid, S., Maier, J., Jenke-Kodama, H., Luxenburger, E., Bernauer, H. S., Wenzel, S. C. (2017), Genomics-guided exploitation of lipopeptide diversity in myxobacteria, *ACS Chem. Biol.*, 12, 779-786.

Weitere Publikationen

Folgende Publikationen entstanden während der Promotionszeit, sind jedoch nicht Bestandteil dieser Dissertation:

Oswald, C., Zaburannyi, N., **Burgard, C.**, Hoffmann, T., Wenzel, S. C., Müller, R. (2014), A highly unusual polyketide synthase directs dawenol polyene biosynthesis in *Stigmatella aurantiaca*, *J. Biotechnol.*, 191, 54-63.

Tagungsbeiträge

Burgard, C., Wenzel, S. C., Müller, R. (2014), Synthetic biotechnology to produce myxobacterial secondary metabolites (Vortrag), 3. Sommersymposium der Graduiertenschule für Naturstoffforschung der Universität des Saarlandes, Saarbrücken

Zusammenfassung

Um ein breites Verständnis der Lipopeptid-Biosynthese in Myxobakterien zu erhalten, wurden im Rahmen dieser Arbeit neue Lipopeptid-Biosynthesewege unter Einsatz eines umfangreichen Screenings myxobakterieller Genome identifiziert und charakterisiert. Auf diesem Weg konnten vier bisher unbekannte Lipopeptid-Gerüste vorhergesagt und im weiteren Verlauf durch Strukturaufklärung bestätigt werden. Daneben konnten anhand detaillierter Sequenzanalysen der beteiligten Biosynthesewege die strukturellen Unterschiede der Lipopeptid-Gerüste auf genetischer Ebene erklärt werden. Diese Untersuchungen haben ebenfalls zur Aufklärung der genetischen Mechanismen beigetragen, welche zur Evolution dieser Biosynthesewege geführt haben.

Darüber hinaus wurden die identifizierten Lipopeptid-Biosynthesewege als Modellsysteme zur Etablierung synthetischer Expressionsplattformen herangezogen. Im Rahmen dieser Arbeit konnte eine flexible Assemblierungsstrategie zur Konstruktion artifizieller Lipopeptid-Gencluster entwickelt und eine Genbibliothek generiert werden, auf deren Basis nicht natürliche Lipopeptid-Biosynthesewege mittels kombinatorischer Biosynthese erzeugt und heterolog exprimiert werden konnten. Diese Studien führten zur Produktion von fünf neuartigen Lipopeptid-Gerüsten und demonstrieren eindrucksvoll die Vorteile synthetisch-biologischer Methoden gegenüber klassischen Ansätzen. Die beschriebene Strategie erlaubt darüber hinaus die schnelle Modifikation der artifiziellen Biosynthesewege.

Abstract

To gain a deep understanding of the lipopeptide biosynthesis in myxobacteria, a comprehensive screening of myxobacterial genomes was initially carried out in the course of this thesis leading to the identification and characterization of novel lipopeptide biosynthetic pathways. By following this strategy, four yet unknown lipopeptide cores were predicted and further structurally characterized to ultimately prove the predicted structures. On the basis of detailed sequence analyses of the underlying biosynthetic pathways, the structural differences of the lipopeptide cores could be rationalized on a genetic basis. These studies also contributed to the elucidation of the genetic mechanisms, by which the different biosynthetic pathways have evolved.

Furthermore, the identified lipopeptide biosynthetic pathways were used as model systems to establish synthetic expression platforms. In the course of this thesis, a versatile assembly strategy for the construction of artificial lipopeptide gene clusters was developed, which allowed the generation and heterologous expression of unnatural lipopeptide biosynthetic pathways based on an established gene library via combinatorial biosynthesis. These studies led to the production of five novel lipopeptide scaffolds and impressively demonstrate the huge potential of synthetic biology techniques compared to classical approaches. Moreover, the described strategy allows the rapid modification of the artificial biosynthetic pathways.

Table of Contents

Danksagung.....	III
Vorveröffentlichungen der Dissertation.....	IV
Zusammenfassung.....	V
Abstract.....	VII
1 Introduction.....	1
1.1 Lipopeptides – A Distinguished Class of Natural Products	1
1.2 Myxobacteria – An Underexplored Source of Microbial Natural Products.....	3
1.3 Natural Product Biosynthesis by Multimodular Enzyme Complexes.....	8
1.3.1 Biochemistry of PKS Machineries	9
1.3.2 Biochemistry of NRPS Machineries	12
1.3.3 PKS/NRPS Hybrid Megasynthetases	16
1.3.4 Bioinformatics Tools to Identify and Characterize Biosynthetic Gene Clusters	18
1.4 Synthetic Biotechnology Approaches in Natural Products Research	21
1.5 Myxochromide Pathways as Model Systems – Current State of Research and Outline of the Presented Work	24
1.5.1 Outline I: Genome-Mining to Find New Lipopeptide Scaffolds in Myxobacteria	29
1.5.2 Outline II: Synthetic Biotechnology to Produce Novel Hybrid Myxochromides.....	30
1.6 References	32
2 Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria	39
2.1 Abstract.....	39
2.2 Introduction	39
2.3 Results and Discussion	41
2.3.1 Discovery of Novel Types of Myxochromide Megasynthetases	41
2.3.2 Genome-Mining for Novel Myxochromide Lipopeptide Cores	42
2.3.3 Comparative Analysis of Myxochromide Megasynthetases	43
2.3.4 Evolutionary Relationship and Distribution of <i>mch</i> Clusters in Myxobacteria	45
2.3.5 Potential Biological Function of Myxochromides	46
2.4 Significance.....	47
2.5 Experimental Procedures.....	47
2.5.1 Identification and Decipherment of the Myxochromide B Biosynthetic Gene Cluster from <i>Myxococcus</i> sp. 171.	47

2.5.2	Screening of Myxobacterial Genome Data for Additional Myxochromide Biosynthetic Gene Clusters and Verification via Production Analysis.....	48
2.5.3	Isolation and Structure Elucidation of Novel Myxochromides.....	48
2.5.4	Detailed <i>in silico</i> Analysis of the 16 <i>mch</i> Clusters.....	49
2.5.5	Fruiting Body Formation and Swarm Expansion Assays	51
2.6	Acknowledgements	51
2.7	Supporting Information	52
2.7.1	Identification and Decipherment of the Myxochromide B Biosynthetic Gene Cluster from <i>Myxococcus</i> sp. 171	52
2.7.1.1	Shotgun Genome Sequencing of <i>Myxococcus</i> sp. 171 and Identification of the Myxochromide B Biosynthetic Gene Cluster	52
2.7.1.2	Construction and Screening of a Cosmid Library from <i>Myxococcus</i> sp. 171.....	52
2.7.1.3	Subcloning and Sequencing of Myxochromide B Biosynthetic Gene Cluster Fragments.....	53
2.7.2	Screening of Myxobacterial Genome Data for Additional Myxochromide Biosynthetic Gene Clusters and Verification via Production Analysis.....	56
2.7.2.1	Identification and Annotation of Additional Myxochromide Biosynthetic Gene Clusters.....	56
2.7.2.2	<i>In silico</i> Prediction of Products of the C-type and D-type Megasynthetases.....	61
2.7.2.3	Analysis of Myxochromide Production Spectra	62
2.7.3	Isolation and Structure Elucidation of Myxochromide C ₃ from <i>M. virescens</i> ST200611.....	75
2.7.3.1	Cultivation of <i>M. virescens</i> ST200611 and Isolation of Myxochromide C ₃	75
2.7.3.2	Structure Elucidation of Myxochromide C ₃	75
2.7.4	Isolation and Structure Elucidation of Myxochromide D ₁ from <i>S. erecta</i> Pde77	85
2.7.4.1	Cultivation of <i>S. erecta</i> Pde77 and Isolation of Myxochromide D ₁	85
2.7.4.2	Structure Elucidation of Myxochromide D ₁	85
2.7.5	Isolation and Structure Elucidation of Myxochromide S ₂ -Abu from the Heterologous Expression Strain <i>M. xanthus</i> DK1622::pTpS-mchS	94
2.7.5.1	Cultivation of <i>M. xanthus</i> DK1622::pTpS-mchS and Isolation of Myxochromide S ₂ -Abu.....	94
2.7.5.2	Structure Elucidation of Myxochromide S ₂ -Abu	94
2.7.6	Isolation and Structure Elucidation of Myxochromide S ₂ -diAbu from the Heterologous Expression Strain <i>M. xanthus</i> DK1622::pTpS-mchS	103

2.7.6.1	Cultivation of <i>M. xanthus</i> DK1622::pTpS-mchS and Isolation of Myxochromide S ₂ -diAbu.....	103
2.7.6.2	Structure Elucidation of Myxochromide S ₂ -diAbu	103
2.7.7	Detailed <i>in silico</i> Analysis of the 16 Analyzed <i>mch</i> Clusters	112
2.7.7.1	Sequence Analysis of Catalytic Domains of the Encoded Megasynthetases.....	112
2.7.7.2	Analysis of Recombination Sites	121
2.7.7.3	Phylogenetic Analysis	126
2.7.7.4	CAI Analysis	134
2.7.8	Fruiting Body Formation and Swarm Expansion Assays with <i>M. xanthus</i> DK1622 and Myxochromide-Deficient as well as Overproducing Mutants	142
2.8	References	145
3	Synthetic Biotechnology to Engineer Myxobacterial Lipopeptide Biosynthesis.....	149
3.1	Abstract	149
3.2	Introduction	149
3.3	Results and Discussion	151
3.3.1	Sequence Requirements for the Design of Artificial Myxochromide Pathways....	151
3.3.2	Functional Sequence Design for Heterologous Expression in <i>M. xanthus</i>	152
3.3.3	Constructional Sequence Design of Artificial Myxochromide Pathways.....	153
3.3.4	Construction and Heterologous Expression of a Synthetic A-type <i>mch</i> Cluster....	156
3.3.5	Adaption of the Constructional Sequence Design to an Alternative Type IIS Restriction Enzyme	158
3.3.6	<i>In vitro</i> Reconstitution of Artificial Hybrid <i>mch</i> Clusters.....	161
3.3.7	Heterologous Expression of Artificial Hybrid <i>mch</i> Clusters in <i>M. xanthus</i> and Production Analysis of Mutant Strains	165
3.3.8	Isolation and Structure Elucidation of Hybrid Myxochromides	169
3.3.9	PCP Inactivation Experiments to Induce Module-Skipping Events	173
3.3.10	Significance	180
3.3.11	Experimental Procedures.....	181
3.3.11.1	Sequence Analysis and Design	181
3.3.11.2	De Novo Gene Synthesis of Artificial Gene Cluster Fragments and Synthetic Vectors.....	181
3.3.11.3	Microbial Strains and Culture Conditions.....	182
3.3.11.4	DNA Isolation, Processing and Analysis	182

3.3.11.5	Heterologous Expression of Artificial <i>mch</i> Pathways in <i>M. xanthus</i>	183
3.3.11.6	Myxochromide Production Analysis.....	183
3.3.11.7	Isolation and Structure Elucidation of Engineered Hybrid Myxochromides.....	184
3.4	Supporting Information	186
3.4.1	Constructional Sequence Design of <i>mch</i> Clusters.....	186
3.4.2	Design of the Cloning Vector pSynbio1 and the Expression Vector pSynbio2.....	194
3.4.3	Generation of <i>mch</i> Cluster Fragments via DNA Synthesis.....	196
3.4.4	Construction of Modified Cloning and Expression Vectors	199
3.4.5	Assembly of Artificial <i>mch</i> Biosynthesis Genes and Generation of a Gene Library.....	201
3.4.6	<i>In vitro</i> Reconstitution of Artificial <i>mch</i> Clusters.....	207
3.4.7	Transfer and Heterologous Expression of Artificial <i>mch</i> Clusters in <i>Myxococcus xanthus</i>	212
3.4.8	Structure Elucidation of Novel Hybrid Myxochromides	214
3.4.8.1	Cultivation of Heterologous Production Strains and Isolation of Myxochromides	214
3.4.8.2	Structure Elucidation of Hybrid Myxochromides	216
3.5	References	257
4	Discussion & Outlook	259
4.1	General Scope of the Present Work.....	259
4.2	Bacterial Secondary Metabolite Pathways – Evolution and Diversification.....	259
4.2.1	Recombination Events Lead to Myxochromide Pathway Diversification	261
4.2.2	‘Module-Skipping’ Lead to Myxochromide Pathway Diversification	267
4.2.3	Concluding Remarks	268
4.3	Synthetic Expression Platforms to Produce Myxobacterial Natural Products	269
4.3.1	Heterologous Expression of Myxobacterial Biosynthetic Gene Clusters	270
4.3.2	Establishment of an Innovative Assembly Strategy for Synthetic Gene Clusters .	276
4.3.3	Synthetic Biotechnology to Engineer Novel Nonribosomal Peptides	281
4.3.4	Concluding Remarks	285
4.4	References	287

1 Introduction

1.1 Lipopeptides – A Distinguished Class of Natural Products

Natural products are defined as chemical compounds that are produced by a biological source,¹ e.g. microorganisms such as bacteria and fungi, but also by higher organisms like plants and animals, and often exhibit interesting biological and pharmacological activities.² These compounds are usually referred to as secondary metabolites since they are not essential for the producer organisms' life cycle. However, microbial secondary metabolites often have important functions in the biology of the producer organisms, e.g. as protective agents against pathogens or predators in defense mechanisms, as signaling molecules in cell development, differentiation, inter- and intraspecies communication and induction of sporulation, as motility factors, and as siderophores for efficient metal uptake.^{3–6} In many cases, the functions of natural products in their native context are not known. A distinguished class of microbial natural products that attracted attention over the past decades are the lipopeptides (LPs), which are composed of a fatty acid moiety linked to a peptide core that can either be linear or cyclic,^{7,8} and can further include several nonproteinogenic or unusual amino acids such as chemically modified (*O*-methylated, *N*-methylated, β -hydroxylated, halogenated amino acids), *D*-configured or β -amino acids. In conjunction with the occurrence of various lipid chains given by differences in the length, the oxidation state and the degree of branching of the acyl chain, microbial LPs constitute a structurally diverse compound class occupying an enormous chemical space (Figure 1).^{6,8–10} Among the different LP classes, the cyclic lipopeptides (CLPs) are most abundant and are produced by various bacteria such as soil-dwelling *Pseudomonas* spp.,⁶ *Bacillus* spp.,⁶ *Streptomyces* spp.,¹¹ *Actinomyces* spp.,¹² as well as cyanobacterial *Microcystis* spp. found in marine habitats.¹³ In addition, fungal species such as *Aspergillus* spp. and *Candida* spp. are capable of producing a vast array of CLPs amongst others.^{14,15} Due to their potent medically relevant activities and their distinct amphiphilic properties, CLPs came recently into the focus of pharmaceutical companies and other industries.^{16,17} The cyclic lipopeptide antibiotic daptomycin (Figure 1A), which is produced by *Streptomyces roseoporus*,^{18,19} is the first approved member of the CLP family for the treatment of systemic infections caused by Gram-positive and often multi-resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA).²⁰ Beside the bactericidal activities,^{18,21,22} CLPs have also been found to exhibit promising antifungal,^{23–26} antiviral,^{27,28} anticancer,^{29,30} immunosuppressant,³¹ immunomodulating,³² and hemolytic properties.^{33,34}

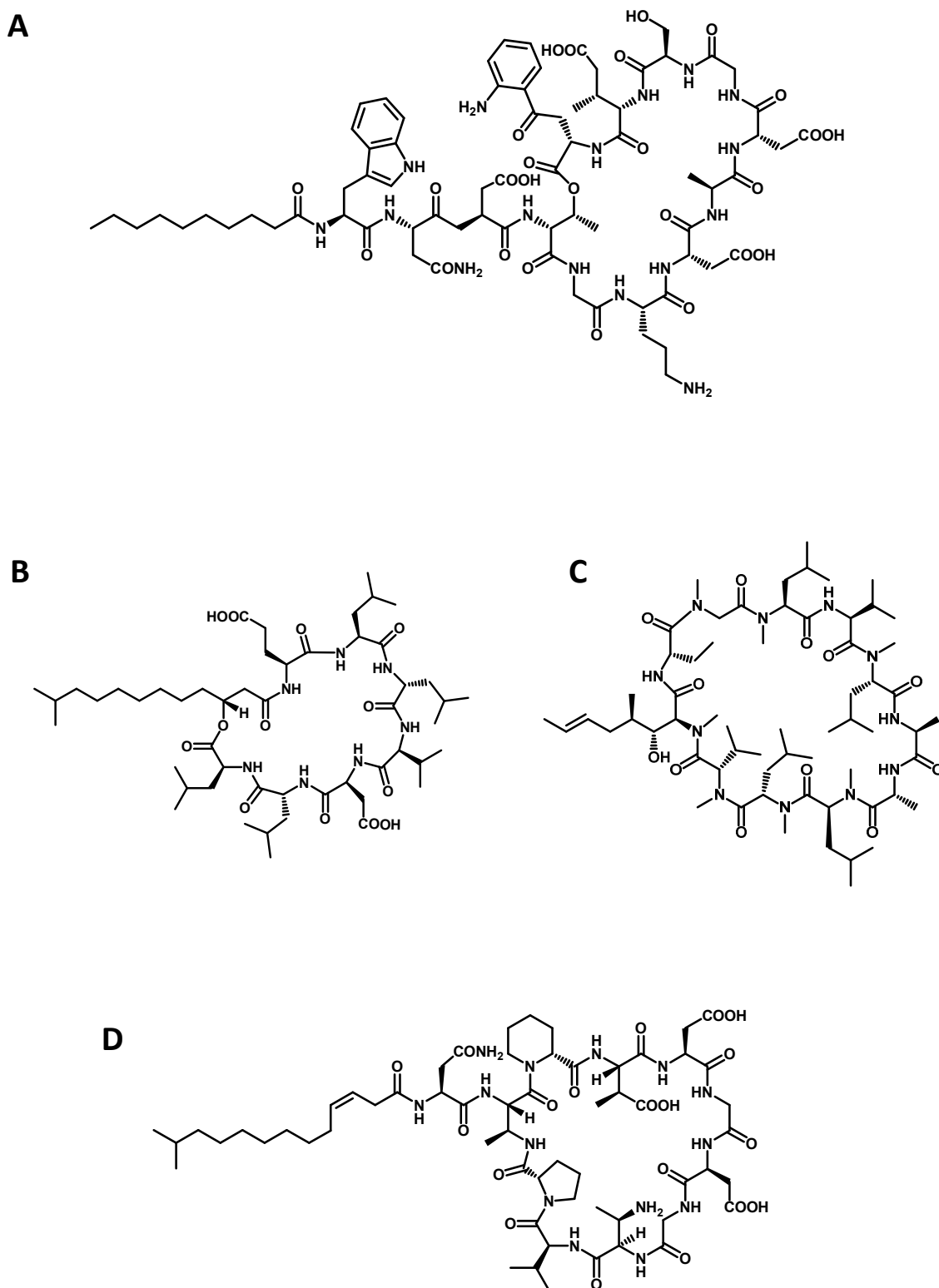


Figure 1. Prominent cyclic lipopeptides of microbial origin. **A:** Structure of the antimicrobial compound daptomycin produced by *Streptomyces roseoporous*. **B:** Structure of surfactin A from *Bacillus subtilis* exhibiting multiple pharmacological activities. **C:** Structure of the immunosuppressive agent cyclosporine A from *Tolypocladium inflatum*. **D:** Structure of the antifungal compound friulimicin B produced by *Actinoplanes friuliensis*.

Interestingly, many CLPs have been recognized as effective biosurfactants,³⁵ which find commercial application in environmental industries for the treatment of soil and water due to their ability to enhance biodegradation and bioavailability by improving the solubility of poorly water-soluble pollutants.³⁶ In addition, CLPs are widely used in other fields such as crop protection,⁷ and as ingredients of cosmetic formulations.^{37,38} However, CLPs also fulfill a wide range of functions in their producer strains or in their natural environment. Recent studies on the versatile functions of microbial CLPs from *Pseudomonas* spp. and *Bacillus* spp. revealed their roles in the protection against predatory microbes, active cell movement on surfaces and the formation of complex biofilms, thereby enabling the microorganisms to colonize novel habitats.⁶ Furthermore, functions as signaling molecules for coordinated cell growth and differentiation, in the biodegradation of xenobiotics and as metal chelators have been recently reported for CLPs.⁶ The latter ability can also essentially contribute to the pharmacological activity of CLPs as recently described for the calcium-dependent compounds daptomycin and friulimicin (Figure 1D).^{39–41} The intriguing structural diversity of CLPs together with their remarkable potential in the application as highly effective surfactants, biological pesticides, cosmeceuticals and in particular as drugs exhibiting novel modes of action, underscore the importance of searching new LP scaffolds in the yet underexplored microbial species.

1.2 Myxobacteria – An Underexplored Source of Microbial Natural Products

Historically, initiatives in natural product discovery from microbial resources particularly focused on well-studied microbes such as actinomycetes, in particular streptomycetes, and fungi.⁴² Unsurprisingly, the majority of bioactive compounds found over the past decades originated from these proficient producers of secondary metabolites and rapidly changed into stagnation as more and more known compounds have been reisolated from related species. The emerging resistance of high-risk pathogens against these formerly useful antibiotics called for the exploitation of yet untapped microbial habitats and ecological niches in order to isolate novel microbes and their corresponding secondary metabolites. Among the underexplored microbial sources for bioactive natural products, the myxobacteria represent a promising resource as proficient producers of pharmacologically interesting natural products. Myxobacteria are Gram-negative, rod-shaped, slime bacteria belonging to the group of δ -proteobacteria (order *Myxococcales*). They can be ubiquitarily found around the globe and occupy diverse habitats such as soil, dung and marine habitats.⁴³ Myxobacteria exhibit unique and characteristic features making them outstanding prokaryotes, which is also reflected by

their exceptional genomic capacities. The myxobacterium *Sorangium cellulosum* So0157-2 possesses the largest bacterial genome discovered so far (14.8 Mbp),⁴⁴ and average genome sizes of 9-15 Mbp with a high GC content of about 70% are usually common for myxobacterial strains.^{45,46} The enormous genetic information used by these bacteria especially attributes to their complex cellular developmental cycle (Figure 2), which is adapted to various environmental conditions, their distinct social behavior, their ability to cooperatively glide on surfaces, their ability to prey on other microbes and their potential to produce a plethora of bioactive secondary metabolites.^{47,48}

Most myxobacteria are chemoorganotrophic microorganisms,⁴⁹ which means that they are reliant on insoluble organic substances to maintain their cellular processes. In order to find nutrients in their habitat, myxobacteria are able to move in swarms in a chemotaxis-like manner, although they do not possess flagella or similar structures for directed movement like other bacteria.⁵⁰ In contrast, flagella-independent swarming motility in myxobacteria is mediated by two different motility systems, which are referred to as social (S-) and adventurous (A-) motility.⁵¹ S-motility plays a crucial role in multicellular swarms (Figure 2A) for both predation and fruiting body formation,⁵²⁻⁵⁴ another feature which was originally thought to be a characteristic of eukaryotic fungi. S-motility is triggered by different extracellular components including exopolysaccharide (EPS) containing fibrils, lipopolysaccharide and retractile type IV pili, which provide direct cell-cell contact.^{55,56} During the swarming process the pili, which are located at the leading pole of the cell, are attached to the EPS of adjacent cells, thereby coordinating swarm movement by multiple cycles of extension and retraction.⁵⁷ If necessary, this machinery can be reassembled at the opposite pole of the cell to change the direction. In contrast, A-motility describes the movement of single cells at the edge of a swarm to explore new habitats, thereby leaving behind an extracellular matrix, that 'paves the way' for the following companions.⁵¹ Although a large set of genes was identified to be involved in the A-motility system over the past decades,^{58,59} the precise mechanisms are still not understood in detail. Two main hypotheses have been proposed, which try to explain the movement of A-motile cells both suggesting the involvement of a proton motive force, which drives different protein motors.^{58,60} There is consensus in both models, that the traits of extracellular slime produced by the gliding bacteria plays a crucial role in bacterial A-motility and in facilitating cell adhesion to the surface they move on. In addition, it has been proposed that the deposited slime also contributes to the social capabilities in myxobacteria.⁶¹ Interestingly, previous studies demonstrated the effect of LPs on motility in various bacterial producer strains.⁶ LP-deficient

mutant strains lost their ability to glide on solid surfaces suggesting the biological function of LPs to help the bacteria to translocate to nutrient-rich habitats, amongst others.

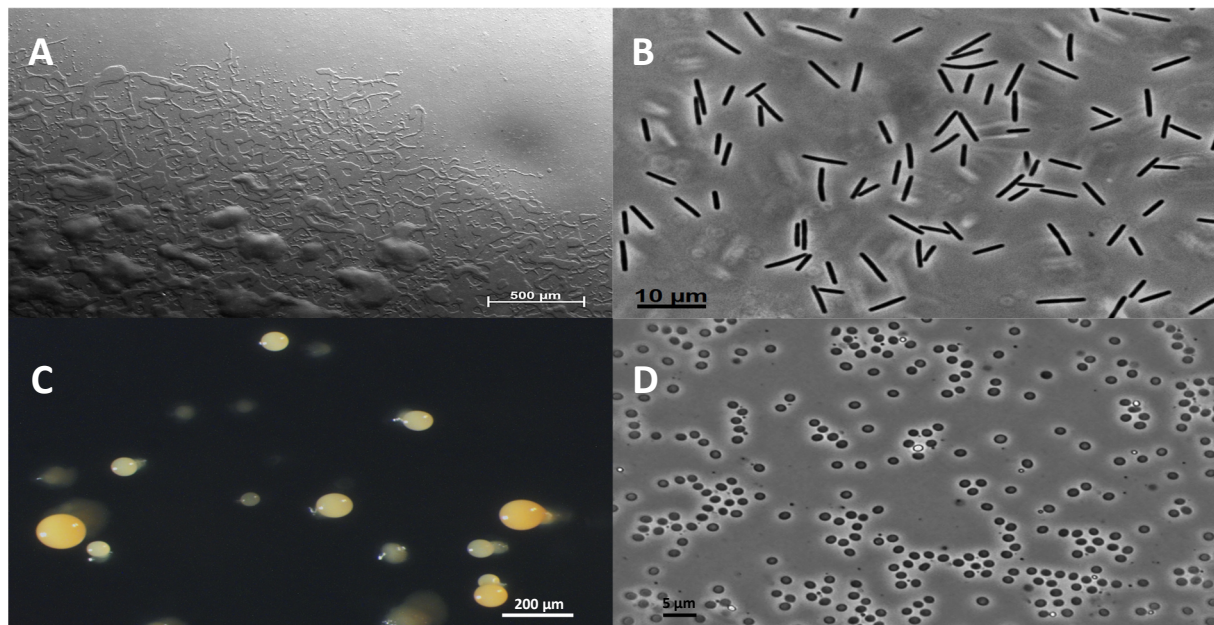


Figure 2. Developmental stages of *Myxococcus xanthus* on solid agar (pictures taken by Ronald Garcia). **A:** Swarming colony showing flare edges. **B:** Vegetative cells. **C:** Developing fruiting bodies on agar. **D:** Slide mount of myxospores. **A/C:** stereo photomicrographs, **B/D:** phase contrast photomicrographs.

Besides their unique swarming motility, myxobacteria are distinguished by a complex life cycle, which largely relies on the supply of nutrients in their habitat. Myxobacteria are capable of coordinated predation by making use of the described motility systems in conjunction with segregation of lytic enzymes in order to degrade suitable biopolymers or to prey on other bacteria or fungi by the action of additionally secreted antibiotics.^{62–64} When the continuous nutrient supply becomes limited, myxobacteria enter several developmental stages, which are characterized by the formation of multicellular fruiting bodies from thousands of vegetative cells (Figures 2B/C). Prerequisites for the development of fruiting bodies are a critical cell density, which requires motility on solid surfaces, recognition of the limited nutrient supply as well as a complex interplay of inter- and intracellular signaling pathways.⁴⁸ In *M. xanthus*, it was shown that a reduced cell movement supports cell aggregation to form fruiting bodies.⁶⁵ In the course of this process, vegetative cells undergo significant morphological changes including the formation of spherical cells, which are enveloped by several lipid layers. Under constant starvation conditions, more and more cells aggregate to form the fruiting bodies, in which approximately 10% of the cells develop into viable myxospores (Figure 2D).⁶⁶ Myxospores are highly resistant to heat, desiccation, pH value, UV radiation and the continued absence of nutrients,⁶⁷ which displays an efficient

strategy to survive under extreme environmental conditions. Most of the cells become lysed during fruiting body development,⁶⁸ thereby providing essential nutrients for the sporulation process and the differentiation of myxospores. When nutrient supply is restored, myxospores germinate and rearrange to vegetative cells, which are again able to glide in cooperative swarms, thereby restarting the myxobacterial growth cycle.⁶⁶ In analogy to the motility systems described above, microbial LPs seem to have biological functions linked to the producer's ability to form fruiting bodies and other biofilms. In a recent study, fruiting body development in a *M. xanthus* mutant strain exhibiting accelerated production levels of the lipohexapeptide myxochromide A was dramatically downregulated.⁶⁹ In *Bacillus* sp. and *Pseudomonas* sp. LPs were shown to likely play essential roles as inducers of multicellularity by stimulating morphological rearrangements.⁶

Taxonomically, myxobacteria display a coherent group comprising three suborders (*Cystobacterineae*, *Nannocystineae* and *Sorangiiineae*), 11 families and 29 genera, which were validated at the time of writing this thesis,⁷⁰ but the aforementioned numbers continue to increase due to innovative biodiversity mining. Traditionally, classification of novel isolates was realized by investigation of the unique phenotypic and morphological features.⁷¹ It turned out soon that this strategy alone is not sufficient to certainly assign the novel strains to the existing families and genera as many strains were misclassified in the past, which needed to be carefully revisited. Today, a combination of physiological, phenotypic and genetic features, especially the use of 16S rRNA genes, complemented by instrumental techniques such as GC-MS analysis of specific fatty acids contributes to the clarification of the phylogenetic position of novel taxa.^{72–74} In addition, analysis of 16S DNA sequences as well as coding DNA sequences or protein sequences allows for the *in silico* reconstruction of phylogenetic trees to illuminate strain phylogenies. Phylogenetic studies can also be applied to the underlying natural product pathways, which direct the production of certain secondary metabolites, by *in silico* analysis of DNA sequences coding for the corresponding pathways. In this way, it is possible to directly link strain phylogeny to pathway phylogeny in order to investigate the evolutionary origins, relationships and histories of these secondary metabolite pathways and the respective producer strains.

Over the past decades, myxobacteria have been recognized as proficient producers of such secondary metabolites, mainly polyketides and peptides, exhibiting an intriguing diversity concerning their chemical structures and bioactivities ranging from antifungal, antibacterial, antiviral and antimalarial to cytotoxic, immunosuppressive and antioxidative properties.⁷⁵ This is in fact expressed by their aforementioned huge genetic capacities pointing up their

biosynthetic potential. Around 10% of a myxobacterial genome encode for natural product biosynthetic pathways resulting in a number of usually 10 to more than 20 biosynthetic pathways per genome.^{45,76} Until today, more than 100 distinct natural product core structures harboring partly unusual structural elements and about 600 derivatives have been isolated and characterized.⁴⁷

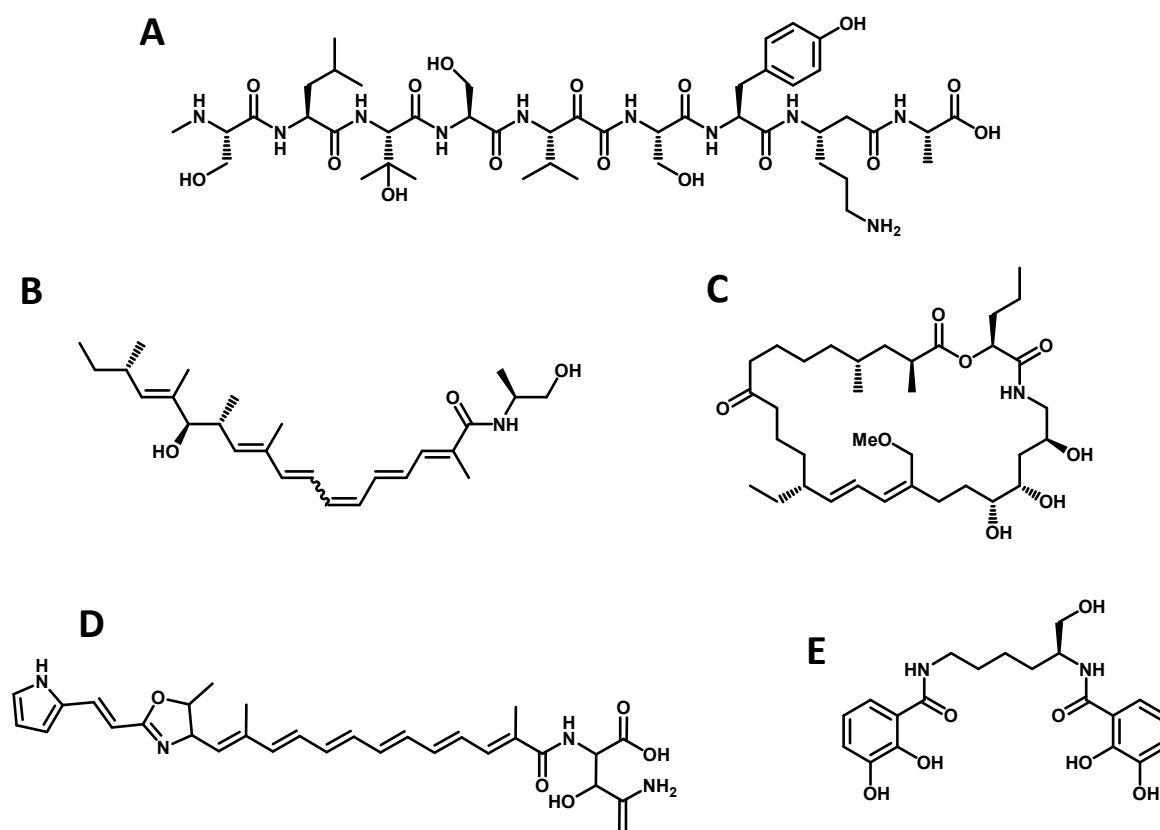


Figure 3. Structures of selected secondary metabolites produced by the model strain *Myxococcus xanthus* DK1622. **A:** myxoprincomide. **B:** myxalamid A. **C:** myxovirescin A. **D:** DKxanthen-534. **E:** myxochelin A.

Most myxobacterial natural products were isolated from *Sorangium*, *Myxococcus*, and *Chondromyces* strains, as compound screening particularly focused on these genera in the past.⁷⁷ The model strain *M. xanthus* DK1622 is e.g. capable of producing a vast array of diverse secondary metabolites such as myxoprincomide,⁷⁸ myxovirescin A,⁷⁹ myxalamides,⁸⁰ DKxanthenes,^{81,82} and myxochelin A (Figure 3),⁸³ exhibiting various natural functions or antibiotic activities. Lipopeptides were found in myxobacteria as well but compared to the extensively studied LP producers *Pseudomonas* spp. and *Bacillus* spp., LP biosynthesis in myxobacteria is currently underexplored. The best studied myxobacterial LP pathways direct the biosynthesis of myxochromides A from *M. xanthus* DK1622,⁸⁴ and myxochromides S

produced by *Stigmatella aurantiaca* DW4/3-1.⁸⁵ Additionally, the structure of another member of this LP family, myxochromide B₃ from an unclassified *Myxococcus* sp. strain, was previously elucidated.⁸⁶ Cystomanamides A-D produced by *Cystobacter fuscus* MCy9118 display a novel class of myxobacterial linear LPs, which was recently structurally characterized.⁸⁷ Both LP families do not exhibit any biological activity under the applied conditions in standard bioactivity screenings, but one has to consider that LPs often exhibit highly specific functions or bioactivities, which are frequently not detected in these screens. Another type of myxobacterial lipopeptides are the lipothiazoles, which have been detected in extracts of *Sorangium cellulosum* GT47.⁸⁸ However, the most prominent myxobacterial secondary metabolite, which was approved by the FDA in 2007 as an agent against advanced breast cancer,⁸⁹ is the macrolide epothilone produced by various *Sorangium* strains.⁹⁰ In addition, five compound classes from myxobacteria targeting different medical indications are currently assessed in preclinical studies and 14 different scaffolds are extensively under investigation concerning their mechanisms of action, thereby being promising candidates for future drug development programs.⁴⁷ Considering the relatively small number of myxobacterial strains screened so far, these examples illustrate the potential of myxobacteria as a still rich source of bioactive natural products, which is far from being exhausted.

1.3 Natural Product Biosynthesis by Multimodular Enzyme Complexes

Nature developed intriguing concepts for the formation of bioactive substances during billions of years, which led to diverse groups of natural products. Many of these concepts include the supply of simple precursor molecules, which can be subsequently transformed into rather complex metabolites either by a subset of single enzymes or by multienzyme complexes. For instance, plant polyphenolic compounds like flavonoids or stilbenes usually derive from simple phenylpropanoids using the Shikimate pathway,⁹¹ whilst terpenes are made of isoprene or its activated forms,⁹² respectively, in multistep enzymatic conversions. In myxobacteria, most of the secondary metabolites characterized so far belong to the class of polyketides and nonribosomal peptides, many of them were shown to exhibit antibiotic and/or cytotoxic activities.⁴⁷ Such polyketides and nonribosomal peptides also consist of simple building blocks that can be assembled to form a vast number of structurally diverse compounds. Condensation of these precursors is mediated by polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS), giant multimodular enzyme complexes, which act as molecular assembly lines in natural product formation.^{93,94} The genes encoding these

assembly lines are usually clustered, thereby simplifying the identification and analysis of gene clusters that are associated with natural product biosynthesis.

1.3.1 Biochemistry of PKS Machineries

Polyketide synthases (PKS) are evolutionary linked to and derived from fatty acid synthases (FAS).⁹⁵ FASs have been intensively studied since their discovery, and it was quickly recognized that these systems can appear either as a set of distinct enzymes or as multienzyme complexes, which convert simple carboxylic acid monomers into saturated fatty acids using an iterative assembly strategy.^{96,97} In general, fatty acids are biosynthesized via successive, decarboxylative Claisen-like condensations of the short-chain carboxylic acids acetyl-CoA and malonyl-CoA.⁹⁸ The dimeric FAS machinery involves seven catalytic functionalities to specifically select the precursor molecules and to further process the growing acyl chains.⁹⁹ Malonyl/acyl transferase (MAT) is responsible for transferring the starter unit acetyl-CoA and the extender units malonyl-CoA onto the acyl carrier protein (ACP). ACPs are four-helix bundles consisting of 80-100 amino acid residues. The substrates are covalently linked to a conserved serine residue of the ACP as a thioester using a phosphopantetheinyl (ppant) linker, which is derived from CoA.¹⁰⁰ The ppant moiety acts as a flexible arm and can subsequently guide the substrates and intermediates to other catalytic functionalities for further processing, thereby providing an efficient biosynthetic system, which is not dependent on and limited to diffusion control as it is the case for isolated enzymes. Phosphopantetheinyl transferases (PPTases) catalyze the posttranslational transfer of the CoA-derived linker molecule onto the ACP.^{101,102} After loading of the starter and extender units, β -ketoacylsynthase (KS) catalyzes C-C bond formations by employing decarboxylative condensations between KS-bound acetyl-CoA starter units or biosynthetic intermediates and extender units, which are bound to the ACP.¹⁰³ The repetitive extension of the growing ACP-bound acyl chain with C_2 carboxylic acid monomers results in poly- β -oxo compounds, which are further processed by a set of reductive enzymes to build up fully reduced fatty acids. This set of reductive domains of FASs consists of three catalytically active enzymes: a ketoreductase (KR), a dehydratase (DH) and an enoylreductase (ER). Stepwise processing of the introduced extender units first leads to the conversion of the β -ketoacyl functionality to β -hydroxy compounds by NAD(P)H-dependent KR catalysis. DHs subsequently catalyze water elimination, thereby forming a C=C double bond, which is finally being reduced to a saturated C-C bond by ERs (Figure 4).⁹⁶ The final product of the FAS is being cleaved off the ACP by a thioesterase (TE). The product spectrum, an iterative FAS is able to produce, is quite limited. Only acetyl-

CoA and malonyl-CoA are being used as starter and extender units usually providing saturated fatty acids with chain lengths of C_{14} - C_{18} .⁹⁹

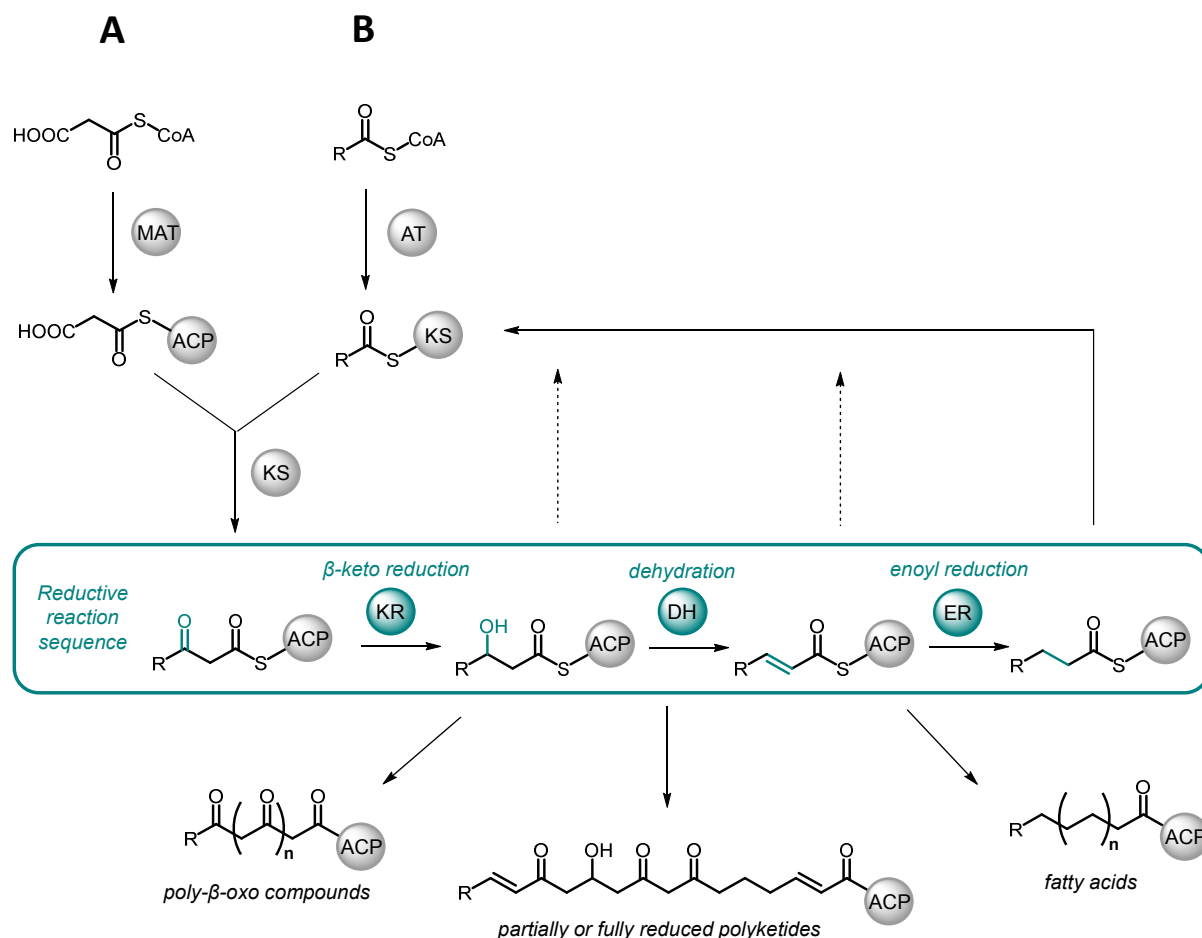


Figure 4. Reaction scheme illustrating the stepwise formation of products derived from FAS and PKS machineries (adapted from Smith and Tsai)⁹⁶. **A:** FAS pathway leading to the formation of fully reduced fatty acids. Condensation of an acetyl-CoA starter unit with the first extender unit malonyl-CoA yields a β -oxo diketide, which is successively reduced by a set of reductive domains consisting of the KR, DH and ER domains prior to the following round of chain elongation. **B:** PKS pathway leading to the formation of various compound species. The β -oxo diketide produced in the first chain extension can undergo several reaction scenarios leading either to the formation of poly- β -oxo compounds or to the biosynthesis of partially or fully reduced polyketide structures depending on whether the reductive domains are present/active or not present/inactive in the corresponding module.

Studies about FASs essentially contributed to the still expanding knowledge in the field of PKS research,^{94,104} and since sophisticated genomics, proteomics and structural methodologies have become routinely applicable, it was possible to gain detailed insights into how such giant megasynthase complexes are capable of producing structurally diverse and complex polyketides. In contrast to FASs, most bacterial PKSs are composed of so-called modules, which contain catalytically active domains fulfilling the tasks described above for FASs and which are arranged as a kind of molecular assembly line.¹⁰⁵ A minimal PKS

module is composed of at least an acyltransferase (AT) domain, a KS domain and an ACP. In such multimodular PKSs, every module is thus responsible for one round of chain elongation during biosynthesis. Consequently, the number of functional modules in the PKS system usually correlates with the number of incorporated building blocks, which is known as the co-linearity rule.¹⁰⁶ Although FASs and PKSs share basically closely related catalytic mechanisms, there are significant differences between both systems concerning building block selection, different degrees of reduction after each elongation step, the use of additional modifying domains/enzymes as well as chain termination mechanisms, thereby leading to an unique structural diversity of polyketides.⁹⁶

Unlike in FASs, where the MAT domain exhibits dual-specificity for acetyl-CoA and malonyl-CoA for both substrate loading and chain extension, the AT domains present in modular PKSs usually have distinct specificities either for malonyl- or methylmalonyl-CoA.¹⁰⁷ In addition to the standard building blocks described, the structural diversity of polyketides can be dramatically increased by selection of various unusual starter and extender units. Besides acetyl-CoA, possible starter units are e.g. isovaleryl-CoA, benzoyl-CoA, cyclohexanyl-CoA, cinnamoyl-CoA and many more.¹⁰⁸ Diversification by using alternative extender units is less common but there are some examples known, e.g. methoxymalonyl-CoA, ethylmalonyl-CoA, hydroxylmalonyl-CoA and aminomalonyl-CoA.^{109–111}

The sequential use of the reductive domains (KR, DH, ER) increases the structural diversity of polyketides as well. The poly- β -oxo compounds produced by multiple rounds of chain extension can either form complex polyphenolic compounds by undergoing orchestrated cyclizations or can be fully or partially reduced by KR, DH and ER domains (Figure 4). While the reductive domains are necessarily required in iterative FASs for fatty acid biosynthesis, it is merely optional in PKS machineries.⁹⁴

In addition, polyketides can be further structurally diversified by the action of unusual domains, enzymatic mechanisms and assembly line organizations.⁹⁴ Several optional domains integrated into PKS assembly lines can be found such as methyltransferases (MT) that catalyze SAM-dependent *C*- or *O*-methylation.¹¹² In addition, tailoring enzymes that act in *trans* transform the often pharmacologically inactive precursor polyketides into potent bioactive natural products. The most common tailoring enzymes are glycosyltransferases that attach sugars – predominantly hexoses, which can also be highly modified – to hydroxyl groups of the polyketide scaffold, thereby increasing the hydrophilicity, oxygenases that introduce hydroxyl or epoxide functionalities and halogenases that catalyze chlorination or bromination, which is often found in polyketides from marine origin.¹¹³ Modifications that are

catalyzed by these dedicated tailoring enzymes usually affect biological activity, metabolic stability and/or solubility. Apart from that, there are several unusual mechanisms, which deviate from the standard ‘textbook’ enzymology of PKS systems, thereby violating the co-linearity rule.¹¹⁴ In some biosynthesis pathways, modules appear to be skipped during chain assembly. Based on a study of an engineered PKS system, module skipping is most likely facilitated by ACP-to-ACP transfer of the nascent intermediate, while the ACP of the skipped module carries over the acyl chain to the downstream extension module.¹¹⁵ In opposite cases, PKS assembly lines contain fewer modules than should be present according to the number of incorporated building blocks in a certain polyketide. This phenomenon, which is called ‘stuttering’, implies the back transfer of the growing polyketide chain from the ACP to the KS domain of the same module, repeated loading of the free ACP domain followed by condensation.¹¹⁴ The affected polyketides are often by-products, while the major product conforms to the co-linearity rule. However, ‘stuttering’ is more appropriately referred to as ‘programmed iteration’, since this process seems to play an important role in the biosynthesis of the major polyketides as well.¹¹⁴

The chain termination also contributes to the intriguing structural diversity of polyketides. Whilst TE domains from FAS systems only hydrolyze the fatty acid-CoA thioester releasing linear products, the TE domains in PKSs often catalyze intramolecular cyclizations to form complex macrolides.¹¹⁶

Furthermore, PKSs can interact with another class of multimodular enzyme complexes, which includes the incorporation of a vast array of amino acids into the acyl chain. The observed structural diversity accounts for numerous functions polyketides have in Nature, which are largely linked to secondary metabolism, thereby discriminating them from fatty acids produced by FASs, which are involved in primary metabolism. The seemingly endless options to combine different domains, modules and subunits in conjunction with the possibility to interact with different classes of molecular assembly lines give rise to an unprecedented structural diversity of polyketides exhibiting diverse biological activities.

1.3.2 Biochemistry of NRPS Machineries

Scientific reports about the biosynthesis of biopolymers that consist of amino acid monomers have traditionally their roots in the field of ribosome research. Protein synthesis is thereby mediated by a huge protein/RNA complex that coordinates a rather complicated interplay between various catalytic centers and enzymes finally facilitating the translation of the information encoded in the mRNA into a polypeptide chain.¹¹⁷ Furthermore, the ribosome is

also able to produce small genome-encoded precursor peptides found in some microorganisms, so-called RiPPs (ribosomally synthesized and posttranslationally modified peptides), which are usually being highly posttranslationally modified to become biologically active.¹¹⁸ In contrast, many bacteria and fungi make use of another biosynthetic machinery to produce structurally diverse peptides. The multimodular nonribosomal peptide synthetases (NRPS) employ a highly similar biosynthetic strategy as already described for PKSs using peptidyl carrier proteins (PCP) as a substrate shuttling platform and canonical as well as non-canonical amino acids as building blocks for the biosynthesis of nonribosomal peptides discriminating them from their counterparts synthesized by the ribosome.⁹³ In analogy to PKS modules, every NRPS module consists of several catalytically active domains, which are responsible for building block activation (A, adenylation domain), substrate transfer (PCP) and condensation/chain elongation of usually *L*-configured amino acids (^LC_L, condensation domain). Similar to PKS machineries, most NRPS assembly lines contain a loading or initiation module, which often harbors only an A domain as well as a PCP domain in order to start the biosynthesis. In addition, termination modules contain TE domains to release the final linear product by hydrolysis. In many cases, the TE domains are also able to catalyze macrocyclizations, which lead to complex macrolactams, macrolactones and related compounds.^{119,120} Similar to short-chain carboxylic acids in PKS biosynthesis, the amino acids used as building blocks in NRPS systems need to be presented in an activated form. Unlike the related AT domains in PKS modules, A domains select the corresponding amino acid and catalyze the formation of an aminoacyl-adenylate (AMP-ester) using adenosine-triphosphate (ATP) and magnesium ions as cofactors.¹²¹ When two adjacent NRPS modules are loaded with amino acids, the downstream C domain catalyzes peptide bond formation, which yields a peptide intermediate that is bound to the PCP of the second downstream module. During this reaction, the upstream PCP-bound donor amino acid is attacked by the nucleophilic amino group of the downstream PCP-bound acceptor amino acid (Figure 5).¹²² In general, A domains specifically recognize amino acids depending on the cavity in their active sites, which is largely determined by a certain set of amino acid residues.

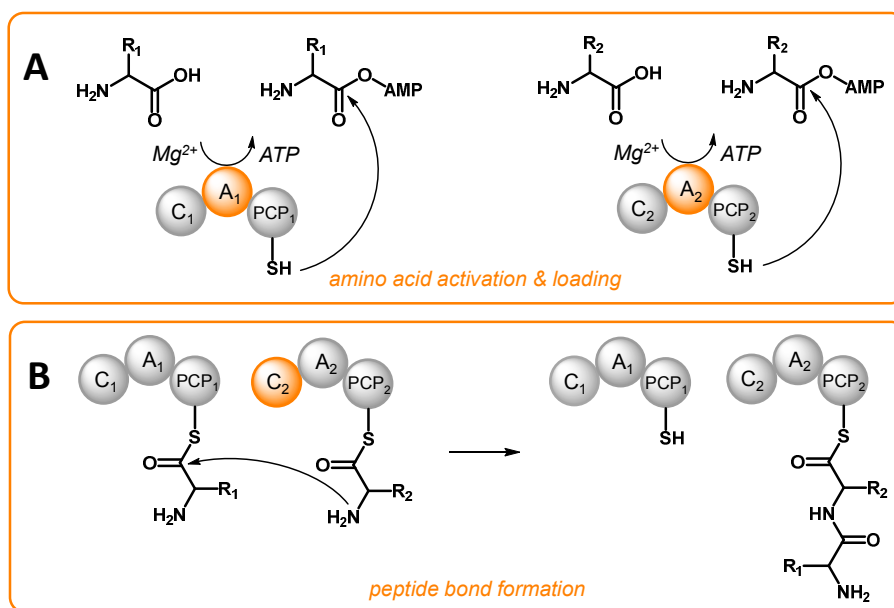


Figure 5. Mechanism of peptide bond formation in NRPSs. Active domains are highlighted in orange. **A:** Adenylation domains catalyze amino acid activation via formation of an aminoacyl-adenylate and subsequent transfer to the adjacent peptidyl carrier protein, thereby consuming ATP. **B:** Condensation of two amino acids attached to neighboring NRPS modules catalyzed by the downstream condensation domain.

This amino acid specificity-conferring code has been intensively investigated by Stachelhaus *et al.* and other research groups and provides an useful tool for the *in silico* prediction of amino acids selected by distinct A domains.^{123,124} However, substrate specificities of A domains are very often not stringent allowing the activation of more than one amino acid by an A domain. A domains of NRPS assembly lines are not only capable of activating and incorporating proteinogenic amino acids but also of an almost uncountable number of non-proteinogenic amino acid precursors. These precursors are usually being synthesized *trans* by a set of dedicated enzymes and often originate from proteinogenic amino acids or other primary metabolites and can subsequently be loaded onto PCP domains.¹²⁵ Prominent examples are 2,3-diaminopropionate (DAP), which is made from *L*-serine and the amino group donor *L*-ornithine,¹²⁶ as well as *L*-*p*-hydroxyphenylglycine, which is biosynthesized in a transamination-like conversion, in which the amino functionality of *L*-tyrosine is transferred to the acceptor *p*-hydroxybenzoylformate.¹²⁷ The β -hydroxylation of amino acids is another possibility to further extend the scope of precursor molecules that can be incorporated by NRPS systems during assembly of the NRPS-derived peptide. Oxidative enzymes that catalyze β -hydroxylation of amino acids are usually cytochrome P450 monooxygenases (CYPs), di-iron monooxygenases or iron/ α -ketoglutarate-dependent dioxygenases and can act either on free amino acids or on PCP-bound amino acids of external A-PCP didomains.^{128–131}

β -hydroxylation has been reported for the proteinogenic amino acids *L*-phenylalanine, *L*-tyrosine, *L*-leucine and *L*-valine, amongst others, as well as for the non-proteinogenic counterpart *L*-*p*-aminophenylalanine.^{131–133} Furthermore, non-proteinogenic amino acids derived from the primary metabolism can be directly introduced into NRPs, e.g. *L*-ornithine.¹⁸ Similar to PKS enzymology, there are additional catalytic domains that can be integrated into the assembly line in order to modify single building blocks or the mature peptide. For instance, *D*-configured amino acids can be found in many NRPS products. Since *D*-configured amino acids usually do not occur in the primary metabolism, they are directly converted from *L*-amino acids by epimerization domains (E), which are located downstream a PCP domain in NRPS modules.^{134,135} In those cases, the C domain of the following module exhibits an altered stereospecificity (^DC_L). Additionally, typical modification reactions comprise *N*- or *C*-methylations catalyzed by methyltransferases (MT), glycosylations catalyzed by glycosyltransferases (GT), hydroxylations and oxidative cross-coupling reactions mainly mediated by CYPs, transaminations (AMT, aminotransferase domain) and heterocyclizations (HC, heterocyclization domain or Cy, cyclisation domain), rarely also cyclopropanations (Cyp, cyclopropanase domain).^{136–138} Heterocyclizations in NRPS systems frequently involve serine, threonine and cysteine residues, which can be converted into 5-membered oxazoline or thiazoline rings, respectively. Oxidase domains (Ox), which often appear to be present together with HC domains are able to catalyze dehydrogenation of the oxazoline or thiazoline yielding the respective oxazole or thiazole.¹³⁹ In addition, reduction by dedicated reductases (Re) leads to the formation of oxazolidine or thiazolidine rings.¹⁴⁰ Moreover, Re domains are found to be involved in reductive product release mechanisms, which facilitates the formation of carboxylic acid and aldehyde moieties, amongst others.¹⁴¹ Module skipping, as already described for PKS biosynthesis, can also occur during NRP formation and contributes to the expansion of secondary metabolites, which can be made by a single assembly line.⁸⁴ Since the PKS and NRPS biosynthetic logic and assembly line organizations share many striking similarities following modular biosynthetic principles, it is not surprising that PKS/NRPS hybrid assembly lines evolved over time and greatly contribute to the incredibly versatile pool of microbial natural products.

1.3.3 PKS/NRPS Hybrid Megasyntetases

In principle, modular PKS and NRPS megaenzymes share the same biosynthetic strategies, which comprise building block selection, priming of a carrier protein and condensation with a downstream extender unit. ACPs and PCPs with their flexible ppant linkers therefore guide the substrates or biosynthetic intermediates along the assembly line by providing them in the active site of the downstream domain that catalyzes condensation of the ACP/PCP-bound templates (C or KS). Since the functions of domains or even modules of PKS and NRPS systems and their arrangement are highly similar, the discovery of hybrid PKS/NRPS assembly lines was the logical consequence.¹⁴² This concept further diversifies microbial natural products that are made either of PKS or NRPS assembly lines and combines enzyme functions and features of both production routes. An example of a prominent PKS/NRPS hybrid system is the heptamodular tubulysin assembly line from the myxobacterium *Angiococcus disciformis*, which directs the biosynthesis of the potent cytotoxic peptides tubulysins exhibiting antimitotic activity (Figure 6).¹⁴³

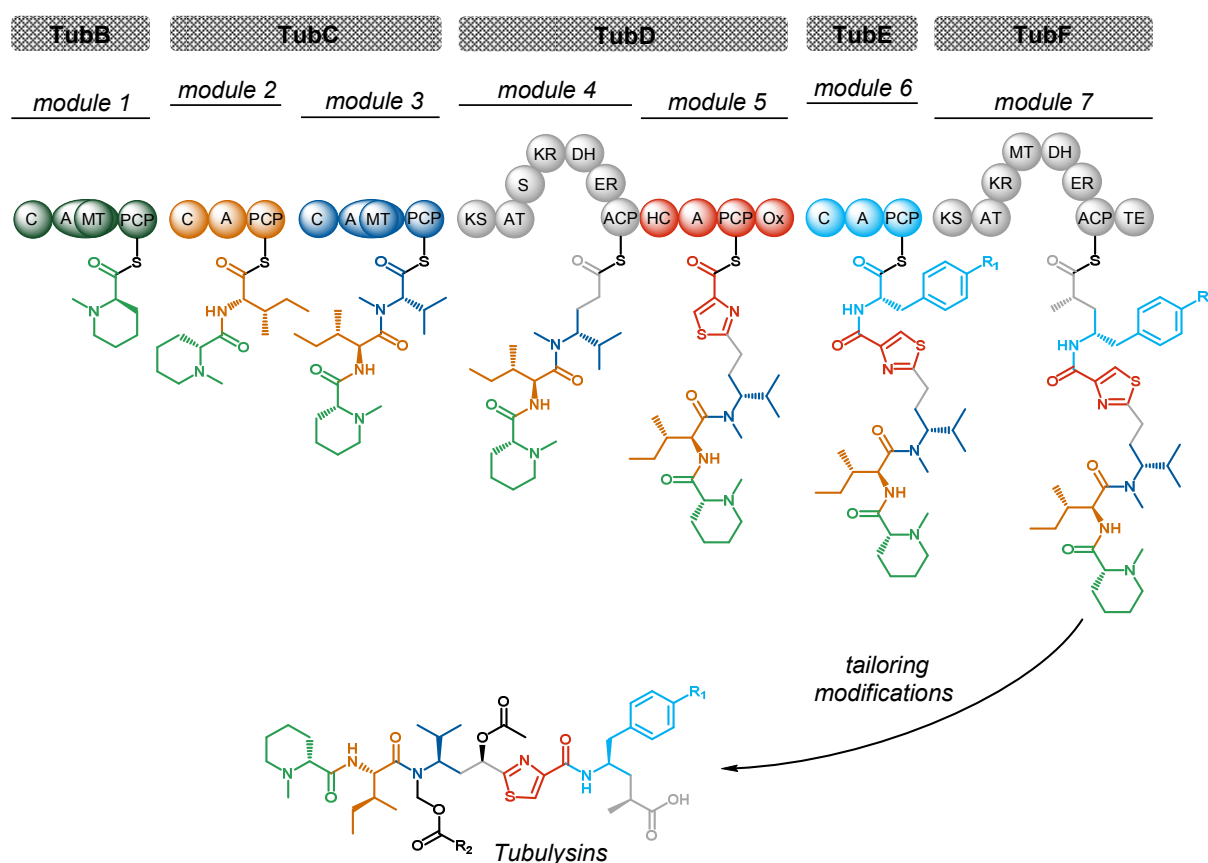


Figure 6. Assembly line architecture and proposed biosynthetic pathway of the tubulysins. The PKS/NRPS hybrid machinery consists of seven modules, two PKS (illustrated in gray) and five NRPS (colored) modules and facilitates the incorporation of acyl units plus a set of proteinogenic, unusual and modified amino acids. The coloring of the catalytic domains corresponds to the biosynthetic intermediate and product structures. Abbreviations of catalytic domains are explained in the main text; S = spacer.

The tubulysin biosynthetic machinery involves five NRPS and two PKS modules distributed over five protein subunits (TubB-F), which catalyze successive elongation of the growing peptide chain and an additional enzyme, which supplies the starter amino acid (TubZ). Several deviations from ‘textbook’ NRPS enzymology, which were discussed above, lead to the production of a highly unusual natural product. The initiation module starts biosynthesis with the incorporation of the unusual amino acid pipecolic acid, which derives from lysine. Pipecolic acid is generated *in situ* by the action of lysine cyclodeaminase TubZ, which is also encoded in the tubulysin biosynthetic gene cluster. The amino acids isoleucine and valine are installed in the peptide chain by two subsequent NRPS modules. Both the first and the third module contain a MT domain integrated into the respective A domains. Accordingly, the corresponding amino acids were found to be methylated.¹⁴³ The most fascinating aspect of PKS/NRPS hybrid systems is how they manage the translocation of the nascent product chain at the interfaces between PKS and NRPS subunits. Transition from module 3 to module 4 constitutes the first NRPS/PKS interface in the assembly line and requires a KS domain to be able to accept the peptidyl donor from module 3 and catalyzes C-C bond formation with a malonyl-CoA extender unit selected by module 4, which contains a fully active reductive loop. In analogy, the downstream HC domain is capable of catalyzing the condensation of the acyl/peptidyl chain with cysteine, thereby forming the thiazoline moiety, which is subsequently oxidized to the thiazole ring by the Ox domain. The minimal NRPS module 6 incorporates an aromatic amino acid residue (either phenylalanine or tyrosine) followed by the condensation of the resulting intermediate with the second malonyl-CoA extender unit, which is fully reduced and methylated at the α -carbon atom via action of the reductive loop and the integrated MT domain. Product release is facilitated by the TE domain, which hydrolyzes the linear PK/NRP hybrid product to form a mature peptide that is subjected to further rare tailoring modifications.¹⁴³

Although many PKS/NRPS hybrid machineries like the tubulysin assembly line have been characterized over the past years, there is still a lack of mechanistic insights into how PKS and NRPS subunits or modules can interact with each other. Several attempts to reprogram these biosynthetic megaenzymes to produce ‘unnatural’ scaffolds failed in most cases, which suggests that intersubunit or intermodule communication is mediated by rather sensitive and fine-tuned protein interactions.¹⁴⁴ Previous studies revealed that the specific intersubunit recognition processes are mediated by so-called docking domains in PKSs and communication-mediating (COM) domains in NRPS systems, which are located at the N- and C-termini of the respective protein subunits.^{145,146} It likely seems to be that a similar

quaternary structure of both the PKS and NRPS subunits is crucial for their correct interaction and orientation in the operative assembly line. Several studies suggested that PKS as well as the cognate FAS subunits occur as homodimeric, ‘double-helical’ proteins,¹⁴⁷ whereas investigations about quaternary structures of NRPS systems are rather inconsistent.^{148,149} Monomeric as well as dimeric structures have been found using sophisticated analytical techniques and it is assumed that NRPS subunits are able to switch between both states depending on whether they interact with other NRPS subunits (monomeric) or PKS subunits (dimeric) in hybrid megasynthetases.¹⁴⁸ By now, several families of docking or COM domains have been identified and many hybrid pathways have been intensively characterized.¹⁴⁴ However, detailed insights into enzymatic mechanisms and particularly into the dynamics and protein-protein interaction networks of PKS and NRPS systems and their hybrids remain largely to be deciphered. State-of-the-art structural techniques like protein NMR, X-ray crystallography and cryo electron microscopy are currently used to study multienzyme architectures as well as domain and module contact regions, interaction surfaces and dynamic mechanisms of precursor processing/product formation. The output of those studies is assumed to have a significant impact on the rational design of PKS, NRPS and PKS/NRPS hybrid assembly lines in order to produce novel metabolites, which would otherwise not be synthesized in Nature, by combinatorial biosynthesis.

1.3.4 Bioinformatics Tools to Identify and Characterize Biosynthetic Gene Clusters

Recent advances in high-throughput shotgun genome sequencing techniques such as Illumina, PacBio, IonTorrent and 454 pyrosequencing essentially contributed to the exploding number of microbial genome sequence data over the past years.¹⁵⁰ The progress achieved so far made the sequence-based analysis of these genomes and, in particular, of biosynthetic gene clusters a routine approach in natural product research as the costs of whole genome sequencing dramatically dropped, while the sequences can be provided in unforeseen short time frames and accuracy. The approach that attempts to translate the underlying molecular genetics into isolated natural products by analyzing pathway-specific sequence motifs is known as ‘genome-mining’ and is widely used in the identification process of biosynthetic gene clusters and their corresponding secondary metabolites.¹⁵¹ The genome-mining process can be generally subdivided into three stages: 1) NGS data acquisition from natural sources, e.g. genome sequencing of microbial producers, 2) Identification of biosynthetic gene clusters and prediction of corresponding product structures using bioinformatics tools and 3) Identification, isolation and structure elucidation of the predicted secondary metabolites.

Genome-mining also allows to back correlate identified product structures to the respective biosynthetic gene clusters to gain deep insights into the enzymology of the biosynthetic processes. The modular architecture and mechanistic similarities of both PKS and NRPS biosynthetic machineries in conjunction with high sequence similarities of PKS and NRPS domains, especially on the protein level, allows for the rapid identification and comparative analysis of microbial biosynthetic gene clusters based on the genome sequence of a certain producer strain. Many *in silico* tools have been developed and refined in the PKS/NRPS field over the past decade to assist genome-mining initiatives. The most prominent tool among those is certainly antiSMASH, which provides a comprehensive platform for the analysis of microbial biosynthetic gene clusters.¹⁵² This tool requires an input DNA sequence and offers the automated genome-wide analysis and annotation of biosynthetic gene clusters based on sequence similarity compared to deposited sequences as well as specific sequence motifs, which are characteristic features of PKS and NRPS domains. These sequence motifs are referred to as core motifs, which are highly or even strictly conserved regions on the protein level of a certain domain and were previously found in obligatory core domains of both PKS and NRPS systems.^{119,153} The antiSMASH tool enables a fast assessment of the biosynthetic potential of a given producer strain, thereby enabling the possibility to further focus on interesting and unusual assembly line organizations. The annotated sequence data can be used in the following to postulate a biosynthesis model for the assembly of a certain secondary metabolite prior to experimental verification, e.g. by gene inactivation or heterologous expression, and provides a basis for the prediction of the secondary metabolite core structure by implementing additional tools covering the substrate specificities of the PKS/NRPS domains such as NRPSpredictor2.¹⁵⁴ This tool relies on different approaches for the analysis of the A domain substrate specificities, which are determined by a set of distinct amino acid residues of A domains. These amino acids have been previously identified by Stachelhaus *et al.*, and Challis *et al.* based on an A domain crystal structure and are widely known as the Stachelhaus code,^{124,155} which was recently further extended to an overall set of 34 amino acids lining the active site of the A domain in a distance of 8 Å around the bound substrate.¹²³ In addition, several other approaches have been developed to predict A domain specificities, which complement the current tool box.¹⁰⁶ In a similar way, AT domain specificities in PKS modules can be predicted using computational tools considering amino acid residues within the AT domain active sites, which e.g. discriminate between malonyl-CoA and methylmalonyl-CoA.¹⁵⁶ Bioinformatics tools allow not only the *in silico* prediction of the building blocks that are incorporated into the assembled product but also the prediction of the

stereochemical course during PKS/NRPS biosynthesis. In NRPS assembly lines, C domains can be usually classified as $^D\text{C}_L$, $^L\text{C}_L$ and starter C domains depending on the observed substrate specificities. $^D\text{C}_L$ domains catalyze the condensation between a *D*-configured donor peptide intermediate and an *L*-configured acceptor substrate, whilst $^L\text{C}_L$ domains facilitate peptide bond formation between donor and acceptor building blocks, which show both *L*-configuration. In contrast, starter C domains can be e.g. found in initiation modules of lipopeptide assembly lines, where they catalyze the covalent linkage of the acyl side chain with the first amino acid precursor, a process known as lipoinitiation.¹⁵⁷ The different types of C domains significantly differ at several positions in their core motifs, which allow assignment of uncharacterized C domains by means of *in silico* predictions.¹⁵⁸ In PKS biosynthesis, stereochemical preferences can be predicted for KR domains, which produce either an (*R*)-3-OH or an (*S*)-3-OH acyl intermediate during catalysis. There are two different types of KR domains, A-type and B-type KR domains, which differ in a highly conserved core motif, thereby making stereopreferences predictable based on sequence analysis.¹⁵⁹ The ACPs and PCPs, which are responsible for the translocation of biosynthetic intermediates in the respective megasynthetases, share a common core motif (GGH(D)SL) harboring the active site serine residue. In NRPSs, a second class of PCPs was detected, that exhibits a slightly modified but absolutely conserved signature motif (GGDSI). These PCP^E domains are specifically found in modules, in which the PCP domain directly interacts with a downstream E domain.¹³⁴ Based on comparative sequence alignments, the two PCP types can be easily differentiated.

The presented sequence-based analysis tool box has become indispensable for the discovery of novel natural products and uncommon enzymatic activities. However, the approaches described here are depending on the sequences of all available biosynthetic gene clusters and domain crystal structures, which serve as references in the existing databases as well as on algorithms developed for *in silico* predictions, which implies that the current models cannot provide the full picture. Thus, *in silico* predictions always need to be critically examined as many predictions and assignments turned out to be wrong in the past. Deviations from ‘textbook’ biosynthetic logic, violation of the co-linearity rule as well as unexpected inactivity of catalytic domains, which cannot be explained by sequence analysis so far significantly attribute to this issue. In addition, genes that encode tailoring enzymes and stand-alone domains or modules are sometimes not present in the corresponding biosynthetic gene cluster, but are located somewhere else in the genome, which makes it difficult to correlate such genes to the gene cluster of interest. As more and more gene cluster sequences will be

deciphered and provided to the scientific community, the developed models will be further refined to improve the accuracy of such predictions.

1.4 Synthetic Biotechnology Approaches in Natural Products Research

The discovery and in-depth analysis of novel natural products and the corresponding biosynthetic pathways from natural sources usually involved a low-throughput, time-consuming and expensive workflow over the past decades. This especially holds true for the investigation of natural product biosynthesis in myxobacteria as most of the candidate producer strains exhibit low growth rates or even cannot readily be cultivated under standard laboratory conditions, and are often not susceptible to genetic manipulations, e.g. to alter the complex product structures or to improve production yields, which are often found to be homeopathic, thereby hampering the discovery process. In addition, and despite of the impressive advances made in the field of microbial genomics, which revealed a by far underestimated biosynthetic capacity of microbes,¹⁶⁰ many microbial biosynthetic gene clusters appear to be inactive under standard cultivation conditions.¹⁶¹ These limitations have been addressed (amongst other approaches) by transferring the gene cluster of interest into a well-studied heterologous host that is, often in contrast to the native producer, amenable to further strain improvement by genetic engineering. Heterologous expression of biosynthetic pathways has been successfully used to modify the natural products produced by the encoded pathways, to essentially increase production titers and to unleash secondary metabolite production by activating silent gene clusters.¹⁶² However, heterologous production of natural products is by no means trivial as the chosen host strain needs to meet several key requirements, which critically influence the outcome of such experiments. The foreign biosynthetic gene cluster must be stably and functionally expressed in the surrogate host, which requires suitable genetic elements such as promoters and terminators, as well as vector systems that allow stable transfer and maintenance.^{162,163} Regarding PKS and NRPS pathways, the heterologous host needs to bear the genetic capacity for posttranslational activation of the encoded megasynthetases, e.g. PPTases,¹⁰² and for the supply of simple precursor molecules, which is particularly important if these precursors derive from building blocks of the primary metabolism and are produced *in situ*. Otherwise, the host strain has to be equipped with essential genes via genetic engineering.¹⁶⁴ In addition, toxic effects of the secondary metabolites produced need to be avoided to protect the heterologous host, which can be realized via expression of additional genes that confer self-resistance.¹⁶⁵ Another factor that seems to have a great impact on the efficient heterologous expression of biosynthetic

gene clusters is the codon usage of the host strain. PKS and NRPS genes found in myxobacterial genomes usually exhibit a high GC content, which requires an appropriate codon usage bias of the selected surrogate host.¹⁶⁶ Taking these requirements into account, the lateral transfer of a target biosynthetic gene cluster into a closely related host strain developed for heterologous secondary metabolite production might be obviously the most convenient way to establish heterologous production platforms.

Once a biosynthetic gene cluster of interest has been selected for heterologous expression, the respective DNA needs to be captured from the native producer strain. Conventionally, the target genes are mobilized from the genomic DNA of the producer strain via library preparation using cosmids, fosmids or bacterial artificial chromosomes. As the screening for correct constructs harboring the target gene cluster or parts thereof is very laborious, recent developments focused on direct cloning methods, which allow recombination-based capture of the target gene cluster and provide possibilities for subsequent downstream engineering. The most widely used cloning techniques encompass the transformation associated recombination (TAR) protocol in yeast,^{167,168} as well as Red/ET recombineering.^{169,170} The latter technology is directly applicable to further vector backbone modifications, e.g. to insert suitable promoters for controlled gene expression, selection markers or transfer cassettes as well as to modify the biosynthesis genes. The large recombinant DNA constructs are subsequently transferred into the host strain, e.g. via electroporation or conjugation, and are either be integrated into the host chromosome via homologous recombination, phage attachment sites or transposon-based approaches, or can be expressed as replicative plasmids.¹⁶⁵ However, heterologous expression of biosynthetic gene clusters often lead to reduced or even abolished secondary metabolite production, since the manifold demands that need to be met are not sufficiently tackled as many factors associated with efficient expression in the host cell remain unknown.

Alternatively, target biosynthetic gene clusters can now be manufactured via DNA synthesis, which provides custom-specific DNA sequences within steadily shorter delivery times and at decreasing costs without being dependent on physically existing DNA.¹⁷¹ The sole input required for *de novo* gene synthesis is a DNA sequence acquired from DNA sequencing, which can be almost arbitrarily redesigned according to specific constructional and functional sequence requirements. The first totally synthesized natural product pathway, that was successfully assembled via conventional cloning and subsequently expressed, was the erythromycin biosynthetic gene cluster reported by Kodumal *et al.*¹⁷² As the synthesis of large genes or even entire gene clusters is still rather challenging, DNA synthesis is often combined

with state-of-the-art DNA assembly strategies that enable construction of the entire gene cluster construct from pre-defined DNA fragments. Most prominent among the recent developments in high-throughput DNA assembly technologies are the recombination-based technologies such as TAR, Gibson assembly and sequence and ligation independent cloning (SLIC).^{173–175} In addition, several restriction/ligation-based methods have been developed such as the Golden Gate cloning method and the ligase cycling reaction (LCR).^{176,177}

The high-throughput synthesis of arbitrary DNA sequences opens up versatile possibilities in the field of ‘Synthetic Biology’ enabling e.g. the translation of biological systems into interchangeable modular building blocks in a ‘bio brick’ manner based on engineering principles for the redesign of metabolic pathways and the creation of designer microorganisms that serve as tailor-made production factories. One approach in natural product research is to reduce the complexity of a natural system, e.g. a given producer organism, and to build up a ‘minimal’ chassis, in which non-essential parts are completely removed and physiological bottlenecks are eliminated. The gene cluster sequence to be heterologously expressed can be modulated towards optimized codon usage in the host strain and non-essential genes and regulatory elements are being removed to bring the artificial system under the control of synthetic regulators that govern transcription and translation.¹⁷⁸ In this context, constructional sequence requirements notably cover the necessary changes to be introduced to allow for efficient construction of the synthetic gene cluster from smaller DNA segments and for downstream engineering. To be not reliant on naturally occurring restriction sites in the native gene cluster sequence, restriction sites engineering can be performed by applying silent mutations to remove disturbing restriction sites or to introduce restriction sites used for pathway assembly and modification, thereby conserving the native protein sequence. Additionally, homology arms can be introduced into the synthetic gene cluster to facilitate integration into the host’s chromosome via homologous recombination. Functional sequence requirements encompass selection of suitable regulatory elements and vector systems, which are known to be functional in the chosen heterologous host, as well as sequence optimization regarding GC content and codon usage, respectively. This approach is referred to as ‘refactoring’ and holds great potential for the future engineering of artificial production platforms. Refactoring of natural product pathways has already been applied to several biosynthetic gene clusters to establish production platforms for combinatorial biosynthesis, to activate orphan gene clusters and to study the effects of codon adaption on the productivity of the underlying pathway. For instance, the erythromycin gene cluster was one of the first pathways, which was shown to be redesigned including restriction sites engineering to

facilitate exchange of domains or modules as well as codon optimization, which essentially boosted production levels in *E. coli*.¹⁷⁹ The spectinabilin biosynthetic gene cluster was refactored via exchange of every native promoter for well-known promoters, which are functional in the heterologous host *Streptomyces lividans*. Decoupling of pathway expression from the native regulation machinery resulted in detectable amounts of spectinabilin, which was not identified in the native producer before.¹⁸⁰ The same research group recently managed to activate another biosynthetic gene cluster in *S. lividans* using a highly similar strategy, thereby resulting in the production of novel polycyclic tetramate macrolactams. In addition, the synthetic expression platform was also used to characterize the underlying biosynthetic pathway via a set of gene deletions.¹⁸¹ The first artificial PKS/NRPS hybrid pathway from a myxobacterium was established initially in *E. coli* and later on in *M. xanthus*. Oßwald *et al.* reconstructed the large 58 kb epothilone pathway from *S. cellulosum* using a flexible modular assembly strategy and analyzed the codon optimized gene cluster variant in *M. xanthus* using a multiplasmid approach.¹⁸² However, production titers were found to be very low, which underpins that codon adaption alone does not necessarily lead to improved secondary metabolite production, if the complete regulatory networks in the heterologous host are not fully understood or if there are additional limiting factors.

In conclusion, the tool box that is currently available allows the synthetic biologist to redesign, construct and optimize a gene cluster of interest in multifaceted ways and in a high-throughput manner. DNA synthesis and assembly of redesigned sequences enables experimental set-ups, in which many different sequence designs can be analyzed in parallel. Dissection of the biosynthetic pathways into modular parts and the generation of synthetic libraries consisting of standardized building blocks for swapping genetic elements and the fine-tuned control of gene expression in appropriate heterologous hosts will essentially accelerate the development of microbial cell factories for manufacturing natural products.

1.5 Myxochromide Pathways as Model Systems – Current State of Research and Outline of the Presented Work

Synthetic biology has a great potential to overcome the described limitations, which are associated with conventional heterologous expression platforms based on native biosynthetic gene clusters. Due to the impressive development of chemical DNA synthesis, it is possible to design and to synthesize optimized gene fragments, which need to be subsequently assembled to the entire biosynthetic gene cluster using state-of-the-art cloning techniques. The native DNA sequences are adapted to the constructional and functional requirements of the

biosynthetic pathway and particularly of the host organism that was chosen for heterologous secondary metabolite production.¹⁸² However, the optimal sequence design of genes as well as the use of not well-established host organisms, such as myxobacteria, still remains a challenge. Therefore, the BMBF-funded ‘SynBioDesign’ project (Synthetic Biology to Design Production Platforms for Complex Natural Products) aimed at a comprehensive understanding of all factors, which are relevant for natural product formation in a heterologous host starting from artificial DNA over the levels of transcription and translation to the point of natural product biosynthesis from simple precursor molecules to systematically optimize artificial production platforms. By combining advanced high-throughput analytical techniques covering transcriptomics and proteomics approaches, it might be possible to illuminate unknown but often limiting factors regarding secondary metabolite production such as the stability of messenger RNA (mRNA), the available pools of aminoacyl-transfer RNAs (aminoacyl-tRNA), and proper co-translational folding of the expressed megasynthetases. Detailed knowledge on these basic cellular features would allow to directly link the gained experimental qualitative and quantitative data to the sequence design, thereby facilitating the optimization process during multiple rounds of adapted sequence design, assembly and heterologous expression of modified gene cluster versions, acquisition of analytical data and eventually data interpretation. In conjunction with modern analytical equipment, this approach would also require a fast and efficient DNA assembly strategy for the construction of many different synthetic gene clusters, which can subsequently be tested in parallel in a high-throughput manner and might lead to significant advances towards yield improvement of synthetic production platforms. In addition, the resulting generic production platforms can also be used for the rational engineering of the underlying pathways to generate chemical diversity. To address these aims, the choice of a suitable biosynthetic pathway that serves as a model system as well as an already well-studied myxobacterial heterologous host is essential. Due to several reasons discussed below, the biosynthesis of the myxobacterial lipopeptides myxochromides appeared to be the ideal model systems in the ‘SynBioDesign’ project to establish synthetic DNA platforms for the fundamental understanding of pathway design and optimization. At the beginning of the project, three myxochromides have been described in the literature, which differ in the number, order, composition and configuration of the amino acids in the peptide core and in the acyl side chains, which can vary in length: the lipohexapeptides myxochromides A₂₋₄ from *Myxococcus xanthus*,⁸⁴ the lipopentapeptides myxochromides S₁₋₃ from *Stigmatella aurantiaca*,⁸⁵ and the lipopeptapeptide myxochromide B₃ from an unclassified *Myxococcus* sp. (Figure 7C).⁸⁶

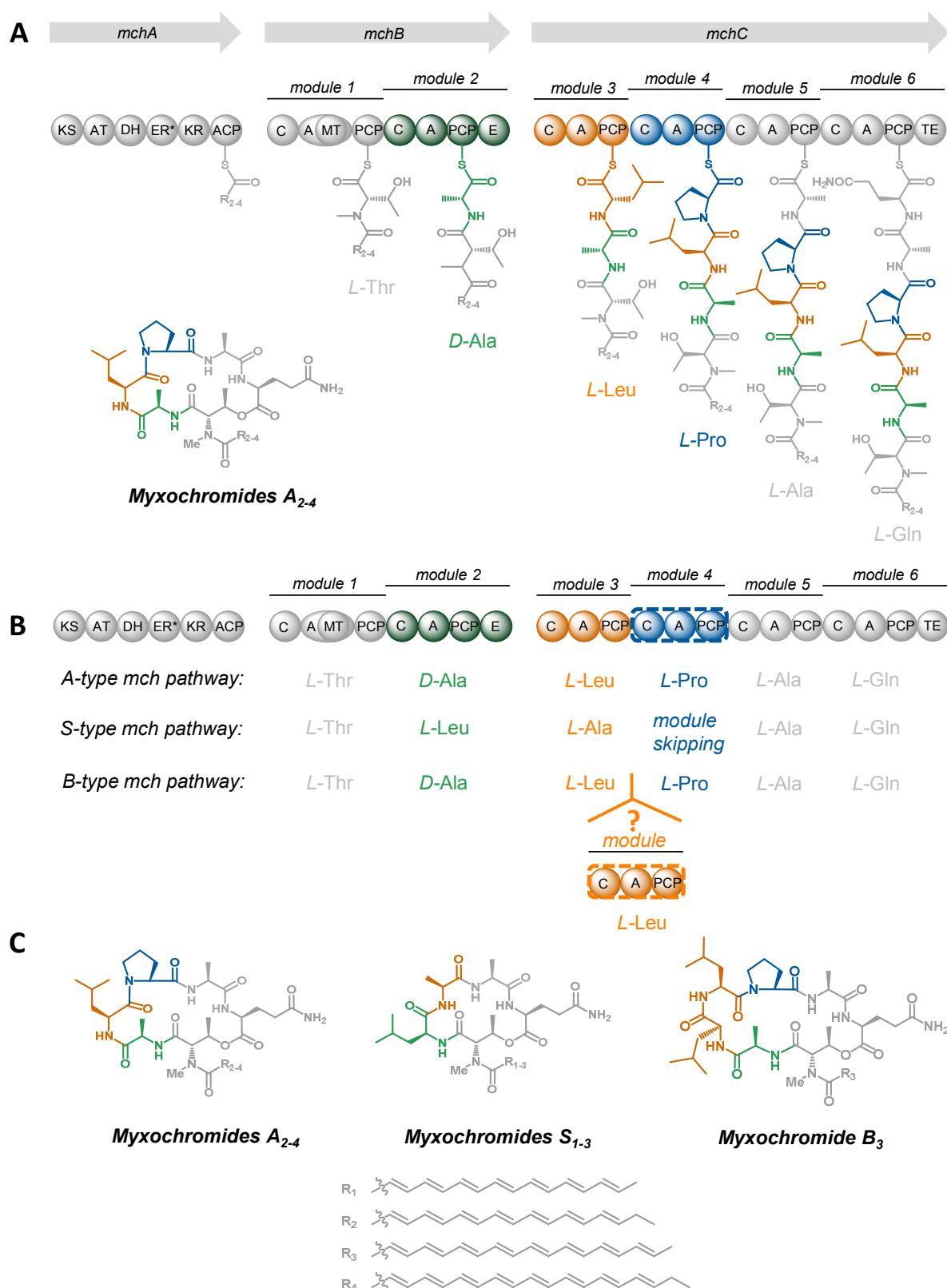


Figure 7. Overview on different myxochromide pathways. The coloring of the catalytic domains corresponds to the biosynthesis products shown in Figure 7C. The PKS/NRPS hybrid machineries (A- and S-type) consist of an iterative PKS module and six NRPSs modules and facilitates myxochromide biosynthesis. Abbreviations of catalytic domains are explained in the main text. **A:** Organization of the myxochromide A biosynthetic gene cluster and model for the biosynthesis of myxochromides A. **B:** Comparison of the A-, S- and the proposed B-type myxochromide pathways. **C:** Structures of different myxochromides.

The biosynthetic machineries of the myxochromide A and S pathways are encoded in relatively small biosynthetic gene clusters (~30 kb) comprising one single operon that consists of three biosynthetic genes *mchA-C*. These genes contain the blueprint of ~1MDa PKS/NRPS hybrid megasynthetases, which combine the biosynthesis of the polyene side chains with the formation of the peptide core. The *mchA* gene encodes for an iterative ‘unimodular’ PKS, which is mainly found in eukaryotic fungi, whereas the *mchB* and *mchC* genes are translated into a hexamodular NRPS system.^{84,85} In myxochromide A biosynthesis, the PKS subunit that contains a full set of reductive domains (KR, ER, DH) starts the biosynthesis with the formation of the lipid chains, which have been found to be fully unsaturated. It is assumed that the ER domain, which might be responsible for the reduction of the enoyl moiety to the saturated lipids is inactive.⁸⁵ Various lengths of the polyene side chain have been detected depending on the number of iterative elongation cycles and the selected starter unit (acetate or propionate). The lipid chains are transferred to the first NRPS subunit (MchB), where *N*-acylation of the first amino acid *L*-threonine is catalyzed by the initial C domain in module 1. The MT domain integrated into the downstream A domain is responsible for *N*-methylation of *L*-threonine and the first biosynthetic intermediate is subsequently translocated to the second module by the PCP. Module 2 also contains a ‘non-standard’ domain (E domain) and catalyzes elongation with the second amino acid (*L*-alanine) along with epimerization to form a racemic mixture of the dipeptide intermediate (*D/L*-Ala-*N*-Me-*L*-Thr), whereas only the *D*-configured intermediate is further processed by the second NRPS subunit (MchC). MchC contains four standard ‘C-A-PCP’ modules responsible for selection, activation and incorporation of the amino acids *L*-leucine, *L*-proline, *L*-alanine and *L*-glutamine. The additional TE domain of the termination module catalyzes cyclization between the hydroxyl group of the *N*-Me-*L*-Thr side chain and the carboxylic acid of *L*-Gln of the linear hexapeptide intermediate to form the cyclic lipohexapeptide products (Figure 7A). Although both myxochromide A and S assembly lines appear to be organized in an identical arrangement of catalytic domains (Figure 7B), studies on the biosyntheses of both types of myxochromides revealed striking differences between the underlying pathways leading to structural diversity.⁸⁴ Most intriguing is an unique ‘module-skipping’ process in the myxochromide S pathway, which results in skipping of module 4 in the MchC NRPS subunit, and which has been known only for PKS systems before. Consequently, the *L*-Pro residue that might be activated by this module, is missing in the final lipopentapeptide products myxochromides S, thereby violating the co-linearity rule. It has been shown that a Ser → Pro loss-of-function mutation of the conserved serine residue in the PCP of module 4 might be

responsible for ‘module-skipping’, as the PCP domain cannot be posttranslationally primed with the CoA-derived phosphopantetheinyl arm. In addition, the A domain of this module has been shown to be still active in *in vitro* studies with expressed A domains.⁸⁴ Another difference is the observed switch in substrate specificities of modules 2 and 3, which was previously explained by point mutations in the corresponding A domains of the S-type myxochromide pathway, thereby probably changing the binding pockets. Interestingly, the *L*-Leu residue that is incorporated by module 2 in the myxochromide S assembly line is not epimerized, although the E domain is present in this module as well. Protein sequence alignments did not reveal any mutations in critical positions of the E domain. It is still not known if the E domain may be inactive or if the more bulky *L*-Leu residue is not properly used as a substrate by the E domain in the myxochromide S pathway.^{84,85} In addition to myxochromides A and S, a third myxochromide type from a *Myxococcus* sp. was isolated and structurally characterized by Ohlendorf *et al.*⁸⁶ Myxochromide B₃ has been found to be the only representative of this new family and no further derivatives exhibiting different acyl chain lengths have been detected. Myxochromide B₃ is a lipopeptide and shares the same peptide core as observed for myxochromides A with an additional leucine residue located adjacent to the leucine residue that is incorporated by module 3 in myxochromide A biosynthesis. Since the corresponding B-type pathway was not identified and further characterized, it was not clear if a module duplication event or the iterative use of only one Leu-incorporating module is responsible for this observation (Figure 7B).⁸⁶ Some isolated myxochromide derivatives have been tested for their antimicrobial and cytotoxic activities, but have been shown to exhibit no interesting pharmacological properties.^{84–86}

Choosing the myxochromide pathways as model systems for the ‘SynBioDesign’ project has several significant advantages: 1) The myxochromide biosynthetic gene clusters are relatively small and encode both PKS and NRPS subunits, 2) heterologous expression of native myxochromide gene clusters has already been successfully demonstrated in various host organisms such as myxobacteria, pseudomonads and burkholderia strains in high yields up to 500 mg/L,^{183–186} 3) the myxochromide pathways do not require any unusual precursor molecules, 4) myxochromides did not show any toxicity to the heterologous hosts tested and 5) different myxochromide families are known and widely distributed among myxobacteria, which provides promising opportunities for combinatorial approaches to generate hybrid natural products. The choice of a suitable heterologous host is as much as important as the selection of a model pathway to be designed. In the ‘SynBioDesign’ project, *M. xanthus* is used as surrogate host for several reasons. *M. xanthus* is one of the best-studied myxobacteria

and is amenable to genetic manipulation. Several genetic and regulatory elements have been characterized and complement the tool box for host strain engineering. Furthermore, *M. xanthus* exhibits short doubling times (~ 5 hours) compared to other myxobacteria, bears the biosynthetic capacities to produce polyketides as well as nonribosomal peptides and is able to posttranslationally activate the corresponding megasynthetases by using two broad spectrum PPTases, which were previously identified.¹⁸⁷ Additionally, *M. xanthus* has already been demonstrated to be a suitable host for heterologous production of natural products, including myxochromides S from *S. aurantiaca* in high production yields.¹⁸⁶ Since all myxochromide gene clusters originate from myxobacterial producers, it is very likely that foreign myxochromide pathways are efficiently expressed in the related *M. xanthus* host due to their similar codon usage.

By combination of synthetic genes of different myxochromide pathways or of differentially optimized gene versions, it might be possible to identify gene cluster features/regions, which have a positive or negative influence on myxochromide biosynthesis and thus on the production yields in case the optimization of the culture conditions ensures that the gene design is the limiting factor. High production yields have already been achieved with the native gene clusters, so that the multifactorial optimization of precursor supply is not further necessary to hold the system manageable. Following, structural genes can be directly modified. In conclusion, acquisition, analysis and implementation of sequence data of different myxochromide pathways and subsequent establishment of synthetic DNA platforms for the production of myxochromides in *M. xanthus* in this context might build the basis for the deduction of conceptional guidelines for designing complex natural product pathways and for the future creation of predictable microbial cell factories.

1.5.1 Outline I: Genome-Mining to Find New Lipopeptide Scaffolds in Myxobacteria

In the course of the ‘SynBioDesign’ project, the initial aim of this work was to exploit myxochromide diversity in myxobacteria in order to achieve a comprehensive understanding of the PKS/NRPS-mediated myxochromide biosynthesis as a broad basis for the sequence design and the creation of flexible synthetic DNA platforms. This part is covered by chapter 2 of this thesis. Sequence data acquisition should be performed by genome sequencing of myxobacterial strains as well as screening of available genome data followed by detailed *in silico* sequence analysis to identify new putative myxochromide biosynthetic gene clusters. By applying the genome-mining approach, identified myxobacterial producer strains should be cultivated and the corresponding myxochromide families, which were predicted, should be

detected, isolated and structurally characterized. This approach was expected to provide first insights into the biosynthesis of the previously described myxochromides B, as the corresponding biosynthetic pathway has not been deciphered yet.⁸⁶ In addition, genome-mining was intended to be used to directly link structural differences of different myxochromide families to the acquired genetic information, complemented by in-depth computational analysis of the evolutionary relationships of different myxochromide biosynthetic gene clusters. Furthermore, selected representatives of different myxochromide families should be analyzed for their natural functions as myxochromides do not exhibit any pharmacological activities in the available test panels so far. These studies are expected to provide an ideal starting point towards the development of synthetic DNA platforms and the rational engineering of the model pathways to generate novel lipopeptides.

1.5.2 Outline II: Synthetic Biotechnology to Produce Novel Hybrid Myxochromides

Based on the valuable insights gained from the genome-mining approach in conjunction with evolutionary aspects of myxobacterial pathway diversification, chapter 3 of this thesis addresses the constructional and functional redesign of the identified myxochromide pathways. Specific sequence requirements should be predefined in order to allow assembly of synthetic expression constructs harboring artificial myxochromide gene clusters and their heterologous expression in the selected host strain *M. xanthus*. In the course of the ‘SynBioDesign’ project, the establishment of a fast and efficient strategy for the assembly of large synthetic gene cluster constructs from synthetic building blocks was a major goal to generate various artificial gene cluster variants in parallel, which is a key prerequisite for the pursued high-throughput workflow. Simultaneously, the assembly strategy should be highly flexible to facilitate engineering efforts to alter myxochromide structures and to generate entirely new lipopeptide scaffolds. Chapter 3 covers the development of such an assembly strategy based on a dedicated gene library that contains synthetic myxochromide biosynthesis genes originating from different myxochromide pathways. The cloning approach should involve the generation of the gene library by using modern restriction/ligation-based techniques followed by the *in vitro* reconstitution of the entire pathways based on the Golden Gate cloning method,¹⁷⁷ both covered by the constructional sequence design. The described strategy should also exhibit broad applicability regarding possible combinations with different cloning techniques and the possibility for specific exchanges on the subunit, module and domain level. Exemplarily, the established assembly strategy should be demonstrated in a structural engineering approach to produce ‘unnatural’ myxochromides. The effects of

artificial gene combinations on the functionality and productivity of the engineered PKS/NRPS assembly lines should be analyzed, which would provide first insights into how complex natural product pathways should be designed, thereby serving as versatile platforms for the ‘SynBioDesign’ project.

1.6 References

- (1) *Nat. Chem. Biol.* **2007**, *3*, 351.
- (2) Cragg, G. M.; Newman, D. J. *Biochim. Biophys. Acta* **2013**, *1830*, 3670–3695.
- (3) Fox, E. M.; Howlett, B. J. *Curr. Opin. Microbiol.* **2008**, *11*, 481–487.
- (4) O'Brien, J.; Wright, G. D. *Curr. Opin. Biotechnol.* **2011**, *22*, 552–558.
- (5) Demain, A. L.; Fang, A. *Adv. Biochem. Eng. Biotechnol.* **2000**, *69*, 1–39.
- (6) Raaijmakers, J. M.; Bruijn, I. de; Nybroe, O.; Ongena, M. *FEMS Microbiol. Rev.* **2010**, *34*, 1037–1062.
- (7) Bonmatin, J.-M.; Laprevote, O.; Peypoux, F. *Comb. Chem. High Throughput Screen.* **2003**, *6*, 541–556.
- (8) Strieker, M.; Marahiel, M. A. *ChemBioChem* **2009**, *10*, 607–616.
- (9) Stein, T. *Mol. Microbiol.* **2005**, *56*, 845–857.
- (10) Roongsawang, N.; Washio, K.; Morikawa, M. *Int. J. Mol. Sci.* **2010**, *12*, 141–172.
- (11) Baltz, R. H.; Miao, V.; Wrigley, S. K. *Nat. Prod. Rep.* **2005**, *22*, 717–741.
- (12) Aretz, W.; Meiwes, J.; Seibert, G.; Vobis, G.; Wink, J. J. *Antibiot.* **2000**, *53*, 807–815.
- (13) Singh, S.; Kate, B. N.; Banerjee, U. C. *Crit. Rev. Biotechnol.* **2005**, *25*, 73–95.
- (14) Rufino, R. D.; Luna, J. M. de; de Campos Takaki, Galba Maria; Sarubbo, L. A. *Electron. J. Biotechnol.* **2014**, *17*, 34–38.
- (15) Emri, T.; Majoros, L.; Toth, V.; Pocs, I. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 3267–3284.
- (16) Hamley, I. W. *Chem. Commun. (Cambridge, U. K.)* **2015**, *51*, 8574–8583.
- (17) Cochrane, S. A.; Vederas, J. C. *Med. Res. Rev.* **2014**, *36*, 4–31.
- (18) Miao, V.; Coeffet-LeGal, M. F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M. *et al. Microbiology* **2005**, *151*, 1507–1523.
- (19) Debono, M.; Abbott, B. J.; Molloy, R. M.; Fukuda, D. S.; Hunt, A. H.; Daupert, V. M.; Counter, F. T.; Ott, J. L.; Carrell, C. B.; Howard, L. C. *J. Antibiot.* **1988**, *41*, 1093–1105.
- (20) Arbeit, R. D.; Maki, D.; Tally, F. P.; Campanaro, E.; Eisenstein, B. I. *Clin. Infect. Dis.* **2004**, *38*, 1673–1681.
- (21) Landman, D.; Georgescu, C.; Martin, D. A.; Quale, J. *Clin. Microbiol. Rev.* **2008**, *21*, 449–465.
- (22) Groupe, V.; Pugh, L. H.; Weiss, D.; Kochi, M. *Proc. Soc. Exp. Biol. Med.* **1951**, *78*, 354–358.
- (23) Debono, M.; Gordee, R. S. *Annu. Rev. Microbiol.* **1994**, *48*, 471–497.
- (24) Ongena, M.; Jacques, P. *Trends Microbiol.* **2008**, *16*, 115–125.
- (25) Singh, P.; Cameotra, S. S. *Trends Biotechnol.* **2004**, *22*, 142–146.
- (26) Denning, D. W. *Lancet (London, England)* **2003**, *362*, 1142–1151.
- (27) Vollenbroich, D.; Ozel, M.; Vater, J.; Kamp, R. M.; Pauli, G. *Biologicals* **1997**, *25*, 289–297.
- (28) Naruse, N.; Tenmyo, O.; Kobaru, S.; Kamei, H.; Miyaki, T.; Konishi, M.; Oki, T. *J. Antibiot.* **1990**, *43*, 267–280.
- (29) Kim, S.-Y.; Kim, J. Y.; Kim, S.-H.; Bae, H. J.; Yi, H.; Yoon, S. H.; Koo, B. S.; Kwon, M.; Cho, J. Y.; Lee, C.-E. *et al. FEBS Lett.* **2007**, *581*, 865–871.
- (30) Yu, Z.; Lang, G.; Kajahn, I.; Schmaljohann, R.; Imhoff, J. F. *J. Nat. Prod.* **2008**, *71*, 1052–1054.
- (31) Borel, J. F.; Feurer, C.; Gubler, H. U.; Stahelin, H. *Agents Actions* **1976**, *6*, 468–475.
- (32) Cardoso, L. S.; Araujo, M. I.; Goes, A. M.; Pacifico, L. G.; Oliveira, R. R.; Oliveira, S. C. *Microb. Cell Fact.* **2007**, *6*, 1.
- (33) Latoud, C.; Peypoux, F.; Michel, G.; Genet, R.; Morgat, J. L. *Biochim. Biophys. Acta* **1986**, *856*, 526–535.
- (34) Quentin, M. J.; Besson, F.; Peypoux, F.; Michel, G. *Biochim. Biophys. Acta* **1982**, *684*, 207–211.
- (35) Mnif, I.; Ghribi, D. *Biopolymers* **2015**, *104*, 129–147.
- (36) Pacwa-Plociniczak, M.; Plaza, G. A.; Piotrowska-Seget, Z.; Cameotra, S. S. *Int. J. Mol. Sci.* **2011**, *12*, 633–654.
- (37) Robinson, L. R.; Fitzgerald, N. C.; Doughty, D. G.; Dawes, N. C.; Berge, C. A.; Bissett, D. L. *Int. J. Cosmet. Sci.* **2005**, *27*, 155–160.
- (38) Lintner, K.; Peschard, O. *Int. J. Cosmet. Sci.* **2000**, *22*, 207–218.
- (39) Schneider, T.; Gries, K.; Josten, M.; Wiedemann, I.; Pelzer, S.; Labischinski, H.; Sahl, H.-G. *Antimicrob. Agents Chemother.* **2009**, *53*, 1610–1618.
- (40) Lakey, J. H.; Maget-Dana, R.; Ptak, M. *Biochim. Biophys. Acta* **1989**, *985*, 60–66.
- (41) Lakey, J. H.; Ptak, M. *Biochemistry* **1988**, *27*, 4639–4645.
- (42) Bérdy, J. *J. Antibiot. (Tokyo)* **2012**, *65*, 441.

-
- (43) Dawid, W. *FEMS Microbiol. Rev.* **2000**, *24*, 403–427.
- (44) Han, K.; Li, Z. F.; Peng, R.; Zhu, L. P.; Zhou, T.; Wang, L. G.; Li, S. G.; Zhang, X. B.; Hu, W.; Wu, Z. H. *et al. Sci. Rep.* **2013**, *3*, 1–7.
- (45) Schneiker, S.; Perlova, O.; Kaiser, O.; Gerth, K.; Alici, A.; Altmeyer, M. O.; Bartels, D.; Bekel, T.; Beyer, S.; Bode, E. *et al. Nat. Biotechnol.* **2007**, *25*, 1281–1289.
- (46) Reichenbach, H. *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 149–156.
- (47) Herrmann, J.; Fayad, A. A.; Müller, R. *Nat. Prod. Rep.* **2017**, *34*, 135–160.
- (48) Munoz-Dorado, J.; Marcos-Torres, F. J.; Garcia-Bravo, E.; Moraleda-Munoz, A.; Perez, J. *Front. Microbiol.* **2016**, *7*, 781.
- (49) Rosenberg, E., Ed., *Myxobacteria: Development and Cell Interactions*, Springer, New York, **1984**, 91–107.
- (50) Mauriello, E. M. F.; Mignot, T.; Yang, Z.; Zusman, D. R. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 229–249.
- (51) Nan, B.; Zusman, D. R. *Annu. Rev. Genet.* **2011**, *45*, 21–39.
- (52) Perez, J.; Jimenez-Zurdo, J. I.; Martinez-Abarca, F.; Millan, V.; Shimkets, L. J.; Munoz-Dorado, J. *Environ. Microbiol.* **2014**, *16*, 2341–2350.
- (53) Shimkets, L. J. *J. Bacteriol.* **1986**, *166*, 842–848.
- (54) Shimkets, L. J. *J. Bacteriol.* **1986**, *166*, 837–841.
- (55) Kearns, D. B.; Shimkets, L. J. *Trends Microbiol.* **2001**, *9*, 126–129.
- (56) Bowden, M. G.; Kaplan, H. B. *Mol. Microbiol.* **1998**, *30*, 275–284.
- (57) Chang, Y.-W.; Rettberg, L. A.; Treuner-Lange, A.; Iwasa, J.; Sogaard-Andersen, L.; Jensen, G. J. *Science* **2016**, *351*, aad2001.
- (58) Islam, S. T.; Mignot, T. *Semin. Cell Dev. Biol.* **2015**, *46*, 143–154.
- (59) Nan, B.; McBride, M. J.; Chen, J.; Zusman, D. R.; Oster, G. *Curr. Biol.* **2014**, *24*, R169–73.
- (60) Nan, B.; Bandaria, J. N.; Moghtaderi, A.; Sun, I.-H.; Yildiz, A.; Zusman, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E1508–13.
- (61) Ducret, A.; Fleuchot, B.; Bergam, P.; Mignot, T. *eLife* **2013**, *2*, e00868.
- (62) Morgan, A. D.; MacLean, R. C.; Hillesland, K. L.; Velicer, G. J. *Appl. Environ. Microbiol.* **2010**, *76*, 6920–6927.
- (63) Müller, S.; Strack, S. N.; Hoefler, B. C.; Straight, P. D.; Kearns, D. B.; Kirby, J. R. *Appl. Environ. Microbiol.* **2014**, *80*, 5603–5610.
- (64) Perez, J.; Moraleda-Munoz, A.; Marcos-Torres, F. J.; Munoz-Dorado, J. *Environ. Microbiol.* **2016**, *18*, 766–779.
- (65) Sliusarenko, O.; Zusman, D. R.; Oster, G. *J. Bacteriol.* **2007**, *189*, 611–619.
- (66) Whitworth, D. E., Ed., *Myxobacteria*, American Society of Microbiology, Washington, **2008**, 43–76.
- (67) O'Connor, K. A.; Zusman, D. R. *J. Bacteriol.* **1991**, *173*, 3318–3333.
- (68) Nariya, H.; Inouye, M. *Cell* **2008**, *132*, 55–66.
- (69) Volz, C.; Kegler, C.; Müller, R. *Chem. Biol.* **2012**, *19*, 1447–1459.
- (70) Wittmann, C.; Liao, J. C., Eds., *Industrial Biotechnology: Microorganisms*; WILEY-VCH, Weinheim, **2017**, 453–485.
- (71) Brenner, D.; Krieg, N.; Staley, J., Eds., *Bergey's Manual of Systematic Bacteriology*, vol. 2, Springer, New York, **2005**, 1059–1144.
- (72) Garcia, R.; Pistorius, D.; Stadler, M.; Müller, R. *J. Bacteriol.* **2011**, *193*, 1930–1942.
- (73) Stackebrandt, E.; Pauker, O.; Steiner, U.; Schumann, P.; Straubler, B.; Heibei, S.; Lang, E. *Syst. Appl. Microbiol.* **2007**, *30*, 109–118.
- (74) Lang, E.; Sproer, C. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 2991–2992.
- (75) Weissman, K. J.; Müller, R. *Nat. Prod. Rep.* **2010**, *27*, 1276–1295.
- (76) Goldman, B. S.; Nierman, W. C.; Kaiser, D.; Slater, S. C.; Durkin, A. S.; Eisen, J.; Ronning, C. M.; Barbazuk, W. B.; Blanchard, M.; Field, C. *et al. Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15200–15205.
- (77) Gerth, K.; Pradella, S.; Perlova, O.; Beyer, S.; Müller, R. *J. Biotechnol.* **2003**, *106*, 233–253.
- (78) Cortina, N. S.; Krug, D.; Plaza, A.; Revermann, O.; Müller, R. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 811–816.
- (79) Simunovic, V.; Zapp, J.; Rachid, S.; Krug, D.; Meiser, P.; Müller, R. *ChemBioChem* **2006**, *7*, 1206–1220.
- (80) Bode, H. B.; Meiser, P.; Klefisch, T.; Cortina, N. S.; Krug, D.; Göhring, A.; Schwär, G.; Mahmud, T.; Elnakady, Y. A.; Müller, R. *ChemBioChem* **2007**, *8*, 2139–2144.

-
- (81) Meiser, P.; Weissman, K. J.; Bode, H. B.; Krug, D.; Dickschat, J. S.; Sandmann, A.; Müller, R. *Chem. Biol.* **2008**, *15*, 771–781.
- (82) Meiser, P.; Bode, H. B.; Müller, R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133.
- (83) Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. *Eur. J. Biochem.* **2000**, *267*, 6476–6485.
- (84) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
- (85) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. *ChemBioChem* **2005**, *6*, 375–385.
- (86) Ohlendorf, B.; Kehraus, S.; König, G. M. *J. Nat. Prod.* **2008**, *71*, 1708–1713.
- (87) Etzbach, L.; Plaza, A.; Garcia, R.; Baumann, S.; Müller, R. *Org. Lett.* **2014**, *16*, 2414–2417.
- (88) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. *Anal. Chem.* **2014**, *86*, 10780–10788.
- (89) Lechleider, R. J.; Kaminskis, E.; Jiang, X.; Aziz, R.; Bullock, J.; Kasliwal, R.; Harapanhalli, R.; Pope, S.; Sridhara, R.; Leighton, J. *et al. Clin. Cancer Res.* **2008**, *14*, 4378–4384.
- (90) Gerth, K.; Bedorf, N.; Hofle, G.; Irschik, H.; Reichenbach, H. *J. Antibiot. (Tokyo)* **1996**, *49*, 560–563.
- (91) Vogt, T. *Mol. Plant* **2010**, *3*, 2–20.
- (92) Gao, Y.; Honzatko, R. B.; Peters, R. J. *Nat. Prod. Rep.* **2012**, *29*, 1153–1175.
- (93) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275–287.
- (94) Hertweck, C. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688–4716.
- (95) Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. *Mol. Biol. Evol.* **2005**, *22*, 2027–2039.
- (96) Smith, S.; Tsai, S. C. *Nat. Prod. Rep.* **2007**, *24*, 1041–1072.
- (97) Beld, J.; Lee, D. J.; Burkart, M. D. *Mol. Biosyst.* **2015**, *11*, 38–59.
- (98) Arnstadt, K. I.; Schindlbeck, G.; Lynen, F. *Eur. J. Biochem.* **1975**, *55*, 561–571.
- (99) Schweizer, E.; Hofmann, J. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 501–517.
- (100) Byers, D. M.; Gong, H. *Biochem. Cell Biol.* **2007**, *85*, 649–662.
- (101) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585–1595.
- (102) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923–936.
- (103) Witkowski, A.; Joshi, A. K.; Smith, S. *Biochemistry* **2002**, *41*, 10877–10887.
- (104) Staunton, J.; Weissman, K. J. *Nat. Prod. Rep.* **2001**, *18*, 380–416.
- (105) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
- (106) Minowa, Y.; Araki, M.; Kanehisa, M. *J. Mol. Biol.* **2007**, *368*, 1500–1517.
- (107) Liou, G. F.; Lau, J.; Cane, D. E.; Khosla, C. *Biochemistry* **2003**, *42*, 200–207.
- (108) Moore, B. S.; Hertweck, C. *Nat. Prod. Rep.* **2002**, *19*, 70–99.
- (109) Chan, Y. A.; Boyne, M. T.; Podevels, A. M.; Klimowicz, A. K.; Handelsman, J.; Kelleher, N. L.; Thomas, M. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14349–14354.
- (110) Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D. *Gene* **2000**, *251*, 81–90.
- (111) Wenzel, S. C.; Williamson, R. M.; Grünanger, C.; Xu, J.; Gerth, K.; Martinez, R. A.; Moss, S. J.; Carroll, B. J.; Grond, S.; Unkefer, C. J. *et al. J. Am. Chem. Soc.* **2006**, *128*, 14325–14336.
- (112) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468–3496.
- (113) Olano, C.; Mendez, C.; Salas, J. A. *Nat. Prod. Rep.* **2010**, *27*, 571–616.
- (114) Moss, S. J.; Martin, C. J.; Wilkinson, B. *Nat. Prod. Rep.* **2004**, *21*, 575–593.
- (115) Thomas, I.; Martin, C. J.; Wilkinson, C. J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2002**, *9*, 781–787.
- (116) Kohli, R. M.; Walsh, C. T. *Chem. Commun.* **2003**, 297–307.
- (117) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.
- (118) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J. *et al. Nat. Prod. Rep.* **2013**, *30*, 108–160.
- (119) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651–2674.
- (120) Keating, T. A.; Ehmman, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T. *ChemBioChem* **2001**, *2*, 99–107.
- (121) Schmelz, S.; Naismith, J. H. *Curr. Opin. Struct. Biol.* **2009**, *19*, 666–671.
- (122) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. *J. Biol. Chem.* **1998**, *273*, 22773–22781.

-
- (123) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
- (124) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493–505.
- (125) Walsh, C. T.; O'Brien, R. V.; Khosla, C. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7098–7124.
- (126) Kevany, B. M.; Rasko, D. A.; Thomas, M. G. *Appl. Environ. Microbiol.* **2009**, *75*, 1144–1155.
- (127) Hubbard, B. K.; Thomas, M. G.; Walsh, C. T. *Chem. Biol.* **2000**, *7*, 931–942.
- (128) Neary, J. M.; Powell, A.; Gordon, L.; Milne, C.; Flett, F.; Wilkinson, B.; Smith, C. P.; Micklefield, J. *Microbiology (Reading, Engl.)* **2007**, *153*, 768–776.
- (129) Strieker, M.; Nolan, E. M.; Walsh, C. T.; Marahiel, M. A. *J. Am. Chem. Soc.* **2009**, *131*, 13523–13530.
- (130) Cryle, M. J. *Metallomics* **2011**, *3*, 323–326.
- (131) Makris, T. M.; Chakrabarti, M.; Münck, E.; Lipscomb, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15391–15396.
- (132) Pohle, S.; Appelt, C.; Roux, M.; Fiedler, H.-P.; Sussmuth, R. D. *J. Am. Chem. Soc.* **2011**, *133*, 6194–6205.
- (133) Puk, O.; Bischoff, D.; Kittel, C.; Pelzer, S.; Weist, S.; Stegmann, E.; Sussmuth, R. D.; Wohlleben, W. *J. Bacteriol.* **2004**, *186*, 6093–6100.
- (134) Linne, U.; Doekel, S.; Marahiel, M. A. *Biochemistry* **2001**, *40*, 15824–15834.
- (135) Stein, T.; Kluge, B.; Vater, J.; Franke, P.; Otto, A.; Wittmann-Liebold, B. *Biochemistry* **1995**, *34*, 4633–4642.
- (136) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. *Nature* **2005**, *436*, 1191–1194.
- (137) van Lanen, S. G.; Dorrestein, P. C.; Christenson, S. D.; Liu, W.; Ju, J.; Kelleher, N. L.; Shen, B. *J. Am. Chem. Soc.* **2005**, *127*, 11594–11595.
- (138) Walsh, C. T.; Chen, H. W.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L. S.; Marshall, C. G.; Miller, D. A.; Patel, H. M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 525–534.
- (139) Schneider, T. L.; Shen, B.; Walsh, C. T. *Biochemistry* **2003**, *42*, 9722–9730.
- (140) Walsh, C. T. *Nat. Prod. Rep.* **2016**, *33*, 127–135.
- (141) Du, L.; Lou, L. *Nat. Prod. Rep.* **2010**, *27*, 255–278.
- (142) Du, L.; Shen, B. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 215–228.
- (143) Sandmann, A.; Sasse, F.; Müller, R. *Chem. Biol.* **2004**, *11*, 1071–1079.
- (144) Weissman, K. J.; Müller, R. *ChemBioChem* **2008**, *9*, 826–848.
- (145) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15585–15590.
- (146) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723–731.
- (147) Staunton, J.; Caffrey, P.; Aparicio, J. F.; Roberts, G. A.; Bethell, S. S.; Leadlay, P. F. *Nat. Struct. Biol.* **1996**, *3*, 188–192.
- (148) Hillson, N. J.; Walsh, C. T. *Biochemistry* **2003**, *42*, 766–775.
- (149) Sieber, S. A.; Linne, U.; Hillson, N. J.; Roche, E.; Walsh, C. T.; Marahiel, M. A. *Chem. Biol.* **2002**, *9*, 997–1008.
- (150) Goodwin, S.; McPherson, J. D.; McCombie, W. R. *Nat. Rev. Genet.* **2016**, *17*, 333–351.
- (151) Ziemert, N.; Alanjary, M.; Weber, T. *Nat. Prod. Rep.* **2016**, *33*, 988–1005.
- (152) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Brucoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W. *et al. Nucleic Acids Res.* **2015**, *43*, W237–W243.
- (153) Ansari, M. Z.; Yadav, G.; Gokhale, R. S.; Mohanty, D. *Nucleic Acids Res.* **2004**, *32*, W405–W413.
- (154) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362–W367.
- (155) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–224.
- (156) Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C. *PLoS ONE* **2013**, *8*, e62136.
- (157) Kraas, F. I.; Helmetag, V.; Wittmann, M.; Strieker, M.; Marahiel, M. A. *Chem. Biol.* **2010**, *17*, 872–880.
- (158) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78–92.
- (159) Caffrey, P. *ChemBioChem* **2003**, *4*, 654–657.
- (160) Wenzel, S. C.; Müller, R. *Nat. Prod. Rep.* **2009**, *26*, 1385–1407.
- (161) Rutledge, P. J.; Challis, G. L. *Nat. Rev. Microbiol.* **2015**, *13*, 509–523.
- (162) Wenzel, S. C.; Müller, R. *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606.
- (163) Stevens, D. C.; Hari, T. P. A.; Boddy, C. N. *Nat. Prod. Rep.* **2013**, *30*, 1391–1411.

-
- (164) Gross, F.; Ring, M. W.; Perlova, O.; Fu, J.; Schneider, S.; Gerth, K.; Kuhlmann, S.; Stewart, A. F.; Zhang, Y.; Müller, R. *Chem. Biol.* **2006**, *13*, 1253–1264.
- (165) Ongley, S.; Bian, X.; Neilan, B. A.; Müller, R. *Nat. Prod. Rep.* **2013**, *30*, 1121–1138.
- (166) Gustafsson, C.; Govindarajan, S.; Minshull, J. *Trends Biotechnol.* **2004**, *22*, 346–353.
- (167) Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. *Biopolymers* **2010**, *93*, 833–844.
- (168) Noskov, V. N.; Kouprina, N.; Leem, S.-H.; Ouspenski, I.; Barrett, J. C.; Larionov, V. *BMC Genomics* **2003**, *4*, 16.
- (169) Zhang, Y.; Muylers, J. P. P.; Testa, G.; Stewart, A. F. *Nat. Biotechnol.* **2000**, *18*, 1314–1317.
- (170) Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F. *et al. Nat. Biotechnol.* **2012**, *30*, 440–446.
- (171) Kosuri, S.; Church, G. M. *Nat. Methods* **2014**, *11*, 499–507.
- (172) Kodumal, S. J.; Patel, K. G.; Reid, R.; Menzella, H. G.; Welch, M.; Santi, D. V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15573–15578.
- (173) Li, M. Z.; Elledge, S. J. *Nat. Methods* **2007**, *4*, 251–256.
- (174) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., III; Smith, H. O. *Nat. Methods* **2009**, *6*, 343–345.
- (175) Kouprina, N.; Larionov, V. *Nat. Protoc.* **2008**, *3*, 371–377.
- (176) Kok, S. de; Stanton, L. H.; Slaby, T.; Durot, M.; Holmes, V. F.; Patel, K. G.; Platt, D.; Shapland, E. B.; Serber, Z.; Dean, J. *et al. ACS Synth. Biol.* **2014**, *3*, 97–106.
- (177) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2008**, *3*, e3647.
- (178) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. *Nat. Rev. Microbiol.* **2016**, *14*, 135–149.
- (179) Menzella, H. G.; Reisinger, S. J.; Welch, M.; Kealey, J. T.; Kennedy, J.; Reid, R.; Tran, C. Q.; Santi, D. V. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 22–28.
- (180) Shao, Z.; Rao, G.; Li, C.; Abil, Z.; Luo, Y.; Zhao, H. *ACS Synth. Biol.* **2013**, *2*, 662–669.
- (181) Luo, Y.; Huang, H.; Liang, J.; Wang, M.; Lu, L.; Shao, Z.; Cobb, R. E.; Zhao, H. *Nat. Commun.* **2013**, *4*, 2894.
- (182) Oßwald, C.; Zipf, G.; Schmidt, G.; Maier, J.; Bernauer, H. S.; Müller, R.; Wenzel, S. C. *ACS Synth. Biol.* **2014**, *3*, 759–772.
- (183) Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, A. F.; Müller, R. *Chem. Biol.* **2005**, *12*, 349–356.
- (184) Stephan, S.; Heinzle, E.; Wenzel, S. C.; Krug, D.; Müller, R.; Wittmann, C. *Process Biochem.* **2006**, *41*, 2146–2152.
- (185) Perlova, O.; Gerth, K.; Kuhlmann, S.; Zhang, Y.; Müller, R. *Microb. Cell Fact.* **2009**, *8*, 1.
- (186) Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. *Nucleic Acids Res.* **2008**, *36*, e113.
- (187) Meiser, P.; Müller, R. *ChemBioChem* **2008**, *9*, 1549–1553.

Chapter 2

Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria

Published online on 27th January 2017:

Christian Burgard,¹ Nestor Zaburannyi,¹ Suvd Nadmid,¹ Josef Maier,² Holger Jenke-Kodama,³ Eva Luxenburger,¹ Hubert S. Bernauer,⁴ Silke C. Wenzel¹
ACS Chemical Biology, DOI: 10.1021/acscchembio.6b00953

Affiliations

¹ Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology at Saarland University, Saarland University Campus, Building E8.1, 66123 Saarbrücken, Germany

² ISTLS – Information Services to Life Sciences, Härlestraße 24/1, 78727 Oberndorf am Neckar/Boll, Germany

³ Microbiology and Biochemistry of Secondary Metabolites Unit, Okinawa Institute of Science and Technology (OIST), 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan

⁴ ATG Biosynthetics GmbH, Weberstraße 40, 79249 Merzhausen, Germany

Reproduced by permission of the American Chemical Society

Contributions to the Presented Work

A. Author's Contribution

The author conceived and performed most of the experiments described in this chapter and interpreted the resulting data. The author contributed to the decipherment of the myxochromide B gene cluster via construction and screening of a cosmid library of the myxobacterial strain *Myxococcus* sp. 171 followed by extensive subcloning and sequencing efforts. Identification and annotation of additional myxochromide gene clusters as well as prediction of two novel myxochromides by genome-mining was carried out by the author. Furthermore, the author conducted the cultivation of the identified producer strains as well as the feeding experiments described and prepared extracts for metabolome analysis. Large-scale cultivation of selected producer strains for subsequent isolation and structure elucidation of four novel myxochromides via NMR spectroscopy and analysis of the absolute configuration was performed by the author. The author also contributed to the detailed *in silico* sequence analysis of selected catalytic domains. Swarm expansion assays described in the manuscript were carried out by the author as well. The author contributed to the conception and writing of the manuscript.

B. Contribution by Co-Workers

Nestor Zaburannyi assembled the genome sequence of the strain *Myxococcus* sp. 171 from Illumina sequencing data, analyzed the myxochromide gene cluster loci in their genomic context and contributed to the annotation and *in silico* analysis of the identified gene clusters and to the identification of the recombination sites described. Suvd Nadmid contributed to the analysis of NMR data. Josef Maier performed advanced CAI analysis and contributed to the annotation of myxochromide gene clusters. Holger Jenke-Kodama conducted phylogenetic analysis of myxochromide gene clusters and the respective producer strains. Eva Luxenburger performed HPLC-MS measurements of crude extracts and analyzed the resulting data. Hubert S. Bernauer and Silke C. Wenzel were responsible for the conception of the project. Furthermore, Silke C. Wenzel was involved in the *in silico* analysis of myxochromide gene clusters and analyzed the recombination sites. All authors contributed to the conception, writing and review of the manuscript.

2 Genomics-Guided Exploitation of Lipopeptide Diversity in Myxobacteria

2.1 Abstract

Analysis of 122 myxobacterial genome sequences suggested 16 strains as producers of the myxochromide lipopeptide family. Detailed sequence comparison of the respective *mch* biosynthetic gene clusters informed a genome-mining approach, ultimately leading to the discovery and chemical characterization of four novel myxochromide core types. The myxochromide megasynthetase is subject to evolutionary diversification, resulting in considerable structural diversity of biosynthesis products. The observed differences are due to the number, type, sequence, and configuration of the incorporated amino acids. The analysis revealed molecular details on how point mutations and recombination events led to structural diversity. It also gave insights into the evolutionary scenarios that have led to the emergence of *mch* clusters in different strains and genera of myxobacteria.

2.2 Introduction

Lipopeptides (LPs) constitute a distinguished class of microbial secondary metabolites with multifaceted biological functions. These include, for example, surfactant, antimicrobial, or cytotoxic activities attracting interest in diverse industrial fields.^{1,2} Bacterial LPs are generally synthesized in a ribosome-independent manner by large multimodular enzymes called nonribosomal peptide synthetases (NRPS).³ Often, they are produced as mixtures of components varying in their lipid tail and/or amino acid composition.^{4,5} Whereas various LP biosynthetic pathways from *Bacillus*, *Pseudomonas*, *Streptomyces*, and other microbes have been intensively studied over the past decades,⁶ LP biosynthesis in myxobacteria is currently underexplored. Most prominent among the very few products characterized so far are the myxochromides, for which three different types have been described: myxochromides A (1) from *Myxococcus xanthus*,⁷ myxochromides B (2) from an unclassified *Myxococcus* sp.,⁸ and myxochromides S (3) from *Stigmatella aurantiaca*.⁹ The cyclic peptide structures are composed of five to seven amino acids and are linked to a polyunsaturated acyl chain, which can vary in length (Figure 1). Biosynthetic studies on myxochromides A (lipohexapeptides) and myxochromides S (lipopentapeptides) revealed that an iterative polyketide synthase (PKS) generates the lipid chains, thereby initiating the assembly of the peptidyl backbone catalyzed by a hexamodular NRPS system.^{7,9} Specifically, it was shown how point mutations switch substrate specificity in NRPS modules and cause an unusual “module-skipping” process during biosynthesis of the myxochromide S pentapeptide core.⁷ However, the

assembly line for production of myxochromides B (lipoheptapeptides) was not deciphered yet. Driven by the progress in bacterial genome-mining in general,¹⁰ as well as myxobacterial genomics,¹¹ and secondary metabolomics,¹² we aimed to achieve a comprehensive understanding of the biosynthesis of this versatile LP family and thereby intended to further exploit myxochromide chemical diversity.

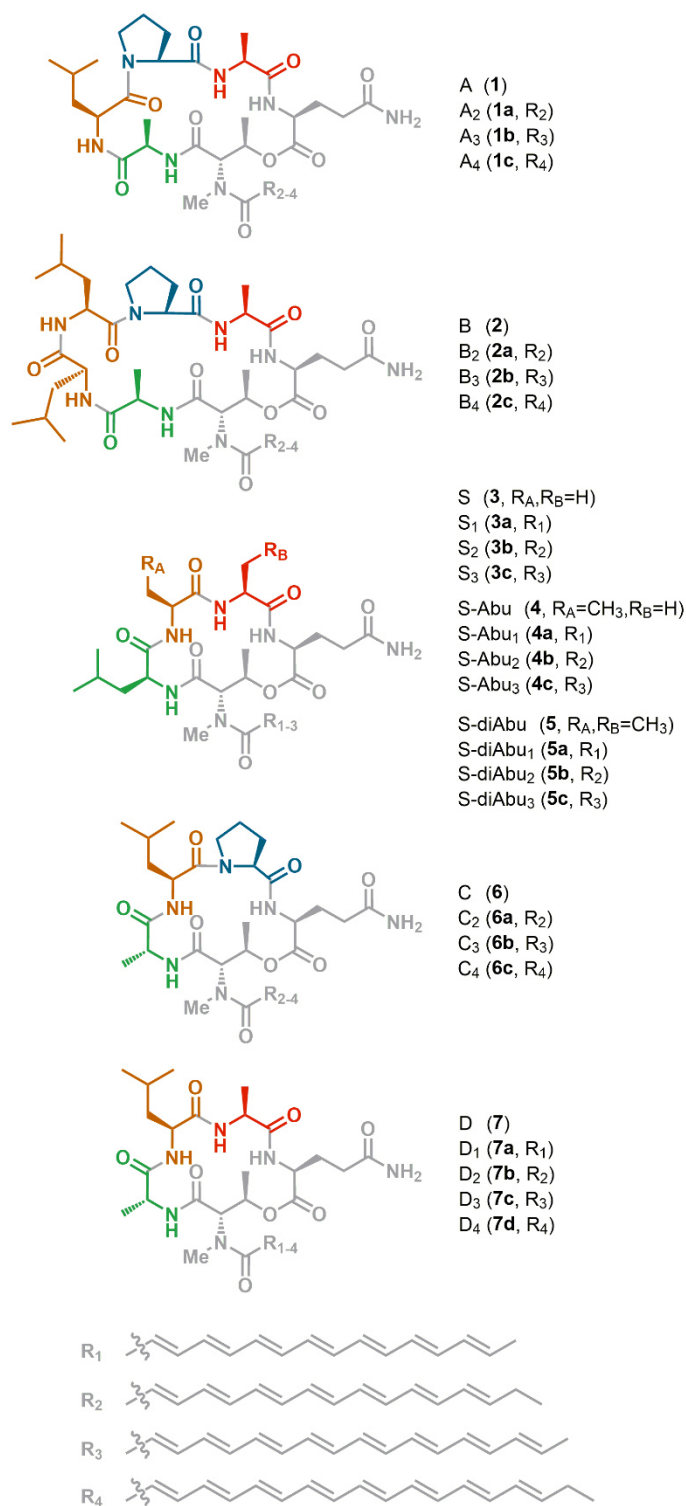


Figure 1. Chemical structures of myxochromides.

2.3 Results and Discussion

2.3.1 Discovery of Novel Types of Myxochromide Megasyntetases

Initially, we targeted to establish a molecular comprehension of how the myxochromide B pathway evolved from the known A- and S-type pathways. The putative B-type biosynthetic gene cluster harbors large sequence repeats and could thus only be deciphered using shotgun genome data of the producer *Myxococcus* sp. 171 in conjunction with significant additional sequencing efforts based on the generation of a cosmid library (Supporting Information Figure S1).

Table1: Myxochromide producers and their biosynthetic gene clusters (*mch* clusters) analyzed in this study.

Strain	Strain abbrev.	Cluster	GenBank Accession ^[a]
<i>Myxococcus fulvus</i> HW-1	Mf1	A-type	KX622592
<i>Myxococcus xanthus</i> DK1622	Mx1	A-type ^[b]	KX622595
<i>Myxococcus xanthus</i> DK897	Mx2	A-type	KX622596
<i>Myxococcus xanthus</i> A47	Mx3	A-type	KX622597
<i>Myxococcus xanthus</i> Mx48	Mx4	A-type	KX622598
<i>Myxococcus</i> sp. 171 ^[c]	M1	B-type	KX622591
<i>Myxococcus virescens</i> ST200611	Mv1	C-type	KX622594
<i>Myxococcus hansupus mixupus</i>	Mh1	C-type	KX622593
<i>Cystobacterineae</i> sp. CcG34 ^[d]	Cy1	D-subtype 2	KX622587
<i>Hyalangium minutum</i> DSM14724	Hm1	D-subtype 1	KX622588
<i>Hyalangium minutum</i> Hym-3	Hm2	D-subtype 1	KX622589
<i>Hyalangium minutum</i> NOCb10	Hm3	D-subtype 1	KX622590
<i>Stigmatella erecta</i> Pde77	Se1	D-subtype 1	KX622602
<i>Stigmatella aurantiaca</i> DW4/3-1	Sa1	S-type ^[e,f]	KX622599
<i>Stigmatella aurantiaca</i> Sga15	Sa2	S-type ^[f]	KX622600
<i>Stigmatella aurantiaca</i> Sga32	Sa3	S-type ^[f]	KX622601

[a] Annotated cluster files were additionally deposited in the MiBIG database (BGC0001417-BGC0001432)^[4]; [b] Characterized in previous study.⁷ [c] Unclassified strain belonging to the genus *Myxococcus*. [d] Unclassified species belonging to the suborder *Cystobacterineae* (16S analysis indicates that this species might belong to the genus *Hyalangium*). [e] Characterized in previous study.⁹ [f] S-type gene clusters were shown to also produce myxochromides S-Abu and S-diAbu.

Sequence analysis of the encoded assembly line revealed an additional leucine-specific NRPS module in comparison to the hexamodular A-type megasynthetase, indicating that the B-type pathway emerged from a module duplication event. To obtain deeper insights into

myxochromide pathway evolution, we screened genome data from 122 myxobacterial strains subjected to automated annotation¹³ and identified numerous additional putative myxochromide biosynthetic gene clusters (*mch* clusters) exclusively from species belonging to the suborder *Cystobacterineae* (Table 1).

Among those, four other A-type pathways (from strains Mf1, Mx2, Mx3, and Mx4) and two other S-type pathways (from strains Sa2 and Sa3) were retrieved. Intriguingly, some of the detected putative *mch* clusters (from strains Mv1, Mh1, Cy1, Hm1, Hm2, Hm3 and Se1) appeared to encode novel types of myxochromide megasynthetases differing in their domain arrangement and functionality from the previously characterized A- and S-types^{7,9} as well as the newly discovered putative B-type megasynthetase. The new *mch* cluster types were designated as C-type and D-type pathways predicted to direct the biosynthesis of two novel pentapeptide cores (see Supporting Information).

2.3.2 Genome-Mining for Novel Myxochromide Lipopeptide Cores

To connect this genetic information with the actual biosynthesis products, the putative producer strains listed in Table 1 (with exception of strain Mh1) were cultivated and analyzed for myxochromide production by HPLC-MS (Supporting Information Figure S3). Indeed, novel myxochromide derivatives exhibiting MS data consistent with the expected myxochromide C and D derivatives could be detected in extracts of the respective strains. The genome-mining approach was pursued by isolation of a representative of the two novel myxochromide types, later designated as myxochromide C₃ from strain Mv1 and myxochromide D₁ from strain Se1. Structure elucidation including assignment of the absolute stereochemistry unambiguously confirmed consistency with the *in silico* predicted structures featuring novel pentapeptide cores (Figure 1). Besides myxochromides C and D, additional putative novel myxochromide types, designated as myxochromides S-Abu and S-diAbu, were detected as minor products from myxochromide S producing *S. aurantiaca* strains. A representative of each myxochromide type was isolated from a previously described high titer myxochromide S heterologous production strain.¹⁵ Structure elucidation revealed that the minor products indeed represent novel lipopentapeptides that differ from myxochromide S by replacement of *L*-alanine with *L*- α -aminobutyric acid (Abu) in one or two positions (Figure 1). Putative novel Abu-containing myxochromide derivatives were later on also detected from A-, B-, C- and D-type producer strains after supplementing the cultures with *L*-Abu (Supporting Information Figure S5). In summary, a combined myxobacterial genomics/secondary metabolomics approach led to the discovery of four novel

myxochromide types (myxochromides C, D, S-Abu, and S-diAbu; Figure 1). Although there is no direct experimental evidence for the correlation of the newly identified “putative” *mch* clusters to myxochromide production (e.g., via gene inactivation or heterologous expression), sequence comparison with described *mch* clusters from Mx1 and Sa1 and myxochromide production analysis (Supporting Information Figure S3) strongly supports the classification of *mch* pathways as shown in Table 1. In the following discussion regarding the *mch* pathway comparison, the term “putative” is avoided for better readability.

2.3.3 Comparative Analysis of Myxochromide Megasyntetases

We next analyzed all respective biosynthetic gene clusters aiming at a deeper understanding of the biochemical and genetic basis for the observed structural diversity. Figure 2 illustrates the different types of myxochromide megasyntetases, which consist of three subunits: an iterative PKS (MchA) and two NRPSs (MchB, MchC).

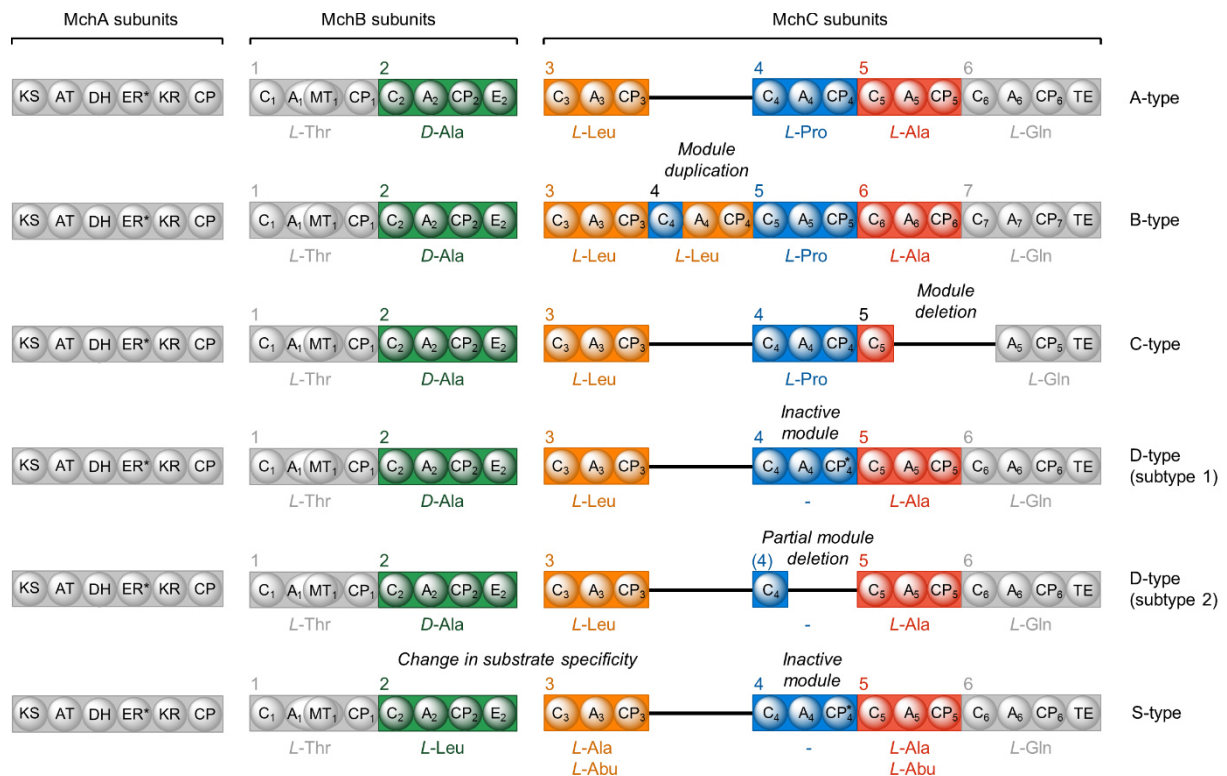


Figure 2. Different types of myxochromide megasyntetases from myxobacteria (for host strains see Table 1). The coloring of the catalytic domains corresponds to the biosynthesis products shown in Figure 1 (for phylogenetic analysis, see Supporting Information Figure S23). As verified for S-type pathways, there are indications for alternative incorporation of α -aminobutyric acid (Abu) by alanine (Ala)-specific modules of A-, B-, C-, and D-type megasyntetases (see Supporting Information Figure S5). Catalytic domain abbreviations: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; CP, acyl carrier protein (in MchA) or peptidyl carrier protein (in MchB/C); C, condensation domain; A, adenylation domain; MT, methyltransferase; E, epimerization domain; TE, thioesterase. Domains marked with an asterisk are supposed to be inactive.

Whereas all MchA homologues as well as all MchB homologues show identical arrangements of catalytic domains, respectively (MchA: KS-AT-DH-ER*-KR-CP and MchB: C₁-A₁-MT₁-CP₁-C₂-A₂-CP₂-E₂; Figure 2), MchC subunits differ significantly. Compared to the A-type assembly line, the B-type megasynthetase features an additional leucine-incorporating module consistent with the biosynthesis of its lipoheptaepptide products (**2a–2c**). On the contrary, the C-type megasynthetase lacks a complete module, resulting in the production of novel lipopentapeptides (**6a–6c**), which do not contain *L*-alanine as predicted from *in silico* analysis and confirmed by genome-mining. Phylogenetic analysis of NRPS condensation (C), adenylation (A), and carrier protein (CP) domains (Supporting Information Figure S23) from all 16 myxochromide megasynthetases indicated that “A-CP-C units” instead of dedicated “C-A-CP modules” are duplicated (A₃-CP₃-C₄ in B-type) or deleted (A₅-CP₅-C₆ in C-type). Detailed sequence analysis of the MchC subunits from *Myxococcus* strains even revealed the corresponding recombination sites, which are located at the 5′ end of the A domain or 3′ end of C domain encoding regions, respectively (Supporting Information Figures S20 and S21). Another recombination event led to partial module deletion in the D-subtype 2 assembly line, which lacks the A₄-CP₄ domains (Supporting Information Figure S22). However, as module 4 in the related D-subtype 1 megasynthetase is already inactive due to the mutated CP₄ core motif (Supporting Information Figure S18), the same myxochromide structures were actually expected and found to be produced by the two different D-type assembly lines. In accordance with the *in silico* analysis, the biosynthesis products represent novel lipopentapeptides (**7a–7d**) lacking *L*-proline as compared to myxochromides A (1). A similar scenario (inactive module 4) causing so-called “module-skipping”¹⁶ was already observed for the myxochromide S megasynthetase,⁷ which also differs from all other myxochromide assembly lines in terms of module 2 and 3 biochemistry (*L*-Leu/*L*-Ala instead of *D*-Ala/*L*-Leu). Here, sequence-based structure prognosis did not allow prediction of accurate stereochemistry and substrate specificity (see Supporting Information). However, the *in silico* analysis suggests that point mutations rather than intragenic A domain swapping¹⁷ cause the observed reversed amino acid order in myxochromides S (3). Overall, the diversity of detected assembly line variations (Figure 2) shows that the myxochromide biosynthetic pathway is subject to significant evolutionary diversification (compare, *e.g.*, cyanobacterial toxin pathways¹⁸ and other expamples^{19–22}). Sequence analysis of *mch* genes revealed that adaptation in local synonymous codon usage is reduced in module 3 and 4 regions. Especially the A domain regions show lower adaptation compared to other modules, which may indicate recent mutational activity and diversification or more recent acquisition (see Supporting

Information).²³ Even further reduced is the codon adaptation of inactive A₄-CP₄ domains from *Stigmatella* producers and also the region encoding the inactive PKS ER* domain in all *mch* pathways (Supporting Information Figure S25).

2.3.4 Evolutionary Relationship and Distribution of *mch* Clusters in Myxobacteria

According to the analyzed myxobacterial genome data, *mch* clusters only occur in strains belonging to the *Cystobacterineae* suborder, more precisely to the *Cystobacteraceae* and *Myxococcaceae* families. Comparative phylogenetic analysis of the 16 *mch* clusters with selected house-keeping genes of their hosts revealed an overall congruence between strain and cluster phylogeny with one exception: The D-type pathway from *S. erecta* Pde77 (Se1), which was likely obtained via horizontal gene transfer from *Hyalangium* species (Figure 3).

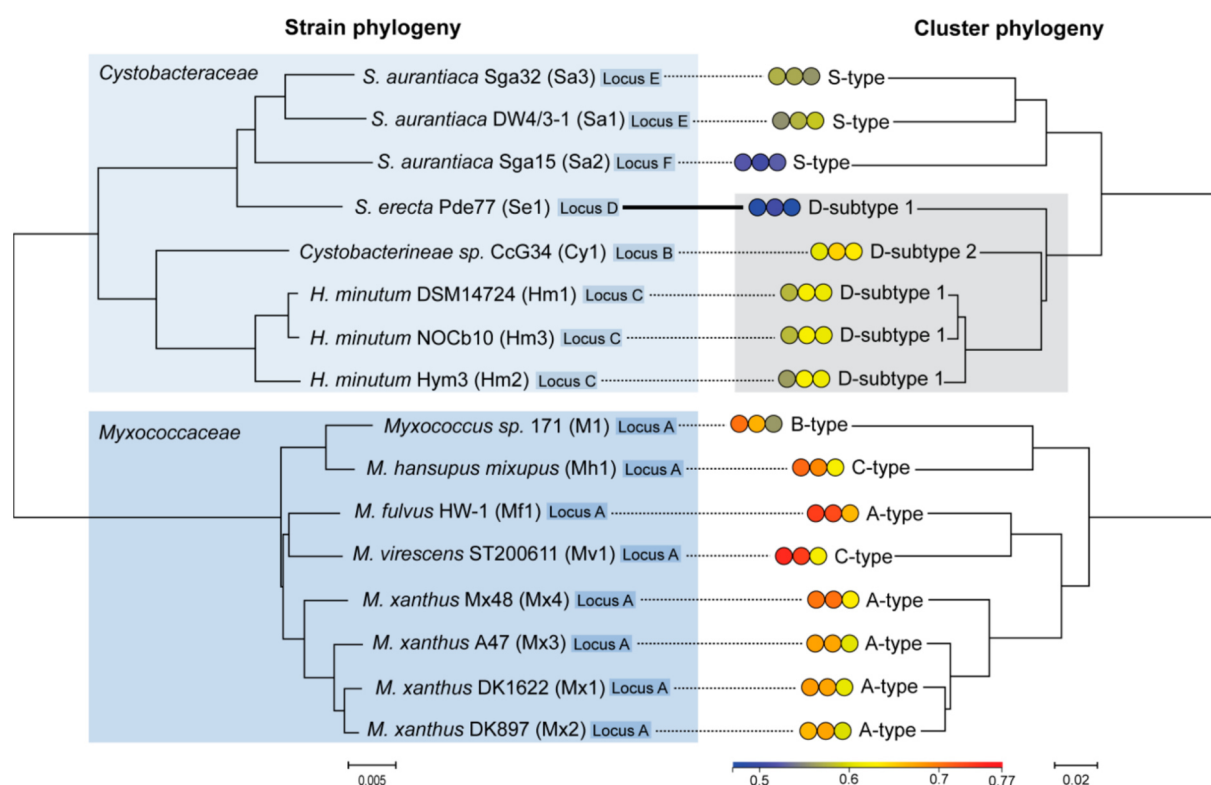


Figure 3. Comparison of myxochromide producer strain and *mch* cluster phylogeny (for details, see Supporting Information). The three megasynthetase coding sequences (CDS) (*mchA*, *mchB*, *mchC*) are shown as disks, colored according to their adaptation of synonymous codon usage to the highly expressed gene sets of their hosts (see Supporting Information Table S12). Common cluster genome environments are indicated (Locus A–F; see Supporting Information Figure S2).

In addition, it can be assumed that *Stigmatella* S-type pathways derive from D-subtype 1 clusters and diversify via additional mutations switching substrate specificities of NRPS

modules 2 and 3 (Figure 2). Within the *Myxococcaceae*, three different pathway types (A, B and C) occur, which deviate in the MchC subunit due to module duplication/deletion events resulting in heptamodular/pentamodular NRPS assembly lines. However, the hexamodular A-type pathway represents the most common variant in this analysis occurring in all *M. xanthus* species and in *M. fulvus* (Table 1; Figure 3). In fact, a study on secondary metabolomes from 98 different *M. xanthus* strains determined myxochromides A as ubiquitous metabolites from this species.²⁴ Overall, our phylogenetic analysis strongly suggests that myxochromide biosynthetic pathways have evolved from a common ancestor, which based on all available data seems to be the A-type cluster. Codon adaptation analysis revealed that *mch* clusters found in *Cystobacteraceae* are much less adjusted to their hosts than clusters from *Myxococcaceae*, which thus seem to have resided in their hosts for a longer time period (Figure 3). The best average adaptation value was obtained for the cluster from strain Mf1 supporting the proposed ancestral role of the A-type pathway. It is likely that the also well adapted C-type cluster of Mv1 evolved in the same genomic context (Supporting Information Figure S2). The less adapted A-type clusters were then distributed to other *M. xanthus* strains, from which the even younger and less adapted B-type cluster of M1 originated. The original C-type cluster was transferred to Mh1. All of these clusters are currently found in the same genomic context. Several independent integration events led, in the sequence suggested by the codon adaptation level, to the origin of the D-subtype 2 cluster of Cy1; the three D-subtype 1 clusters of Hm1, Hm2, and Hm3; the two S-type clusters from Sa1 and Sa3; and lately to the quite little adapted S-type cluster of Sa3. Most recently, the D-subtype 1 cluster was integrated into Se1, where it shows the least adaptation.

2.3.5 Potential Biological Function of Myxochromides

The widespread occurrence of myxochromides among different myxobacterial species suggests a relevant biological function for the producer strain as known from other bacterial LPs.²⁵ As myxochromides did not exhibit significant antimicrobial or cytotoxic activities,^{8,9} they might play a role in the developmental life cycle of myxobacteria. This was analyzed in established assays with the myxochromide producer *M. xanthus* DK1622 and respective myxochromide overproduction and knockout mutants (Supporting Information Figures S26–27). It could be demonstrated that myxochromide overproduction has a negative effect on fruiting body development, which correlates with observations from previous studies.²⁶ This finding may be explained by myxochromide surfactant properties, which probably increase cell motility and hamper cell aggregation. As no difference between the *mch* cluster

knockout mutant and the wildtype strain could be observed, myxochromides do not seem to be essential for fruiting body formation and swarming of *M. xanthus*. However, specific biological functions of myxochromides, possibly even different functions for structural variants, can currently not be excluded.

2.4 Significance

In conclusion, we demonstrate how myxobacterial genome-mining can provide a broad picture regarding the distribution, genetic basis, and evolution of secondary metabolite pathways. Additionally, our approach enabled the discovery and isolation of novel derivatives of the selected lipopeptide compound class by prioritizing potential producer strains for chemical analysis. Detailed biosynthetic pathway comparison revealed an impressive diversification of the involved megasynthetases among various producer strains and allowed rationalization of the observed structural differences of the corresponding products. Our comprehensive study thus provides a broad picture on the interrelationship of genetic changes as causative agents for chemical diversification of lipopeptides from a whole taxon of microorganisms. The described genetic changes leading to natural product pathway evolution (recombination sites, point mutations, codon adaptation) represent valuable information for the future engineering of microbial NRPS pathways via synthetic biology approaches.^{27,28}

2.5 Experimental Procedures

2.5.1 Identification and Decipherment of the Myxochromide B Biosynthetic Gene Cluster from *Myxococcus* sp. 171.

Shotgun genome sequence data of strain *Myxococcus* sp. 171 (M1) were generated by employing Illumina sequencing technology, and the putative myxochromide B biosynthetic gene cluster was identified by comparative sequence analysis using the published myxochromide A gene cluster from *M. xanthus* DK1622 as reference (Genbank accession number CP000113). The identified myxochromide B gene cluster region contained several sequence gaps. For the decipherment of the entire gene cluster sequence, a cosmid library of strain M1 comprising 3072 clones was constructed on the basis of the SuperCos 1 cosmid vector (Agilent Technologies). In a PCR-based approach, the library was screened for cosmids harboring the myxochromide B gene cluster or parts thereof. Two of the identified cosmids (Cos8F11 and Cos1P19) were used as starting constructs to subclone fragments covering the unresolved gene cluster region (see Supporting Information Figure S1). The inserts from four of the generated plasmids (pASK_mchB6, pBCSK_mchB5, pBCSK_mchB4, and pBCSK_mchB3) were completely sequenced, which allowed for

decipherment of the repetitive myxochromide B gene cluster regions. For additional details, see Supporting Information Chapter 2.7.1.

2.5.2 Screening of Myxobacterial Genome Data for Additional Myxochromide Biosynthetic Gene Clusters and Verification via Production Analysis

Genome data from 122 myxobacterial species including representatives of the three suborders *Cystobacterineae* (68 strains), *Sorangiiineae* (43 strains), and *Nannocystineae* (11 strains) were screened for putative myxochromide biosynthetic gene clusters by using the antiSMASH 3.0.4 tool;¹³ all strains are listed in Supporting Information Chapter 2.7.2. In total, 16 myxochromide biosynthetic gene clusters (*mch* clusters) were identified in strains belonging to the suborder *Cystobacterineae*. The genomic context of the *mch* cluster was analyzed in each host strain, revealing six different chromosomal loci (A–F; see Figure 3 and Supporting Information Figure S2). The automatic annotation of the *mch* gene clusters from the antiSMASH 3.0.4 analysis¹³ was manually revised based on reported core motifs and structural data of other PKS/NRPS megasynthetases. The gene cluster data were submitted to GenBank (accession numbers are listed in Table 1) and deposited in the MiBIG database.¹⁴ The domain organization of the encoded megasynthetases was compared and revealed novel types of putative myxochromide biosynthetic pathways (C- and D-type clusters; Table 1 and Figure 2), for which putative pathway products were predicted based on the *in silico* data. Myxochromide production analysis of the identified strains listed in Table 1 (except Mh1) was carried out on a 50 mL scale under routine cultivation conditions. Crude extracts were prepared and subjected to HPLC-MS analysis including FT-ICR-MS² measurements to characterize the myxochromide production profiles (Supporting Information Figures S3 and S4). Selected producer strains (Mx1, M1, Mv1, Se1, and Sa1) were additionally grown in media supplemented with *L*- α -aminobutyric acid (*L*-Abu), and the effect on myxochromide production profiles was analyzed by HPLC-MS (Supporting Information Figure S5). For additional details, see Supporting Information Chapter 2.7.2.3.

2.5.3 Isolation and Structure Elucidation of Novel Myxochromides

Representatives of four novel myxochromide types (C-type, D-type, S-Abu-type, and S-diAbu type) were isolated from culture extracts of selected producer strains via a combination of size exclusion chromatography and reversed-phase HPLC. The planar structures were elucidated based on 1D (¹H) and 2D (¹H–¹H COSY, HSQC, HMBC, and ROESY) NMR spectroscopy as well as HR-ESI-MS data. The absolute configuration of the amino acid residues was

determined by HPLC-MS analysis of the *L*- and *D*-FDLA (1-fluoro-2,4- dinitrophenyl-5-*L*-/*D*-leucinamide) derivatives of the acid hydrolysate of myxochromides in comparison with corresponding derivatives of *L*-configured amino acid standards (Marfey's method²⁹). The following myxochromide derivatives were purified and structurally characterized.

Myxochromide C₃ (6b). A total of 2.3 mg of **6b** was isolated from a 20 L culture of *Myxococcus virescens* ST200611 (Mv1). The molecular formula was established to be C₄₂H₅₈N₆O₈ (*m/z* 775.43921 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S6 and S7, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **6b** hydrolysate is shown in Supporting Information Figure S8. For details, see Supporting Information Chapter 2.7.3.

Myxochromide D₁ (7a). A total of 2.7 mg of **7a** was isolated from a 20 L culture of *Stigmatella erecta* Pde77 (Se1). The molecular formula was established to be C₃₈H₅₄N₆O₈ (*m/z* 723.40748 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S9 and S10, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **7a** hydrolysate is shown in Supporting Information Figure S11. For details, see Supporting Information Chapter 2.7.4.

Myxochromide S₂-Abu (4b). A total of 5.3 mg of **4b** was isolated from a 9 L culture of the heterologous production strain *M. xanthus* DK1622::pTpS-mchS.¹⁵ The molecular formula was established to be C₄₀H₅₈N₆O₈ (*m/z* 751.43940 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S12 and S13, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **4b** hydrolysate is shown in Supporting Information Figure S14. For details, see Supporting Information Chapter 2.7.5.

Myxochromide S₂-diAbu (5b). A total of 6.4 mg of **5b** was isolated from a 4.5 L culture of the heterologous production strain *M. xanthus* DK1622::pTpS-mchS.¹⁵ The molecular formula was established to be C₄₁H₆₀N₆O₈ (*m/z* 765.45520 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S15 and S16, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **5b** hydrolysate is shown in Supporting Information Figure S17. For details, see Supporting Information Chapter 2.7.6.

2.5.4 Detailed *in Silico* Analysis of the 16 *mch* Clusters

The *mch* clusters listed in Table 1 were analyzed according to different aspects using established bioinformatics tools. Protein sequence alignments of selected catalytic domains were performed using the Geneious alignment tool integrated into Geneious software version 9.1.2.³⁰ As shown in Supporting Information Figure S18, these include alignments of C

domains from modules 2 and 3 (C₂ and C₃) and E domains from module 2 (E₂) as well as PCP domains from modules 2, 3, and 4 (PCP₂, PCP₃, and PCP₄). The reported C, E, and PCP domain core motifs^{31,32} and additional residues to distinguish ^LC_L and ^DC_L domain subtypes³³ were analyzed. In addition, substrate specificities of all A domains from the 16 myxochromide megasynthetases (in total 94 A domains) were analyzed using the antiSMASH 3.0¹³ and NRPSpredictor2 analysis tools.³⁴ The substrate predictions together with the A domain 8 Å signatures and Stachelhaus codes^{35,36} are shown in Supporting Information Figure S19. In order to identify recombination sites from domain duplication and deletion events, alignments with selected regions of *mchC* genes were performed using the Geneious alignment tool integrated into Geneious software version 9.1.2.³⁰ The outcome of this analysis is illustrated in Supporting Information Figure S20 for the “module duplication” (myxochromide B pathway), Supporting Information Figure S21 for the “module deletion” (myxochromide C pathway), and Supporting Information Figure S22 for the “partial module deletion” (myxochromide D-subtype 2 pathway). The analysis of the recombination events was supported by findings from phylogenetic analysis of myxochromide megasynthetase NRPS domains. Phylogenetic trees for A, C, and PCP domains, which were performed based on DNA sequences, are shown in Supporting Information Figure S23. In addition, phylogenetic relationships of all myxochromide megasynthetase subunits were analyzed based on DNA and protein sequences (*mchA*/MchA, *mchB*/MchB, and *mchC*/MchC; see Supporting Information Figure S24). Phylogenetic reconstitutions were carried out by applying the distance-based neighbor-joining method using the modules “neighbor” and “dnadist” or “protdist” of the PHYLIP package.³⁷ The sequences were aligned using the GUIDANCE2 Server³⁸ by applying the MAFFT algorithm or using ClustalX.³⁹ For the comparison of strain and *mch* cluster phylogenies illustrated in Figure 3, the *mchA*, *mchB*, and *mchC* nucleotide sequences of the myxochromide megasynthetase encoding genes were concatenated (cluster phylogeny), and the nucleotide sequences of the 16S rRNA gene were concatenated with nucleotide sequences of 15 selected genes encoding highly conserved proteins (strain phylogeny). The phylogenetic analysis was complemented by codon adaptation index (CAI⁴⁰) analysis, which was performed based on a set of selected high expression genes from the analyzed myxochromide producer strains (“hxp2 gene sets”, each harboring 343–405 genes). On the basis of these gene sets, codon usage tables were calculated by the EMBOSS program cusp.⁴¹ The respective sequence-specific CAI for the *mch* cluster genes was calculated by a Perl program as described by Sharp and Li⁴⁰ and is illustrated for *mchA*, *mchB*, and *mchC* in Figure 3 (for CAI values see Supporting Information

Table S12, Chapter 2.7.7.4). In addition, local CAI values along the CDS of *mchA*, *mchB*, and *mchC* were calculated for overlapping CDS regions of 101 codons using a Perl program and were visualized as plotted color-shaded vertical lines along the CDS. For comparison of the local CAI distribution in the *mchA*, *mchB*, and *mchC* gene sets from the 16 analyzed myxochromide producers, codons were aligned to orthologous positions based on protein sequence alignments performed with ClustalX.³⁹ Alignments of *mchC* were modified according to the identified recombination sites. The local CAI values along the *mchA*, *mchB*, and *mchC* CDS sequences of the 16 analyzed myxochromide pathways are illustrated in Supporting Information Figure S25. For additional details, see Supporting Information Chapter 2.7.7.4.

2.5.5 Fruiting Body Formation and Swarm Expansion Assays

The myxobacterial model strain and myxochromide A producer *M. xanthus* DK1622 wild type (Mx1) was analyzed in established developmental assays in comparison to three mutants. These include *M. xanthus* DK1622::pMch22a,⁷ which is deficient in myxochromide A production, *M. xanthus* DK1622::pMch70a (Wenzel et al., unpublished), which produces about 10-fold higher amounts of myxochromides A compared to the wildtype, and *M. xanthus* DK1622::pTpsmchS,¹⁵ which produces myxochromides S in around 50-fold higher amounts in addition to wildtype levels of myxochromides A. The results from the comparative fruiting body formation assays are illustrated in Supporting Information Figure S26. The outcome of the comparative swarm expansion assays, which were performed according to the procedure described by Kaiser et al.⁴² is shown in Supporting Information Figure S27. For additional details, see Supporting Information Chapter 2.7.8.

2.6 Acknowledgements

The authors would like to thank D. Auerbach for Marfey's analysis measurements, I. Kochems for conducting fruiting body assays, R. Garcia for microbiological assistance, and F. Yan for his comments on this manuscript. We acknowledge R. Müller and G. Velicer for providing myxobacterial genome data for *mch* cluster screening and analysis. Furthermore, we gratefully acknowledge R. Müller for constantly supporting and discussing this study and for manuscript revision. This work was generously supported by a grant from the German Federal Ministry of Education and Research (FKZ: 031A155). We thank the "SynBioDesign" teams from ATG Biosynthetics GmbH and PharmBioTec GmbH for supporting this study.

2.7 Supporting Information

2.7.1 Identification and Decipherment of the Myxochromide B Biosynthetic Gene Cluster from *Myxococcus* sp. 171

2.7.1.1 Shotgun Genome Sequencing of *Myxococcus* sp. 171 and Identification of the Myxochromide B Biosynthetic Gene Cluster

Draft genome sequence of the myxochromide B producer *Myxococcus* sp. 171 was determined by employing Illumina sequencing technology at Seq-It GmbH (Kaiserslautern, Germany). Paired-end sequencing library was prepared from the strain's total DNA. This library was then sequenced to a mean genome coverage of 357x; mean read length was 251 bp; mean distance between reads (fragment size) was 378 bp. Raw sequencing data were then assembled by Abyss-pe software 1.3.6 to yield 61 contigs in 47 scaffolds. The estimated genome size of *Myxococcus* sp. 171 was 9,603,576 bp. The putative myxochromide B biosynthetic gene cluster was identified by comparative sequence analysis using the myxochromide A biosynthetic gene cluster from *Myxococcus xanthus* DK1622 (Genbank accession number CP000113) as a reference.⁴³ However, due to repetitive sequence elements, the sequence of the *mchC* gene was not completely resolved (see Figure S1).

2.7.1.2 Construction and Screening of a Cosmid Library from *Myxococcus* sp. 171

To elucidate the entire gene cluster sequence, a cosmid library of the producer strain was constructed to enable subcloning and sequencing of unsolved gene cluster fragments. Chromosomal DNA from *Myxococcus* sp. 171 was prepared by standard phenol/chloroform/isoamylalcohol extraction.⁴⁴ Isolated genomic DNA was subsequently used for the construction of a cosmid library according to the SuperCos 1 Cosmid Vector Kit protocol (Agilent Technologies). Cosmid packaging reactions were carried out using the Gigapack III Packaging Extract Kit (Agilent Technologies) according to the manufacturer's protocol. The resulting cosmid library consists of 3072 clones (8×384 microtiter plates, MTPs), which were subsequently screened for cosmids harboring the putative myxochromide B biosynthetic gene cluster or fragments thereof. A PCR-based screening approach was applied using three oligonucleotide pairs specific for both ends of the gene cluster as well as for an internal region around the *mchB/mchC* interface (see Figure S1). PCRs were performed with Taq polymerase (Fermentas) using different cosmid DNA pools or DNA from single cosmids as templates and the following primer pairs (expected PCR product sizes are indicated):

Oligonucleotide	Sequence	Amplicon size
P1	5'TGCGAGGATGATTCTGGGC-3'	} 340 bp
P2	5'-TGTCAACAGTTCCCAGAGG-3'	
P3	5'-AGACCCTCGAGCGGCTGAAC-3'	} 411 bp
P4	5'-TCCCACAGAAAGGCCGAACG-3'	
P5	5'-ATGGGCGTGGGGGTGGTGTAG-3'	} 299 bp
P6	5'-GTTGATTTCACGCGACTTCTGG-3'	

Initially, the 3072 library clones were pooled and inoculated as 16 sets, each representing 192 cosmid clones from a half MTP. The combined clones were grown in 1.5 mL LB medium (tryptone 10 g/L, yeast extract 5 g/L, NaCl 5 g/L) amended with kanamycin (50 µg/mL) at 37 °C overnight to isolate the respective cosmid DNA mixtures via standard alkaline lysis.⁴⁴ Cosmid sets revealing PCR products with primer pairs P1/P2 as well as P5/P6 were further screened by at first preparing 8 subsets of 24 clones, from which in the next step (in case of positive PCR results) the individual cosmid clones were analyzed. However, as this strategy did not lead to the identification of cosmids harboring the entire putative myxochromide B biosynthetic gene cluster, which would result in PCR products with the P1/P2 as well as P5/P6 primer set, the screening approach was altered to detect cosmid inserts covering one half of the gene cluster including the unsolved region of *mchC*. A number of cosmids yielding PCR products with P3/P4 and either P1/P2 or P5/P6 were identified and further analyzed by end-sequencing of the inserts using primers T3 (5'-ATTAACCCTCACTAAAGGGA-3') and T7 (5'-TAATACGACTCACTATAGGG-3'). Based on the obtained data, cosmids Cos1P19 and Cos8F11, both harboring the unsolved *mchC* region and combined harboring the entire gene cluster, were selected for subsequent subcloning approaches.

2.7.1.3 Subcloning and Sequencing of Myxochromide B Biosynthetic Gene Cluster Fragments

To decipher the sequence of the unsolved *mchC* region, detailed restriction analysis of cosmids Cos1P19 and Cos8F11 was performed. As illustrated in Figure S1B, seven plasmids harboring smaller gene cluster fragments including (part of) the gap region, were constructed by using standard cloning methods. First of all, a 8.9 kb *NotI* fragment was subcloned from

cosmid Cos8F11 to reveal plasmids pBCSK_mchB1-R1 and pBCSK_mchB1-R2 (harboring the insert in different directions). Based on these constructs, three smaller gene cluster fragments were subcloned and completely sequenced from the following three plasmids: pBCSK_mchB3 (4.3 kb *SpeI/PvuII* fragment from pBCSK_mchB1-R1 ligated into pBCSK(+) linearized with *SpeI/EcoRV*), pBCSK_mchB4 (1.5 kb *PvuII* fragment from pBCSK_mchB1-R1 ligated into pBCSK(+) linearized with *EcoRV*) and pBCSK_mchB5 (3.1 kb *SpeI/PvuII* fragment from pBCSK_mchB1-R2 ligated into pBCSK(+) linearized with *SpeI/EcoRV*). In parallel, a 9.0 kb *EcoRI/HindIII* fragment from cosmid Cos1P19 was subcloned into pASK-IBA6 (Iba Life Sciences) to reveal plasmid pASK_mchB2. Based on this construct, a 2.7 kb *StuI/HindIII* fragment was subcloned into pASK-IBA6 and completely sequenced from plasmid pASK_mchB6.

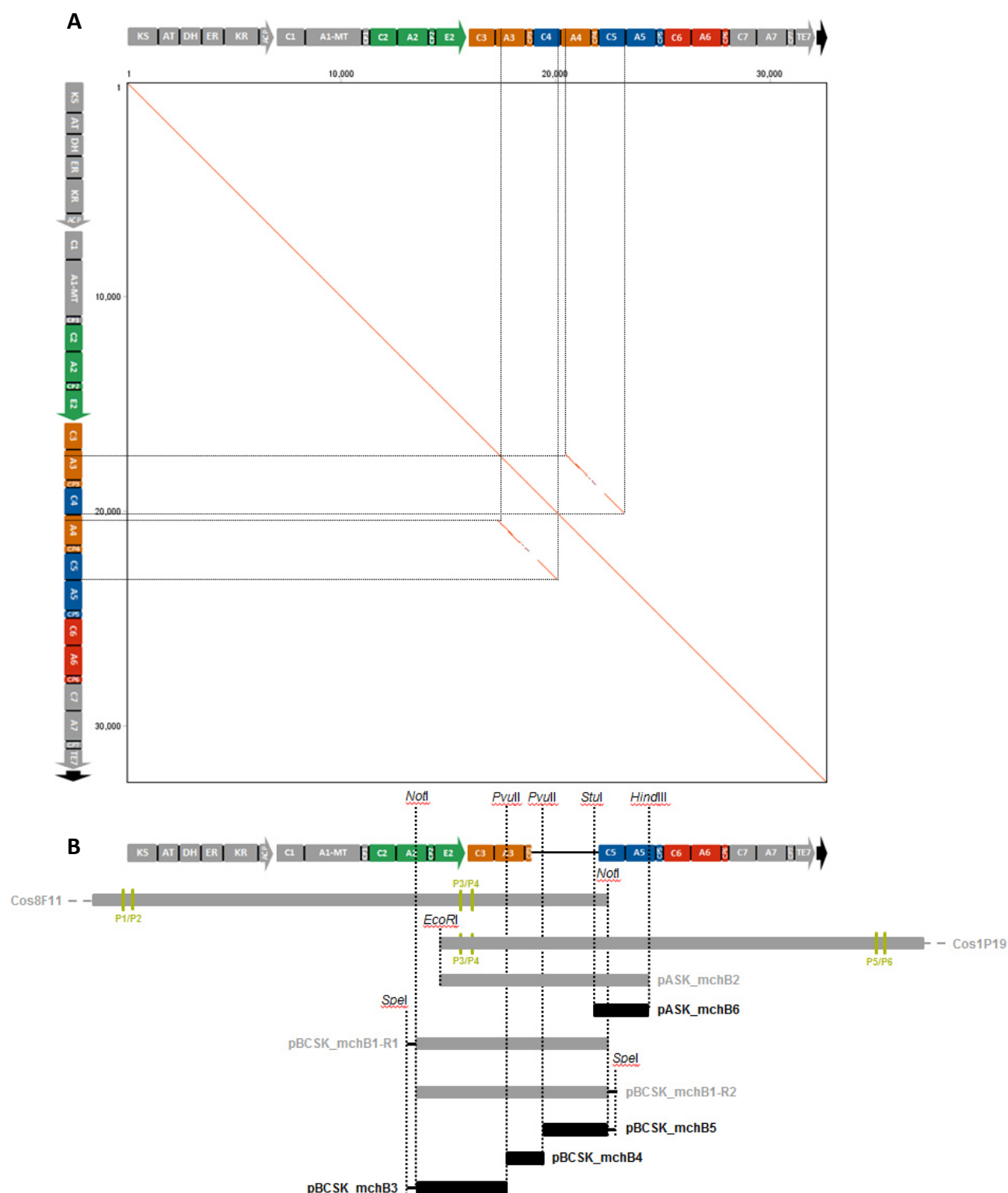


Figure S1. Illustration of repetitive regions within the *mchC* gene and subcloning strategy to decipher the complete putative myxochromide B biosynthetic gene cluster sequence. **A:** Dotplot of the *mchABCD* operon of the myxochromide B pathway performed with the Geneious 9.1.2 software⁴⁵ based on the EMBOSS 6.5.7 tool dottup.⁴¹ Large sequence repeats can be detected within the *mchC* gene, a region which could not completely deciphered based on shotgun genome sequence data. **B:** Subcloning approach based on two cosmids, which were identified by PCR-based screening of a cosmid library using three primer sets (P1/P2, P3/P4, P5/P6, highlighted in green). Seven plasmids harboring smaller gene cluster fragments were constructed (for details see chapter 2.7.1.3). Fragments shown in black were completely sequenced to reconstitute the repetitive region of *mchC*.

2.7.2 Screening of Myxobacterial Genome Data for Additional Myxochromide Biosynthetic Gene Clusters and Verification via Production Analysis

2.7.2.1 Identification and Annotation of Additional Myxochromide Biosynthetic Gene Clusters

For the identification of additional putative myxochromide biosynthetic gene clusters (*mch* clusters), 122 myxobacterial genome sequences were screened by using the antiSMASH 3.0.4 tool.¹³ The analyzed species include 68 representatives of the suborder *Cystobacterineae*, 43 representatives of the suborder *Sorangiiineae* and 11 representatives of the suborder *Nannocystineae* (Table S1). In addition to the previously described A-type *mch* cluster from *M. xanthus* DK1622,⁷ the S-type *mch* cluster from *S. aurantiaca* DW4/3-1,⁹ and the B-type *mch* cluster from *Myxococcus* sp. 171 (see chapter 2.7.1), 13 putative *mch* clusters were identified, from which 7 represent a new type of *mch* gene cluster (B-type, C-type, D-subtype 1, D-subtype 2; see Table 1 in the manuscript). The 16 identified *mch* clusters were exclusively found in myxobacterial strains, which belong to the suborder *Cystobacterineae*.

The *mch* clusters are organized in a four-gene operon (*mchA-mchD*) encoding an iterative type I polyketide synthase (MchA), two nonribosomal peptide synthetases (MchB and MchC) as well as a conserved hypothetical protein with unknown function (MchD), which was not reported as part of the myxochromide pathway in previous studies.^{7,9} Based on *in-silico* protein sequence analysis using the Pfam database,⁴⁶ the hypothetical protein MchD was predicted to be an integral membrane protein, since it contains the conserved DUF2269 domain. The MchD protein could potentially act as a transmembrane anchor for the myxochromide assembly lines. Association of a polyketide megasynthetase with the bacterial cell membrane was previously described for the myxovirescin biosynthetic machinery.⁴⁷

Analysis of the genomic context of the 16 *mch* gene clusters revealed their loci in the genomes of the respective producer strains (see Figure S2). The A-type, B-type and C-type *mch* operons of the analyzed *Myxococcus* species (Mf1, Mx1, Mx2, Mx3, Mx4, M1, Mh1, Mv1) share basically the same set of genes surrounding the *mch* gene clusters with only little differences (see Figure S2A). In contrast, the *mch* gene cluster loci from *Stigmatella* (Sa1, Sa2, Sa3, Se1) and *Hyalangium* strains (Hm1, Hm2, Hm3) as well as from the unclassified strain (Cy1) significantly differ from the loci, which were observed for *Myxococcus* species. Interestingly, the *mch* cluster loci from *Stigmatella* sp. and *Hyalangium* sp. also partly differ within the same family from each other, except the S-type *mch* gene clusters from Sa1 and Sa3, which are integrated in the same genomic locus.

To prepare the *mch* gene cluster sequences for detailed *in-silico* analyses, manual gap closing for the *mch* gene clusters from Mx2, Mx3 and Mv1 was performed by amplifying and

sequencing respective PCR products covering the sequence gap (not shown). Based on the automatic annotation of the *mch* gene clusters from the antiSMASH 3.0.4 analysis,¹³ the borders of catalytic domains from the polyketide synthase MchA and the two multimodular nonribosomal peptide synthetases MchB and MchC were manually revised and core motifs, which characterize these domains, were annotated as well.^{31,32} The considerably increased sequence information of *mch* gene clusters, in addition to the novel identified putative types of *mch* gene clusters (B-type, C-type, D-subtype1, D-subtype2) allowed for detailed characterization of the different pathways in order to explain the structural diversity of the myxochromide family on a genetic basis (see chapter 2.7.7).

Table S1. Myxobacterial genomes screened for *mch* gene clusters. Strains, which harbor a *mch* gene cluster, are highlighted in boldface.

No.	Strain	Suborder	Family	Genus
1	<i>Anaeromyxobacter dehalogenans</i> 2CP-1	Cystobacterineae	Cystobacteraceae	<i>Anaeromyxobacter</i>
2	<i>Anaeromyxobacter dehalogenans</i> 2CP-C	Cystobacterineae	Cystobacteraceae	<i>Anaeromyxobacter</i>
3	<i>Anaeromyxobacter</i> sp. Fw109-5	Cystobacterineae	Cystobacteraceae	<i>Anaeromyxobacter</i>
4	<i>Anaeromyxobacter</i> sp. K	Cystobacterineae	Cystobacteraceae	<i>Anaeromyxobacter</i>
5	<i>Anaeromyxobacter</i> sp. PSR-1	Cystobacterineae	Cystobacteraceae	<i>Anaeromyxobacter</i>
6	<i>Angiococcus disciformis</i> AngGT8	Cystobacterineae	Cystobacteraceae	<i>Angiococcus</i>
7	<i>Archangium gephyra</i> Ar8082	Cystobacterineae	Cystobacteraceae	<i>Archangium</i>
8	<i>Archangium gephyra</i> DSM 2261	Cystobacterineae	Cystobacteraceae	<i>Archangium</i>
9	<i>Archangium</i> sp. Ar3548	Cystobacterineae	Cystobacteraceae	<i>Archangium</i>
10	<i>Cystobacter armeniacae</i> Cba6	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
11	<i>Cystobacter ferrugineus</i> Cbfe23	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
12	<i>Cystobacter fuscus</i> DSM 2262	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
13	<i>Cystobacter fuscus</i> SBCb021	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
14	<i>Cystobacter</i> sp. MCy9104	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
15	<i>Cystobacter</i> sp. SBCb004	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
16	<i>Cystobacter velatus</i> Cbv34	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
17	<i>Cystobacter violaceus</i> Cb vi76	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
18	<i>Cystobacter violaceus</i> Cbvi35	Cystobacterineae	Cystobacteraceae	<i>Cystobacter</i>
19	<i>Cystobacterineae</i> sp. SBAr001 ¹⁾	Cystobacterineae	-	-
20	<i>Cystobacterineae</i> sp. SBCy008 ¹⁾	Cystobacterineae	-	-
21	<i>Cystobacterineae</i> sp. SBCy012 ¹⁾	Cystobacterineae	-	-
22	<i>Cystobacterineae</i> sp. SBCy016 ¹⁾	Cystobacterineae	-	-
23	<i>Cystobacterineae</i> sp. SBCy017 ¹⁾	Cystobacterineae	-	-
24	<i>Cystobacterineae</i> sp. SBCy018 ¹⁾	Cystobacterineae	-	-
25	<i>Cystobacterineae</i> sp. SBCy027 ¹⁾	Cystobacterineae	-	-
26	<i>Cystobacterineae</i> sp. SBCy030 ¹⁾	Cystobacterineae	-	-
27	<i>Cystobacterineae</i> sp. SBCy048 ¹⁾	Cystobacterineae	-	-
28	<i>Cystobacterineae</i> sp. SBCy050 ¹⁾	Cystobacterineae	-	-
29	<i>Hyalangium minutum</i> DSM 14724 (Hm1)	Cystobacterineae	Cystobacteraceae	<i>Hyalangium</i>
30	<i>Hyalangium minutum</i> Hym3 (Hm2)	Cystobacterineae	Cystobacteraceae	<i>Hyalangium</i>
31	<i>Hyalangium minutum</i> NOCb10 (Hm3)	Cystobacterineae	Cystobacteraceae	<i>Hyalangium</i>
32	<i>Melittangium boletus</i> Meb2	Cystobacterineae	Cystobacteraceae	<i>Melittangium</i>
33	<i>Melittangium lichenicola</i> Mel 24	Cystobacterineae	Cystobacteraceae	<i>Melittangium</i>
34	<i>Stigmatella aurantiaca</i> DW4/3-1 (Sa1)	Cystobacterineae	Cystobacteraceae	<i>Stigmatella</i>
35	<i>Stigmatella aurantiaca</i> Sga15 (Sa2)	Cystobacterineae	Cystobacteraceae	<i>Stigmatella</i>
36	<i>Stigmatella aurantiaca</i> Sga32 (Sa3)	Cystobacterineae	Cystobacteraceae	<i>Stigmatella</i>
37	<i>Stigmatella erecta</i> Pde77(Se1)	Cystobacterineae	Cystobacteraceae	<i>Stigmatella</i>

38	<i>Vulgatibacter incomptus</i> DSM 27710	Cystobacterineae	Vulgatibacteraceae	<i>Vulgatibacter</i>
39	<i>Aggregicoccus edoensis</i> MCy10622	Cystobacterineae	Myxococcaceae	<i>Aggregicoccus</i>
40	<i>Coralloccoccus coralloides</i> Ccc1071	Cystobacterineae	Myxococcaceae	<i>Coralloccoccus</i>
41	<i>Coralloccoccus coralloides</i> DSM 2259	Cystobacterineae	Myxococcaceae	<i>Coralloccoccus</i>
42	<i>Coralloccoccus coralloides</i> ST201330	Cystobacterineae	Myxococcaceae	<i>Coralloccoccus</i>
43	<i>Coralloccoccus</i> sp. Ccc127	Cystobacterineae	Myxococcaceae	<i>Coralloccoccus</i>
44	<i>Coralloccoccus</i> sp. MCy10984	Cystobacterineae	Myxococcaceae	<i>Coralloccoccus</i>
45	<i>Myxococcus fulvus</i> 124B02	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
46	<i>Myxococcus fulvus</i> HW-1 (Mf1)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
47	<i>Myxococcus fulvus</i> Mxf50	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
48	<i>Myxococcus fulvus</i> Mxf65	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
49	<i>Myxococcus fulvus</i> SBMx122	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
50	<i>Myxococcus fulvus</i> SBMx132	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
51	<i>Myxococcus hansupus mixupus</i> (Mh1)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
52	<i>Myxococcus</i> sp. 171 (M1)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
53	<i>Myxococcus</i> sp. MCy10608	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
54	<i>Myxococcus stipitatus</i> DSM 14675	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
55	<i>Myxococcus virescens</i> ST200611 (Mv1)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
56	<i>Myxococcus xanthus</i> DK 1622 (Mx1)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
57	<i>Myxococcus xanthus</i> DK897 (Mx2)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
58	<i>Myxococcus xanthus</i> DZ2	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
59	<i>Myxococcus xanthus</i> DZF1	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
60	<i>Myxococcus xanthus</i> MxA47 (Mx3)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
61	<i>Myxococcus xanthus</i> Mxx48 (Mx4)	Cystobacterineae	Myxococcaceae	<i>Myxococcus</i>
62	<i>Pyxidicoccus fallax</i> And48	Cystobacterineae	Myxococcaceae	<i>Pyxidicoccus</i>
63	<i>Pyxidicoccus</i> sp. SBCy002	Cystobacterineae	Myxococcaceae	<i>Pyxidicoccus</i>
64	Unclassified sp. And30 ¹⁾	Cystobacterineae	-	-
65	Unclassified sp. Ang983 ¹⁾	Cystobacterineae	-	-
66	Unclassified sp. CcG34¹⁾ (Cy1)	Cystobacterineae	-	-
67	Unclassified sp. SBCy006 ¹⁾	Cystobacterineae	-	-
68	Unclassified sp. SBMx152 ¹⁾	Cystobacterineae	-	-
69	<i>Aetherobacter fasciculatus</i> SBSr002	Sorangiineae	Polyangiaceae	<i>Aetherobacter</i>
70	<i>Aetherobacter rufus</i> SBSr003	Sorangiineae	Polyangiaceae	<i>Aetherobacter</i>
71	<i>Aetherobacter</i> sp. SBSr001	Sorangiineae	Polyangiaceae	<i>Aetherobacter</i>
72	<i>Aetherobacter</i> sp. SBSr008	Sorangiineae	Polyangiaceae	<i>Aetherobacter</i>
73	<i>Byssovorax cruenta</i> Byc1	Sorangiineae	Polyangiaceae	<i>Byssovorax</i>
74	<i>Chondromyces apiculatus</i> DSM 436	Sorangiineae	Polyangiaceae	<i>Chondromyces</i>
75	<i>Chondromyces catenulatus</i> SBCm007	Sorangiineae	Polyangiaceae	<i>Chondromyces</i>
76	<i>Chondromyces crocatus</i> Cm c5	Sorangiineae	Polyangiaceae	<i>Chondromyces</i>
77	<i>Chondromyces pediculatus</i> Cmp5	Sorangiineae	Polyangiaceae	<i>Chondromyces</i>
78	<i>Jahnella</i> sp. SBSr007	Sorangiineae	Polyangiaceae	<i>Jahnella</i>
79	<i>Polyangium spumosum</i> Plsm9	Sorangiineae	Polyangiaceae	<i>Polyangium</i>
80	<i>Sorangiineae</i> sp. SBSr022	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
81	<i>Sorangium cellulosum</i> SBSr026	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
82	<i>Sorangium cellulosum</i> So ce56	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
83	<i>Sorangium cellulosum</i> So0157-2	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
84	<i>Sorangium cellulosum</i> Soce10	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
85	<i>Sorangium cellulosum</i> Soce1128	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
86	<i>Sorangium cellulosum</i> Soce1525	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
87	<i>Sorangium cellulosum</i> Soce1875	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
88	<i>Sorangium cellulosum</i> Soce26	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
89	<i>Sorangium cellulosum</i> Soce307	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
90	<i>Sorangium cellulosum</i> Soce340	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
91	<i>Sorangium cellulosum</i> Soce377	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
92	<i>Sorangium cellulosum</i> Soce38	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
93	<i>Sorangium cellulosum</i> Soce439	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
94	<i>Sorangium cellulosum</i> Soce690	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
95	<i>Sorangium cellulosum</i> Soce836	Sorangiineae	Polyangiaceae	<i>Sorangium</i>
96	<i>Sorangium cellulosum</i> Soce960	Sorangiineae	Polyangiaceae	<i>Sorangium</i>

97	<i>Sorangium cellulosum</i> Soce969	<i>Sorangiineae</i>	<i>Polyangiaceae</i>	<i>Sorangium</i>
98	<i>Sorangium cellulosum</i> SoceGT47	<i>Sorangiineae</i>	<i>Polyangiaceae</i>	<i>Sorangium</i>
99	<i>Sorangium nigrum</i> Soce487	<i>Sorangiineae</i>	<i>Polyangiaceae</i>	<i>Sorangium</i>
100	<i>Sandaracinus amylolyticus</i> DSM 53668	<i>Sorangiineae</i>	<i>Sandaracinaceae</i>	<i>Sandaracinus</i>
101	<i>Sandaracinus</i> sp. SBSa001	<i>Sorangiineae</i>	<i>Sandaracinaceae</i>	<i>Sandaracinus</i>
102	<i>Labilithrix luteola</i> strain DSM 27648	<i>Sorangiineae</i>	<i>Labilitrichaceae</i>	<i>Labilithrix</i>
103	<i>Minicystis</i> sp. SBNa008	<i>Sorangiineae</i>	-	-
104	Unclassified sp. MSr10681 ²⁾	<i>Sorangiineae</i>	-	-
105	Unclassified sp. MSr3139 ²⁾	<i>Sorangiineae</i>	-	-
106	Unclassified sp. NOSO-3 ¹⁾	<i>Sorangiineae</i>	-	-
107	Unclassified sp. SBSr015 ²⁾	<i>Sorangiineae</i>	-	-
108	Unclassified sp. SBSr017 ²⁾	<i>Sorangiineae</i>	-	-
109	Unclassified sp. SBSr044 ²⁾	<i>Sorangiineae</i>	-	-
110	Unclassified sp. SBSr060 ²⁾	<i>Sorangiineae</i>	-	-
111	Unclassified sp. SBSr071 ²⁾	<i>Sorangiineae</i>	-	-
112	<i>Nannocystineae</i> sp. SBNc001	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	
113	<i>Nannocystis exedens</i> Nae485	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	<i>Nannocystis</i>
114	<i>Nannocystis exedens</i> Nae487	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	<i>Nannocystis</i>
115	<i>Enhygromyxa salina</i> DSM 15201	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	<i>Enhygromyxa</i>
116	<i>Enhygromyxa salina</i> SBCm009	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	<i>Enhygromyxa</i>
117	<i>Haliangium ochraceum</i> DSM 14365	<i>Nannocystineae</i>	<i>Haliangiaceae</i>	<i>Haliangium</i>
118	<i>Haliangium ochraceum</i> GS1	<i>Nannocystineae</i>	<i>Haliangiaceae</i>	<i>Haliangium</i>
119	<i>Kofleria flava</i> Plvt1	<i>Nannocystineae</i>	<i>Kofleriaceae</i>	<i>Kofleria</i>
120	<i>Plesiocystis pacifica</i> SIR-1	<i>Nannocystineae</i>	<i>Nannocystaceae</i>	<i>Plesiocystis</i>
121	Unclassified sp. Nc005 ³⁾	<i>Nannocystineae</i>	-	-
122	Unclassified sp. Nc018 ³⁾	<i>Nannocystineae</i>	-	-

1) unclassified species belonging to the suborder *Cystobacterineae*

2) unclassified species belonging to the suborder *Sorangiineae*

3) unclassified species belonging to the suborder *Nannocystineae*

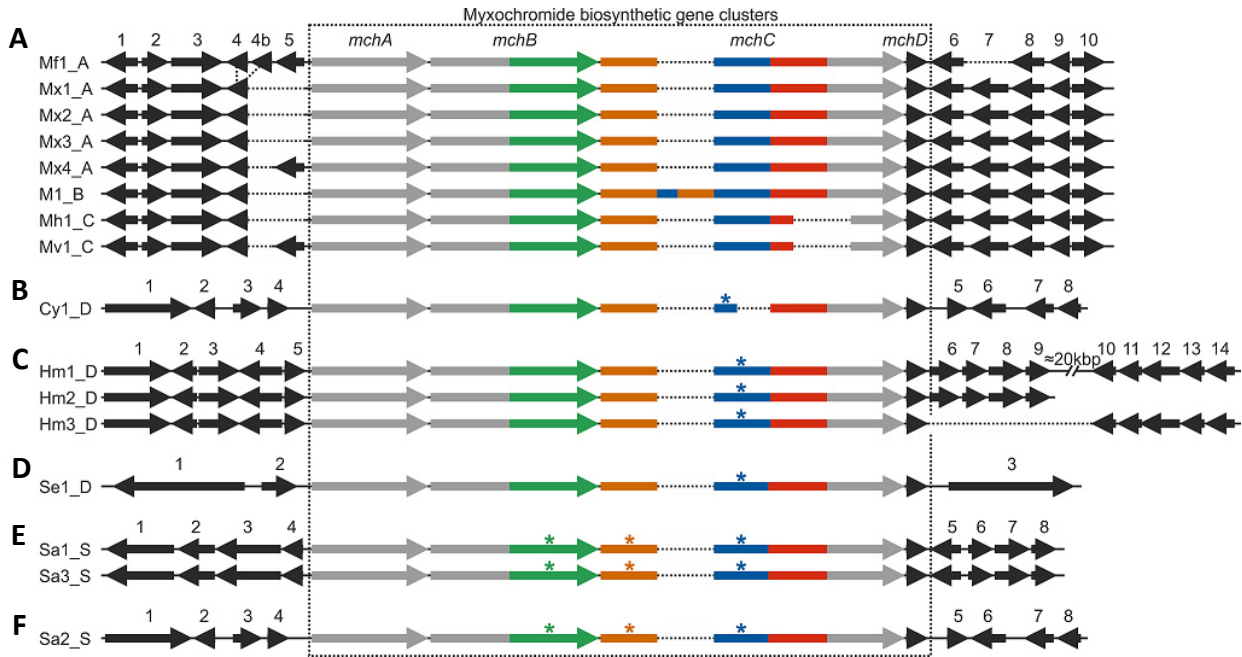


Figure S2. Myxochromide biosynthetic gene cluster loci from the 16 analyzed myxobacterial producer strains. The (putative) *mch* gene clusters (highlighted by a dashed rectangle) represent four-gene operons (*mchA-D*) encoding an iterative polyketide synthase (MchA), two multimodular nonribosomal peptide synthetases (MchB and MchC) and a conserved hypothetical protein (MchD). **A:** A-type *mch* gene clusters from *Myxococcus fulvus* HW-1 (Mf1_A), *M. xanthus* DK1622 (Mx1_A), *M. xanthus* DK897 (Mx2_A), *M. xanthus* A47 (Mx3_A) and *M. xanthus* Mx48 (Mx4_A), B-type *mch* gene cluster from *Myxococcus* sp. 171 (M1_B) and C-type *mch* gene cluster from *M. hansupus* mixupus (Mh1_C) and *M. virescens* ST200611 (Mv1_C). Minor differences in the flanking chromosomal regions can be observed (4b, 5, 7), which encode: glycosyl hydrolase (1), multidrug ABC transporter ATP-binding protein (2), membrane protein (3), uncharacterized protein (4/4b), AraC family transcriptional regulator (5), ATP-dependent RNA helicase RhIE (6), elongation factor G (7), cyclic nucleotide-binding protein (8), peptidyl-prolyl cis-trans isomerase (9) and metallophosphatase (10). **B:** D-type *mch* gene cluster from *Cystobacterineae* sp. CcG34 (Cy1_D). Flanking regions encode: two-component sensor histidine kinase (1), membrane protein (2), uncharacterized protein (3), uncharacterized protein (4), uncharacterized protein (5), aldo/keto reductase (6), uncharacterized protein (7) and uncharacterized protein (8). **C:** D-type *mch* gene clusters from *Hyalangium minutum* DSM14724 (Hm1_D), *H. minutum* Hym-3 (Hm2_D), *H. minutum* NoCb10 (Hm3_D). Flanking regions encode: uncharacterized protein (1), uncharacterized protein (2), pilus assembly protein PilZ (3), amino acid dehydrogenase (4), tRNA-Thr (5), uncharacterized protein (6), ABC transporter ATP-binding protein (7), ABC transporter permease (8), uncharacterized protein (9), benzoate-CoA ligase family protein (10), 8-amino-7-oxononanoate synthase (11), phosphopantetheine-binding protein (12), acyl-CoA dehydrogenase (13) and acyl-CoA synthetase (14). The downstream region 6-9 identified in Hm1_D/Hm2_D plus additional ~20 kb are missing from cluster downstream DNA in Hm3_D. **D:** D-type *mch* gene cluster from *Stigmatella erecta* Pde77 (Se1_D). Flanking regions encode putative homologous of the dawenol (Daw) and myxalamid (Mxa) biosynthetic pathways,^{48,49} polyketide synthases homologous to Daw3 (1), Daw2 (2) and MxaF (3). **E:** S-type *mch* gene clusters from *S. aurantiaca* DW4/3-1 (Sa1_S) and *S. aurantiaca* Sga32 (Sa3_S). Flanking regions encode: cation/H(+) antiporter (1), hydrolase (2), ABC transporter permease (3), ABC transporter ATP-binding protein (4), alpha/beta hydrolase (5), uncharacterized protein (6), coproporphyrinogen III oxidase (7) and uncharacterized protein (8). **F:** S-type *mch* gene cluster from *S. aurantiaca* Sga15 (Sa2_S). Flanking regions encode: uncharacterized protein (1), TetR family transcriptional regulator (2), 3-oxoadipate enol-lactonase (3), PHB depolymerase esterase (4), glutathione S-transferase (5), uncharacterized protein (6), protein kinase (7) and protein kinase (8).

2.7.2.2 *In-silico* Prediction of Products of the C-type and D-type Megasyntetases

In the course of the genome mining approach described in chapter 2.7.2.1, 7 putative *mch* gene clusters covering 4 novel types (B-type, C-type, D-subtype1, D-subtype2) were identified, which deviate from the previously described myxochromide A and S pathways in terms of the modular organization of the respective assembly lines.^{7,9}

In the putative C-type *mch* gene clusters from Mh1 and Mv1, the (compared to the A-type *mch* gene cluster) corresponding A₅-PCP₅-C₆ unit is deleted, which led to recombination of the ‘hybrid’ C₅-A₆-PCP₆-TE₆ module. *In-silico* prediction of the substrate specificities of all adenylation domains present in the putative C-type *mch* gene clusters (see chapter 2.7.7) revealed the same amino acid specificities as already predicted and experimentally verified for myxochromides A.⁷ As a result, it was expected that myxochromides C might have the same peptide core structure as myxochromides A but lack the alanine residue between the proline and glutamine residues. Since there are only minor differences in the protein sequences of the epimerization domains of module 2 from A-type and putative C-type *mch* gene clusters, it was assumed that the alanine residue, that might be introduced by module 2, has *D* configuration in the final product, which was already shown for myxochromides A.⁷ Theoretical masses of proposed myxochromides C₁-C₄ were predicted, for which extracts of Mv1 (strain Mh1 was not available) were screened by HPLC-MS analysis (see chapter 2.7.2.3).

Sequence analysis of the putative D-type *mch* gene clusters (subtype 1) from Hm1, Hm2, Hm3 and Se1 revealed the same critical mutation within the core motif of the peptidyl carrier protein of module 4, which was already described for myxochromide S biosynthesis (see chapter 2.7.7).⁷ Here, this loss of function mutation led to ‘module-skipping’ resulting in the production of a pentapeptide core. The same scenario was now postulated for the putative D-subtype 1 pathways, which were predicted to generate pentapeptide cores lacking the proline residue compared to myxochromides A. Similar products were expected from the putative D-subtype 2 pathway from Cy1, which actually lacks two catalytic domains (A₄-PCP₄) obviously also resulting in a non-functional module 4. Specificity predictions of the adenylation domains from both subtypes indicate that both putative myxochromide D assembly lines might recruit and incorporate the same amino acid residues as previously described for the myxochromide A biosynthetic machinery (see chapter 2.7.7).⁷ In comparison to myxochromides S, the amino acid specificities of modules 2 and 3 might be interchanged (see chapter 2.7.7) leading to the production of a novel pentapeptide. Based on the high protein sequence similarity between the epimerization domains from A-type and putative D-type pathways, the absolute configuration of myxochromides D might be the same

as described for myxochromides A with one alanine residue (incorporated by module 2) showing *D* configuration. Theoretical masses of proposed myxochromides D₁-D₄ were predicted, for which extracts of Hm1, Hm2, Hm3, Se1 and Cy1 were screened by HPLC-MS analysis (see chapter 2.7.2.3).

2.7.2.3 Analysis of Myxochromide Production Spectra

To verify myxochromide production in the (putative) producer strains identified in the genome-based screening (Table S1), the strains (except strain Mh1, which was not available) were routinely grown in 50 mL liquid medium (300 mL shaking flasks) including 2% of amberlite XAD-16 adsorber resin at 30 °C and 180 rpm for up to 7 days. The strains Mf1, Mx1, Mx2, Mx3 and Mx4 were grown in CTT medium (casitone 1%, Tris-HCl [pH 8.0] 10 mM, K₂HPO₄/KH₂PO₄ buffer [pH 7.6] 1 mM, MgSO₄ × 7 H₂O 8 mM, pH adjusted to 7.6), strains Sa1, Sa2 and Sa3 were grown in tryptone medium (tryptone 1%, MgSO₄ × 7 H₂O 0.2%, pH adjusted to 7.2), strains Hm1, Hm2 and Hm3 were grown in MD1G medium (casitone 0.3%, CaCl₂ × 2 H₂O 0.05%, MgSO₄ × 7 H₂O 0.1%, glucose 0.35%, HEPES 0.11%, pH adjusted to 7.0), strain Se1 was grown in VY/2 medium (Baker's yeast 0.5%, CaCl₂ × 2 H₂O, HEPES 0.11%, vitamin B₁₂ 0.5 µg/mL, pH adjusted to 7.0), strain Cy1 was grown in H medium (soy flour 0.2%, yeast extract 0.2%, glucose 0.2%, starch 0.8%, CaCl₂ × 2 H₂O 0.1%, MgSO₄ × 7 H₂O 0.1%, HEPES 0.11%, pH adjusted to 7.4), strain Mv1 was grown in CY medium (casitone 0.3%, yeast extract 0.1%, CaCl₂ × 2 H₂O 0.1%, MgSO₄ × 7 H₂O 0.2%, vitamin B₁₂ 0.5 µg/mL, pH adjusted to 7.4) and strain M1 was grown in a peptone medium (casitone 0.3%, yeast extract 0.3%, CaCl₂ × 2 H₂O 0.07%, MgSO₄ × 7 H₂O 0.2%, glucose 0.1%, starch 1%, pH adjusted to 7.2).

Cells and XAD-16 were harvested by centrifugation at 8,000 rpm and 4 °C for 10 min and subsequently extracted twice with 50 mL of a mixture of methanol and acetone (1:1). The extracts were evaporated to dryness, dissolved in methanol and subjected to HPLC-MS analysis using a Dionex Ultimate 3000 RSLC system coupled to a Bruker maXis 4G TOF mass spectrometer. Separation was performed using a Waters BEH C18, 100 × 2.1 mm, 1.7 µm d_p column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0-0.5 min 5% B, 0.5-18.5 min 5-95% B, 18.5-20.5 min 95% B. Full scan mass spectra were acquired in positive ESI mode in a range from 150-2500 *m/z*. All MS² experiments of target masses were performed on a Bruker solariX XR FT-ICR (7T) mass spectrometer. Initially, precursors were isolated in the quadrupole with an isolation width of 5 Da and subsequently fragmented in the collision cell by applying an energy of 20 eV. Occurred fragments were then transferred to the

ICR cell and detected in a mass range from 100-1600 m/z by acquiring a 490 ms transient. The chromatograms and selected MS² fragment spectra obtained from the HPLC-MS analysis are illustrated in Figures S3/S4. A peak list, in which identified MS² fragments are summarized, is shown in Table S2. Myxochromides with different polyunsaturated acyl side chains can be distinguished by MS² fragmentation yielding a highly characteristic *N*-Methreonine-acyl chain fragment as most prominent ion (Figure S4).^{7,9}

All 15 analyzed (putative) producer strains were shown to produce specific types of myxochromides as illustrated in Figure S3 and summarized below.

Myxochromides A

The previously described lipohexapeptides myxochromides A₂, A₃ and A₄ (**1a-c**⁷; for MS² spectra see Figure S4A) were detected in extracts of *Myxococcus* strains Mf1, Mx1, Mx2, Mx3 and Mx4. The most prominent derivative was in most cases **1b**, in case of strain Mx2 **1a** under the applied cultivation conditions (Figure S3). Thus, all *M. xanthus* species analyzed in this study were verified as A-type producers, which correlates with a previous secondary metabolome analysis of 98 *M. xanthus* strains revealing myxochromides A to be ubiquitous metabolites in this species.²⁴

Myxochromides B

The lipopeptapeptides myxochromides B₂, B₃ and B₄ (**2a-c**; for MS² spectra see Figure S4B) were detected in extracts of strain M1. Whereas production and structure elucidation of the most prominent derivative **2b** was previously reported,⁸ two additional minor derivatives **2a** and **2c** were identified in this study (Figure S3). Based on the analytical HPLC-MS data including high-resolution MS² spectra, **2a** and **2c** were assigned as myxochromides B₂ and B₄.

Myxochromides C – novel lipopentapeptides

From the putative myxochromide C producers assigned based on genomic data (Mv1 and Mh1, Table 1), only strain Mv1 was available for production analysis. As discussed in chapter 2.7.2.2, sequence analysis of the encoded putative myxochromide megasynthetase allowed for structure prediction of potential biosynthesis products designated as myxochromides C. Respective myxochromide derivatives with [M+H]⁺ masses and MS² spectra corresponding with the expected products were detected in the Mv1 extract (**6a-c**; Figures S3 and S4C). Isolation and structure elucidation of the major derivative **6b** revealed that myxochromides C indeed represent novel lipopentapeptides (chapter 2.7.3). In addition to myxochromide C₃ (**6b**), two minor derivatives (**6a** and **6c**) were detected and designated as myxochromides C₂ and C₄ based on the corresponding MS² spectra.

Myxochromides D – novel lipopentapeptides

Based on the genome data, five strains were expected to produce another novel lipopentapeptide type designated as myxochromides D (Cy1, Hm1, Hm2, Hm3 and Se1; Table 1). As discussed in chapter 2.7.2.2, sequence analysis of the encoded putative myxochromide megasynthetase allowed for structure prediction of potential novel D-type biosynthesis products. Respective myxochromide derivatives with $[M+H]^+$ masses and MS² spectra corresponding with the expected products were detected in extracts of the five analyzed producers (**7a-d**; Figures S3 and S4D). Isolation and structure elucidation of the major derivative **7a** from strain Se1 revealed that myxochromides D indeed represent novel lipopentapeptides (chapter 2.7.4). Additionally to myxochromide D₁ (**7a**), three other derivatives (**7b-d**) were detected and designated as myxochromides D₂₋₄ based on the corresponding MS² spectra. In contrast to strain Se1, strains Cy1 and Hm1-3 were shown to produce **7b** and **7c** as major myxochromide derivatives.

Myxochromides S

The previously described lipopentapeptides myxochromides S₁, S₂ and S₃ (**3a-c**⁹; for MS² spectra see Figure S4E) were detected in extracts of *Stigmatella aurantiaca* strains Sa1-3 with **3a** as major derivative under the applied cultivation conditions (Figure S3). Myxochromides S and D feature the same amino acid composition (but different amino acid sequence) and can be distinguished due to different retention times and MS² spectra.

Myxochromides S-Abu and S-diAbu – novel lipopentapeptides from S-type producers

Prior to this study, myxochromide producers were described to generate single myxochromide peptide cores (A-type, B-type or S-type⁷⁻⁹). Interestingly, our analysis of S-type producer strains revealed additional biosynthesis products (Figure S3), which according to the analytical data, seem to represent novel lipopentapeptides designated as myxochromides S-Abu (**4a-c**; for MS² spectra see Figure S4F) and myxochromides S-diAbu (**5a-c**; for MS² spectra see Figure S4G). A representative of each type (S₂-Abu and S₂-diAbu) was isolated from a previously described heterologous high-titer myxochromide S production strain (chapter 2.7.5 and 2.7.6)¹⁵. Structure elucidation revealed that the novel pentapeptide cores differ from myxochromides S by replacement of one or two *L*-alanine (*L*-Ala) residues with *L*-α-aminobutyric acid (*L*-Abu). This result correlates with the observation that production of **4** and **5** can be increased when supplementing the production cultures with *L*-Abu (see below).

Novel lipopeptides detected in A-,B-,C- and D-type producers after L-Abu feeding

The production of additional Abu-containing myxochromides S-Abu and S-diAbu in S-type producer strains indicated that Ala-specific modules from the S-type assembly line can

alternatively incorporate Abu. To analyze whether similar variations can be observed for other (putative) myxochromide megasynthetase types as well, Abu feeding studies were performed with a representative producer strain for each myxochromide type: Mx1, M1, Mv1, Se1 and Sa1. The strains were routinely grown in duplicates under the same conditions as described above (XAD-16 was supplemented 6 h before cell harvest). One duplicate was supplemented with 4.2 mg *L*- α -aminobutyric acid three times per day for up to 7 days. Cell harvest, extract preparation and HPLC-MS-MS analysis was performed as described above. As illustrated in Figure S5, myxochromide production profiles from each of the analyzed strains clearly change when supplementing the cultures with *L*-Abu which indicates the production of novel Abu-containing myxochromides as already shown for Sa1 (myxochromides S-Abu and S-diAbu). In case of A-, B and D-type producers, the detected novel derivatives (compounds **8**, **9**, **10**, **11**, **13**, **14**) revealed mass shifts in $[M+H]^+$ corresponding to one ($+CH_2$, $m/z +14$) or two ($+C_2H_4$, $m/z +28$) additional methylene groups (Table S3). This points to the production of modified peptide cores, which contain either one or two Abu instead of Ala residues. In case of the C-type producer, only derivatives with mass shifts corresponding to one additional methylene group were detected (**12**; Figure S5). This was expected as the myxochromide C peptide core harbors only one Ala residue, which could be substituted with Abu (Table S3). Our data suggest that Ala-incorporating modules from all types of myxochromide megasynthetase can make use of *L*-Abu as alternative substrate. Feeding production cultures with *L*-Abu directed the biosynthesis towards novel derivatives (**8-14**) and further expanded the structural diversity of myxochromide peptide cores.

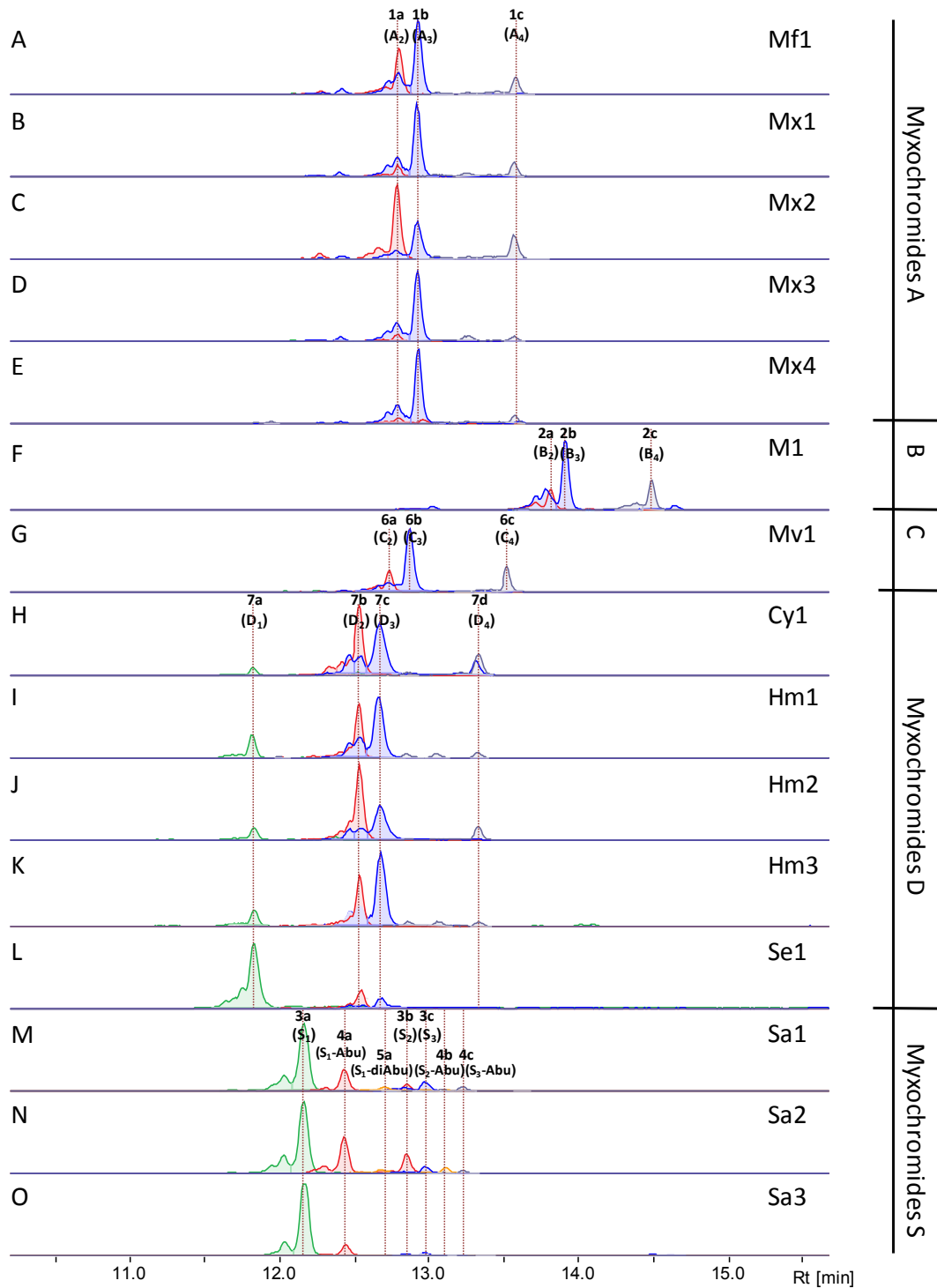


Figure S3. HPLC-MS analysis of myxochromide production in 15 myxobacterial producer strains (strain Mh1 was not available). Extracted ion chromatograms (EICs) for ± 0.02 m/z corresponding to the $[M+H]^+$ ions of myxochromides are shown. **A-E:** Detection of myxochromides A in Mf1, Mx1, Mx2, Mx3 and Mx4; A₂ ($[M+H]^+ = 834.47600$), A₃ ($[M+H]^+ = 846.47600$) and A₄ ($[M+H]^+ = 860.49165$). **F:** Detection of myxochromides B in M1; B₂ ($[M+H]^+ = 947.56007$), B₃ ($[M+H]^+ = 959.56007$), B₄ ($[M+H]^+ = 973.57572$). **G:** Detection of myxochromides C in Mv1; C₂ ($[M+H]^+ = 763.43889$), C₃ ($[M+H]^+ = 775.43889$), C₄ ($[M+H]^+ = 789.45454$). **H-L:** Detection of myxochromides D in Cy1, Hm1, Hm2, Hm3 and Se1; D₁ ($[M+H]^+ = 723.40759$), D₂ ($[M+H]^+ = 737.42324$), D₃ ($[M+H]^+ = 749.42324$), D₄ ($[M+H]^+ = 763.43889$). **M-O:** Detection of myxochromides S, S-Abu and S-diAbu in Sa1, Sa2 and Sa3; S₁ ($[M+H]^+ = 723.40759$), S₂+S₁-Abu ($[M+H]^+ = 737.42324$), S₂-Abu+S₁-diAbu ($[M+H]^+ = 751.43889$), S₃ ($[M+H]^+ = 749.42324$), S₃-Abu ($[M+H]^+ = 763.43889$).

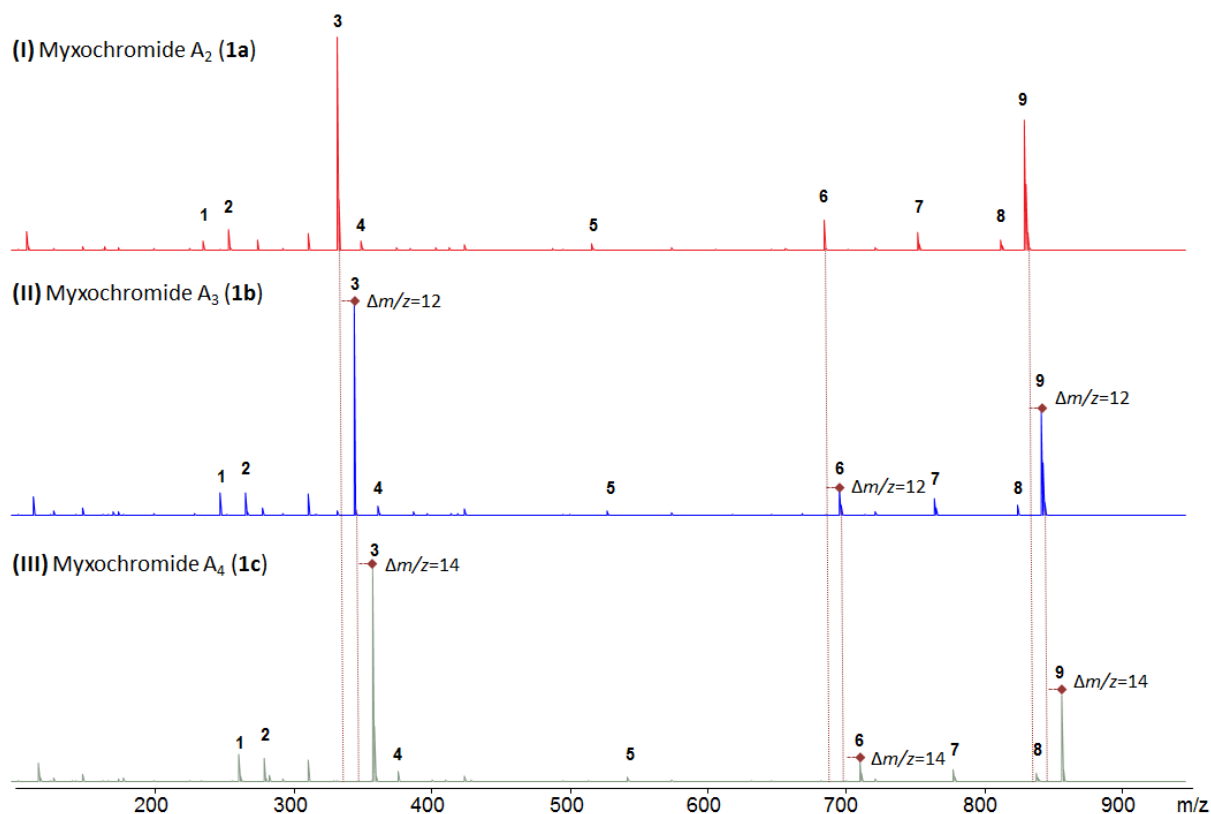
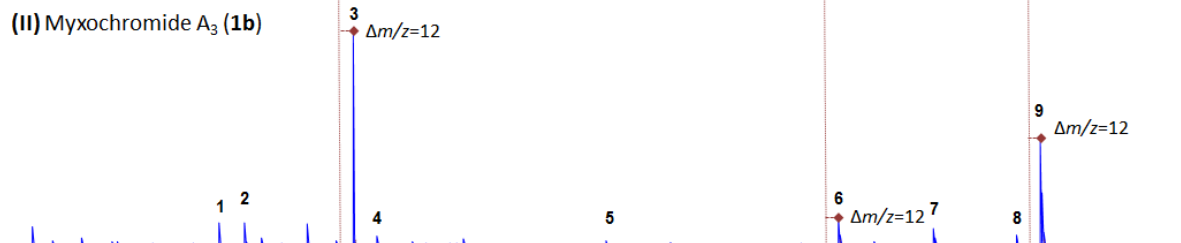
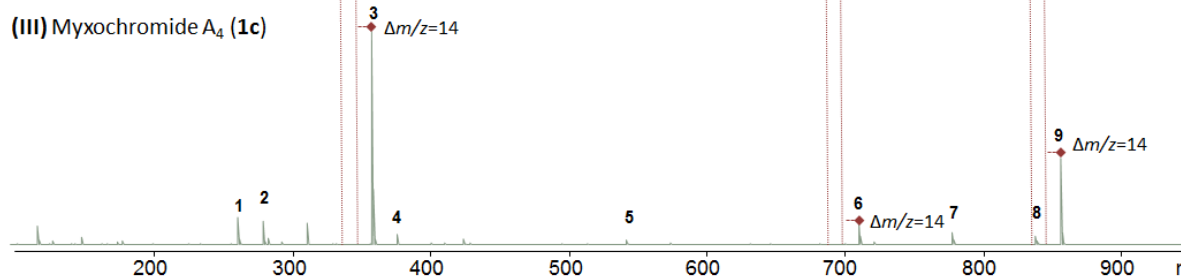
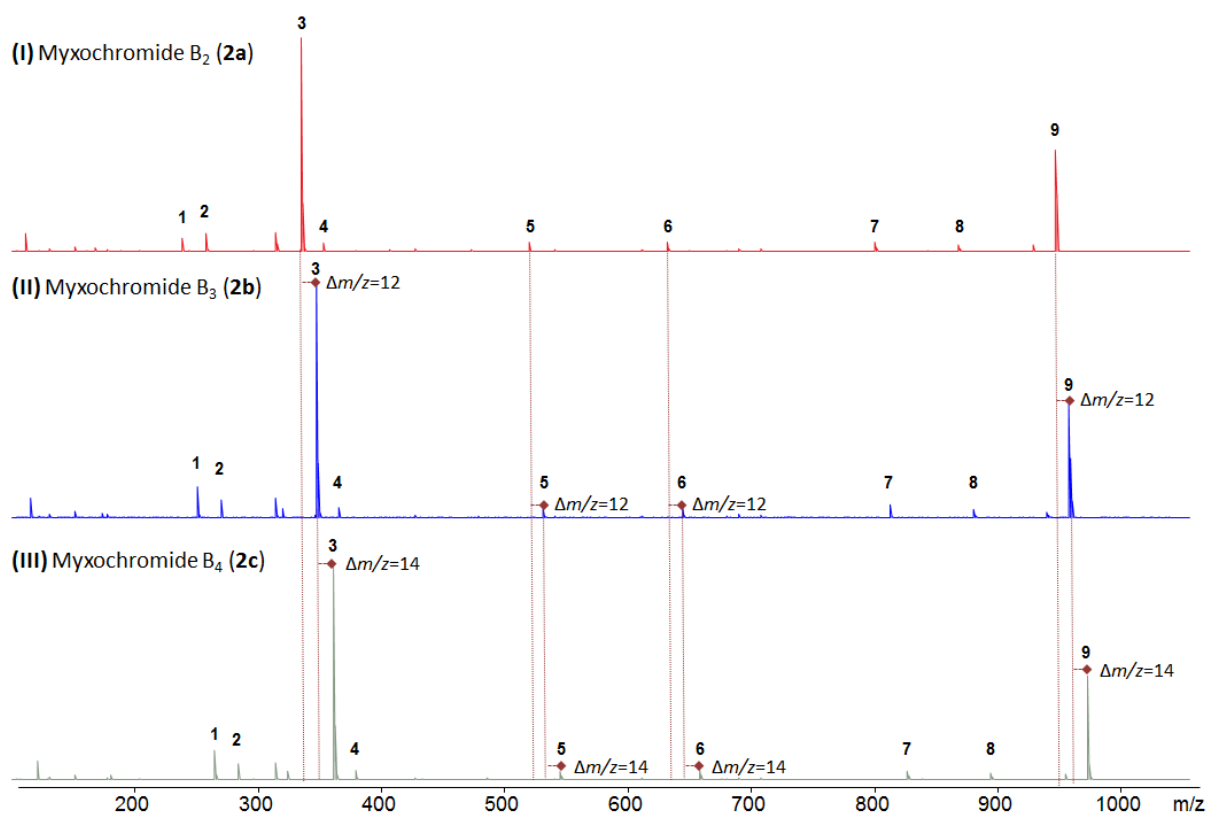
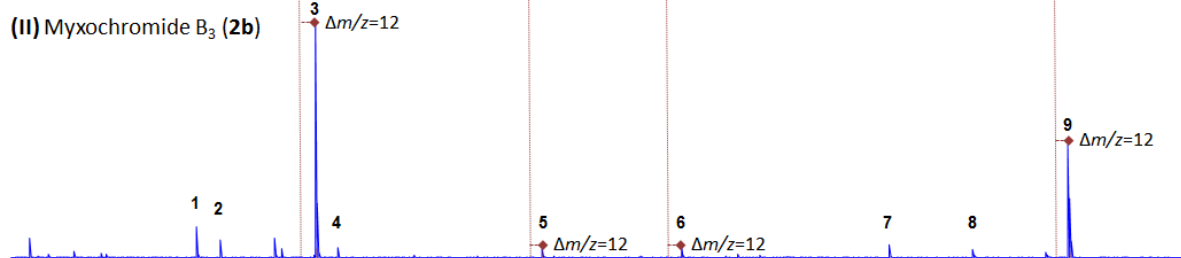
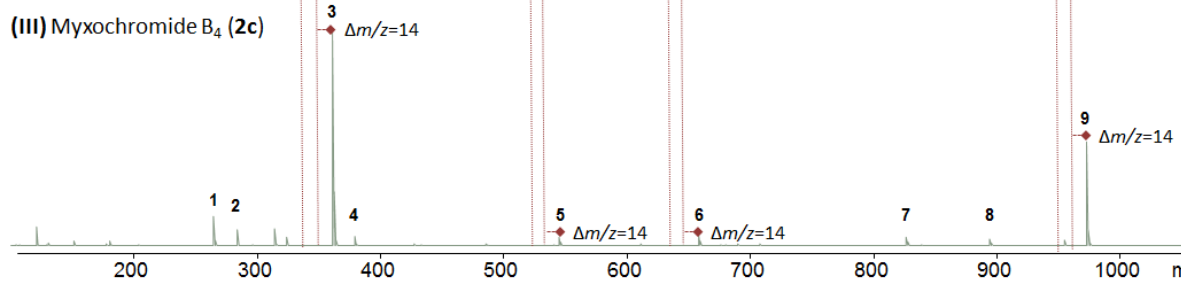
A**(I) Myxochromide A₂ (1a)****(II) Myxochromide A₃ (1b)****(III) Myxochromide A₄ (1c)****B****(I) Myxochromide B₂ (2a)****(II) Myxochromide B₃ (2b)****(III) Myxochromide B₄ (2c)**

Figure S4 (continued on next page)

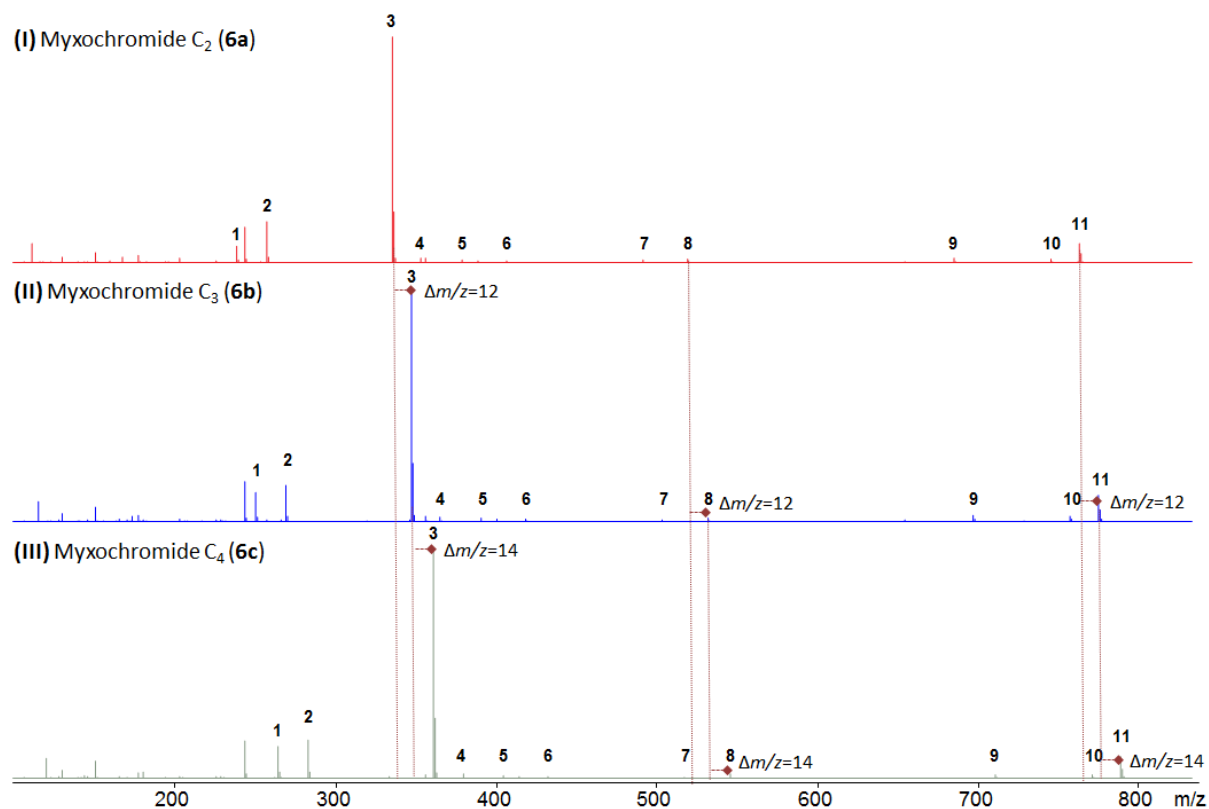
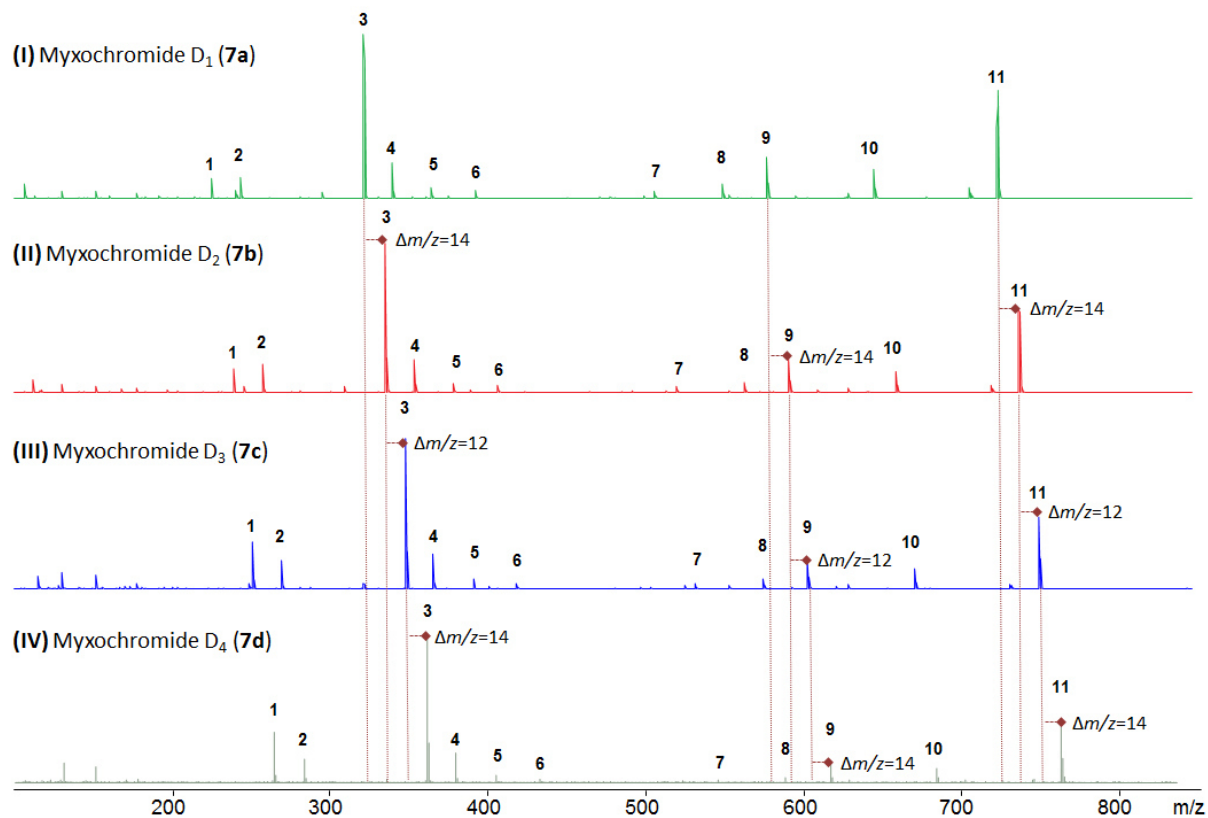
C**D**

Figure S4 (continued on next page)

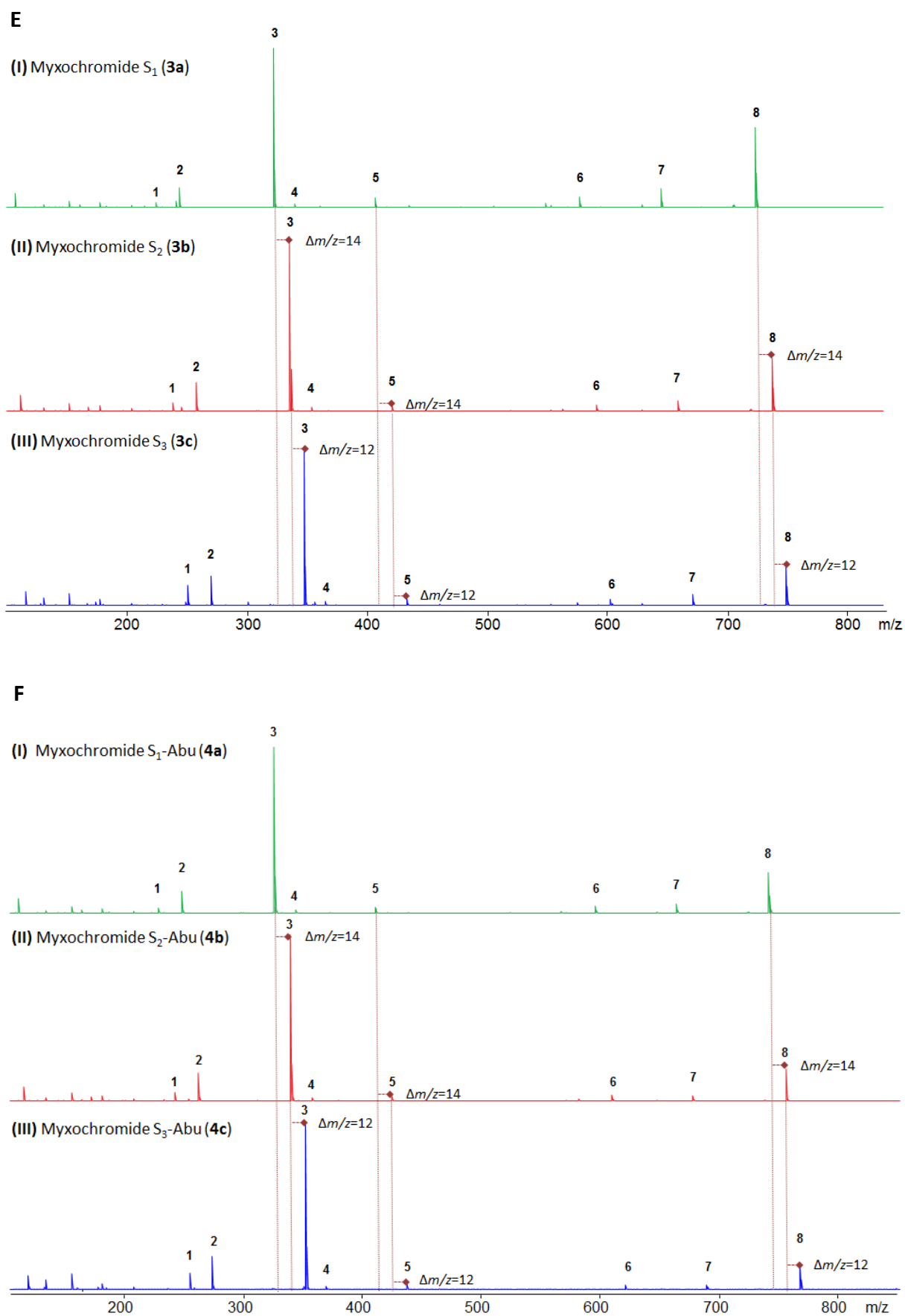


Figure S4 (continued on next page)

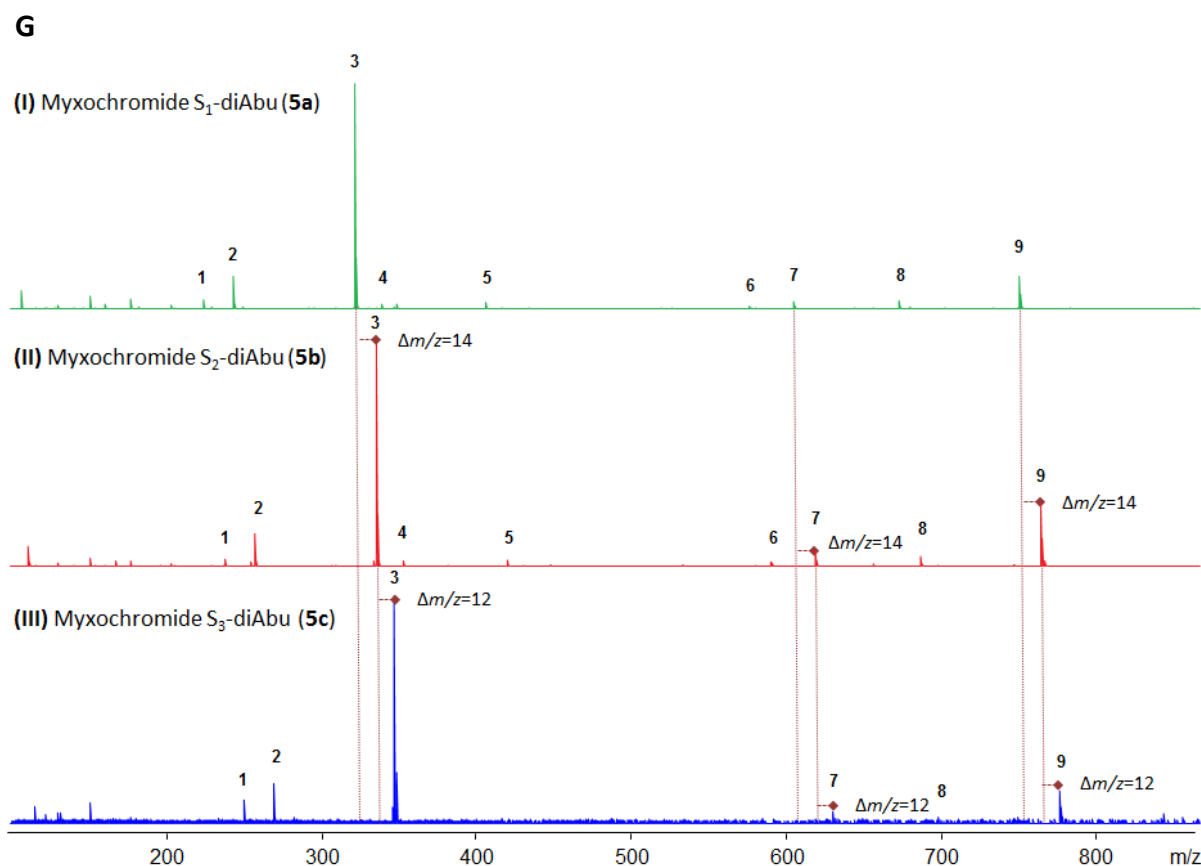


Figure S4. Annotated FT-ICR MS² spectra of different myxochromide derivatives (for m/z values see peak list table below). Full scan mass spectra were acquired in positive ESI mode in a range from 150-2500 m/z . **A:** Myxochromides A from an extract of strain Mx2. (I) MS² of m/z 834.47647 identified as myxochromide A₂. (II) MS² of m/z 846.47327 identified as myxochromide A₃. (III) MS² of m/z 860.48891 identified as myxochromide A₄. **B:** Myxochromides B from an extract of strain M1. (I) MS² of m/z 947.55939 identified as myxochromide B₂. (II) MS² of m/z 959.55927 identified as myxochromide B₃. (III) MS² of m/z 973.57529 identified as myxochromide B₄. **C:** Myxochromides C from an extract of strain Mv1. (I) MS² of m/z 763.43784 identified as myxochromide C₂. (II) MS² of m/z 775.43855 identified as myxochromide C₃. (III) MS² of m/z 789.45341 identified as myxochromide C₄. **D:** Myxochromides D from extracts of strains Se1 and Hm2. (I) MS² of m/z 723.40758 identified as myxochromide D₁ (Se1 extract). (II) MS² of m/z 737.42373 identified as myxochromide D₂ (Hm2 extract). (III) MS² of m/z 749.42344 identified as myxochromide D₃ (Hm2 extract). (IV) MS² of m/z 763.43957 identified as myxochromide D₄ (Hm2 extract). **E:** Myxochromides S from an extract of strain Sa1. (I) MS² of m/z 723.40774 identified as myxochromide S₁. (II) MS² of m/z 737.42319 identified as myxochromide S₂. (III) MS² of m/z 749.42391 identified as myxochromide S₃. **F:** Myxochromides S-Abu from extracts of strains Sa1 and Sa2. (I) MS² of m/z 737.42350 identified as myxochromide S₁-Abu (Sa1 extract). (II) MS² of m/z 751.43762 from myxochromide S₂-Abu (Sa2 extract). (III) MS² of m/z 763.43973 identified as myxochromide S₃-Abu (Sa1 extract). **G:** Myxochromides S-diAbu from an extract of strain Sa1 and authentic myxochromide S₂-diAbu reference substance. (I) MS² of m/z 737.42350 identified as myxochromide S₁-diAbu (Sa1 extract). (II) MS² of m/z 751.43762 from myxochromide S₂-diAbu (authentic reference). (III) MS² of m/z 763.43973 identified as myxochromide S₃-diAbu (Sa1, *L*- α -aminobutyric acid feeding).

Table S2. MS² fragments and retention times of different myxochromide derivatives. The *m/z* values correspond to the fragments, which are assigned in Figure S4. The characteristic threonine-polyketide fragments are highlighted in boldface.

Peak no.	-	Myxochromide A ₂ (12.79 min)	Myxochromide A ₃ (12.93 min)	Myxochromide A ₄ (13.57 min)
1	-	239.14381	251.14349	265.15923
2	-	258.14965	270.14935	284.16498
3	-	336.19667	348.19623	362.21196
4	-	354.20745	366.20673	380.22245
5	-	520.31822	532.31681	546.33271
6	-	688.40764	700.40549	714.42130
7	-	756.43011	768.42712	782.44278
8	-	816.46591	828.46263	842.47881
9	-	834.47647	846.47327	860.48891
Peak no.	-	Myxochromide B ₂ (13.81 min)	Myxochromide B ₃ (13.95 min)	Myxochromide B ₄ (14.49 min)
1	-	239.14373	251.14368	265.15948
2	-	258.14956	270.14948	284.16525
3	-	336.19659	348.19647	362.21227
4	-	354.20719	366.20709	380.22293
5	-	520.31746	532.31760	546.33374
6	-	633.40192	645.40167	659.41719
7	-	801.49081	813.49011	827.50707
8	-	869.51339	881.51228	895.52855
9	-	947.55939	959.55927	973.57529
Peak no.	-	Myxochromide C ₂ (12.72 min)	Myxochromide C ₃ (12.84 min)	Myxochromide C ₄ (13.49 min)
1	-	239.14404	251.14419	265.15975
2	-	258.14991	270.15003	284.16562
3	-	336.19692	348.19699	362.21250
4	-	354.20753	366.20768	380.22318
5	-	389.22336	401.22370	415.23896
6	-	407.23391	419.23411	433.24940
7	-	492.32269	504.32307	518.33814
8	-	520.31753	532.31795	546.33333
9	-	685.39215	697.39174	711.40673
10	-	745.42777	757.42799	771.44344
11	-	763.43784	775.43855	789.45341
Peak no.	Myxochromide D ₁ (11.83 min)	Myxochromide D ₂ (12.53 min)	Myxochromide D ₃ (12.69 min)	Myxochromide D ₄ (13.33 min)
1	225.12799	239.14373	251.14371	265.15994
2	244.13387	258.14958	270.14952	284.16579
3	322.18091	336.19660	348.19653	362.21288
4	340.19146	354.20718	366.20719	380.22360
5	365.22306	379.23876	391.23872	405.25525
6	393.21812	407.23388	419.23381	433.25008
7	506.30200	520.31787	532.31761	546.33514
8	549.34396	563.36006	575.35988	589.37602
9	577.33916	591.35495	603.35488	617.37101
10	645.36089	659.37692	671.37679	685.39284
11	723.40758	737.42373	749.42344	763.43957

Peak no.	Myxochromide S ₁ (12.17 min)	Myxochromide S ₂ (12.85 min)	Myxochromide S ₃ (12.99 min)	-
1	225.12809	239.14411	251.14376	-
2	244.13391	258.14997	270.14958	-
3	322.18095	336.19704	348.19662	-
4	407.27029	421.28618	433.28603	-
5	435.26516	449.28960	461.28083	-
6	478.30728	492.32320	504.32348	-
7	506.30228	520.31808	532.31799	-
8	549.34461	563.36014	575.36022	-
9	577.33940	591.35497	603.35523	-
10	645.36161	659.37674	671.37717	-
11	723.40774	723.42319	749.42391	-
Peak no.	Myxochromide S ₁ -Abu (12.43 min)	Myxochromide S ₂ -Abu (13.09 min)	Myxochromide S ₃ -Abu (13.22 min)	-
1	225.12800	239.14402	251.14376	-
2	244.13387	258.14987	270.14960	-
3	322.18092	336.19687	348.19668	-
4	340.19148	354.20723	366.20726	-
5	407.27016	421.28595	433.28579	-
6	591.35482	605.36993	617.37078	-
7	659.37678	673.39153	685.39273	-
8	737.42350	751.43762	763.43973	-
Peak no.	Myxochromide S ₁ -diAbu (12.70 min)	Myxochromide S ₂ -diAbu (13.37 min)	Myxochromide S ₃ -diAbu (13.48 min)	-
1	225.12844	239.14292	251.14285	-
2	244.13429	258.14875	270.14861	-
3	322.18135	336.19561	348.19567	-
4	340.19179	354.20621	-	-
5	407.27042	421.28482	-	-
6	577.37548	591.39058	-	-
7	605.37051	619.38558	631.38563	-
8	673.39260	687.40784	699.40533	-
9	751.43863	765.45465	777.45432	-

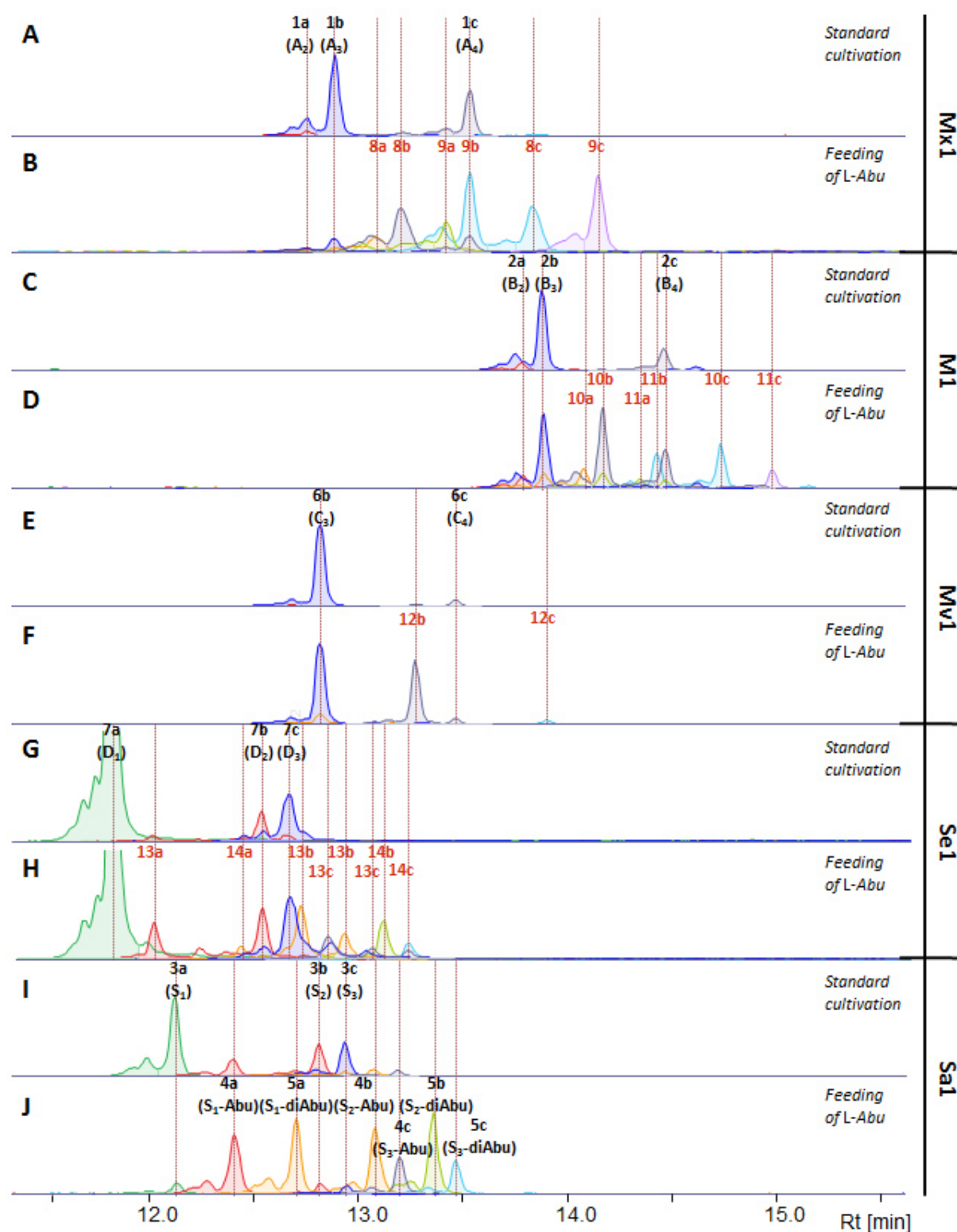
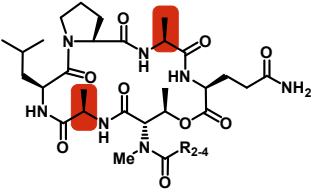
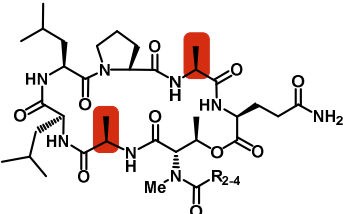
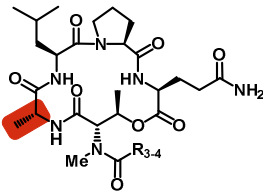
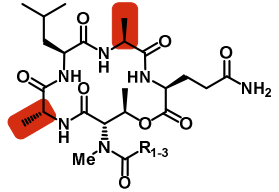
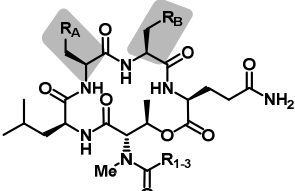


Figure S5. HPLC-MS analysis of myxochromide production in selected producer strains after feeding of *L*-α-aminobutyric acid (*L*-Abu) compared to standard cultivation experiments. The following strains were analyzed: Myxochromide A producer Mx1 (A/B), myxochromide B producer M1 (C/D), myxochromide C producer Mv1 (E/F), myxochromide D producer Se1 (G/H) and myxochromide S/S-Abu/S-diAbu producer Sa1 (I/J). Extracted ion chromatograms (EICs) for ± 0.02 m/z corresponding to the $[M+H]^+$ ions of known myxochromides (compounds 1-7) and novel myxochromide derivatives detected after *L*-Abu feeding (compounds 8-14, labelled in red) are shown. For further details see Table S3.

Table S3. Myxochromide derivatives detected in feeding experiments with *L*- α -aminobutyric acid (*L*-Abu). The novel derivatives (compounds **8-14**) are expected to harbor additional methylene groups in positions highlighted in red due to Abu instead of alanine incorporation.

Strain	Myxochromide	Compound	Formula	[M+H] ⁺ calc.	Rt [min]	Structures
<i>M. xanthus</i> DK1622 (Mx1)	A ₂	1a	C ₄₄ H ₆₃ N ₇ O ₉	834,47600	12,80	 <p>Myxochromides A₂₋₄ (1a-c)</p>
	A ₃	1b	C ₄₅ H ₆₃ N ₇ O ₉	846,47600	12,93	
	A ₄	1c	C ₄₆ H ₆₅ N ₇ O ₉	860,49165	13,58	
	[A ₂ + CH ₂]	8a	C ₄₅ H ₆₅ N ₇ O ₉	848,49165	13,08	
	[A ₃ + CH ₂]	8b	C ₄₆ H ₆₅ N ₇ O ₉	860,49165	13,20	
	[A ₄ + CH ₂]	8c	C ₄₇ H ₆₇ N ₇ O ₉	874,50730	13,85	
	[A ₂ + C ₂ H ₄]	9a	C ₄₆ H ₆₇ N ₇ O ₉	862,50730	13,43	
	[A ₃ + C ₂ H ₄]	9b	C ₄₇ H ₆₇ N ₇ O ₉	874,50730	13,56	
	[A ₄ + C ₂ H ₄]	9c	C ₄₈ H ₆₉ N ₇ O ₉	888,52295	14,16	
<i>Myxococcus</i> sp. 171 (M1)	B ₂	2a	C ₅₀ H ₇₄ N ₈ O ₁₀	947,56007	13,82	 <p>Myxochromides B₂₋₄ (2a-c)</p>
	B ₃	2b	C ₅₁ H ₇₄ N ₈ O ₁₀	959,56007	13,92	
	B ₄	2c	C ₅₂ H ₇₆ N ₈ O ₁₀	973,57572	14,50	
	[B ₂ + CH ₂]	10a	C ₅₁ H ₇₆ N ₈ O ₁₀	961,57572	14,08	
	[B ₃ + CH ₂]	10b	C ₅₂ H ₇₆ N ₈ O ₁₀	973,57572	14,17	
	[B ₄ + CH ₂]	10c	C ₅₃ H ₇₈ N ₈ O ₁₀	987,59137	14,73	
	[B ₂ + C ₂ H ₄]	11a	C ₅₂ H ₇₈ N ₈ O ₁₀	975,59137	14,35	
	[B ₃ + C ₂ H ₄]	11b	C ₅₃ H ₇₈ N ₈ O ₁₀	987,59137	14,42	
	[B ₄ + C ₂ H ₄]	11c	C ₅₄ H ₈₀ N ₈ O ₁₀	1001,60702	14,99	
<i>M. virescens</i> ST200611 (Mv1)	C ₃	6b	C ₄₂ H ₅₈ N ₆ O ₈	775,43889	12,86	 <p>Myxochromides C₃₋₄ (6b-c)</p>
	C ₄	6c	C ₄₃ H ₆₀ N ₆ O ₈	789,45454	13,50	
	[C ₃ + CH ₂]	12b	C ₄₃ H ₆₀ N ₆ O ₈	789,45454	13,28	
	[C ₄ + CH ₂]	12c	C ₄₄ H ₆₂ N ₆ O ₈	803,47019	13,90	
<i>S. erecta</i> Pde77 (Se1)	D ₁	7a	C ₃₈ H ₅₄ N ₆ O ₈	723,40759	11,83	 <p>Myxochromides D₁₋₃ (7a-c)</p>
	D ₂	7b	C ₃₉ H ₅₆ N ₆ O ₈	737,42324	12,54	
	D ₃	7c	C ₄₀ H ₅₆ N ₆ O ₈	749,42324	12,68	
	[D ₁ + CH ₂]	13a	C ₃₉ H ₅₆ N ₆ O ₈	737,42324	12,00	
	[D ₂ + CH ₂]	13b	C ₄₀ H ₅₈ N ₆ O ₈	751,43889	12.71/12.91	
	[D ₃ + CH ₂]	13c	C ₄₁ H ₅₈ N ₆ O ₈	763,43889	12.84/13.04	
	[D ₁ + C ₂ H ₄]	14a	C ₄₀ H ₅₈ N ₆ O ₈	751,43889	12,42	
	[D ₂ + C ₂ H ₄]	14b	C ₄₁ H ₆₀ N ₆ O ₈	765,45454	13,09	
<i>S. aurantiaca</i> DW4/3-1 (Sa1)	S ₁	3a	C ₃₈ H ₅₄ N ₆ O ₈	723,40759	12,17	 <p>Myxochromides S₁₋₃ (3a-c; R_{A,B}=H) Myxochromides E₁₋₃ (4a-c; R_A=CH₃, R_B=H) Myxochromides F₁₋₃ (5a-c; R_{A,B}=CH₃)</p>
	S ₂	3b	C ₃₉ H ₅₆ N ₆ O ₈	737,42324	12,85	
	S ₃	3c	C ₄₀ H ₅₆ N ₆ O ₈	749,42324	12,97	
	E ₁	4a	C ₃₉ H ₅₆ N ₆ O ₈	737,42324	12,41	
	E ₂	4b	C ₄₀ H ₅₈ N ₆ O ₈	751,43889	13,09	
	E ₃	4c	C ₄₁ H ₅₈ N ₆ O ₈	763,43889	13,20	
	F ₁	5a	C ₄₀ H ₅₈ N ₆ O ₈	751,43889	12,71	
	F ₂	5b	C ₄₁ H ₆₀ N ₆ O ₈	765,45454	13,36	
	F ₃	5c	C ₄₂ H ₆₀ N ₆ O ₈	777,45454	13,47	

2.7.3 Isolation and Structure Elucidation of Myxochromide C₃ from *M. virescens* ST200611

2.7.3.1 Cultivation of *M. virescens* ST200611 and Isolation of Myxochromide C₃

The putative myxochromide C producer strain *Mycrococcus virescens* ST200611 (Mv1) was cultivated in 20 L (10x 2 L) CYS medium (casitone 0.25%, yeast extract 0.1%, CaCl₂ × 2 H₂O 0.1%, starch 0.25%, HEPES 0.24%, vitamin B₁₂ 500 µg/L, pH adjusted to 7.0) including 2% XAD-16 resin for 4 days at 30 °C and 180 rpm. Cells and XAD-16 amberlite adsorber resin were harvested by centrifugation at 10,500 rpm and 4 °C for 15 min and were five times extracted with 1 L of a mixture of methanol and acetone (1:1). The organic solvents were removed under reduced pressure and the residue was six times extracted with 200 mL of ethyl acetate. After removal of the solvent, the crude extract was dissolved in 15 mL of methanol for subsequent separation via size exclusion chromatography using Sephadex LH-20 and methanol as solvent. The collected fractions were concentrated and analyzed by HPLC-MS analysis for the presence of myxochromides C. Analysis was performed on a Dionex Ultimate 3000 RSLC system coupled to a Bruker amaZon speed mass spectrometer using a Waters BEH C18, 50 × 2.1 mm, 1.7 µm d_p column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0-0.5 min 5% B, 0.5-9.5 min 5-95%, 9.5-10.5 min 95% B, 10.5-10.8 min 95-5% B, 10.8-12.5 min 5% B. Full scan mass spectra were acquired in positive ESI mode ranging from 200-2000 *m/z*. Sephadex fractions containing the target compound were combined, evaporated and dissolved in methanol for subsequent separation via reverse phase HPLC. A Dionex UltiMate 3000 system equipped with a Luna 5u C18(2) 100A column (250 × 10 mm, Phenomenex) was used. At constant flow rate (5.0 mL/min), the following multi-step gradient was applied (A: deionized water, B: acetonitrile): 0-5 min 10-45% B, 5-30 min 45-65% B, 30-40 min 65-80% B, 40-41 min 80-10% B, 41-47 min 10% B. UV traces were recorded by a diode array detector (DAD) with specified wave lengths (210, 266 and 410 nm) with myxochromides showing good UV absorption at 410 nm. A total amount of 2.3 mg of pure myxochromide C₃ (R_t = 27.7 min) was isolated.

2.7.3.2 Structure Elucidation of Myxochromide C₃

Structure elucidation of myxochromide C₃ was achieved using 1D and 2D NMR spectroscopy as well as HR-MS data. NMR spectra were acquired in CD₃OD at a Bruker Ascend 700 or 500 MHz spectrometer equipped with a 5 mm TXI cryoprobe. 1D ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra were recorded using standard pulse programs and are illustrated in Figure S7. Carbon chemical shifts were extracted from 2D NMR data. NMR

spectroscopic data are listed in the Table S4. HR-ESI-MS data were obtained on a Bruker Maxis 4G mass spectrometer. Full scan mass spectra were acquired in a range from 150-2500 m/z in a positive mode. HR-ESI-MS of myxochromide C₃ gave a quasimolecular ion at m/z 775.43921 $[M+H]^+$ consistent with a molecular formula C₄₂H₅₇N₆O₈ (775.43889, calculated for C₄₂H₅₈N₆O₈, $\Delta m/z$ 0.413 ppm). The ¹H NMR spectrum exhibited signals corresponding to five α -CH protons (δ_H 4.2-5.5), four CH₃ groups (δ_H 0.9-1.3), three CH₂ groups (δ_H 1.3-2.2) and one methine proton (δ_H 1.55) together with a *N*-Me group (δ_H 3.39, 3H, s). Moreover, a number of downfield signals belonging to the unsaturated polyketide side chain (δ_H 5.7-7.3) and a CH₃ signal (δ_H 1.78, 3H, d) were observed. 2D NMR data revealed the presence of *N*-Me-threonine, alanine, leucine, proline and glutamine residues and HMBC/ROESY data established the cyclic pentapeptide with the unsaturated polyketide side chain (Figure S6). The NMR data and core structure of myxochromide C₃ were in accordance with the one of myxochromide A, which differs by its additional alanine residue.⁷

For the assignment of the absolute configuration, Marfey's method based on amino acid derivatization was applied.²⁹ 0.3 mg of pure myxochromide C₃ was hydrolyzed with 37% HCl (0.2 mL) in a 1.5 mL glass vial for 3 days at 110°C. The hydrolysate was evaporated to dryness and dissolved in H₂O (100 μ L). A 50 μ L aliquot was supplemented with 1N NaHCO₃ (20 μ L) and 1% 1-fluoro-2,4-dinitrophenyl-5-*L/D*-leucinamide (*L*-FDLA or *D*-FDLA) solution in acetone (20 μ L), and the mixtures were heated to 40 °C for 8 h at 700 rpm. After cooling down to room temperature, the solutions were neutralized with 2N HCl (20 μ L), evaporated to dryness and the derivatized amino acids were dissolved in 300 μ L acetonitrile. An amino acid standard mix (Sigma Aldrich) as well as *N*-Me-*L*-Threonine (Sigma Aldrich) were derivatized via the same procedure and all samples were analyzed on a Dionex Ultimate 3000 RSLC system coupled to a Bruker Maxis 4G mass spectrometer. Separation was performed using a Waters BEH C18, 100 \times 2.1 mm, 1.7 μ m d_p column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0 min 5% B, 0-1 min 5-10% B, 1-15 min 10-35% B, 15-22 min 35-55% B, 22-25 min 55-80% B, 25-26 min 80% B, 26-26.5 min 80-5% B, 26.5-31 min 5% B. Full scan mass spectra were acquired in a range from 100-1000 m/z . The chromatograms obtained from the HPLC-MS analysis are illustrated in Figure S8 and stereochemical assignments are illustrated in Table S5.

Comparison of the retention times and m/z values of derivatized standard amino acids and the hydrolyzed lipopeptide revealed the presence of a *D*-configured alanine residue (C13) in myxochromide C₃. The amino acids proline (C2), leucine (C7), *N*-Me-threonine (C16) and

glutamine (C20), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. Since the alanine activating module contains an epimerization domain, the detection of *D*-alanine was expected and was already observed for myxochromides A and B.^{7,8}

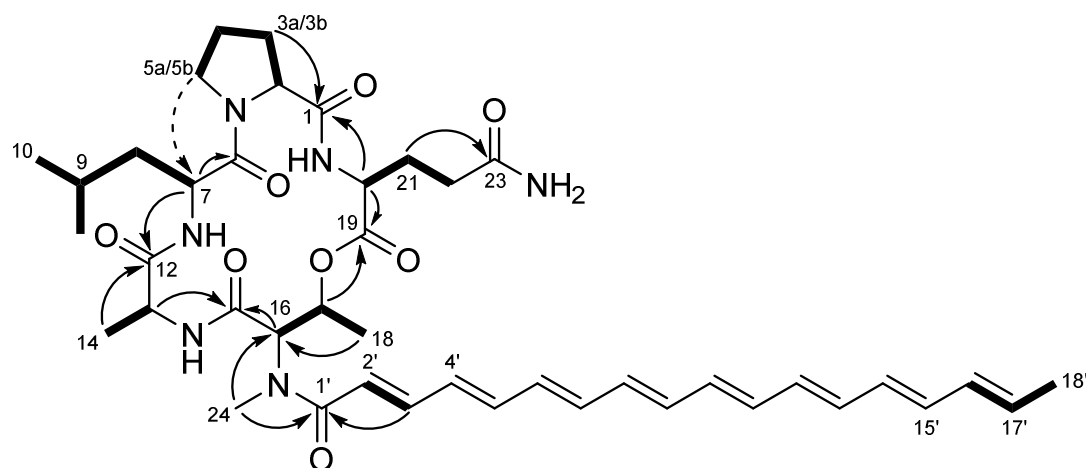


Figure S6. Structure of myxochromide C₃ showing COSY (bold line), ROESY (dashed arrow) and key HMBC (plain arrow) correlations.

Table S4. NMR spectroscopic data of myxochromide C₃.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c	ROESY ^{d, e}
<i>L</i> -Pro	1	174.6			
	2	63.3	4.31, dd (3.1, 9.1)	1, 3, 4, 5	7
	3a	32.7	2.16, m	1, 4, 5	
	3b		2.43, m	1, 2, 4, 5	
	4	23.6	2.00, m	2, 3, 5	
	5a	47.9	3.52, m	2, 3, 4	
	5b		3.75, m	2, 3, 4	
<i>L</i> -Leu	6	174.2			
	7	49.6	4.68, dd (2.7, 11.4)	6, 8, 9, 12	2, 10, 11
	8a	42.4	1.32, m	7	
	8b		1.61, m	6, 7	
	9	25.6	1.55, m	7, 8, 10, 11	
	10	23.7	0.93, d (6.2)	8, 9, 11	
	11	21.9	0.92, d (6.2)	10	
<i>D</i> -Ala	12	174.7			
	13	51.7	4.20, m	12, 14, 15	
	14	17.1	1.32, m	12, 13	
<i>N</i> -Me- <i>L</i> -Thr	15	170.9			
	16	61.3	5.50, d (2.9)	1', 15, 17, 24	2', 24
	17	72.4	6.02, m	18, 19	
	18	16.6	1.19, d (6.7)	16, 17	24
	24	35.0	3.39, s	1', 16	2', 18, 20
<i>L</i> -Gln	19	171.4			
	20	55.0	4.40, dd (5.3, 6.9)	1, 19, 21, 22	
	21	29.0	2.09, m	20, 22, 23	
	22	31.6	2.22, t (7.8)	20, 21, 23	
	23	177.2			
Side chain	1'	171.2			
	2'	120.4	6.66, d (14.6)	1', 4'	16, 24
	3'	145.1	7.34, dd (11.4, 14.6)	1', 2', 5'	
	4'	138.5	6.54, m	<i>f</i>	
	5'	142.0	6.72, m	3'	
	6'-14'	<i>f</i>	<i>f</i>	<i>f</i>	
	15'	135.1	6.25, m	<i>f</i>	
	16'	133.2	6.13, m	15', 18'	
	17'	131.2	5.76, m	15', 18'	
	18'	18.4	1.78, d (6.7)	15', 16', 17'	

^a acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 700 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d proton showing ROESY correlations to indicated protons.^e acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^f overlapped signals.

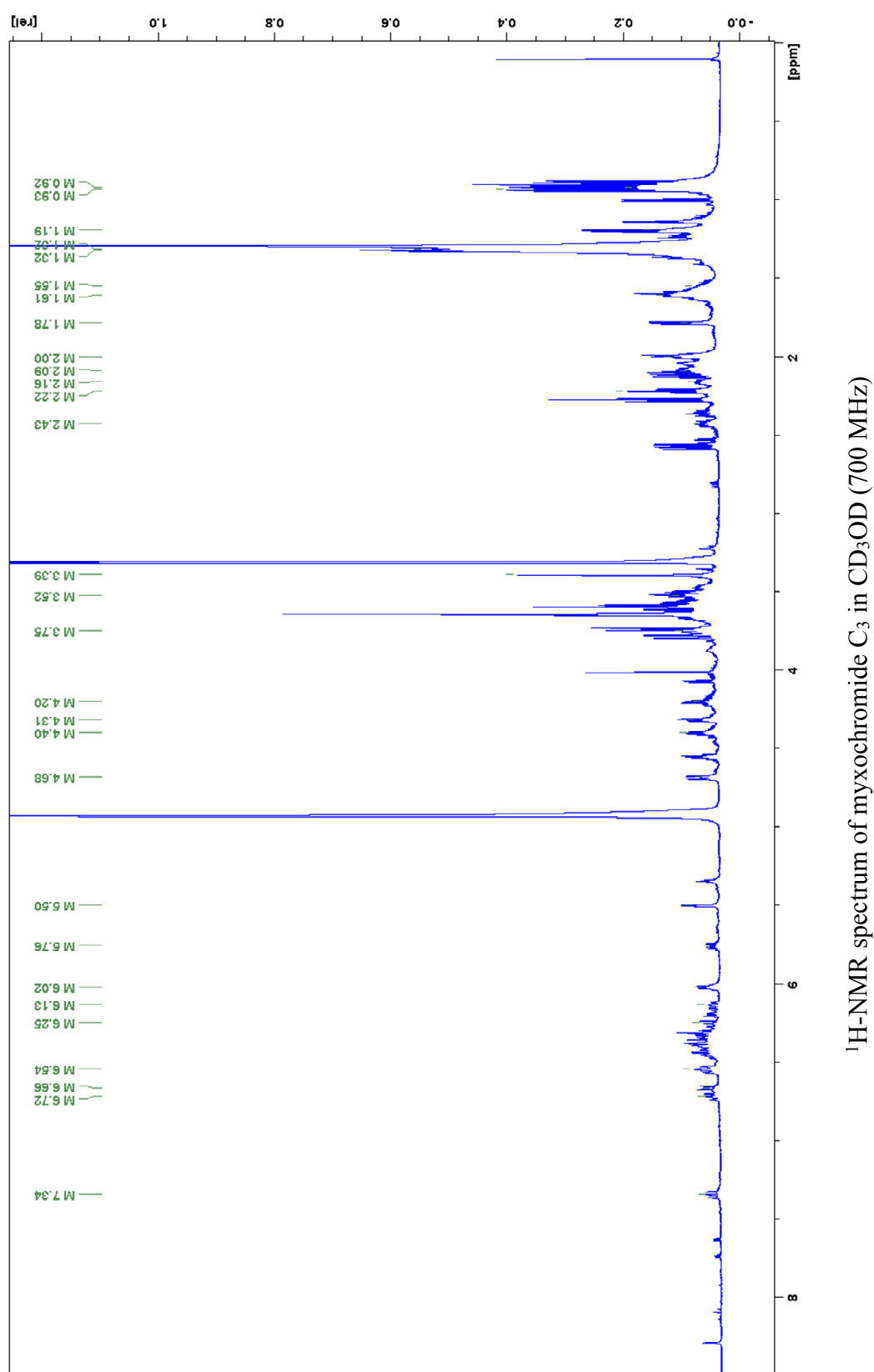


Figure S7 (continued on next page)

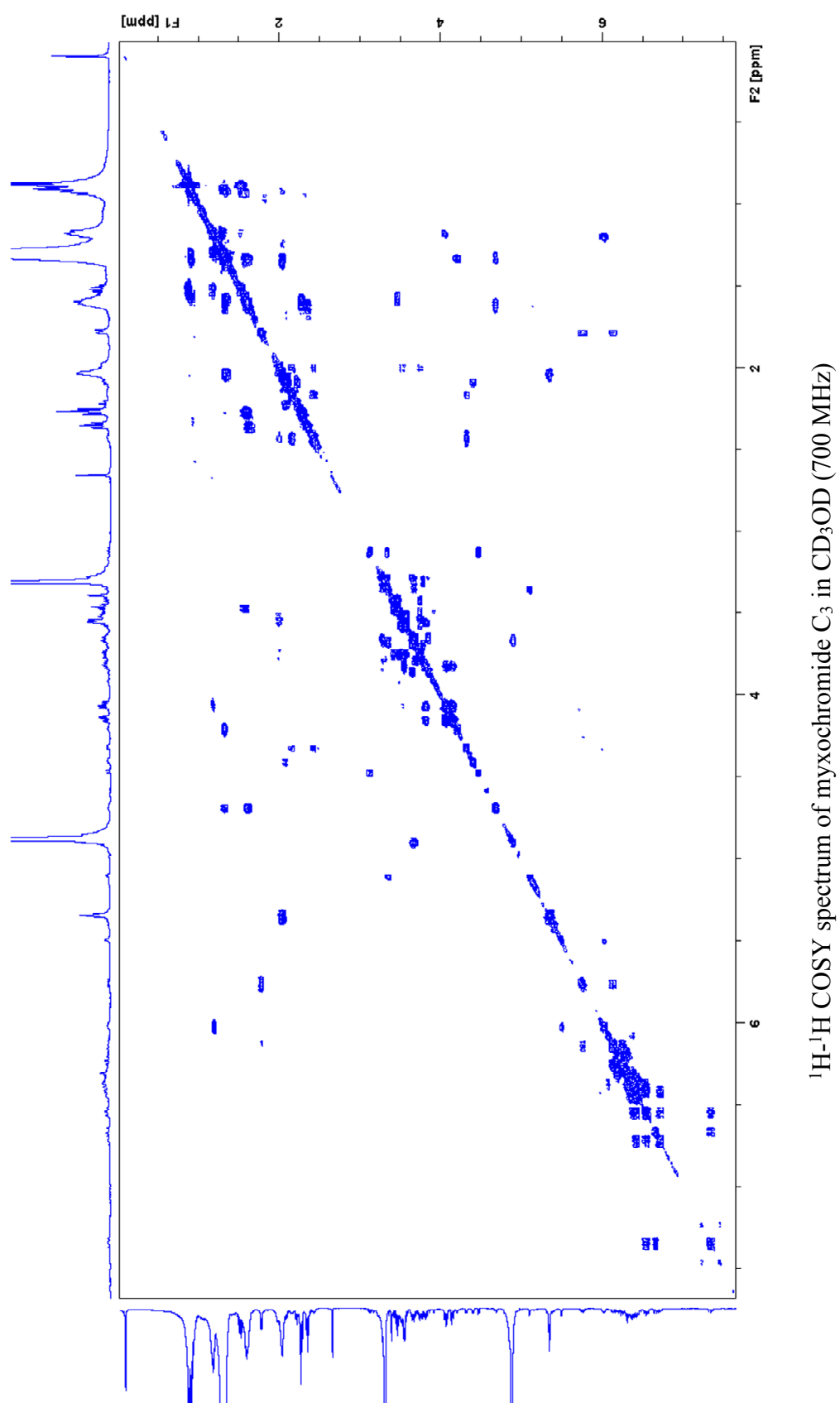
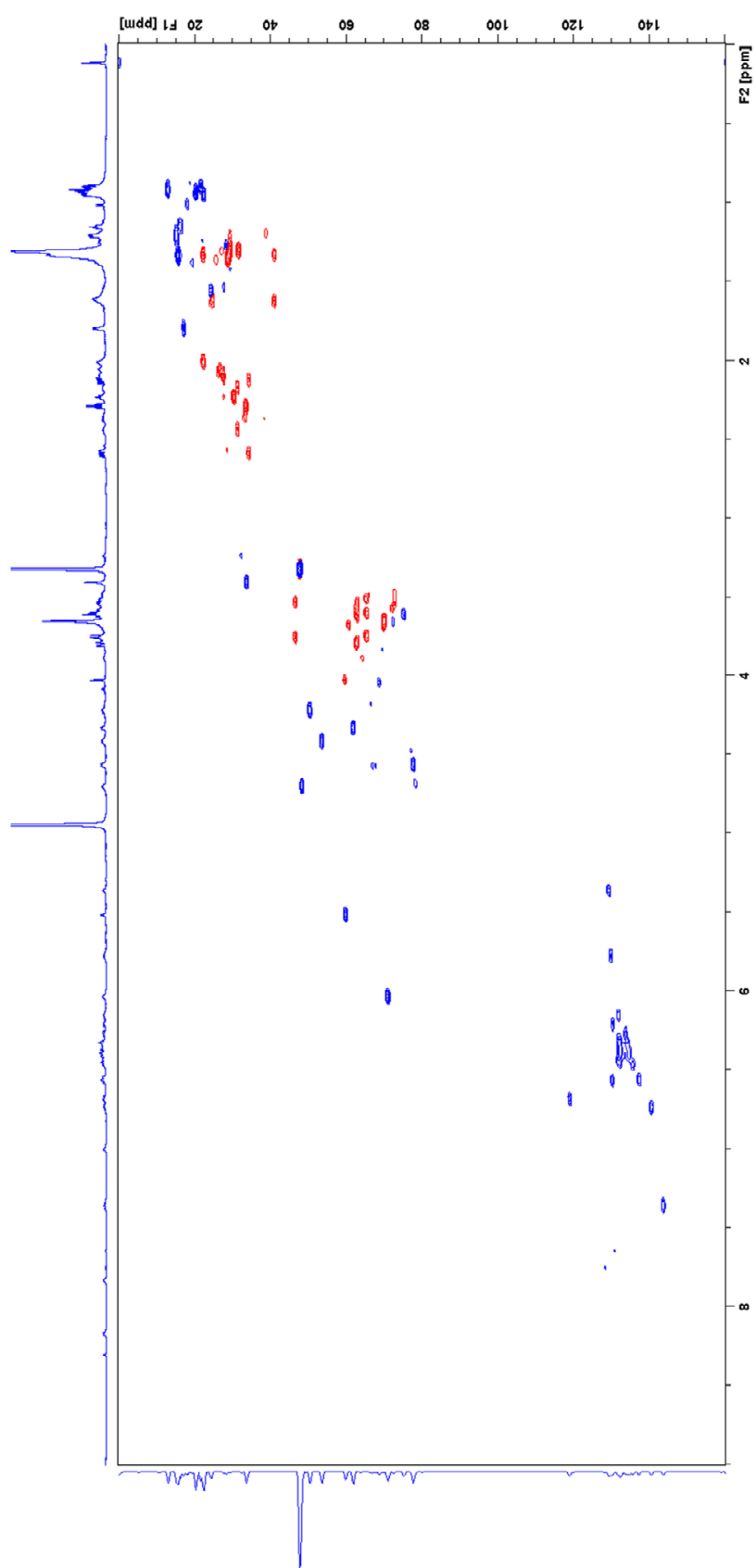
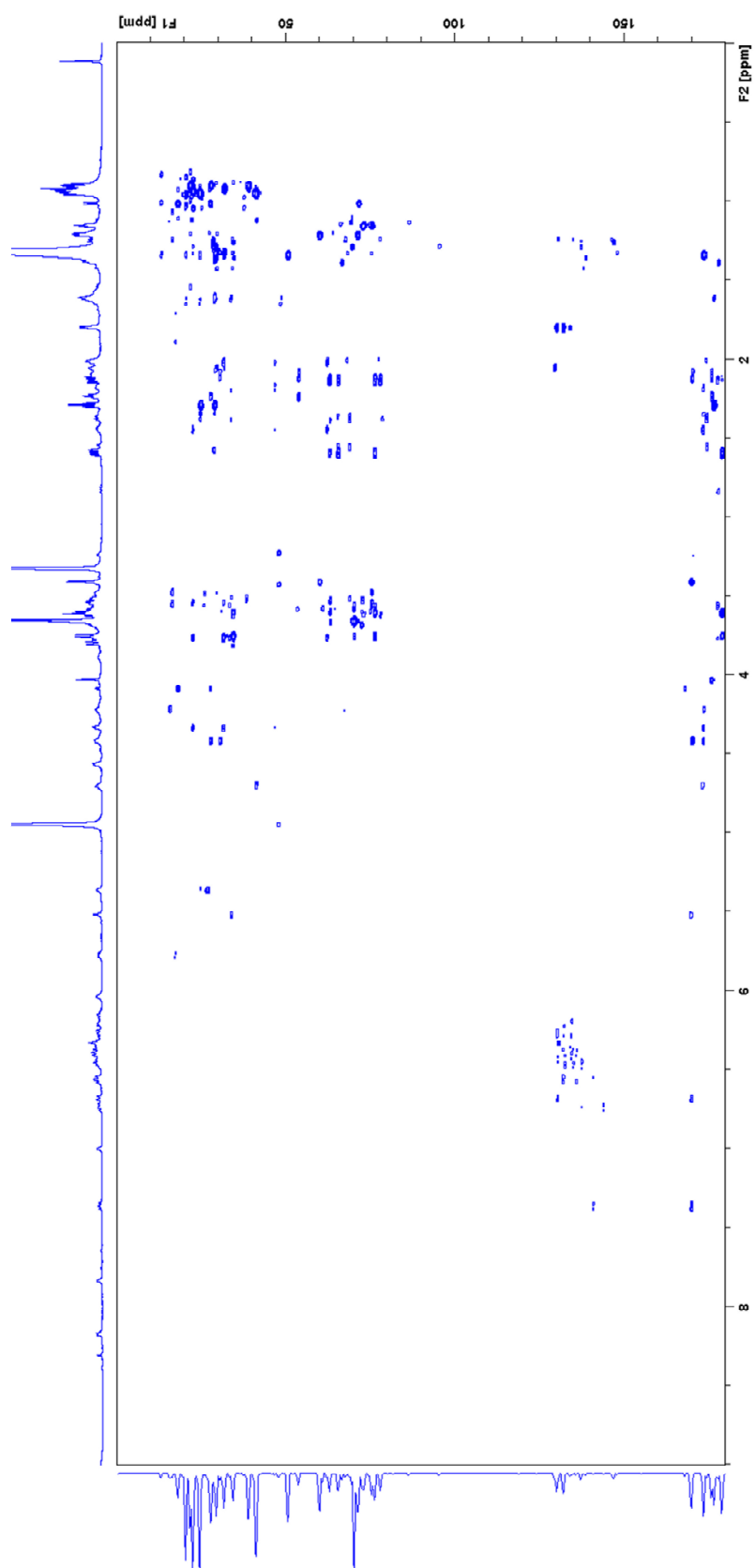


Figure S7 (continued on next page)



HSQC spectrum of myxochromide C_3 in CD_3OD (700 MHz)

HMBC spectrum of myxochromide C₃ in CD₃OD (700 MHz)

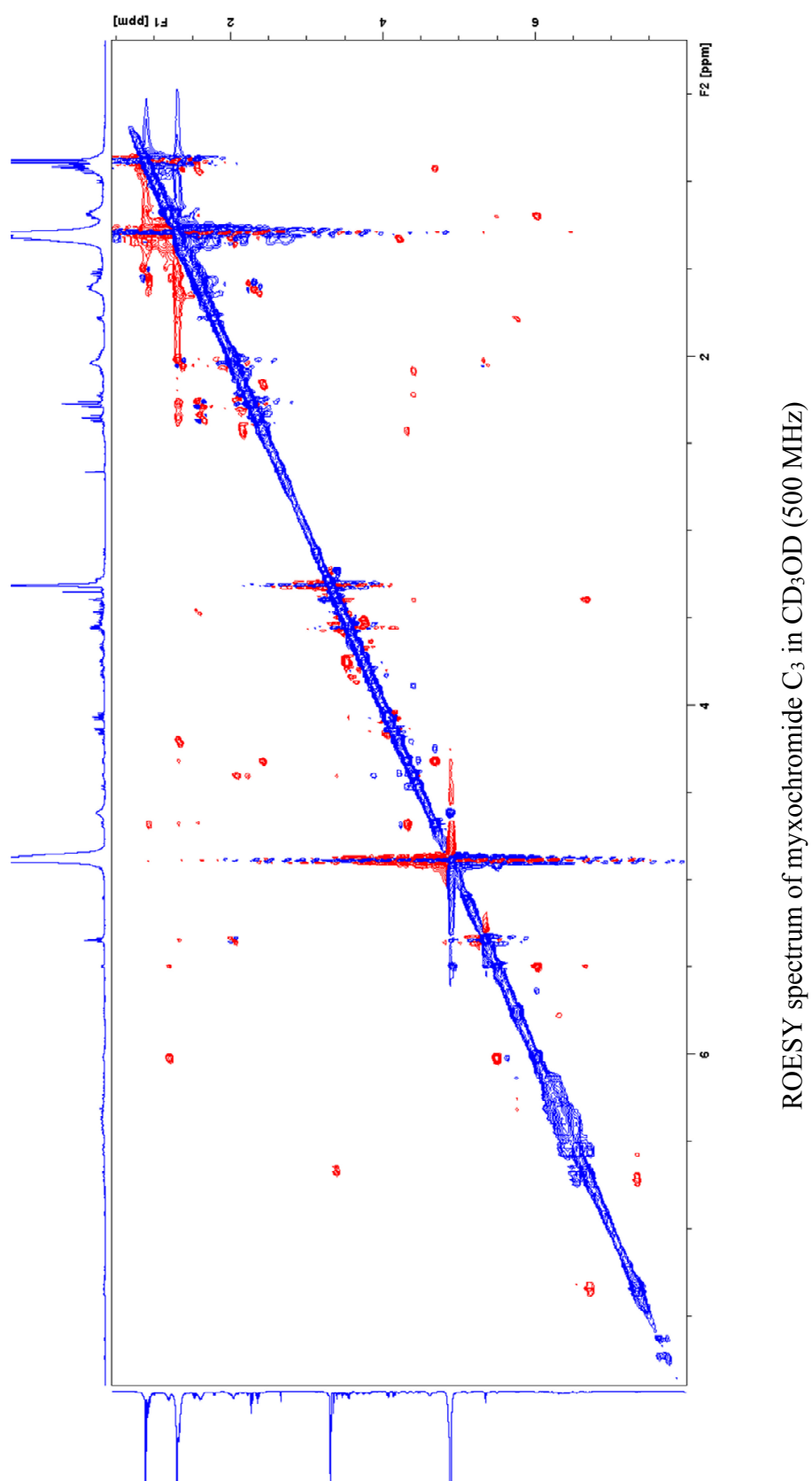


Figure S7. NMR spectra of myxochromide C₃.

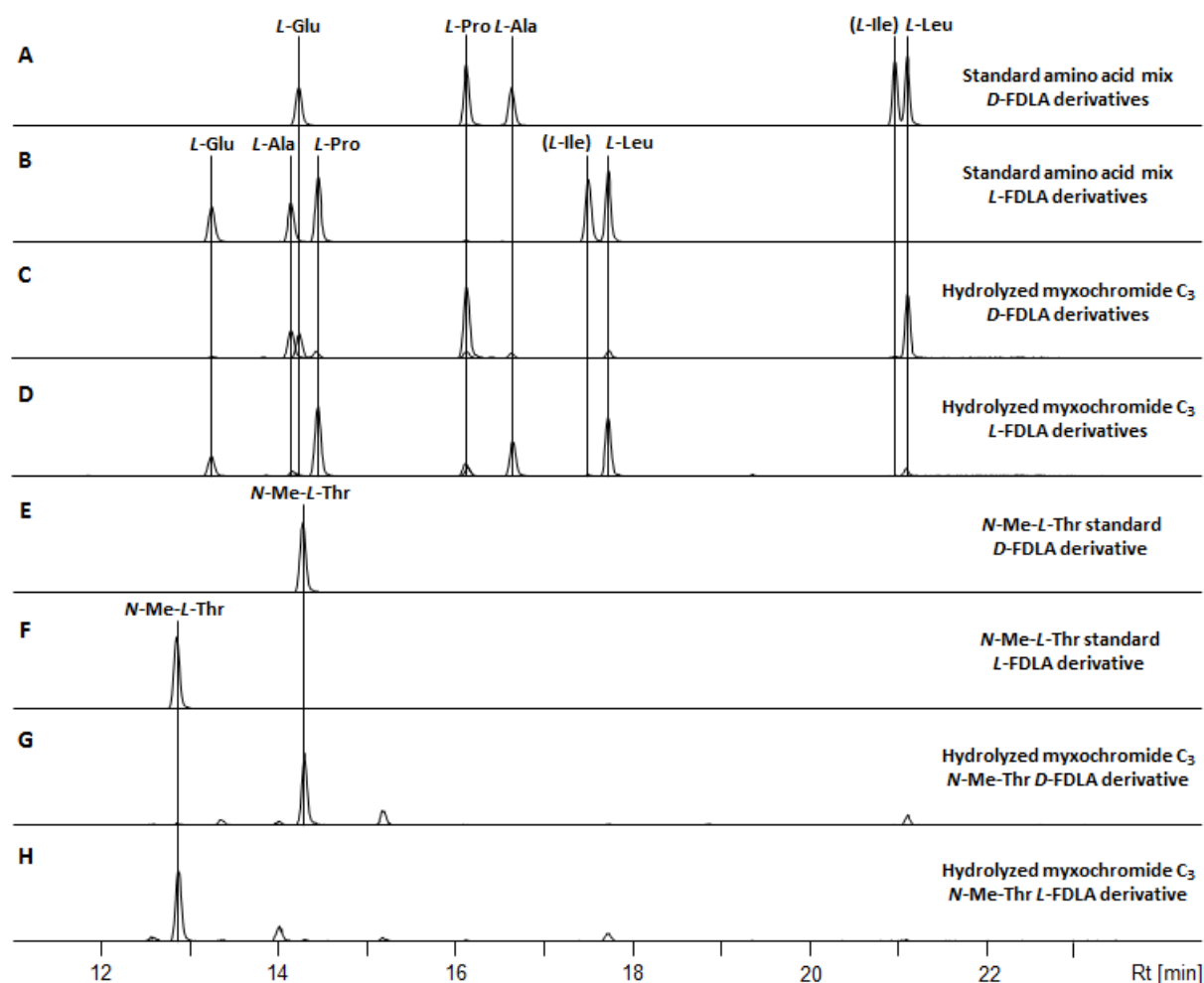


Figure S8. Results of the analysis of the absolute configuration by advanced Marfey's method.²⁹ Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide C₃ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide C₃ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative.

Table S5. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide C₃ (inverse correlation of retention times (t_R) of *D*-configured Ala-FDLA derivatives from the peptide hydrolysate compared to *L*-configured Ala-FDLA standards are shown in bold).

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	t_R [min]	m/z $[M+H]^+$	t_R [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.2	442.1617	14.2	442.1619	L
Glu- <i>L</i> -FDLA	13.2	442.1619	13.2	442.1620	
Pro- <i>D</i> -FDLA	16.1	410.1711	16.1	410.1723	L
Pro- <i>L</i> -FDLA	14.4	410.1720	14.4	410.1722	
Ala- <i>D</i> -FDLA	16.6	384.1554	14.1	384.1561	D
Ala- <i>L</i> -FDLA	14.1	384.1561	16.6	384.1563	
Leu- <i>D</i> -FDLA	21.1	426.2027	21.1	426.2042	L
Leu- <i>L</i> -FDLA	17.7	426.2038	17.7	426.2039	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.3	428.1818	14.3	428.1822	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.8	428.1831	12.8	428.1825	

2.7.4 Isolation and Structure Elucidation of Myxochromide D₁ from *S. erecta* Pde77

2.7.4.1 Cultivation of *S. erecta* Pde77 and Isolation of Myxochromide D₁

The putative myxochromide D producer strain *Stigmatella erecta* Pde77 (Se1) was cultivated in 20 L (10x 2 L) VY/2 medium (Baker's yeast 0.5%, CaCl₂ × 2 H₂O 0.05%, HEPES 0.11%, vitamin B₁₂ 500 µg/L, pH adjusted to 7.0) including 2% XAD-16 resin for 7 days at 30 °C and 180 rpm. Cells and XAD-16 were harvested by centrifugation at 8,000 rpm and 4 °C for 15 min. Extraction and isolation of the target compound myxochromide D₁ using size exclusion chromatography and semi-preparative HPLC was achieved as described for myxochromide C₃ (see chapter 2.7.3.1). A total amount of 2.7 mg of pure myxochromide D₁ (R_t = 22.4 min) was isolated.

2.7.4.2 Structure Elucidation of Myxochromide D₁

Structure elucidation of myxochromide D₁ was achieved as described for myxochromide C₃ (see chapter 2.7.3.2). 1D ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra are shown in Figure S10. Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in the Table S6. Myxochromide D₁ showed quasimolecular ion at *m/z* 723.40748 [M+H]⁺ by HR-ESI-MS, which corresponds to the molecular formula C₃₈H₅₄N₆O₈ (723.40759, calculated for C₃₈H₅₅N₆O₈, Δ*m/z* -0.152 ppm). Its ¹H NMR spectrum closely resembled to that of myxochromide C₃. In addition to the common structural parts, analysis of 2D NMR spectra corroborated the presence of two alanine residues, one of which substitutes the proline in myxochromide C₃. Key HMBC correlations established the amino acid sequence and finalized its planar structure as depicted in Figure S9. Length of the polyene side chain was deduced based on the HR-MS data and molecular formula. For the assignment of the absolute configuration of myxochromide D₁ the same procedure, hydrolysis and Marfey analysis of the obtained amino acids,²⁹ was applied as described for myxochromide C₃ in chapter 2.7.3.2. The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S11 and stereochemical assignments are illustrated in Table S7. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that one of the two alanine residues (C2 and C11) from myxochromide D₁ has *D* configuration. The second alanine residue as well as the amino acids leucine (C5), *N*-Me-threonine (C14) and glutamine (C18), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. According to the domain organization of the myxochromide D assembly line, which harbors an epimerization domain in module 2, the *D*-configured alanine was assigned to C11. This also correlates with the structures of myxochromides A, B and C (this study; see chapter 2.7.3.2)^{7,8}.

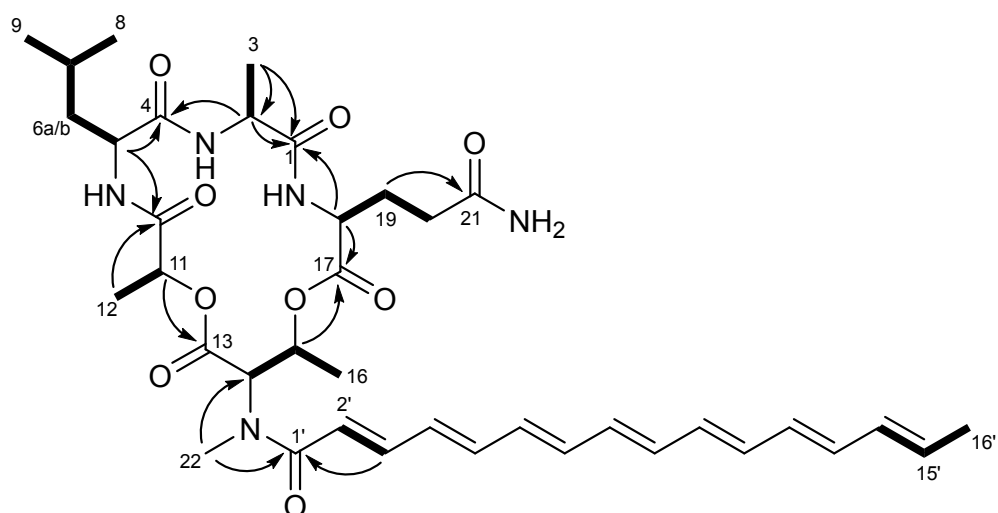


Figure S9. Structure of myxochromide D₁ showing COSY (bold line) and key HMBC (arrow) correlations.

Table S6. NMR spectroscopic data of myxochromide D₁ (CD₃OD).

Moiety	Position	δ_C^a	δ_H^b (<i>J</i> in Hz)	HMBC ^c	ROESY ^{b, d, e}
<i>L</i> -Ala	1	173.0			
	2	50.8	4.27, q (6.8)	1, 3, 4	
	3	18.6	1.37, d (6.8)	1, 2	
<i>L</i> -Leu	4	174.7			
	5	54.4	4.16, dd (5.4, 10.0)	4, 6, 7, 10	8, 9
	6a	40.7	1.64, m	7	
	6b		1.71, m		
	7	26.2	1.72, m	5, 6, 8, 9	
	8	21.4	0.92, d (6.3)	6, 7	5
	9	23.2	0.98, d (6.3)	6, 7	5
<i>D</i> -Ala	10	175.9			
	11	50.3	4.33, m	10, 12, 13	13
	12	16.7	1.30, m	10, 11	
<i>N</i> -Me- <i>L</i> -Thr	13	170.0			
	14	59.5	5.44, d (2.0)	1', 13, 15, 22	2', 22
	15	72.8	5.47, m	16, 17	
	16	17.2	1.15, d (5.4)	14, 15	
	22	35.2	3.07, s	1', 14	2', 14, 16, 18
<i>L</i> -Gln	17	171.0			
	18	54.0	3.98, dd (5.3, 9.4)	1, 17, 19, 20	
	19a	26.3	2.22, m	18, 20, 21	
	19b		2.13, m	18, 20, 21	
	20a	32.2	2.26, m	18, 19, 21	
	20b		2.21, m	18, 19, 21	
	21	177.7			
Side chain	1'	170.7			
	2'	120.2	6.60, d (14.5)	1'	14, 22
	3'	145.0	7.34, dd (11.4, 14.6)	1', 2', 5'	
	4'	138.6	6.53, m	<i>e</i>	
	5'	141.9	6.71, m	<i>e</i>	
	6'-12'	<i>e</i>	<i>e</i>	<i>e</i>	
	13'	135.3	6.26, m	<i>e</i>	
	14'	133.3	6.13, m	13', 16'	
	15'	131.4	5.77, m	13'	
	16'	18.4	1.79, d (6.3)	13', 14', 15'	

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d proton showing ROESY correlations to indicated protons.^e overlapped signals.

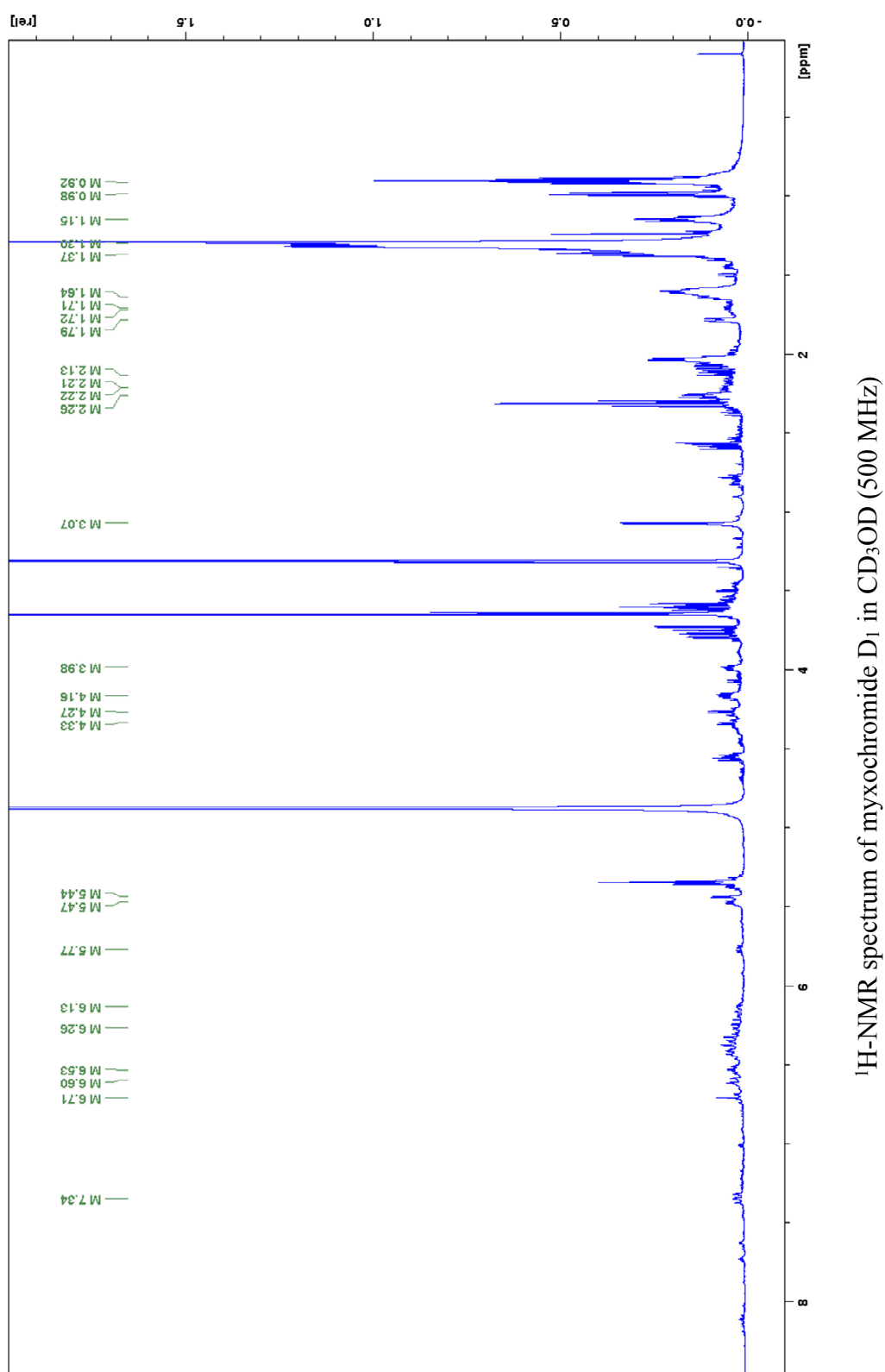


Figure S10 (continued on next page)

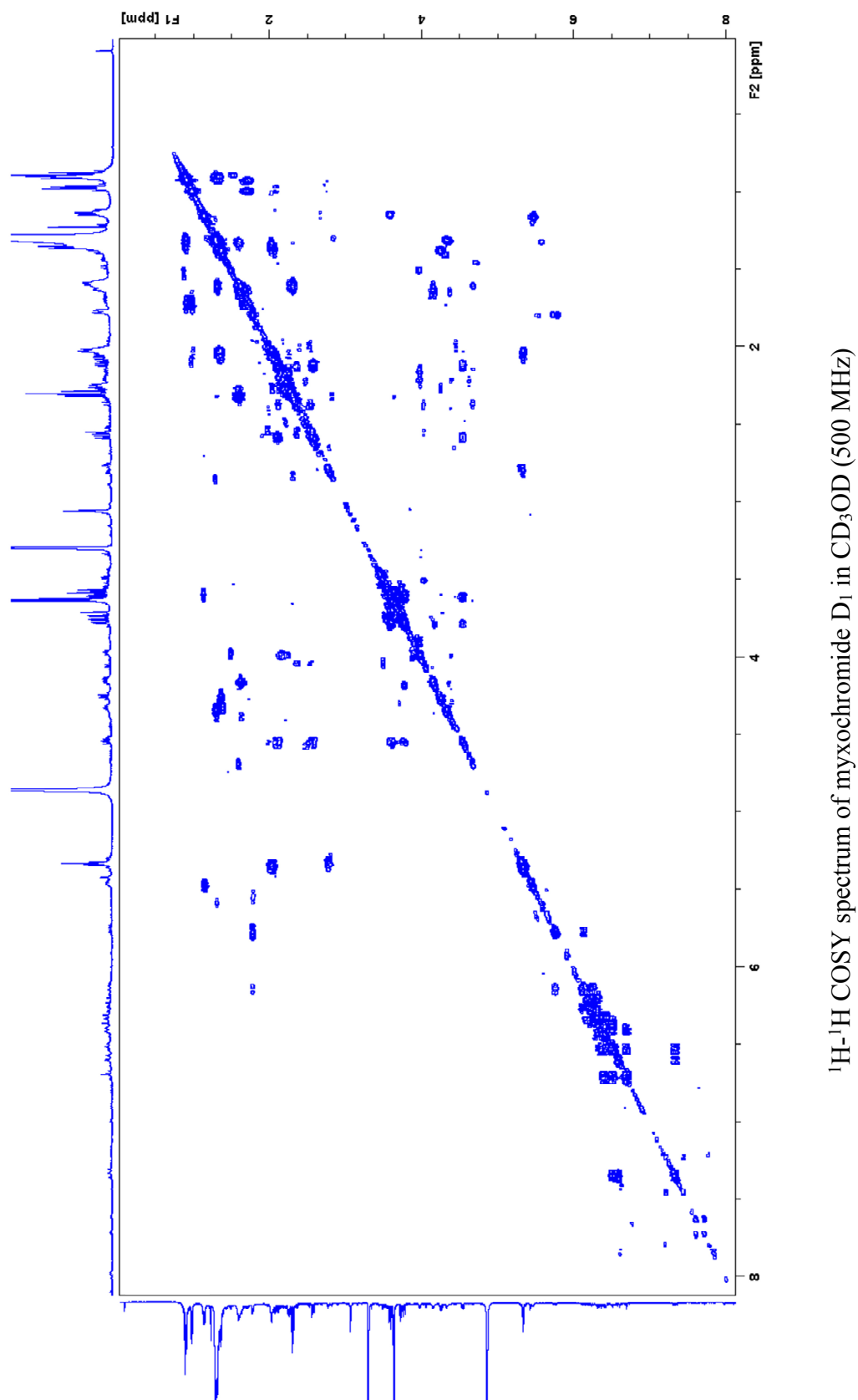


Figure S10 (continued on next page)

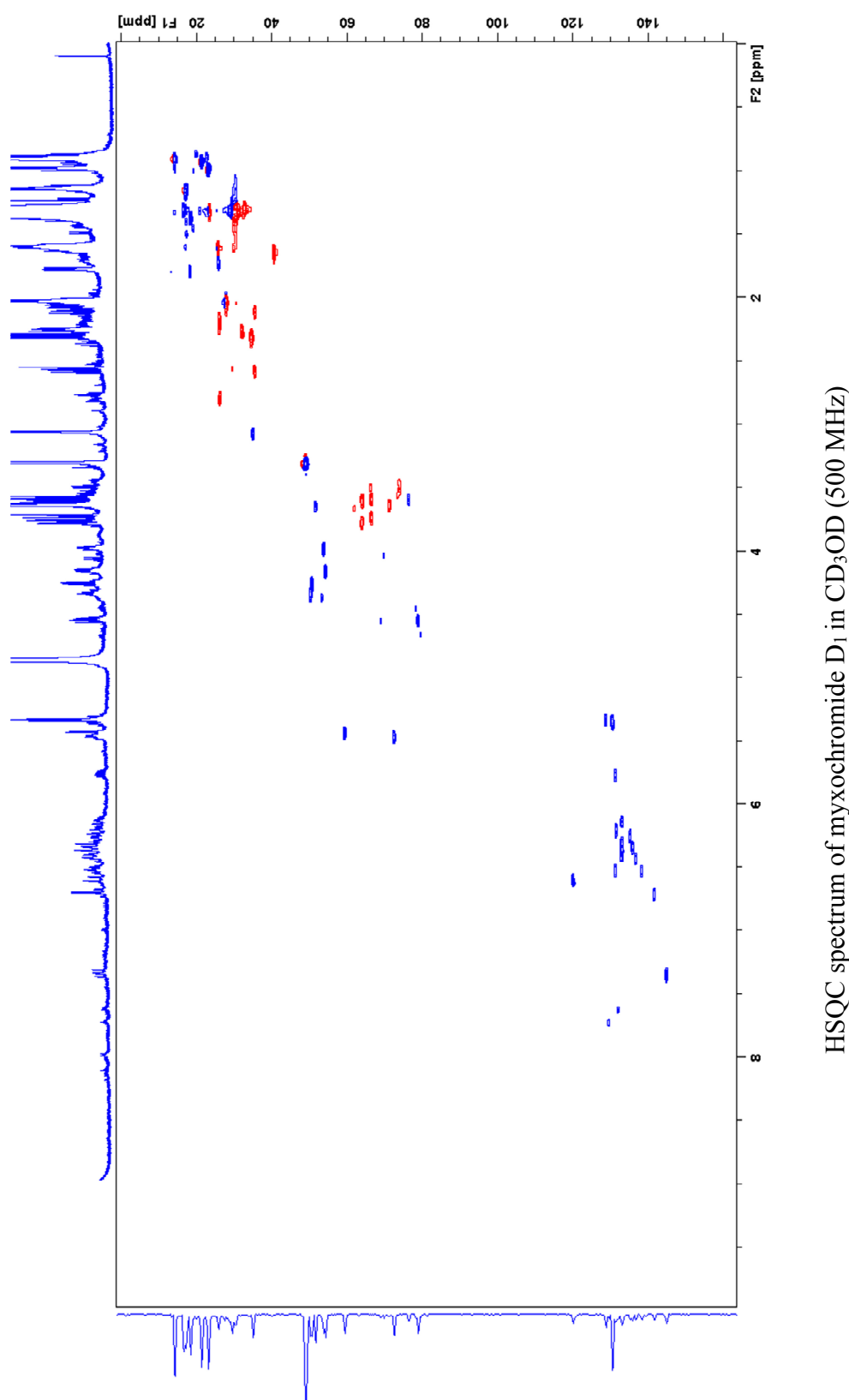


Figure S10 (continued on next page)

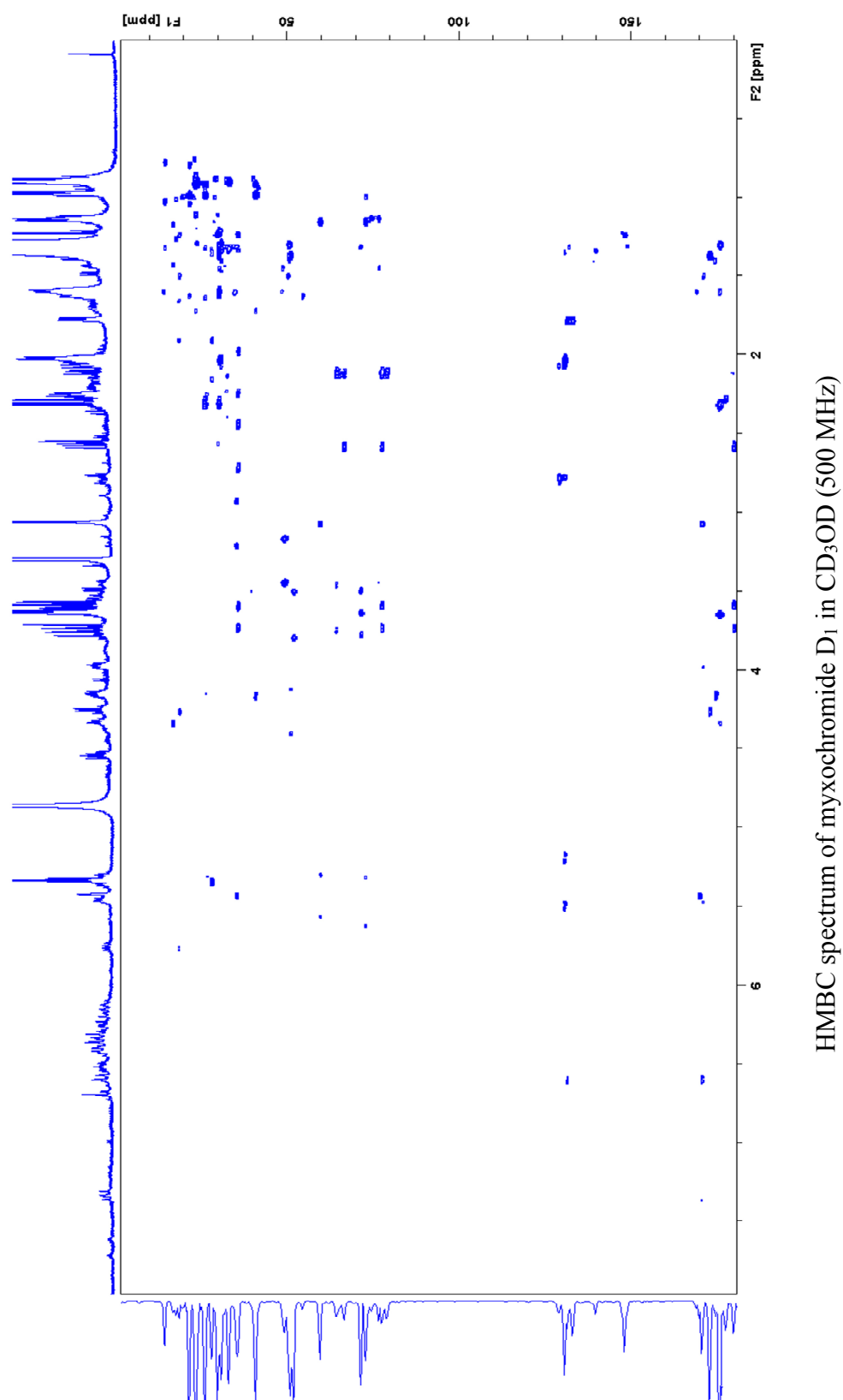


Figure S10 (continued on next page)

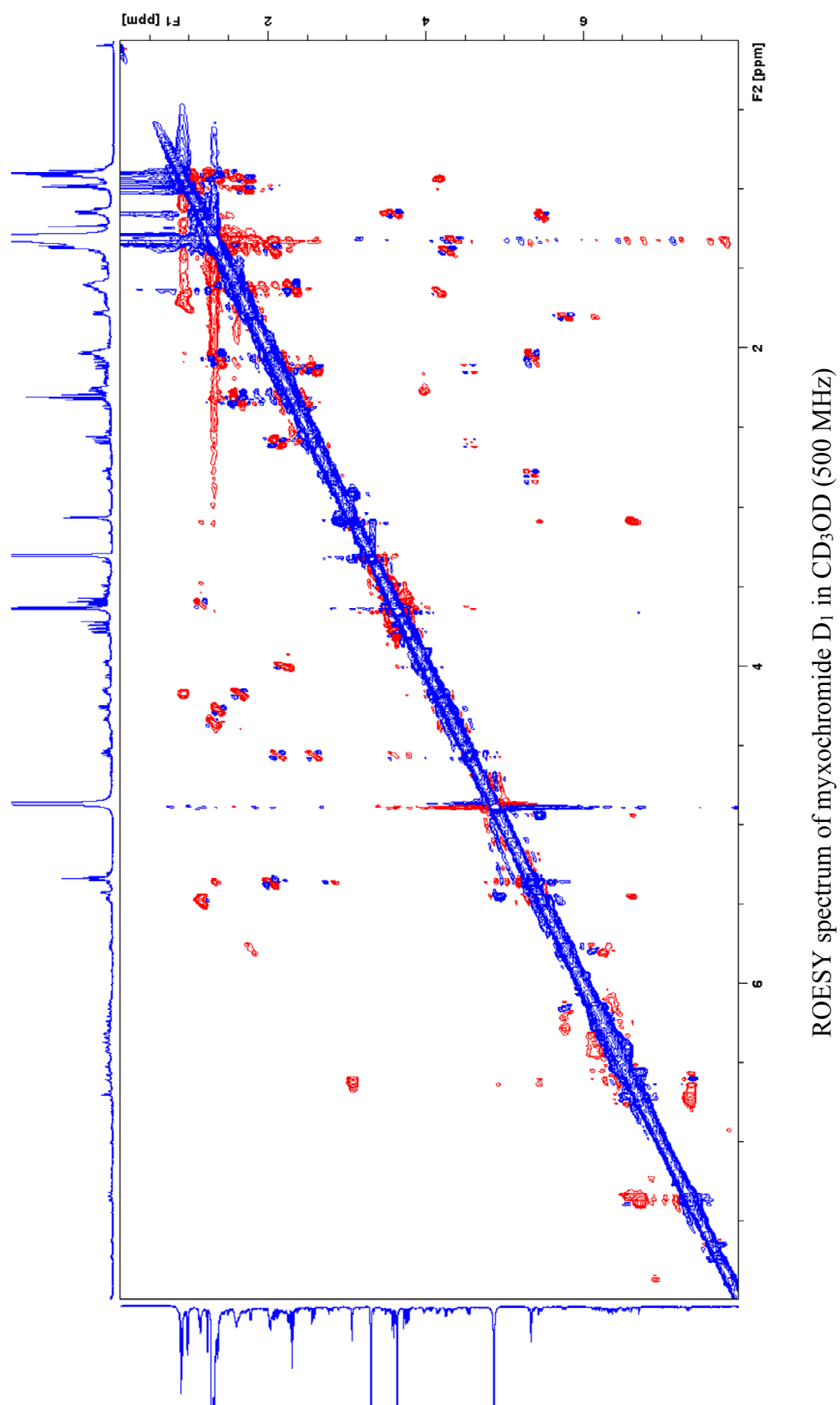


Figure S10. NMR spectra of myxochromide D₁.

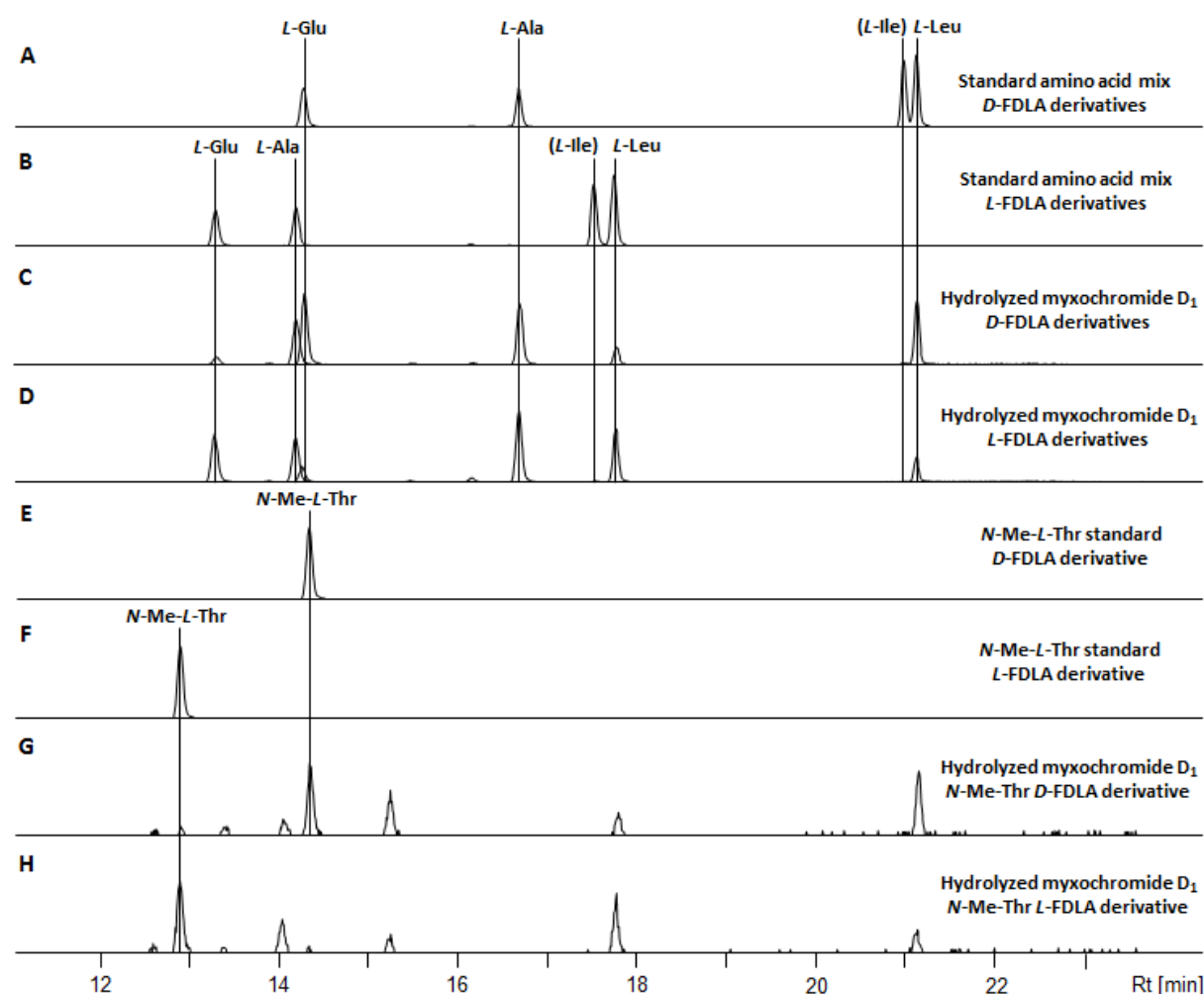


Figure S11. Results of the analysis of the absolute configuration by advanced Marfey's method.²⁹ Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide D₁ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide D₁ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative.

Table S7. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide D₁ (inverse correlation of retention times (t_R) of *D*-configured Ala-FDLA derivatives from the peptide hydrolysate compared to *L*-configured Ala-FDLA standards are shown in bold).

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	t_R [min]	m/z $[M+H]^+$	t_R [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.2	442.1617	14.2	442.1618	L
Glu- <i>L</i> -FDLA	13.2	442.1619	13.2	442.1622	
Ala- <i>D</i> -FDLA	16.6	384.1554	16.6	410.1723	L
Ala- <i>L</i> -FDLA	14.1	384.1561	14.1	384.1563	
Ala- <i>D</i> -FDLA	16.6	384.1554	14.1	384.1561	D
Ala- <i>L</i> -FDLA	14.1	384.1561	16.6	384.1565	
Leu- <i>D</i> -FDLA	21.1	426.2027	21.1	426.2042	L
Leu- <i>L</i> -FDLA	17.7	426.2038	17.7	426.2038	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.3	428.1818	14.3	428.1817	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.8	428.1831	12.8	428.1838	

2.7.5 Isolation and Structure Elucidation of Myxochromide S₂-Abu from the Heterologous Expression Strain *M. xanthus* DK1622::pTpS-mchS

2.7.5.1 Cultivation of *M. xanthus* DK1622::pTpS-mchS and Isolation of Myxochromide S₂-Abu

The myxochromide S overproducing mutant *M. xanthus* DK1622::pTpS-mchS was cultivated in 9 L (6x 1.5 L) CTT medium amended with kanamycin (50 mg/L) including 2% XAD-16 adsorber resin for 5 days at 30 °C and 180 rpm.¹⁵ Cell harvest, extraction and isolation of the target compound myxochromide S₂-Abu using size exclusion chromatography was achieved as described for myxochromide C₃ (see chapter 2.7.3.1). Semi-preparative HPLC was performed on a Dionex UltiMate 3000 system equipped with a Kinetex 5u Biphenyl 100Å column (250 × 10 mm, Phenomenex). At constant flow rate (5.0 mL/min), the following multi-step gradient was applied (A: deionized water, B: acetonitrile): 0-1 min 1% B, 1-8 min 1-40% B, 8-52 min 40-57% B, 52-54 min 57-95% B, 54-57 min 95% B, 57-58 min 95-1% B, 58-62 min 1% B. UV traces were recorded by a diode array detector (DAD) with specified wave lengths (210, 266 and 410 nm) with myxochromides showing good UV absorption at 410 nm. A total amount of 5.3 mg of pure myxochromide S₂-Abu (*R*_t = 37.1 min) was isolated.

2.7.5.2 Structure Elucidation of Myxochromide S₂-Abu

Structure elucidation of myxochromide S₂-Abu was achieved as described for myxochromide C₃ (see chapter 2.7.3.2). 1D ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra are shown in Figure S13. Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S8. HR-ESI-MS of myxochromide S₂-Abu gave a quasimolecular ion at *m/z* 751.43940 [M+H]⁺ ascribable to a molecular formula C₄₀H₅₈N₆O₈ (751.43889, calculated for C₄₀H₅₉N₆O₈, Δ*m/z* 0.679 ppm). The COSY spectrum supported by HSQC and HMBC data showed presence of spin systems corresponding to *N*-Me-threonine, glutamine, alanine and leucine residues as well as an unusual amino acid (α-aminobutyric acid, Abu) and an polyene side chain. Amino acid sequence was established by means of key HMBC correlations and final structure was elucidated as shown in Figure S12. For the assignment of the absolute configuration of myxochromide S₂-Abu the same procedure, hydrolysis and Marfey analysis of the obtained amino acids,²⁹ was applied as described for myxochromide C₃ (see chapter 2.7.3.2). The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S14 and stereochemical assignments are illustrated in Table S9. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that all amino acids of the myxochromide S₂-Abu peptide core show *L*-configuration, while glutamine was converted to glutamic acid during hydrolysis.

This correlates with the major product, myxochromides S,⁹ as well as with myxochromide S₂-diABu produced by the same pathway (see chapter 2.7.6). Despite the presence of an epimerization domain in the second module of the assembly line, *L*-configured amino acids are incorporated into this position of the peptide core.

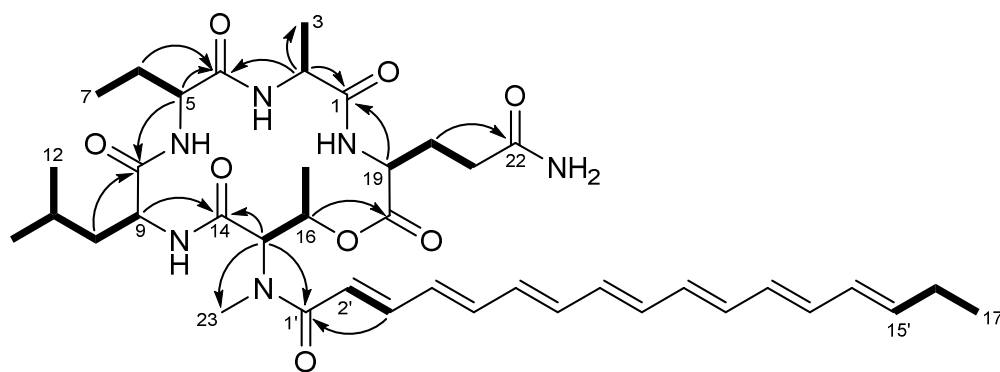


Figure S12. Structure of myxochromide S₂-Abu showing selected COSY (bold line) and key HMBC (arrow) correlations.

Table S8. NMR spectroscopic data of myxochromide S₂-Abu (CD₃OD).

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c	ROESY ^d
<i>L</i> -Ala	1	173.2			
	2	53.1	3.77, q (7.4)	1, 3, 4	
	3	15.4	1.65, d (7.3)	1, 2	
<i>L</i> -Abu	4	173.3			
	5	58.2	3.52, dd (3.9, 10.7)	4, 6, 7, 8	
	6a	21.9	1.85, m	4, 5, 7	
	6b		2.03, m	5, 7	
	7	11.3	0.92, d (7.3)	5, 6, 8, 9	
<i>L</i> -Leu	8	176.2			
	9	54.3	4.07, m	8, 10, 11, 14	
	10a	40.8	1.47, m	8, 9, 13	
	10b		1.60, m	9, 13	
	11	25.6	1.75, m	10, 12, 13	
	12	22.2	0.97, d (6.6)	10, 11, 13	
	13	22.9	1.01, d (6.6)	10, 11, 12	
<i>N</i> -Me- <i>L</i> -Thr	14	171.4			
	15	59.6	5.56, d (4.2)	1', 14, 16, 23	2', 23
	16	73.9	5.52, m	17, 18	
	17	16.5	1.25, d (6.3)	15, 16	
	23	35.1	3.26, s	1', 15	2', 15, 17, 19
<i>L</i> -Gln	18	171.0			
	19	51.9	4.66, dd (3.6, 9.8)	1, 18, 20, 21	
	20a	27.8	1.92, m	18, 19, 21, 22	
	20b		2.31, m		
	21a	31.9	2.31, m	19, 20, 22	
	21b		2.25, m		
	22	178.2			
Side chain	1'	170.6			
	2'	119.8	6.58, d (14.6)	1'	23
	3'	145.4	7.30, dd (11.3, 14.6)	1', 2', 5'	
	4'	138.6	6.52, m	<i>e</i>	
	5'	142.0	6.69, dd (11.1, 14.6)	<i>e</i>	
	6'-12'	<i>e</i>	<i>e</i>	<i>e</i>	
	13'	135.5	6.26, m	<i>e</i>	
	14'	131.0	6.12, m	13', 16'	
	15'	131.0	5.80, dt (6.7, 15.0)	13', 16', 17'	
	16'	26.7	2.14, m	17'	
	17'	13.7	1.02, t (7.5)	16'	

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d proton showing ROESY correlations to indicated protons.^e overlapped signals.

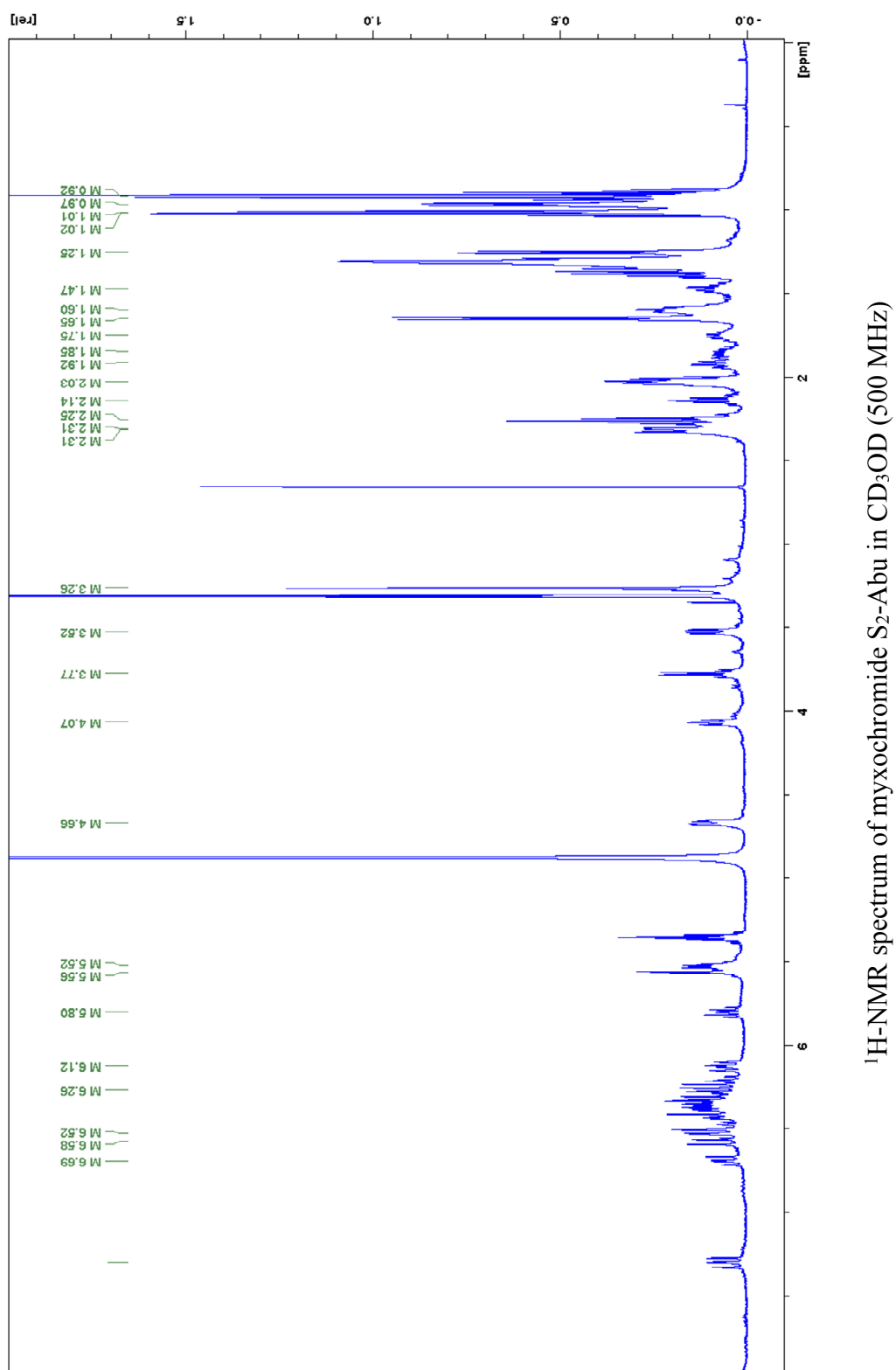


Figure S13 (continued on next page)

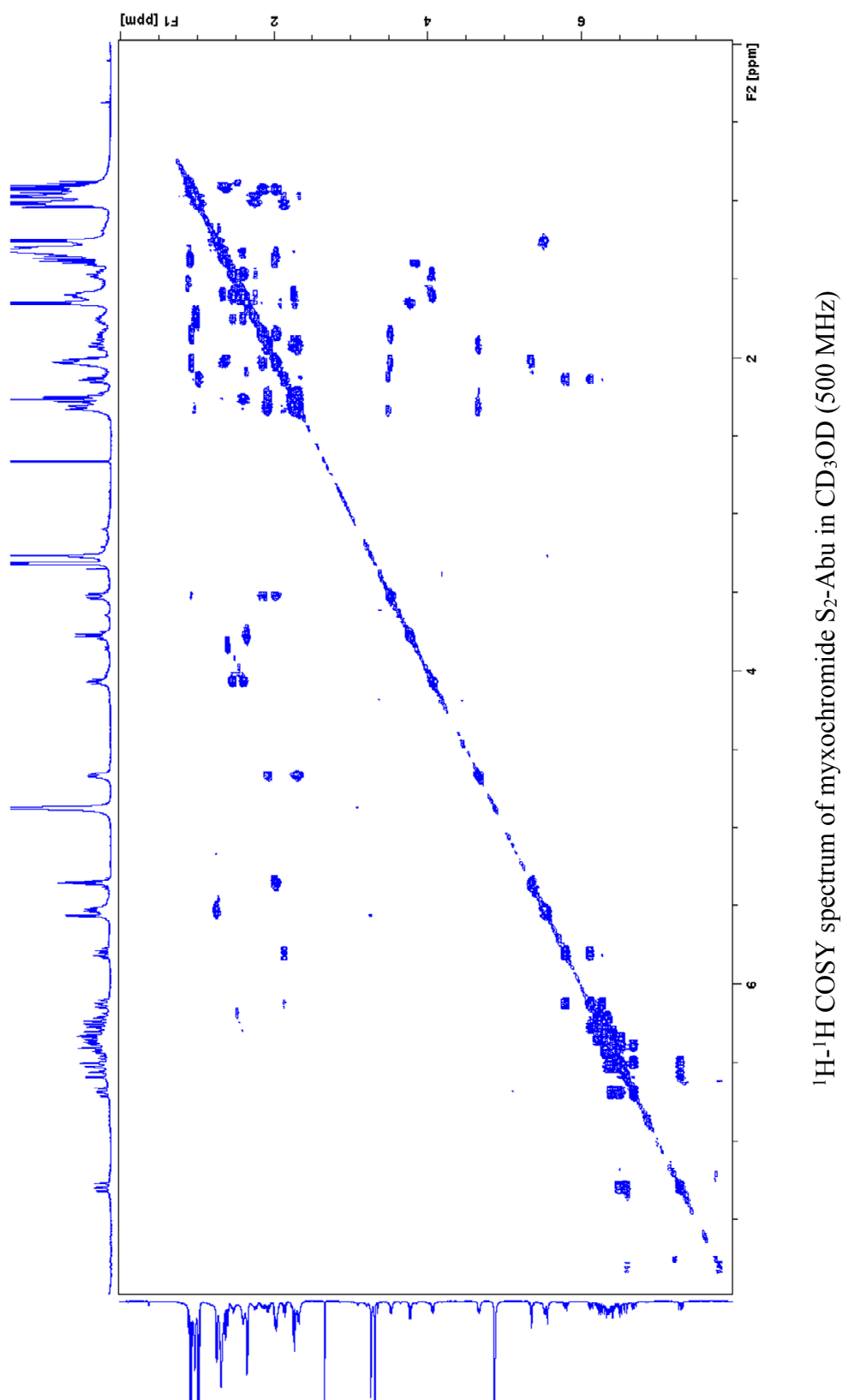
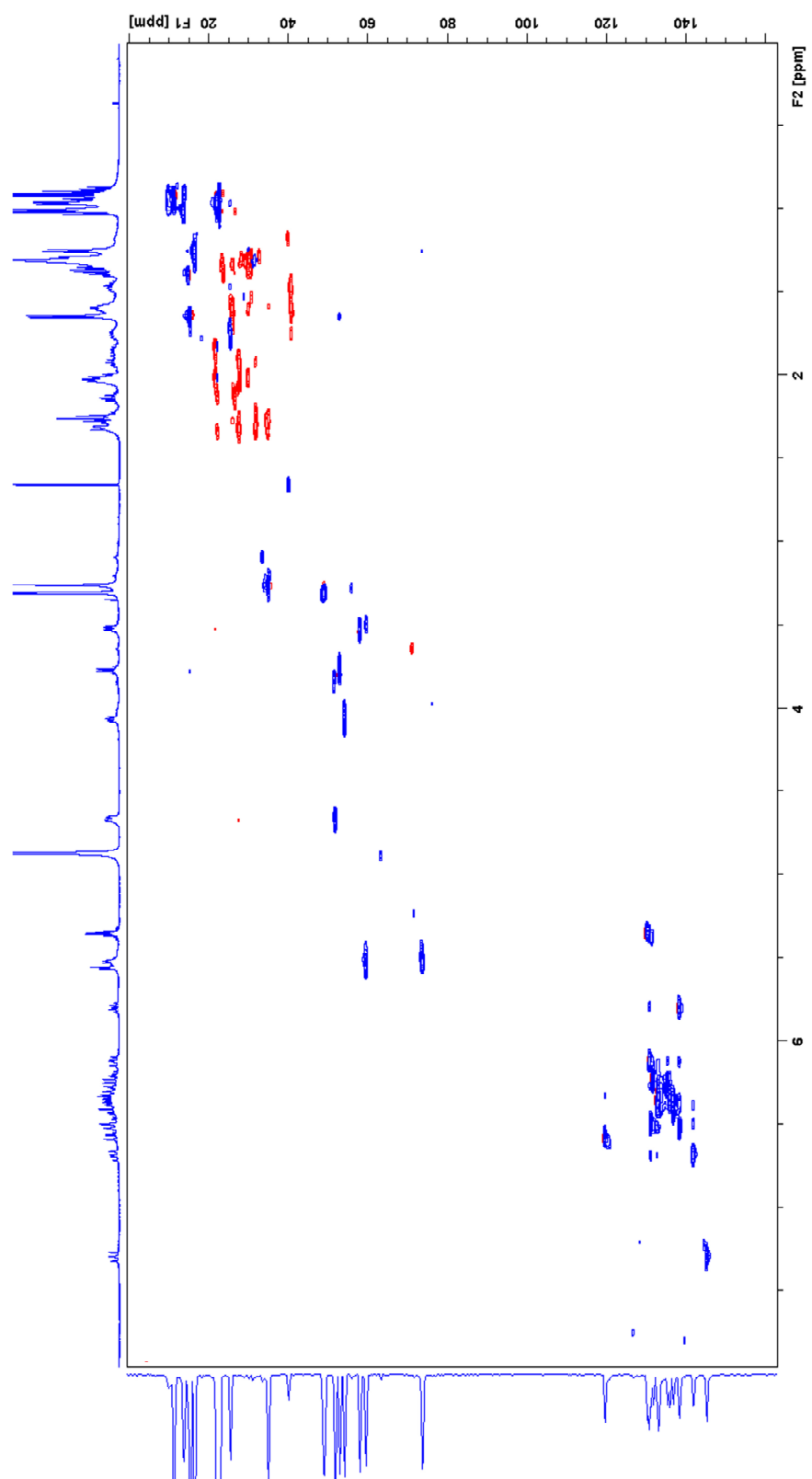


Figure S13 (continued on next page)

HSQC spectrum of myxochromide S₂-Abu in CD₃OD (500 MHz)

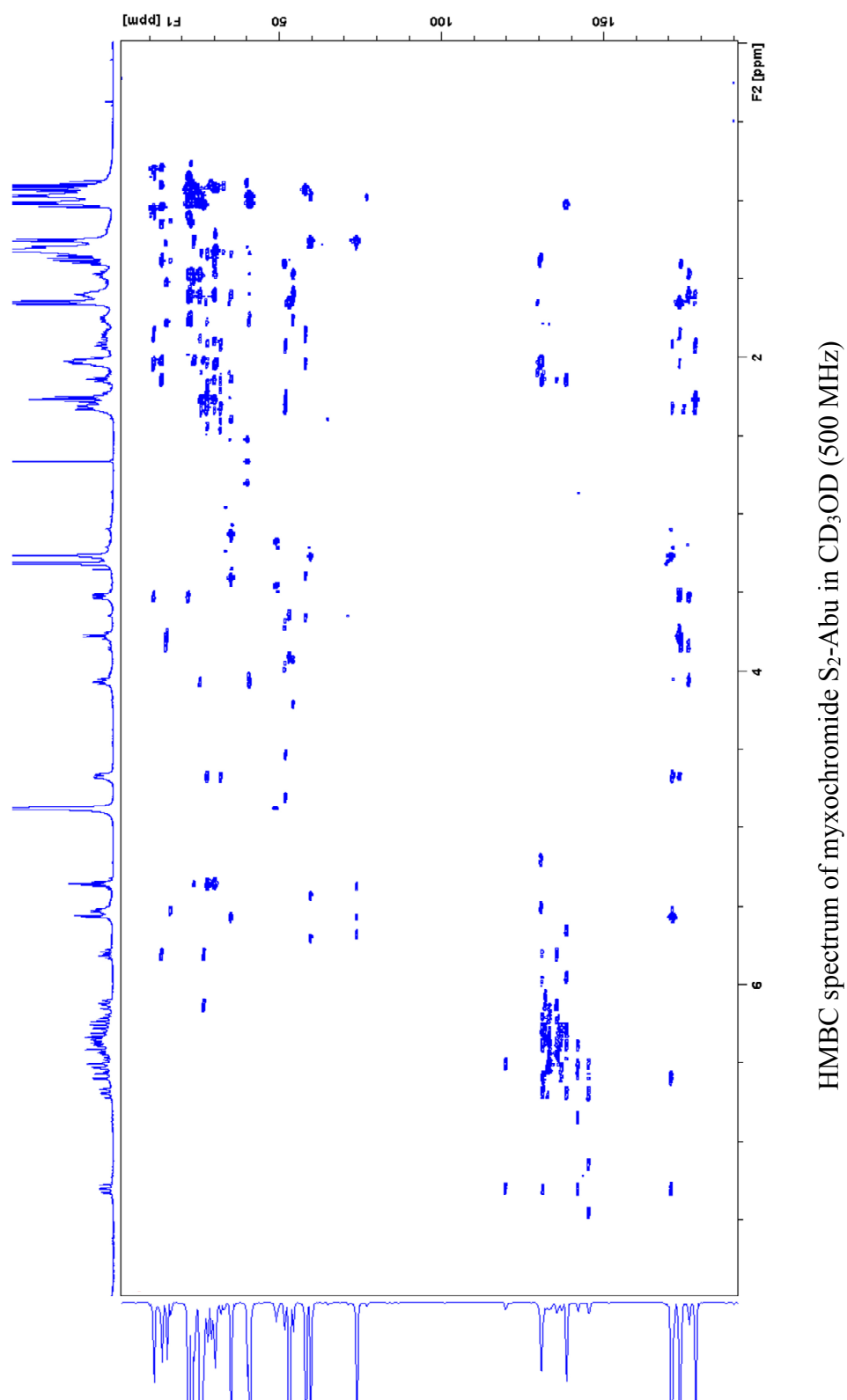


Figure S13 (continued on next page)

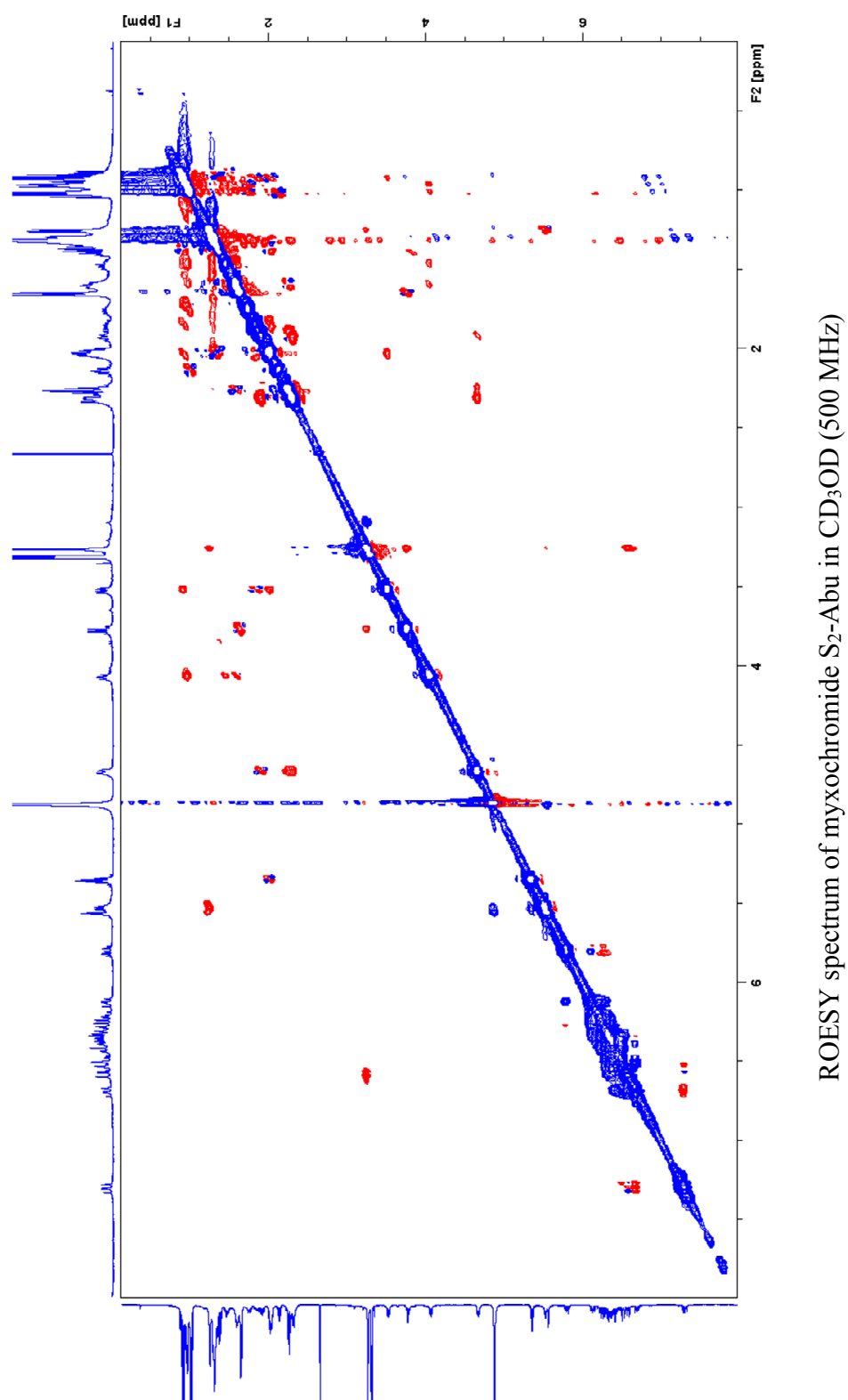


Figure S13. NMR spectra of myxochromide S₂-Abu.

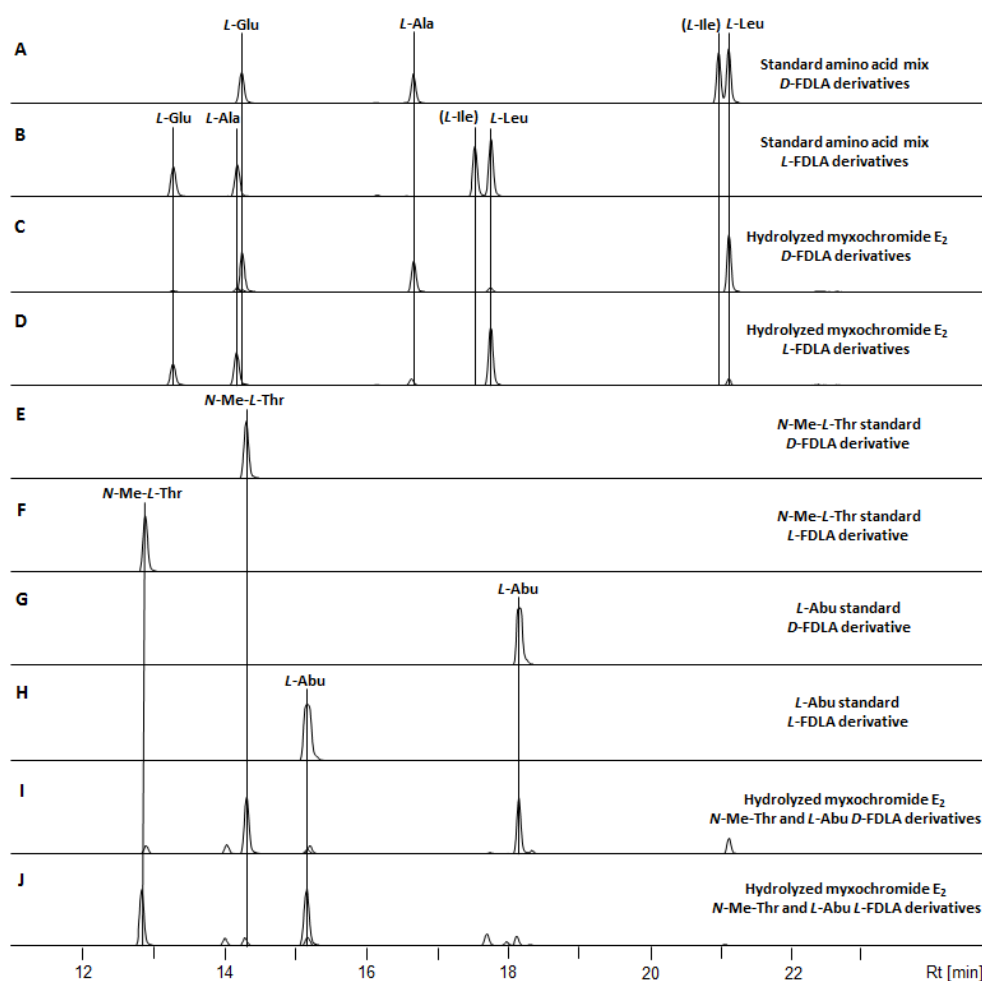


Figure S14. Results of the analysis of the absolute configuration by advanced Marfey's method.²⁹ Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide S₂-Abu derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide S₂-Abu derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Standard solution of *L*-aminobutyric acid derivatized with *D*-FDLA. **H:** Standard solution of *L*-aminobutyric acid derivatized with *L*-FDLA. **I:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative and for the *L*-aminobutyric acid *D*-FDLA derivative. **J:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative and for the *L*-aminobutyric acid *L*-FDLA derivative.

Table S9. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide S₂-Abu.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	<i>t_R</i> [min]	<i>m/z</i> [M+H] ⁺	<i>t_R</i> [min]	<i>m/z</i> [M+H] ⁺	
Glu- <i>D</i> -FDLA	14.2	442.1617	14.2	442.1615	L
Glu- <i>L</i> -FDLA	13.2	442.1619	13.2	442.1615	
Abu- <i>D</i> -FDLA	18.2	398.1740	18.2	398.1730	L
Abu- <i>L</i> -FDLA	15.2	398.1739	15.2	398.1725	
Ala- <i>D</i> -FDLA	16.6	384.1554	16.6	384.1568	L
Ala- <i>L</i> -FDLA	14.1	384.1561	14.1	384.1568	
Leu- <i>D</i> -FDLA	21.1	426.2027	21.1	426.2040	L
Leu- <i>L</i> -FDLA	17.7	426.2038	17.7	426.2038	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.3	428.1818	14.3	428.1826	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.8	428.1831	12.8	428.1828	

2.7.6 Isolation and Structure Elucidation of Myxochromide S₂-diAbu from the Heterologous Expression Strain *M. xanthus* DK1622::pTpS-mchS

2.7.6.1 Cultivation of *M. xanthus* DK1622::pTpS-mchS and Isolation of Myxochromide S₂-diAbu

The myxochromide S overproducing mutant *M. xanthus* DK1622::pTpS-mchS was cultivated in 4.5 L (3x 1.5 L) CTT medium amended with kanamycin (50 mg/L) including 2% XAD-16 adsorber resin for 4 days at 30 °C and 180 rpm.¹⁵ After 6, 12, 24, 30, 36, 48, 54, 60, 66 and 72 h, 8.5 mg of *L*- α -aminobutyric acid were added to the cultures. Cell harvest and extraction of the target compound myxochromide S₂-diAbu was performed as described for myxochromide C₃ (see chapter 2.7.3.1) and isolation was achieved using semi-preparative HPLC as described for myxochromide S₂-Abu (see chapter 2.7.5.1). A total amount of 6.4 mg of pure myxochromide S₂-diAbu (R_t = 39.7 min) was isolated.

2.7.6.2 Structure Elucidation of Myxochromide S₂-diAbu

Structure elucidation of myxochromide S₂-diAbu was achieved as described for myxochromide C₃ (see chapter 2.7.3.2). 1D ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra are shown in Figure S16. Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S10. HR-ESI-MS of myxochromide S₂-diAbu displayed a quasimolecular ion at m/z 765.45520 [$M+H$]⁺ which is 14 Da heavier than that of myxochromide S₂-Abu consistent with a molecular formula C₄₁H₆₀N₆O₈ (765.45454, calculated for C₄₁H₆₁N₆O₈, $\Delta m/z$ 0.862 ppm). Analysis of 2D NMR data revealed the presence of two Abu spin systems, one of which was replacing the alanine residue in myxochromide S₂-Abu. Consequently, its final structure was established as depicted in Figure S15.

For the assignment of the absolute configuration of myxochromide S₂-diAbu the same procedure, hydrolysis and Marfey analysis of the obtained amino acids,²⁹ was applied as described for myxochromide C₃ (see chapter 2.7.3.2). The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S17 and stereochemical assignments are illustrated in Table S11. Comparison of the retention times and masses of derivatized amino acids and the hydrolyzed lipopeptide revealed that all amino acids of the myxochromide S₂-diAbu peptide core show *L*-configuration, while glutamine was converted to glutamic acid during hydrolysis. This correlates with the major product, myxochromides S,⁹ as well as with myxochromide S₂-Abu produced by the same pathway. Despite the presence of an epimerization domain in the second module of the assembly line, *L*-configured amino acids are incorporated into this position of the peptide core.

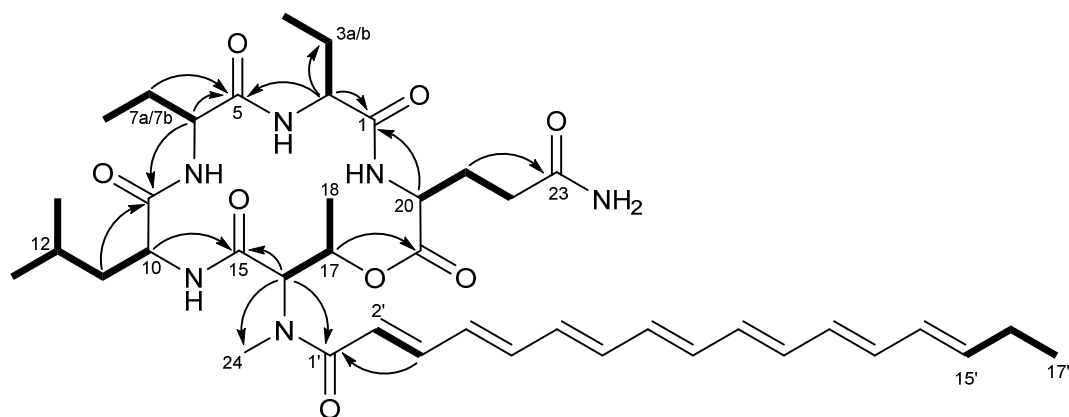


Figure S15. Structure of myxochromide S₂-diAbu showing selected COSY (bold line) and key HMBC (arrow) correlations.

Table S10. NMR spectroscopic data of myxochromide S₂-diAbu (CD₃OD).

Moiety	Position	δ_C^a	δ_H^b (<i>J</i> in Hz)	HMBC ^c	ROESY ^d
<i>L</i> -Abu (1)	1	172.8			
	2	59.8	3.50, dd (4.3, 11.4)	1, 3, 4, 5	
	3a	22.3	2.12, m	1, 2, 4	
	3b		2.33, m	2, 4	
	4	11.5	0.96, t (7.5)	2, 3	
<i>L</i> -Abu (2)	5	173.3			
	6	58.0	3.56, dd (4.1, 10.7)	5, 7, 8, 9	
	7a	21.9	1.86, m	5, 6, 8	
	7b		2.03, m	6, 8	
	8	11.4	0.93, t (7.5)	6, 7	
<i>L</i> -Leu	9	176.2			
	10	54.3	4.08, dd (6.1, 8.4)	9, 11, 12, 15	
	11a	40.8	1.47, m	9, 10, 12, 13, 14	
	11b		1.61, m	9, 10, 12, 13, 14	
	12	25.6	1.75, m	10, 11, 13, 14	
	13	22.1	0.97, d (6.5)	11, 12	
	14	22.9	1.01, d (6.5)	11, 12	
	15	171.4			
<i>N</i> -Me- <i>L</i> -Thr	16	59.6	5.57, d (4.2)	1', 15, 17, 24	2', 24
	17	73.8	5.53, m	18, 19	
	18	16.4	1.25, d (6.3)	16, 17	
	24	35.1	3.26, s	1', 16	2', 18
	19	171.1			
<i>L</i> -Gln	20	51.8	4.67, dd (3.6, 9.8)	1, 19, 21, 22	21
	21a	27.6	1.92, m	19, 20, 22, 23	
	21b		2.31, m	20, 22, 23	
	22a	32.0	2.25, m	21, 22	
	22b		2.30, m	21, 22	
	23	178.2			
	24	170.7			
Side chain	1'	170.7			
	2'	119.7	6.58, d (14.5)	1'	24
	3'	145.4	7.30, dd (11.4, 14.5)	1', 2', 5'	
	4'	138.6	6.52, m	<i>e</i>	
	5'	142.0	6.69, dd (11.2, 14.5)	<i>e</i>	
	6'-12'	<i>e</i>	<i>e</i>	<i>e</i>	
	13'	135.5	6.26, m	<i>e</i>	
	14'	131.0	6.12, m	13', 16'	
	15'	131.0	5.80, dt (6.7, 15.0)	13', 16', 17'	
	16'	26.8	2.14, m	17'	
	17'	13.8	1.02, t (7.5)	16'	

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d proton showing ROESY correlations to indicated protons.^e overlapped signals.

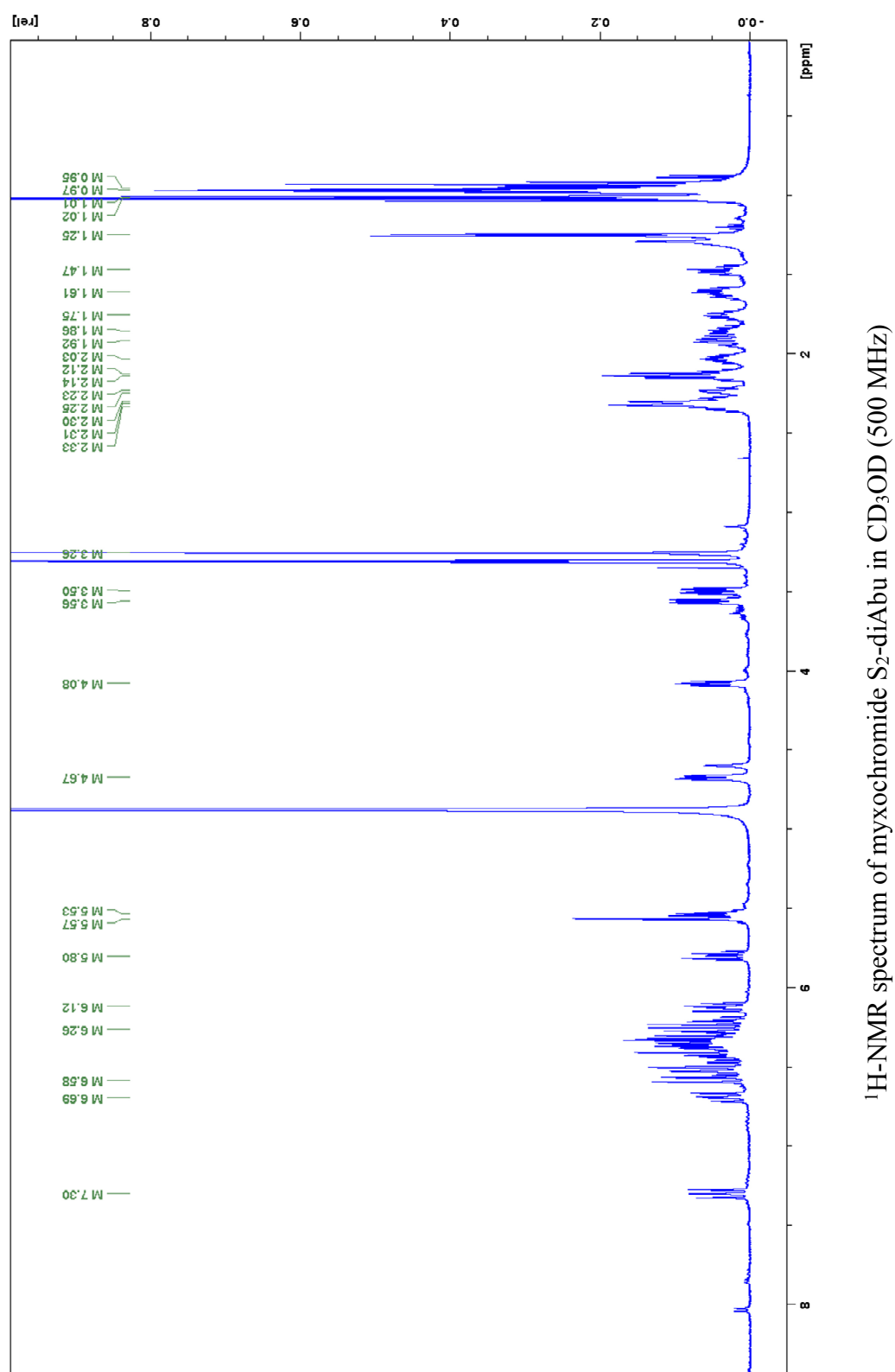


Figure S16 (continued on next page)

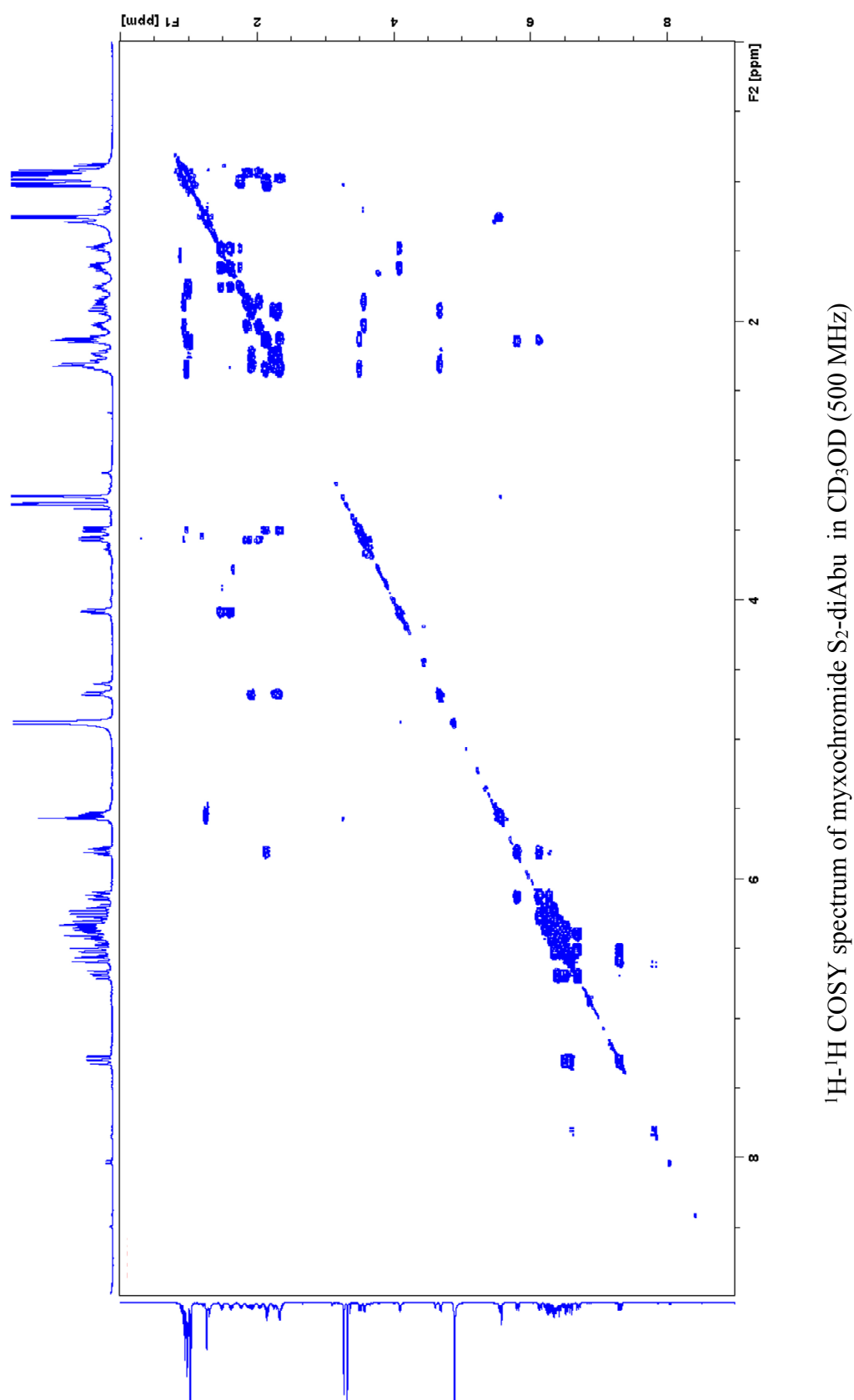


Figure S16 (continued on next page)

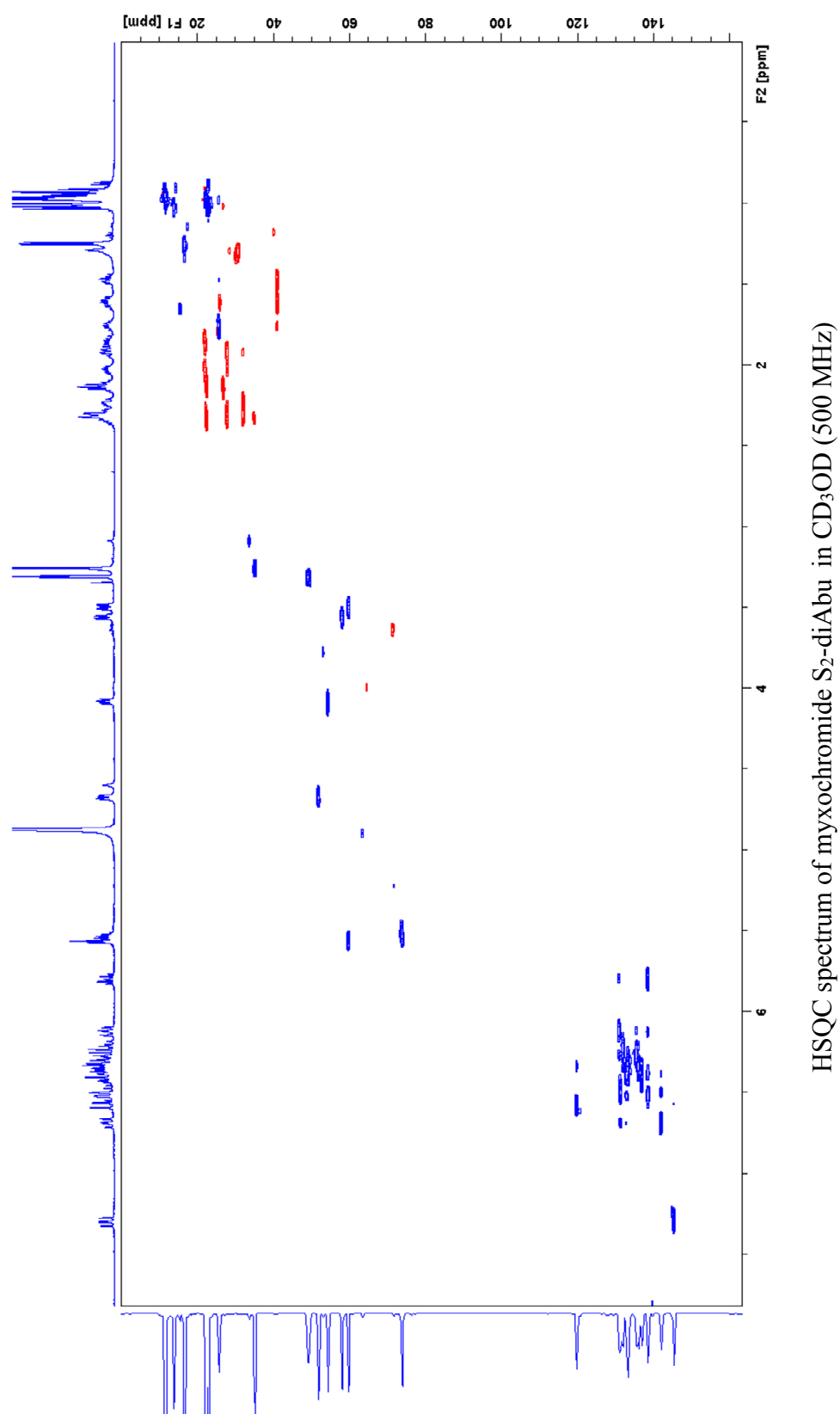


Figure S16 (continued on next page)

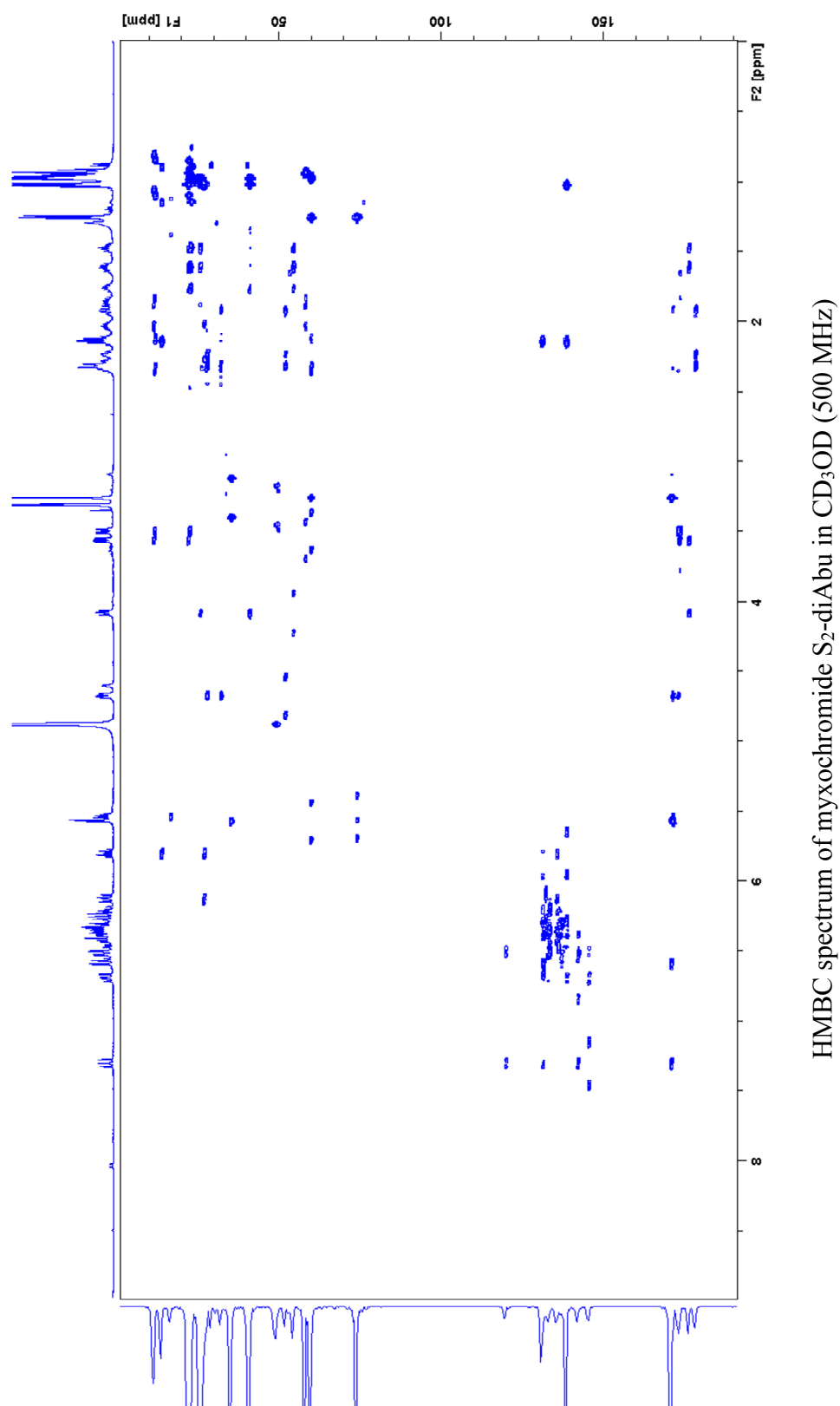
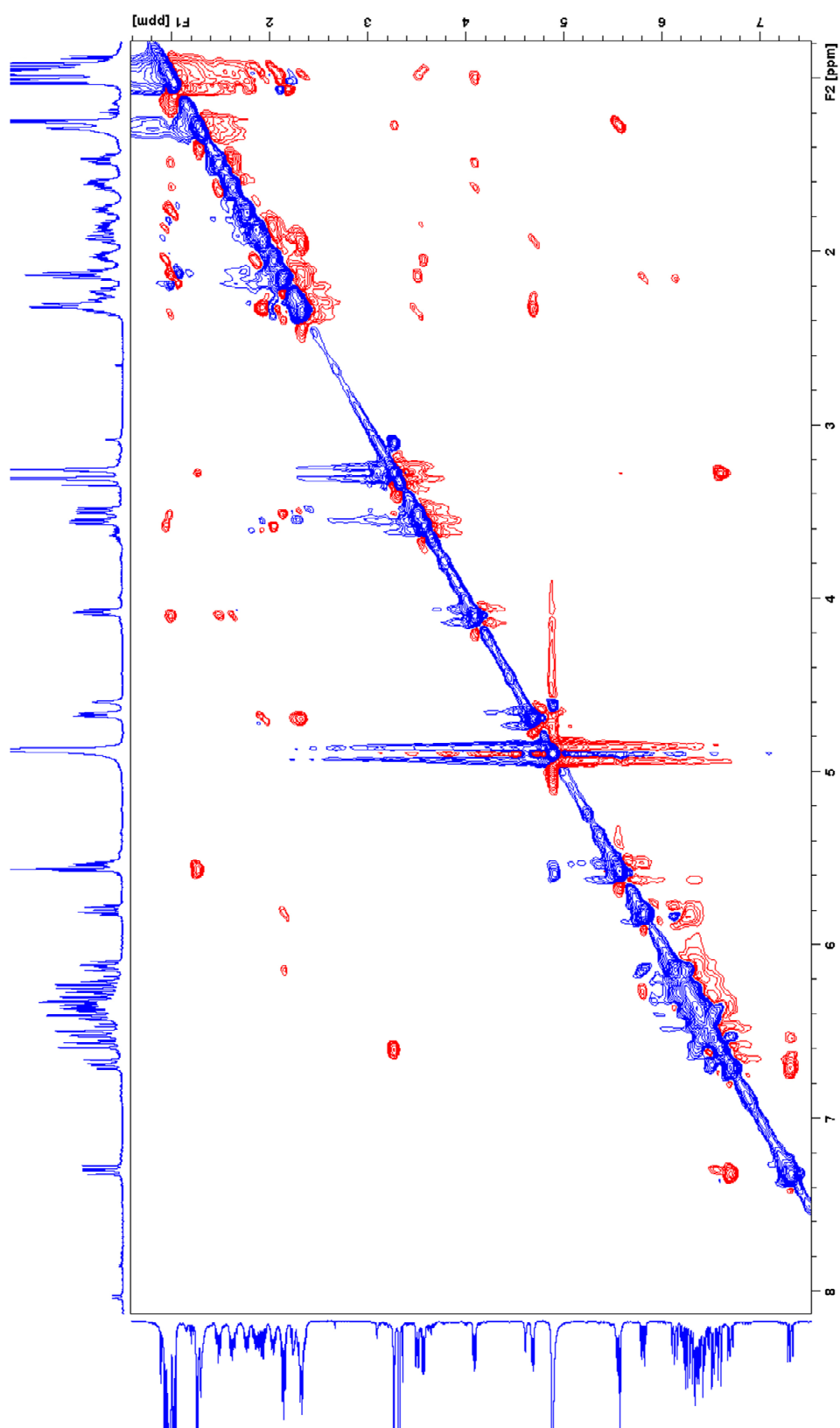


Figure S16 (continued on next page)

ROESY spectrum of myxochromide S₂-diAbu in CD₃OD (500 MHz)**Figure S16.** NMR spectra of myxochromide S₂-diAbu.

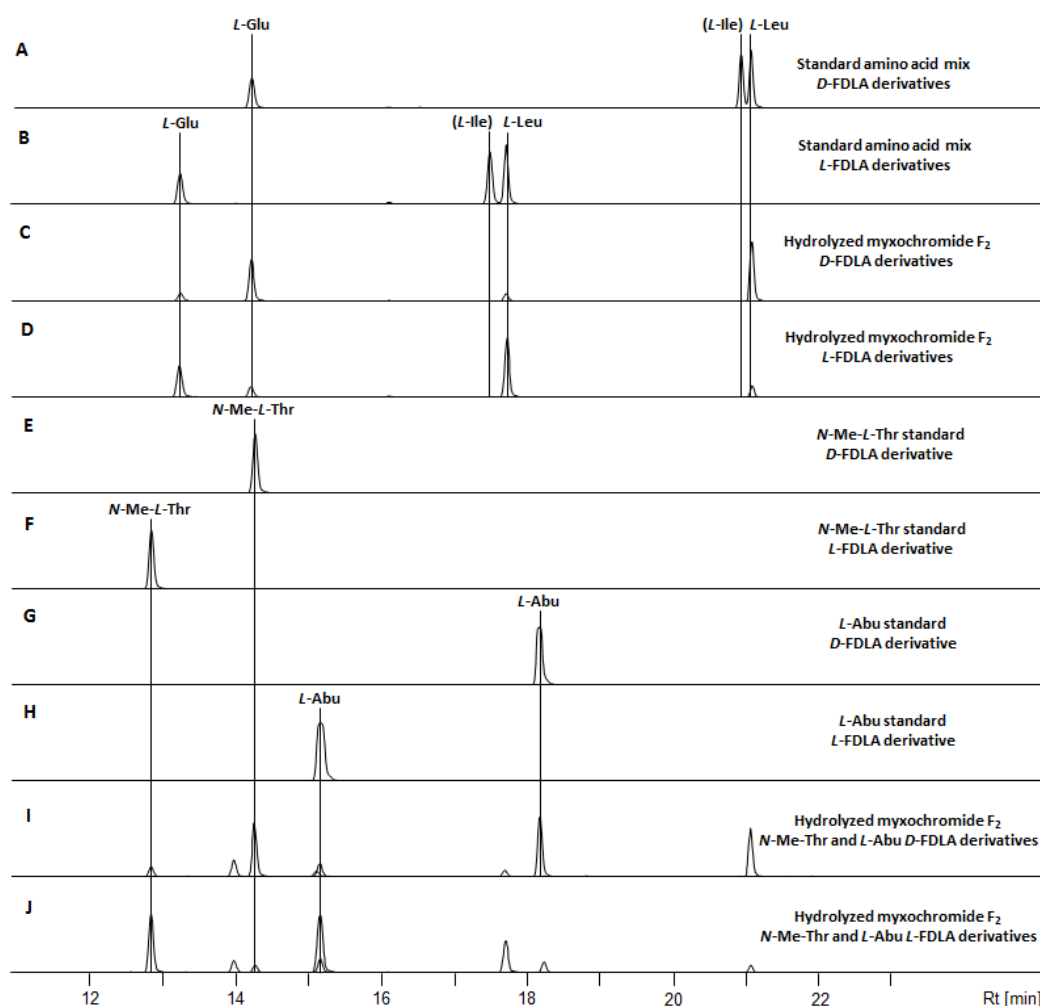


Figure S17. Results of the analysis of the absolute configuration by advanced Marfey's method.²⁹ Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide S_2 -diAbu derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide S_2 -diAbu derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Standard solution of *L*-aminobutyric acid derivatized with *D*-FDLA. **H:** Standard solution of *L*-aminobutyric acid derivatized with *L*-FDLA. **I:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative and for the *L*-aminobutyric acid *D*-FDLA derivative. **J:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative and for the *L*-aminobutyric acid *L*-FDLA derivative.

Table S11. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide S_2 -diAbu.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	tr [min]	m/z $[M+H]^+$	tr [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.2	442.1617	14.2	442.1628	L
Glu- <i>L</i> -FDLA	13.2	442.1619	13.2	442.1619	
Abu- <i>D</i> -FDLA	18.2	398.1740	18.2	398.1723	L
Abu- <i>L</i> -FDLA	15.2	398.1739	15.2	398.1721	
Abu- <i>D</i> -FDLA	18.2	398.1740	18.2	398.1723	L
Abu- <i>L</i> -FDLA	15.2	398.1739	15.2	398.1721	
Leu- <i>D</i> -FDLA	21.1	426.2027	21.1	426.2041	L
Leu- <i>L</i> -FDLA	17.7	426.2038	17.7	426.2037	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.3	428.1818	14.3	428.1822	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.8	428.1831	12.8	428.1825	

2.7.7 Detailed *in silico* Analysis of the 16 Analyzed *mch* Clusters

2.7.7.1 Sequence Analysis of Catalytic Domains of the Encoded Megasynthetases

The applied genome mining approach identified 16 (putative) *mch* clusters encoding seven different (sub)types of myxochromide megasynthetases (Figure 2), which could be correlated to distinct myxochromide production profiles (Figure S3). The detected myxochromide products differ in the number, type, sequence and configuration of the incorporated amino acids (Figure 1). In order to explain the observed structural diversity, catalytic domains of the involved NRPSs were analyzed in detail to evaluate *in silico* their substrate specificities and/or general activity. Protein sequence alignments of selected PCP, E and C domains were performed using the Geneious alignment tool integrated into Geneious software version 9.1.2⁴⁵; see Figure S18. In addition, A domain substrate specificities predicted with antiSMASH 3.0¹³, as well as the corresponding 8Å signatures and Stachelhaus codes^{35,36}, both retrieved with the NRPSpredictor2 analysis tool³⁴, were compared for each set of orthologous A domains (Figure S19). In the following, the *in silico* analysis data are discussed in the light of observed structural differences.

Lack of proline incorporation in myxochromides D and S (S-Abu/S-diAbu)

Myxochromide D and myxochromide S megasynthetases generate lipopentapeptide products lacking the proline residue compared to myxochromides A, B and C. In case of the D-subtype 2 megasynthetase this result can be explained by the absence of two essential catalytic domains (A₄-PCP₄) due to partial module deletion (see chapter 2.7.7.2). In contrast, the S-type and D-subtype 1 megasynthetases still harbor the complete C₄-A₄-PCP₄ domain set. However, detailed sequence analysis of PCP₄ domains revealed a mutated core motif lacking the highly conserved serine residue (Figure S18F). As expected and shown previously for PCP₄ from the Sa1 S-type pathway,⁷ this mutation impairs posttranslational activation by 4-phosphopantetheinyl transferases (PPTases)⁵⁰. Therefore, PCP₄ is not converted into its active *holo* form and not able to bind amino acid substrates or biosynthetic intermediates. Consequently, due to the PCP₄ “loss of function mutation”, a unique NRPS module-skipping process takes place during myxochromides S and D biosynthesis.

Only L-configured amino acids in myxochromides S (S-Abu/S-diAbu)

All analyzed myxochromide megasynthetase types harbor an E domain in module 2, which correlates with the presence of a *D*-configured alanine in myxochromides A, B, C and D as verified for myxochromide A₃ from Mx1,⁷ B₃ from M1,⁸ C₃ from Mv1 and D₁ from Se1. However, lipopentapeptide products from S-type pathways seem to exclusively contain *L*-configured amino acids as shown for myxochromides S₁,⁹ S₂-Abu and S₂-diAbu from the Sa1

megasynthetase. More precisely, they contain a *L*-leucine residue in the second position of the assembled peptide cores, for which a *D*-configured amino acid would be predicted by sequence analysis. According to textbook biochemistry, after chain elongation with *L*-leucine the generated PCP₂-bound dipeptidyl intermediate is expected to be epimerized by the action of E₂ and only *D*-configured intermediates are expected to be processed by the downstream C₃ domain. Such C domains, which are *D*-specific for the peptidyl donor and *L*-specific for the aminoacyl acceptor are described as so-called ^DC_L domains.⁵¹ Interestingly, although C₃ domains from S-type megasynthetases feature the sequence motifs reported for ^DC_L domains (Figure S18B;³³) only *L*-configured ‘leucine intermediates’ (peptidyl donors) seem to be processed. Whether ‘*D*-leucine intermediates’ are actually provided by S-type MchB subunits remains uncertain. Sequence analysis revealed that PCP₂ domains show the characteristic core motif of specialized PCP^E domains (Figure S41A), which are required as partner for E domains to be active.³² Sequence comparison of the E₂ domains from S-type pathways with (certainly functional) E₂ domains from A-, B-, C- and D-type megasynthetases revealed a few deviations from the reported E domain core motifs (core E1-E7;³¹; see Figure S18C), which probably have no significant effect on activity.⁵² In case of E domains from Sa1 and Sa2 S-type producer strains, these also include a mutation of one of the two key active site residues described from structural studies on the E domain of tyrocidine synthetase A (His743 and Glu882; ⁵³: The glutamate from core motif E4, which is assumed to act as catalytic base during the epimerization process, is replaced with an aspartate, which still might fulfill the same function. In terms of codon usage adaptation, E domains of S-type producers (Sa1-3) are compared to surrounding functional domains not obviously lower adapted (Figure S25B), in contrast to the lower CAI values of the inactive ER* domain in *mchA* (Figure S25A). In summary, based on the present *in silico* data, we speculate that E domains from S-type megasynthetases are possibly still active, but might not exert their function on the more bulky leucine residue. Alternatively, the downstream C₃ domain is not able to process *D*-configured ‘leucine intermediates’ (peptidyl donors). This finding illustrates a limitation of sequence-based structure prediction approaches, which would point to the epimerization and thus incorporation of a *D*-configured amino acid during myxochromide S (S-Abu/S-diAbu) biosynthesis.

Different amino acid sequence in myxochromide S and aminobutyric acid containing derivatives (myxochromides S-Abu/S-diAbu)

Whereas myxochromides A, B, C and D all harbor *D*-Ala/*L*-Leu in the module 2/3 derived peptide core positions, myxochromides S consist of *L*-Leu/*L*-Ala indicating that substrate specificities of modules 2 and 3 have switched. Based on phylogenetic analysis, intragenic A domain swapping, e.g. as proposed for the mycosubtilin/iturin A operons,¹⁷ could be excluded as the molecular reason for this phenomenon (Figure S23A). *In silico* analysis of residues from the A domain substrate binding pocket (8Å signature and Stachelhaus code)^{35,36} clearly illustrate that A₂/A₃ from S-type pathways deviate from orthologous A domains of other myxochromide megasynthetases (Figure S19). Although no distinct substrate specificities could be predicted by the applied analysis tools, these point mutations (and probably additional mutations in the involved C domains) seem to cause the observed structural variation. Additionally identified myxochromide S-Abu and S-diAbu derivatives showed that modules 3 and 5 of the S-type megasynthetases can also incorporate α -aminobutyric acid besides alanine. A similar substrate tolerance, which could not be predicted from the *in silico* specificity analysis, is assumed for A₂/A₅ from A- or D-type, A₂/A₆ from B-type and A₂ from C-type pathways based on the detection of novel myxochromide derivatives in feeding experiments with α -aminobutyric acid (Figure S5/Table S3). Substrate specificities of threonine, proline and glutamine incorporating modules correlate well with the *in silico* predictions for all myxochromide pathways (Figure S19).

Overall, the detailed sequence analysis of catalytic NRPS domains (C, A, PCP, E) demonstrated how point mutations during natural gene cluster evolution lead to structural diversity of the biosynthesis products.

A

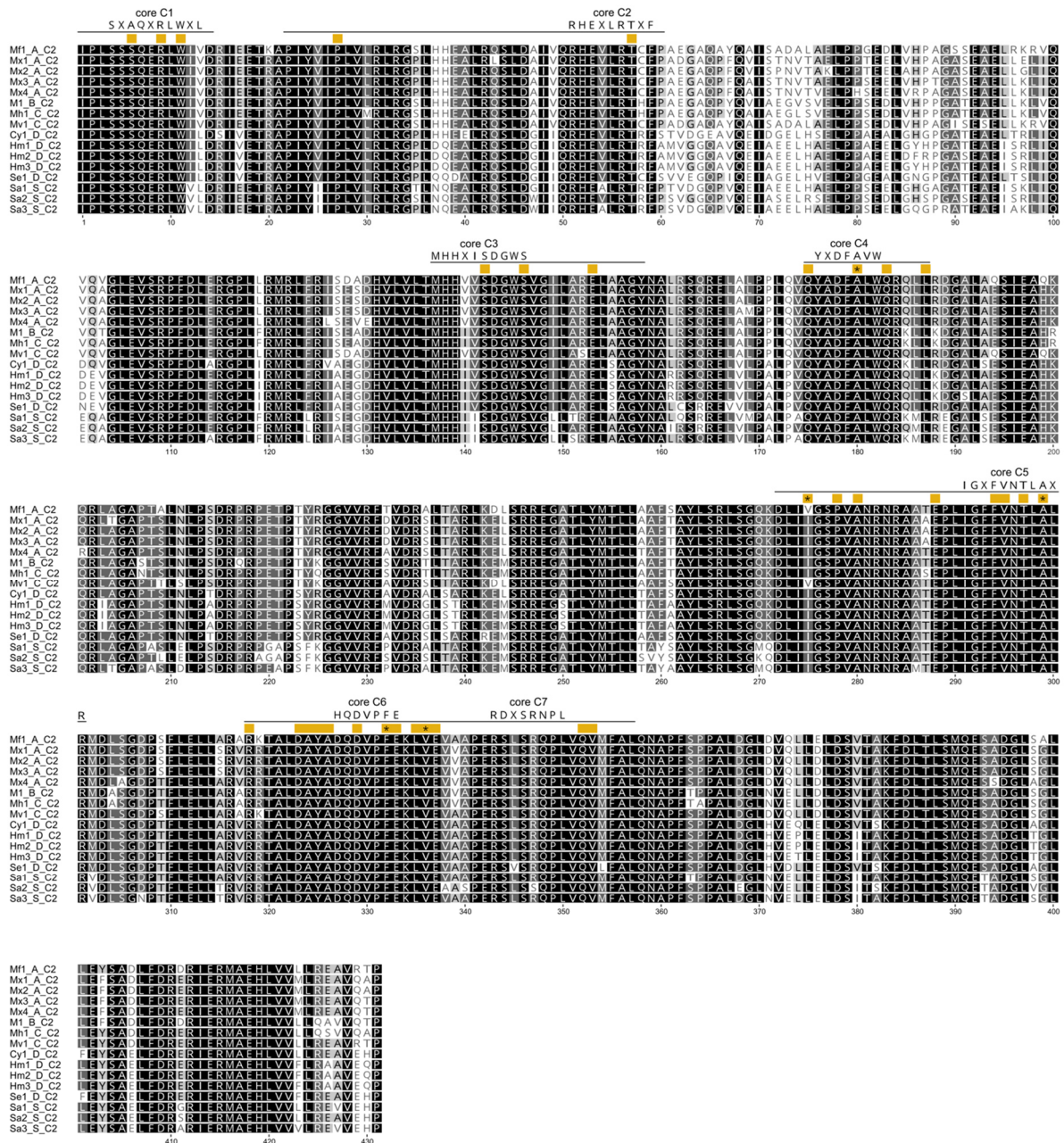


Figure S18 (continued on next page)

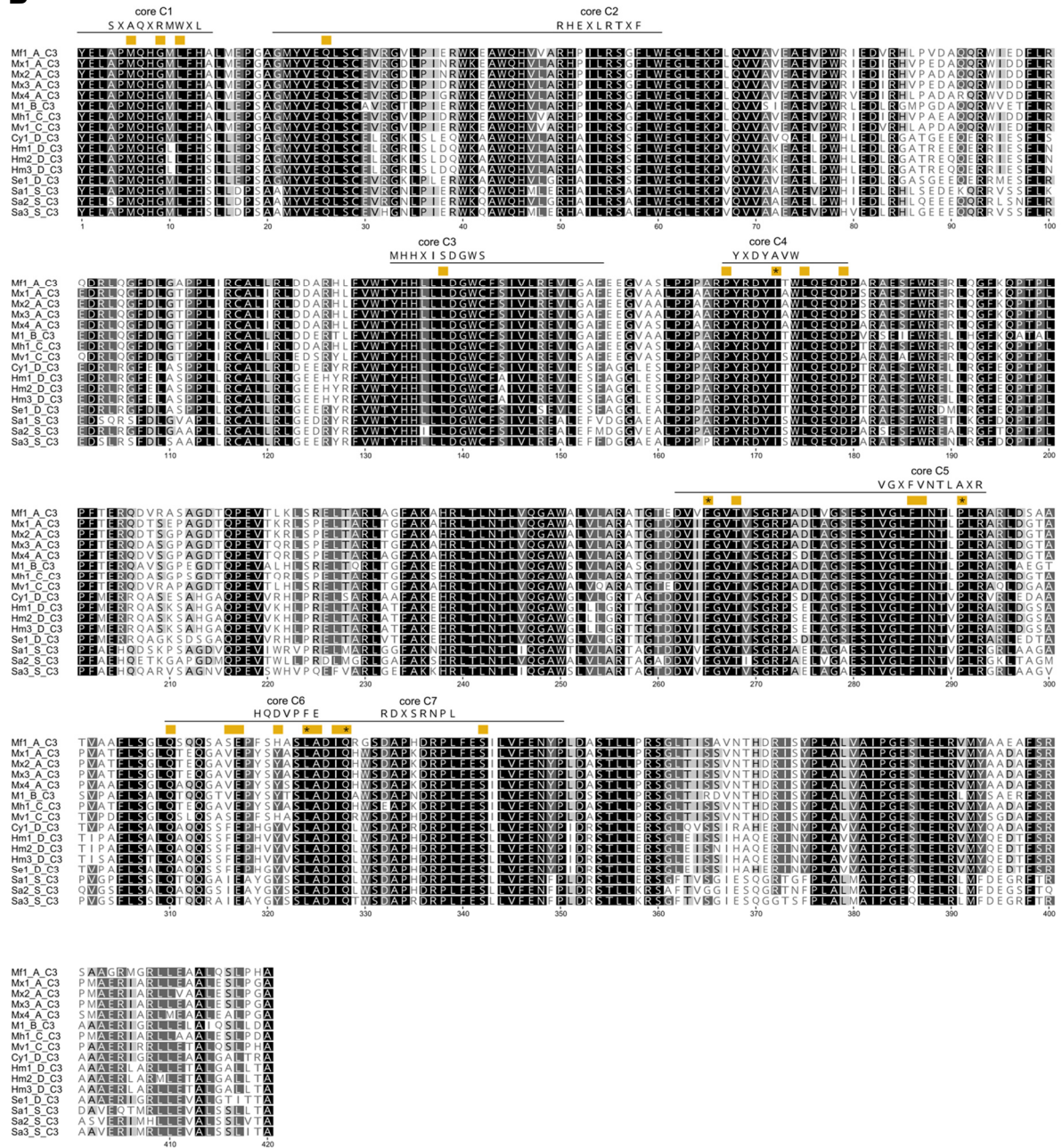
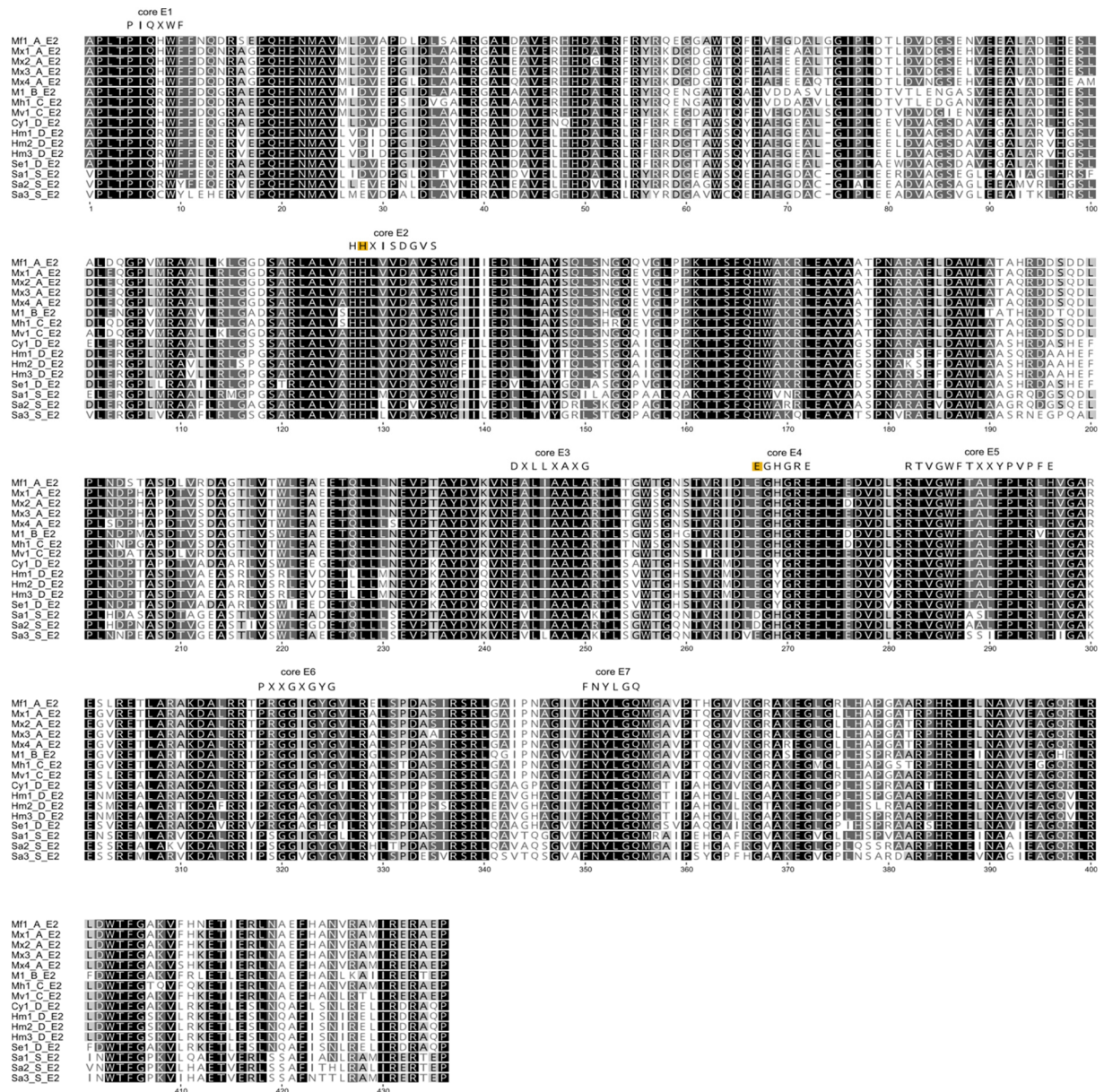
B

Figure S18 (continued on next page)

C



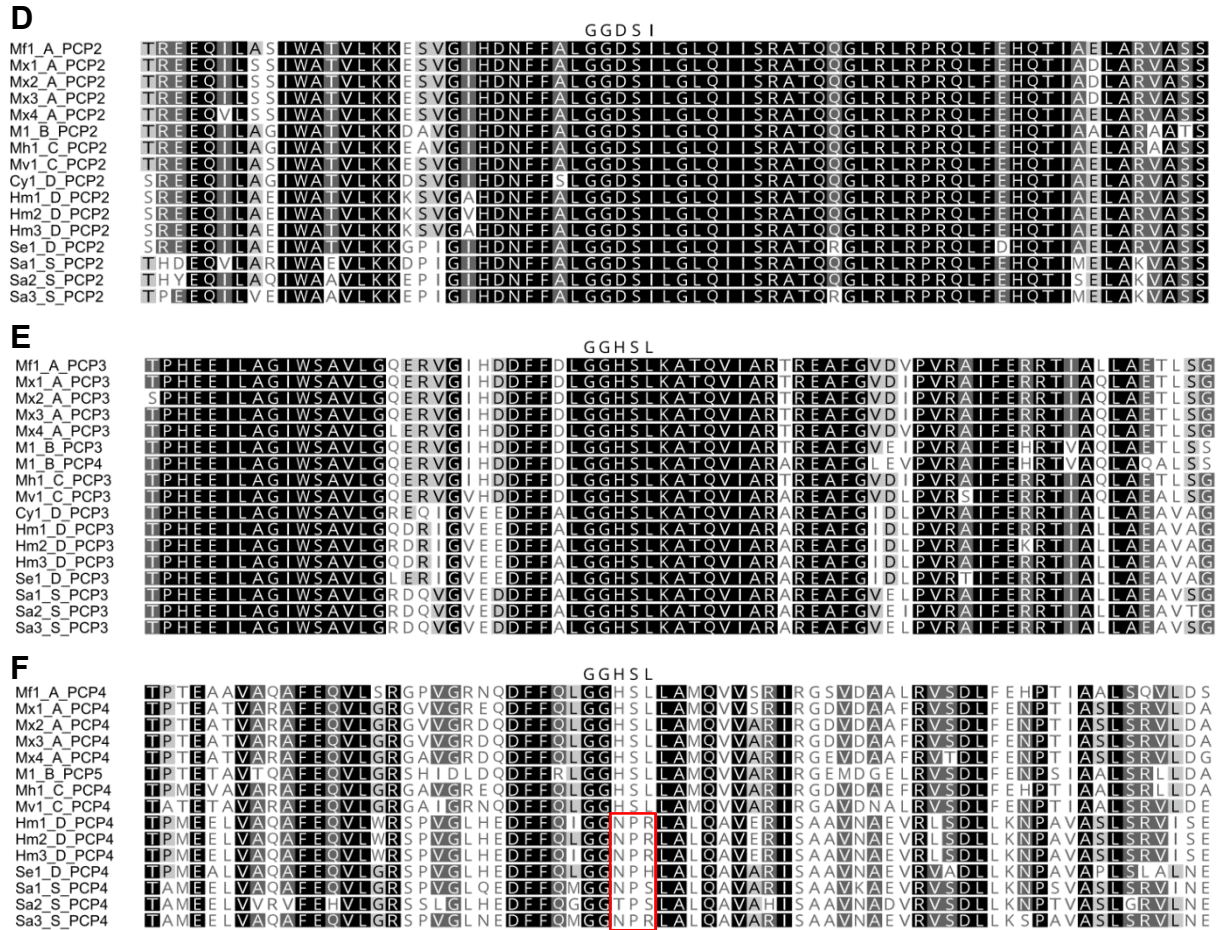


Figure S18. Sequence alignments of selected NRPS condensation (C), epimerization (E) and peptidyl carrier protein (PCP) domains from the 16 analyzed myxochromide megasynthetases. **A/B:** Two groups of homologous C domains according to Figure S23B are compared. The consensus sequences of the seven conserved core motifs (core C1-C7) described by Marahiel *et al.* are illustrated on top of the alignment.³¹ The lines indicate the extended core regions, which were analyzed by Rausch *et al.* to distinguish different C domain subtypes, whereas yellow bars indicate significant specificity determining positions between ¹C_L, ^DC_L and starter C domain subtypes (those marked with an asterisk represent the most significant positions)³³. **A:** C domains from module 2 harbor the described residues typical for ¹C_L domains. **B:** C domains from modules 3 harbor the described residues typical for ^DC_L domains (except in pos. 316 of C3 from Cy1, Hm1, Hm2, Hm3, Se1, Sa1, Sa2 and Sa3). **C:** Sequence alignment of the module 2 E domains. The consensus sequences of the seven conserved core motifs (core E1-E7) described by Marahiel *et al.* are illustrated on top of the alignment.³¹ Key active site residues described from structural analysis of the E domain of tyrocidine synthetase A (His743 and Glu882)⁵³ located in core E2 and E4 are highlighted with yellow bars. **D-F:** Three groups of homologous PCP domains according to Figure S23C are compared. **D:** The characteristic [GGDSI] core motif described for PCP domains associated with epimerization (E) domains, so-called PCP^E domains,³² can be detected in all PCP domains from modules 2. **E:** The characteristic [GGHSL] core motif described for PCP domains from ordinary NRPS elongation modules, so-called PCP^C domains,³² can be detected in all PCP domains from modules 3 (and the M1_B_PCP4 homolog). **F:** PCP domains from modules 4 (module 5 for M1_B) can be subdivided into two groups: PCP domains from active modules incorporating proline during myxochromide A, B and C biosynthesis harbor an intact [GGHSL] core motif. PCP domains from inactive modules from myxochromide D and S megasynthetases show several mutations within the core motif (see red box) and lack the highly conserved Ser residue required for posttranslational activation (which was biochemically analyzed for Sa1 S_PCP4 in previous studies ⁷).

A

A		*	*	*	*	*	*	*	*	*	*	*	
Mf1_A_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mx1_A_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mx2_A_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mx3_A_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mx4_A_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
M1_B_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mh1_C_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Mv1_C_A1	L T S	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Cy1_D_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Hm1_D_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Hm2_D_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Hm3_D_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Se1_D_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Sa1_S_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Sa2_S_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						
Sa3_S_A1	L T T	H F D F S V W E G N Q I	F G G E	I N M Y G I	T E T T V H V T Y	D F W N I G M V H K	Thr / Thr / Thr / Thr						

B

[illegible]

C

[illegible]

D

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Figure S19 (continued on next page)

From Pro-incorporating modules:

Mf1_A_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mx1_A_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mx2_A_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mx3_A_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mx4_A_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
M1_B_A5	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mh1_C_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro
Mv1_C_A4	LWQAFDVSFQESFLITAGDHNHYGPTESHVVTAL	DVQFI AHVVK	Pro / Pro / Pro / Pro

From inactive modules:

Hm1_D_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pro / Pro
Hm2_D_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pro / Pro
Hm3_D_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pro / Pro
Se1_D_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pip / -
Sa1_S_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pip / -
Sa2_S_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / Pip / -
Sa3_S_A4	LWHGTGASLQESFLITAGGERCOYSVAEAYVVAAH	GAQFI AQVVK	(c) / Pro / mPro / -

E

Mf1_A_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	Phe / Ala / Ala / Ala
Mx1_A_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	Phe / Ala / Ala / Ala
Mx2_A_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(d) / Ala / Ala / Ala
Mx3_A_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	Phe / Ala / Ala / Ala
Mx4_A_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	Phe / Ala / Ala / Ala
M1_B_A6	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(e) / Ala / Ala / Ala
Cy1_D_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(a) / Ala / Ala / Ala
Hm1_D_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(a) / Ala / Ala / Ala
Hm2_D_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(a) / Ala / Ala / Ala
Hm3_D_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(a) / Ala / Ala / Ala
Se1_D_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(a) / Ala / Ala / Ala
Sa1_S_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(d) / Ala / Ala / Ala
Sa2_S_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(d) / Ala / Ala / Ala
Sa3_S_A5	VMQSF DVS VWE MVL T SAGEVNAYGPT E V S I S T L	DVWV L A A I K	(d) / Ala / Ala / Ala

F

Mf1_A_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mx1_A_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mx2_A_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mx3_A_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mx4_A_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
M1_B_A7	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mh1_C_A5	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Mv1_C_A5	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Cy1_D_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Hm1_D_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Hm2_D_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Hm3_D_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Se1_D_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Sa1_S_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Sa2_S_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln
Sa3_S_A6	HWMTFDASVWEAQLCVGG EWNLYGPT EATIDSLA	DAWQCGLIDK	(f) / Gln / Gln / Gln

Figure S19. Substrate specificity analysis of the 94 adenylation (A) domains from the 16 analyzed myxochromide megasynthetases. Six groups of homologous A domains according to Figure S23A are compared (A-F). The two sequence columns represent the 8Å signature (a set of 34 active site residues as defined by Rausch *et al.*³⁵ and the Stachelhaus code (ten specificity-conferring residues defined by Stachelhaus *et al.*³⁶, nine of which are part of the 8Å signature as labeled with asterisks). The 8Å signature/Stachelhaus code for each A domain was retrieved via the NRPSpredictor2 analysis tool.³⁴ Predicted substrate specificities, shown in the four columns to the right, were retrieved from reports of the applied antiSMASH 3.0 gene cluster analysis.¹³ They include substrate predictions based on the NRPSpredictor2 method³⁴ (1st column) / Stachelhaus code³⁶ (2nd column) / method of Minowa *et al.*⁵⁴ (3rd column) / consensus of the three approaches (4th column; “-” indicates no consensus). In some cases, no single substrates but (only) classes/clusters of several amino acids were predicted by NRPSpredictor2 as indicated by (a-f): (a) Apolar, aliphatic (Gly, Ala, Val, Leu, Ile, Abu, Iva), (b) Aliphatic, branched hydrophobic (Val, Leu, Ile, Abu, Iva), (c) Hydrophobic aliphatic (Ala, Gly, Val, Leu, Ile, Abu, Iva, Ser, Thr, Hpg, Dhpg, Cys, Pro, Pip), (d) Unpolar aromatic ring (Phe, Trp), (e) Tiny, hydrophilic, transition to aliphatic (Gly, Ala), (f) Aliphatic chain with H-bond donor (Asp, Asn, Glu, Gln, Aad).

2.7.7.2 Analysis of Recombination Sites

Comparison of the domain organization of the different myxochromide megasynthetase (sub)types suggests that three different recombination events took place in the course of *mch* gene cluster diversification (Figure 2). Based on data from phylogenetic analysis of A-, C- and (P)CP domain regions (Figure S23), the three novel cluster types (B-type, C-type and D-subtype 2) result from duplication or deletion of specific catalytic domain regions. To detect possible recombination sites, alignments with selected regions of *mchC* genes were performed using the Geneious alignment tool integrated into Geneious software version 9.1.2⁴⁵ (Figures S20-22). In the following, the *in silico* analysis data are discussed for each of the three novel megasynthetase types.

Myxochromide B megasynthetase ('module duplication')

Phylogenetic analysis revealed that the heptamodular myxochromide B megasynthetase contains a duplicate of A₃, CP₃ and C₄ compared to other myxochromide assembly lines (Figure S23). This indicates that the B-type *mch* cluster evolved from an A-type cluster by duplication of the A₃-CP₃-C₄ region. Sequence alignments point to a recombination site that is located at the 5' end of the A₃/A₄ domain regions, more precisely and on protein level located between the first α -helix/ β -sheet according to structural data of A domain regions from other NRPS systems (Figure S20).^{55,56} This result is in accordance with detected repeat regions in the B-type *mch* cluster dot plot (Figure S1), which additionally indicates mutational activities in the N-terminal C domain regions after the duplication event. Furthermore, the analysis of the local codon usage adaptation along the catalytic domains shows a highly similar pattern in the 'duplicated' A₃-CP₃-C₄ assembly line region (with exception of the ultimate N-terminal A₃ region; Figure S25C, line M1). The 'duplicated module' introduces a second *L*-leucine residue into the myxochromide B heptapeptide core.

Myxochromide C megasynthetase ('module deletion')

Phylogenetic analysis revealed that the pentamodular myxochromide C megasynthetase lacks A₅, CP₅ and C₆ compared to other myxochromide assembly lines (Figure S23). This indicates that the C-type *mch* cluster evolved from the A-type cluster by deletion of the A₅-CP₅-C₆ region. Sequence alignments point to a recombination site that is located at the 3' end of the C₅/C₆ domain regions, more precisely and on protein level between the last β -sheet/ α -helix according to structural data of C domain regions from other NRPS systems (Figure S21).^{55,57} Deletion of the A₅-CP₅-C₆ assembly line region (compared to the A-type megasynthetase) causes the lack of *L*-alanine in myxochromide C lipopentapeptide cores compared to myxochromides A.

Myxochromide D-subtype 2 megasynthetase ('partial module deletion')

Phylogenetic analysis revealed that the myxochromide D-subtype 2 megasynthetase lacks A₄ and CP₄ compared to other myxochromide assembly lines (Figure S23). This indicates that the D-subtype 2 *mch* cluster evolved from the D-subtype 1 cluster by deletion of the A₄-CP₄ region. Sequence alignments point to a recombination site that is located around the C₄-A₄/CP₄-C₅ (linker) regions (Figure S22).^{55,57} Deletion of the A₄-CP₄ causes the lack of *L*-proline in the myxochromide D lipopentapeptide products compared to myxochromides A. Interestingly, we see also less codon usage adaptation in this inactive A₄-CP₄ region in all *Cystobacterineae* strains compared with the following A₅-CP₅ or preceding A₃-CP₃ region (see Figure S25C).

In summary, the three discussed recombination events (followed by additional mutational activities) established novel and in all cases functional myxochromide assembly lines directing the production of altered myxochromide peptide cores. The observed 'natural' assembly line diversification results from duplication/deletion of 'A-CP(-C) units' instead of dedicated 'C-A-CP' modules (for further discussion see chapter 2.7.7.4).

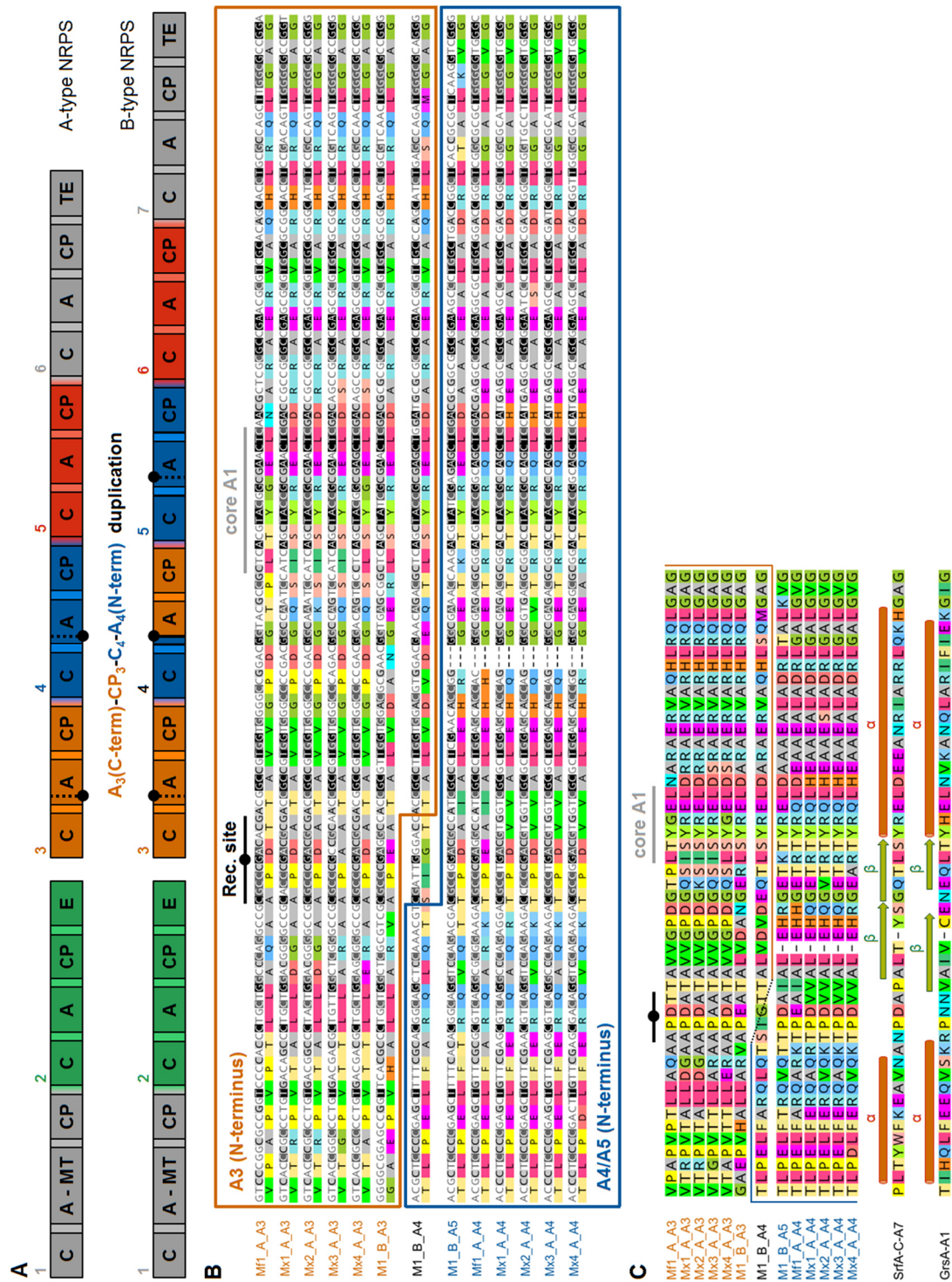


Figure S20. Module duplication resulting in a B-type NRPS assembly line. **A:** Illustration of the identified recombination site at regions encoding the A3/A4 N-termini of the A-type pathway. **B:** Translation alignment of the regions encoding the A3/A4 N-termini (first 147/144 nt, 49/48 aa) from A-type pathways from strains Mf1, Mx1, Mx2, Mx3 and Mx4 with the respective region of A3/A4/A5 from the B-type pathway from strain M1. The identified recombination site is indicated with a black line. **C:** Protein alignment of the regions analyzed in **B** with the respective regions from the surfactin synthetase subunit SrfA-C (PDB:2VSQ,²⁸) and the gramicidin S synthetase subunit GrsA (PDB:1AMU,²⁹). According to this comparison, the ‘fusion site’ after the module duplication event resulting in a B-type pathway is located between the first A domain α -helix and β -sheet.

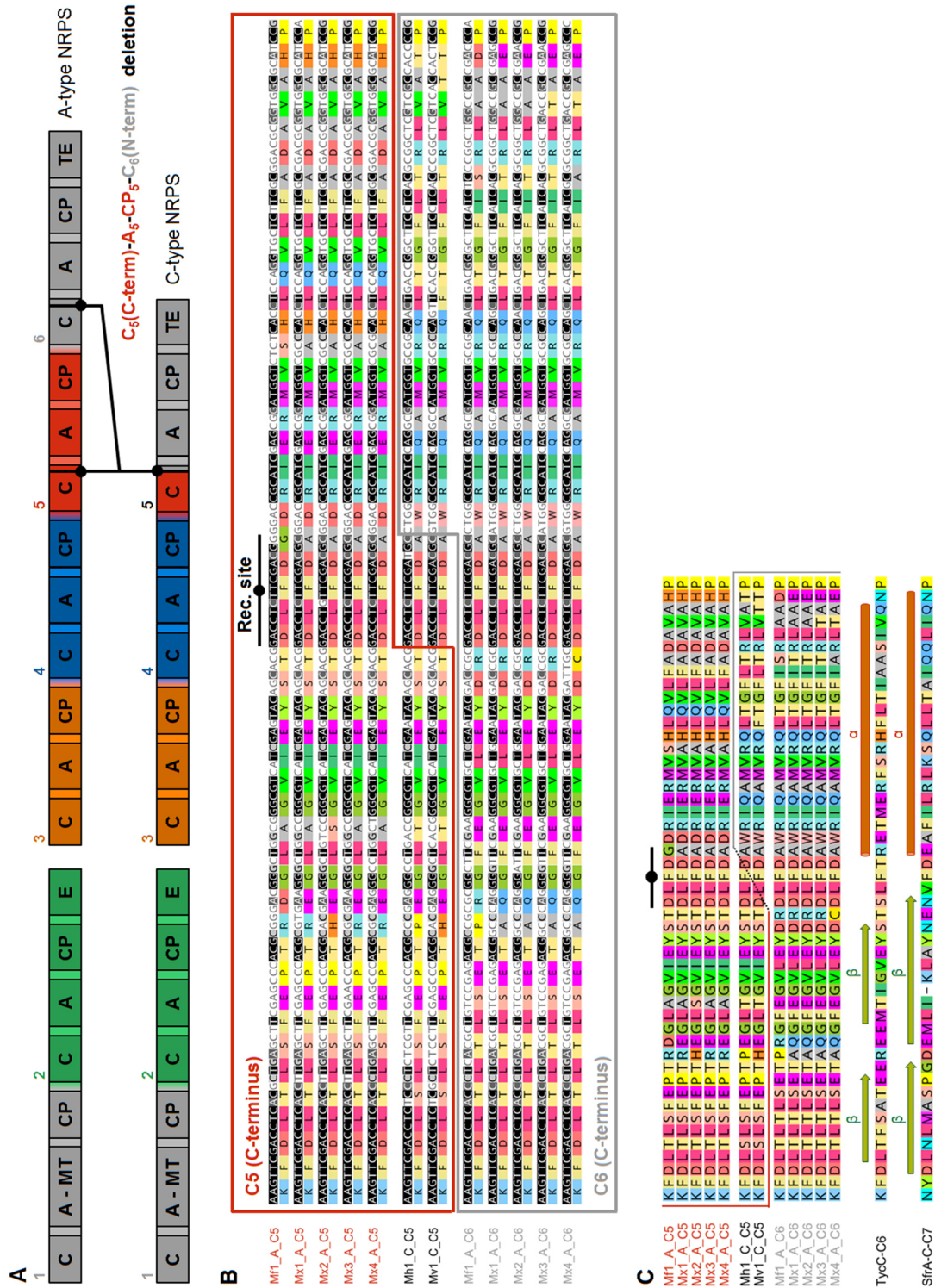


Figure S21. Module deletion resulting in a C-type NRPS assembly line. **A:** Illustration of the identified recombination site at regions encoding the C5/C6 C-termini of the A-type pathway. **B:** Translation alignment of the regions encoding the C5/C6 C-termini (last 147 nt / 49 aa) from A-type pathways from strains Mf1, Mx1, Mx2, Mx3 and Mx4 with the respective region of C5 from C-type pathways from strains Mh1 and Mv1. The identified recombination site is indicated with a black line. **C:** Protein alignment of the regions analyzed in **B** with the respective regions of the tyrocidine synthetase subunit TycC (PDB:2JGP,³⁰) and the surfactin synthetase subunit SrfA-C (PDB:2VSQ,²⁸). According to this comparison, the ‘fusion site’ after the module deletion event resulting in C-type pathways is located between the last C domain β -sheet and α -helix.

2.7.7.3 Phylogenetic Analysis

Phylogenetic reconstructions were mainly performed on the basis of coding DNA sequences. The sequences were aligned using the GUIDANCE2Server³⁸ applying the MAFFT algorithm with the codon option. Columns with a GUIDANCE score below 0.93 were removed from the alignments. Reconstructions were done by the distance-based neighbor-joining method using the modules “dnadist” and “neighbor” of the PHYLIP package³⁷ and applying the F84 model of nucleotide substitution. The reliability of tree topologies was evaluated by means of the bootstrap method based on 1000 pseudo-replicates using the module “seqboot”. Majority consensus trees were calculated using the “consense” module. For the megasynthetase subunits (MchA, MchB and MchC), phylogenetic reconstructions were also performed based on amino acid sequences. The sequences were aligned using ClustalX,³⁹ and manually edited to remove areas that could not be aligned with confidence. The phylogenetic reconstruction was conducted using the modules “protdist” and “neighbor” of the PHYLIP package applying the JTT model of amino acid substitution and a gamma distribution to represent among-site rate heterogeneity. Reliability of branching topologies was evaluated by the bootstrap method based on 1000 pseudo-replicates of the alignment. To reconstruct the phylogeny of the complete clusters, the nucleotide sequences of mchA, mchB and mchC were concatenated and aligned using the GUIDANCE2Server with the same settings described above. The neighbor-joining tree was calculated based on the F84 model of nucleotide substitution and its reliability tested by bootstrapping 1000 pseudo-replicates. For the reconstruction of the phylogeny of the myxobacterial producer strains, the nucleotide sequences of the 16S small ribosomal subunit rRNA genes were concatenated with the nucleotide sequences of 15 selected genes encoding highly conserved proteins, a subset of 24 genes previously suggested for use in bacterial taxonomy.⁵⁸ The 15 genes were small ribosomal subunit proteins S3, S5, S7, S9, S10, S12 and S19, large ribosomal subunit proteins L2, L4, L6, L11, L14 and L16, elongation factor 4 and phenylalanine-tRNA ligase subunit alpha. The subset was selected on uniqueness and completeness of all 16 orthologs, sequence assembly without ambiguous or missing nucleotides and gap-free alignments of identical length. The sequences (supplemental Microsoft Excel file “Supplemental_file_16_gene_sets.xlsx”) were aligned using ClustalX. Phylogenetic reconstruction was conducted using the modules “dnadist” and “neighbor” and was based on the F84 model of nucleotide substitution. Tree reliability was again tested by the bootstrap method with 1000 pseudo-replicates. A comparison of strain and cluster phylogenies is illustrated and discussed in the main text (Figure 3).

Phylogeny of NRPS domains (A, C and PCP)

The phylogenetic tree of adenylation (A) domains (see Figure S23A) shows a clear separation of the domains according to the module they belong to. The phylogenetic relationships within each main branch correspond to the strain phylogeny of the producers. Such a pattern is typical for orthologous genes, which share the same ancestor and were diversified by speciation or separation into strains. The only exceptions are sequences from *Stigmatella erecta* Pde77 (Se1), which have a tendency not to cluster together with the other *Stigmatella* strains Sa1, Sa2 and Sa3 but instead with the *Hyalangium minutum* strains Hm1, Hm2 and Hm3 and the unclassified strain Cy1. However, this phylogenetic pattern does not exclude the possibility of independent heterologous integration events by horizontal gene transfer (see chapter 2.7.7.4 and main text). The A domain phylogeny also revealed that the A domains of NRPS modules 2 and 3 from the S-type megasynthetases fall together on the same branch with the respective counterparts of the other megasynthetases (see the green and orange box in Figure S23A). Therefore, all A domains of the second NRPS module as well as all A domains of the third NRPS module appear to be orthologs and thus the change in substrate specificity of the A domains belonging to S-type megasynthetases is probably the result of mutations in the active center (see chapter 2.7.7.1). Furthermore, the A domain phylogeny shows that the heptamodular B-type megasynthetase harbors a ‘duplicate’ of A₃ (A₃/A₄; see the orange box in Figure S23A). A similar observation was made in the phylogeny of peptidyl carrier protein (PCP) domains, which revealed an additional copy of PCP₃ (PCP₃/PCP₄; see the orange box in Figure S23C). However, the evolutionary relationships are different in the case of condensation (C) domains. Here, the phylogeny revealed that the B-type megasynthetase contains a ‘duplicate’ of C₄ (C₄/C₅; see the blue box in Figure S23B). These data suggest that the additional module from the heptamodular B-type assembly line originated from an ‘A₃-PCP₃-C₄’ instead of a ‘C₃-A₃-PCP₃’ duplication event. Corresponding recombination sites were identified (see chapter 2.7.7.2).

Phylogeny of NRPS subunits (MchA, MchB and MchC)

The phylogenies of the whole megasynthetase subunits MchA, MchB and MchC are shown in Figure S24. The tree topologies were the same for reconstructions based on nucleotide or amino acid sequences. For each type of subunit, the phylogenetic relationships follow the strain phylogenies. However, as in the case of single domain phylogenies, there is always one exception: The subunits from the *Stigmatella erecta* strain (Se1) mainly cluster together with *Hyalangium minutum* strains but not with the other *Stigmatella* sp. strains, which could be explained by a horizontal gene transfer of the *mch* cluster to the Se1 strain.

A

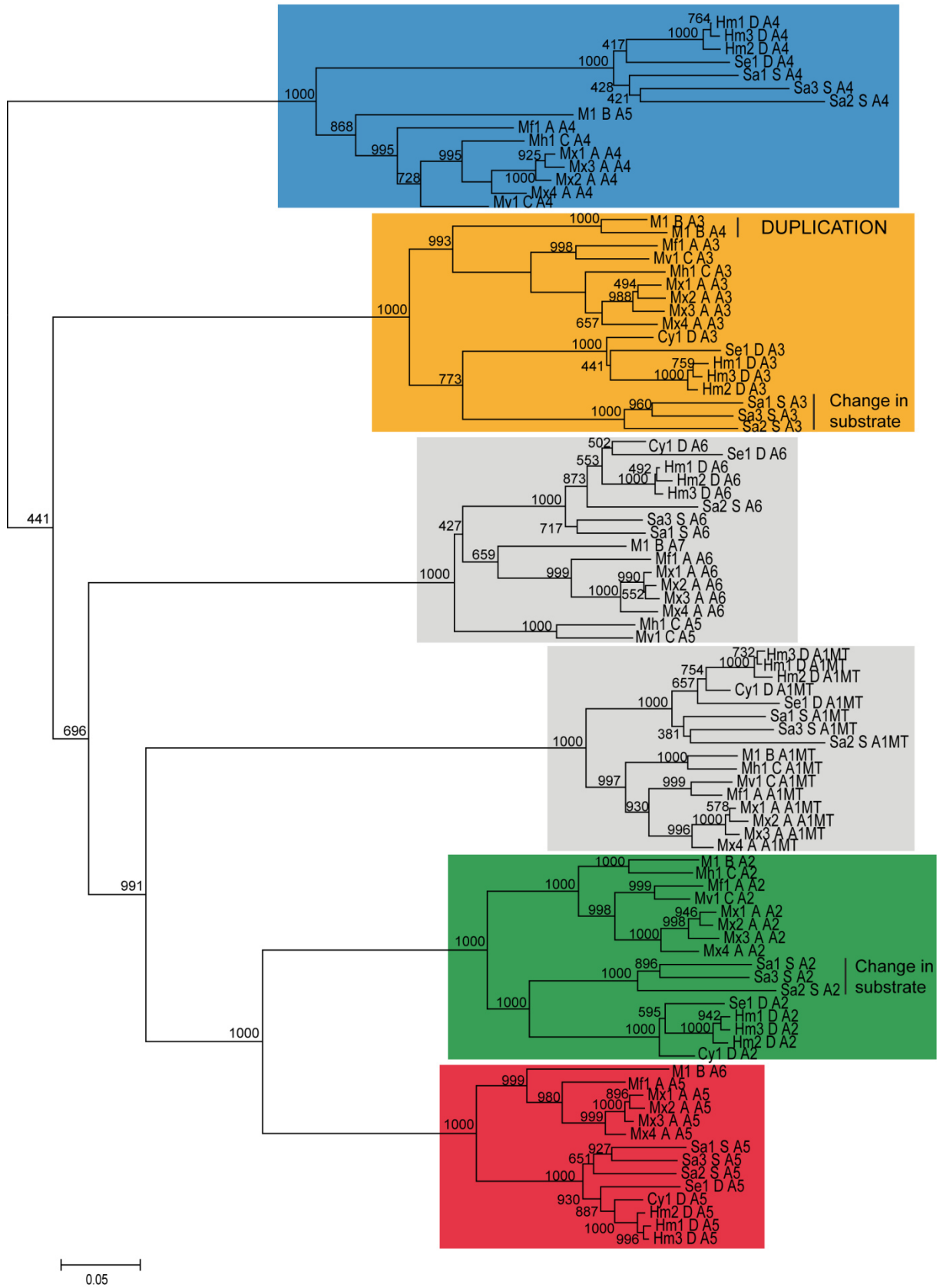


Figure S23 (continued on next page)

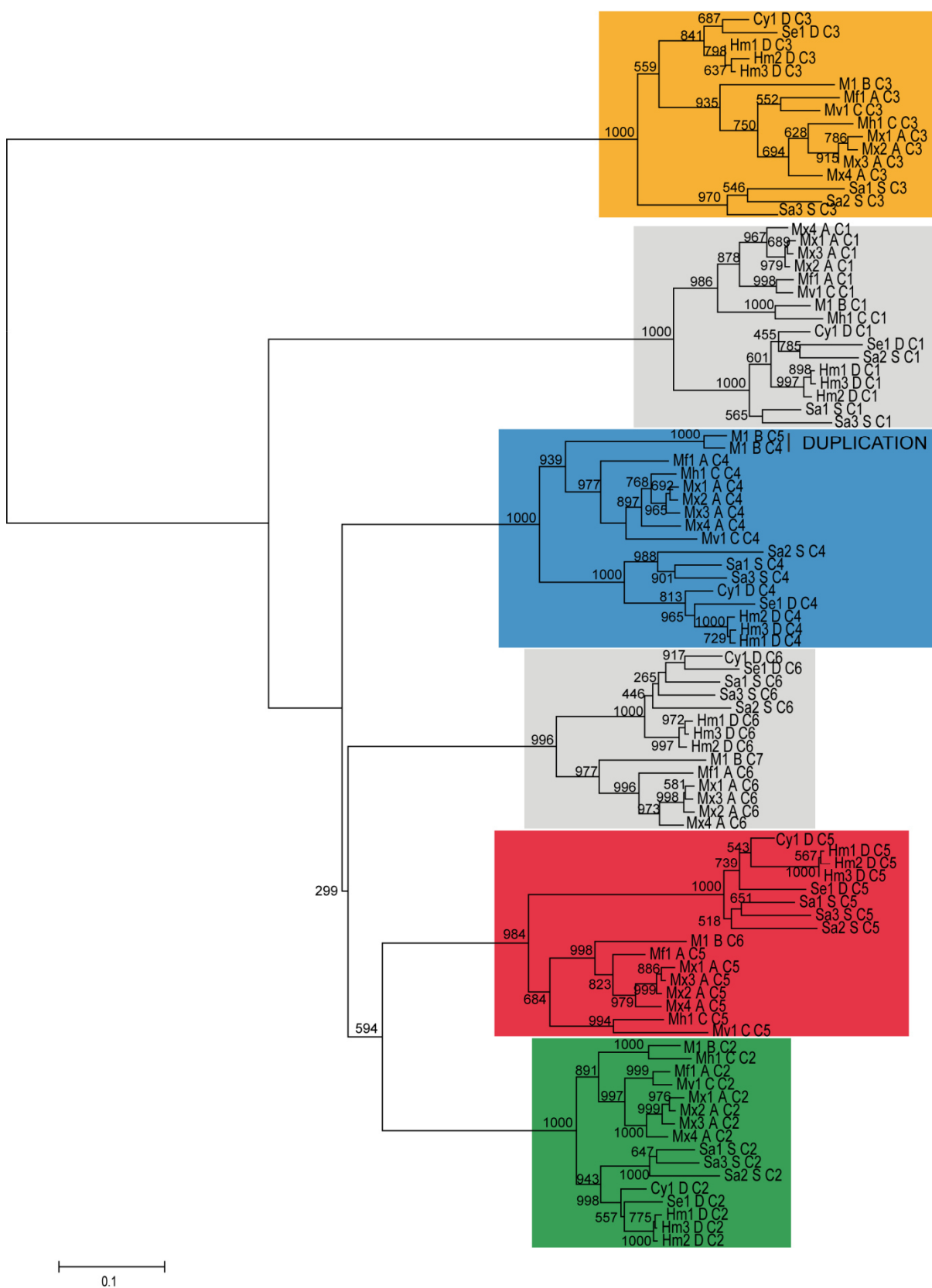
B

Figure S23 (continued on next page)

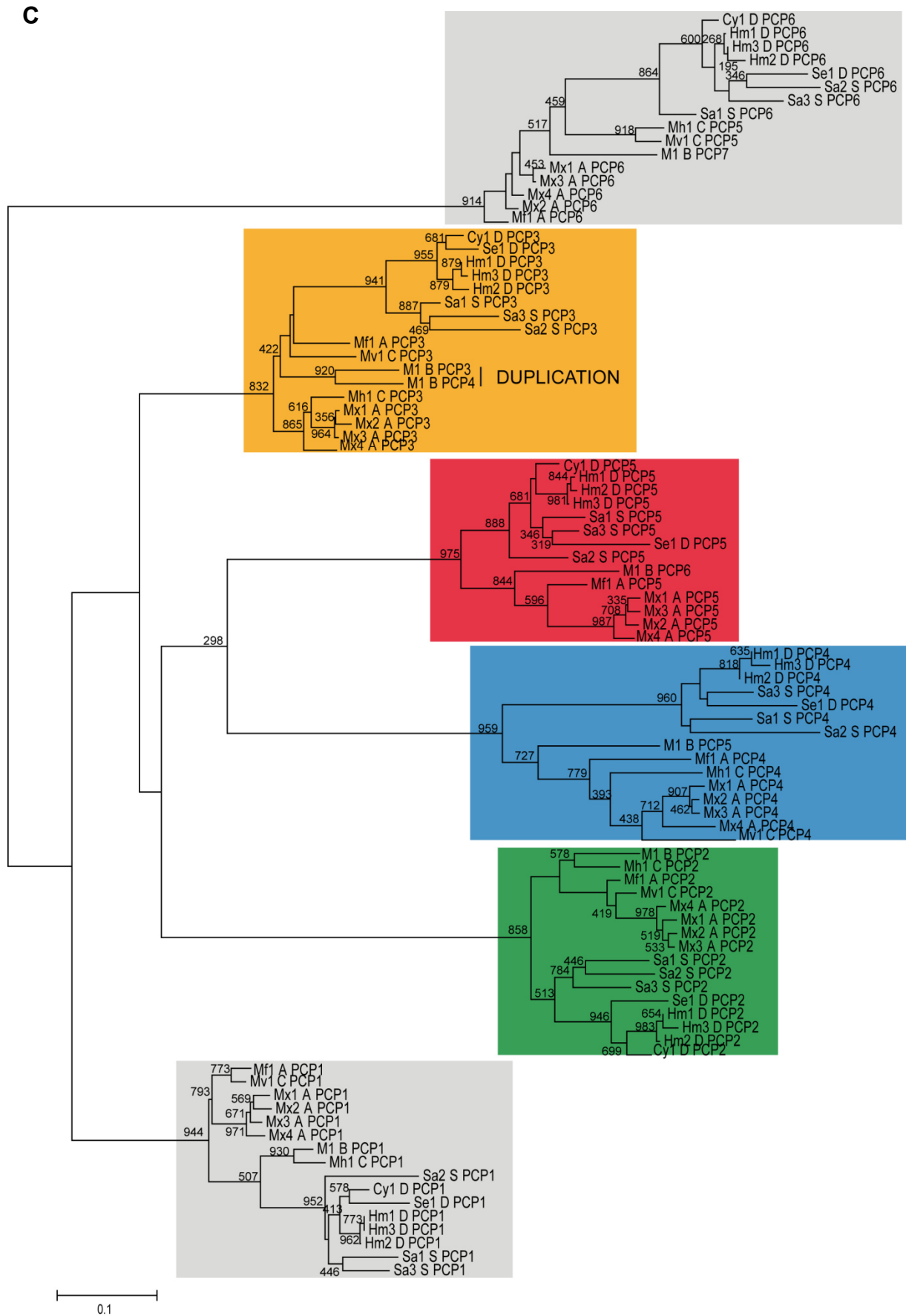


Figure S23. Phylogeny of NRPS domains from the 16 analyzed myxochromide megasynthetases based on DNA sequences. **A:** Adenylation (A) domains. **B:** Condensation (C) domains. **C:** Peptidyl carrier protein (PCP) domains.

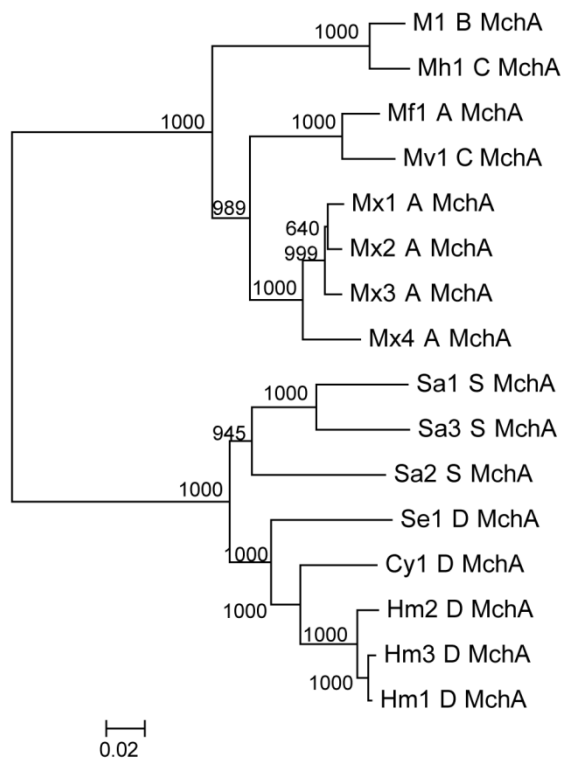
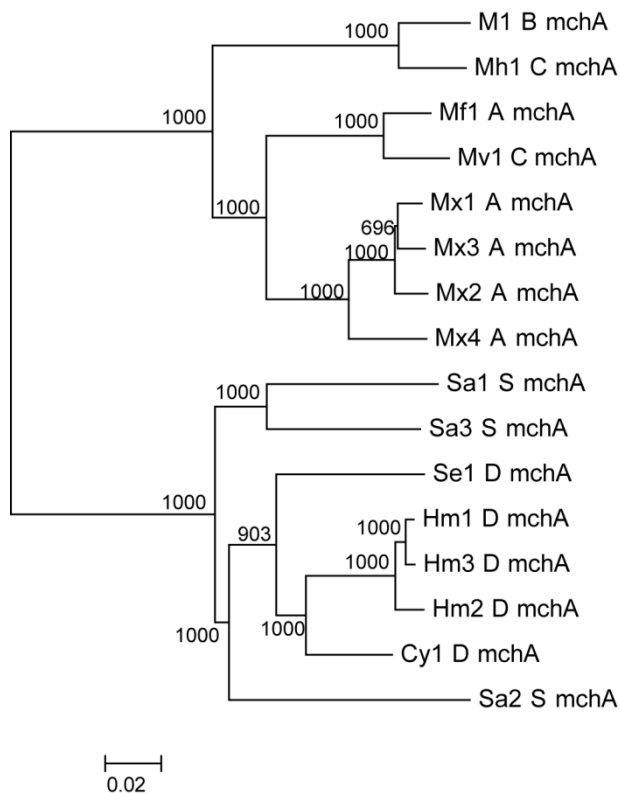
A

Figure S24 (continued on next page)

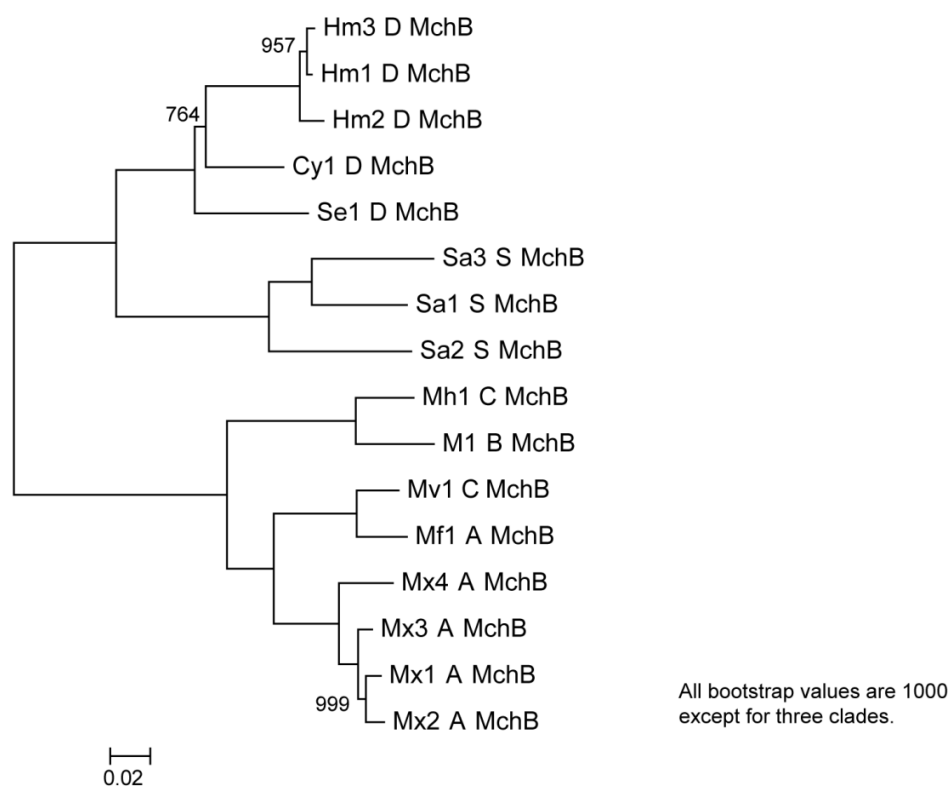
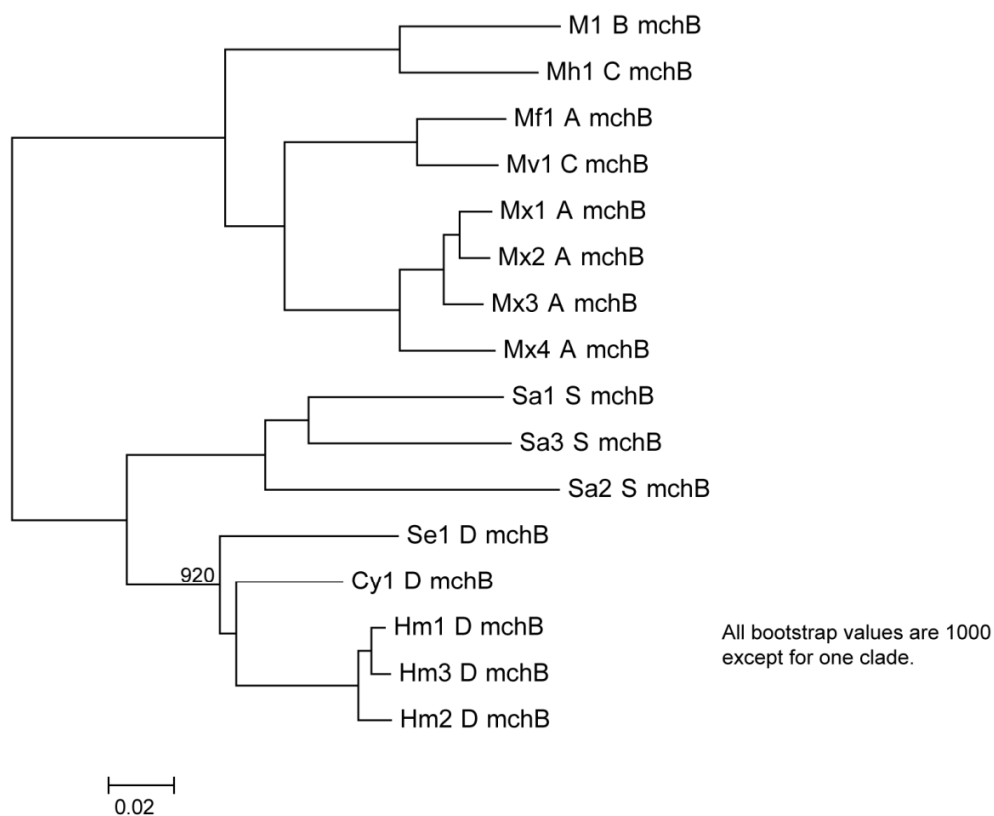
B

Figure S24 (continued on next page)

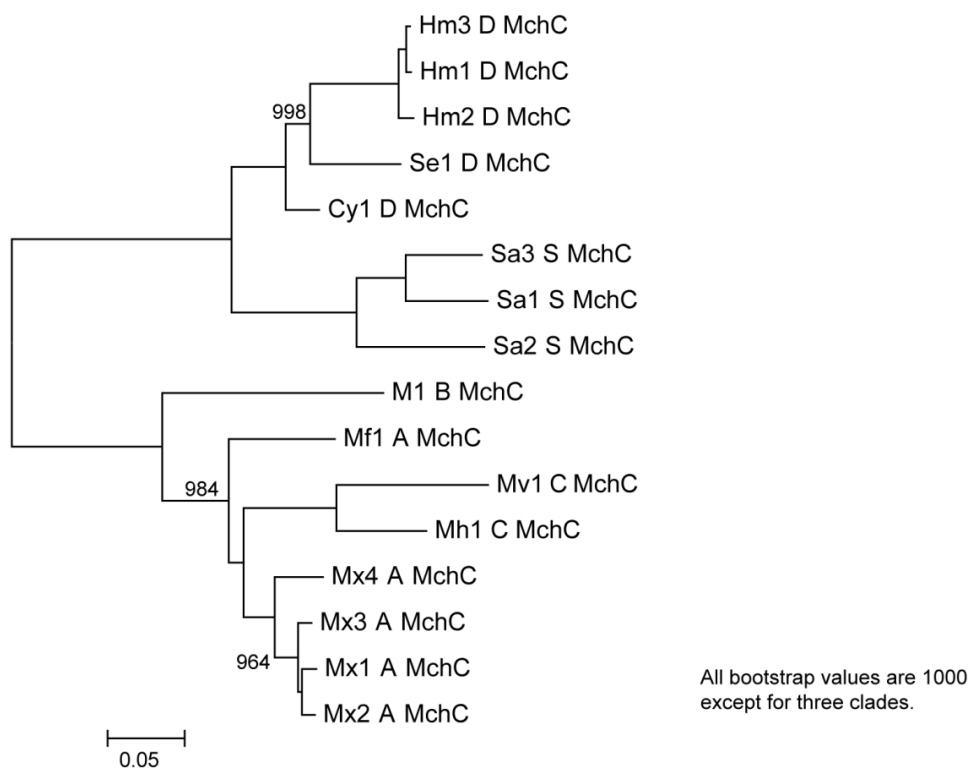
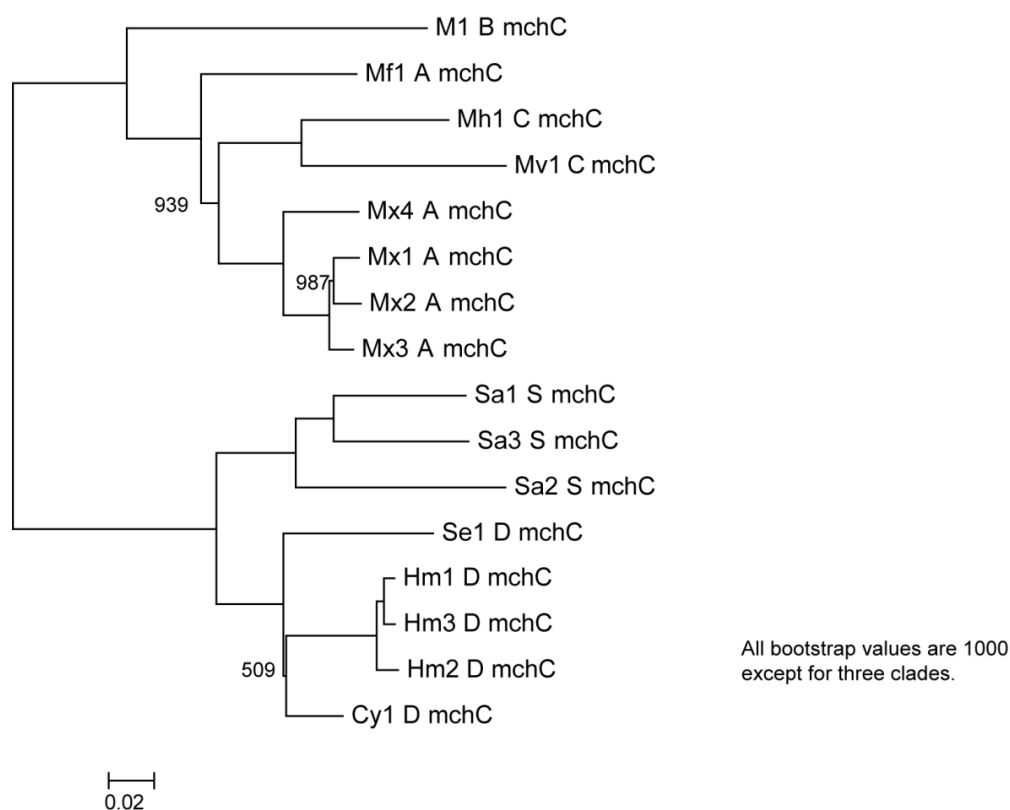
C

Figure S24. Phylogeny of subunits of the 16 analyzed NRPS megasynthetases based on DNA and protein sequences. **A:** PKS subunits mchA (DNA) and MchA (protein). **B:** NRPS subunits mchB (DNA) and MchB (protein). **C:** NRPS subunits mchC (DNA) and MchC (protein).

2.7.7.4 CAI Analysis

Adaptation of the specific synonymous codon usage of CDS to their corresponding genome's codon usage of highly expressed (hxp) genes correlates with the expression level of the CDS, which is usually robustly measured by calculating the codon adaptation index (CAI).⁵⁹ Additionally, the codon usage of horizontally acquired gene clusters is likely to differ from the new host genome at the time of gene transfer. Consequently, the CAI value of such gene clusters is expected to be subsequently adjusted over evolutionary time scales. “Old” and conserved genes have therefore generally higher codon adaptation compared with newly acquired and “young” genes.^{23,60,61} However, mutational hotspots in rapidly evolving genes, e.g. after a recent genome rearrangement or due to new evolutionary pressures, may initially change the codon adaptation to lower values due to higher pressure for selection of the encoded amino acid sequence. After this process yields mutational equilibrium, codon optimised genes may re-evolve again via silent mutations.^{62,63} We therefore analyzed global and local codon usage adaptation to gain additional insights into cluster evolution.

Selection of hxp2 gene sets

We previously selected a modified hxp gene set (hxp2) for the *Myxococcales* Mx1 (*Myxococcus xanthus* DK1622), *Sorangium cellulosum* So ce56 and *Chondromyces crocatus* Cm c5,⁶⁴ based on estimations of translation efficiency in *E. coli* and identification of functional orthologs and paralogs.⁶⁵ We used the established hxp2 protein set of Mx1 to search with locally installed NCBI BLAST⁶⁶ for respective homologous and paralogous protein sequences from all 16 strains investigated here. We manually reviewed the list, calculated intermediary CAI values, as described below, from the initial list members and removed paralogs having low codon adaptation manually, which may represent protein family members expressed alternatively during non-optimal growth conditions with changed codon usage or generally lowly expressed family members. The selected hxp2 gene sets are provided in the Microsoft Excel document “Supplemental_file_hxp2_gene_sets.xlsx”. For the 16 strains, following numbers of hxp2 proteins were selected: Mx1 (353)⁶⁴, Mx2 (346), Mx3 (353), Mx4 (358), Mf1 (369), Mv1 (356), Mh1 (374), M1 (364), Sa1 (377), Sa2 (378), Sa3 (367), Se1 (391), Cy1 (405), Hm1 (378), Hm2 (385), Hm3 (387). Within the header line of the provided FASTA files, the categories of functionally related highly-expressed protein subsets are annotated: hxp2 = A (ATP synthase and adenylate kinases), C (chaperones), D (DNA-related), E (export-related), F (fatty acid metabolism), G (glycolysis and basic carbohydrate metabolism), H (respiration), K (citrate cycle and related carboxylases), N (nitrogen metabolism), O (major outer membrane proteins), P (pentose phosphate cycle), R

(transcription), S (detoxification), T (translation). Additionally the CAI value calculated with the codon usage table of the respective strain-specific hxp2 sequences is indicated by the tag “cai_hxp2”.

Calculation of codon adaptation index (CAI) values

The codon usage tables for the selected set of hxp2 cds sequences were calculated by the EMBOSS program cusp⁴¹ after removing FASTA headers and sequence line breaks by a Perl pipeline program. The respective sequence-specific CAI was calculated by a Perl program as described by Sharp and Li.⁵⁹ The CAI was calculated for the myxochromide cluster CDS sequences. A color scale blue-yellow-red (CAI 0.48 to 0.77) was used to illustrate the level of adaptation in codon usage (see Figure 3 and Table S12). The color scale and the respectively colored CDS symbols of Figure 3 and Table S12 were generated with R.⁶⁷

Local CAI values along the CDS of *mchA*, *mchB* and *mchC* were calculated for overlapping CDS regions of 101 codons using a Perl program. For each codon, the local CAI represents the CAI value of a window of 101 codons centered at the actual codon. At the start and the end of the CDS, the window size was reduced for the missing codons on the respective 5'- or 3'-sides. The local CAI values were visualized as plotted color-shaded vertical lines along the three CDS for all 16 strains, where the codons were aligned to orthologous positions based on protein sequence alignments performed with ClustalX.³⁹ The alignment was additionally corrected according to the identified recombination sites for deletions/insertions within the MchC protein sequences (see Figures S20-22). The color scale was constructed for the CAI range of 0.2 to 0.91 (710 steps) as follows: 0.2-0.4, black to blue (201 colors); 0.401-0.6, blue to yellow (200 colors), 0.601-0.8, yellow to red (200 colors), 0.801-0.910 red to dark-red (110 colors). The respective Figures S25A (*mchA*), S25B (*mchB*) and S25C (*mchC*) were plotted using R.⁶⁷

Table S12. Codon adaptation of the *mch* CDS of the myxochromide gene clusters to the *hxp2* gene sets of their host genomes. The three disks represent the megasynthetase CDS *mchA*, *mchB* and *mchC* and are colored for their respective CAI value according to the displayed color scale. The disks were used accordingly in Figure 3.

Cluster		<i>mchA</i>	<i>mchB</i>	<i>mchC</i>	<i>mchD</i>
Mx1_A		0.690	0.684	0.618	0.748
Mx2_A		0.673	0.684	0.610	0.772
Mx3_A		0.685	0.682	0.614	0.729
Mx4_A		0.713	0.714	0.643	0.804
Mf1_A		0.746	0.738	0.672	0.799
Mv1_C		0.760	0.746	0.631	0.777
Mh1_C		0.719	0.702	0.632	0.825
M1_B		0.714	0.677	0.570	0.767
Sa1_S		0.566	0.586	0.597	0.606
Sa2_S		0.531	0.517	0.535	0.608
Sa3_S		0.584	0.579	0.568	0.618
Se1_D		0.490	0.516	0.484	0.383
Cy1_D		0.617	0.658	0.644	0.561
Hm1_D		0.588	0.630	0.621	0.514
Hm2_D		0.572	0.636	0.624	0.487
Hm3_D		0.590	0.632	0.620	0.502

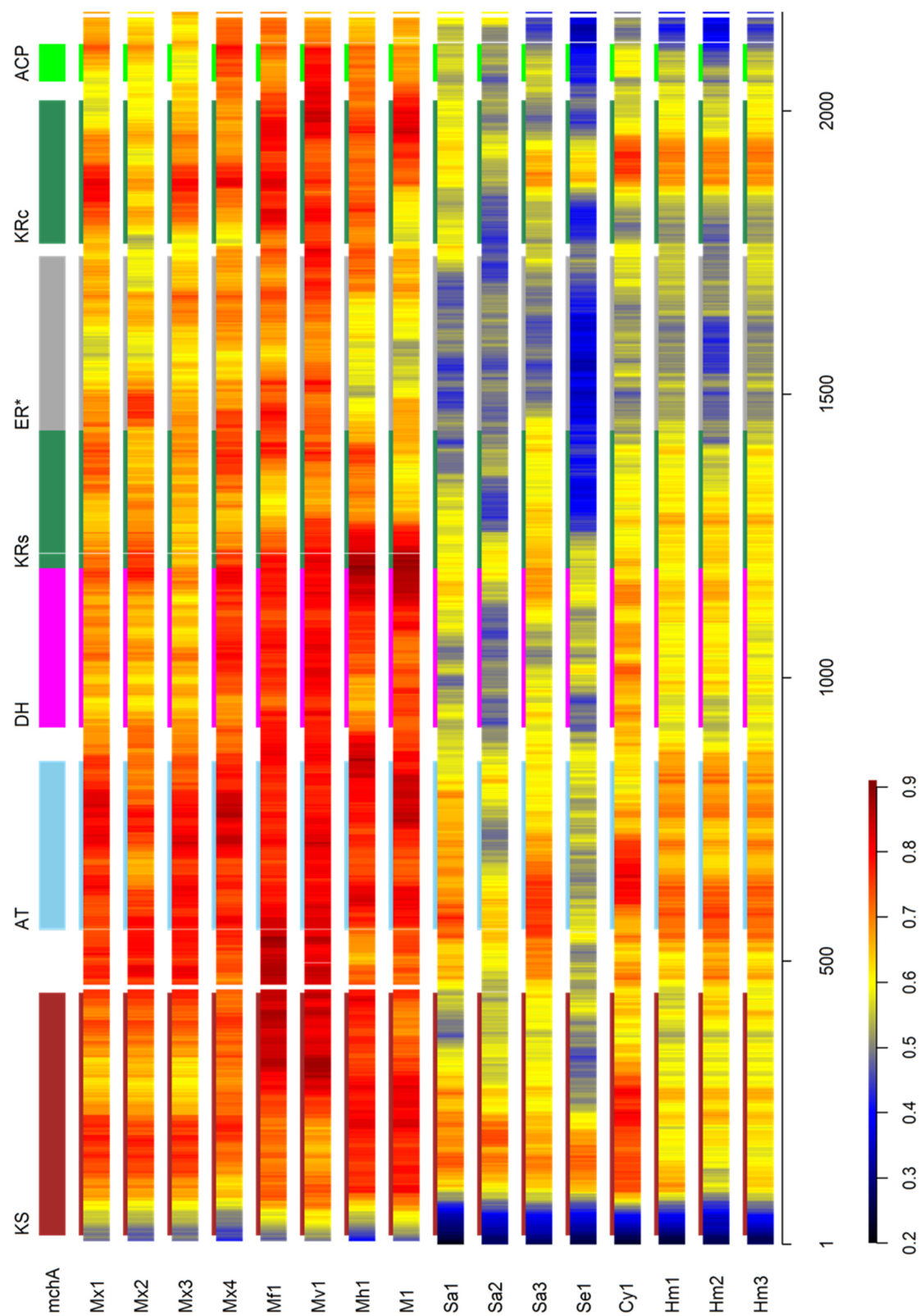
A

Figure S25 (continued on next page)

B

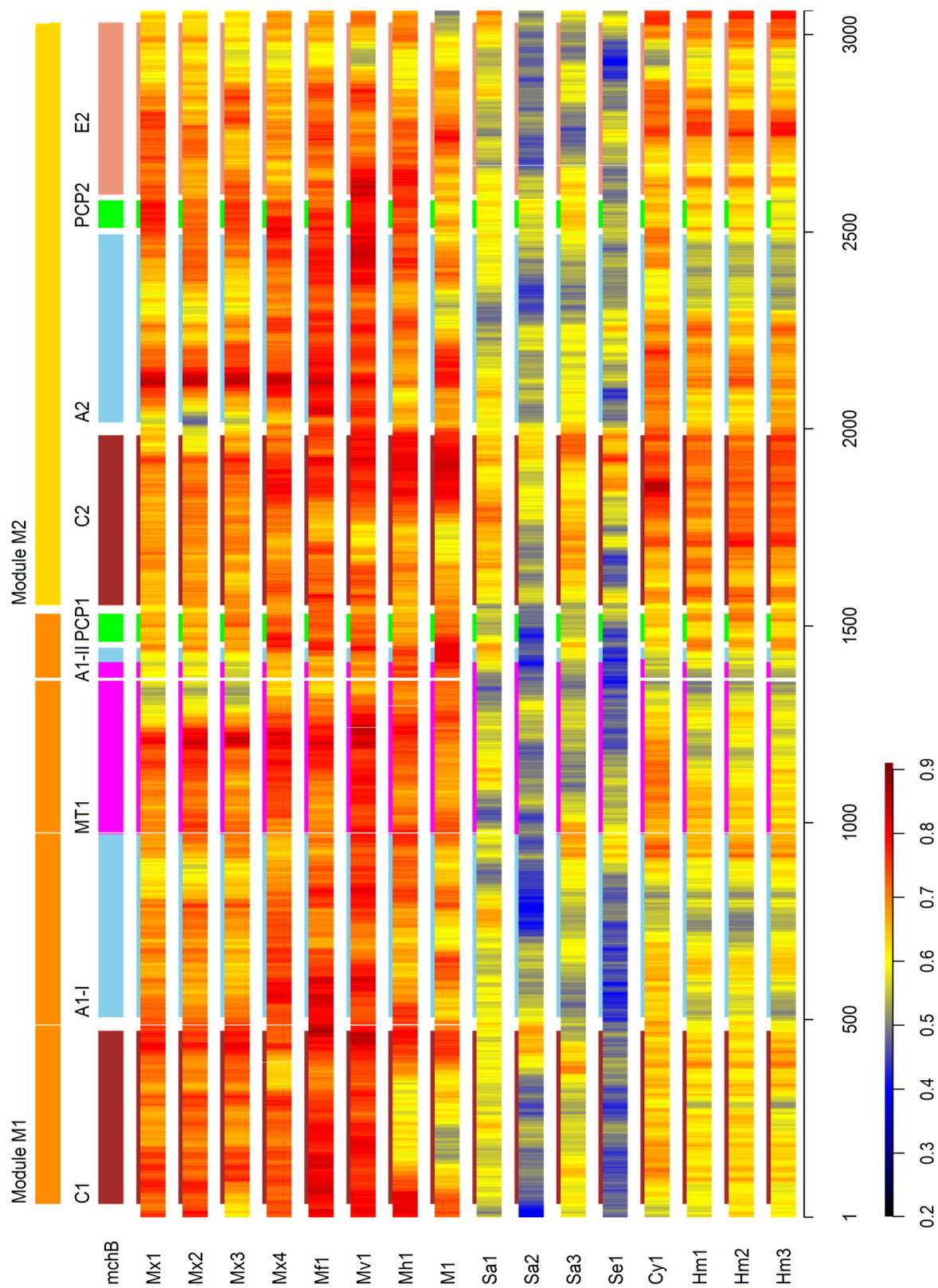


Figure S25 (continued on next page)

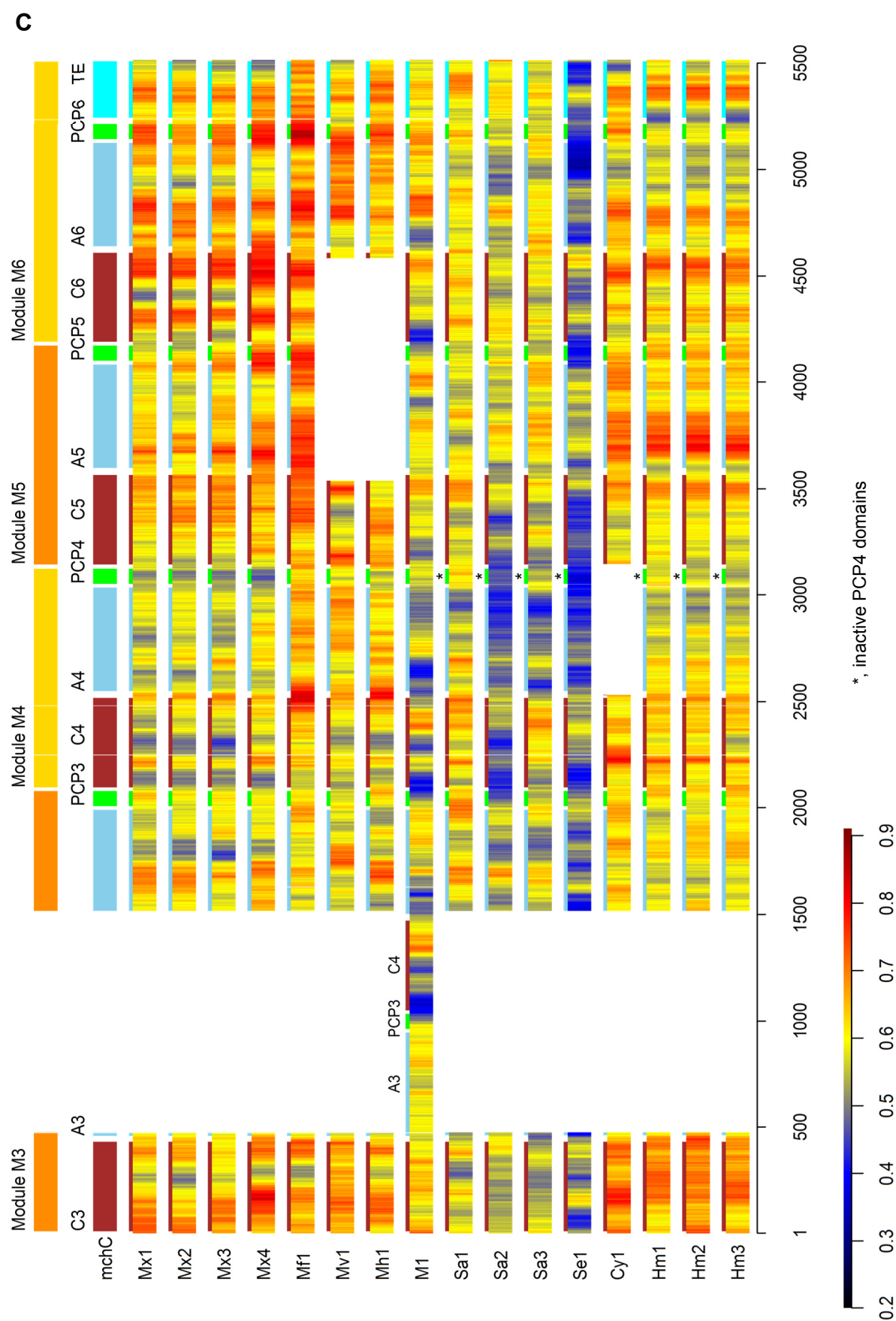


Figure S25: Local codon adaption index (CAI) values for a shifting window of 101 codons along the *mchA* (A), *mchB* (B) and *mchC* (C) CDS sequences of the 16 analyzed myxochromide pathways. Colors represent CAI values according to the shown color scale.

Extended discussion of the CAI analysis

The adaptation of codon usage generally follows sequence phylogeny (Figure 3) and is different for the six separate genomic *mch* cluster locations A-F (see Figure S2), which represent independent cluster integration events by horizontal gene transfer. As we have no contradicting data, we assume that the expression levels and biological function of the myxochromide gene clusters within their host strains are comparable. Therefore, our data suggest that the clusters with higher codon usage adaptation represent more ancient versions regarding genomic residence time compared with lower adaptation for more recently acquired clusters. A combination of horizontal gene transfer events and inheritance by strain divergence would agree with the general complex pattern of megasynthetase evolution observed throughout all kingdoms of life.^{21,68}

Cluster history in *Myxococcaceae*

All *Myxococcaceae* gene clusters are located in the same genomic integration site (Locus A; Figure S2) and show higher levels of codon usage adaptation compared with the *Cystobacteraceae* strains, especially for the structurally conserved *mchA* and *mchB* CDS. The highest adaptation is found in the two clusters of types A and C of the Mf1 and Mv1 strains, respectively. Myxochromide clusters may therefore have been acquired initially by Mf1 in the past, as indicated by the highest codon usage, the most common and intact cluster structure and a 5'-UTR region in front of the *mchA* CDS without deletions of the upstream neighboring AraC transcription factor (gene 5 in Figure S2). Next, the C-type cluster of Mv1 may have evolved, as suggested by its close association to the Mf1 cluster in phylogeny, harboring also an intact AraC transcription factor and having an exceptionally high CAI value for *mchA* and *mchB* as well. The deletion in *mchC* found in Mv1 leads to a reduced CAI because of loss of the highly adapted A₅-PCP₅-C₆ region. These two prototype-clusters were subsequently inherited by or distributed to other *Myxococcaceae* by more recent homologous recombination events. They therefore do not show the same level of codon adaptation as observed in the primordial cluster versions, which have resided in their genomes for the longest time periods. M1 acquired or evolved the B-type cluster presumably even more recently, where the module duplication in *mchC* triggered mutational events reducing codon usage adaptation of *mchC* (see Figure S25C, line M1). Additionally, the *mchC* CDS CAI value was reduced also by duplication of the exceptionally low adapted A₃-PCP₃-A₄ region of M1, which also shows generally lower adaptation in all strains. With the exception of Mx4 all other cluster integration sites in *Myxococcaceae* lost the upstream AraC transcription factor gene by a deletion event, which may have occurred during a more recent homologous recombination event, corroborating the proposed sequence of events.

Cluster history in Cystobacteraceae

In contrast to the *Myxococcaceae* strains, the *Cystobacteraceae* strains show 5 independent genomic integration sites (Figure S2, Locus B-F), which may correspond to 5 instances of horizontal gene transfer at different time points. Consequently, for all 5 types of integration sites, different levels of adaptations are found. These are generally much lower for the conserved *mchA* and *mchB* CDS (Figure S25A,B). If the interpretation of cluster CAI values as genomic residence time indicator is valid, the sequence of integration events was 1) introduction of the D-type cluster into Cy1 locus B, 2) followed by independent integration of D-type clusters in Hm1, Hm2 and Hm3 in locus C, 3) acquirement of S-type clusters by Sa1 and Sa3 in locus E, 4) integration of S-type cluster in Sa2 in locus F and most recently transfer of the D-type cluster to Se1 to locus D.

Inactive regions show lower CAI values

Inactive domains may have released evolutionary pressure to optimize codon usage and may therefore show reduced CAI values. The inactive ER* domains in the MchA proteins show reduced CAI throughout all 16 strains (Figure S25A). Additionally, the inactive A₄-PCP₄ region in the MchC proteins within the *Stigmatella* strains Sa1, Sa2, Sa3 and Se1 show reduced CAI levels when compared to the neighboring A₃-PCP₃ and A₅-PCP₅ regions (Figure S25C). Similar observations are made for the *Hyalangium* strains, Hm1, Hm2 and Hm3, whereas in Cy1 this region was even deleted. Generally, codon usage adaptation in the regions of modules 3 and 4 (A₃-PCP₃-C₄-A₄-PCP₄) of *mchC* is found at relatively low levels for all strains indicating these as presumably younger regions with ongoing mutational changes. This suggestion is corroborated by the structural changes observed in that region (duplication of the A₃-PCP₃-C₄ region in M1, deletion of the A₄-PCP₄ region in Cy1, change of the specificity of A₃ domain in S-type clusters).

Duplication event in M1 strain

The local CAI values along the *mchC* CDS of strain M1 (Figure S25C, line M1) show that the duplicated region A₃-PCP₃-C₄ most likely originated from the M1 A₄-PCP₄-C₅ region (orthologous to A₃-PCP₃-C₄ in other strains), which is also suggested by domain phylogeny (as described above) and recombination site analysis (as described above). The local CAI value of the N-terminal regions of the duplicated A₃ domain in M1, however, has substantially changed when compared with its template region. Additionally, several regions along the CDS show lower adaptation compared with all other *Myxococcaceae* strains. This finding could indicate ongoing adaptation to the new heptapeptide assembly regime due to co-

evolutionary changes not only in the duplicated domains C₄ and C₅, but also at the N-termini of the subsequent domains A₅, C₆, C₇ and A₇.

Relation to general megasynthetase cluster evolution

The observations described above are in line with a recent discussion of megasynthetase gene clusters evolution throughout all kingdoms of life,⁶⁸ which summarizes that gene clusters evolve by combined horizontal and vertical gene transfer and are formed by a series of gene duplications, insertions/deletions, recombinations, mutation and reshuffling events. These may follow common rules, specific for co-evolving sets of clusters or sub-clusters, which mix by recombination events and evolve independently.²¹ E.g., for myxochromide clusters the *mchA-mchB* CDS and the *mchC* CDS may constitute two such independently evolving sub-clusters, where *mchC* may be formed by family specific recombination rules, e.g. recombinations occur in the three known cases always near the regions encoding C-A domain interfaces. As in the case of polyketide synthase encoding gene clusters in *Streptomyces avermitilis*,⁶⁹ natural biorecombinatorics may be confined to regions that account for the structural diversity of the products, as it is the case here for *mchC* A-domain-driven recombinations.

2.7.8 Fruiting Body Formation and Swarm Expansion Assays with *M. xanthus* DK1622 and Myxochromide-Deficient as well as Overproducing Mutants

In order to obtain first insights into a possible biological function of myxochromides for their natural producers, fruiting body assays and swarm expansion assays with the myxobacterial model strain *M. xanthus* DK1622 and mutants thereof were performed. In addition to *M. xanthus* DK1622 wild type (WT, characterized as myxochromide A producer)⁷, a myxochromide A-deficient mutant *M. xanthus* DK1622::pMch22a⁷, as well as two myxochromide overproducing mutants were analyzed: *M. xanthus* DK1622::pMch70a (Wenzel *et al.*, unpublished) and *M. xanthus* DK1622::pTps-mchS.¹⁵ *M. xanthus* DK1622::pMch70a was generated by insertion of the constitutive Tn5 promoter upstream the myxochromide A biosynthetic gene cluster, which led to an about 10-fold increase of myxochromide A production compared to *M. xanthus* DK1622 WT. *M. xanthus* DK1622::pTpS-mchS was generated via transposon mediated integration of the SaI myxochromide S biosynthetic gene cluster under control of the constitutive Tn5 promoter and produces myxochromides A at the same level as the wild type plus myxochromides S/S-Abu/S-diAbu at about 50-fold higher production yields.

Fruiting body formation assays

Comparative fruiting body assays were performed as follows: Routinely grown liquid cultures were washed twice with MC7 buffer (0.5 M Mops buffer, 50 mM $\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$, pH 7.0) and $5 \cdot 10^9$ cells were spotted onto ‘agar wells’ of a 24-well microtiter plate. Each well contained 600 μl TPM agar (Tris-HCl [pH 8.0] 10 mM, $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ buffer [pH 7.6] 1mM, $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$ 8 mM, pH adjusted to 7.6). Microscopic images were recorded after 6, 24, 48 and 72 h incubation at 30 °C by using a Zeiss binocular microscope with 10 \times magnification (see Figure S26). In this assay, no significant difference in fruiting body development was observed between the *M. xanthus* WT and the myxochromide deficient mutant strain indicating that myxochromides are not essential for fruiting body formation. However, both myxochromide overproducers showed a clear deviation from the standard development process as formation of fruiting bodies was significantly delayed and the total number of fruiting bodies was severely reduced relative to the wild type strain. This phenotype seems to be more pronounced in *M. xanthus* DK1622::pTpS-mchS, which produces the highest amounts of myxochromides. High (and constitutive) myxochromide production therefore seems to have a negative impact on fruiting body formation. This finding correlates with a recent study on enhancer binding proteins, in which a mutant with altered secondary metabolite profile (increased myxochromide production, decreased DKxanthene and myxovirescin production) showed similar effects on fruiting body development (*M. xanthus* DK1622hsfA::kan)²⁶. The observed negative effects on fruiting body formation due to myxochromide overproduction could e.g. be explained by myxochromide surfactant properties or myxochromide-mediated increased swarming activity and cell movement, which might hamper cell aggregation.⁷⁰

Swarm expansion assays

Comparative swarm expansion assays were performed in triplicates on 1% CTT agar at 21 °C as described by Kaiser *et al.*,⁴² and were monitored for 2 weeks. The daily increase in swarm expansion and selected images are illustrated in Figure S27. In this assay, no significant difference in swarm expansion between the four analyzed strains was observed. After seven days slightly higher colony diameters were detected in myxochromide overproducing strains, but the myxochromide-deficient mutant also showed marginal higher swarm expansion compared to WT. Based on this assay there is no clear indication that myxochromides contribute to swarming motility as described e.g. for lipopeptides from *Pseudomonas* sp. and *B. subtilis*.^{71,72}

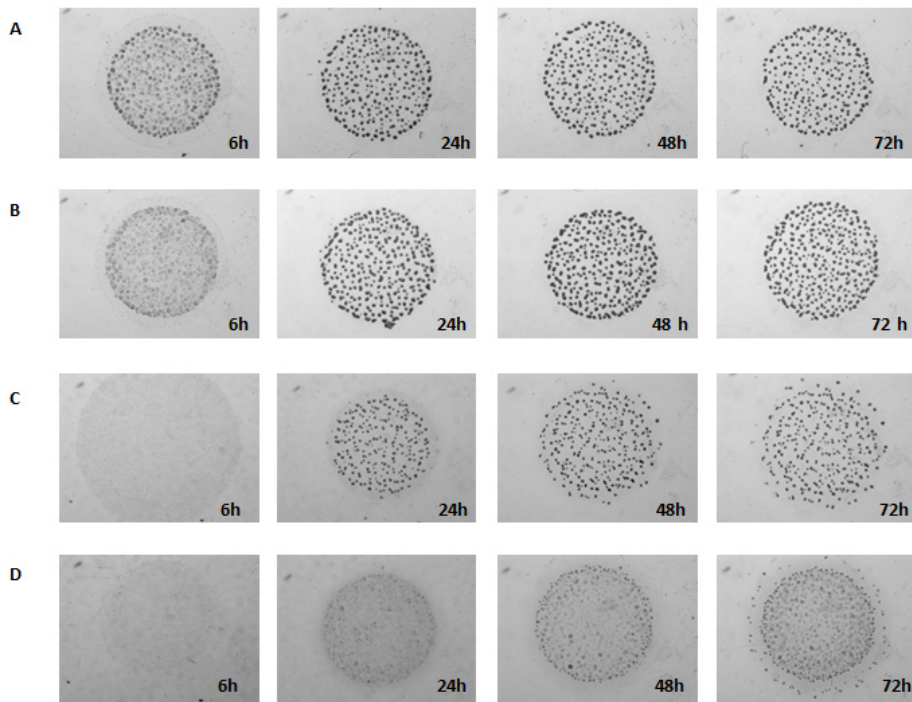


Figure S26. Microscopic images of fruiting body development of different *M. xanthus* strains on TPM agar at 6, 24, 48 and 72 h after inoculation. **A:** *M. xanthus* DK1622 wild type (WT). **B:** *M. xanthus* DK1622::pMch22a (myxochromide deficient). **C:** *M. xanthus* DK1622::pMch70a (~10-fold higher myxochromide production than WT). **D:** *M. xanthus* DK1622::pTpS-mchS (~50-fold higher myxochromide production than WT).

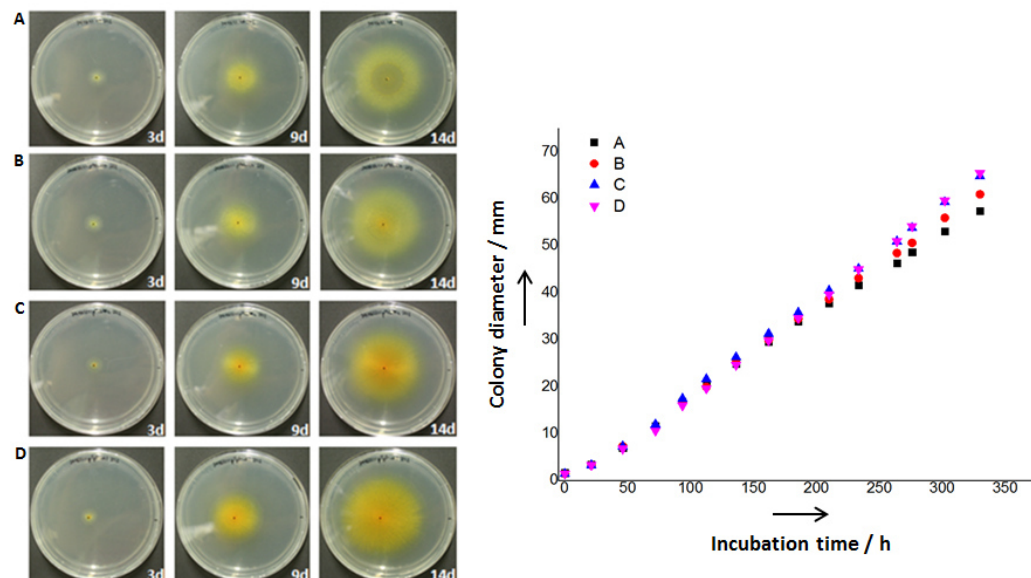


Figure S27. Photographic images of swarming colonies of different *M. xanthus* strains on 1 % CTT agar at 3, 9 and 14 d after inoculation and daily increase in diameter of swarms. **A:** *M. xanthus* DK1622 wild type (WT). **B:** *M. xanthus* DK1622::pMch22a (myxochromide deficient). **C:** *M. xanthus* DK1622::pMch70a (~10-fold higher myxochromide production than WT). **D:** *M. xanthus* DK1622::pTpS-mchS (~50-fold higher myxochromide production than WT).

2.8 References

- (1) Cochrane, S. A.; Vederas, J. C. *Med. Res. Rev.* **2014**, *36*, 4–31.
- (2) Hamley, I. W. *Chem. Commun. (Cambridge, U. K.)* **2015**, *51*, 8574–8583.
- (3) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275–287.
- (4) Strieker, M.; Marahiel, M. A. *ChemBioChem* **2009**, *10*, 607–616.
- (5) Bonmatin, J.-M.; Laprevote, O.; Peypoux, F. *Comb. Chem. High Throughput Screen.* **2003**, *6*, 541–556.
- (6) Baltz, R. H.; Miao, V.; Wrigley, S. K. *Nat. Prod. Rep.* **2005**, *22*, 717–741.
- (7) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
- (8) Ohlendorf, B.; Kehraus, S.; König, G. M. *J. Nat. Prod.* **2008**, *71*, 1708–1713.
- (9) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. *ChemBioChem* **2005**, *6*, 375–385.
- (10) Ziemert, N.; Alanjary, M.; Weber, T. *Nat. Prod. Rep.* **2016**, *33*, 988–1005.
- (11) Wenzel, S. C.; Müller, R. *Nat. Prod. Rep.* **2009**, *26*, 1385–1407.
- (12) Krug, D.; Müller, R. *Nat. Prod. Rep.* **2014**, *31*, 768–783.
- (13) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W. *et al. Nucleic Acids Res.* **2015**, *43*, W237–W243.
- (14) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; Bruijn, I. de; Chooi, Y. H.; Claesen, J.; Coates, R. C. *et al. Nat. Chem. Biol.* **2015**, *11*, 625–631.
- (15) Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. *Nucleic Acids Res.* **2008**, *36*, e113.
- (16) Moss, S. J.; Martin, C. J.; Wilkinson, B. *Nat. Prod. Rep.* **2004**, *21*, 575–593.
- (17) Tsuge, K.; Akiyama, T.; Shoda, M. *J. Bacteriol.* **2001**, *183*, 6265–6273.
- (18) Dittmann, E.; Fewer, D. P.; Neilan, B. A. *FEMS Microbiol. Rev.* **2013**, *37*, 23–43.
- (19) Yue, Q.; Chen, L.; Zhang, X.; Li, K.; Sun, J.; Liu, X.; An, Z.; Bills, G. F. *Eukaryot. Cell* **2015**, *14*, 698–718.
- (20) Roongsawang, N.; Washio, K.; Morikawa, M. *Int. J. Mol. Sci.* **2010**, *12*, 141–172.
- (21) Medema, M. H.; Cimerancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. *PLoS Comput. Biol.* **2014**, *10*, e1004016.
- (22) Bushley, K. E.; Ripoll, D. R.; Turgeon, B. G. *BMC Evol. Biol.* **2008**, *8*, 328.
- (23) Yin, H.; Ma, L.; Wang, G.; Li, M.; Zhang, Z. *Gene* **2016**, *590*, 29–34.
- (24) Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Müller, R. *Appl. Environ. Microbiol.* **2008**, *74*, 3058–3068.
- (25) Raaijmakers, J. M.; Bruijn, I. de; Nybroe, O.; Ongena, M. *FEMS Microbiol. Rev.* **2010**, *34*, 1037–1062.
- (26) Volz, C.; Kegler, C.; Müller, R. *Chem. Biol.* **2012**, *19*, 1447–1459.
- (27) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. *Nat. Prod. Rep.* **2016**, *33*, 317–347.
- (28) Baltz, R. H. *ACS Synth. Biol.* **2014**, *3*, 748–758.
- (29) Bhushan, R.; Bruckner, H. *Amino Acids* **2004**, *27*, 231–247.
- (30) Kears, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. *et al. Bioinformatics* **2012**, *28*, 1647–1649.
- (31) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651–2674.
- (32) Linne, U.; Doekel, S.; Marahiel, M. A. *Biochemistry* **2001**, *40*, 15824–15834.
- (33) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78–92.
- (34) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362–W367.
- (35) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
- (36) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493–505.
- (37) Felsenstein, J. *Cladistics*, *5*, 164–166.
- (38) Sela, I.; Ashkenazy, H.; Katoh, K.; Pupko, T. *Nucleic Acids Res.* **2015**, *43*, W7–W14.
- (39) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
- (40) Sharp, P. M.; Li, W.-H. *Nucleic Acids Res.* **1987**, *15*, 1281–1295.
- (41) Rice, P.; Longden, I.; Bleasby, A. *Trends Genet.* **2000**, *16*, 276–277.
- (42) Kaiser, D.; Warrick, H. *J. Bacteriol.* **2011**, *193*, 5898–5904.

-
- (43) Goldman, B. S.; Nierman, W. C.; Kaiser, D.; Slater, S. C.; Durkin, A. S.; Eisen, J.; Ronning, C. M.; Barbazuk, W. B.; Blanchard, M.; Field, C. *et al. Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15200–15205.
- (44) Green, M. R.; Sambrook, J. *Molecular cloning: A laboratory manual*, 4th ed. / Michael R. Green, Joseph Sambrook; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2012.
- (45) Kears, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. *et al. Bioinformatics* **2012**, *28*, 1647–1649.
- (46) Finn, R. D.; Coghill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A. *et al. Nucleic Acids Res.* **2016**, *44*, D279–285.
- (47) Simunovic, V.; Gherardini, F. C.; Shinkets, L. J. *J. Bacteriol.* **2003**, *185*, 5066–5075.
- (48) Oßwald, C.; Zaburanyi, N.; Burgard, C.; Hoffmann, T.; Wenzel, S. C.; Müller, R. *J. Biotechnol.* **2014**, *191*, 54–63.
- (49) Silakowski, B.; Nordsiek, G.; Kunze, B.; Blöcker, H.; Müller, R. *Chem. Biol.* **2001**, *8*, 59–69.
- (50) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923–936.
- (51) Clugston, S. L.; Sieber, S. A.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2003**, *42*, 12095–12104.
- (52) Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, *39*, 5775–5787.
- (53) Samel, S. A.; Czodrowski, P.; Essen, L.-O. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 1442–1452.
- (54) Minowa, Y.; Araki, M.; Kanehisa, M. *J. Mol. Biol.* **2007**, *368*, 1500–1517.
- (55) Tanovic, A.; Samel, S. A.; Essen, L. O.; Marahiel, M. A. *Science* **2008**, *321*, 659–663.
- (56) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997**, *16*, 4174–4183.
- (57) Samel, S. A.; Schoenafinger, G.; Knappe, T. A.; Marahiel, M. A.; Essen, L. O. *Structure* **2007**, *15*, 781–792.
- (58) Lang, J. M.; Darling, A. E.; Eisen, J. A. *PLoS ONE* **2013**, *8*, e62510.
- (59) Sharp, P. M.; Li, W. H. *Nucleic Acids Res.* **1987**, *15*, 1281–1295.
- (60) Lawrence, J. G.; Ochman, H. *J. Mol. Evol.* **1997**, *44*, 383–397.
- (61) Lawrence, J. G.; Ochman, H. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9413–9417.
- (62) Akashi, H. *Genetics* **1994**, *136*, 927–935.
- (63) Ran, W.; Kristensen, D. M.; Koonin, E. V. *mBio* **2014**, *5*, e00956-14.
- (64) Zaburanyi, N.; Bunk, B.; Maier, J.; Overmann, J.; Müller, R. *Appl. Environ. Microbiol.* **2016**, *82*, 1945–1957.
- (65) Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S. *Cell* **2014**, *157*, 624–635.
- (66) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.
- (67) R Development Core Team. R: a language and environment for statistical computing. <http://www.r-project.org/>.
- (68) Wang, H.; Sivonen, K.; Fewer, D. P. *Curr. Opin. Genet. Dev.* **2015**, *35*, 79–85.
- (69) Jenke-Kodama, H.; Borner, T.; Dittmann, E. *PLoS Comput. Biol.* **2006**, *2*, e132.
- (70) Sliusarenko, O.; Zusman, D. R.; Oster, G. *J. Bacteriol.* **2007**, *189*, 611–619.
- (71) Bruijn, I. de; Raaijmakers, J. M. *J. Bacteriol.* **2009**, *191*, 1910–1923.
- (72) Luo, C.; Zhou, H.; Zou, J.; Wang, X.; Zhang, R.; Xiang, Y.; Chen, Z. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 1897–1910.

Chapter 3

Synthetic Biotechnology to Engineer Myxobacterial Lipopeptide Biosynthesis

Christian Burgard,¹ Fu Yan,¹ Alexander Popoff,¹ Eva Luxenburger,¹ Gregor Zipf,² Hubert S. Bernauer,² Rolf Müller,¹ Silke C. Wenzel¹

Affiliations

¹ Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology at Saarland University, Saarland University Campus, Building E8.1, 66123 Saarbrücken, Germany

² ATG Biosynthetics GmbH, Weberstraße 40, 79249 Merzhausen, Germany

Contributions to the Presented Work

A. Author's Contribution

The author conceived and performed most of the experiments described in this chapter and interpreted the resulting data. The author designed artificial DNA sequences of five known myxochromide biosynthetic gene clusters and established the assembly strategy for artificial gene cluster constructs described, which include the generation of a myxochromide gene library. The author assembled artificial gene constructs harboring synthetic versions of the myxochromide A pathway based on the use of two different type IIS restriction enzymes. Heterologous expression of these constructs in *Myxococcus xanthus* was conducted by the author. In addition, assembly and heterologous expression in *M. xanthus* of five artificial hybrid myxochromide gene clusters as well as seven myxochromide A gene clusters harboring inactive PCP domains at different positions were carried out by the author as well. Furthermore, the author conducted the cultivation of the heterologous mutant strains and prepared extracts for metabolome analysis. The author generated heterologous overproduction mutants for three artificial myxochromide hybrid gene clusters by implementation of the strong constitutive *tn5* promoter. Large-scale cultivation of heterologous production strains for subsequent isolation and structure elucidation of five novel hybrid myxochromides via NMR spectroscopy and analysis of the absolute configuration was performed by the author. The author conceived and wrote this chapter.

B. Contribution by Co-Workers

Fu Yan assembled artificial gene constructs as well as entire expression constructs harboring synthetic versions of the native myxochromide pathways and conducted heterologous expression of these constructs in *M. xanthus*. Furthermore, Fu Yan generated the myxochromide A overproducing mutant via Red/ET recombineering. Alexander Popoff reisolated and repurified myxochromides SB as well as SD and measured and processed the NMR data of the hybrid myxochromides. Eva Luxenburger performed HPLC-MS measurements of crude extracts. Gregor Zipf performed the *in silico* restriction sites engineering on the native myxochromide gene clusters. Hubert S. Bernauer, Rolf Müller and Silke C. Wenzel were responsible for the conception of the project. Proofreading of this chapter was done by Silke C. Wenzel and Rolf Müller.

3 Synthetic Biotechnology to Engineer Myxobacterial Lipopeptide Biosynthesis

3.1 Abstract

The intriguing structural diversity of the myxochromide lipopeptide family originates from evolutionary diversification of the underlying myxochromide megasynthetases, making these pathways a promising model system to further increase the chemical diversity of lipopeptides in myxobacteria using biocombinatorial approaches. In this study, the redesign and assembly of artificial A-, B-, C-, D- and S-type myxochromide biosynthetic gene clusters (*mch* clusters) and hybrid combinations thereof based on synthetic gene fragments is described. A versatile assembly strategy was established which relies on type IIS restriction enzymes and allows for directed exchanges of gene segments coding for PKS/NRPS subunits, modules or even single domains. Heterologous expression of five artificial hybrid *mch* clusters in *Myxococcus xanthus* and secondary metabolite profile analysis of mutant strains revealed the production of engineered hybrid myxochromides, which were subsequently isolated and structurally characterized. Stereochemical analysis of hybrid myxochromides also contributed to the identification of the relevant determinant controlling the stereospecificity of the processed lipopeptide intermediates. Finally, mutated PCP domains based on the inactive module 4 of the native myxochromide S pathway were engineered in every module of the artificial A-type *mch* pathway to evaluate if ‘module-skipping’ scenarios can be induced at different positions of the assembly line. The presented work thus reflects the potential of synthetic DNA platforms not only for pathway assembly and engineering but also for the elucidation of biosynthetic mechanisms.

3.2 Introduction

The generation of novel secondary metabolite analogues via genetic alterations of the corresponding biosynthetic pathways continues to be a promising tool in natural products research to increase the chemical diversity of a certain compound class,^{1,2} as chemical synthesis remains rather challenging or even impossible due to the impressive structural complexity of most natural products. However, rational engineering of polyketide synthase (PKS),³ and nonribosomal peptide synthetase (NRPS) directed biosynthesis,⁴ is often limited by a lack of detailed structural information of the involved megasynthetases, e.g. including the flexible linker regions between individual assembly line subunits, modules and domains,^{5,6} that are important for their functional and dynamical interplay within the large megasynthetase complexes. In addition, several experimental hurdles further impede the success of combinatorial biosynthesis approaches: (1) cultivation and genetic manipulation of

native producer strains is often laborious or in some cases not feasible at all, (2) fermentation yields in the native producers are too low, and (3) the large size of PKS/NRPS biosynthetic gene clusters, which can span from 10 to more than 100 kb,⁷ makes direct cloning from natural sources, rational engineering and subsequent implantation into suitable host genomes difficult. Taken this into account, it is not surprising that only a few examples have been extensively described demonstrating the efficient expression of rationally altered (and closely related) PKS/NRPS biosynthetic gene clusters and the production of the corresponding ‘unnatural’ secondary metabolite derivatives,^{8–11} although the combinatorial reprogramming of PKS/NRPS systems already started more than 30 years ago.¹² Recent advances in the field of synthetic biology have the potential to address these challenges to generate customized production platforms for complex natural products.^{13–16} However, gaining detailed knowledge about relevant factors for the redesign of biosynthetic gene clusters for improved expression of natural product assembly lines is highly desirable as current gene optimization strategies were not as successful as for standard genes. The BMBF funded project ‘SynBioDesign’ basically aimed at understanding the principles of redesigning artificial biosynthetic pathways for heterologous expression in a selected host organism based on synthetic DNA by combining and implementing comprehensive analyses on the genome, transcriptome, proteome and metabolome levels of various gene cluster variants in order to experimentally verify, understand and further improve the *in silico* sequence design. A critical step towards this ambitious task is the development of a flexible and efficient assembly strategy for complex biosynthetic gene clusters to generate numerous variants within a relatively short time frame. This would also allow any desired genetic modification of the gene clusters in order to use this platform to further expand the chemical diversity of the produced compound class by combinatorial approaches. To address these demands, we aimed at generating flexible synthetic DNA platforms based on the well-studied *mch* lipopeptide pathways, as they are encoded in one single operon comprising medium sized four-gene clusters (~ 30 kb). Furthermore, different myxochromides have been heterologously expressed at high yields and did not show any toxicity to their host. In addition, engineering of the precursor supply is not necessary as myxochromides contain only proteinogenic amino acids. Detailed sequence analyses of 16 PKS/NRPS *mch* clusters from various myxobacterial species producing this versatile class of lipopeptides revealed a high degree of sequence homology on the nucleotide and protein level.¹⁷ Moreover, studies on the strain and *mch* cluster phylogeny revealed the close evolutionary relationship between the different *mch* cluster types, which exhibit different product spectra, thereby qualifying these pathways as ideal model systems for

combinatorial studies in the ‘SynBioDesign’ project. Thus, we intended to further increase the structural diversity of the myxochromide lipopeptide family and to study the effects of artificial gene combinations on myxochromide biosynthesis by establishing flexible synthetic DNA platforms for the different *mch* cluster types (A-, B-, C-, D- and S-type).

3.3 Results and Discussion

3.3.1 Sequence Requirements for the Design of Artificial Myxochromide Pathways

To assemble and functionally express an artificial *mch* pathway in a myxobacterial heterologous host, several key requirements for the sequence design need to be considered.

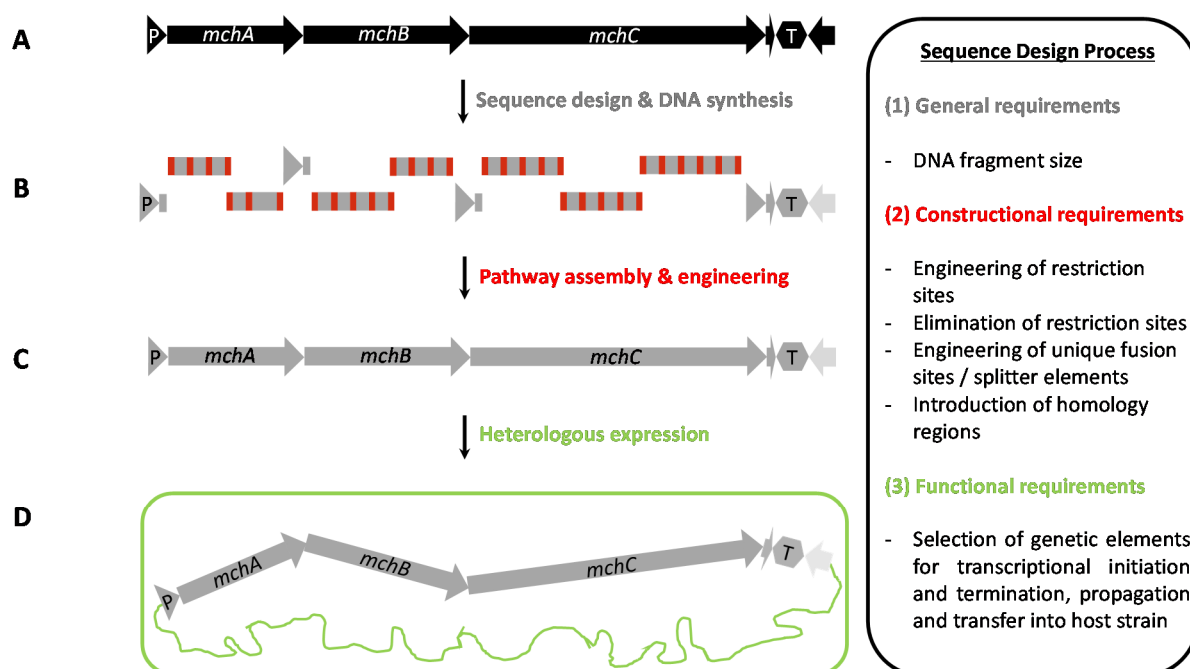


Figure 1. Sequence requirements considered for the generation of synthetic *mch* expression platforms. **A:** Gene organization of the native myxochromide A biosynthetic gene cluster from *M. xanthus* DK1622. **B:** De novo redesign and DNA synthesis of 11 building blocks for pathway assembly. **C:** In vitro reconstitution of the artificial A-type pathway. **D:** Integration into the host genome and functional expression of the artificial A-type pathway.

General sequence requirements encompass the demands on DNA synthesis (1). As the size of a biosynthetic gene cluster is far too large to be synthesized in one piece via standard gene synthesis approaches, it has to be subdivided into smaller DNA segments. Constructional sequence requirements consider the specific demands of the envisaged assembly strategy to construct the desired expression constructs (2). The assembly and downstream modification of entire *mch* cluster constructs particularly required the choice and engineering of unique

restriction sites (R-sites) between the defined synthetic DNA fragments and the elimination of appropriate R-sites along the whole *mch* cluster sequences as well as the introduction of a homology region for the specific integration of the artificial pathways into the genome of the heterologous host. Functional sequence requirements are defined by the host strain selected for heterologous expression and comprise the choice of suitable genetic elements to achieve gene expression such as promoter and terminator sequences as well as elements to ensure construct propagation and stability (3). An overview on the applied strategy is exemplified for the A-type *mch* pathway from *M. xanthus* DK1622 (Figure 1).¹⁸

3.3.2 Functional Sequence Design for Heterologous Expression in *M. xanthus*

According to the current state of research, *Myxococcus xanthus* seems to be the most promising heterologous host for the expression of myxobacterial natural product pathways and numerous studies have been previously reported, which support this assumption.¹⁹ Since the A- and S-type *mch* pathways were already known to be well expressed in *M. xanthus* DK1622,^{18,20} this myxobacterial strain was selected as a heterologous host for functional expression of the artificial *mch* cluster constructs. A myxochromide A-deficient mutant of *M. xanthus* DK1622 (*M. xanthus* DK1622 Δ *mchA-tet*, Wenzel *et al.*, unpublished), in which the native A-type *mch* cluster was replaced by an (oxy-)tetracycline resistance cassette, was already available in the group as heterologous production strain.

To deliver and functionally express the synthetic *mch* pathways in *M. xanthus* DK1622 Δ *mchA-tet*, the expression vector pSynbio2 was designed and manufactured by DNA synthesis (Supporting Information Figure S2). The minimal vector backbone includes a p15A low-copy origin of replication to ensure stability of the large *mch* cluster constructs during propagation in *E. coli*, an antibiotic resistance gene (kanamycin, *kanR*) suitable for selection of *M. xanthus* DK1622 Δ *mchA-tet* transformants, an origin of transfer (*oriT*), to allow for conjugation as an alternative strategy to transformation via electroporation, and a multiple cloning site (MCS), which is composed of all the R-sites needed for pathway assembly and engineering (Supporting Information Table S2). In addition, unique R-sites were introduced between the genetic elements (*PacI*, *PmeI*, *SwaI*) to allow for the exchange or addition of vector backbone elements by conventional cloning techniques, *e.g.* to investigate other chromosomal integration sites. Integration of the artificial *mch* constructs into the former *mchA* locus in the genome of *M. xanthus* DK1622 Δ *mchA-tet* was intended via single crossover by using the downstream helicase gene *rhIE* as homologous region (which is part of

the 3'*mchC-mchD-rhlE* terminator fragments, see Section 'Constructional Sequence Design of Artificial Myxochromide Pathways').

To achieve gene expression in *M. xanthus*, we decided to place all artificial *mch* pathways under the control of the native promoter from the A-type *mch* cluster from *M. xanthus* DK1622. Moreover, to ensure the transcriptional termination of artificial *mch* pathways originating from different myxobacterial species (B-type *mch* pathway from *Myxococcus* sp.,¹⁷ C-type *mch* pathway from *Myxococcus virescens*,¹⁷ D-type *mch* pathway from *Stigmatella erecta*,¹⁷ S-type *mch* pathway from *Stigmatella aurantica*,²¹) in *M. xanthus*, the 3'*mchC-mchD-rhlE* terminator fragments were designed in a way that the terminator sequence (plus the *rhlE* gene) from the A-type *mch* cluster was integrated downstream of the *mchD* genes from the other *mch* pathways to generate hybrid terminator fragments, which is schematically illustrated in Figure 2.

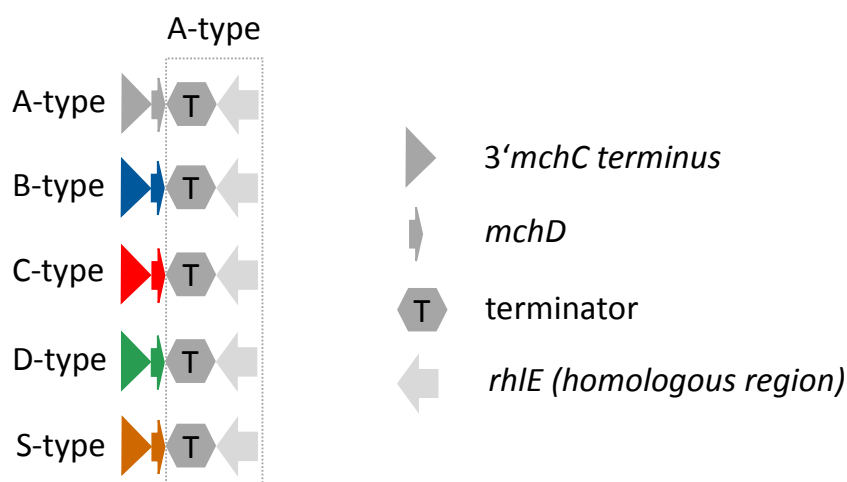


Figure 2. Design of hybrid terminator fragments for artificial A-, B-, C-, D- and S-type *mch* pathways. The individual 3'*mchC-mchD-rhlE* terminator fragments contain the native 3'*mchC* ends including the *mchD* gene of each *mch* cluster type, which are genetically fused to the downstream helicase gene (*rhlE*) including the terminator sequence from the A-type *mch* pathway (illustrated in gray) as a homologous region to allow for specific integration into the host genome and for functional termination of transcription in *M. xanthus*.

3.3.3 Constructional Sequence Design of Artificial Myxochromide Pathways

To address the generation of synthetic DNA platforms for the production of complex nonribosomal peptides, the development of a versatile assembly strategy for large and GC-rich myxochromide gene clusters was required, which includes the possibility to (1) assemble numerous *mch* cluster variants fast and efficiently using a combination of conventional cloning techniques and modern DNA assembly strategies based on the use of type IIS restriction enzymes, (2) to allow for flexible interchanges of *mch* cluster segments on the domain, module and subunit level based on the generation of a *mch* gene library to generate

novel ‘unnatural’ peptide cores via combinatorial biosynthesis and (3) to allow for the interchangeability of genetic elements for vector backbone modifications. Initially, the A-type *mch* cluster from *M. xanthus* DK1622,¹⁸ was subjected to the sequence design process. As the size of a DNA fragment to be synthesized is still a limiting factor of DNA fabrication, the A-type *mch* operon, which is approximately 30 kb in size, was rationally dissected into seven segments comprising the promoter-5′*mchA* (P5*mchA*) and 3′*mchC*-*mchD*-terminator-*rhIE* (T3*mchC*) regions, the three 5′/3′ end truncated biosynthesis genes *mchA*, *mchB* and *mchC* and the two intergenic linker regions 3′*mchA*-5′*mchB* (3A5B) and 3′*mchB*-5′*mchC* (3B5C) between the biosynthesis genes. The large *mchC* gene (13.4 kb) was subdivided into three fragments (MchC_A_AarI_fragABCE, MchC_A_AarI_fragD and MchC_A_AarI_fragF) for gene synthesis ranging in size from 6.2 to 9.4 kb. To meet the objectives described above, we aimed at establishing a three-step assembly strategy comprising the assembly of the truncated biosynthesis genes, the generation of a gene library for combinatorial approaches and the assembly of entire *mch* cluster constructs using a combination of conventional restriction/ligation cloning techniques and state-of-the-art assembly technologies.

To allow for the future interchangeability of every desired PKS or NRPS domain, so-called splitter elements (SE) were designed and implemented into the interdomain linker regions within the truncated biosynthesis gene fragments *mchA*-*C* (except the linker between the PKS enoylreductase domain and the acyl carrier protein domains). The SEs basically consist of recognition sequences of a type IIS restriction enzyme (Figure 3).

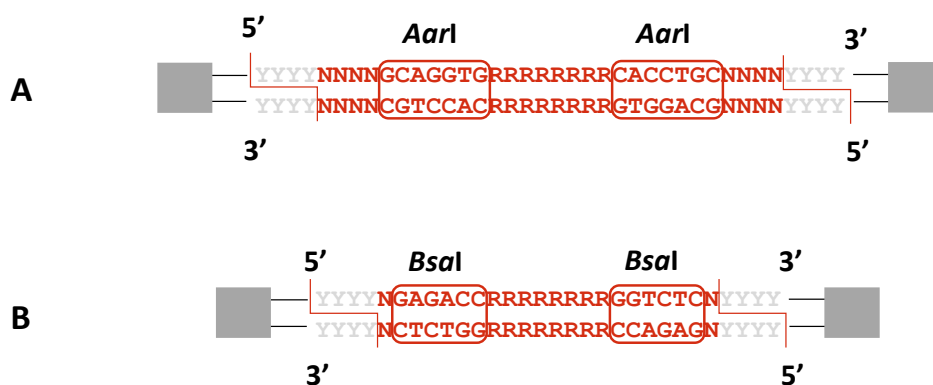


Figure 3. Structure and composition of splitter elements (SE) used for *mch* cluster engineering. The SE were placed between each PKS/NRPS domain fragment (except between the ER-ACP domain fragments, illustrated as dark gray boxes) and contain two recognition sites for a type IIS restriction enzyme (highlighted in red boxes), which hydrolyzes the DNA double strand outside of this recognition sequence, thereby generating 4 bp 5′ overhangs (shown in light gray). The introduced sequence in the middle (shown in red, R) represents the recognition site of a unique conventional type II restriction enzyme to allow for gene assemblies and fragment exchanges. **A:** SE based on the type IIS restriction enzyme *AarI* (32-34 bp in size). **B:** SE based on the type IIS restriction enzyme *BsaI* (26-28 bp in size).

These enzymes exhibit the special ability to hydrolyze the DNA double strand outside of their asymmetric recognition sequence,²² which enables the selection and design of unique overhangs between the domain fragments of each biosynthesis gene. This approach allows for the directed and seamless reassembly of the biosynthesis genes at the second stage of the assembly process. In the first instance, suitable fusion sites in the interdomain linkers along the A-type *mch* cluster were defined and introduced SEs were equipped with the recognition sequences of the type IIS restriction enzyme *AarI*.²³ Moreover, an additional R-site for a unique ‘standard’ type II restriction enzyme between the *AarI* recognition sites was introduced for downstream domain or module exchanges and for the assembly of large biosynthesis genes (*mchC*) from smaller synthetic DNA fragments. Based on the domain organization of the *mch* biosynthesis genes, the *mchA* gene was equipped with 4 SEs, *mchB* with 6 and in *mchC*, 12 SEs were introduced with an overall size of 32-34 bp. After gene assembly using these conventional R-sites within the SEs, SEs need to be eliminated by a ‘desplitting’ process, which basically describes the hydrolysis of the biosynthesis genes using *AarI*. The released SEs can easily be removed using PCR purification columns and the single domain fragments can be rejoined in a one-pot multi-fragment ligation exploiting the unique 4 bp overhangs as compatible fusion sites. Following this approach led to the generation of a gene library containing fully ‘desplitted’ gene constructs. The last stage of the assembly process is the stitching of the three ‘desplitted’ biosynthesis gene fragments together with four additional fragments harboring the promoter, terminator and intergenic linker regions described above to assemble functional *mch* cluster constructs. In addition to the internal SEs between the domain encoding fragments, external SEs were implemented at the 5’ and 3’ ends of each synthetic cluster fragment in a way that the generated 4 bp overhangs are part of another unique R-sites, which were engineered within the coding sequence by silent mutations (*BsiWI*, *MreI*, *MluI*, *NotI*, *SphI*, *AgeI*) whenever possible (Supporting Information Table S5). This concept allows in parallel the assembly of entire *mch* cluster constructs from the seven building blocks either in a one-pot fashion or their conventional stitching as a backup strategy, if the one-pot assembly fails. In total, R-sites of a set of 25 enzymes used for gene assembly, domain/module engineering and pathway reconstitution were eliminated along the A-type *mch* cluster sequence, including R-sites for the exchange of vector backbone elements (*PacI*, *PmeI*, *SwaI*) on the designed pSynbio2 vector.

3.3.4 Construction and Heterologous Expression of a Synthetic A-type *mch* Cluster

The *de novo* synthesis of the redesigned A-type *mch* cluster fragments based on the *AarI* design approach was accomplished by a gene synthesis company commissioned by our bioinformatics collaborator ATG:Biosynthetics GmbH. The synthesis of the genes *mchA-C* required high delivery times (~ 3-5 months) due to their large size and complexity (GC-rich sequences), whereas the significantly smaller promoter (*P5mchA*), terminator (*T3mchC*) and intergenic linker fragments (*3A5B/3B5C*) were delivered in reasonable time frames (several weeks). In the case of the large *mchC* gene, we received three individual DNA fragments, since synthesis of this gene in one piece was not successful. Thus, we followed the envisaged assembly strategy and first assembled the full-length *mchC* gene in two steps using the unique R-sites within the introduced SEs for the construction of pGH-MchC_A_AarI_fragABCEF and subsequent complementation with pGH-MchC_A_AarI_fragD to construct pGH-MchC_A_AarI_SE. Since each synthetic fragment was delivered in the standard pGH vector backbone, the seven building blocks were subcloned into our cloning vector pSynbio1 that was subjected to the R-sites engineering and thus does not contain any recognition sequences of the restriction enzymes needed for ‘desplitting’ or pathway reconstitution/engineering, which is a prerequisite for the downstream stages of the assembly process. After *mchC* gene assembly and subcloning, the truncated biosynthesis genes *mchA-C* were successfully ‘desplitted’ via hydrolysis with *AarI* followed by religation of the single domain fragments after removal the SEs. Although the ‘desplitting’ process of the biosynthesis gene fragments was initially successful, cloning efficiencies were insufficiently low in terms of establishing a fast and efficient assembly strategy. ‘Desplitting’ of the smaller *mchA* and *mchB* gene constructs already required some screening effort (one correct clone out of 24) and ‘desplitting’ of the large *mchC* gene containing 12 SEs was difficult to achieve at all. In this case, up to 100 colonies needed to be tested to find a correct clone harboring a fully ‘desplitted’ version of this gene construct. In addition, partially ‘desplitted’ *mchC* constructs were often observed, which still contained some SEs despite of the performed purification step using the PCR column. In some other cases, fully ‘desplitted’ but shortened *mchC* constructs were detected, in which several domain fragments were deleted. This can happen if three out of the four nucleotides making up the fusion sites are complementary to each other, so that they can anneal and subsequently be ligated.²⁴ The type IIS restriction enzyme *AarI* additionally requires a synthetic oligonucleotide to gain full activity and thus seemed to be the bottleneck, although it was recently used in studies describing the similar Golden Gate assembly strategy.^{24,25} Constructs, which were thought to be fully ‘desplitted’ based on restriction analyses, were sequenced by employing Illumina sequencing technology in order

to guarantee a verified gene library. However, we continued with the third stage of the assembly strategy and tried to assemble the first artificial A-type *mch* pathway from the seven building blocks plus the expression vector pSynbio2 (which contains a MCS flanked by two SEs generating unique fusion sites to capture the promoter-5' *mchA* and 3' *mchC-mchD*-terminator-*rhIE* fragments and thus acts as acceptor vector) via separate hydrolysis of each construct using *AarI* followed by a directed one-pot religation of the linearized synthetic fragments to form functional *mch* cluster constructs. Unfortunately, one-pot assemblies of artificial *mch* pathways failed, which is most likely due to the diverse size distribution of the synthetic fragments (150 bp up to 13.4 kb). The assembly of ~ 35 kb constructs from non-standardized DNA fragments (in terms of fragment size) might thus be hard to achieve in a highly efficient and flexible way. In contrast, the similar Golden Gate Cloning platform for the assembly of large DNA constructs from modular building blocks, which relies on type IIS restriction enzymes and their special features, has recently become popular describing the custom-specific generation of Transcription Activator-Like Effector Nuclease (TALEN) libraries used for *in vivo* gene editing.^{26,27} In these and other studies, modules exhibiting similar sequence lengths in equimolar amounts were used for one-pot restriction/ligation assemblies and relatively small constructs were assembled (up to 10 kb). Taken this into account, fragment and final construct sizes might be the relevant factors limiting the success of this assembly strategy. Thus, the first artificial A-type *mch* pathway was stepwise constructed from the seven building blocks via conventional cloning using the engineered unique R-sites (R_{x1} , R_{x6} , R_{y1} , R_{y8} , R_{z1} , R_{z14}) plus the two R-sites flanking the multiple cloning site of the expression vector pSynbio2 (R_L , R_R), which revealed the intermediate constructs pSyn2-ca1, pSyn2-ca2, pSyn2-ca3, pSyn2-ca4, pSyn2-ca5, pSyn2-ca6 and the final expression construct pSynMch1 harboring the entire synthetic *mchA* cluster (Supporting Information Table S9). The expression construct was verified by a combination of restriction analyses and Illumina sequencing prior to transformation into *M. xanthus* DK1622. Integration of the artificial A-type *mch* pathway into the former locus of the corresponding native pathway in *M. xanthus* was analyzed by PCR (Supporting Information Figure S4) and genotypically verified mutant strains were cultivated for subsequent preparation of culture extracts. As controls, the wild type strain as well as the myxochromide A-deficient mutant *M. xanthus* DK1622 $\Delta mchA-tet$ were cultivated in parallel. Culture extracts were analyzed for myxochromide production by HPLC-MS and target masses for myxochromides A₂, A₃ and A₄ were detected in the extracts of the wild type strain as well as of the mutant strains harboring the artificial version of the A-type *mch* cluster (experimental data not shown). Myxochromide

production titers were comparable in both strains showing that the synthetic DNA platform is principally a robust and competitive system for further studies. However, the assembly strategy and thus the constructional sequence design needed to be modified to significantly improve the robustness and efficiency of the described procedure.

3.3.5 Adaption of the Constructional Sequence Design to an Alternative Type IIS Restriction Enzyme

To have an efficient assembly strategy in hand for which the presented concept still builds the basis, the constructional sequence design was adapted to the alternative type IIS restriction enzyme *BsaI*, which is also known from published DNA assembly protocols,^{24,25} and which does not require any additives to become active. However, the use of *BsaI* has some limitations regarding the strategic possibilities of the envisaged assembly strategy, e.g. the *BsaI* enzyme only skips one nucleotide until it hydrolyzes the DNA double strand compared to four nucleotides, which are skipped by *AarI*. Thus, it was not possible anymore to design the SE-derived 4 bp overhangs in a way that they are simultaneously part of the ‘conventional’ R-site recognition sequences at the 5’/3’ ends of the synthetic DNA fragments, which allowed the construction of synthetic *mch* clusters via both one-pot and conventional assemblies. Thus, we only focused on the conventional construction of entire artificial *mch* constructs by using the unique R-sites at the 5’/3’ ends of each synthetic DNA fragment. Furthermore, and in addition to the A-type *mch* pathway, the strategy was extended to the B-, C-, D- and S-type *mch* clusters in order to generate a diversified gene library for combinatorial biosynthesis.

First, we decided to redesign the whole A-type *mch* cluster based on the sequence requirements, which were determined by the adaption to the *BsaI* design plus a minimal set of cluster fragments from the other described *mch* pathways, which allows their rational recombination to assemble a maximum number of different *mch* cluster variants to produce novel myxochromides (Supporting Information Figure S3). Thus, the *mchA* gene from the A-type pathway encoding the polyketide synthase that directs the biosynthesis of the polyunsaturated side chain, was subjected to the sequence design process and was subsequently used for the assembly of all *mch* pathways described in this study (plus the corresponding promoter-5’*mchA* and the 3’*mchA*-5’*mchB* intergenic linker fragment). Since the MchB subunits of the A-, B- and C-type *mch* pathways share a relatively high degree of sequence homology, originate from *Myxococcus* sp. and activate the same amino acids for incorporation into the growing peptide chain (in contrast to the MchB subunit from the S-type

mch pathway), the *mchB* gene from the A-type *mch* cluster was redesigned and used for the construction of artificial A-, B- and C-type *mch* clusters. To assemble an artificial D-type *mch* cluster, the corresponding *mchB* gene from *Stigmatella* sp. was redesigned and, in addition, the S-type *mchB* gene since the corresponding S-type MchB subunit exhibits a different amino acid specificity. Due to various recombination events and the resulting differences in the *mchC* genes of the *mch* clusters, we redesigned the *mchC* genes from all *mch* pathways (including the corresponding 3'*mchB*-5'*mchC* intergenic linker and terminator fragments). Sequence alignments of the selected *mch* genes were performed in order to find suitable positions for the *BsaI*-based SEs (or fusion sites respectively) between each domain encoding fragment. We intended to choose fusion sites, which were located at the same positions among the different *mch* genes to allow for directed exchanges of domains or modules originating from *mch* genes from different *mch* clusters. In some cases, it was not possible to detect identical 4 bp combinations in the interdomain linker regions among the different *mch* genes. Alternative fusion sites were then chosen, which were located either within the upstream or downstream domain based on revised domain annotations. However, overall 15 out of 72 fusion sites needed to be modified by silent point mutations (except one fusion site mutation which affected the amino acid sequence of the corresponding protein) to be identical and located at the same positions among all *mch* genes (Supporting Information Table S4). In the B-type *mch* gene, three additional fusion sites were selected and appropriate SEs were introduced (15 instead of 12 as for the A-, D- and S-type *mchC* genes) due to the duplicated module described in Chapter 2. Analogously, the C-type *mch* gene was only equipped with 9 SEs as a result of the module deletion.

In analogy to the *AarI*-based design process, the *mch* genes were truncated at their 5' and 3' ends and were further subdivided into smaller DNA fragments due to their large size (overall 11 synthetic cluster fragments for the A-, D- and S-type pathways; 10 cluster fragments for the C-type pathway and 12 cluster fragments for the B-type pathway). The unique R-sites at the 5' and 3' ends, which were previously defined, were introduced at the same positions as described for the *AarI* design. Since the one-pot assembly of entire *mch* cluster constructs did not work efficiently using *AarI* most likely due to the diverse size distribution of the synthetic cluster fragments, we decided to assemble the expression constructs via conventional cloning methods using these unique R-sites. Thus, only the truncated *mch* gene fragments were equipped with external SEs at their 5'/3' ends to allow for the religation of the gene fragments into the pSynbio1 cloning vector after the 'desplitting' process.

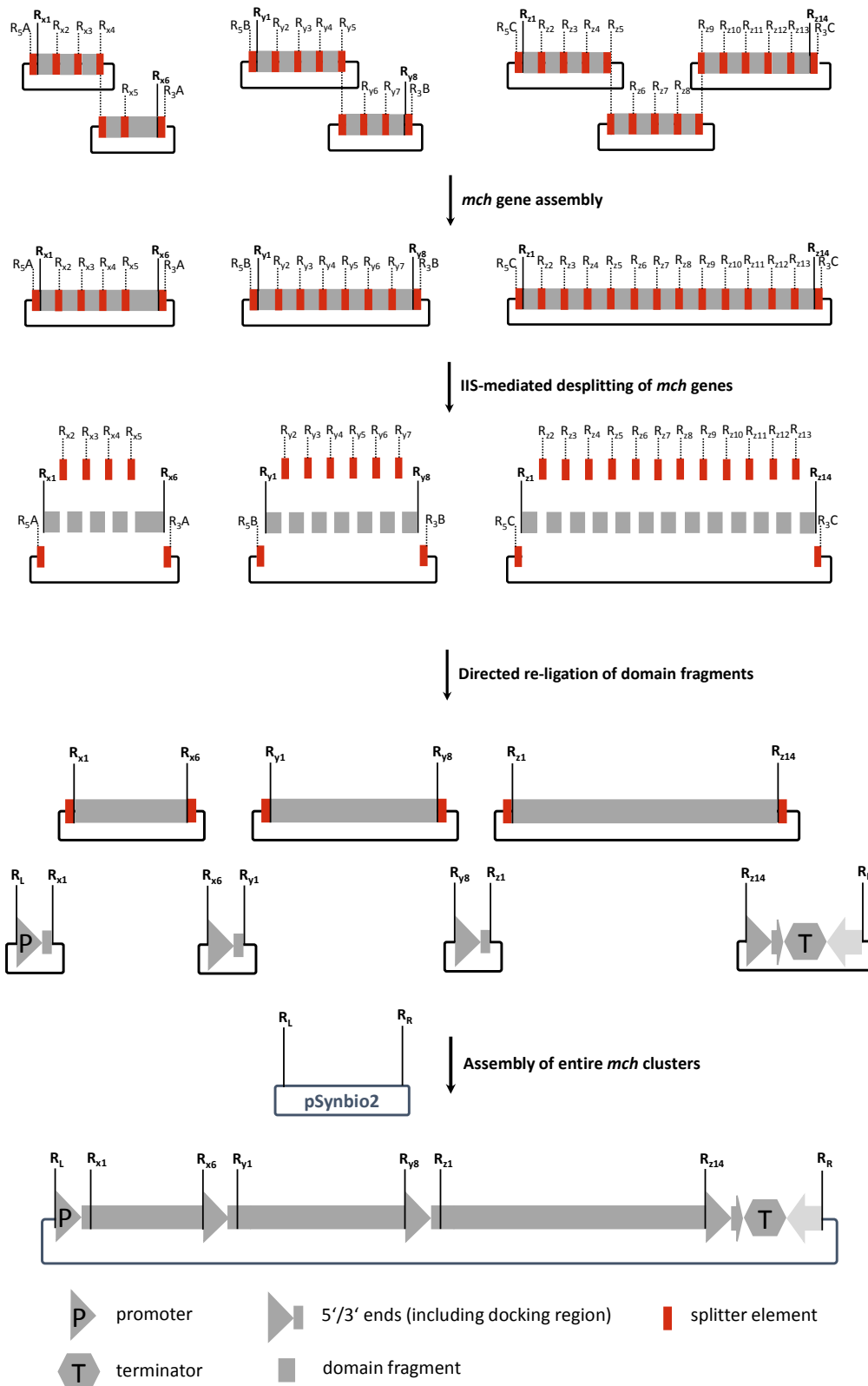


Figure 4. Assembly strategy for the generation of artificial *mch* clusters. The assembly process involves overall three steps starting with the construction of the biosynthesis genes from smaller synthetic DNA fragments. ‘Desplitting’ of the biosynthesis genes using type IIS restriction enzymes followed by directed religation of the single domain fragments results in the generation of a gene library consisting of splitter-free gene constructs. To reconstitute the entire *mch* clusters, seven synthetic building blocks comprising the gene constructs from the gene library as well as four additional cluster fragments harboring the promoter, intergenic linker and terminator sequences are stepwise cloned into the expression vector pSynbio2 using the unique R-sites at the 5' and 3' ends of each synthetic cluster fragment (R_L, R_{x1}, R_{x6}, R_{y1}, R_{y8}, R_{z1}, R_{z14}, R_R).

In addition, R-sites engineering was performed for the selected *mch* cluster fragments to eliminate perturbing R-sites. The whole assembly strategy starting from gene assemblies over ‘desplitting’ of *mch* gene constructs to the reconstitution of artificial *mch* clusters for heterologous expression is exemplified for the A-type *mch* cluster in Figure 4 (for a detailed overview on the designed *mch* cluster fragments see Supporting Information Figure S3).

In addition to the *mch* cluster fragments, the cloning (pSynbio1_AarI) and expression vectors (pSynbio2_AarI) also needed to be modified to meet the constructional sequence requirements of the *BsaI* design. To eliminate a *BsaI* R-site in the ampicillin resistance gene in pSynbio1, a small synthetic analogous fragment was designed, in which this R-site was removed by silent point mutation and which can be easily exchanged by restriction/ligation cloning (Supporting Information Table S7). In pSynbio2, a modified MCS equipped with flanking *BsaI* R-sites instead of *AarI* sites was designed and exchanged using unique R-sites on the original vector backbone (Supporting Information Table S7).

3.3.6 *In vitro* Reconstitution of Artificial Hybrid *mch* Clusters

Despite the modification and expansion of the described constructional sequence design, the basic assembly strategy remained unaffected as we still intended to construct the *mch* pathways via the established three-step approach starting with the assembly of the split biosynthesis genes using the unique R-sites within the SEs followed by subcloning of these fragments into the cloning vector pSynbio1, which revealed the following constructs: pSyn1-MchA_A_SE, pSyn1-MchB_A_SE, pSyn1-MchB_D_SE, pSyn1-MchB_S_SE, pSyn1-MchC_A_SE, pSyn1-MchC_B_SE, pSyn1-MchC_C_SE, pSyn1-MchC_D_SE and pSyn1-MchC_S_SE. At the second stage, hydrolysis of these constructs with *BsaI* and subsequent religation of domain fragments gave the fully ‘desplitted’ genes providing a versatile gene library for the combinatorial reprogramming of myxochromide assembly lines. Compared to the ‘desplitting’ procedure using *AarI*, ‘desplitting’ reactions with the *BsaI* type IIS restriction enzyme appeared to be much more efficient. The ‘desplitted’ versions of the smaller *mchA* (pSyn1-MchA_A) and *mchB* (pSyn1-MchB_A, pSyn1-MchB_D and pSyn1-MchB_S) gene constructs were easily obtained without putting significant effort in the screening of correct clones (cloning efficiencies were ~ 50-70%). However, cloning efficiencies dropped noticeably together with the number of SEs or construct size, respectively. ‘Desplitting’ of the largest *mchC* gene construct (pSyn1-MchC_B_SE) was not successful, although ‘desplitting’ reactions were optimized in terms of enzyme and substrate concentrations, incubation time and temperature needed for the efficient religation to form the fully ‘desplitted’ construct.

Thus, we designed four modified *mchC* gene fragments from the B-type *mch* pathway (MchC_B_fragA_woSE, MchC_B_fragB_woSE, MchC_B_fragC_woSE and MchC_B_fragD_woSE), in which the SEs were completely removed (except one *BsaI* site plus the pre-defined unique R-site at the 5'/3' ends of each fragment for gene assembly as a kind of partial SE) in order to reduce the complexity of the 'desplitting' reaction. The *mchC* gene was constructed from the modified gene fragments via conventional cloning methods using the unique R-sites, thereby generating SEs at the ligation sites. Removal of the resulting three SEs via hydrolysis using *BsaI* efficiently worked, thereby revealing pSyn1-MchC_B, which underpins the assumption that the efficiency of the 'desplitting' process is mainly limited by the number of SEs. 'Desplitting' of the *mchC* gene constructs from the A-, C-, D- and S-type *mch* pathways harboring 10 to 12 SEs was still possible and revealed pSyn1-MchC_A, pSyn1-MchC_C, pSyn1-MchC_D and pSyn1-MchC_S, even though the cloning efficiencies were lowered compared to *mchA/mchB* construct 'desplittings' (Supporting Information Table S8). A number of 12 SEs thus seems to be the benchmark in the described experimental set-up, for which reasonable cloning efficiencies can be realized.

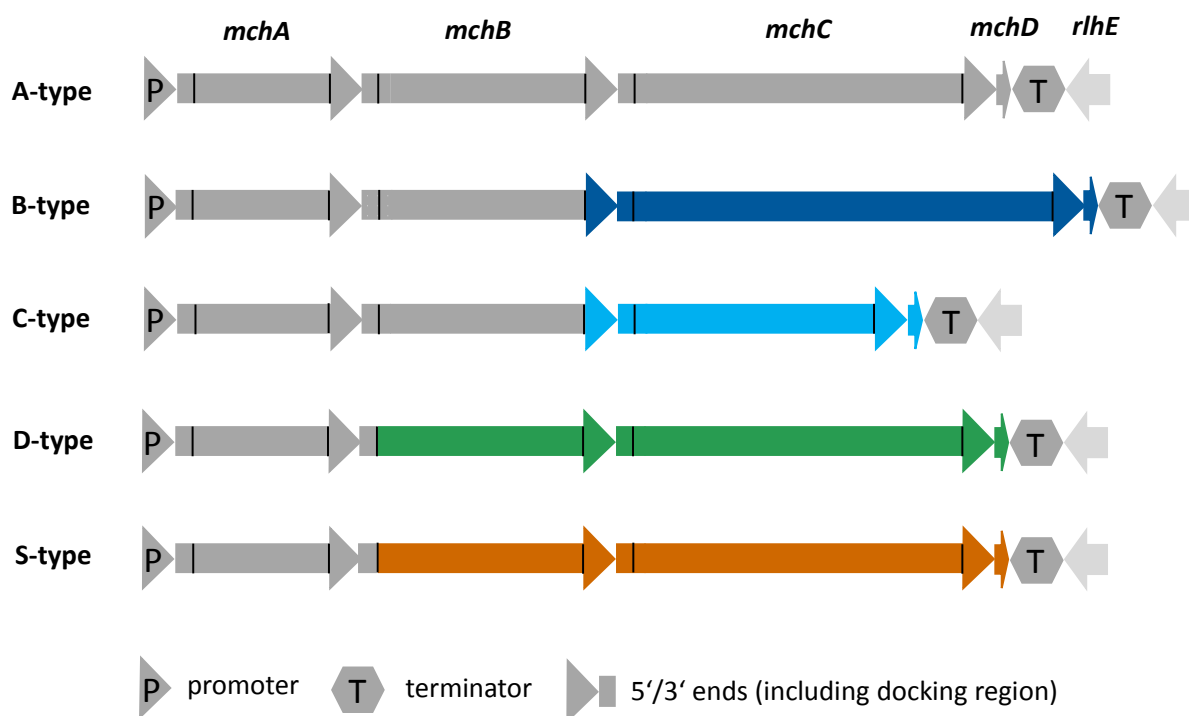


Figure 5. Schematic representation of the five artificial hybrid *mch* clusters generated by rational recombination of the seven synthetic building blocks. The illustrated pathways are assumed to direct the biosynthesis of the naturally occurring myxochromides A, B, C, D and S. The color code illustrates the origin of the single redesigned *mch* cluster fragments: synthetic building blocks from the A-type *mch* pathway are shown in dark gray, from the B-type *mch* cluster in blue, from the C-type *mch* cluster in red, from the D-type pathway in green and from the S-type pathway in orange. Black lines indicate the positions of unique R-sites for the conventional assembly of the *mch* cluster constructs from the synthetic building blocks.

On the basis of the generated *mch* gene library, we first intended to create synthetic DNA platforms for the production of the native myxochromide scaffolds, which serve as a starting point for combinatorial engineering to increase the chemical diversity of myxochromides. By rationally recombining the seven synthetic building blocks including the different *mch* genes from the gene library, five hybrid *mch* pathways were constructed via conventional cloning, which were assumed to be responsible for the production of the native A-, B-, C-, D- and S-type myxochromides (Figure 5, work by Dr. Fu Yan, unpublished).

Heterologous expression of the artificial *mch* pathways in *M. xanthus* followed by genotypic verification of selected mutant strains and HPLC-MS analysis of culture extracts revealed that all pathways were correctly integrated into the host chromosome and were functionally expressed. The corresponding A-, B-, C-, D- and S-type myxochromides were detected in culture extracts of the mutant strains (data not shown; work by Dr. Fu Yan, unpublished).

We then intended to go one step further to generate entirely new myxochromide peptide cores, which have not (yet) been found in Nature, by exchanging the MchB/MchC subunits, which might change the amino acid order in the produced lipopeptide compounds.

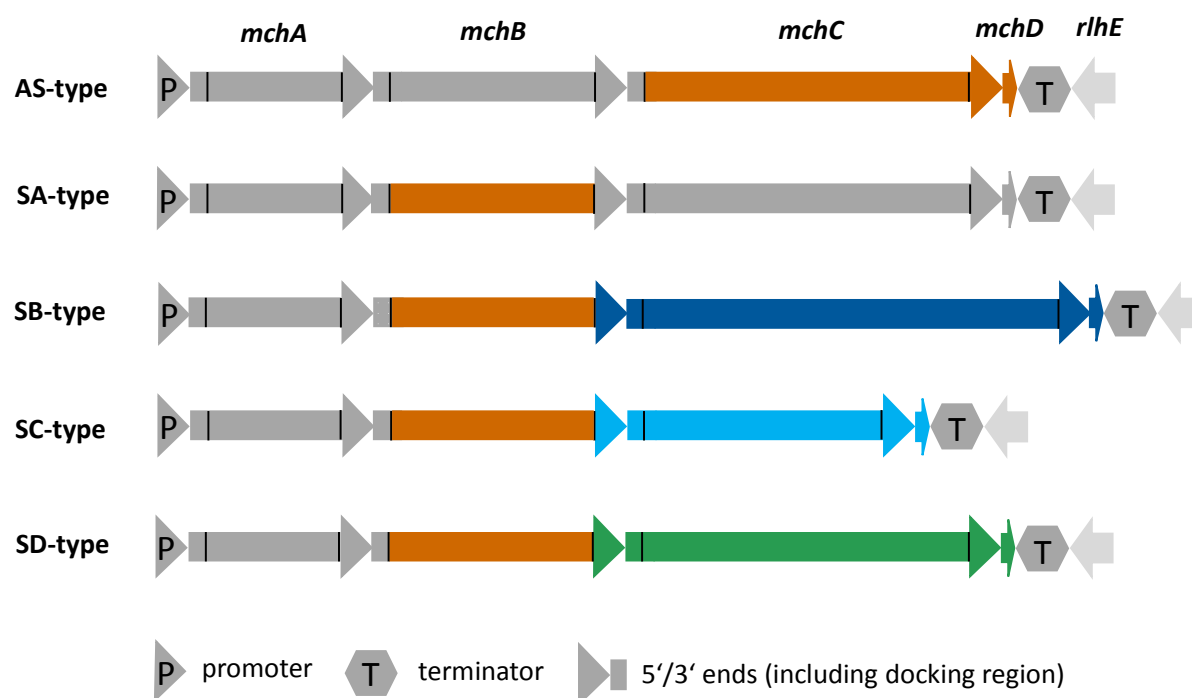


Figure 6. Schematic representation of the five artificial hybrid *mch* clusters generated by rational swaps of the *mch* genes from the gene library. The illustrated pathways are assumed to direct the biosynthesis of novel hybrid myxochromide cores (AS-, SA-, SB-, SC- and SD-type). The color code illustrates the origin of the single redesigned *mch* cluster fragments: synthetic building blocks from the A-type *mch* pathway are shown in dark gray, from the B-type *mch* cluster in blue, from the C-type *mch* cluster in red, from the D-type pathway in green and from the S-type pathway in orange. Black lines indicate the positions of unique R-sites for the conventional assembly of the *mch* cluster constructs from the synthetic building blocks.

Using the gene library consisting of the ‘desplitted’ biosynthetic *mch* genes, five different *mch* gene combinations can be deduced, which in conjunction with the available promoter-5′*mchA* (P5*mchA*), intergenic linker 3′*mchA*-5′*mchB* (3A5B), 3′*mchB*-5′*mchC* (3B5C) and 3′*mchC*-*mchD*-terminator-*rhIE* (T3*mchC*) fragments might encode biosynthetic machineries producing novel myxochromide cores (Figure 6). Alternative gene combinations would inevitably lead to *mch* pathways, which are assumed to be responsible for the production of the native myxochromide scaffolds.

Since we intended to focus on the engineering of the NRPS (MchB/MchC) subunits to switch amino acid positions in the peptide core, the PKS (MchA) encoding *mchA_A* gene from the A-type *mch* pathway was consistently used as a standard building block for the assembly of the mentioned hybrid *mch* clusters together with the P5*mchA_A* promoter fragment and the 3A_A5B_A intergenic linker fragment, which derived from the A-type *mch* cluster as well. We decided to use the 3A_A5B_A linker fragment from the A-type *mch* cluster to ensure native docking interactions at the corresponding PKS/NRPS interface in the megasynthetase complex rather than using hybrid linker fragments which combine the 3′ end of the *mchA_A* and the 5′ end of the *mchB_S* fragment.

Similar considerations were made for the integration of the 3*mchA*-5*mchB* (3B5C) linker fragments. We intended to maintain the native docking regions between *mchB_A*/*mchB_S* and the different *mchC* gene fragments and thus used the individual linker fragments harboring the 3′*mchB* and 5′*mchC* ends originating from the different *mch* cluster types (3B_A5C_A, 3B_B5C_B, 3B_C5C_C and 3B_D5C_D). If one NRPS encoding gene (*mchB*/*mchC*) derived from the A-type *mch* pathway, the corresponding A-type intergenic linker fragment (3B_A5C_A) was used (Figure 6). Based on the established gene library, the different hybrid *mch* clusters (Figure 6) were constructed stepwise in four to seven cloning steps using the unique R-sites at the 5′/3′ ends of the synthetic *mch* cluster fragments, thereby yielding the final expression constructs, which harbor the artificial AS- (pSynMch8), SA- (pSynMch9), SB- (pSynMch11), SC- (pSynMch10) and SD-type (pSynMch12) hybrid *mch* clusters (Supporting Information Table S9).

All expression constructs were fully sequenced by employing Illumina sequencing technology. In some cases, frameshift mutations were detected in common parts of the expression constructs, which needed to be repaired prior to heterologous expression (Supporting Information Table S9).

3.3.7 Heterologous Expression of Artificial Hybrid *mch* Clusters in *M. xanthus* and Production Analysis of Mutant Strains

To heterologously produce the assumed novel hybrid myxochromides in *M. xanthus* DK1622, the expression constructs pSynMch8 (AS-type), pSynMch9 (SA-type), pSynMch10 (SC-type), pSynMch11 (SB-type) and pSynMch12 (SD-type) harboring the artificial hybrid *mch* clusters were subsequently transformed into the heterologous host strain *M. xanthus* DK1622 $\Delta mchA-tet$ lacking the native A-type *mch* cluster. Integration into the host genome via single crossover was achieved using the *rhIE* gene, which is part of the terminator fragments (3'*mchC-mchD*-terminator-*rhIE*) and originated from the native A-type *mch* pathway as a homologous region (see Chapter 3.3.2). Correct integration into the former locus of the A-type *mch* cluster was analyzed by PCR using two sets of primers at the 5' and 3' ends of the synthetic *mch* clusters generating mutant-specific PCR products (Supporting Information Figure S4) and selected mutant strains harboring the hybrid *mch* clusters were grown in duplicates in 50 mL CTT medium (casitone 1%, Tris-HCl [pH 8.0] 10 mM, K₂HPO₄/KH₂PO₄ buffer [pH 7.6] 1 mM, MgSO₄ × 7 H₂O 8 mM, pH adjusted to 7.6) at 30°C and analyzed for myxochromide production via HPLC-MS (Figure 7). To relatively compare hybrid myxochromide production titers with the native production levels of myxochromide A in *M. xanthus* DK1622, the wild type strain was analyzed as well. The myxochromide A-deficient mutant strain *M. xanthus* DK1622 $\Delta mchA-tet$ served as a negative control. Production of novel lipopeptide cores (Figure 8) was successfully demonstrated and is summarized below:

Myxochromides AS – novel engineered lipopentapeptides

Myxochromide derivatives with [M+H]⁺ masses corresponding to the expected hybrid myxochromides AS₂, AS₃ and AS₄ (**1a-c**, Figures 7/8) were detected in extracts of *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch8$. The most prominent derivative was **1a** under the applied cultivation conditions, whereas derivatives **1b** and **1c** were produced in lower yields but at similar levels compared to each other (Figure 7).

Myxochromides SA – novel engineered lipohexapeptides

Myxochromide derivatives with [M+H]⁺ masses corresponding to the expected novel hybrid lipohexapeptides myxochromides SA₂, SA₃ and SA₄ (**2a-c**, Figures 7/8) were detected in extracts of *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch9$. Compound **2b** was found to be the major derivative under the applied cultivation conditions, whereas derivatives **2a** and **2c** were produced in significantly lower amounts (Figure 7).

Myxochromides SB – novel engineered lipopeptptides

Myxochromide derivatives with $[M+H]^+$ masses corresponding to the assumed hybrid lipopeptptides myxochromides SB₂, SB₃ and SB₄ (**3a-c**, Figures 7/8) were detected in extracts of *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch11$. The major derivative was **3c** under the applied cultivation conditions followed by **3b** and **3a**, which were produced in significantly lower amounts (Figure 7).

Myxochromides SC – novel engineered lipopentapeptides

Myxochromide derivatives with $[M+H]^+$ masses corresponding to the expected novel lipopentapeptides myxochromides SC₂, SC₃ and SC₄ (**4a-c**, Figures 7/8) were detected in extracts of *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch10$. The production profile was found to be highly similar to that of the myxochromide SB producing mutant strain (Figure 7). Compound **4c** was identified as most prominent derivative under the applied cultivation conditions, whereas derivatives **4a** and **4b** were produced as minor products.

Myxochromides SD – novel engineered lipopentapeptides

Myxochromide derivatives with $[M+H]^+$ masses corresponding to the expected hybrid myxochromides SD₂, SD₃ and SD₄ (**5a-c**, Figures 7/8) were detected in extracts of *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch12$. In comparison to the other hybrid myxochromides, production of **5a-c** was found to be significantly lower, and was also observed for several independent clones. The major derivative was **5b** under the applied cultivation conditions and was roughly produced at similar levels as some minor compounds in the other extracts. Derivatives **5a** and **5c** were produced in even lower yields (Figure 7).

The successful heterologous expression in *M. xanthus* demonstrated the functionality of the artificially recombined hybrid *mch* clusters, which were constructed based on synthetic DNA. NRPS subunits from different *mch* pathways, which were rationally recombined in a non-native way, were obviously able to successfully interact with each other. In addition, non-native biosynthetic intermediates, which are not biosynthesized by the native *mch* biosynthetic machineries, were transferred to and processed by the downstream domains of the hybrid *mch* pathways, thereby providing fully functional synthetic DNA platforms for the production of entirely novel myxochromide lipopeptide cores.

When comparing the production titers of the synthetic hybrid *mch* clusters in the heterologous mutant strains with the native myxochromide A production in *M. xanthus* DK1622, it turned out that the corresponding hybrid myxochromides AS₂₋₄, SA₂₋₄, SB₂₋₄ and SC₂₋₄ were produced at different production levels in this initial cultivation experiment (Figure 7/Table1).

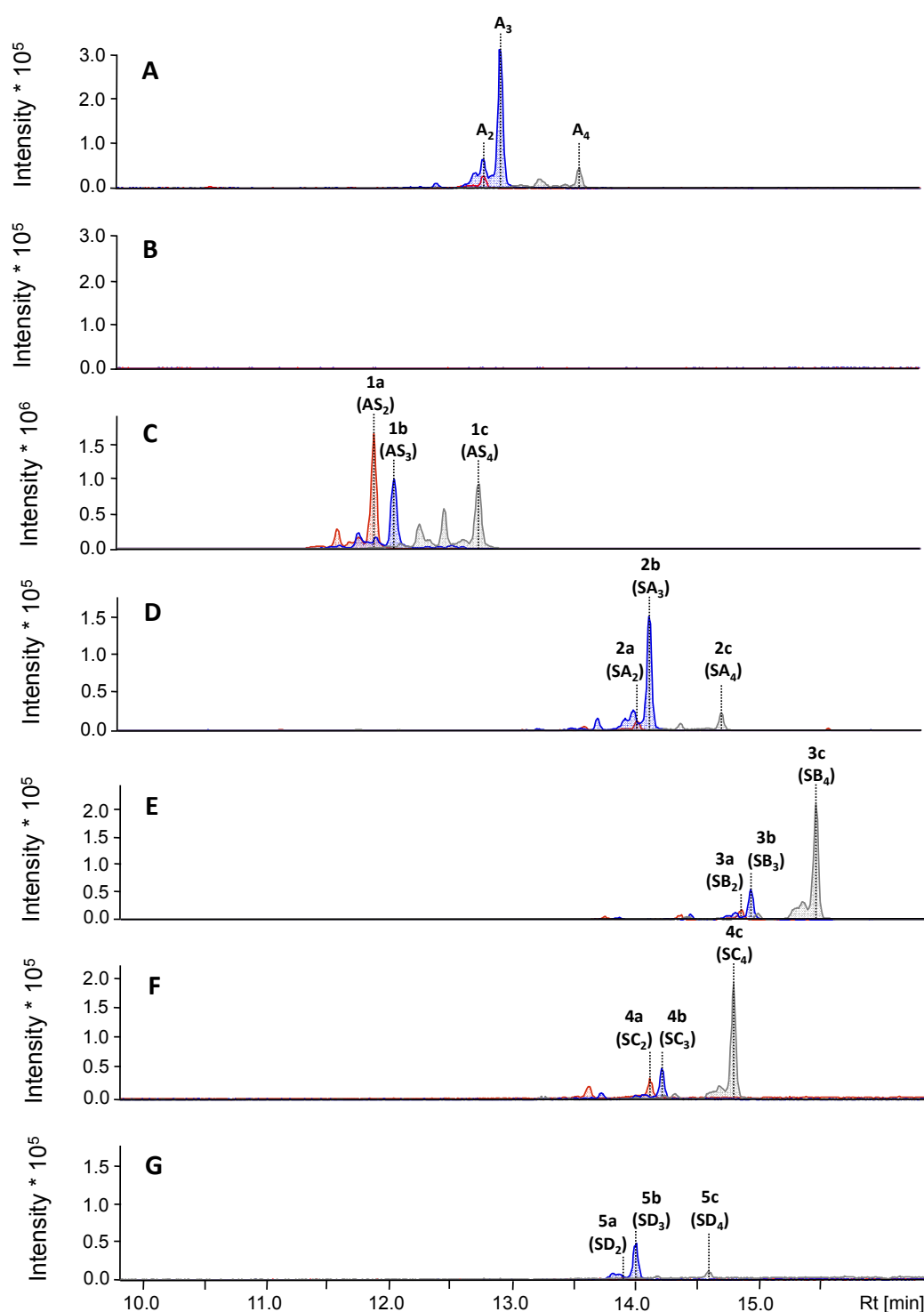


Figure 7. HPLC-MS analysis of hybrid myxochromide production in myxobacterial mutant strains. Extracted ion chromatograms (EICs) for ± 0.02 m/z corresponding to the $[M+H]^+$ ions of myxochromides are shown. **A:** Detection of myxochromides A in *M. xanthus* DK1622 wild type; A₂ ($[M+H]^+ = 834.47655$), A₃ ($[M+H]^+ = 846.47655$) and A₄ ($[M+H]^+ = 860.49220$). **B:** No myxochromide production in *M. xanthus* DK1622 $\Delta mchA$ -tet. **C:** Detection of myxochromides AS in *M. xanthus* DK1622 $\Delta mchA$ -tet::pSynMch8; AS₂ ($[M+H]^+ = 695.37684$), AS₃ ($[M+H]^+ = 707.37684$), AS₄ ($[M+H]^+ = 721.39249$). **D:** Detection of myxochromides SA in *M. xanthus* DK1622 $\Delta mchA$ -tet::pSynMch9; SA₂ ($[M+H]^+ = 876.52350$), SA₃ ($[M+H]^+ = 888.52350$), SA₄ ($[M+H]^+ = 902.53915$). **E:** Detection of myxochromides SB in *M. xanthus* DK1622 $\Delta mchA$ -tet::pSynMch11; SB₂ ($[M+H]^+ = 989.60757$), SB₃ ($[M+H]^+ = 1001.60757$), SB₄ ($[M+H]^+ = 1015.62322$). **F:** Detection of myxochromides SC in *M. xanthus* DK1622 $\Delta mchA$ -tet::pSynMch10; SC₂ ($[M+H]^+ = 805.48639$), SC₃ ($[M+H]^+ = 817.48639$), SC₄ ($[M+H]^+ = 831.50204$). **G:** Detection of myxochromides SD in *M. xanthus* DK1622 $\Delta mchA$ -tet::pSynMch12; SD₂ ($[M+H]^+ = 779.47074$), SD₃ ($[M+H]^+ = 791.47074$), SD₄ ($[M+H]^+ = 805.48639$).

Table 1: Chromatographic data of detected mxochromide derivatives.

Compound	R _t [min]	Peak area	Intensity
Myxochromide A ₂	12.8	68958	25765
Myxochromide A ₃	12.9	1009846	319991
Myxochromide A ₄	13.6	128337	42760
Myxochromide AS ₂	11.9	5448946	1659018
Myxochromide AS ₃	12.1	3684638	997237
Myxochromide AS ₄	12.7	3746519	925440
Myxochromide SA ₂	14.0	34121	13702
Myxochromide SA ₃	14.1	450801	153900
Myxochromide SA ₄	14.7	59352	23614
Myxochromide SB ₂	14.8	49470	17853
Myxochromide SB ₃	14.9	178012	56136
Myxochromide SB ₄	15.5	639513	206972
Myxochromide SC ₂	14.1	69915	31898
Myxochromide SC ₃	14.2	137546	50067
Myxochromide SC ₄	14.8	551340	186612
Myxochromide SD ₂	13.9	n.d. ^{a)}	3252
Myxochromide SD ₃	14.0	138982	46565
Myxochromide SD ₄	14.6	n.d. ^{a)}	11506

^{a)} not determined; signal-to-noise ratio too low

Whilst myxochromides SA₂₋₄, SB₂₋₄ and SC₂₋₄ were produced at approximately 50-70% relative to the production of myxochromides A₂₋₄ in the wild type strain, the hybrid myxochromides AS₂₋₄ were produced at significantly higher titers (increased production by a factor of 10 relative to myxochromide A₂₋₄ production). Consequently, these hybrid *mch* pathways might be efficiently expressed in the host strain since the A-type *mch* pathway from *M. xanthus* DK1622 is a prominent example for a well-expressed myxobacterial PKS/NRPS gene cluster showing relatively high production yields. However, production of the SD-type myxochromide family was found to be significantly reduced (<12% relative to myxochromide A production levels). Analysis of the codon usage in the D-type *mchC* gene in correlation to the host genome's codon usage suggested that the D-type *mchC* gene might be suitably expressed in *M. xanthus* DK1622, although codon usage of this gene is slightly lower adapted to the host genome compared to the S-type *mchC* gene, which was shown to be efficiently

expressed as part of the AS-type hybrid *mch* cluster (data on codon usage analysis not shown). This assumption is supported by the fact that the production titer of the synthetic D-type *mch* cluster is in the same range as observed for the native A-type *mch* cluster as well as for the synthetic SA-, SB- and SC-type *mch* clusters described above (production profile not shown). Thus, we speculate that the chimeric MchB_{SD} protein consisting of the MchB_S subunit linked to the C-terminus of the MchB_D subunit could potentially somehow influence proper protein-protein interactions between the NRPS subunits, thereby affecting the efficiency of the SD-type hybrid *mch* pathway. However, it cannot be excluded that the D-type MchC subunit is not able to efficiently process the non-native biosynthetic intermediate, which is supplied by the upstream MchB subunit or that the D-type *mchC* gene is not properly expressed for other reasons. Interestingly, analysis of the hybrid myxochromide production profiles also revealed that the different hybrid *mch* pathways show significant differences in the preference for the polyunsaturated side chain generated by the PKS subunit, although the PKS encoding *mchA* gene from the A-type *mch* pathway was specifically used for all hybrid *mch* pathways, thereby providing identical pairs of interaction partners (*mchA_A/mchB_A* or *mchA_A/mchB_S*). In addition, the preference for a specific polyketide side chain varies not only between the different mutant strains but sometimes also within the same mutant strain when changing the culture conditions (e.g. the culture volume as observed for some hybrids when upscaling the production for compound isolation). For those reasons, it is not possible to draw rational conclusions from this initial cultivation experiment regarding the biosynthetic mechanisms which determine the preference for the polyene side chains. To ultimately prove the existence of the five novel hybrid myxochromide cores, isolation and structure elucidation was carried out for one prioritized derivative of each myxochromide type (Figure 8).

3.3.8 Isolation and Structure Elucidation of Hybrid Myxochromides

Since the production profiles of the hybrid myxochromides significantly changed in some cases regarding the major derivatives as well as the production yields, the native promoter was exchanged for the strong constitutive *Tn5* promoter, which was recently established in *M. xanthus*,²⁰ in the SA-, SB- and SD-type *mch* pathways to improve production titers and to simplify the isolation of the corresponding hybrid myxochromides. To structurally characterize the novel hybrid myxochromides in detail, one representative of each hybrid myxochromide type was isolated from the corresponding heterologous production mutant strains and their planar structures were solved by NMR spectroscopy. The structure elucidation process indeed confirmed five novel lipopeptide scaffolds belonging to the

myxochromide family (Figure 8). The underlying hybrid *mch* biosynthetic machineries are obviously capable of directing the biosynthesis of the engineered myxochromides. The number and order of the amino acids of the generated peptide cores unambiguously conform to the expected and artificially engineered hybrid structures. In comparison to the naturally occurring myxochromide lipopeptide cores, single amino acids of the peptide backbones were successfully switched at several positions (modules 2/3), thereby creating entirely new scaffolds based on the rational recombination of the synthetic *mch* biosynthetic genes from the established gene library. Elucidation of the absolute configuration of the characterized hybrid myxochromides also revealed the reason for the incorporation of only *L*-configured amino acids by the native S-type assembly line despite the presence of an epimerization (E) domain in module 2, which led to the incorporation of *D*-configured alanine in case of A-, B-, C- and D-type *mch* pathways.^{17,21} Whereas the presence of *D*-alanine in myxochromides A,¹⁸ B,²⁸ C and D correlates with the presence of the epimerase domain in module 2 (E₂) of the underlying *mch* biosynthetic machineries, the amino acids in myxochromides S are invariably *L*-configured, although the leucine residue, which is incorporated by module 2, is predicted to be *D*-configured as well. In theory, several assumptions can be made to explain this observation: (1) the E₂ domains are somehow inactive in the native S-type *mch* pathways, although mutations which were detected might not be critical for exhibiting catalytic activity,^{17,29,30} (2) the E₂ domains of the *mch* pathways are evolutionary adapted to act on alanine and the more bulky leucine substrate cannot be converted by the E₂ domain due to its increased steric hindrance or (3) the downstream condensation domains of modules 3 (C₃) in the S-type *mch* pathways, which were predicted to be *D*-specific for the peptidyl donor and *L*-specific for the aminoacyl acceptor (^DC_L domain),³¹ are actually *L*-specific condensation domains for both donor and acceptor (^LC_L domain) and are not able to further process *D*-configured peptidyl intermediates.¹⁷

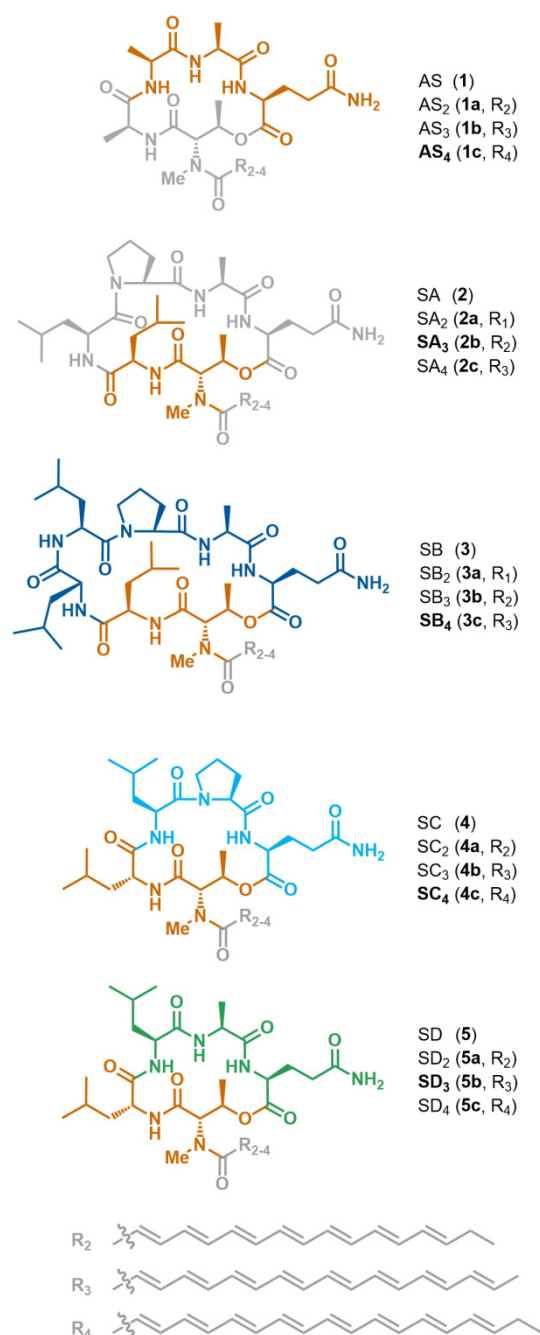


Figure 8. Chemical structures of novel hybrid myxochromides. Structurally characterized myxochromide derivatives are highlighted in bold. Minor derivatives, which were not characterized in detail, were postulated based on HPLC-MS analysis. The corresponding assembly lines are shown in Figure 9.

These hypotheses were verified by determination of the absolute configuration of the five hybrid myxochromides using Marfey's method,³² exploiting the non-native combinations of A-type E₂ domains with S-type C₃ domains and vice versa (plus combinations of S-type E₂ domains with B-, C- and D-type C₃ domains). In myxochromide AS₄, all amino acids were found to be *L*-configured including the alanine residue, which is incorporated by module 2. In

contrast, 1:1 ratios of *D*-leucine/*L*-leucine were detected in hydrolysates of myxochromides SA₃, SC₄ and SD₃, whereas 1:2 ratios of *D*-leucine/*L*-leucine were found in the hydrolysate of myxochromide SB₄ (Supporting Information Chapter 3.4.8). According to the biosynthetic logic of the underlying hybrid *mch* pathways (Figure 9), the detected *D*-configured leucine residues in myxochromides SA₃, SB₄, SC₄ and SD₃ must originate from epimerization by the obviously active E₂ domains of the corresponding modules 2. Consequently, these results exclude the hypotheses that the E₂ domains of the native S-type *mch* pathways are inactivated by mutations or that the E₂ domains cannot exert their catalytic function on the more bulky leucine substrate. In fact, the sequence-based models for predicting the specificities of NRPS condensation domains (^DC_L/^LC_L) seems to be not accurate enough to reliably predict the preferences of condensation domains for *D*- or *L*-configured peptide intermediates, as the C₃ domains of the S-type *mch* pathways were obviously predicted as ^DC_L domains by mistake. In conclusion, the E₂ domains of all *mch* pathways are indeed active and are not specific for alanine alone. The C₃ domains of the downstream modules are the key determinants of the observed stereochemistry in both native and hybrid myxochromides, thereby fulfilling their role as gatekeepers in the *mch* megasynthetases.

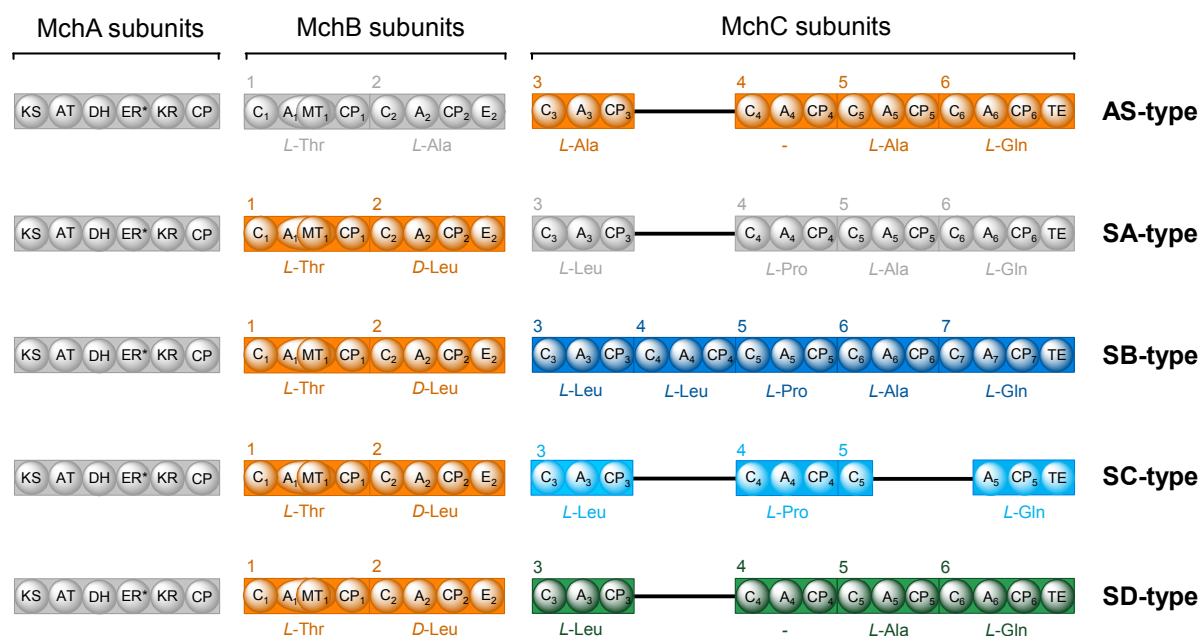


Figure 9. Schematic representation of hybrid myxochromide assembly lines. The coloring of the PKS and NRPS subunits corresponds to the biosynthesis products shown in Figure 8. Catalytic domain abbreviations: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; CP, acyl carrier protein (in MchA) or peptidyl carrier protein (in MchB/C); C, condensation domain; A, adenylation domain; MT, methyltransferase; E, epimerization domain; TE, thioesterase. Domains marked with an asterisk are supposed to be inactive. The corresponding myxochromide structures are shown in Figure 8.

3.3.9 PCP Inactivation Experiments to Induce Module-Skipping Events

To further exploit the tool box given by the established highly flexible assembly strategy, we intended to not only recombine *mch* genes from the different *mch* pathways but also to exchange selected catalytic domains using the unique R-sites within the SEs. We initially focused on the NRPS encoding *mchC* genes from the A-type *mch* pathway as well as from the S-type *mch* cluster and targeted the ‘module-skipping’ process, which was recently described to occur in myxochromide S biosynthesis.²¹ Based on sequence alignments, it is assumed that the observed ‘module-skipping’ process is the result of a mutation (Ser → Pro) in the PCP core motif of module 4, which leads to the loss of the catalytically active serine residue, thereby inactivating the PCP domain, which cannot be post-translationally modified.^{33–35} In addition, several other mutations were detected in the core motif (GGHSL → GGNPS) in comparison with the PCP core motif, which corresponds to the described core motif sequence of PCP domains.³⁶ We therefore aimed at answering the question whether it is possible to induce ‘module-skipping’ in the A-type *mch* pathway by mutating PCP domain core motifs in a similar way as detected in the S-type *mch* cluster. By applying this approach to module 4 of the A-type *mch* cluster, a D-type *mch* pathway might emerge in terms of the activated amino acids, which would be incorporated by the corresponding assembly line and can be used to mimic evolutionary diversification by turning one *mch* pathway into another. The opposite strategy, the reactivation of module 4 in the S-type *mch* cluster by using artificial PCP domains harboring an active serine residue, would potentially result in an entirely new *mch* pathway responsible for the production of a novel myxochromide scaffold. To achieve this, two PCP₄ encoding fragments from the A-type pathway were designed, which harbor either the same core motif as the S-type PCP₄ domain (Figure 10A) or which harbor a Ser → Ala mutation as an alternative (Figure 10A). The latter mutation was recently described to inactivate or to reduce the flexible motion of PCP domains of NRPS megasynthetases in activity-related and structural studies.^{37,38} In addition, a synthetic PCP₄ encoding fragment based on the S-type *mch* pathway was designed in a way that the core motif from the A-type *mch* cluster was reconstituted by applying several mutations and which was assumed to restore the activity of the corresponding S-type PCP₄ domain (Figure 10B). Mutations were rationally introduced at certain positions within the core motif to meet the constructional sequence requirements for the gene cluster assemblies, which basically means that the generation of R-sites used for pathway assembly and engineering by the exchange of single nucleotides needs to be avoided. The synthetic fragments were flanked with SEs, which allow the exchange of the PCP domain fragments based on the *mch* genes from the gene library using the unique R-sites and the directed religation of the domain fragments after the

‘desplitting’ process by exploiting the unique fusion sites. To further expand the scope of this approach, appropriate mutated PCP domain fragments were designed for the other modules of the synthetic A-type *mch* cluster (CP1, CP2, CP3, CP5, CP6) as well. In those cases, only one version harboring the described Ser → Ala mutation was designed. By following this strategy, we aimed at inducing ‘module-skipping’ at different positions of the synthetic A-type *mch* cluster to verify the hypothesis that ‘module-skipping’ can potentially be induced by point mutations, thereby further increasing the chemical diversity of myxochromides.

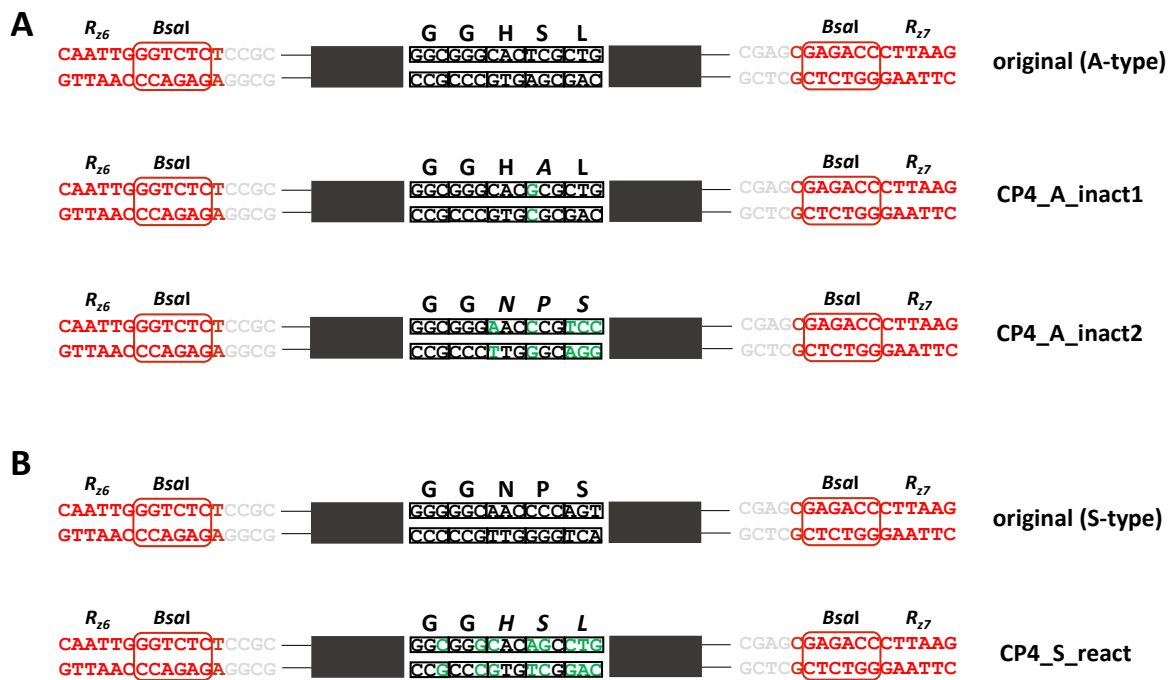


Figure 10. Design of synthetic DNA fragments encoding mutated PCP domains (illustrated as dark gray boxes). The core motifs of the artificial PCP domains are particularly highlighted. DNA fragments are flanked with SEs harboring unique R-sites for the exchange of domain fragments in the *mch* genes from the gene library (*R₂₆*/*R₂₇*). R-sites are highlighted in red, unique overhangs (fusion sites) in light grey and mutated nucleotides in green (referred to the native nucleotide sequences of the corresponding original PCP domains). **A:** Synthetic PCP₄ domain fragments from the A-type *mch* pathway harboring either a Ser → Ala mutation (CP4_A_inact1) or the S-type derived amino acid mutations (CP4_A_inact2) in the core motif. For comparison, the original PCP₄ domain fragment from the synthetic A-type *mch* pathway is also shown. **B:** Synthetic PCP₄ domain fragment from the S-type *mch* pathway harboring mutations to reconstitute the core motif observed in the native A-type PCP₄ domain. For comparison, the original PCP₄ domain fragment from the synthetic S-type *mch* pathway is shown.

The respective modified *mch* gene constructs were constructed following a three-step cloning approach, which is exemplarily shown for a PCP₄ exchange in the *mchC* gene (Figure 11). The mutated PCP domain fragments described above were cloned into the artificial *mchB* (pSyn1-MchB_A_SE) and *mchC* (pSyn1-MchC_A_SE or pSyn1-MchC_S_SE) gene constructs harboring the SEs, followed by ‘desplitting’ of these constructs and subsequent

religation of the domain fragments, which resulted in the generation of fully ‘desplitted’ constructs complementing the *mch* gene library. Integration of the modified *mch* biosynthesis genes harboring the described mutations into a synthetic A-type *mch* cluster, which is under the control of the strong, constitutive *Tn5* promoter (pSynMch13, constructed by Dr. Fu Yan, unpublished) was achieved using the unique R-sites at the 5’ and 3’ ends of the generated constructs (R_{y1}/R_{y8} for *mchB* gene constructs and R_{z1}/R_{z14} for *mchC* gene constructs).

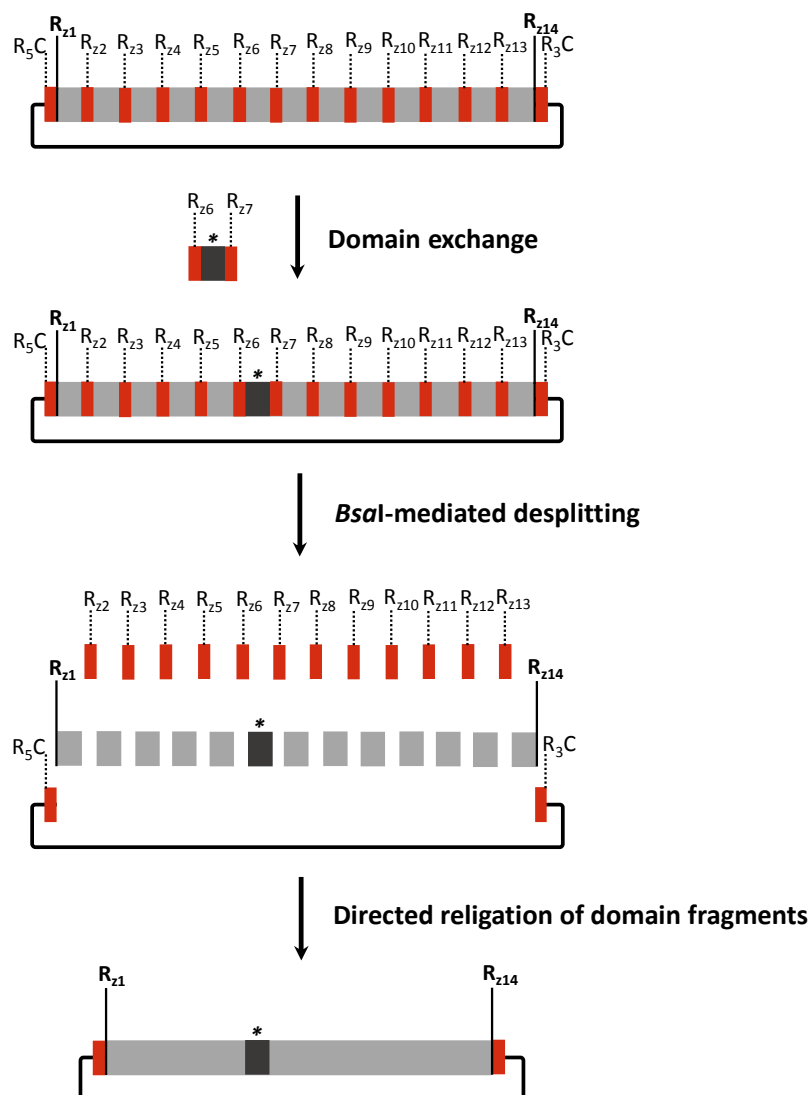


Figure 11. Strategy to generate synthetic *mch* gene constructs harboring mutated PCP domains. Synthetic PCP domain encoding fragments harboring mutations in the core motif (illustrated as dark grey boxes and marked with an asterisk) were cloned into the *mch* gene constructs from the gene library using the unique R-sites within the SEs. ‘Desplitting’ of these constructs using *BsaI* results in splitter-free *mch* gene constructs, which can be cloned into the *mch* expression constructs by conventional cloning using the unique R-sites at the 5’ and 3’ ends (R_{z1}/R_{z14}).

A list of the generated *mch* gene constructs as well as entire expression constructs describing their construction is given in Tables S6 and S7 (Supporting Information). Results from the Sanger sequencing of the PCP domain encoding regions revealed that the original PCP₂ domain fragment harboring the active GGHSL core motif was still integrated in the pSyn1-MchB_A_CP2inact1_SE construct. Furthermore, Illumina sequencing results of the entire expression constructs indicated a 1 bp deletion in constructs pSynMch20 (harboring the MchC_A_CP4inact1 fragment) and pSynMch21 (harboring the MchC_A_CP4inact2 fragment) causing a frameshift mutation and a 1.2 kb domain fragment duplication in the pSynMch22 construct (harboring the MchC_A_CP5inact1 fragment). In the latter case, duplication of the C₅ condensation domain happened most likely during the ‘desplitting’ procedure due to incorrect ligations of non-complementary overhangs. The 1 bp deletion in constructs pSynMch20 and pSynMch22 was most likely acquired during the last assembly step. The affected expression constructs were thus not considered for heterologous expression experiments. After sequence verification, the remaining expression constructs harboring inactive PCP domains in modules 1 (pSynMch17), 3 (pSynMch19) and 6 (pSynMch23) of the synthetic A-type *mch* pathway and the reactivated PCP₄ domain in module 4 (pSynMch24) of the synthetic S-type *mch* cluster were transformed into the heterologous host *M. xanthus* DK1622 $\Delta mchA-tet$ for subsequent heterologous expression of the modified *mch* cluster constructs. Several clones were routinely cultivated and prepared extracts were analyzed for the expected myxochromides by HPLC-MS (data not shown). The respective myxochromides lacking either *N*-Me-*L*-threonine (incorporated by module 1), *L*-leucine (incorporated by module 2) or *L*-glutamine (incorporated by module 6) compared to the original lipohexapaptides myxochromides A were not detected in culture extracts of the mutant strains. In these cases, the addressed modules are located at the N-termini (module 1 in MchB and module 3 in MchC) as well as at the C-termini (module 6 in MchC) of the NRPS subunits, which represent the interfaces between the individual proteins making up the entire assembly lines (MchA/MchB and MchB/MchC) or the termination points of the biosynthesis respectively (MchC). One can assume that the transfer of biosynthetic intermediates across PKS/NRPS subunits (from the PKS (MchA) subunit to module 1 of the first NRPS (MchB) subunit and from module 2 of the first NRPS to module 3 of the second NRPS (MchC)) is more critical than substrate shuttling between ‘internal’ modules, which are part of one distinct subunit. However, mutated PCP domains, which were introduced in the A-type *mch* pathway, only harbor a few mutations exclusively in the core motif, whereas the surrounding interdomain linker regions remained unchanged. Thus, direct domain-domain interactions,

intermodule communication or even interactions between the Mch subunits might not be directly affected when mediated by significant conformational changes. In fact, details on the ‘module-skipping’ mechanism, e.g. from structural studies, is unfortunately not available thus far and the MchC subunit of the S-type *mch* pathway is currently, to the best of our knowledge, the only reported example for a NRPS, for which module-skipping was described.¹⁸ It is very likely that the biosynthetic intermediate in the native S-type *mch* pathway is transferred from the module 3 PCP domain to the PCP domain of module 5, thereby skipping the entire fourth module, as proposed recently.²¹ This would suggest that the PCP₃ domain must directly interact with the C₅ domain, which represents a noncognate domain-domain interaction that needs to be evolutionary arranged. This would have strong implications for the engineered hybrid assembly lines, which is discussed below:

Skipping of module 1 in the synthetic A-type mch pathway

In case the first module in the MchB subunit responsible for the incorporation of *N*-Me-*L*-threonine is skipped, the ACP domain from the PKS subunit must directly interact with the donor site of the C₂ domain from module 2, which represents a noncognate interaction. Sequence analysis of the C domains from *mch* pathways revealed that all orthologue C₂ domains fall into the class of ^LC_L domains (see Chapter 2), which significantly differ from so-called starter C domains,³¹ which initiate NRPS biosynthesis. It can be assumed that the interaction of the ACP domain with the downstream starter C₁ domain is highly specific and the noncognate ACP-C₂ interaction is thus not facilitated in the engineered PKS/NRPS assembly line, which would explain abolishment of myxochromide production. The crucial role of tightly synchronized interactions between an ACP domain and a downstream starter C domain and the high specificity of the C domain for the incoming acyl intermediate from *in trans* acting PKS machineries has been recently described in the biosynthesis of the lipopeptide antibiotic CDA.³⁹ Furthermore, in the myxochromide A lipohexapeptide core, the *N*-Me-*L*-threonine moiety is connected over its side chain hydroxyl group with the carboxyl group of the *L*-glutamine residue to form the cyclic depsipeptide structure.¹⁸ From a (bio)chemical point of view, it is rather unlikely that the TE domain of the termination module responsible for the cyclization reaction is able to cyclize the linear lipopeptide without the *N*-Me-*L*-threonine moiety in a similar manner, which could potentially affect proper termination of the biosynthesis. It is more likely that the linear peptide intermediate is hydrolyzed and thus cleaved off the assembly line (spontaneous hydrolysis or catalyzed by the

TE domain), which may reduce the production yield significantly. Unfortunately, such linear biosynthetic intermediates were not detected in culture extracts.

Skipping of module 3 in the synthetic A-type mch pathway

Similarly, skipping of module 3 in the MchC subunit would result in a noncognate interaction between the PCP₂ domain and the donor site of the downstream C₄ domain. According to the phylogenetic C domain sequence analysis (see Chapter 2),³¹ the C₄ domain was predicted to be a ^LC_L domain, whereas the E domain of module 2 provides a racemic mixture of the biosynthetic intermediate *N*-Me-*L*-Thr-(*D/L*)-Ala. Consequently, the C₄ (^LC_L) domain might choose the *L*-configured intermediate, but is most likely not able to further process non-native dipeptide intermediate. Furthermore, it is not clear how the transfer of biosynthetic building blocks is mechanistically realized across individual NRPS subunits communicating via specific com-domains,⁴⁰ which may additionally be problematic if the first module at the N-terminus is skipped. As already mentioned above, proper termination of the biosynthesis via cyclization of the mature peptide could be impaired as the TE domain is possibly not able to exert its function on the non-native intermediate. This might lead to hydrolysis of the biosynthetic intermediate, but linear peptide precursors were not detected in culture extracts.

Skipping of module 6 in the synthetic A-type mch pathway

The termination module of the A-type *mch* pathway harbors a thioesterase (TE) domain, which catalyzes the cyclization of the linear lipopeptide, thereby releasing the final product from the assembly line. It was shown that extensive domain-domain interactions between PCP and TE domains are crucial for the proper synchronization of this final step of the NRP biosynthesis.^{41–43} Recent structural studies on the surfactin termination module suggested that tremendous conformational rearrangements are necessary to guide the PCP-bound peptide intermediate to the adjacent TE domain due to long distances between these domains.³⁸ In the case of the inactive module 6 in the engineered A-type *mch* pathway, it is assumed that the PCP₅ domain, which is even more spatially separated from the downstream TE₆ domain, is not able to interact with the TE₆ acceptor site making the engineered assembly line unable to produce the expected lipopentapeptide core. If the noncognate PCP₅-TE₆ interaction is not functional in the engineered assembly line, the linear lipopentapeptide would most likely be hydrolyzed from the NRPS. However, linear biosynthetic intermediates were not found in culture extracts. In addition, the important role of native PCP-TE interactions is underpinned by the module deletion event, which was observed in the native C-type *mch* pathway (see

Chapter 2).¹⁷ The A₅-PCP₅-C₆ unit was deleted in the course of pathway diversification rather than the C₅-A₅-PCP₅ module, thereby conserving the cognate PCP₆-TE₆ and native PCP₄-C₅ interaction.

Reactivation of module 4 in the synthetic S-type *mch* pathway

Surprisingly, the installation of a synthetic S-type PCP₄ domain harboring the restored GGHSL core motif in the synthetic S-type *mch* pathway did not lead to the production of a novel lipohexapeptide core. In addition, module 4 is obviously not skipped in this engineered pathway as myxochromide S biosynthesis products were not detected as well. This unexpected result only allows highly speculative conclusions about the functionality of the underlying assembly line due to the lack of mechanistic and structural insights. A vague hypothesis could be that the native S-type *mch* assembly line undergoes significant structural changes or domain rearrangements to allow the lipopeptide intermediate to get transferred from module 3 to module 5, thereby skipping module 4. The Ser → Pro mutation may play an important role in stabilizing this conformation in conjunction with additional mutations in the interdomain and intermodule linkers around module 4. To re-verify the reactivation of this module, new mutant strains should be obtained and analyzed to exclude unexpected phenotypic characteristics, which could also be responsible for this observation.

Overall, induction of module skipping by applying point mutations to PCP domains does not seem to be a tool that can be generally used in NRPS engineering. Several key factors might play a role in the ‘module-skipping’ process observed in myxochromide S biosynthesis,²¹ including e.g. mutational changes in the surrounding linker regions of the PCP₄ domain and/or a specific function of the proline residue in the PCP₄ core motif. As codon usage analysis indicated that the region around modules 3 and 4 (A₃-PCP₃-C₄-A₄-PCP₄) is particularly less adapted (see Chapter 2) and thus a hotspot for mutational activity,¹⁷ it is likely that structural changes of the encoded native S-type *mch* assembly line took place in this region, thereby enabling the megasynthetase to skip an entire module. However, as neither linear biosynthetic intermediates nor complete myxochromide derivatives were detected in extracts of the heterologous mutants, it has to be further analyzed whether ‘module-skipping’ indeed takes place in the engineered *mch* clusters.

3.3.10 Significance

In conclusion, the establishment of synthetic DNA platforms based on a dedicated gene library enables an unprecedented high degree of flexibility towards the engineering of artificial biosynthetic gene clusters. By following the predefined constructional and functional sequence requirements, it is possible to design and exchange any gene cluster fragment of interest ranging from domains, modules and entire subunits to other functional building blocks for specific engineering efforts based on the described innovative DNA assembly strategy including the ‘desplitting’ approach. In conjunction with the detailed sequence analysis of evolutionary related pathways, the established platforms allow for the recombination of PKS/NRPS subunits to generate non-native hybrid pathways on a rational basis for subsequent functional expression and hybrid compound production in a selected heterologous host strain. This approach led to the identification of five novel ‘unnatural’ myxochromide peptide scaffolds, thereby further increasing the structural diversity of this compound class. Moreover, chemical characterization of novel hybrid metabolites can provide deep insights into the biosynthetic mechanisms and allows straightforward experimental verification of biosynthetic hypotheses based on *in silico* predictions. Here, the role of the condensation domain as a gatekeeper for stereospecific processing of biosynthetic intermediates has been successfully illuminated and gave rational explanations for the observed stereochemistry also in the native lipopeptide cores, which could not deduced from *in silico* analyses of the native *mch* pathways. In this study, we demonstrated how synthetic biology approaches essentially contribute to the rational engineering and elucidation of microbial natural product biosynthesis pathways. Our strategy can generally be applied to any desired biosynthetic gene cluster in order to alter product structures or to improve production yields on a broad scale, which might also be of interest for the development of natural product derived drugs with improved pharmacological properties driven by structure-activity relationship studies. It also sets the stage for future initiatives aiming at understanding the general principles of the sequence modulation process for the design of improved complex biosynthetic gene clusters.

3.3.11 Experimental Procedures

3.3.11.1 Sequence Analysis and Design

The *mch* clusters from strains *Myxococcus xanthus* DK1622 (A-type), *Myxococcus* sp. 171 (B-type), *Myxococcus virescens* ST200611 (C-type), *Stigmatella erecta* Pde77 (D-type) and *Stigmatella aurantiaca* DW4/3-1 (S-type) were extensively analyzed *in silico* to meet the demands for the construction of a gene library as well as entire gene cluster constructs and for engineering of artificial *mch* pathways. To ensure the directed assembly of *mch* cluster constructs by conventional restriction/ligation cloning, six restriction enzyme sites were engineered within the coding sequence at the 5' and 3' ends of the biosynthesis genes *mchA*, *mchB* and *mchC* by point mutations (Supporting Information Table S5). A set of 19 restriction enzymes were further selected for the 'desplitting' procedure as well as for pathway engineering to allow for module and domain swaps, and the corresponding recognition sequences were eliminated along the whole gene cluster sequences by silent mutations as well (Supporting Information Table S2). Comparative sequence analysis of the *mch* clusters listed in Table S1 was carried out by using the Geneious alignment tool integrated into Geneious software version 9.1.2,⁴⁴ to identify suitable positions for the SEs. Based on these sequence alignments, 4 bp fusion sites were selected and designed to be unique within all *mch* genes and between each domain fragment to enable the Golden Gate-based generation of a *mch* gene library using the type IIS restriction enzyme *BsaI*. SE sequences harboring the pre-defined recognition sequences of conventional type II enzymes are summarized in Table S3 (Supporting Information).

3.3.11.2 *De Novo* Gene Synthesis of Artificial Gene Cluster Fragments and Synthetic Vectors

The artificial gene cluster fragments needed for the assembly of entire synthetic *mch* pathways were designed as described in chapter 3.3.3 and were generated via DNA synthesis (ATG Biosynthetics GmbH, Merzhausen). The > 30 kb *mch* pathways were split into seven synthetic building blocks. Additionally, biosynthesis genes (6 up to 16 kb) were subdivided into two to four fragments (2.1-5.9 kb). In addition to the *mchA-C* gene fragments, artificial sequences were synthesized, which contain the promoter (native promoter from A-type pathway and synthetic *Tn5* promoter), intergenic linkers (3*mchA*-5*mchB* and 3*mchB*-5*mchC*) as well as terminator (including the *mchD* gene and the downstream helicase gene *rhIE*) sequences. Since the domain organization of the MchA subunits appears to be basically identical in *mch* pathways, only the *mchA* gene of the A-type *mch* cluster was synthesized and

used for the assembly of all artificial *mch* cluster constructs described in this study. Furthermore, the MchB subunits from the A-, B-, C- and D-type assembly lines are identical in terms of domain arrangement and amino acid substrate selection (but different from the S-type MchB). Thus, a synthetic version of the A-type *mchB* gene was also used for the construction of B-, C- and D-type pathways. Due to various recombination events and the resulting differences in the *mchC* genes of the *mch* clusters, synthetic *mchC* fragments from all *mch* pathways (including the corresponding terminator fragments) were designed. The fragments generated by DNA synthesis were additionally flanked by unique R-sites to allow for the subcloning of individual fragments or assembled genes into the cloning vector pSynbio1. The vector backbones pSynbio1 and pSynbio2 were provided as circularized vectors by the gene synthesis company. The synthetic constructs used in this study are listed in Table S4 (Supporting Information). The constructs were delivered in a standard cloning vector (pGH or pUC57) harboring an ampicillin resistance gene. Genetic features of the cloning and expression vectors are illustrated in Figures S1-S2 (Supporting Information).

3.3.11.3 Microbial Strains and Culture Conditions

Escherichia coli strains (HS996 or DH10 β) were routinely grown in LB medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) at 30 or 37°C overnight. For myxochromide production analysis, *Myxococcus xanthus* DK1622 wildtype strain,²¹ the myxochromide A-deficient mutant strain *M. xanthus* DK1622 Δ *mchA-tet* (Wenzel *et al.*, unpublished) and its heterologous expression mutants (harboring artificial *mch* clusters) were routinely grown in 300 mL shaking flasks on a 50 mL scale in CTT medium (casitone 1%, Tris-HCl [pH 8.0] 10 mM, K₂HPO₄/KH₂PO₄ buffer [pH 7.6] 1 mM, MgSO₄ × 7 H₂O 8 mM, pH adjusted to 7.6) at 30°C and 180 rpm for 4-5 days. For the isolation of hybrid myxochromides, fermentations on a 1 L scale were performed under the same conditions in 5L shaking flasks. Cultures were amended with antibiotics if necessary in the following final concentrations: ampicillin 100 µg/mL, kanamycin 50 µg/mL, oxytetracycline 10 µg/mL.

3.3.11.4 DNA Isolation, Processing and Analysis

Enzymes used in this study were purchased from Fermentas. Oligonucleotides used in polymerase chain reactions (PCR) were obtained from Sigma. DNA was extracted and purified from *E. coli* strains by using a standard alkaline lysis protocol,⁴⁵ or by using the GeneJET Plasmid Miniprep Kit from Thermo Fisher Scientific. For standard cloning procedures, plasmid DNA, which was hydrolyzed for subsequent ligation into an acceptor

vector, was separated by agarose gel electrophoresis and cleaved DNA fragments were purified from the agarose gel by using the peqGOLD Gel Extraction Kit from PeqLab. Acceptor DNA was further dephosphorylated using the shrimp alkaline phosphatase from Fermentas. Ligations were carried out using the T4 ligase from Fermentas. Ligations were routinely dialyzed using MF-nitrocellulose membranes (0.025 μm) from Merck Millipore prior to transformation via electroporation in either *E. coli* or *M. xanthus* host strains. ‘Desplitting’ reactions were purified by applying the QIAquick PCR Purification Kit from Qiagen with a cut-off of 100 bp in order to remove the released SEs. PCRs were performed using the PCR machine Mastercycler pro from Eppendorf. All methods were performed according to the manufacturers or to standard protocols.⁴⁵ For details regarding the cloning vectors, constructed plasmids, expression constructs and applied PCR conditions see Supporting Information.

3.3.11.5 Heterologous Expression of Artificial *mch* Pathways in *M. xanthus*

The *mch* expression constructs were transformed into *M. xanthus* DK1622 $\Delta mchA$ -*tet* (Wenzel *et al.*, unpublished) via electroporation using established standard protocols.²⁰ Mutants resistant to kanamycin were selected and further analyzed for correct integration of the *mch* pathways into the former *mchA* gene cluster locus in *M. xanthus* DK1622 $\Delta mchA$ -*tet* via single crossover. To isolate chromosomal DNA, cells from a CTT agar plate were re-suspended in 100 μL water and heated at 100°C for 10 min. 1 μL of this mixture was used as template for PCR. Detailed information on applied PCR conditions and used oligonucleotides is given in the Supporting Information (Figure S4, Tables S10 and S11).

3.3.11.6 Myxochromide Production Analysis

Myxochromide production analysis in the heterologous production strains was carried out on a 50 mL scale under routine cultivation conditions. Cells and Amberlite XAD-16 adsorber resin were harvested by centrifugation at 8,000 rpm and 4 °C for 10 min and subsequently extracted twice with 50 mL of a mixture of methanol and acetone (1:1). The extracts were evaporated to dryness, dissolved in methanol and subjected to HPLC-MS analysis using a Dionex Ultimate 3000 RSLC system coupled to a Bruker maXis 4G TOF mass spectrometer. Separation was performed using a Waters BEH C18, 100 \times 2.1 mm, 1.7 μm d_p column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0-0.5 min 5% B, 0.5-18.5 min 5-95% B,

18.5-20.5 min 95% B. Full scan mass spectra were acquired in positive ESI mode in a range from 150-2500 m/z .

3.3.11.7 Isolation and Structure Elucidation of Engineered Hybrid Myxochromides

Selected representatives of the engineered hybrid myxochromides (AS-type, SA-type, SB-type, SC-type, and SD type) were isolated from culture extracts of the heterologous production strains via semi-preparative reversed-phase HPLC. The planar structures were elucidated based on 1D (^1H) and 2D (^1H - ^1H COSY, HSQC, HMBC, and ROESY) NMR spectroscopy as well as HR-ESI-MS data. The absolute configuration of the amino acid residues was determined by HPLC-MS analysis of the *L*- and *D*-FDLA (1-fluoro-2,4-dinitrophenyl-5-*L*-/*D*-leucinamide) derivatives of the acid hydrolysate of myxochromides in comparison with corresponding derivatives of *L*-configured amino acid standards (Marfey's method).³² The following myxochromide derivatives were purified and structurally characterized:

Myxochromide AS₄ (1c). A total of 4.5 mg of **1c** was isolated from a 18 L culture (18x 1L) of *Myxococcus xanthus* DK1622 $\Delta mchA\text{-}tet::p\text{SynMch8}$. The molecular formula was established to be $\text{C}_{38}\text{H}_{52}\text{N}_6\text{O}_8$ (m/z 721.39373 $[\text{M} + \text{H}]^+$). NMR spectra and selected correlations are illustrated in Supporting Information Figures S5 and S6, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **1c** hydrolysate is shown in Supporting Information Figure S7. For details, see Supporting Information.

Myxochromide SA₃ (2b). A total of 7.2 mg of **2b** was isolated from a 9 L culture (9x 1L) of *Myxococcus xanthus* DK1622 $\Delta mchA\text{-}tet::p\text{SynMch14}$. The molecular formula was established to be $\text{C}_{48}\text{H}_{69}\text{N}_7\text{O}_9$ (m/z 888.52409 $[\text{M} + \text{H}]^+$). NMR spectra and selected correlations are illustrated in Supporting Information Figures S8 and S9, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **2b** hydrolysate is shown in Supporting Information Figure S10. For details, see Supporting Information.

Myxochromide SB₄ (3c). A total of 7.5 mg of **3c** was isolated from a 9 L culture (9x 1L) of *Myxococcus xanthus* DK1622 $\Delta mchA\text{-}tet::p\text{SynMch15}$. The molecular formula was established to be $\text{C}_{54}\text{H}_{80}\text{N}_8\text{O}_{10}$ (m/z 1015.62390 $[\text{M} + \text{H}]^+$). NMR spectra and selected correlations are illustrated in Supporting Information Figures S11 and S12, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **3c** hydrolysate is shown in Supporting Information Figure S13. For details, see Supporting Information.

Myxochromide SC₄ (4c). A total of 0.5 mg of **4c** was isolated from a 18 L culture (18x 1L) of *Myxococcus xanthus* DK1622 $\Delta mchA$ -tet::pSynMch11. The molecular formula was established to be C₄₆H₆₆N₆O₈ (m/z 831.50222 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S14 and S15, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **4c** hydrolysate is shown in Supporting Information Figure S16. For details, see Supporting Information.

Myxochromide SD₃ (5b). A total of 0.7 mg of **5b** was isolated from a 9 L culture (9x 1L) of *Myxococcus xanthus* DK1622 $\Delta mchA$ -tet::pSynMch16. The molecular formula was established to be C₄₄H₆₄N₆O₈ (m/z 791.47051 [M + H]⁺). NMR spectra and selected correlations are illustrated in Supporting Information Figures S17 and S18, and HPLC-MS analysis of *L*- and *D*-FDLA derivatives from the **5b** hydrolysate is shown in Supporting Information Figure S19. For details, see Supporting Information.

3.4 Supporting Information

3.4.1 Constructional Sequence Design of *mch* Clusters

Synthetic DNA platforms for the production of novel myxochromide peptide scaffolds were created based on the native *mch* pathways, which were characterized in detail in Chapter 2.¹⁷ The artificial *mch* cluster sequences were adapted to the highly flexible assembly strategy based on type IIS restriction enzymes developed in this work. Native *mch* clusters, which were subjected to the constructional sequence design process, are shown in Table S1.

Table S1: Overview on myxochromide biosynthetic gene clusters (*mch* clusters) which were subjected to the gene design process in this study.

Producer strain	Strain abbrev.	Cluster	GenBank Accession ^[a]
<i>Myxococcus xanthus</i> DK1622	Mx1	A-type ^[b]	KX622595
<i>Myxococcus</i> sp. 171 ^[c]	M1	B-type ^[d]	KX622591
<i>Myxococcus virescens</i> ST200611	Mv1	C-type ^[e]	KX622594
<i>Stigmatella erecta</i> Pde77	Se1	D-subtype 1 ^[f]	KX622602
<i>Stigmatella aurantiaca</i> DW4/3-1	Sa1	S-type ^[g]	KX622599

[a] Annotated cluster files were additionally deposited in the MiBIG database. [b] Characterized in previous study.¹⁸ [c] Unclassified strain belonging to the genus *Myxococcus*. [d] Characterized in Chapter 2.¹⁷ [e] Characterized in Chapter 2.¹⁷ [f] Characterized in Chapter 2.¹⁷ [g] Characterized in previous study.²¹

To allow for the interchangeability of modules or even single domains, splitter elements (SE) were introduced between each domain of the biosynthesis genes *mchA-C*. The SE are composed of type IIS R-sites (*AarI/BsaI*) at both sites and harbor an additional spacer sequence in-between, which displays the recognition sequence of a type II restriction enzyme suitable for domain and module swaps or for the introduction of additional cluster fragments. The 4 bp overhangs, which are generated by hydrolysis with *AarI/BsaI*, are selected and designed to be unique between each domain fragment along the whole cluster sequence, and can thus be used as unique fusion sites to allow for a directed reassembly of the biosynthesis genes. Selected type II restriction enzymes are listed in Table S2. Sequences of designed SE are shown in Table S3. Locations of the unique fusion sites are summarized in Table S4. Locations of unique R-sites, which were engineered in the coding sequence (CDS) at the 5' and 3' ends of each *mch* gene fragment, are shown in Table S5.

Table S2: Restriction enzyme sites used for pathway assembly and engineering. To allow for the assembly and interchangeability of *mch* cluster parts, the recognition sequences of 6 type II restriction enzymes were introduced into the coding sequence of *mch* genes (highlighted in red) and recognition sequences of another 19 restriction enzymes were eliminated from the CDS sequences.

Restriction enzyme	Recognition sequence	Function
<i>AarI</i>	CACCTGC GTGGACG	RIIS
<i>AflIII</i> (<i>BspTI</i>)	CTTAAG GAATTC	R _{x5} , R _{y6} , R _{z7}
<i>AgeI</i>	ACCGGT TGGCCA	R _{z14}
<i>AseI</i> (<i>VspI</i>)	ATTAAT TAATTA	R _{z9}
<i>AsiSI</i> (<i>SfaAI</i>)	GCGATCGC CGCTAGCG	R _{x3} , R _{y3} , R _{z3}
<i>AvrII</i> (<i>XmaII</i>)	CCTAGG GGATCC	R _{y7} , R _{z11}
<i>BamHI</i>	GGATCC CCTAGG	R _{z10}
<i>BsaI</i>	GGTCTC CCAGAG	RIIS
<i>BsiWI</i>	CGTACG GCATGC	R _{x1}
<i>DraI</i>	TTTAAA AAATTT	Destruction of cloning vector backbone
<i>EcoRI</i>	GAATTC CTTAAG	R _{z4}
<i>HindIII</i>	AAGCTT TTCGAA	R _{z12}
<i>KpnI</i>	GGTACC CCATGG	R _L
<i>MfeI</i> (<i>MunI</i>)	CAATTG GTTAAC	R _{z6}
<i>MluI</i>	ACGCGT TGCGCA	R _{y1} , R _{z15}
<i>MreI</i>	CGCCGGCG GCGGCCGC	R _{x6} , R _{z17}
<i>NdeI</i>	CATATG GTATAC	R _{x4} , R _{y2} , R _{z5}
<i>NheI</i>	GCTAGC CGATCG	R _{z13}
<i>NotI</i>	GCGGCCGC CGCCGGCG	R _{y8} , R _{z16}
<i>PacI</i>	TTAATTAA AATTAATT	Vector backbone modification
<i>PmeI</i> (<i>MssI</i>)	GTTTAAAC CAAATTTG	Vector backbone modification
<i>PvuI</i>	CGATCG GCTAGC	R _R
<i>SpeI</i> (<i>BcuI</i>)	ACTAGT TGATCA	R _{x2} , R _{y5} , R _{z2}
<i>SphI</i>	GCATGC CGTACG	R _{z1}
<i>SwaI</i> (<i>SmiI</i>)	ATTTAAAT TAAATTTA	Vector backbone modification
<i>XbaI</i>	TCTAGA AGATCT	R _{x6} , R _{y4} , R _{z8}

Table S3. Artificial splitter element (SE) sequences of *mch* genes derived from A-, B-, C-, D- and S-type *mch* clusters based on the *BsaI* design. SEs were introduced between each domain fragment of the *mch* genes (except between the KR and ACP domain of *mchA*). For the *mchA* gene, only the PKS CDS sequence from the A-type producer strain (Mx1) was subjected to the sequence design. For the *mchB* gene, only the NRPS CDS sequences from the A-, D- and S-type producer strains (Mx1, Se1, Sa1) were subjected to the sequence design. Unique fusion sites are highlighted in light gray, the *BsaI* recognition sequence in red, and the unique R-sites within the SEs in black. The spacer nucleotide which is skipped by the type IIS restriction enzyme is shown in dark grey.

Linker region	Splitter element sequence (including fusion sites)	Function
<i>mchA</i> (A-type)		
N-Term	NA- GGTCTC CGCAA-KS	<i>BsiWI</i>
KS-AT	KS-CACCT GAGACC ACTAGT GGTCTC CCACC-AT	<i>SpeI</i>
AT-DH	AT-GGCAG GAGACC GCGATCGC GGTCTC TGGCA-DH	<i>AsiSI</i>
DH-ER	DH-ACGGAG GAGACC CATAT GGTCTC GACGG-ER	<i>NdeI</i>
ER-KR/ACP	ER-CGTTC GAGACC CTTAAG GGTCTC TCGTT-KR/ACP	<i>AflIII</i>
C-Term	KR/ACP-ATCGC GAGACC -CA	<i>MreI</i>
<i>mchB</i> (A-, D- and S-type)		
N-Term	NB- GGTCTC CTTCG-C1	<i>MluI</i>
C1-A1	C1-GCGCC GAGACC CATAT GGTCTC CGCGC-A1	<i>NdeI</i>
A1-CP1	A1-CGCGG GAGACC GCGATCGC GGTCTC TCGCG-CP1	<i>AsiSI</i>
CP1-C2	CP1-AGCGA GAGACC TCTAGAG GGTCTC GAGCG-C2	<i>XbaI</i>
C2-A2	C2-CAGCC GAGACC ACTAGT GGTCTC GCAGC-A2	<i>SpeI</i>
A2-CP2	A2-GAAGT GAGACC CTTAAG GGTCTC CGAAG-CP2	<i>AflIII</i>
CP2-E2	CP2-GCAGG GAGACC CTAGG GGTCTC TGCAG-E2	<i>AvrII</i>
C-Term	E2-TCCCC GAGACC -CB	<i>NotI</i>
<i>mchC</i> (A-, D- and S-type)		
N-Term	Nc- GGTCTC TAGCA-C3	<i>SphI</i>
C3-A3	C3-GAGCC GAGACC ACTAGT GGTCTC GGAGC-A3	<i>SpeI</i>
A3-CP3	A3-GGAGC GAGACC GCGATCGC GGTCTC GGGAG-CP3	<i>AsiSI</i>
CP3-C4	CP3-GGACT GAGACC GAATTC GGTCTC CGGAC-C4	<i>EcoRI</i>
C4-A4	C4-ACACC GAGACC CATAT GGTCTC TACAC-A4	<i>NdeI</i>
A4-CP4	A4-CCGCA GAGACC CAATT GGTCTC TCCGC-CP4	<i>MfeI</i>
CP4-C5	CP4-CGAGC GAGACC CTTAAG GGTCTC TCGAG-C5	<i>AflIII</i>
C5-A5	C5-CGGAC GAGACC TCTAGAG GGTCTC TCGGA-A5	<i>XbaI</i>
A5-CP5	A5-TGGAG GAGACC ATTAAT GGTCTC ATGGA-CP5	<i>AseI</i>
CP5-C6	CP5-CGCCA GAGACC GGATCC GGTCTC ACGCC-C6	<i>BamHI</i>
C6-A6	C6-CGACC GAGACC CCTAGG GGTCTC ACGAC-A6	<i>AvrII</i>

A6-CP6	A6-GCGAT GAGACC AAGCTT GGTCTC CGCGA-CP6	<i>HindIII</i>
CP6-TE	CP6-CGCTG GAGACC GCTAGC GGTCTC TCGCT-TE	<i>NheI</i>
C-Term	TE-GCTCC GAGACC -Cc	<i>AgeI</i>
<i>mchC</i> (B-type)		
N-Term	Nc- GGTCTC TAGCA-C3	<i>SphI</i>
C3-A3	C3-GAGCC GAGACC ACTAGT GGTCTC GGAGC-A3	<i>SpeI</i>
A3-CP3	A3-GGAGC GAGACC GCGATCGC GGTCTC GGGAG-CP3	<i>AsiSI</i>
CP3-C4	CP3-GGACT GAGACC GAATTC GGTCTC CGGAC-C4	<i>EcoRI</i>
C4-A4	C4-ACCTC GAGACC ACGCGT GGTCTC GACCT-A4	<i>MluI</i>
A4-CP4	A4-TCGCC GAGACC GCGGCCGC GGTCTC GTCGC-CP4	<i>NotI</i>
CP4-C5	CP4-TCTCC GAGACC CGCCGGCG GGTCTC GTCTC-C5	<i>MreI</i>
C5-A5	C5-ACACC GAGACC CATATG GGTCTC TACAC-A5	<i>NdeI</i>
A5-CP5	A5-CCGCA GAGACC CAATTG GGTCTC TCCGC-CP5	<i>MfeI</i>
CP5-C6	CP5-CGAGC GAGACC CTTAAG GGTCTC TCGAG-C6	<i>AflIII</i>
C6-A6	C6-CGGAC GAGACCT CTAGAG GGTCTC TCGGA-A6	<i>XbaI</i>
A6-CP6	A6-TGGAG GAGACC ATTAAT GGTCTC ATGGA-CP6	<i>AseI</i>
CP6-C7	CP6-CGCCA GAGACC GGATCC GGTCTC ACGCC-C7	<i>BamHI</i>
C7-A7	C7-CGACC GAGACC CCTAGGG GGTCTC ACGAC-A7	<i>AvrII</i>
A7-CP7	A7-GCGAT GAGACC AAGCTT GGTCTC CGCGA-CP7	<i>HindIII</i>
CP7-TE	CP7-CGCTG GAGACC GCTAGC GGTCTC TCGCT-TE	<i>NheI</i>
C-Term	TE-GCTCC GAGACC -Cc	<i>AgeI</i>
<i>mchC</i> (C-type)		
N-Term	Nc- GGTCTC TAGCA-C3	<i>SphI</i>
C3-A3	C3-GAGCC GAGACC ACTAGT GGTCTC GGAGC-A3	<i>SpeI</i>
A3-CP3	A3-GGAGC GAGACC GCGATCGC GGTCTC GGGAG-CP3	<i>AsiSI</i>
CP3-C4	CP3-GGACT GAGACC GAATTC GGTCTC CGGAC-C4	<i>EcoRI</i>
C4-A4	C4-ACACC GAGACC CATATG GGTCTC TACAC-A4	<i>NdeI</i>
A4-CP4	A4-CCGCA GAGACC CAATTG GGTCTC TCCGC-T4	<i>MfeI</i>
CP4-C5	CP4-CGAGC GAGACC CTTAAG GGTCTC TCGAG-C5	<i>AflIII</i>
C5-A6	C5-CGGAC GAGACCT CTAGAG GGTCTC TCGGA-A6	<i>XbaI</i>
A6-CP6	A6-GCGAT GAGACC AAGCTT GGTCTC CGCGA-CP6	<i>HindIII</i>
CP6-TE	CP6-CGCTG GAGACC GCTAGC GGTCTC TCGCT-C6	<i>NheI</i>
C-Term	TE-GCTCC GAGACC -Cc	<i>AgeI</i>

Table S4. Locations of selected/designed unique fusion sites between the domains of *mch* genes derived from A-, B-, C-, D- and S-type *mch* clusters based on the *BsaI* design. Locations are referring to the individual *mch* genes retrieved from the GenBank *mch* cluster entries. Unique fusion sites were designed between each domain fragment of the *mch* genes (except between the KR and CP domain of *mchA*). For the *mchA* gene, only the PKS CDS sequence from the A-type producer strain (Mx1) was subjected to the sequence design. For the *mchB* gene, only the NRPS CDS sequences from the A-, D- and S-type producer strains (Mx1, Se1, Sa1) were subjected to the sequence design. Unique fusion sites are highlighted in gray. Mutated nucleotides/amino acids are shown in italics.

Linker	Fusion site		Location (within gene)	Relevant Codons	Amino acids
<i>mchA</i> (A-type) ^[1]					
N-Term	R ₅ A	GCAA	191 – 194	GGC AAG ACG	G K T
KS-AT	R _{x2}	CACC	1498 – 1501	GAG CAC CGG	E H R
AT-DH	R _{x3}	GGCA	2645 – 2648	TGG CAG CGT	W Q R
DH-ER	R _{x4}	ACGG	4295 – 4298	CTC GAC GGC	L D G
ER- KR/ACP	R _{x5}	CGTT	5186 – 5189	CTG ACG TTC	L T F
C-Term	R ₃ A	ATCG	6361 – 6364	GAA ATC GAG	E I E
<i>mchB</i> (A-type) ^[1]					
N-Term	R ₅ B	TTCG	232 – 235	TTC GGA ATG	F G L
C1-A1	R _{y2}	GCGC	1304 – 1307	CGG GGC GCC	R G A
A1-CP1	R _{y3}	CGCG	4282 – 4285	CGG CGC GCC	R R A
CP1-C2	R _{y4}	AGCG	4604 – 4607	ATG GAG CGC	M E R
C2-A2	R _{y5}	CAGC	5826 – 5829	TTC AGC GCG	F S A
A2-CP2	R _{y6}	GAAG	7437 - 7440	AAG AAG CTT	K K L
CP2-E2	R _{y7}	GCAG	7739 – 7742	GGC AGC GCG	G S A
C-Term	R ₃ B	TCCC	9080 – 9083	GAC TTC CCG	D F P
<i>mchB</i> (D-type) ^[2]					
N-Term	R ₅ B	TTCG	232 – 235	TTC GAG ATA	F E I
C1-A1	R _{y2}	GCGC	1304 – 1307	CGC GGC GCG	R G A
A1-CP1	R _{y3}	CGCG	4312 – 4315	CGG CGC GCC (CGG CGA GCC)	R R A (R R A)
CP1-C2	R _{y4}	AGCG	4634 – 4637	ATG GAG CGG	M E R
C2-A2	R _{y5}	CAGC	5856 – 5859	TAC AGC GCG	Y S A
A2-CP2	R _{y6}	GAAG	7467 – 7470	AAG AAG CTG	K K L
CP2-E2	R _{y7}	GCAG	7769 – 7772	ACG GGC AGT (ACG GGG AGT)	T G S (T G S)
C-Term	R ₃ B	TCCC	9107 – 9110	GAC TTC CCA	D F P
<i>mchB</i> (S-type) ^[3]					
N-Term	R ₅ B	TTCG	232 – 235	TTC GAG ATG	F E M

C1-A1	R _{y2}	GCGC	1304 – 1307	CGC GGC GCG	R G A
A1-CP1	R _{y3}	CGCG	4312 – 4315	CGG CGC GCG	R R A
CP1-C2	R _{y4}	AGCG	4634 – 4637	ATG GAG CGG	M E R
C2-A2	R _{y5}	CAGC	5856 – 5859	TAC AGC GCG	Y S A
A2-CP2	R _{y6}	GAAG	7467 – 7470	AAG AAG CTC	K K L
CP2-E2	R _{y7}	GCAG	7769 – 7772	ACG GGC AGC	T G S
C-Term	R _{3B}	TCCC	9107 – 9110	GAT TTC CCA	D F P
<i>mchC</i> (A-type) ^[1]					
N-Term	R _{5C}	AGCA	50 – 53	CAG CAC GGC	Q H G
C3-A3	R _{z2}	GAGC	1240 – 1243	GCG GAG CGC	A E R
A3-CP3	R _{z3}	GGAG	2853 – 2856	CTG GAG GCC	L E A
CP3-C4	R _{z4}	GGAC	3147 – 3150	GGG GAC ACG (GGA GAC ACG)	G D T (G D T)
C4-A4	R _{z5}	ACAC	4325 – 4328	GAC ACG GAC	D T D
A4-CP4	R _{z6}	CCGC	5960 – 5963	CCC GCC CCC	P A P
CP4-C5	R _{z7}	CGAG	6270 – 6273	CCC GAG GGC	P E G
C5-A5	R _{z8}	CGGA	7469 – 7472	ACG GAC CTC	T D L
A5-CP5	R _{z9}	TGGA	9083 – 9086	GTG GAC AAG	V D K
CP5-C6	R _{z10}	CGCC	9400 – 9403	GTC CGC CTG (GTC CGA CTG)	V R L (V R L)
C6-A6	R _{z11}	CGAC	10605 – 10608	CGC GAC CTC	R D L
A6-CP6	R _{z12}	GCGA	12235 – 12238	AGC GCG ATG	S A M
CP6-TE	R _{z13}	CGCT	12527 – 12530	CCG CTC ACC	P L T
C-Term	R _{3C}	GCTC	13227 – 13230	GCG GCG CTC	A A L
<i>mchC</i> (B-type) ^[4]					
N-Term	R _{5C}	AGCA	50 – 53	CAG CAC GGC	Q H G
C3-A3	R _{z2}	GAGC	1240 – 1243	GCG GAG CGC	A E R
A3-CP3	R _{z3}	GGAG	2853 – 2856	CTG GAG GCC	L E A
CP3-C4	R _{z4}	GGAC	3147 – 3150	GGG GAC ACG (GGA GAC ACG)	G D T (G D T)
C4-A4	R _{z15}	ACCT	4414 – 4417	CAA ACC TCC	Q T S
A4-CP4	R _{z16}	TCGC	5971 – 5974	GAG TCG CAG	E S Q
CP4-C5	R _{z17}	TCTC	6239 – 6242	ATC TCT CGC	I S R
C5-A5	R _{z5}	ACAC	7439 – 7442	GAC ACC GAC	D T D
A5-CP5	R _{z6}	CCGC	9074 – 9077	CCC GCG CCG	P A P
CP5-C6	R _{z7}	CGAG	9381 – 9384	CCC GAG GGC (CCC GAA GGC)	P E G (P E G)
C6-A6	R _{z8}	CGGA	10580 – 10583	ACG GAC CTC	T D L
A6-CP6	R _{z9}	TGGA	12194 – 12197	GTG GAC AAG	V D K

CP6-C7	R _{z10}	CGCC	12511 – 12514	GTC CGC CTG (GTC CGG TTG)	VRL (VRL)
C7-A7	R _{z11}	CGAC	13716 – 13719	GTC GAC CTC	VDL
A7-CP7	R _{z12}	GCGA	15346 – 15349	GGT GCG ATG	GAM
CP7-TE	R _{z13}	CGCT	15638 – 15641	CCG CTG ACC	PLT
C-Term	R _{3C}	GCTC	16335 - 16338	CCG GCG CTC	PAL
<i>mchC</i> (C-type) ^[5]					
N-Term	R _{5C}	AGCA	50 – 53	CAG CAC GGC	QHG
C3-A3	R _{z2}	GAGC	1240 – 1243	GCC GAG CGC	AER
A3-CP3	R _{z3}	GGAG	2853 – 2856	CAG GAG ACG	QET
CP3-C4	R _{z4}	GGAC	3147 – 3150	GAG GAC ACG	EDT
C4-A4	R _{z5}	ACAC	4325 – 4328	GAC ACC GAC	DTD
A4-CP4	R _{z6}	CCGC	5960 – 5963	CCC GCG CCT	PAP
CP4-C5	R _{z7}	CGAG	6270 – 6273	CCC GAG GGC (CCG GAG GGC)	PEG (PEG)
C5-A6	R _{z8}	CGGA	7469 – 7472	ACG GAC CTC	TDL
A6-CP6	R _{z12}	GCGA	9100 – 9103	GGC GCG ATG (GGC GCG CTG)	GAM (GAL)
CP6-TE	R _{z13}	CGCT	9392 – 9395	CCG CTG ACG	PLT
C-Term	R _{3C}	GCTC	10098 - 10101	CCG GAG CTC	PEL
<i>mchC</i> (D-type) ^[2]					
N-Term	R _{5C}	AGCA	50 – 53	CAG CAC GGC	QHG
C3-A3	R _{z2}	GAGC	1240 – 1243	GCC GAG CGG	AER
A3-CP3	R _{z3}	GGAG	2880 – 2883	ACG GAG CAC (ACA GAG CAC)	TEH (TEH)
CP3-C4	R _{z4}	GGAC	3156 – 3159	GGG GAC GGG (GGT GAC GGG)	GDG (GDG)
C4-A4	R _{z5}	ACAC	4334 – 4337	AAC ACC GAC	NTD
A4-CP4	R _{z6}	CCGC	5969 – 5972	CCC GCG CCT	PAP
CP4-C5	R _{z7}	CGAG	6276 – 6279	CCC GAG AGC (CCG GAG AGC)	PES (PES)
C5-A5	R _{z8}	CGGA	7475 – 7478	ACG GAC CTC	TDL
A5-CP5	R _{z9}	TGGA	9089 – 9092	GTG GAC AAG	VDK
CP5-C6	R _{z10}	CGCC	9406 – 9409	GTG CGC CTG (GTG CGG CTG)	VRL (VRL)
C6-A6	R _{z11}	CGAC	10611 – 10614	CTC GAC CTG	LDL
A6-CP6	R _{z12}	GCGA	12241 – 12244	GGG GCG ATG (GGG ACG ATG)	GTM (GAM)
CP6-TE	R _{z13}	CGCT	12533 – 12536	CCG CTC ACG (CCT CTC ACG)	PLT (PLT)

C-Term	R ₃ C	GCTC	13239 – 13242	CCT GAG CTC (CCT GAG CTT)	P E L (P E L)
<i>mchC</i> (S-type) ^[3]					
N-Term	R ₅ C	AGCA	50 – 53	CAG CAC GGC	Q H G
C3-A3	R _{z2}	GAGC	1240 – 1243	GTC GAG CAG	V E Q
A3-CP3	R _{z3}	GGAG	2862 – 2865	AAG GAG TTG	K E L
CP3-C4	R _{z4}	GGAC	3156 – 3159	GGG GAC GTC	G D V
C4-A4	R _{z5}	ACAC	4337 – 4340	AAC ACC GAC	N T D
A4-CP4	R _{z6}	CCGC	5972 – 5975	CCC GCG CCC	P A P
CP4-C5	R _{z7}	CGAG	6279 – 6282	CCC GAG AGC (CCG GAG AGC)	P E S (P E S)
C5-A5	R _{z8}	CGGA	7478 – 7481	ACG GAC CTC	T D L
A5-CP5	R _{z9}	TGGA	9092 – 9095	GTG GAC AAG	V D K
CP5-C6	R _{z10}	CGCC	9409 – 9412	GTG CGC CTG (GTG CGG CTG)	V R L (V R L)
C6-A6	R _{z11}	CGAC	10614 – 10617	CTC GAC CTG	L D L
A6-CP6	R _{z12}	GCGA	12244 – 12247	GCC GCG ATG	A A M
CP6-TE	R _{z13}	CGCT	12536 – 12539	CCG CTC ACG	P L T
C-Term	R ₃ C	GCTC	13242 – 13245	CCA GAG CTC	P E L

[1] Location of fusion sites based on *mch* cluster retrieved from GeneBank entry KX622595. [2] Location of fusion sites based on *mch* cluster retrieved from GeneBank entry KX622602. [3] Location of fusion sites based on *mch* cluster retrieved from GeneBank entry KX622599. [4] Location of fusion sites based on *mch* cluster retrieved from GeneBank entry KX622591. [5] Location of fusion sites based on *mch* cluster retrieved from GeneBank entry KX622594.

Table S5. Locations of the unique R-sites at the 5' and 3' ends of the *mch* biosynthesis gene fragments. R-sites engineering was performed within the CDS sequences of the *mch* gene fragments to allow for conventional stitching of the artificial *mch* clusters by applying point mutations. Locations are referring to the individual *mch* genes retrieved from the GenBank *mch* cluster entries. Unique R-sites (R_{x1}, R_{x6}, R_{y1}, R_{y8}, R_{z1}, R_{z14}) are highlighted in gray. Mutated nucleotides/amino acids are shown in italics.

Terminus	Gene	R-sites	Location (within gene)	Relevant Codons	Amino acids
A-type <i>mch</i> cluster^[1]					
N-Term	<i>mchA</i>	R _{x1}	197 – 202	ACG TAC GTG	T Y V
C-Term	<i>mchA</i>	R _{x6}	6326 – 6333	GCG CCG GCG (GCG CCC GCG)	A P A (A P A)
N-Term	<i>mchB</i>	R _{y1}	254 – 259	CAC GCG TCC	H A S
C-Term	<i>mchB</i>	R _{y8}	9052 – 9059	GCG GCC GCG (GCC GCG GCG)	A A A (A A A)
N-Term	<i>mchC</i>	R _{z1}	56 – 61	GGC ATG CTC	G M L
C-Term	<i>mchC</i>	R _{z14}	13202 – 13207	GAC CGG TGT (GAC CGC AGT)	D R C (D R S)

B-type <i>mch</i> cluster ^[2]					
N-Term	<i>mchC</i>	R _{z1}	56 – 61	GGC ATG CTC	G M L
C-Term	<i>mchC</i>	R _{z14}	16310 – 16315	GAC CGG TGC (GAC CGT GGC)	D R C (D R G)
C-type <i>mch</i> cluster ^[3]					
N-Term	<i>mchC</i>	R _{z1}	56 – 61	GGC ATG CTC	G M L
C-Term	<i>mchC</i>	R _{z14}	10073 – 10078	GAC CGG TGC (GAC CGG AGC)	D R C (D R S)
D-type <i>mch</i> cluster ^[4]					
N-Term	<i>mchB</i>	R _{y1}	254 – 259	CAC GCG TGC (CAC CCG AGC)	H A C (H P S)
C-Term	<i>mchB</i>	R _{y8}	9079 – 9086	GCG GCC GCG (GCC GCG ACG)	A A A (A A T)
N-Term	<i>mchC</i>	R _{z1}	56 – 61	GGC ATG CTC	G M L
C-Term	<i>mchC</i>	R _{z14}	13214 – 13219	GAC CGG TAT (GAC AGG GAT)	D R Y (D R D)
S-type <i>mch</i> cluster ^[5]					
N-Term	<i>mchB</i>	R _{y1}	254 – 259	CAC GCG TGC (CAC CCG AGC)	H A C (H P S)
C-Term	<i>mchB</i>	R _{y8}	9079 – 9086	GCG GCC GCA (GCA GCC GCA)	A A A (A A A)
N-Term	<i>mchC</i>	R _{z1}	56 – 61	GGC ATG CTG	G M L
C-Term	<i>mchC</i>	R _{z14}	13217 – 13222	GAC CGG TAC (GAC AGG GAC)	D R Y (D R D)

[1] Location of R-sites based on *mch* cluster retrieved from GeneBank entry KX622595. [2] Location of R-sites based on *mch* cluster retrieved from GeneBank entry KX622591. [3] Location of R-sites based on *mch* cluster retrieved from GeneBank entry KX622594. [4] Location of R-sites based on *mch* cluster retrieved from GeneBank entry KX622602. [5] Location of R-sites based on *mch* cluster retrieved from GeneBank entry KX622599.

3.4.2 Design of the Cloning Vector pSynbio1 and the Expression Vector pSynbio2

For the assembly of large biosynthesis gene fragments and for ‘desplitting’ processes, the cloning vector pSynbio1, which is a derivative of the pGH vector, was designed and manufactured by DNA synthesis. The high-copy vector backbone pSynbio1 is composed of a minimal set of genetic elements needed for the amplification and selection in *E. coli*. These include the *oriV* origin of vegetative replication from the broad-range RK2 plasmid,⁴⁶ the *trfA* gene, whose gene product binds to and activates *oriV*, which leads to unidirectional replication,^{47–49} and an ampicillin resistance gene (*ampR*). To meet the constructional requirements, the recognition sequences needed for cloning and assembly of the expression

constructs (Table S2) were calculated out of the vector sequence. Recognition sequences of the type II enzymes *KpnI* and *PmeI* were introduced into the vector backbone to allow for the subcloning of synthetic fragments. The genetic map of pSynbio1 is illustrated in Figure S1.

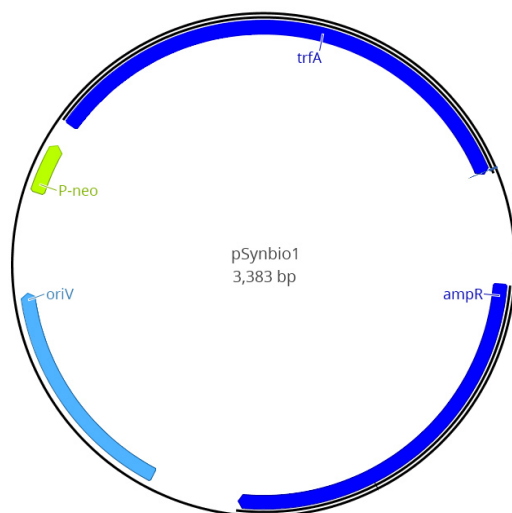


Figure S1. Genetic map of the cloning vector pSynbio1.

The expression vector pSynbio2 was designed and manufactured by DNA synthesis for the functional expression of artificial *mch* clusters in the heterologous host *M. xanthus* DK1622 (see Section ‘Functional Sequence Design for Heterologous Expression’). The genetic map of the pSynbio2 vector showing the unique R-sites for modification of the vector backbone is illustrated in Figure S2. Information on the different genetic element are given in Chapter 3.3.2.

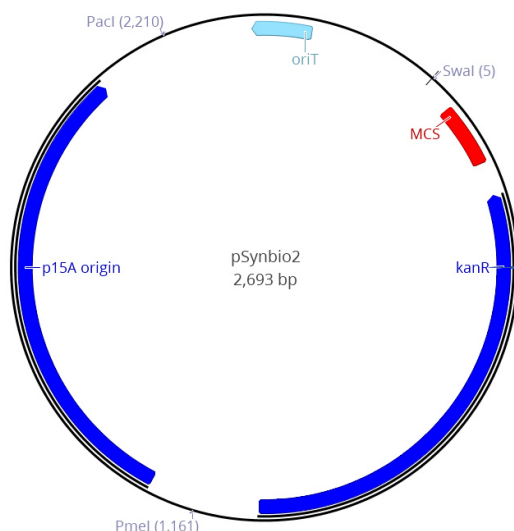


Figure S2. Genetic map of the expression vector pSynbio2.

3.4.3 Generation of *mch* Cluster Fragments via DNA Synthesis

Artificial *mch* cluster fragments, which were designed and synthesized in this study, are shown in Table S6. Cluster fragments were delivered either in the pGH vector (pUC derivative) or in the pUC57 vector backbone harboring an ampicillin resistance gene.

Table S6. Artificial DNA fragments from *mch* gene clusters generated via DNA synthesis.

Construct Name	Description (Location in GenBank file)	Size [bp]	Flanking R-sites
A-type <i>mch</i> cluster fragments from <i>M. xanthus</i> DK1622^[1]			
pGH-P-5mchA_Ab	Promotor fragment (6607-7208 nt)	3539	<i>KpnI</i> - <i>BsiWI</i> / <i>PmeI</i>
pGH-3mchA-5mchB_Ab	Linker fragment (13332-13720 nt)	3326	<i>KpnI</i> / <i>MreI</i> - <i>MluI</i> / <i>PmeI</i>
pGH-3mchB-5mchC_Ab	Linker fragment (22513-22701nt)	3126	<i>KpnI</i> / <i>NotI</i> - <i>SphI</i> / <i>PmeI</i>
pGH-T-3mchC_Ab	Terminator fragment (35867-38440 nt)	5547	<i>KpnI</i> / <i>AgeI</i> - <i>PvuI</i> / <i>PmeI</i>
pGH-MchA_A_fragA	<i>mchA</i> gene fragment (7197-11354 nt)	7116	<i>KpnI</i> / <i>BsiWI</i> - <i>NdeI</i> / <i>PmeI</i>
pGH-MchA_A_fragA_dcm	<i>mchA</i> gene fragment (7197-8507 nt)	4260	<i>KpnI</i> / <i>BsiWI</i> - <i>SpeI</i>
pGH-MchA_A_fragB	<i>mchA</i> gene fragment (11301-13394 nt)	5063	<i>KpnI</i> / <i>NdeI</i> - <i>MreI</i> / <i>PmeI</i>
pGH-MchB_A_fragA	<i>mchB</i> gene fragment (13693-19364 nt)	8628	<i>KpnI</i> / <i>MluI</i> - <i>SpeI</i> / <i>PmeI</i>
pGH-MchB_A_fragA_dcm	<i>mchB</i> gene fragment (13693-14768 nt)	4026	<i>KpnI</i> / <i>MluI</i> - <i>NdeI</i>
pGH-MchB_A_fragB	<i>mchB</i> gene fragment (19287-22592 nt)	6275	<i>KpnI</i> / <i>SpeI</i> - <i>NotI</i> / <i>PmeI</i>
pGH-MchC_A_fragA	<i>mchC</i> gene fragment (22690-27042 nt)	7295	<i>KpnI</i> / <i>SphI</i> - <i>NdeI</i> / <i>PmeI</i>
pGH-MchC_A_fragA_dcm	<i>mchC</i> gene fragment (22690-23883 nt)	4138	<i>KpnI</i> / <i>SphI</i> - <i>SpeI</i>
pGH-MchC_A_fragB	<i>mchC</i> gene fragment (26965-31798 nt)	7797	<i>KpnI</i> / <i>NdeI</i> - <i>AseI</i> / <i>PmeI</i>
pGH-MchC_A_fragC	<i>mchC</i> gene fragment (31723-35966 nt)	7207	<i>KpnI</i> / <i>AseI</i> - <i>AgeI</i> / <i>PmeI</i>
pGH-MchC_A_fragC_dcm	<i>mchC</i> gene fragment (31723-32043 nt)	3264	<i>AseI</i> - <i>BamHI</i>
pUC57-CP1_A_inact1	Carrier protein fragment (17743-18068 nt)	3078	<i>KpnI</i> / <i>AsiSI</i> - <i>XbaI</i> / <i>PmeI</i>
pUC57-CP2_A_inact1	Carrier protein fragment (20898-21203 nt)	3056	<i>KpnI</i> / <i>AflII</i> - <i>AvrII</i> / <i>PmeI</i>
pUC57-CP3_A_inact1	Carrier protein fragment (25494-25787 nt)	3050	<i>KpnI</i> / <i>AsiSI</i> - <i>EcoRI</i> / <i>PmeI</i>
pGH-CP4_A_inact1	Carrier protein fragment (28601-28914 nt)	3079	<i>KpnI</i> / <i>MfeI</i> - <i>AflII</i> / <i>PmeI</i>

pGH-CP4_A_inact2	Carrier protein fragment (28601-28914 nt)	3079	<i>KpnI/MfeI-AflII/PmeI</i>
pUC57-CP5_A_inact1	Carrier protein fragment (31724-32042 nt)	3071	<i>KpnI/AseI-BamHI/PmeI</i>
pUC57-CP6_A_inact1	Carrier protein fragment (34876-35171 nt)	3046	<i>KpnI/HindIII-NheI/PmeI</i>
B-type <i>mch</i> cluster fragments from <i>Myxococcus</i> sp. 171^[4]			
pGH-3mchB-5mchC_Bb	Linker fragment (22380-22568 nt)	3126	<i>KpnI/NotI-SphI/PmeI</i>
pGH-T-3mchC_Bb	Terminator fragment (38817-39474 nt) ^[7]	5547	<i>KpnI/AgeI-PvuI/PmeI</i>
pGH-MchC_B_fragA	<i>mchC</i> gene fragment (22557-26924 nt)	7406	<i>KpnI/SphI-MluI/PmeI</i>
pGH-MchC_B_fragB	<i>mchC</i> gene fragment (26921-31584 nt)	7709	<i>KpnI/MluI-MfeI/PmeI</i>
pGH-MchC_B_fragC	<i>mchC</i> gene fragment (31581-34704 nt)	6141	<i>KpnI/MfeI-AseI/PmeI</i>
pGH-MchC_B_fragD	<i>mchC</i> gene fragment (34701-38845 nt)	7210	<i>KpnI/AseI-AgeI/PmeI</i>
pGH-MchC_B_fragA_woSE	<i>mchC</i> gene fragment (22557-26924 nt)	7332	<i>KpnI/SphI-MluI/PmeI</i>
pGH-MchC_B_fragB_woSE	<i>mchC</i> gene fragment (26921-31584 nt)	7633	<i>KpnI/MluI-MfeI/PmeI</i>
pGH-MchC_B_fragC_woSE	<i>mchC</i> gene fragment (31581-34704 nt)	6093	<i>KpnI/MfeI-AseI/PmeI</i>
pGH-MchC_B_fragD_woSE	<i>mchC</i> gene fragment (34701-38845 nt)	7114	<i>KpnI/AseI-AgeI/PmeI</i>
C-type <i>mch</i> cluster fragments from <i>M. virescens</i> ST200611^[5]			
pGH-3mchB-5mchC_Cb	Linker fragment (23091-23279 nt)	3126	<i>KpnI/NotI-SphI/PmeI</i>
pGH-T-3mchC_Cb	Terminator fragment (33291-33948 nt) ^[7]	5525	<i>KpnI/AgeI-PvuI/PmeI</i>
pGH-MchC_C_fragA	<i>mchC</i> gene fragment (23268-27546 nt)	7311	<i>KpnI/SphI-NdeI/PmeI</i>
pGH-MchC_C_fragA_dcm	<i>mchC</i> gene fragment (23268-24461 nt)	4142	<i>KpnI/SphI-SpeI</i>
pGH-MchC_C_fragB	<i>mchC</i> gene fragment (27543-33319 nt)	8860	<i>KpnI/NdeI-AgeI/PmeI</i>
D-type <i>mch</i> cluster fragments from <i>S. erecta</i> Pde77^[2]			
pGH-3mchB-5mchC_Db	Linker fragment (21027-21217 nt)	3568	<i>KpnI/NotI-SphI/PmeI</i>
pGH-T-3mchC_Db	Terminator fragment (34370-35036 nt) ^[7]	5550	<i>KpnI/AgeI-PvuI/PmeI</i>
pGH-MchB_D_fragA	<i>mchB</i> gene fragment (12179-17807 nt)	8660	<i>KpnI/MluI-SpeI/PmeI</i>
pGH-MchB_D_fragB	<i>mchB</i> gene fragment (17804-21058 nt)	6226	<i>KpnI/SpeI-NotI/PmeI</i>
pGH-MchC_D_fragA	<i>mchC</i> gene fragment (21206-25493 nt)	7319	<i>KpnI/SphI-NdeI/PmeI</i>
pGH-MchC_D_fragB	<i>mchC</i> gene fragment	7793	<i>KpnI/NdeI-VspI/PmeI</i>

	(25490-30248 nt)		
pGH-MchC_D_fragC	<i>mchC</i> gene fragment (30245-34398 nt)	7213	<i>KpnI/VspI-AgeI/PmeI</i>
S-type <i>mch</i> cluster fragments from <i>S. aurantiaca</i> DW4/3-1^[3]			
pGH-3mchB-5mchC_Sb	Linker fragment (22680-22869 nt)	3127	<i>KpnI/NotI-SphI/PmeI</i>
pGH-T-3mchC_Sb	Terminator fragment (36025-36701 nt) ^[7]	5544	<i>KpnI/AgeI-PvuI/PmeI</i>
pGH-MchB_S_fragA	<i>mchB</i> gene fragment (13833-19460 nt)	8650	<i>KpnI/MluI-SpeI/PmeI</i>
pGH-MchB_S_fragA_dcm	<i>mchB</i> gene fragment (13833-14908 nt)	4020	<i>KpnI-NdeI</i>
pGH-MchB_S_fragB	<i>mchB</i> gene fragment (19457-22711 nt)	6272	<i>KpnI/SpeI/NotI/PmeI</i>
pGH-MchC_S_fragA	<i>mchC</i> gene fragment (22858-27148 nt)	7329	<i>KpnI/SphI/NdeI/PmeI</i>
pGH-MchC_S_fragA_dcm	<i>mchC</i> gene fragment (22858-24051 nt)	4138	<i>KpnI/SpeI</i>
pGH-MchC_S_fragB	<i>mchC</i> gene fragment (27145-31903 nt)	7800	<i>KpnI/NdeI/VspI/PmeI</i>
pGH-MchC_S_fragC	<i>mchC</i> gene fragment (31900-36053 nt)	7211	<i>KpnI/VspI/AgeI/PmeI</i>
pGH-MchC_S_fragC_dcm	<i>mchC</i> gene fragment (31900-32220 nt)	3264	<i>AseI/BamHI</i>
pGH-CP4_S_react	Carrier protein fragment (28780-29090 nt)	3076	<i>KpnI/MfeI-AflII/PmeI</i>
A-type <i>mch</i> cluster fragments from <i>M. xanthus</i> DK1622^[6]			
pGH-P-5mchA_A_AarI	Promotor fragment (6603-7206 nt)	4003	<i>KpnI-BsiWI/PmeI</i>
pGH-3mchA-5mchB_A_AarI	Linker fragment (13335-13718 nt)	3783	<i>KpnI/MreI-MluI/PmeI</i>
pGH-3mchA-5mchB_A_AarI	Linker fragment (22517-22699 nt)	3582	<i>KpnI/NotI-SphI/PmeI</i>
pGH-T-3mchC_A_AarI	Terminator fragment (35842-38443 nt)	6001	<i>KpnI/AgeI-PvuI/PmeI</i>
pGH-MchA_A_AarI_SE	<i>mchA</i> gene fragment (7203-13468 nt)	9665	<i>KpnI/BsiWI-MreI/PmeI</i>
pGH-MchB_A_AarI_SE	<i>mchB</i> gene fragment (13715-22746 nt)	12431	<i>KpnI/MluI-NotI/PmeI</i>
pGH-MchC_A_AarI_fragABCE	<i>mchC</i> gene fragment (22696-26974 nt, ABC) (+ 30110-32063 nt, E)	9404	<i>KpnI/SphI-BamHI/MluI</i>
pGH-MchC_A_AarI_fragD	<i>mchC</i> gene fragment (26971-30113 nt)	6186	<i>NdeI-XbaI</i>
pGH-MchC_A_AarI_fragF	<i>mchC</i> gene fragment (32060-35845 nt)	6863	<i>BamHI-MluI</i>

Cloning vectors, expression vectors, and genetic elements			
pSynbio1_AarI	Cloning vector	3383	-
pSynbio2_AarI	Expression vector	2700	-
pGH-Amp_Synbio1mut	<i>ampR</i> gene fragment	3587	<i>PstI</i> - <i>MscI</i>
pGH-MCS_Synbio2	Multiple cloning site	3415	<i>SwaI</i> - <i>HindIII</i>
pUC57-PTn5-5mchA_Ab	<i>Tn5</i> promoter fragment	3086	<i>KpnI</i> - <i>BsiWI</i> / <i>PmeI</i>

[1] *BsaI* design based on *mch* cluster retrieved from GeneBank entry KX622595. [2] *BsaI* design based on *mch* cluster retrieved from GeneBank entry KX622602. [3] *BsaI* design based on *mch* cluster retrieved from GeneBank entry KX622599. [4] *BsaI* design based on *mch* cluster retrieved from GeneBank entry KX622591. [5] *BsaI* design based on *mch* cluster retrieved from GeneBank entry KX622594. [6] *AarI* design based on *mch* cluster retrieved from GeneBank entry KX622595. [7] Location is referring to the 3' *mchC*-*mchD* region (without terminator-*rhIE* region).

3.4.4 Construction of Modified Cloning and Expression Vectors

To adapt the original cloning (pSynbio1_AarI) and expression (pSynbio2_AarI) vectors to the *BsaI* design, modified synthetic DNA fragments were designed, in which either a *BsaI* R-site was eliminated by applying a point mutation (Amp_Synbio1mut) or the MCS was equipped with *BsaI* sites. The synthetic fragments were cloned into the original vector backbones and revealed pSynbio1 and pSynbio2 (Table S7).

Table S7: Cloning and expression vectors generated in this study.

Plasmid	Construction
pSynbio1	666 bp <i>PstI</i> / <i>MscI</i> Amp_Synbio1mut fragment from pGH-Amp_Synbio1mut ligated into pSynbio1_AarI hydrolyzed with <i>PstI</i> / <i>MscI</i> to generate pSynbio1
pSynbio2	495 bp <i>SwaI</i> / <i>HindIII</i> MCS_Synbio2 fragment from pGH-MCS_Synbio2 ligated into pSynbio2_AarI hydrolyzed with <i>SwaI</i> / <i>HindIII</i> to generate pSynbio2

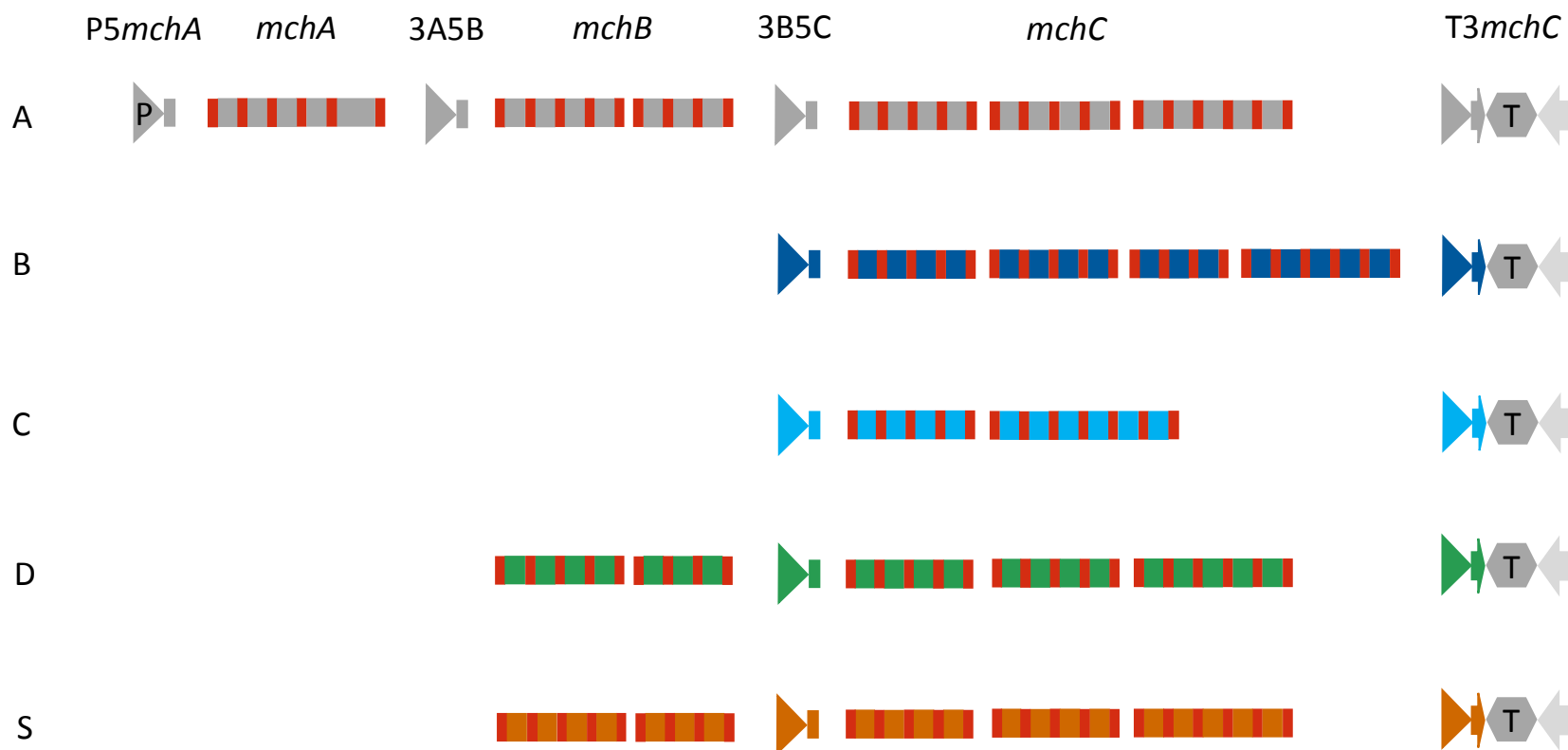


Figure S3. Design of *mch* cluster fragments from the different *mch* pathways. To generate artificial hybrid *mch* pathways for the production of naturally occurring and entirely new myxochromide cores, a minimal set of *mch* cluster fragments were designed and generated via DNA synthesis. The promoter fragment from the A-type *mch* cluster (P5*mchA_A*) as well as the A-type *mchA_A* gene and *mchA_A/mchB_A* intergenic linker fragment (3A_A5B_A) were used for the construction of all hybrid *mch* pathways described in this study. Large biosynthetic genes were subdivided into two (*mchB_A*, *mchB_S*, *mchB_D*, *mchC_C*), three (*mchC_A*, *mchC_D*, *mchC_S*) or even four (*mchC_B*) individual gene fragments.

3.4.5 Assembly of Artificial *mch* Biosynthesis Genes and Generation of a Gene Library

The large artificial biosynthesis genes *mchA-C* from the different *mch* clusters were assembled using traditional restriction/ligation techniques (Figure 4).⁴⁵ To release the SEs between the domain fragments, the gene constructs were hydrolyzed with *BsaI*. Directed religation of the single domain fragments resulted in fully desplitted *mch* gene constructs. The plasmids, which were constructed in this study to form functional *mch* genes, are shown in Table S8.

Table S8: Constructs for *mch* gene library generated in this study.

Plasmid	Construction
A-type <i>mch</i> gene constructs^[2]	
pSyn1-MchA_A_SE	<p>(1) 2122 bp <i>NdeI/PmeI</i> MchA_A_fragB fragment from pGH-MchA_A_fragB ligated into pGH-MchA_A_fragA hydrolyzed with <i>NdeI/PmeI</i> to generate pGH-MchA_A_fragAB</p> <p>(2) 6302 bp <i>KpnI/PmeI</i> MchA_A_fragAB fragment from pGH-MchA_A_fragAB ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchA_A_SE_pre</p> <p>(3) 1336 bp <i>KpnI/SpeI</i> MchA_A_fragA_dcm fragment from pGH-MchA_A_fragA_dcm ligated into pSyn1-MchA_A_SE_pre to generate pSyn1-MchA_A_SE</p>
pSyn1-MchA_A	Hydrolysis of pSyn1-MchA_A_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchB_A_SE	<p>(1) 3335 bp <i>SpeI/PmeI</i> MchB_A_fragB fragment from pGH-MchB_A_fragB ligated into pGH-MchB_A_fragA hydrolyzed with <i>SpeI/PmeI</i> to generate pGH-MchB_A_fragAB</p> <p>(2) 9028 bp <i>KpnI/PmeI</i> MchB_A_fragAB fragment from pGH-MchB_A_fragAB ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchB_A_SE_pre</p> <p>(3) 1100 bp <i>KpnI/NdeI</i> MchB_A_fragA_dcm fragment from pGH-MchB_A_fragA_dcm ligated into pSyn1-MchB_A_SE_pre to generate pSyn1-MchB_A_SE</p>
pSyn1-MchB_A	Hydrolysis of pSyn1-MchB_A_SE by <i>BsaI</i> followed by

	religation to remove SE ^[1]
pSyn1-MchC_A_SE	<p>(1) 4862 bp <i>NdeI/PmeI</i> MchC_A_fragB fragment from pGH-MchC_A_fragB ligated into pGH-MchC_A_fragA hydrolyzed with <i>NdeI/PmeI</i> to generate pGH-MchC_A_fragAB</p> <p>(2) 9237 bp <i>KpnI/PmeI</i> MchC_A_fragAB fragment from pGH-MchC_A_fragAB ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_A_fragAB</p> <p>(3) 4272 bp <i>AseI/PmeI</i> MchC_A_fragC fragment from pGH-MchC_A_fragC ligated into pSyn1-MchC_A_fragAB digested with <i>AseI/PmeI</i> to generate pSyn1-MchC_A_SE_pre</p> <p>(4) 1219 bp <i>KpnI/SpeI</i> MchC_A_fragA_dcm fragment from pGH-MchC_A_fragA_dcm ligated into pSyn1-MchC_A_SE_pre to generate pSyn1-MchC_A_SE_pre2</p> <p>(5) 344 bp <i>AseI/BamHI</i> MchC_A_fragC_dcm fragment from pGH-MchC_A_fragC_dcm ligated into pSyn1-MchC_A_SE_pre2 to generate pSyn1-MchC_A_SE</p>
pSyn1-MchC_A	Hydrolysis of pSyn1-MchC_A_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchB_A_CP1inact1_SE	350 bp <i>AsiSI/XbaI</i> MchB_A_CP1inact1 fragment from pUC57-MchB_A_CP1inact1 ligated into pSyn1-MchB_A_SE to generate pSyn1-MchB_A_CP1inact1_SE
pSyn1-MchB_A_CP1inact1	Hydrolysis of pSyn1-MchB_A_CP1inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchB_A_CP2inact1_SE	330 bp <i>AflIII/AvrII</i> MchB_A_CP2inact1 fragment from pUC57-MchB_A_CP2inact1 ligated into pSyn1-MchB_A_SE to generate pSyn1-MchB_A_CP2inact1_SE
pSyn1-MchB_A_CP2inact1	Hydrolysis of pSyn1-MchB_A_CP2inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchC_A_CP3inact1_SE	322 bp <i>AsiSI/EcoRI</i> MchC_A_CP3inact1 fragment from pUC57-MchC_A_CP3inact1 ligated into pSyn1-MchC_A_SE to generate pSyn1-MchC_A_CP3inact1_SE
pSyn1-MchC_A_CP3inact1	Hydrolysis of pSyn1-MchC_A_CP3inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]

pSyn1-MchC_A_CP4inact1_SE	338 bp <i>MfeI/AflIII</i> MchC_A_CP4inact1 fragment from pUC57-MchC_A_CP4inact1 ligated into pSyn1-MchC_A_SE to generate pSyn1-MchC_A_CP4inact1_SE
pSyn1-MchC_A_CP4inact1	Hydrolysis of pSyn1-MchC_A_CP4inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchC_A_CP4inact2_SE	338 bp <i>MfeI/AflIII</i> MchC_A_CP4inact2 fragment from pUC57-MchC_A_CP4inact2 ligated into pSyn1-MchC_A_SE to generate pSyn1-MchC_A_CP4inact2_SE
pSyn1-MchC_A_CP4inact2	Hydrolysis of pSyn1-MchC_A_CP4inact2_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchC_A_CP5inact1_SE	344 bp <i>AseI/BamHI</i> MchC_A_CP5inact1 fragment from pUC57-MchC_A_CP5inact1 ligated into pSyn1-MchC_A_SE to generate pSyn1-MchC_A_CP5inact1_SE
pSyn1-MchC_A_CP5inact1	Hydrolysis of pSyn1-MchC_A_CP5inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchC_A_CP6inact1_SE	320 bp <i>HindIII/NheI</i> MchC_A_CP6inact1 fragment from pUC57-MchC_A_CP6inact1 ligated into pSyn1-MchC_A_SE to generate pSyn1-MchC_A_CP6inact1_SE
pSyn1-MchC_A_CP6inact1	Hydrolysis of pSyn1-MchC_A_CP6inact1_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
B-type <i>mch</i> gene constructs	
pSyn1-MchC_B_SE	<p>(1) 2954 bp <i>KpnI/MluI</i> MchC_B_fragA fragment from pGH-MchC_B_fragA_woSE and 2947 bp <i>MluI/PmeI</i> MchC_B_fragB fragment from pGH-MchC_B_fragB_woSE ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_B_fragAB</p> <p>(2) 4854 bp <i>KpnI/AseI</i> MchC_B_fragC fragment from pGH-MchC_B_fragC_woSE and 4269 <i>AseI/PmeI</i> MchC_B_fragD fragment from pGH-MchC_B_fragD_woSE ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_B_fragCD</p> <p>(3) 7318 bp <i>MunI/PmeI</i> MchC_B_fragA fragment from pSyn1-MchC_B_fragCD ligated into pSyn1-MchC_B_fragAB hydrolyzed with <i>MunI/PmeI</i> to generate pSyn1-</p>

	<p>MchC_B_SE_pre</p> <p>(4) 3193 bp <i>MunI/AseI</i> MchC_B_fragC fragment from pGH-MchC_B_fragC ligated into pSyn1-MchC_B_SE_pre to generate pSyn1-MchC_B_SE</p>
pSyn1-MchC_B	Hydrolysis of pSyn1-MchC_B_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
C-type <i>mch</i> gene constructs	
pSyn1-MchC_C_SE	<p>(1) 5923 bp <i>NdeI/PmeI</i> MchC_C_fragB fragment from pGH-MchC_C_fragB ligated into pGH-MchC_C_fragA hydrolyzed with <i>NdeI/PmeI</i> to generate pGH-MchC_C_SE</p> <p>(2) 10296 bp <i>KpnI/PmeI</i> MchC_C_SE fragment from pGH-MchC_C_SE ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_C_SE_pre</p> <p>(3) 1211 bp <i>KpnI/SpeI</i> MchC_C_fragA_dcm fragment from pGH-MchC_C_fragA_dcm ligated into pSyn1-MchC_C_SE_pre hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_C_SE</p>
pSyn1-MchC_C	Hydrolysis of pSyn1-MchC_C_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
D-type <i>mch</i> gene constructs	
pSyn1-MchB_D_SE	5719 bp <i>KpnI/SpeI</i> MchB_D_fragA fragment from pGH-pMchB_D_fragA and 3332 bp <i>SpeI/PmeI</i> MchB_D_fragB fragment from pGH-pMchB_D_fragB ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchB_D_SE
pSyn1-MchB_D	Hydrolysis of pSyn1-MchB_D_SE by <i>BsaI</i> followed by religation to remove SE ^[1]
pSyn1-MchC_D_SE	<p>(1) 4386 bp <i>KpnI/NdeI</i> MchC_D_fragA fragment from pGH-MchC_D_fragA and 4859 bp <i>NdeI/PmeI</i> MchC_D_fragB fragment from pGH-MchC_D_fragB ligated into pSynbio1 hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-MchC_D_fragAB</p> <p>(2) 4278 bp <i>AseI/PmeI</i> MchC_D_fragC fragment from pGH-pMchC_D_fragC ligated into pSyn1-MchB_D_fragAB hydrolyzed with <i>AseI/PmeI</i> to generate pSyn1-MchC_D_SE</p>

pSyn1-MchC_D	Hydrolysis of pSyn1-MchC_D_SE by <i>Bsa</i> I followed by religation to remove SE ^[1]
S-type <i>mch</i> gene constructs	
pSyn1-MchB_S_SE	<p>(1) 3328 bp <i>Spe</i>I/<i>Pme</i>I MchB_S_fragB fragment from pGH-MchB_S_fragB ligated into pGH-pMchB_S_fragA hydrolyzed with <i>Spe</i>I/<i>Pme</i>I to generate pGH-MchB_S_fragAB</p> <p>(2) 9051 bp <i>Kpn</i>I/<i>Pme</i>I MchB_S_fragAB fragment from pGH-MchB_S_fragAB ligated into pSynbio1 hydrolyzed with <i>Kpn</i>I/<i>Pme</i>I to generate pSyn1-MchB_S_SE_pre</p> <p>(3) 1100 bp <i>Kpn</i>I/<i>Nde</i>I MchB_S_fragA_dcm fragment from pGH-MchB_S_fragA_dcm ligated into pSyn1-MchB_S_SE_pre to generate pSyn1-MchB_S_SE</p>
pSyn1-MchB_S	Hydrolysis of pSyn1-MchB_S_SE by <i>Bsa</i> I followed by religation to remove SE ^[1]
pSyn1-MchC_S_SE	<p>(1) 4857 bp <i>Nde</i>I/<i>Pme</i>I MchC_S_fragB fragment from pGH-MchC_S_fragB ligated into pGH-MchC_S_fragA hydrolyzed with <i>Nde</i>I/<i>Pme</i>I to generate pGH-MchC_S_fragAB</p> <p>(2) 9234 bp <i>Kpn</i>I/<i>Pme</i>I MchC_S_fragAB fragment from pGH-MchC_S_fragAB ligated into pSynbio1 hydrolyzed with <i>Kpn</i>I/<i>Pme</i>I to generate pSyn1-MchC_S_fragAB</p> <p>(3) 4276 bp <i>Ase</i>I/<i>Pme</i>I MchC_S_fragC fragment from pGH-MchC_S_fragC ligated into pSyn1-MchC_S_fragAB hydrolyzed with <i>Ase</i>I/<i>Pme</i>I to generate pSyn1-MchC_S_SE_pre</p> <p>(4) 1219 bp <i>Kpn</i>I/<i>Spe</i>I MchC_S_fragA_dcm fragment from pGH-MchC_S_fragA_dcm ligated into pSyn1-MchC_S_SE_pre to generate pSyn1-MchC_S_SE_pre2</p> <p>(5) 344 bp <i>Ase</i>I/<i>Bam</i>HI MchC_S_fragC_dcm fragment from pGH-MchC_S_fragC_dcm ligated into pSyn1-MchC_S_SE_pre2 to generate pSyn1-MchC_S_SE</p>
pSyn1-MchC_S	Hydrolysis of pSyn1-MchC_S_SE by <i>Bsa</i> I followed by religation to remove SE ^[1]
pSyn1-	335 bp <i>Mfe</i> I/ <i>Afl</i> III MchC_S_CP4react fragment from pUC57-

MchC_S_CP4react_SE	MchC_S_CP4act ligated into pSyn1-MchC_S_SE to generate pSyn1-MchC_S_CP4react_SE
pSyn1-MchC_S_CP4react	Hydrolysis of pSyn1-MchC_S_CP4react_SE by <i>Bsa</i> I followed by religation to remove SE ^[1]
A-type <i>mch</i> gene constructs^[3]	
pSyn1-MchA_A_AarI_SE	6303 bp <i>Kpn</i> I/ <i>Pme</i> I MchA_A_AarI_SE fragment from pGH-MchA_A_AarI_SE ligated into pSynbio1_AarI hydrolyzed with <i>Kpn</i> I/ <i>Pme</i> I to generate pSyn1-MchA_A_AarI_SE
pSyn1-MchA_A_AarI	Hydrolysis of pSyn1-MchA_A_AarI_SE by <i>Aar</i> I followed by religation to remove SE ^[1]
pSyn1-MchB_A_AarI_SE	9069 bp <i>Kpn</i> I/ <i>Pme</i> I MchB_A_AarI_SE fragment from pGH-MchB_A_AarI_SE ligated into pSynbio1_AarI hydrolyzed with <i>Kpn</i> I/ <i>Pme</i> I to generate pSyn1-MchB_A_AarI_SE
pSyn1-MchB_A_AarI	Hydrolysis of pSyn1-MchB_A_AarI_SE by <i>Aar</i> I followed by religation to remove SE ^[1]
pSyn1-MchC_A_AarI_SE	<p>(1) 3944 bp <i>Bam</i>HI/<i>Mlu</i>I MchC_A_AarI_fragF fragment from pGH-MchC_A_AarI_fragF ligated into pGH-MchC_A_AarI_fragABCE hydrolyzed with <i>Bam</i>HI/<i>Mlu</i>I to generate pGH-MchC_A_AarI_fragABCEF</p> <p>(2) 3256 bp <i>Nde</i>I/<i>Xba</i>I MchC_A_AarI_fragD fragment from pGH-MchC_A_AarI_fragD ligated into pGH-MchC_A_AarI_fragABCEF hydrolyzed with <i>Nde</i>I/<i>Xba</i>I to generate pGH-MchC_A_AarI_SE</p> <p>(3) 13637 bp <i>Kpn</i>I/<i>Pme</i>I MchC_A_AarI_SE fragment from pGH-MchC_A_AarI_SE ligated into pSynbio1_AarI hydrolyzed with <i>Kpn</i>I/<i>Pme</i>I to generate pSyn1-MchC_A_AarI_SE</p>
pSyn1-MchC_A_AarI	Hydrolysis of pSyn1-MchC_A_AarI_SE by <i>Aar</i> I followed by religation to remove SE ^[1]
pSyn1-P-5mchA_A_AarI	641 bp <i>Kpn</i> I/ <i>Pme</i> I P-5mchA_A_AarI fragment from pGH-P-5mchA_A_AarI ligated into pSynbio1_AarI hydrolyzed with <i>Kpn</i> I/ <i>Pme</i> I to generate pSyn1-P-5mchA_A_AarI
pSyn1-3mchA-5mchB_A_AarI	421 bp <i>Kpn</i> I/ <i>Pme</i> I 3mchA-5mchB_A_AarI fragment from pGH-3mchA-5mchB_A_AarI ligated into pSynbio1_AarI

	hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-3mchA-5mchB_A_AarI
pSyn1-3mchB-5mchC_A_AarI	220 bp <i>KpnI/PmeI</i> 3mchB-5mchC_A_AarI fragment from pGH-3mchB-5mchC_A_AarI ligated into pSynbio1_AarI hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-3mchB-5mchC_A_AarI
pSyn1-T-3mchC_A_AarI	2639 bp <i>KpnI/PmeI</i> T-3mchC_A_AarI fragment from pGH-T-3mchC_A_AarI ligated into pSynbio1_AarI hydrolyzed with <i>KpnI/PmeI</i> to generate pSyn1-T-3mchC_A_AarI

[1] SE = splitter elements. [2] Based on *BsaI* design. [3] Based on *AarI* design.

3.4.6 *In Vitro* Reconstitution of Artificial *mch* Clusters

Artificial *mch* cluster constructs were constructed using traditional restriction/ligation cloning techniques (Figure 4).⁴⁵ Constructed *mch* clusters encompass hybrid gene clusters for the production of novel myxochromides as well as artificial A-type *mch* clusters harboring inactivated PCP domains in modules 1-6. In addition, an artificial S-type *mch* cluster was constructed, which harbors a reactivated PCP domain in module 4. Construction of the mentioned *mch* clusters is described in Table S9.

Table S9: Constructs for *mch* cluster assemblies generated in this study.

Plasmid	Construction
pSynMch1	(1) 621 bp <i>KpnI/BsiWI</i> P5mchA _A fragment from pSyn1-P-5mchA_A_AarI ligated into pSynbio2 hydrolyzed with <i>KpnI/BsiWI</i> to generate pSyn2-ca1
	(2) 6134 bp <i>BsiWI/MreI</i> mchA _A fragment from pSyn1-MchA_A_AarI ligated into pSyn2-ca1 hydrolyzed with <i>BsiWI/MreI</i> to generate pSyn2-ca2
	(3) 386 bp <i>MreI/MluI</i> 3A _A 5B _A fragment from pSyn1-3mchA-5mchB_A_AarI ligated into pSyn2-ca2 hydrolyzed with <i>MreI/MluI</i> to generate pSyn2-ca3
	(4) 8803 bp <i>MluI/NotI</i> mchB _A fragment from pSyn1-MchB_A_AarI ligated into pSyn2-ca3 hydrolyzed with <i>MluI/NotI</i> to generate pSyn2-ca4
	(5) 186 bp <i>NotI/SphI</i> 3B _A 5C _A fragment from pSyn1-3mchB-5mchC_A_AarI ligated into pSyn2-ca4 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca5

	<p>(6) 13150 bp <i>SphI/AgeI mchC_A</i> fragment from pSyn1-MchC_A_AarI ligated into pSyn2-ca5 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca6</p> <p>(7) 2616 bp <i>AgeI/PvuI T3mchC_A</i> fragment from pSyn1-T-3mchC_A_AarI ligated into pSyn2-ca6 hydrolyzed with <i>AgeI/PvuI</i> to generate pSynMch1</p>
pSynMch2	<p>(1) 606 bp <i>KpnI/BsiWI P5mchA_A</i> fragment from pGH-P-5mchA_Ab ligated into pSynbio2 hydrolyzed with <i>KpnI/BsiWI</i> to generate pSyn2-ca7</p> <p>(2) 6134 bp <i>BsiWI/MreI mchA_A</i> fragment from pSyn1-MchA_A ligated into pSyn2-ca7 hydrolyzed with <i>BsiWI/MreI</i> to generate pSyn2-ca8</p> <p>(3) 386 bp <i>MreI/MluI 3A_A5B_A</i> fragment from pGH-3mchA-5mchB_Ab ligated into pSyn2-ca8 hydrolyzed with <i>MreI/MluI</i> to generate pSyn2-ca9</p> <p>(4) 8803 bp <i>MluI/NotI mchB_A</i> fragment from pSyn1-MchB_A ligated into pSyn2-ca9 hydrolyzed with <i>MluI/NotI</i> to generate pSyn2-ca10</p> <p>(5) 186 bp <i>NotI/SphI 3B_A5C_A</i> fragment from pGH-3mchB-5mchC_Ab ligated into pSyn2-ca10 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca11</p> <p>(6) 13150 bp <i>SphI/AgeI mchC_A</i> fragment from pSyn1-MchC_A ligated into pSyn2-ca11 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca12</p> <p>(7) 2601 bp <i>AgeI/PvuI T3mchC_A</i> fragment from pGH-T-3mchC_Ab ligated into pSyn2-ca12 hydrolyzed with <i>AgeI/PvuI</i> to generate pSynMch2_pre</p> <p>(8) 15917 bp <i>KpnI/NotI P5mchA_A-mchA_A-3A_A5B_A-mchB_A</i> fragment from pSyn2-ca10 ligated into pSynMch2_pre hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch2</p>
pSynMch3	<p>(1) 186 bp <i>NotI/SphI 3B_B5C_B</i> fragment from pGH-3mchB-5mchC_Bb ligated into pSynbio2 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca13</p> <p>(2) 2601 bp <i>AgeI/PvuI T3mchC_B</i> fragment from pGH-T-3mchC_Bb ligated into pSyn2-ca13 hydrolyzed with <i>AgeI/PvuI</i> to generate pSyn2-ca14</p> <p>(3) 16250 bp <i>SphI/AgeI mchC_B</i> fragment from pSyn1-MchC_B ligated into pSyn2-ca14 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca15</p> <p>(4) 15909 bp <i>KpnI/NotI P5mchA_A-mchA_A-3A_A5B_A-mchB_B</i> fragment from pSynMch2 ligated into pSyn2-ca15 hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch3</p>
pSynMch4	<p>(1) 186 bp <i>NotI/SphI 3B_C5C_C</i> fragment from pGH-3mchB-5mchC_Cb</p>

	<p>ligated into pSyn2-ca10 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca16</p> <p>(2) 10013 bp <i>SphI/AgeI</i> <i>mchC_C</i> fragment from pSyn1-MchC_C ligated into pSyn2-ca16 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca17</p> <p>(3) 2601 bp <i>AgeI/PvuI</i> T3<i>mchC_C</i> fragment from pGH-T-3mchC_Cb ligated into pSyn2-ca17 hydrolyzed with <i>SphI/AgeI</i> to generate pSynMch4</p>
pSynMch5	<p>(1) 205 bp <i>NotI/SphI</i> 3B_D5C_D fragment from pGH-3mchB-5mchC_Db ligated into pSynbio2 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca18</p> <p>(2) 2610 bp <i>AgeI/PvuI</i> T3<i>mchC_D</i> fragment from pGH-T-3mchC_Db ligated into pSyn2-ca18 hydrolyzed with <i>AgeI/PvuI</i> to generate pSyn2-ca19</p> <p>(3) 13154 bp <i>SphI/AgeI</i> <i>mchC_D</i> fragment from pSyn1-MchC_D ligated into pSyn2-ca19 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca20</p> <p>(4) 8826 bp <i>MluI/NotI</i> <i>mchB_D</i> fragment from pSyn1-MchB_D ligated into pSyn2-ca20 hydrolyzed with <i>MluI/NotI</i> to generate pSyn2-ca21</p> <p>(5) 7110 bp <i>KpnI/MluI</i> P5<i>mchA_A-mchA_A-3A_A5B_A</i> from pSynMch2 ligated into pSyn2-ca21 hydrolyzed with <i>KpnI/MluI</i> to generate pSynMch5</p>
pSynMch6	<p>(1) 8830 bp <i>MluI/NotI</i> <i>mchB_S</i> fragment from pSyn1-MchB_S ligated into pSyn2-ca9 hydrolyzed with <i>MluI/NotI</i> to generate pSyn2-ca22</p> <p>(2) 187 bp <i>NotI/SphI</i> 3B_S5C_S fragment from pGH-3mchB-5mchC_Sb ligated into pSyn2-ca22 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca23</p> <p>(3) 13157 bp <i>SphI/AgeI</i> <i>mchC_S</i> fragment from pSyn1-MchC_S ligated into pSyn2-ca23 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca24</p> <p>(4) 2620 bp <i>AgeI/PvuI</i> T3<i>mchC_S</i> fragment from pGH-T-3mchC_Sb ligated into pSyn2-ca24 hydrolyzed with <i>SphI/AgeI</i> to generate pSynMch6</p>
pSynMch8	<p>(1) 13165 bp <i>SphI/AgeI</i> <i>mchC_S</i> fragment from pSyn1-MchC_S ligated into pSyn2-ca11 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca31</p> <p>(2) 2620 bp <i>AgeI/PvuI</i> T3<i>mchC_S</i> fragment from pGH-T-3mchC_Sb ligated into pSyn2-ca31 hydrolyzed with <i>AgeI/PvuI</i> to generate pSynMch8_pre</p> <p>(3) 15917 bp <i>KpnI/NotI</i> P5<i>mchA_A-mchA_A-3A_A5B_A-mchB_A</i> fragment from pSyn2-ca10 ligated into pSynMch8_pre hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch8</p>
pSynMch9	<p>(1) 186 bp <i>NotI/SphI</i> 3B_A5C_A fragment from pGH-3mchB-5mchC_Ab into pSyn2-ca22 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca32</p>

	<p>(2) 13150 bp <i>SphI/AgeI mchC_A</i> fragment from pSyn1-MchC_A ligated into pSyn2-ca32 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca33</p> <p>(3) 2608 bp <i>AgeI/PvuI T3mchC_A</i> fragment from pGH-T-3mchC_Ab ligated into pSyn2-ca33 hydrolyzed with <i>AgeI/PvuI</i> to generate pSynMch9_pre</p> <p>(4) 15944 bp <i>KpnI/NotI P5mchA_A-mchA_A-3A_A5B_A-mchB_S</i> fragment from pSyn2-ca22 ligated into pSynMch9_pre hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch9</p>
pSynMch10	<p>(1) 186 bp <i>NotI/SphI 3B_C5C_C</i> fragment from pGH-3mchB-5mchC_Cb into pSyn2-ca22 hydrolyzed with <i>NotI/SphI</i> to generate pSyn2-ca34</p> <p>(2) 10021 bp <i>SphI/AgeI mchC_C</i> fragment from pSyn1-MchC_C ligated into pSyn2-ca34 hydrolyzed with <i>SphI/AgeI</i> to generate pSyn2-ca35</p> <p>(3) 2601 bp <i>AgeI/PvuI T3mchC_C</i> fragment from pGH-T-3mchC_Cb ligated into pSyn2-ca35 hydrolyzed with <i>AgeI/PvuI</i> to generate pSynMch10_pre</p> <p>(4) 15944 bp <i>KpnI/NotI P5mchA_A-mchA_A-3A_A5B_A-mchB_S</i> fragment from pSyn2-ca22 ligated into pSynMch10_pre hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch10</p>
pSynMch11	19037 bp <i>NotI/PvuI 3B_B5C_B-mchC_B-T3mchC_B</i> fragment from pSyn2-ca15 ligated into pSynMch10 hydrolyzed with <i>NotI/PvuI</i> to generate pSynMch11
pSynMch12	15952 bp <i>NotI/PvuI 3B_D5C_D-mchC_D-T3mchC_D</i> fragment from pSyn2-ca20 ligated into pSynMch10 hydrolyzed with <i>NotI/PvuI</i> to generate pSynMch12
pSynMch13	<p>(1) The native promoter of A-type <i>mch</i> cluster on pSynMch2 was replaced with <i>cmR-ccdB</i> cassette by Red/ET recombination to generate pSynMch2-cmccdB</p> <p>(2) The <i>cmR-ccdB</i> cassette on pSynMch2-cmccdB was substituted with overexpression promoter <i>Ptn5</i> by Red/ET recombination</p>
pSynMch14	8830 bp <i>MluI/NotI mchB_S</i> fragment from pSyn1-MchB_S ligated into pSynMch13 hydrolyzed with <i>NotI/PvuI</i> to generate pSynMch14
pSynMch15	21643 bp <i>KpnI/NotI 3B_B5C_B-mchC_B-T3mchC_B</i> fragment from pSyn2-ca15 ligated into pSynMch14 hydrolyzed with <i>KpnI/NotI</i> to generate pSynMch15
pSynMch16	18558 bp <i>KpnI/NotI 3B_D5C_D-mchC_D-T3mchC_D</i> fragment from pSyn2-ca20 ligated into pSynMch14 hydrolyzed with <i>KpnI/NotI</i> to generate

	pSynMch16
pSynMch17	8803 bp <i>MluI/NotI</i> <i>mchB_A</i> fragment from pSyn1-MchB_A_CP1inact1 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch17
pSynMch19	13150 bp <i>SphI/AgeI</i> <i>mchC_A</i> fragment from pSyn1-MchC_A_CP3inact1 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch19
pSynMch20	13150 bp <i>SphI/AgeI</i> <i>mchC_A</i> fragment from pSyn1-MchC_A_CP4inact1 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch20
pSynMch21	13150 bp <i>SphI/AgeI</i> <i>mchC_A</i> fragment from pSyn1-MchC_A_CP4inact2 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch21
pSynMch22	13150 bp <i>SphI/AgeI</i> <i>mchC_A</i> fragment from pSyn1-MchC_A_CP5inact1 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch22
pSynMch23	13150 bp <i>SphI/AgeI</i> <i>mchC_A</i> fragment from pSyn1-MchC_A_CP6inact1 ligated into pSynMch13 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch23
pSynMch24	13165 bp <i>SphI/AgeI</i> <i>mchC_S</i> fragment from pSyn1-MchC_S_CP4react1 ligated into pSynMch6 hydrolyzed with <i>MluI/NotI</i> to generate pSynMch24

3.4.7 Transfer and Heterologous Expression of Artificial *mch* Clusters in *Myxococcus xanthus*

To heterologously express the artificial *mch* gene clusters in the myxochromide-deficient mutant DK1622 $\Delta mchA-tet$ (Wenzel *et al.*, unpublished), the expression constructs (pSynMch1-6, pSynMch8-12) were transformed by electroporation into the heterologous host strain using established standard procedures. Integration of the artificial *mch* clusters into the former *mchA* locus in the host chromosome was achieved via homologous recombination (single crossover) using the helicase gene *rhIE* as homologous region. The resulting genotype of the mutant strains as well as the genotype of the heterologous host are exemplified for the myxochromide AS production strain *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch8$ and illustrated in Figure S4. Oligonucleotides used for genotypic verification of the resulting mutant strains and PCR conditions are listed in Tables S10 and S11.

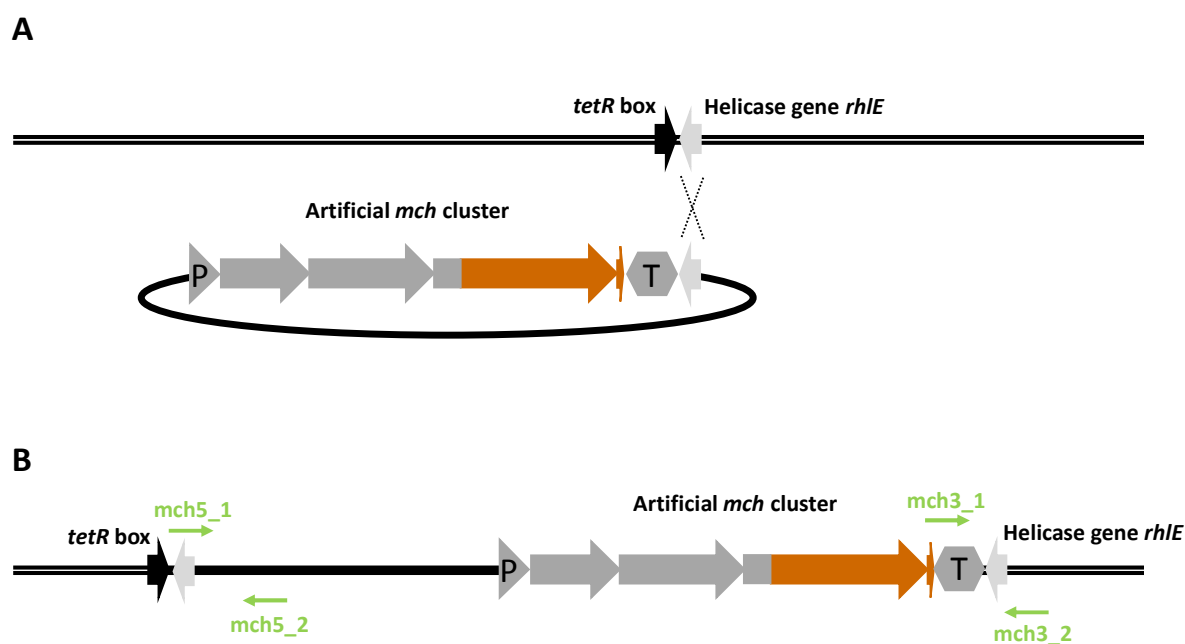


Figure S4. Specific integration of artificial *mch* clusters into the former myxochromide A gene cluster locus of *M. xanthus* DK1622 $\Delta mchA-tet$. **(A)** Genotype representation of the myxochromide A-deficient mutant *M. xanthus* DK1622 $\Delta mchA-tet$. Integration of *mch* cluster constructs was achieved via single crossover using the helicase gene *rhIE*, which is part of the synthetic terminator fragments, as homologous region. **(B)** Genotype representation of mutant strains harboring artificial *mch* clusters. Mutant strains were verified by PCR; binding sites of oligonucleotides for amplification of specific diagnostic markers are shown in green.

Table S10: Oligonucleotides used for PCR amplification of specific markers to verify integration of artificial *mch* clusters into the host chromosome of *M. xanthus* DK1622 $\Delta mchA$ -tet.

Oligonucleotide	Sequence
mch5_1	CGGAGAACTGTGAATGCGC
mch5_2	GTTTCATTTGATGCTCGATG
mch3_1	CGCCGGACGCATGACTCAC
mch3_2	AGAGGCACTCCAGGCCTCTTA

Oligonucleotide pairs	Amplicon size	Present in heterologous host	Present in production strain
mch5_1 / mch5_2	1572 bp	-	+
mch3_1 / mch3_2	1563 bp	-	+

Table S11: Protocol for polymerase chain reaction used to amplify specific markers to verify integration of artificial *mch* clusters into the host chromosome of *M. xanthus* DK1622 $\Delta mchA$ -tet.

PCR step	Temperature [°C]	Time [min]	Cycles
Initialization	95	5	
Denaturation	95	0.5	30
Annealing	58	0.5	
Extension	72	1.5	
Final extension	72	10	

3.4.8 Structure Elucidation of Novel Hybrid Myxochromides

3.4.8.1 Cultivation of Heterologous Production Strains and Isolation of Myxochromides

The heterologous producer strains *M. xanthus* DK1622 $\Delta mchA-tet::pSynMch8$ (Myxochromides AS) and DK1622 $\Delta mchA-tet::pSynMch11$ (Myxochromides SC) were cultivated in 18 L (18x 1 L) and the producer strains DK1622 $\Delta mchA-tet::pSynMch14$ (Myxochromides SA), DK1622 $\Delta mchA-tet::pSynMch15$ (Myxochromides SB) and DK1622 $\Delta mchA-tet::pSynMch16$ (Myxochromides SD) were cultivated in 9L (10x 1 L) CTT medium (casitone 1%, Tris-HCl [pH 8.0] 10 mM, K₂HPO₄/KH₂PO₄ buffer [pH 7.6] 1 mM, MgSO₄ × 7 H₂O 8 mM, pH adjusted to 7.6) including 2% XAD-16 resin for 5-6 days at 30 °C and 180 rpm.

For the isolation of myxochromides AS, SA and SC, cells and XAD-16 Amberlite adsorber resin were harvested by centrifugation at 10,500 rpm and 4 °C for 15 min and were five times extracted with 1 L of a mixture of methanol and acetone (1:1). The organic solvents were removed under reduced pressure and the residues were five times extracted with 200 mL of ethyl acetate. After removal of the solvent, the crude extracts were dissolved in up to 10 mL of methanol for subsequent separation via reverse phase HPLC. A Dionex UltiMate 3000 system equipped with a Luna 5u C18(2) 100A column (250 × 10 mm, Phenomenex) was used. At constant flow rate (5.0 mL/min), the following multi-step gradient was applied for isolation of myxochromides AS₄ and SC₄ (A: deionized water, B: acetonitrile): 0-5 min 10-45% B, 5-30 min 45-65% B, 30-40 min 65-80% B, 40-41 min 80-10% B, 41-47 min 10% B. Myxochromide SC₄ was further purified by applying the following modified gradient (A: deionized water, B: acetonitrile): 0-5 min 10% B, 5-50 min 10-95% B, 50-55 min 95% B, 55-56 min 95-10% B, 56-60 min 10%B. For separation of myxochromide SA, the following modified gradient was applied (A: deionized water, B: acetonitrile): 0-4 min 5% B, 4-8 min 5-65%, 8-41 min 65-95% B, 41-43 min 95% B, 43-45 min 95-5% B, 45-51 min 5% B).

For the isolation of myxochromides SB and SD, cells and XAD adsorber resin were placed in a glass column over glass wool and a sand layer. Myxochromides were extracted by pouring 600 mL *n*-hexane, 900 mL dichloromethane, 600 mL ethyl acetate, 600 mL acetone and 600 mL methanol through the packed column. The fractions were concentrated and analyzed for target myxochromides via HPLC-MS. Separation was performed on a Dionex UltiMate 3000 system using a Waters BEH C18 (100 × 2.1 mm, 1.7 μm) column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0-0.5 min 5% B, 0.5-18.5 min 5-95% B, 18.5-20.5 min 95% B. Full scan mass spectra were acquired in positive ESI mode in a range from 200-2000 *m/z*.

After removal of the solvent, myxochromides were dissolved in 3 mL of methanol for further separation via reverse phase HPLC. Myxochromides SB were purified on a Dionex Ultimate 3000 system equipped with an EclipseC8 column (250 x 10 mm, 4 μ m) at constant flowrate (5 mL/min) by applying the following gradient (A: deionized water, B: acetonitrile): 0-2 min 5 % B, 2-10 min 5-65 % B, 10-30 min 65-70 % B, 30-31 min 70-95% B, 31-34 min 95 % B, 34-35 min 95-5 % B, 35-38 min 5 % B. Myxochromides SD were purified on a Dionex Ultimate 3000 system equipped with a Jupiter column (250 x 10 mm, 4 μ m) at constant flowrate (5 mL/min) by applying the following modified gradient (A: deionized water, B: acetonitrile): 0-2 min 5 % B, 2-10 min 5-66 % B, 10-30 min 66-68 % B, 30-31 min 68-95% B, 31-34 min 95 % B, 34-35 min 95-5 % B, 35-38 min 5 % B.

UV traces were recorded by a diode array detector (DAD) with specified wave lengths (210, 300 and 410 nm) with myxochromides showing good UV absorption at 410 nm. Retention times (R_t) and yields of the isolated compounds are shown in Table S12.

Table S12: Retention times and total amounts of hybrid myxochromides isolated in this study.

Mutant strain	Isolated compound	R_t [min]	Yield [mg]
DK1622 $\Delta mchA-tet::pSynMch8$	Myxochromide AS ₄	27.6	4.5
DK1622 $\Delta mchA-tet::pSynMch14$	Myxochromide SA ₃	18.9	7.2
DK1622 $\Delta mchA-tet::pSynMch15$	Myxochromide SB ₄	23.0	7.5
DK1622 $\Delta mchA-tet::pSynMch11$	Myxochromide SC ₄	39.2	0.5
DK1622 $\Delta mchA-tet::pSynMch16$	Myxochromide SD ₃	19.6	0.7

3.4.8.2 Structure Elucidation of Hybrid Myxochromides

Structure elucidation of myxochromide AS₄, myxochromide SA₃, myxochromide SB₄, myxochromide SC₄ and myxochromide SD₃ was achieved using 1D and 2D NMR spectroscopy as well as HR-MS data. NMR spectra were acquired in CD₃OD at a Bruker Ascend 700 or 500 MHz spectrometer equipped with a 5 mm TXI cryoprobe. 1D ¹H and 2D ¹H–¹H COSY, HSQC, HMBC (and if necessary) ROESY spectra were recorded using standard pulse programs and are illustrated in Figures S5, S8, S11, S14 and S17. Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in the Tables S14, S16, S18, S20 and S22. HR-ESI-MS data were obtained on a Bruker Maxis 4G mass spectrometer. Full scan mass spectra were acquired in a range from 150–2500 *m/z* in a positive mode. HR-ESI-MS data of hybrid myxochromides is shown in Table S13.

Table S13: HR-ESI-MS data of isolated hybrid myxochromides.

Compound	Formula	[M+H] ⁺ calc.	[M+H] ⁺ exp.	Δ <i>m/z</i> [ppm]
Myxochromide AS ₄	C ₃₈ H ₅₂ N ₆ O ₈	721.39194	721.39373	2.48
Myxochromide SA ₃	C ₄₈ H ₆₉ N ₇ O ₉	888.52295	888.52409	1.28
Myxochromide SB ₄	C ₅₄ H ₈₀ N ₈ O ₁₀	1015.62267	1015.62390	1.21
Myxochromide SC ₄	C ₄₆ H ₆₆ N ₆ O ₈	831.50149	831.50222	0.88
Myxochromide SD ₃	C ₄₄ H ₆₄ N ₆ O ₈	791.47019	791.47051	0.40

For the assignment of the absolute configuration, Marfey's method based on amino acid derivatization was applied.³² 0.1–0.3 mg of pure compound was hydrolyzed with 37% HCl (0.2 mL) in a 1.5 mL glass vial for 3 days at 110°C. The hydrolysate was evaporated to dryness and dissolved in H₂O (100 μL). A 50 μL aliquot was supplemented with 1N NaHCO₃ (20 μL) and 1% 1-fluoro-2,4-dinitrophenyl-5-*L/D*-leucinamide (*L*-FDLA or *D*-FDLA) solution in acetone (20 μL), and the mixtures were heated to 40 °C for 8 h at 700 rpm. After cooling down to room temperature, the solutions were neutralized with 2N HCl (20 μL), evaporated to dryness and the derivatized amino acids were dissolved in 300 μL acetonitrile. An amino acid standard mix (Sigma Aldrich) as well as *N*-Me-*L*-Threonine (Sigma Aldrich) were derivatized via the same procedure and all samples were analyzed on a Dionex Ultimate 3000 RSLC system coupled to a Bruker Maxis 4G mass spectrometer. Separation was performed using a Waters BEH C18, 100 × 2.1 mm, 1.7 μm d_p column. At a flow rate of 0.6 mL/min, the following gradient was applied (A: deionized water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid): 0 min 5% B, 0–1 min 5–10% B, 1–15 min 10–35% B, 15–22

min 35-55% B, 22-25 min 55-80% B, 25-26 min 80% B, 26-26.5 min 80-5% B, 26.5-31 min 5% B. Full scan mass spectra were acquired in a range from 100-1000 m/z .

Structure of myxochromide AS₄

Structure elucidation of myxochromide AS₄ was achieved using 1D ¹H and 2D ¹H-¹H COSY, HSQC and HMBC spectra (Figure S6). Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S14. The ¹H NMR spectrum exhibited signals corresponding to five α -CH protons (δ_H 3.8-5.6), four CH₃ groups (δ_H 1.3-1.7) and two CH₂ groups (δ_H 2.0-2.3) together with a *N*-Me group (δ_H 3.29, 3H, s). Moreover, a number of downfield signals belonging to the unsaturated polyketide side chain (δ_H 5.8-7.3) and a CH₃ signal (δ_H 1.04, 3H, t) were observed. 2D NMR data revealed the presence *N*-Me-threonine, glutamine, alanine and a polyene side chain. Amino acid sequence was established by means of key HMBC correlations and final structure was elucidated as shown in Figure S5. For the assignment of the absolute configuration of myxochromide AS₄, hydrolysis and Marfey analysis of the obtained amino acids,³² was applied as described above. The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S7 and stereochemical assignments are illustrated in Table S15. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that all amino acids of the myxochromide AS₄ peptide core show *L*-configuration, while glutamine was converted to glutamic acid during hydrolysis. This correlates with the assumption that in the underlying hybrid pathway, the condensation domain of module 3 specifically processes the *L*-configured aminoacyl donor (^LC_L domain), although the presence of an epimerization domain in module 2 of the assembly line points to the incorporation of *D*-Ala into this position of the peptide core (see Chapter 3.3.8).

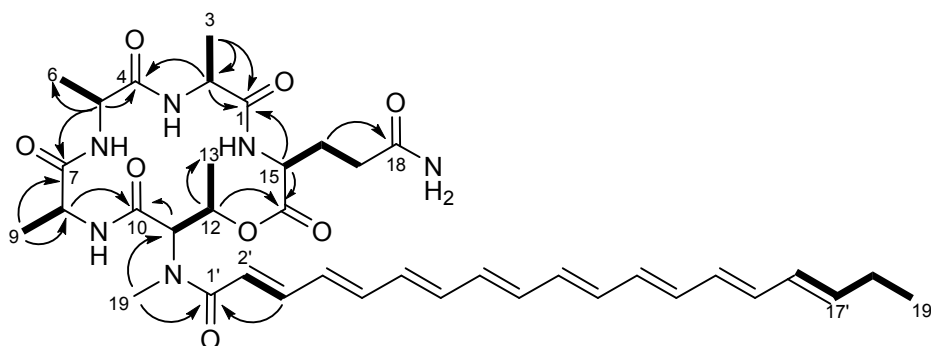
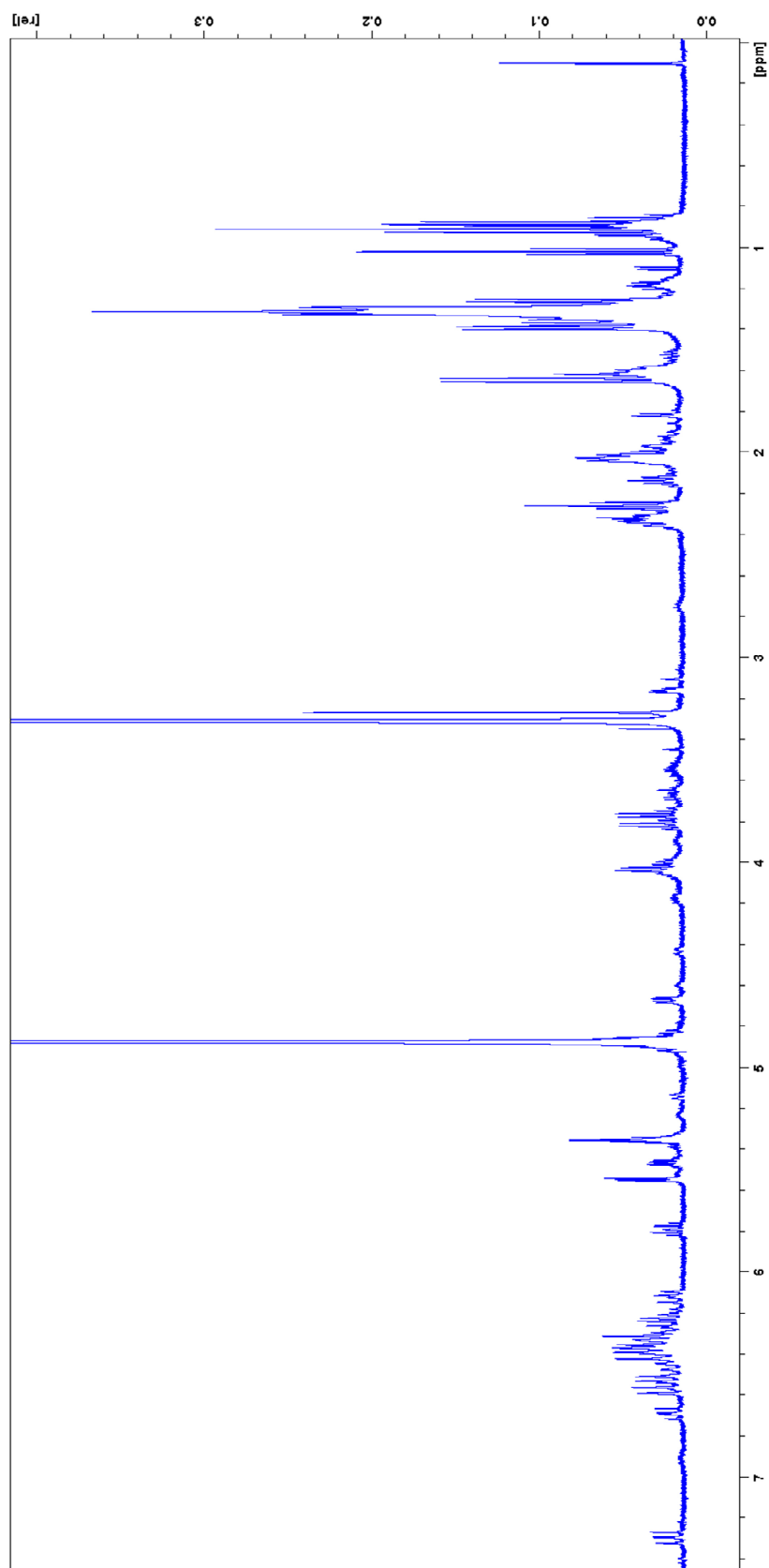


Figure S5. Structure of myxochromide AS₄ showing selected COSY (bold line) and key HMBC (arrow) correlations.

Table S14. NMR spectroscopic data of myxochromide AS₄.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c
<i>L</i> -Ala (1)	1	173.2		
	2	53.0	3.77, <i>q</i> (7.3)	1, 3, 4
	3	15.3	1.65, <i>d</i> (7.3)	1, 2
<i>L</i> -Ala (2)	4	173.7		
	5	51.6	3.82, <i>q</i> (7.0)	4, 6, 7
	6	14.9	1.39, <i>d</i> (7.0)	4, 5
<i>L</i> -Ala (3)	7	176.6		
	8	51.1	4.03, <i>q</i> (7.3)	7, 9, 10
	9	16.1	1.32, <i>m</i>	7, 8
<i>N</i> -Me- <i>L</i> -Thr	10	171.1		
	11	59.5	5.55, <i>d</i> (4.2)	1', 10, 12, 19
	12	74.0	5.46, <i>m</i>	13, 14
	13	16.5	1.26, <i>d</i> (6.5)	11, 12
	19	35.0	3.27, <i>s</i>	1', 11
<i>L</i> -Gln	14	170.9		
	15	51.9	4.67, <i>dd</i> (3.2, 9.6)	1, 14, 16, 17
	16a	27.8	1.93, <i>m</i>	15, 17, 18
	16b		2.04, <i>m</i>	
	17a	31.9	2.26, <i>m</i>	14, 16
	17b		2.32, <i>m</i>	
	18	178.2		
Side chain	1'	170.6		
	2'	119.7	6.57, <i>d</i> (14.6)	1'
	3'	145.2	7.29, <i>dd</i> (11.4, 14.6)	1', 2', 5'
	4'	138.4	6.52, <i>m</i>	
	5'	141.8	6.69, <i>m</i>	
	6'-14'	<i>d</i>	<i>d</i>	
	15'	135.1	6.24, <i>m</i>	
	16'	130.9	6.12, <i>m</i>	15', 18'
	17'	138.1	5.79, <i>m</i>	15', 18', 19'
	18'	26.7	2.14, <i>m</i>	16', 17', 19'
	19'	13.7	1.02, <i>t</i> (7.2)	18'

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d overlapped signals.



^1H -NMR spectrum of myxochromide AS_4 in CD_3OD (500 MHz)

Figure S6 (continued on next page)

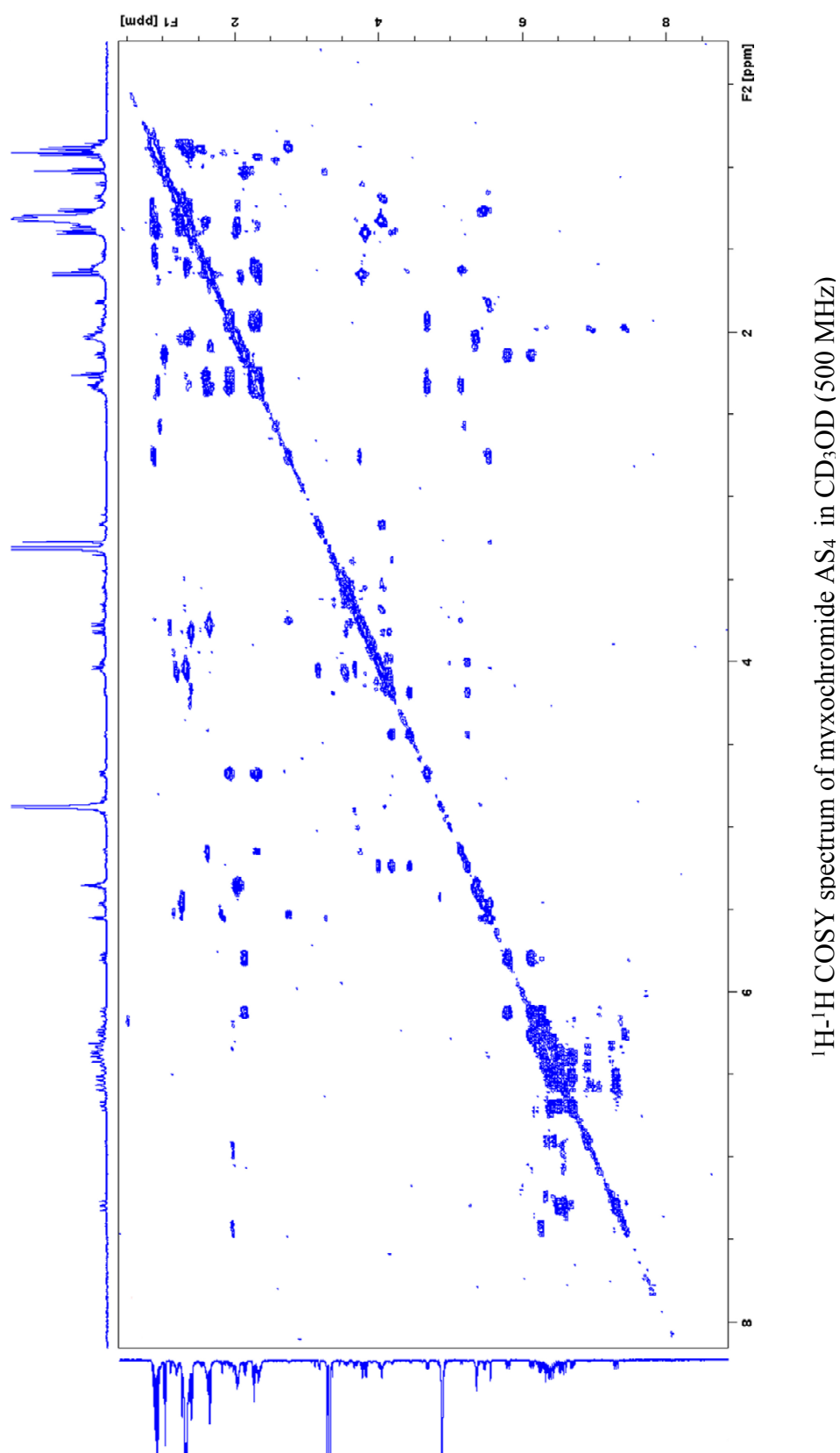


Figure S6 (continued on next page)

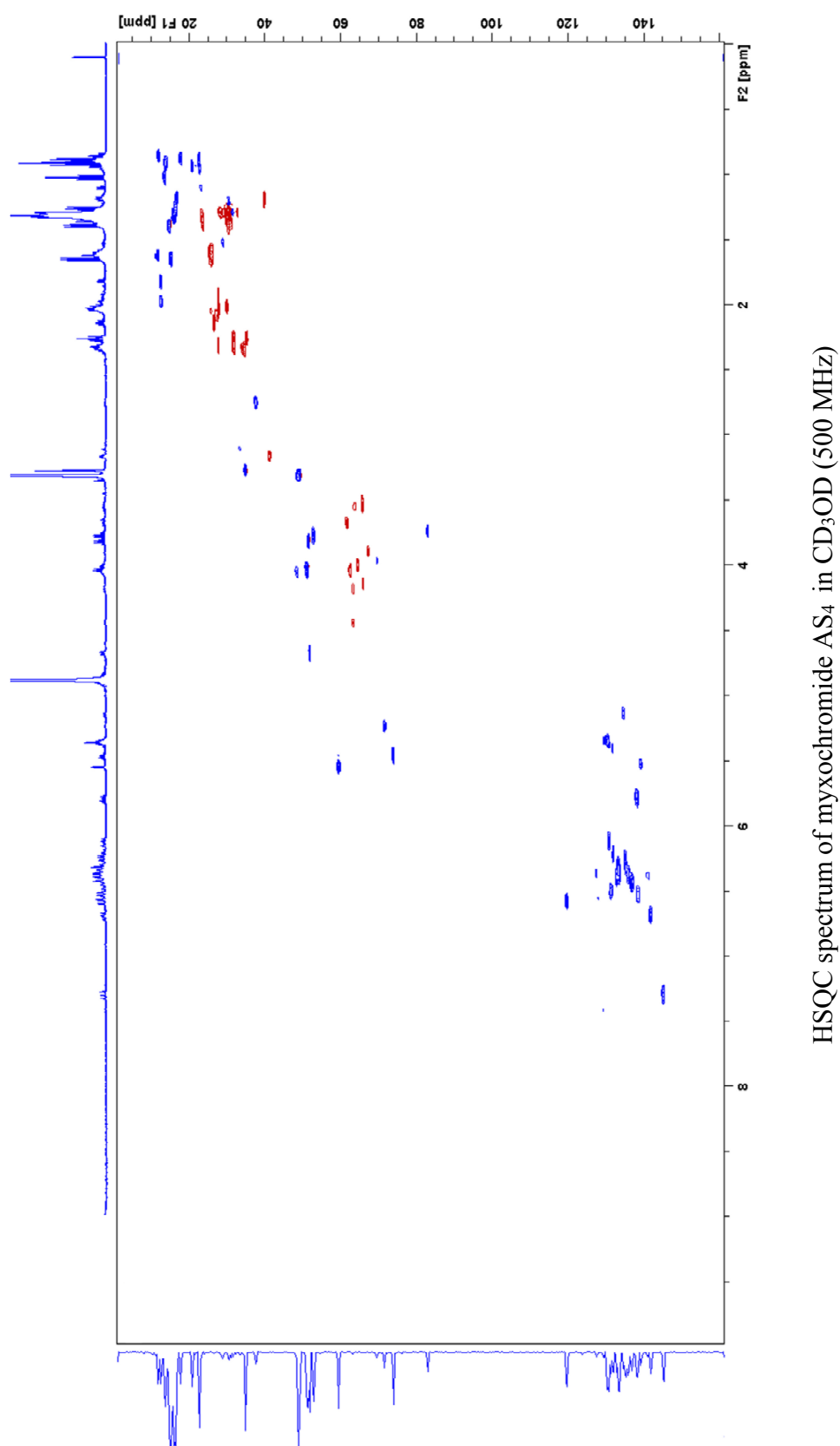


Figure S6 (continued on next page)

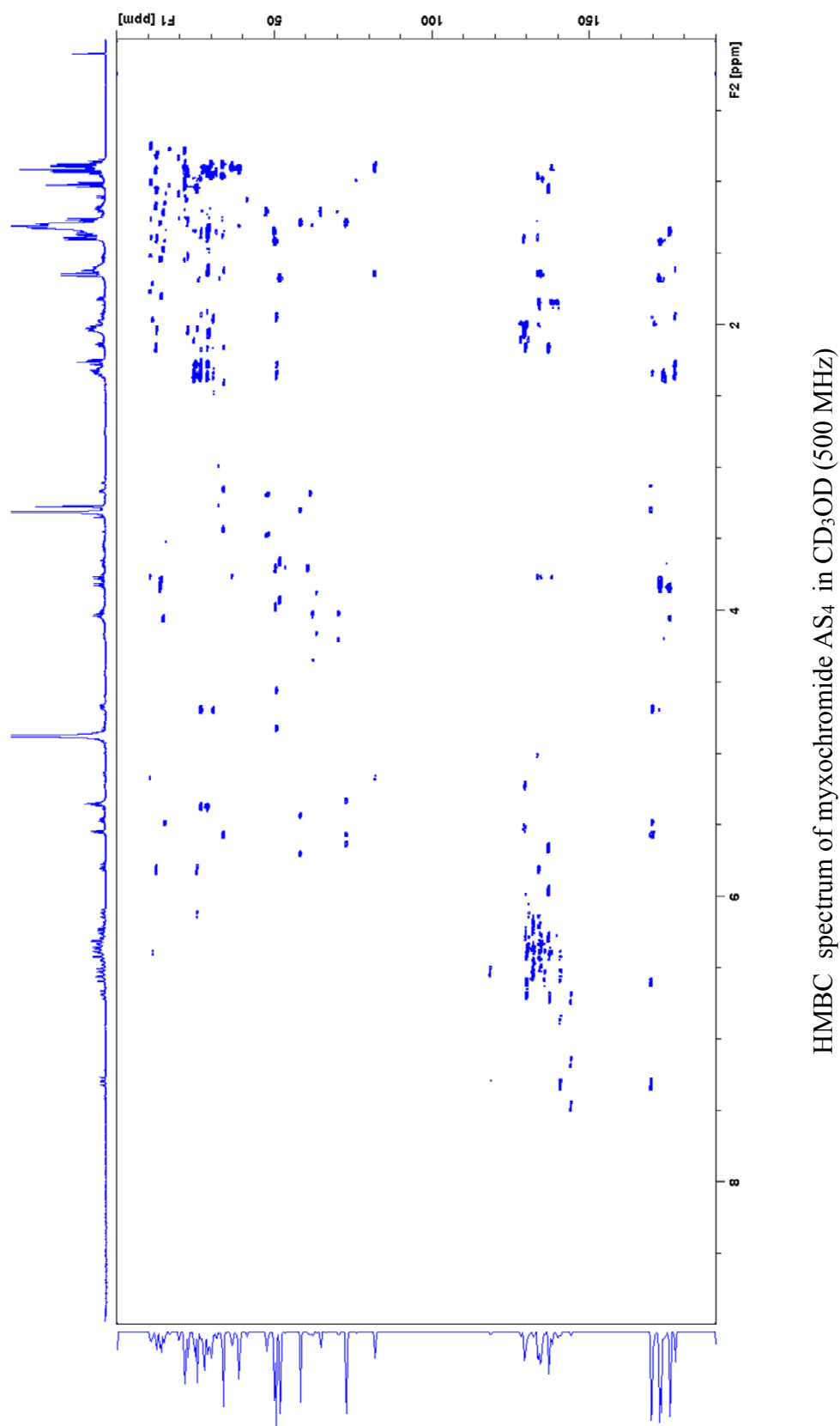


Figure S6. NMR spectra of myxochromide AS₄.

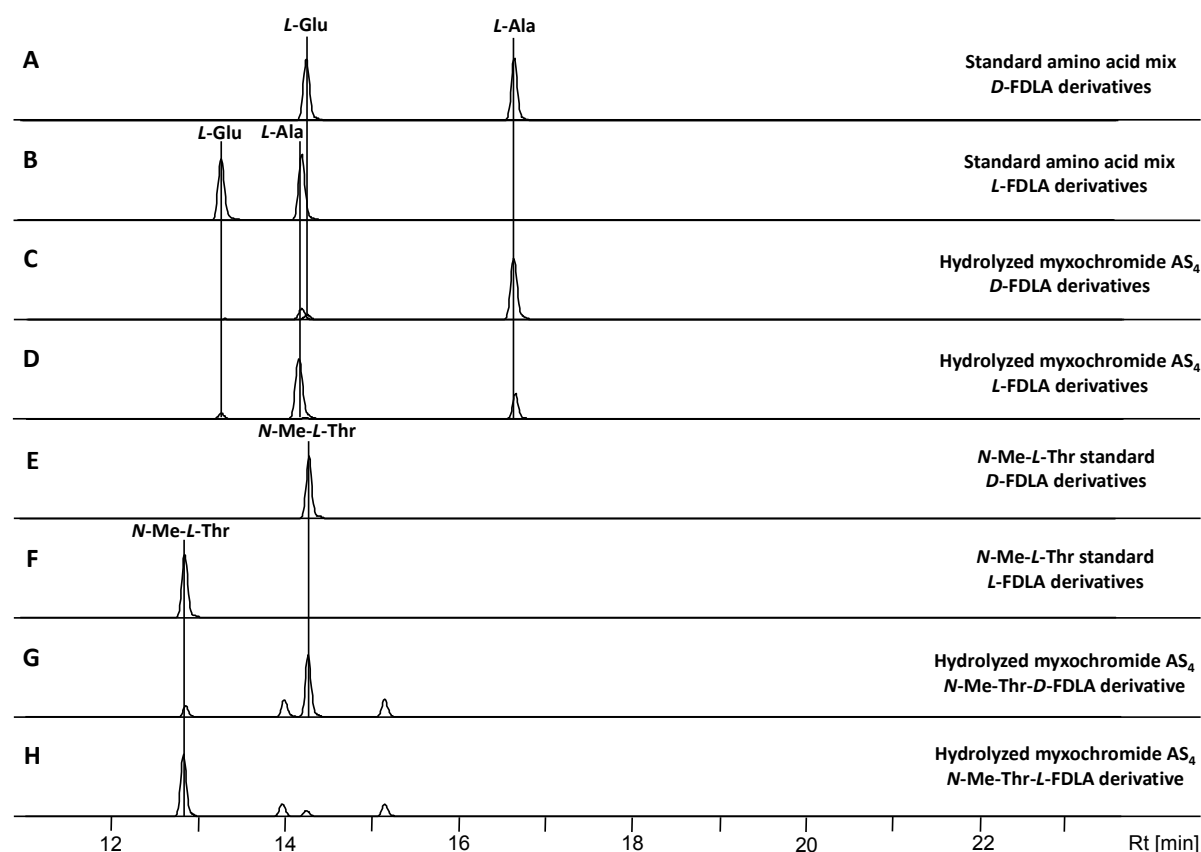


Figure S7. Results of the analysis of the absolute configuration by advanced Marfey's method.³² Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide AS₄ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide AS₄ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative.

Table S15. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide AS₄.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	R _t [min]	m/z $[M+H]^+$	R _t [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.3	442.1578	14.3	442.1563	L
Glu- <i>L</i> -FDLA	13.3	442.1579	13.3	442.1574	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1517	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1518	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1517	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1518	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1517	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1518	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.4	428.1782	14.3	428.1774	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.9	428.1786	12.8	428.1776	

Structure of myxochromide SA₃

Structure elucidation of myxochromide SA₃ was achieved using 1D ¹H and 2D ¹H-¹H COSY, HSQC and HMBC spectra (Figure S9). Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S16. The COSY spectrum supported by HSQC and HMBC data showed presence of spin systems corresponding to *N*-Me-threonine, glutamine, alanine, proline and leucine residues as well as a polyene side chain. Amino acid sequence was established by means of key HMBC correlations and final structure was elucidated as shown in Figure S8. For the assignment of the absolute configuration of myxochromide SA₃, Marfey analysis of the obtained amino acids,³² was applied as described above. The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S10 and stereochemical assignments are illustrated in Table S17. Comparison of the retention times and *m/z* values of derivatized standard amino acids and the hydrolyzed lipopeptide revealed the presence of a *D*-configured leucine residue (C16) in myxochromide SA₃. The amino acids alanine (C2), proline (C5), another leucine (C10), *N*-Me-threonine (C22) and glutamine (C26), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. These findings demonstrate that the epimerization domain of module 2 in the underlying hybrid assembly line is not specific for alanine, but also exerts its function on the more bulky leucine residue. The downstream condensation domain from module 3 originating from the A-type *mch* pathway is obviously a ^DCL-type domain, thereby processing the *D*-configured dipeptide intermediate.

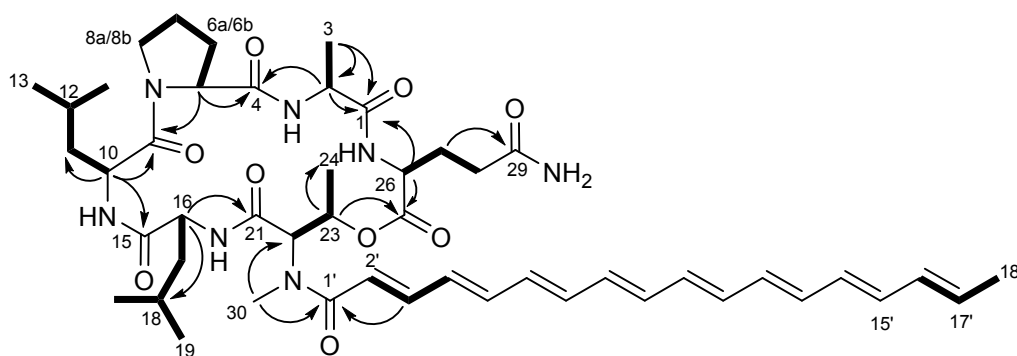


Figure S8. Structure of myxochromide SA₃ showing selected COSY (bold line) and key HMBC (arrow) correlations.

Table S16: NMR spectroscopic data of myxochromide SA₃.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c
<i>L</i> -Ala	1	172.9		
	2	52.1	3.84, <i>q</i> (7.3)	1, 3, 4
	3	15.8	1.56, <i>d</i> (7.3)	1, 2
<i>L</i> -Pro	4	174.3		
	5	63.5	4.00, <i>dd</i> (9.6, 6.9)	4, 6, 9
	6a	29.8	1.89, <i>m</i>	5, 7
	6b		2.24, <i>m</i>	5, 7
	7a	26.4	2.02, <i>m</i>	6, 8
	7b		2.16, <i>m</i>	6, 8
	8a	47.8	3.55, <i>m</i>	7
	8b		3.74, <i>m</i>	7
<i>L</i> -Leu	9	172.0		
	10	52.1	4.40, <i>m</i>	9, 11, 12, 15
	11a	38.0	1.51, <i>m</i>	10
	11b		1.56, <i>m</i>	10
	12	25.8	1.53, <i>m</i>	10, 11, 13, 14
	13	23.4	0.98, <i>d</i> (6.9)	11, 12, 14
	14	23.4	0.98, <i>d</i> (6.9)	13
<i>D</i> -Leu	15	174.8		
	16	51.7	4.78, <i>t</i> (7.5)	15, 18, 21
	17a	40.4	1.49, <i>m</i>	18, 19, 20
	17b		1.57, <i>m</i>	18, 19, 20
	18	25.9	1.52, <i>m</i>	
	19	22.6	0.94, <i>m</i>	20
	20	22.6	0.94, <i>m</i>	19
<i>N</i> -Me- <i>L</i> -Thr	21	169.0		
	22	60.7	5.57, <i>d</i> (3.3)	1', 21, 23
	23	72.2	5.73, <i>m</i>	24, 25
	24	16.3	1.09, <i>d</i> (6.6)	21, 22
	30	35.2	3.21, <i>s</i>	1', 22
<i>L</i> -Gln	25	171.5		
	26	54.0	4.51, <i>dd</i> (8.4, 8.1)	1, 25, 27, 28
	27a	28.8	1.80, <i>m</i>	29
	27b		2.34, <i>m</i>	29
	28a	32.1	2.23, <i>m</i>	29
	28b		2.32, <i>m</i>	29
	29	177.2		
Side chain	1'	170.9		
	2'	120.0	6.60, <i>d</i> (14.7)	1'
	3'	144.8	7.35, <i>dd</i> (14.7, 11.5)	1'
	4'	131.3	6.49, <i>m</i>	
	5'-15'	<i>d</i>	<i>d</i>	<i>d</i>
	16'	133.1	6.14, <i>m</i>	

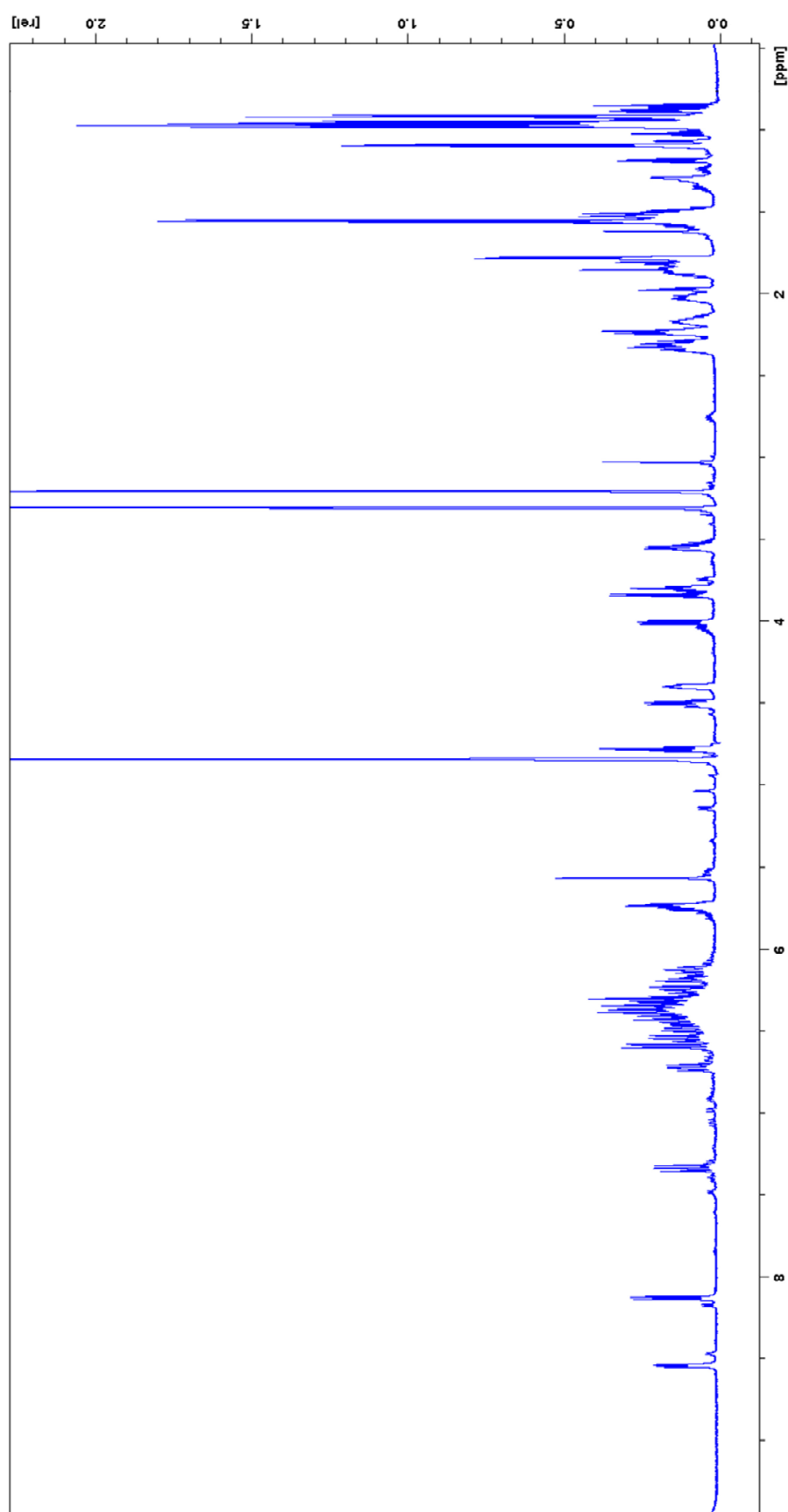
17'	131.1	5.75, <i>m</i>	18'
18'	18.2	1.78, <i>d</i> (6.6)	17'

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.

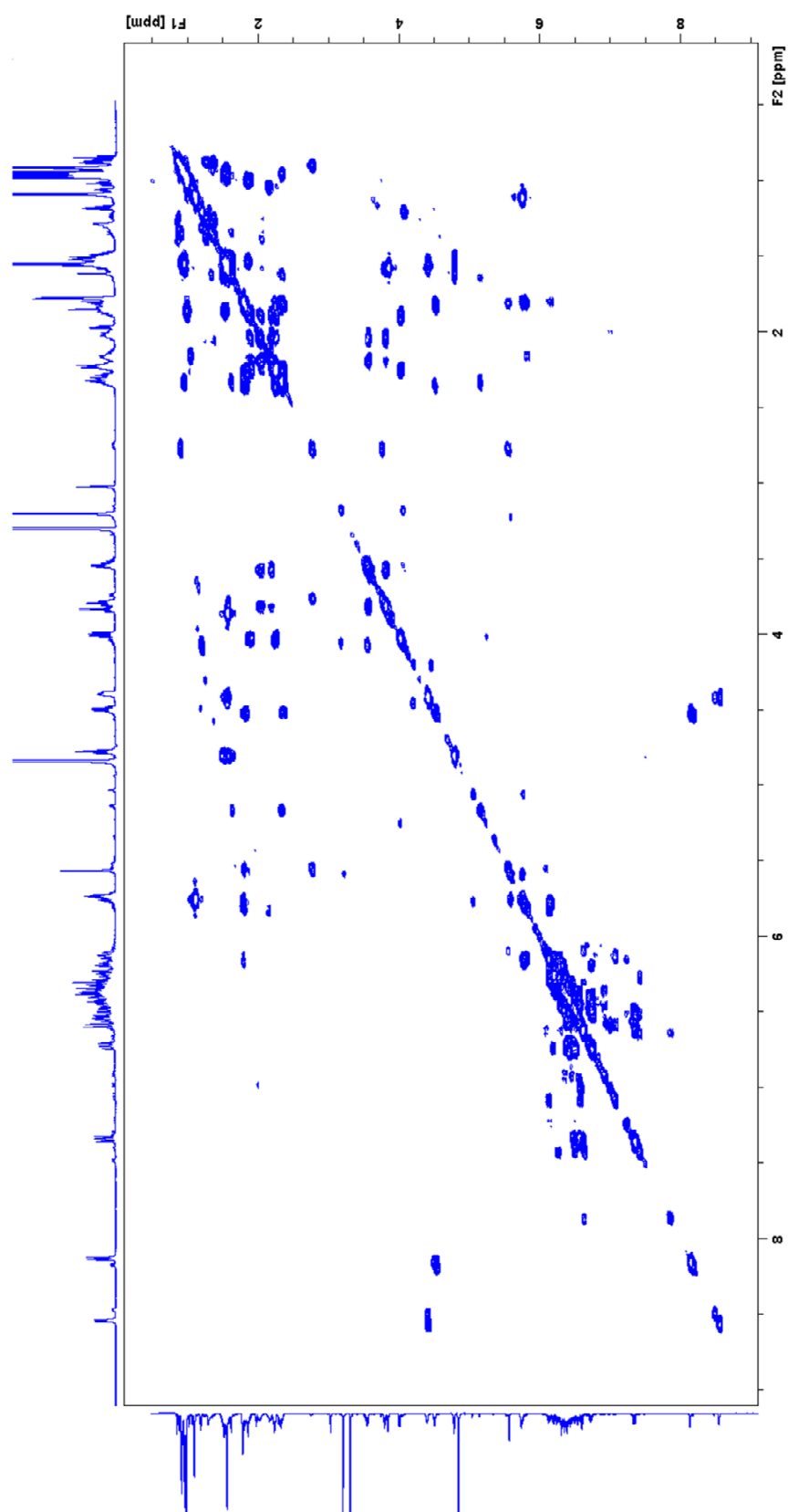
^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.

^c proton showing HMBC correlations to indicated carbons.

^d overlapped signals.



^1H -NMR spectrum of myxochromide SA_3 in CD_3OD (500 MHz)



^1H - ^1H COSY spectrum of myxochromide SA_3 in CD_3OD (500 MHz)

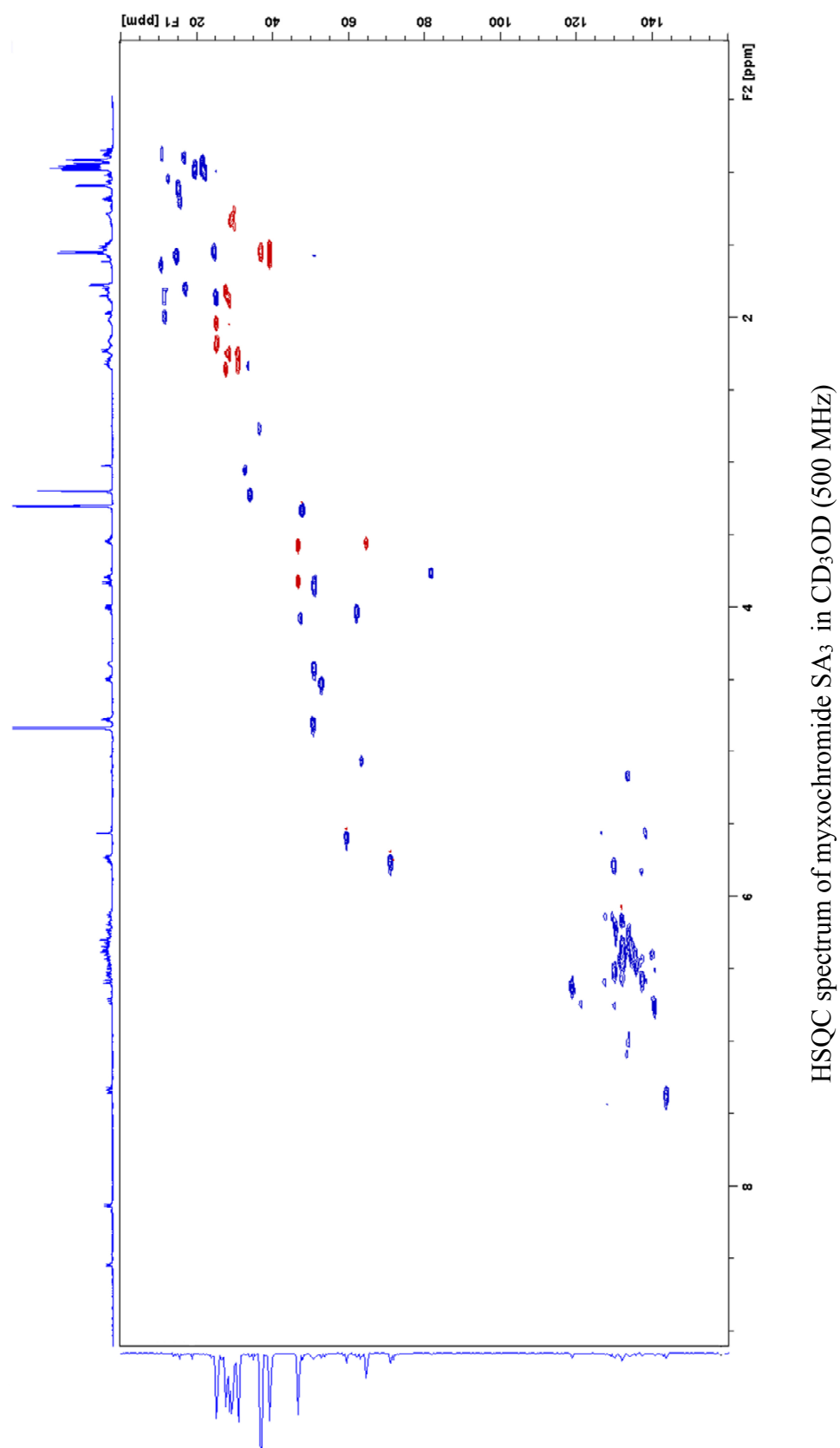


Figure S9 (continued on next page)

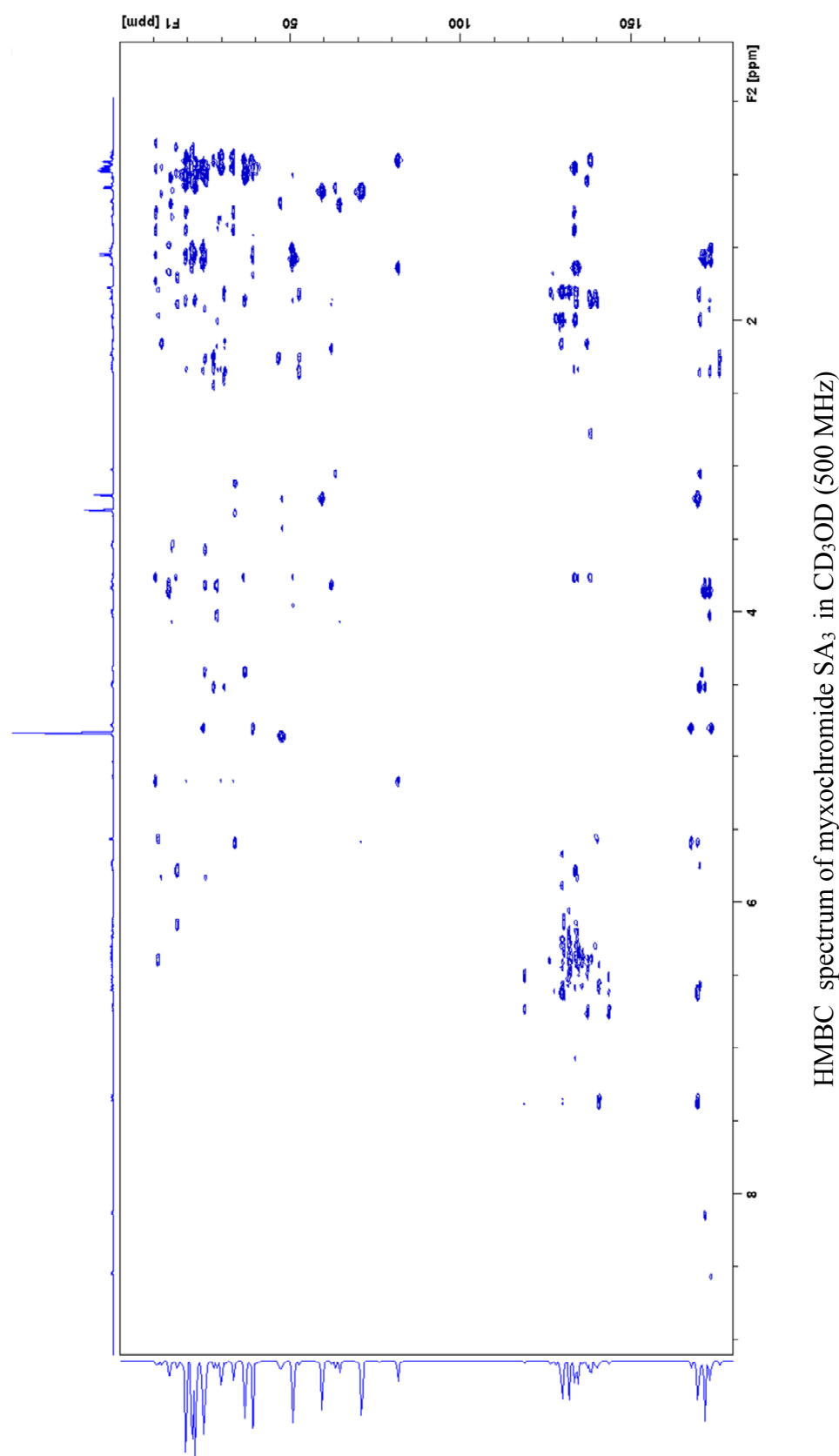


Figure S9. NMR spectra of myxochromide SA₃.

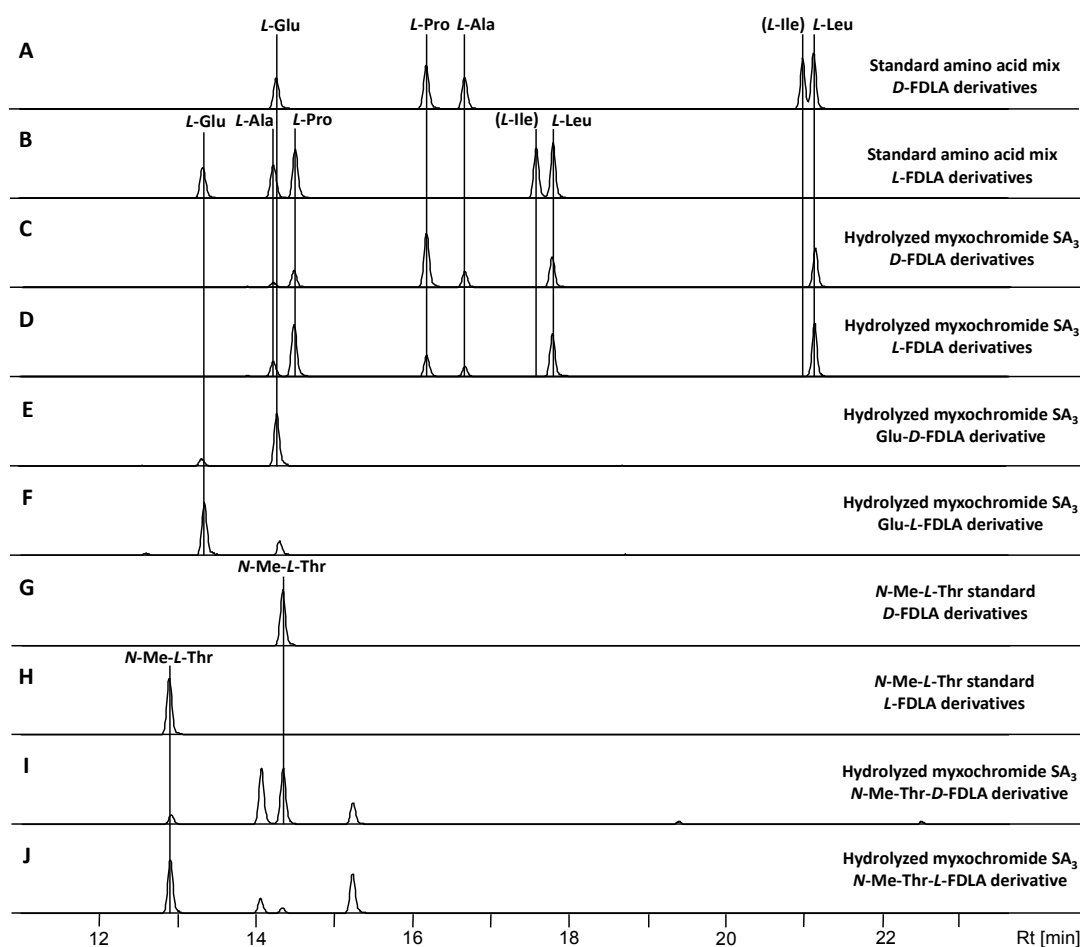


Figure S10. Results of the analysis of the absolute configuration by advanced Marfey's method.³² Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide SA₃ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide SA₃ derivatized with *L*-FDLA reagent. **E:** Same sample as in **C** analyzed for the *L*-glutamic acid *D*-FDLA derivative. **F:** Same sample as in **D** analyzed for the *L*-glutamic acid *L*-FDLA derivative. **G:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **H:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **I:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **J:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA.

Table S17. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide SA₃.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	tr [min]	m/z $[M+H]^+$	tr [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.3	442.1578	14.3	442.1576	L
Glu- <i>L</i> -FDLA	13.3	442.1579	13.3	442.1570	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1515	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1511	
Pro- <i>D</i> -FDLA	16.2	410.1675	16.2	410.1676	L
Pro- <i>L</i> -FDLA	14.5	410.1675	14.5	410.1673	
Leu- <i>D</i> -FDLA	21.1	426.1989	21.1	426.1993	L
Leu- <i>L</i> -FDLA	17.8	426.1988	17.8	426.1986	
Leu- <i>D</i> -FDLA	21.1	426.1989	17.8	426.1993	D
Leu- <i>L</i> -FDLA	17.8	426.1988	21.1	426.1982	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.4	428.1782	14.4	428.1786	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.9	428.1786	12.9	428.1785	

Structure of myxochromide SB₄

Structure elucidation of myxochromide SB₄ was achieved using ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra (Figure S12). Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in the Table S18. The ¹H NMR spectrum closely resembled to that of myxochromide SA₃. In addition to the common structural parts, analysis of 2D NMR spectra corroborated the presence of an additional leucine residue compared to myxochromide SA₃. Key HMBC correlations established the amino acid sequence and finalized its planar structure as depicted in Figure S11. Length of the polyene side chain was deduced based on the HR-MS data and molecular formula. For the assignment of the absolute configuration of myxochromide SB₄, hydrolysis and Marfey analysis of the obtained amino acids,³² was applied as described above. The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S13 and stereochemical assignments are illustrated in Table S19. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that one of the three leucine residues (C10, C16 and C22) from myxochromide SA₃ has *D* configuration. The remaining leucine residues as well as the amino acids alanine (C2), proline (C5), *N*-Me-threonine (C28) and glutamine (C32), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. According to the domain organization of the underlying hybrid assembly line, which harbors an epimerization domain in module 2, the *D*-configured leucine was assigned to C22. This also correlates with the structure of myxochromide SA₃ and identifies the condensation domain of module 3 originating from the B-type *mch* pathway as a ^DC_L domain.

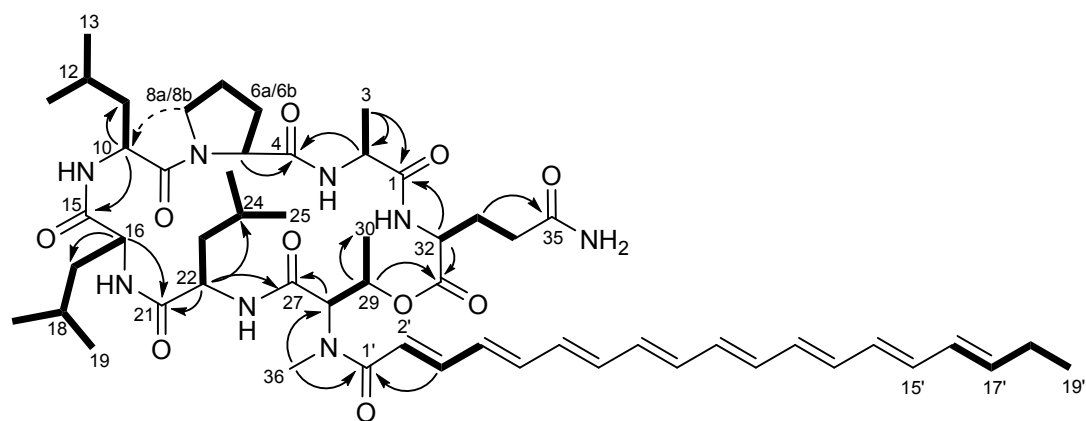


Figure S11. Structure of myxochromide SB₄ showing selected COSY (bold line), ROESY (dashed arrow) and key HMBC (arrow) correlations.

Table S18. NMR spectroscopic data of myxochromide SB₄.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c	ROESY ^{d, e}
<i>L</i> -Ala	1	174.5			
	2	50.3	4.26 <i>m</i>	1,3	
	3	15.5	1.42 <i>d</i> (7.4)	1,2	
<i>L</i> -Pro	4	174.0			
	5	63.6	4.11 <i>m</i>	4, 6a/b,7a/b	10
	6a	30.5	1.90 <i>m</i>	4,5,7a/b	
	6b	30.5	2.35 <i>m</i>	4, 5,7a/b	
	7a	25.9	2.05 <i>m</i>	5,6a/b,8a/b	
	7b	25.9	2.13 <i>m</i>	5,6a/b,8a/b	
	8a	48.3	3.68 <i>m</i>	6a/b,7a/b	10
	8b	48.3	3.85 <i>m</i>	6a/b,7a/b,9	10
<i>L</i> -Leu	9	174.0			
	10	49.6	4.95 <i>dd</i> (^f)	9,11,12,15	8a/b, 5
	11a	41.3	1.51 <i>m</i>	10,12	
	11b	41.3	1.64 <i>m</i>	10,12	
	12	25.8	1.65 <i>m</i>	13,14	
	13	21.2	0.99 <i>m</i>	11,12,14	
	14	23.5	0.95 <i>m</i>	11,12,13	
<i>L</i> -Leu	15	174.4			
	16	54.5	4.27 <i>m</i>	15,17a/b,18,21	
	17a	41.8	1.65 <i>m</i>	19,20	
	17b	41.8	1.76 <i>m</i>	19,20	
	18	25.9	1.74 <i>m</i>		
	19	21.1	0.91 <i>d</i> (6.1)	17a/b	
	20	23.1	1.01 <i>m</i>	17a/b	
<i>D</i> -Leu	21	174.1			
	22	52.6	4.68 <i>m</i>	21,23a/b,24	
	23a	44.3	1.59 <i>m</i>	22,24,25,26	
	23b	44.3	1.44 <i>m</i>	22,24,25,26	
	24	25.7	1.52 <i>m</i>		
	25	22.8	0.94 <i>m</i>	23a/b,24,25	
	26	22.8	0.94 <i>m</i>	23a/b,24,25	
<i>N</i> -Me- <i>L</i> -Thr	27	168.6			
	28	61.0	5.57 <i>d</i> (3.5)	1',27,29,36	
	29	71.0	5.83 <i>m</i>	30,31	
	30	16.2	1.09 <i>d</i> (6.6)	27,28,29	
	36	35.5	3.23 <i>s</i>	1',28	
<i>L</i> -Gln	31	170.7			
	32	55.4	4.12 <i>m</i>	31,33,34	
	33a	28.5	1.92 <i>m</i>	32,34,35	
	33b	28.5	2.16 <i>m</i>	32,34,35	
	34a	31.9	2.29 <i>m</i>	32,33a/b,35	
	34b	31.9	2.53 <i>m</i>	32,33a/b,35	

	35	177.4		
Side chain	1'	170.8		
	2'	120.0	6.69 <i>d</i> (14.9)	1',3'
	3'	145.0	7.36 <i>dd</i> (14.5,11.2)	1',2',4'
	4'	138.4	6.55 <i>m</i>	5'
	5'	141.6	6.72 <i>m</i>	3',4'
	6'-14'	<i>f</i>	<i>f</i>	
	15'	135.1	6.24 <i>m</i>	17'
	16'	130.9	6.12 <i>dd</i> (9.8,15.0)	17'
	17'	138.1	5.79 <i>m</i>	15',18',16'
	18'	26.5	2.14 <i>m</i>	16',17',
	19'	13.6	1.02 <i>t</i> (7.4)	18',17'

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.

^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.

^c proton showing HMBC correlations to indicated carbons.

^d proton showing ROESY correlations to indicated protons.

^e acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.

^f overlapped signals.

^g correlation obtained from HMBC spectra in (CD₃)₂SO (data not shown).

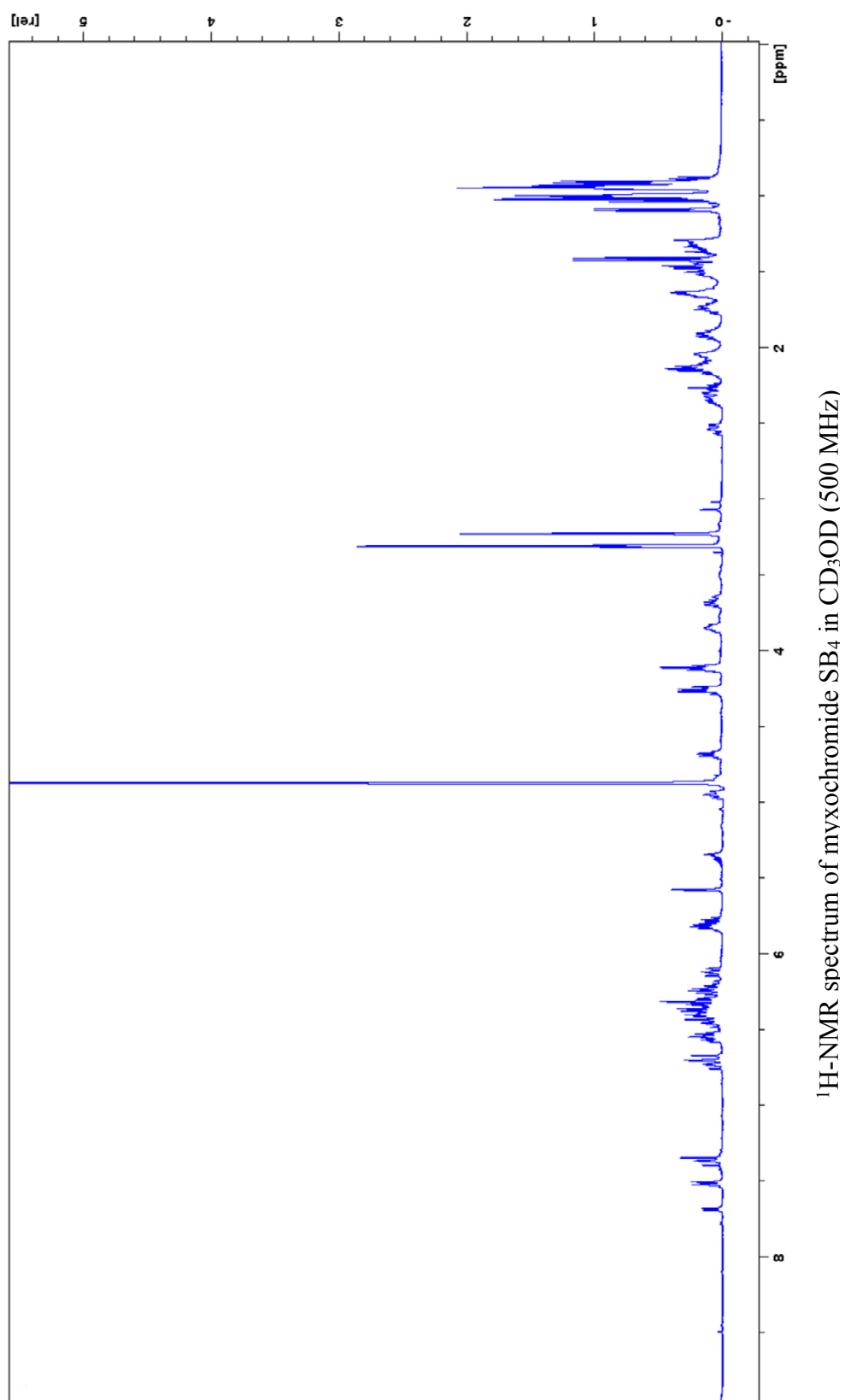


Figure S12 (continued on next page)

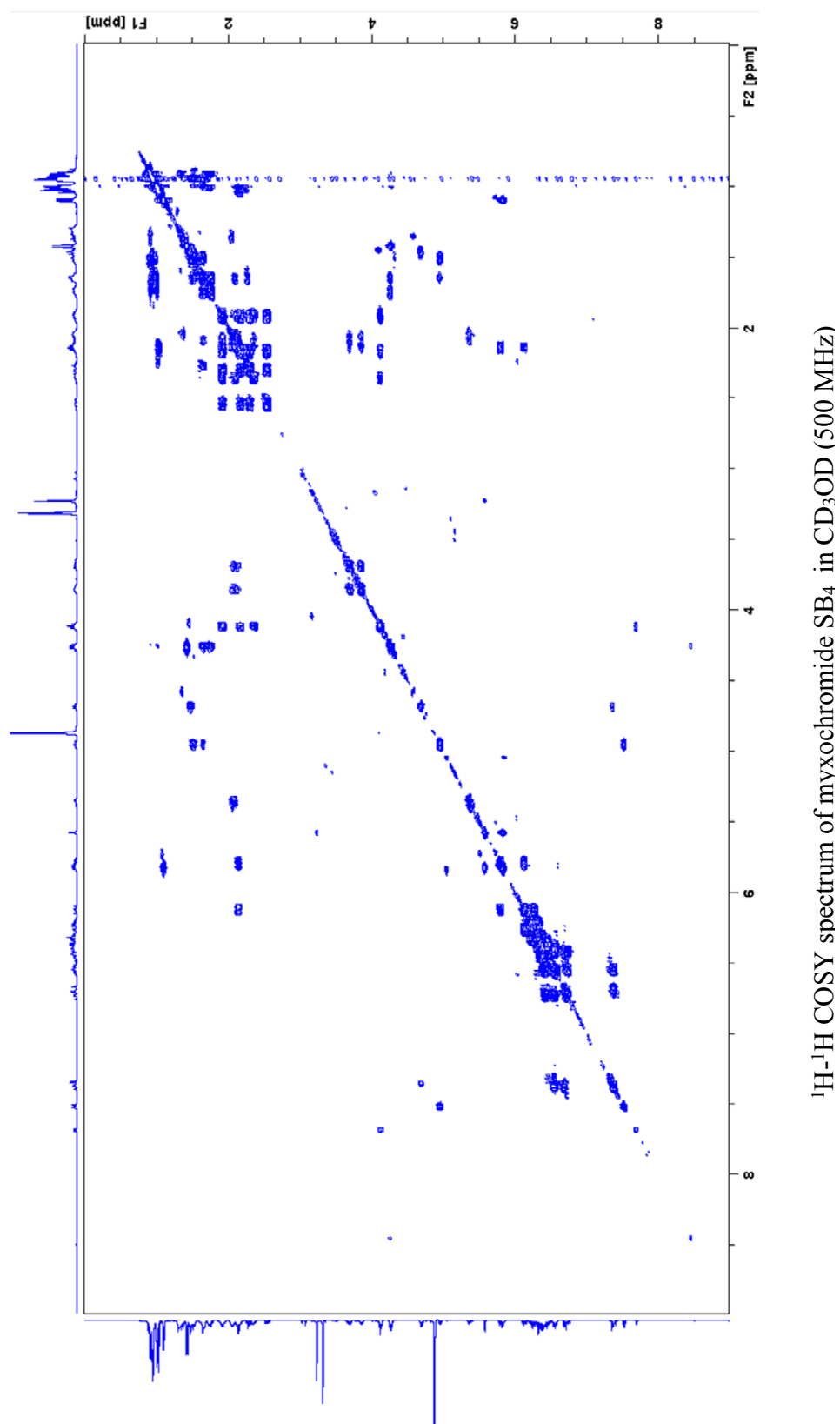


Figure S12 (continued on next page)

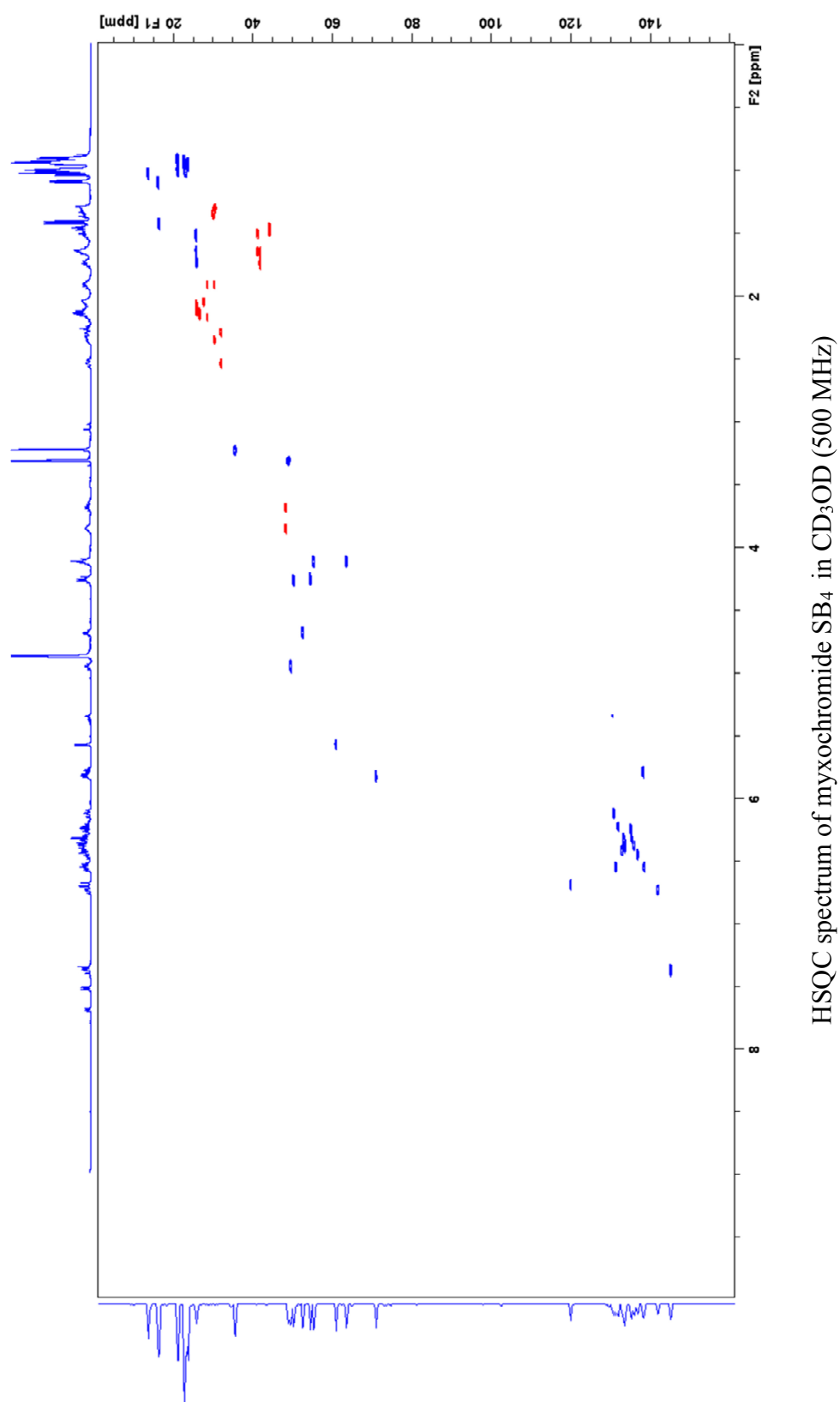
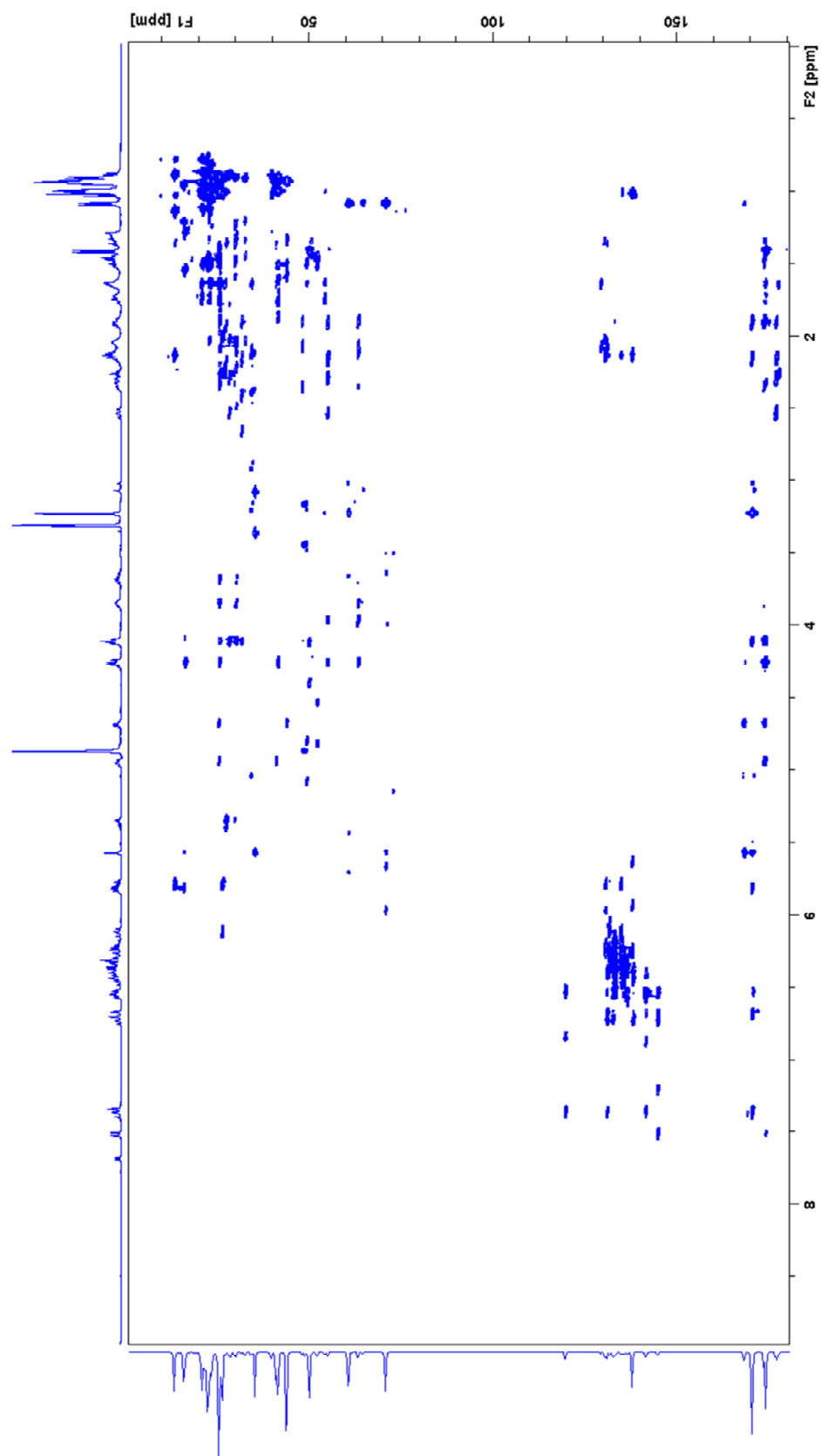


Figure S12 (continued on next page)

HMBC spectrum of myxochromide SB₄ in CD₃OD (500 MHz)

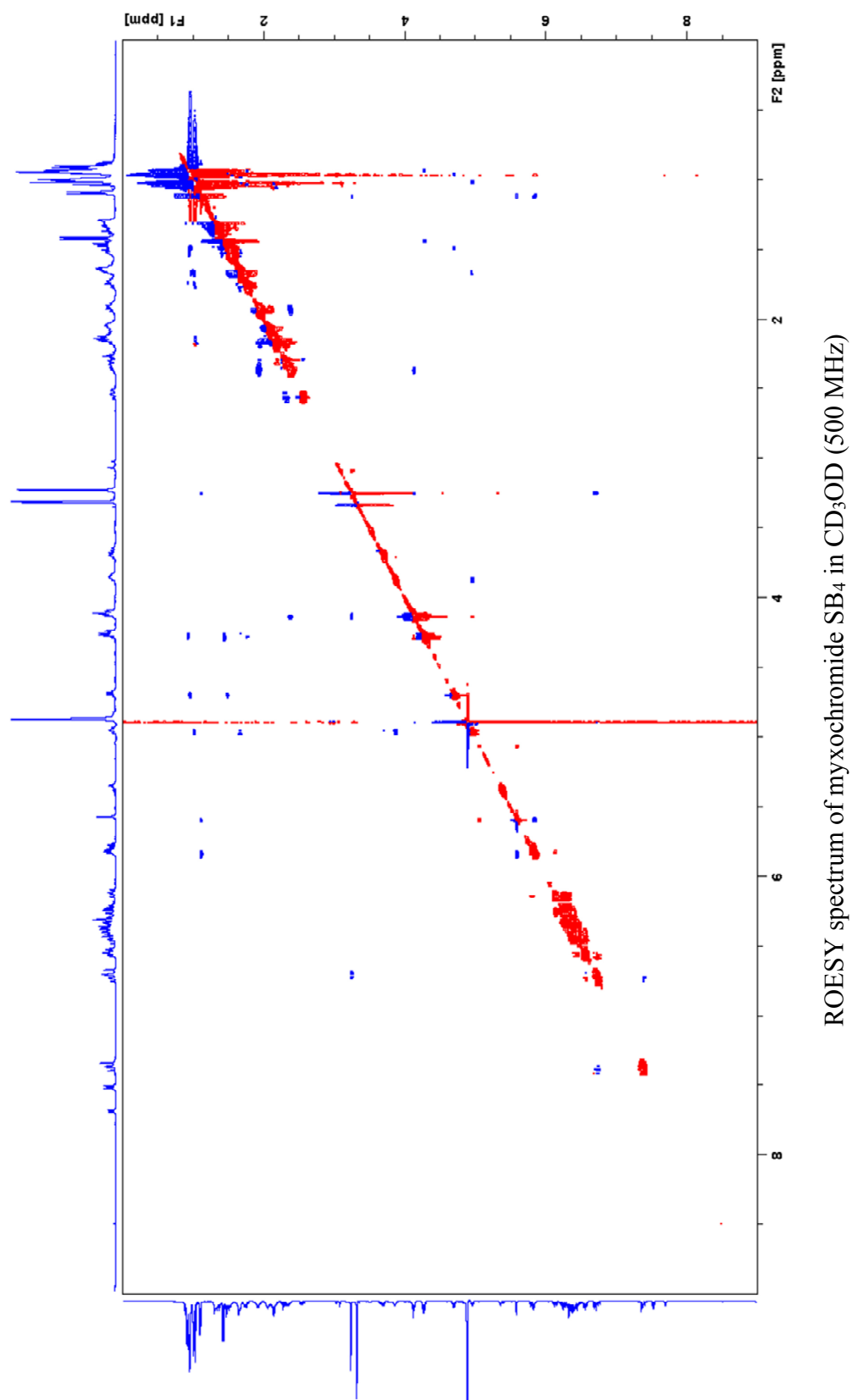


Figure S12. NMR spectra of myxochromide SB₄.

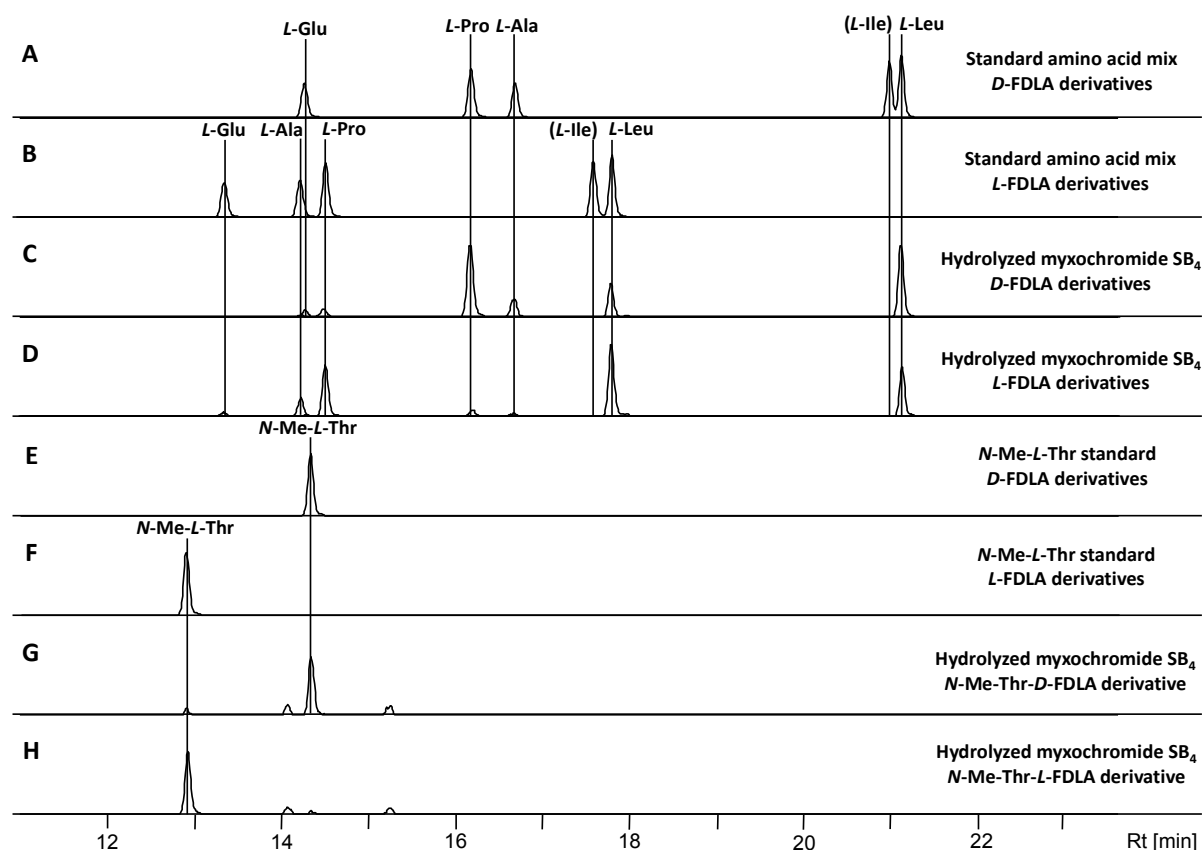


Figure S13. Results of the analysis of the absolute configuration by advanced Marfey's method.³² Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide SB₄ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide SB₄ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative.

Table S19. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide SB₄.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	<i>t_R</i> [min]	<i>m/z</i> [M+H] ⁺	<i>t_R</i> [min]	<i>m/z</i> [M+H] ⁺	
Glu- <i>D</i> -FDLA	14.3	442.1578	14.3	442.1572	L
Glu- <i>L</i> -FDLA	13.3	442.1579	13.3	442.1582	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1513	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1525	
Pro- <i>D</i> -FDLA	16.2	410.1675	16.2	410.1672	L
Pro- <i>L</i> -FDLA	14.5	410.1675	14.5	410.1676	
Leu- <i>D</i> -FDLA	21.1	426.1989	21.1	426.1986	L
Leu- <i>L</i> -FDLA	17.8	426.1988	17.8	426.1985	
Leu- <i>D</i> -FDLA	21.1	426.1989	21.1	426.1986	L
Leu- <i>L</i> -FDLA	17.8	426.1988	17.8	426.1985	
Leu- <i>D</i> -FDLA	21.1	426.1989	17.8	426.1991	D
Leu- <i>L</i> -FDLA	17.8	426.1988	21.1	426.1988	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.4	428.1782	14.4	428.1775	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.9	428.1786	12.9	428.1779	

Structure of myxochromide SC₄

Structure elucidation of myxochromide SC₄ was achieved using 1D ¹H and 2D ¹H-¹H COSY, HSQC, HMBC and ROESY spectra (Figure S15). Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S20. The COSY spectrum supported by HSQC and HMBC data showed presence of spin systems corresponding to *N*-Me-threonine, glutamine, proline and leucine residues as well as a polyene side chain. Amino acid sequence was established by means of key HMBC correlations and final structure was elucidated as shown in Figure S14. For the assignment of the absolute configuration of myxochromide SC₄, hydrolysis and Marfey analysis of the obtained amino acids,³² was applied as described above.³² The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S16 and stereochemical assignments are illustrated in Table S21. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that one of the two leucine residues (C7 and C13) from myxochromide SC₄ is *D*-configured. The second leucine residue as well as the amino acids proline (C2), *N*-Me-threonine (C19) and glutamine (C23), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. According to the domain organization of the underlying hybrid assembly line, which harbors an epimerization domain in module 2, the *D*-configured leucine was assigned to C13. This also correlates with the structures of myxochromide SA₃ myxochromide SB₄ and identifies the condensation domain of module 3 originating from the C-type *mch* pathway as a ^DC_L domain.

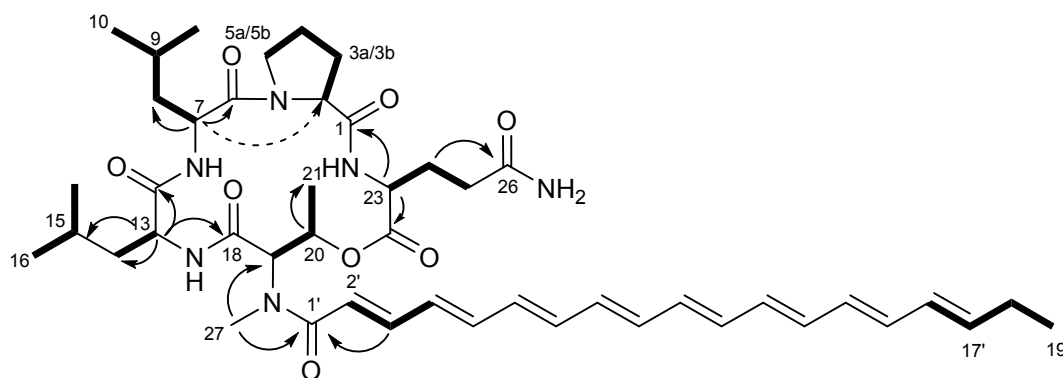


Figure S14. Structure of myxochromide SC₄ showing selected COSY (bold line), ROESY (dashed arrow) and key HMBC (arrow) correlations.

Table S20. NMR spectroscopic data of myxochromide SC₄.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c	ROESY ^{d,e,g}
<i>L</i> -Pro	1	174.2			
	2	63.0	4.34, <i>m</i>	1, 3, 7	7
	3a	32.6	2.18, <i>m</i>	4	
	3b		2.42, <i>m</i>	4	
	4	23.4	2.00, <i>m</i>		
	5a	47.7	3.59, <i>m</i>		
	5b		3.74, <i>m</i>		
<i>L</i> -Leu	6	173.9 ^f			
	7	49.7	4.66, <i>m</i>	6, 8, 9	2
	8a	42.5	1.55, <i>m</i>		
	8b		1.62, <i>m</i>		
	9	25.6	1.53, <i>m</i>		
	10	23.3	0.93, <i>m</i>	8, 9	
	11	23.3	0.93, <i>m</i>	8, 9	
<i>D</i> -Leu	12	173.9 ^f			
	13	54.4	4.26, <i>m</i>	12, 14, 15, 18	
	14a	40.8	1.39, <i>m</i>	12, 15	
	14b		1.75, <i>m</i>	12, 15	
	15	25.6	1.53, <i>m</i>		
	16	21.5	0.91, <i>m</i>	15	
	17	21.5	0.91, <i>m</i>	15	
<i>N</i> -Me- <i>L</i> -Thr	18	170.8			
	19	61.6	5.41, <i>m</i>	18	
	20	72.1	5.97, <i>m</i>		
	21	16.8	1.24, <i>d</i> (6.5)	19, 20	
	27	34.8	3.40, <i>m</i>	1', 19	
<i>L</i> -Gln	22	171.1			
	23	54.4	4.42, <i>m</i>	1, 24, 25	
	24a	28.8	2.03, <i>m</i>	26	
	24b		2.11, <i>m</i>	26	
	25a	31.7	2.22, <i>m</i>	26	
	25b		2.22, <i>m</i>	26	
Side chain	26	176.9			
	1'	170.8			
	2'	119.8	6.65, <i>d</i> (14.8)		
	3'	144.9	7.34, <i>dd</i> (14.8, 11.7)		
	4'	141.8	6.54, <i>m</i>		
	5'-14'	^f	^f		
	15'	134.9	6.27, <i>m</i>		
	16'	130.7	6.12, <i>dd</i> (15.0, 10.0)		
	17'	138.1	5.79, <i>dt</i> (15.0, 6.7)	15'	
	18'	26.6	2.14, <i>m</i>	17', 19'	
	19'	13.6	1.02, <i>t</i> (7.4)	18'	

^a acquired at 175 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.

^b acquired at 700 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.

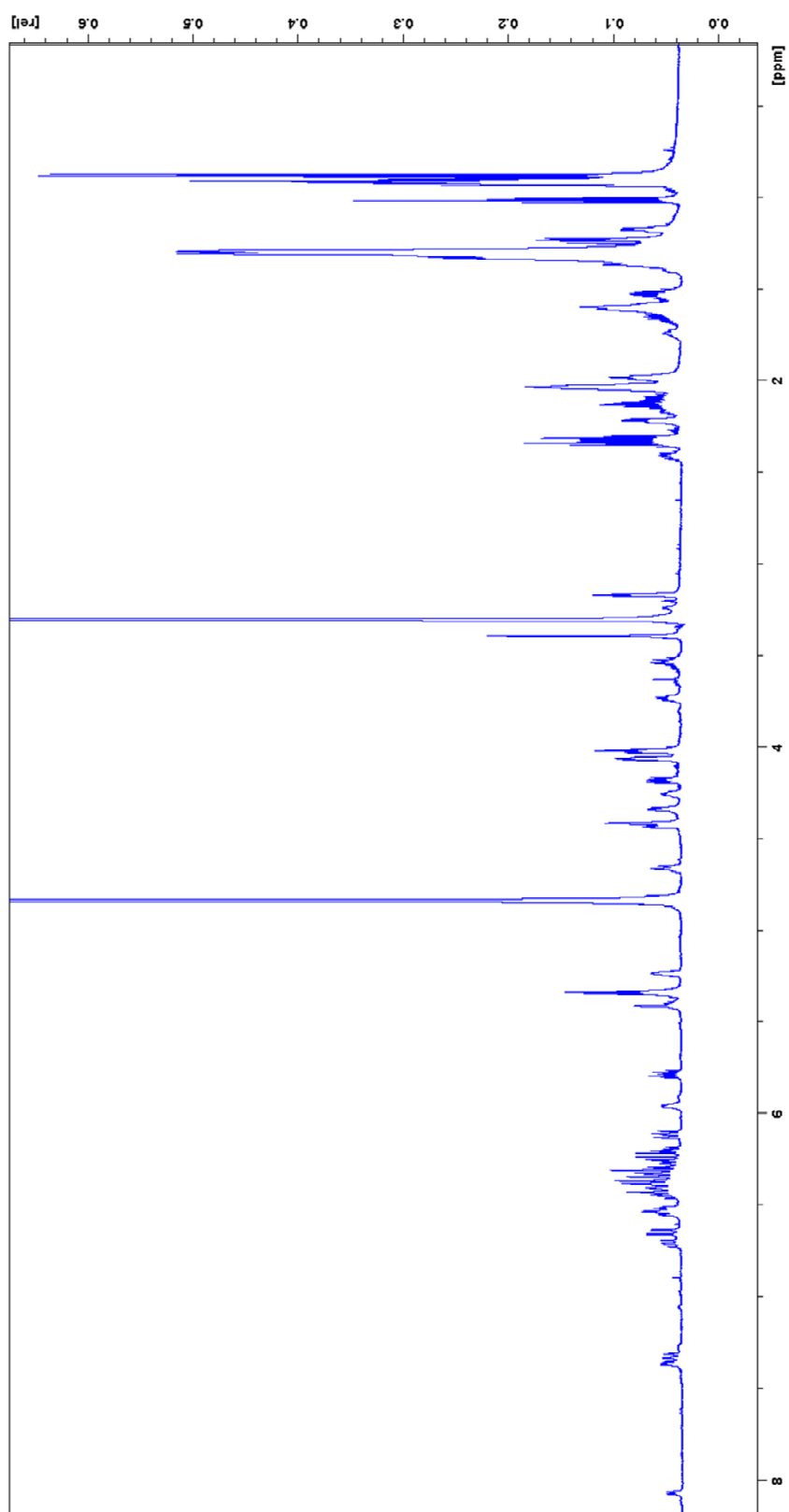
^c proton showing HMBC correlations to indicated carbons.

^d proton showing ROESY correlations to indicated protons.

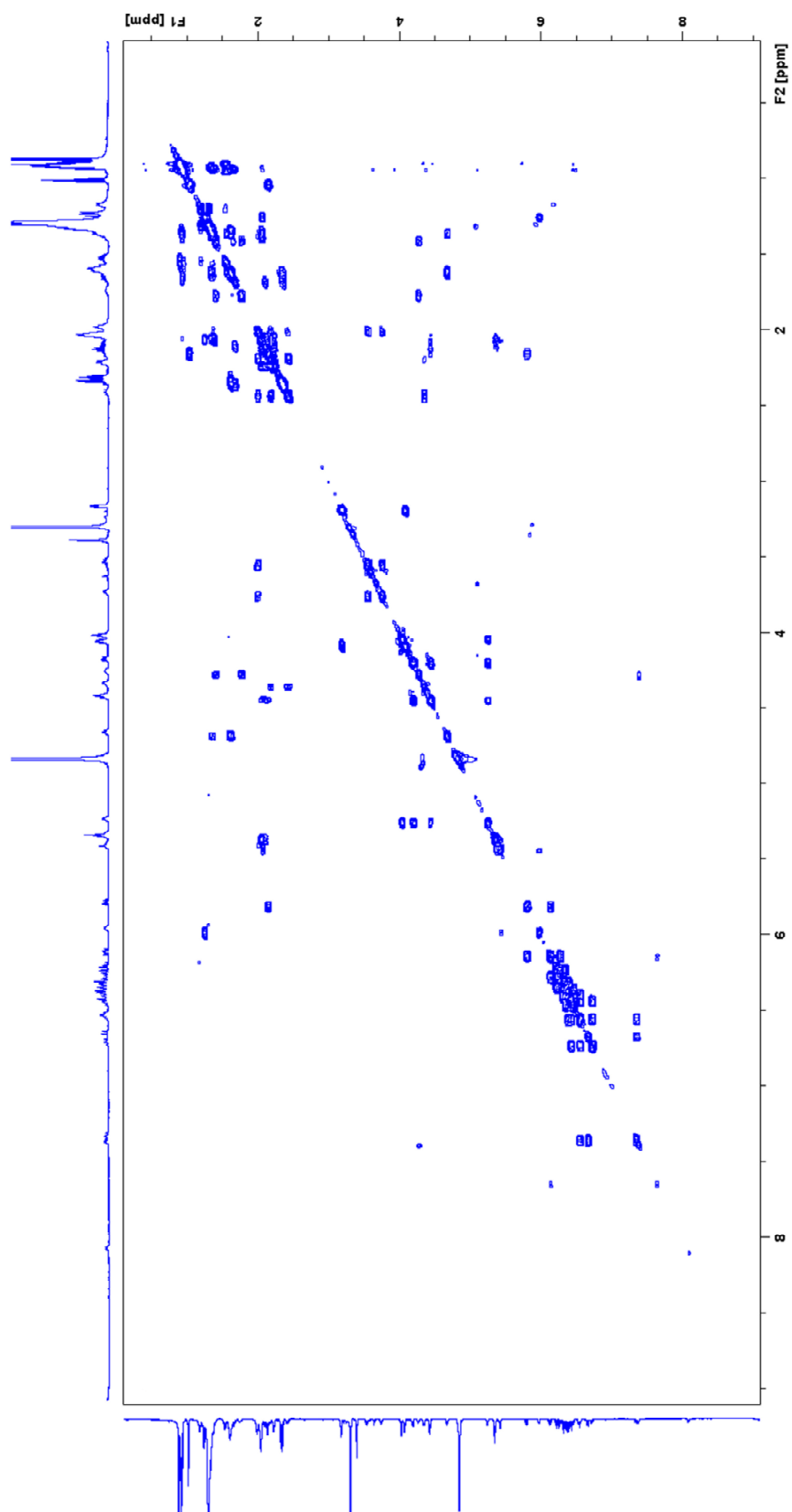
^e acquired at 700 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.

^f overlapped signals.

^g only relevant correlations listed



^1H -NMR spectrum of myxochromide SC_4 in CD_3OD (500 MHz)



^1H - ^1H COSY spectrum of myxochromide SC₄ in CD₃OD (500 MHz)

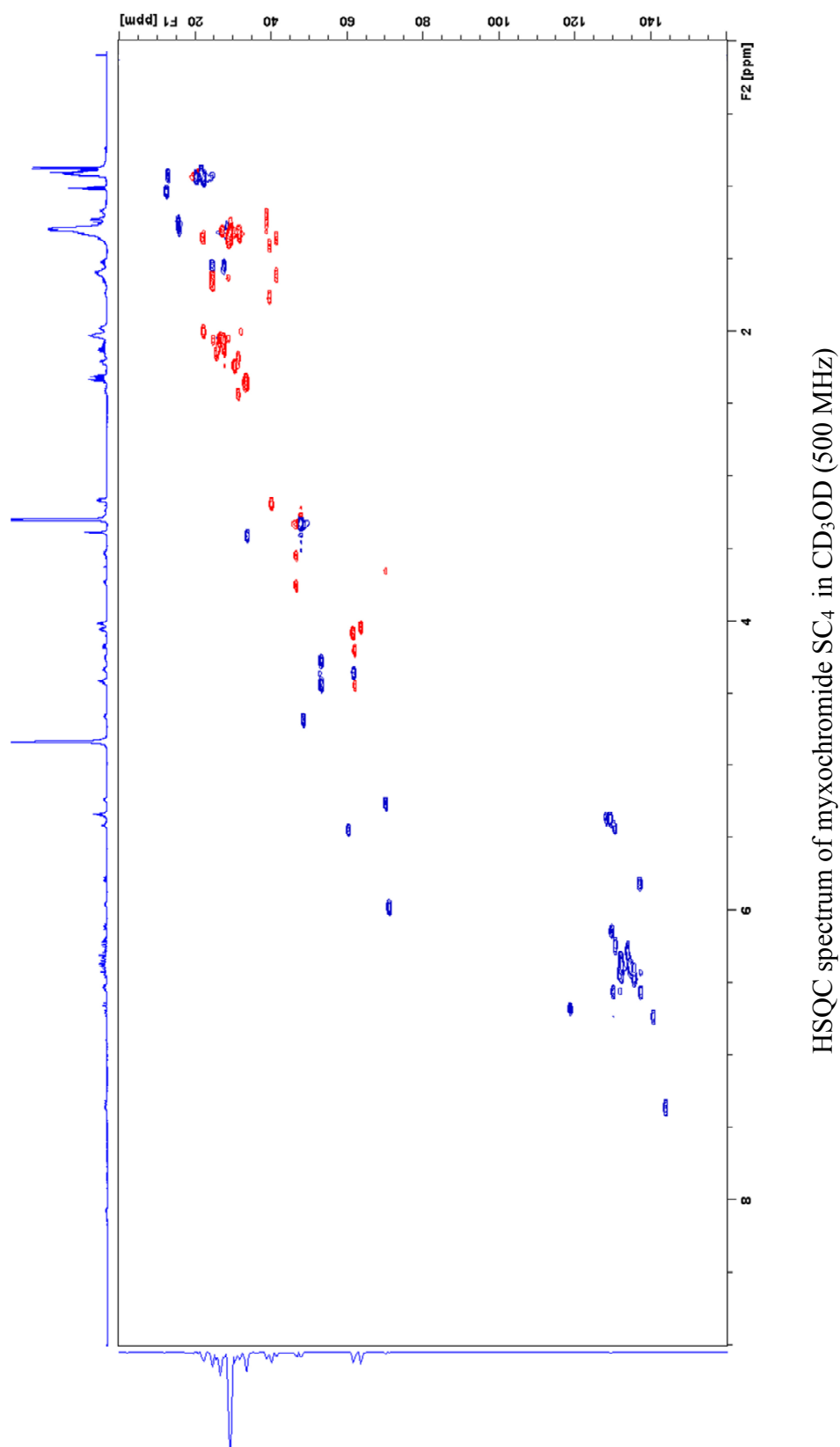


Figure S15 (continued on next page)

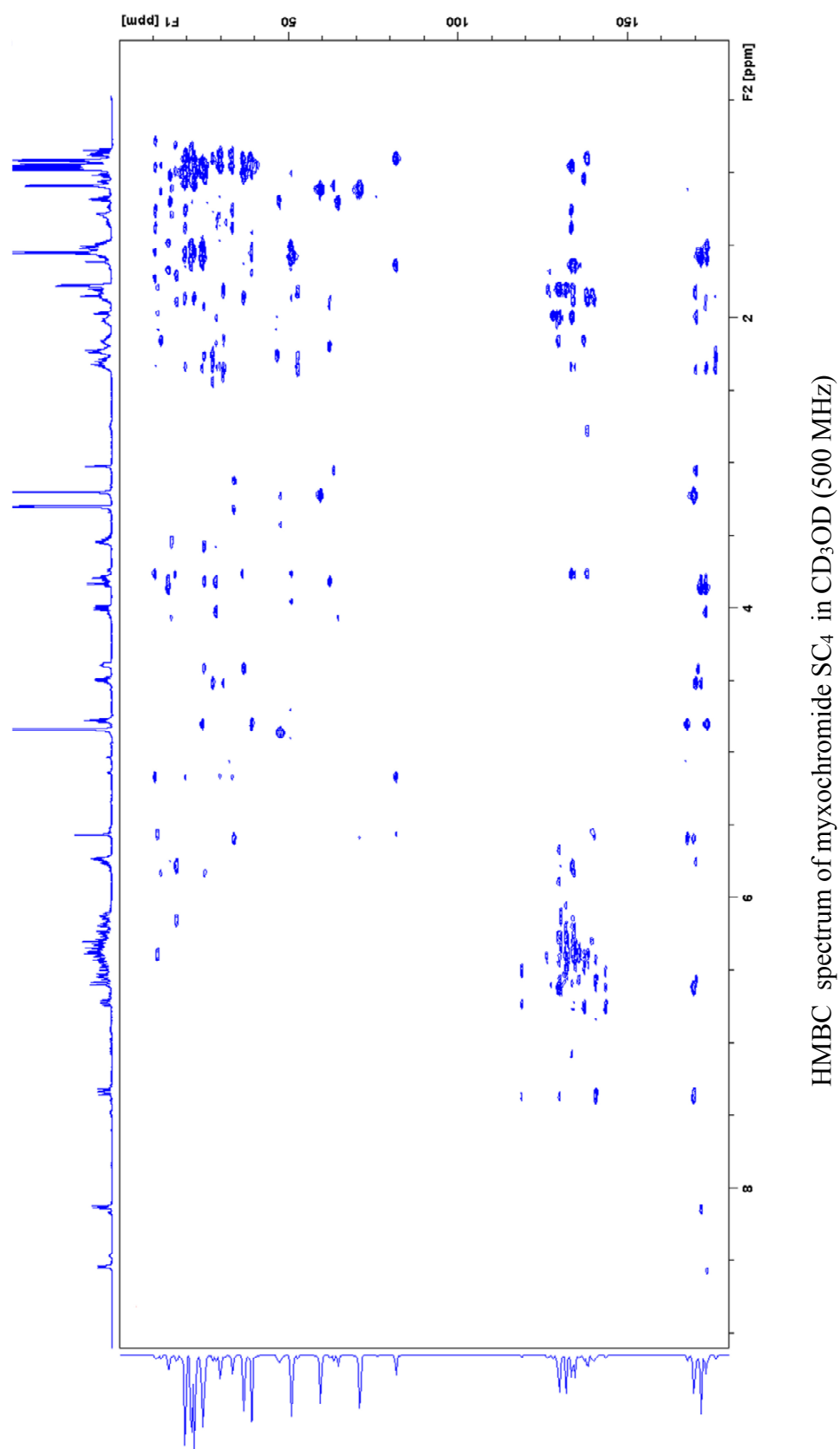


Figure S15 (continued on next page)

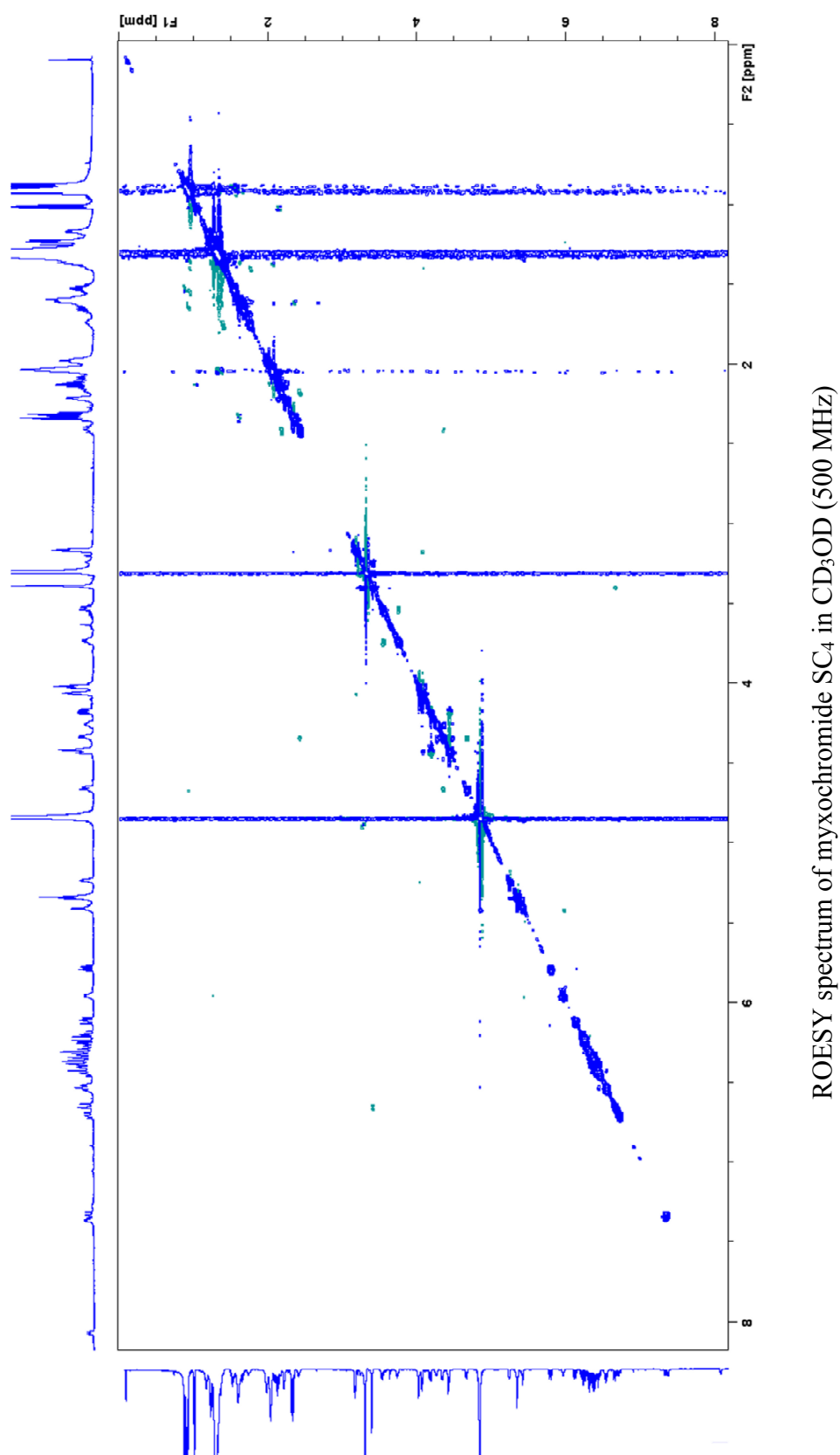


Figure S15. NMR spectra of myxochromide SC₄.

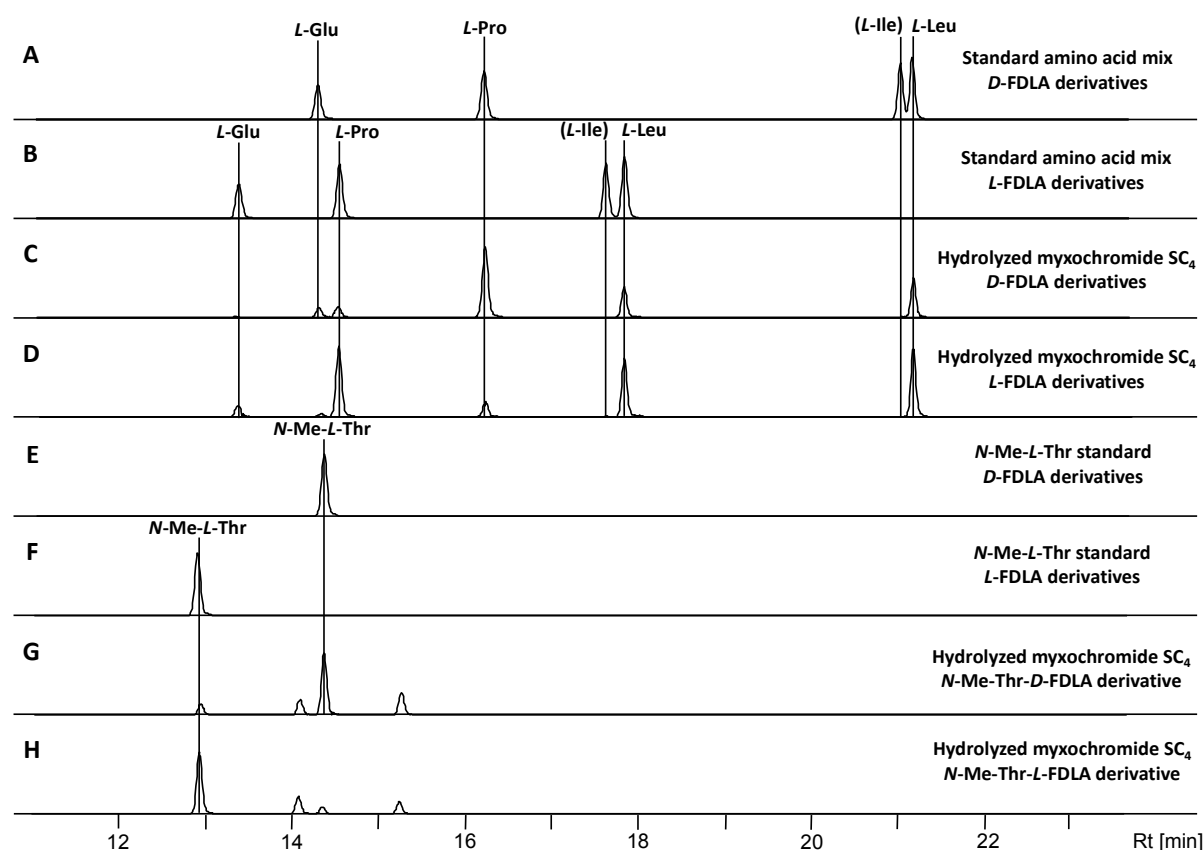


Figure S16. Results of the analysis of the absolute configuration by advanced Marfey's method.³² Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide SC₄ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide SC₄ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative

Table S21. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide SC₄.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	t_R [min]	m/z $[M+H]^+$	t_R [min]	m/z $[M+H]^+$	
Glu- <i>D</i> -FDLA	14.3	442.1578	14.3	442.1572	L
Glu- <i>L</i> -FDLA	13.3	442.1579	13.3	442.1572	
Pro- <i>D</i> -FDLA	16.2	410.1675	16.2	410.1679	L
Pro- <i>L</i> -FDLA	14.5	410.1675	14.5	410.1668	
Leu- <i>D</i> -FDLA	21.1	426.1989	21.1	426.1990	L
Leu- <i>L</i> -FDLA	17.8	426.1988	17.8	426.1982	
Leu- <i>D</i> -FDLA	21.1	426.1989	17.8	426.1981	D
Leu- <i>L</i> -FDLA	17.8	426.1988	21.1	426.1980	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.4	428.1782	14.4	428.1780	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.9	428.1786	12.9	428.1777	

Structure of myxochromide SD₃

Structure elucidation of myxochromide SD₃ was achieved using 1D ¹H and 2D ¹H-¹H COSY, HSQC and HMBC spectra (Figure S18). Carbon chemical shifts were extracted from 2D NMR data. NMR spectroscopic data are listed in Table S22. The COSY spectrum supported by HSQC and HMBC data revealed the presence of *N*-Me-threonine, glutamine, alanine and leucine residues as well as and a polyene side chain. Amino acid sequence was established by means of key HMBC correlations and final structure was elucidated as shown in Figure S17. For the assignment of the absolute configuration of myxochromide SD₃, hydrolysis and Marfey analysis of the obtained amino acids,³² was applied as described above. The chromatograms obtained from HPLC-MS analysis are illustrated in Figure S19 and stereochemical assignments are illustrated in Table S23. Comparison of the retention times and masses of derivatized standard amino acids and the hydrolyzed lipopeptide revealed that one of the two leucine residues (C5 and C11) from myxochromide SD₃ is *D*-configured. The second leucine residue as well as the amino acids alanine (C2), *N*-Me-threonine (C17) and glutamine (C21), which was converted to glutamic acid during hydrolysis, were found to be *L*-configured. This is in accordance with the absolute configurations observed in myxochromides SA₃, SB₄ and SC₄. According to the domain organization of the underlying hybrid assembly line, which harbors an epimerization domain in module 2, the *D*-configured leucine was assigned to C11. This also correlates with the structures of myxochromide SA₃ myxochromide SB₄ and identifies the condensation domain of module 3 originating from the *D*-type *mch* pathway as a ^DC_L domain.

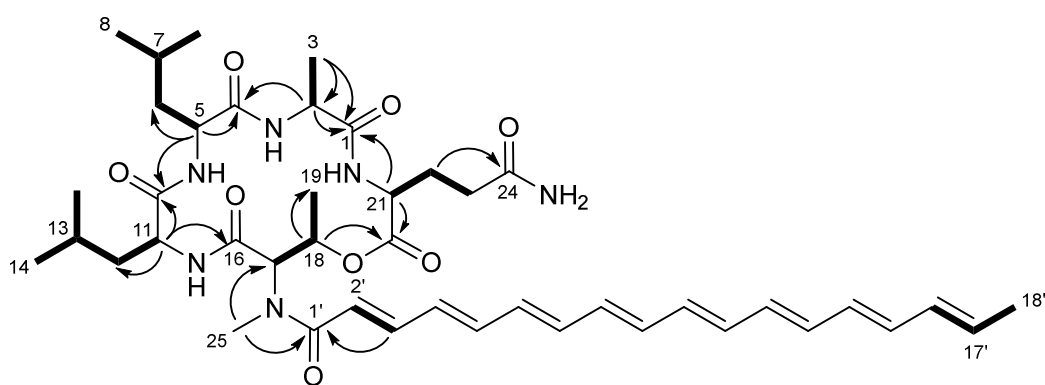


Figure S17. Structure of myxochromide SD₃ showing selected COSY (bold line) and key HMBC (arrow) correlations.

Table S22. NMR spectroscopic data of myxochromide SD₃.

Moiety	Position	δ_C^a	δ_H^b (J in Hz)	HMBC ^c
<i>L</i> -Ala	1	172.7		
	2	50.6	4.21 <i>m</i>	1,3,4
	3	18.7	1.37 <i>d</i> (6.7)	1,3
<i>L</i> -Leu	4	174.7		
	5	54.2	4.19 <i>m</i>	4,6a/b,7, 10 ^e
	6a	40.6	1.58 <i>m</i>	
	6b		1.65 <i>m</i>	
	7	25.9	1.71 <i>m</i>	
	8	23.2	0.99 <i>d</i> (6.3)	6,7,9
	9	21.1	0.91 <i>d</i> (6.5)	6,7,8
	10	175.2		
	11	53.1	4.39 <i>m</i>	10,12,13,16
<i>D</i> -Leu	12a	40.6	1.49 <i>m</i>	11
	12b	40.6	1.60 <i>m</i>	11
	13	25.7	1.56 <i>m</i>	
	14	22.6	0.92 <i>m</i>	12,13,15
	15	22.6	0.92 <i>m</i>	12,13,14
	16	169.8		
	17	59.5	5.43 <i>m</i>	16,18
	18	72.4	5.52 <i>m</i>	19,20
	19	17.2	1.15 <i>d</i> (6.9)	18,20
<i>N</i> -Me- <i>L</i> -Thr	25	35.1	3.02 <i>s</i>	1',17
	20	170.9		
	21	53.9	3.95 <i>m</i>	1,20,22a/b,23a/b
	22a	26.3	2.13 <i>m</i>	20,21,23a/b,24
	22b	26.3	2.25 <i>m</i>	20,21,23a/b,24
	23a	32.1	2.16 <i>m</i>	24
	23b	32.1	2.27 <i>m</i>	24
	24	177.7		
	24	177.7		
Side chain	1'	170.6		
	2'	120.1	6.59 <i>m</i>	1',4'
	3'	144.9	7.35 <i>m</i>	1'
	4'	131.5	6.53 <i>m</i>	
	5'-15'	<i>dd</i>	<i>dd</i>	
	16'	133.1	6.15 <i>m</i>	
	17'	131.1	5.75 <i>m</i>	
	18'	18.3	1.78 <i>d</i> (7.1)	17',16'

^a acquired at 125 MHz and assigned from 2D NMR spectra, referenced to solvent signal CD₃OD at δ 49.15 ppm.^b acquired at 500 MHz, referenced to solvent signal CD₃OD at δ 3.31 ppm.^c proton showing HMBC correlations to indicated carbons.^d overlapped signals.^e HMBC acquired with 2k F1 resolution

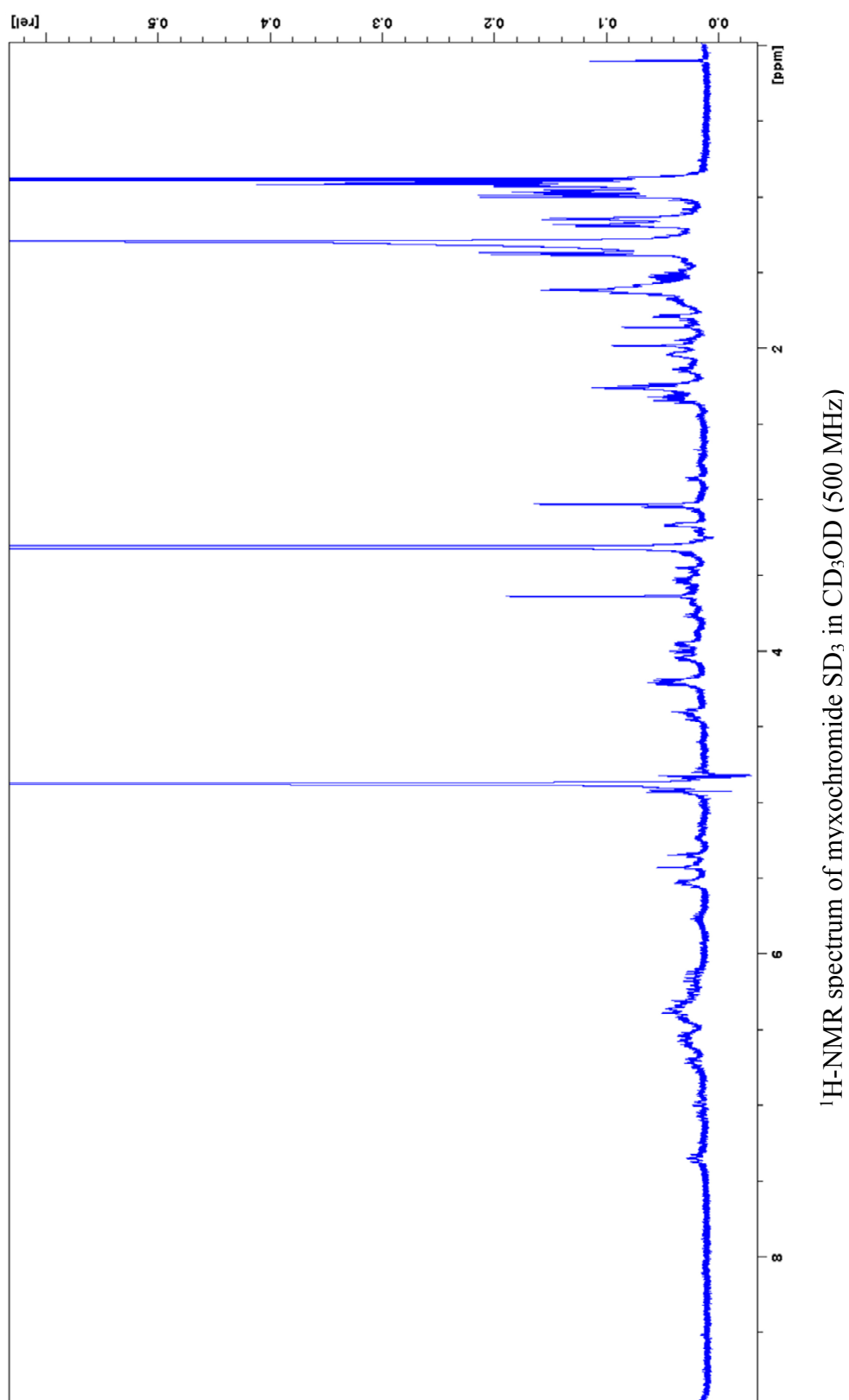
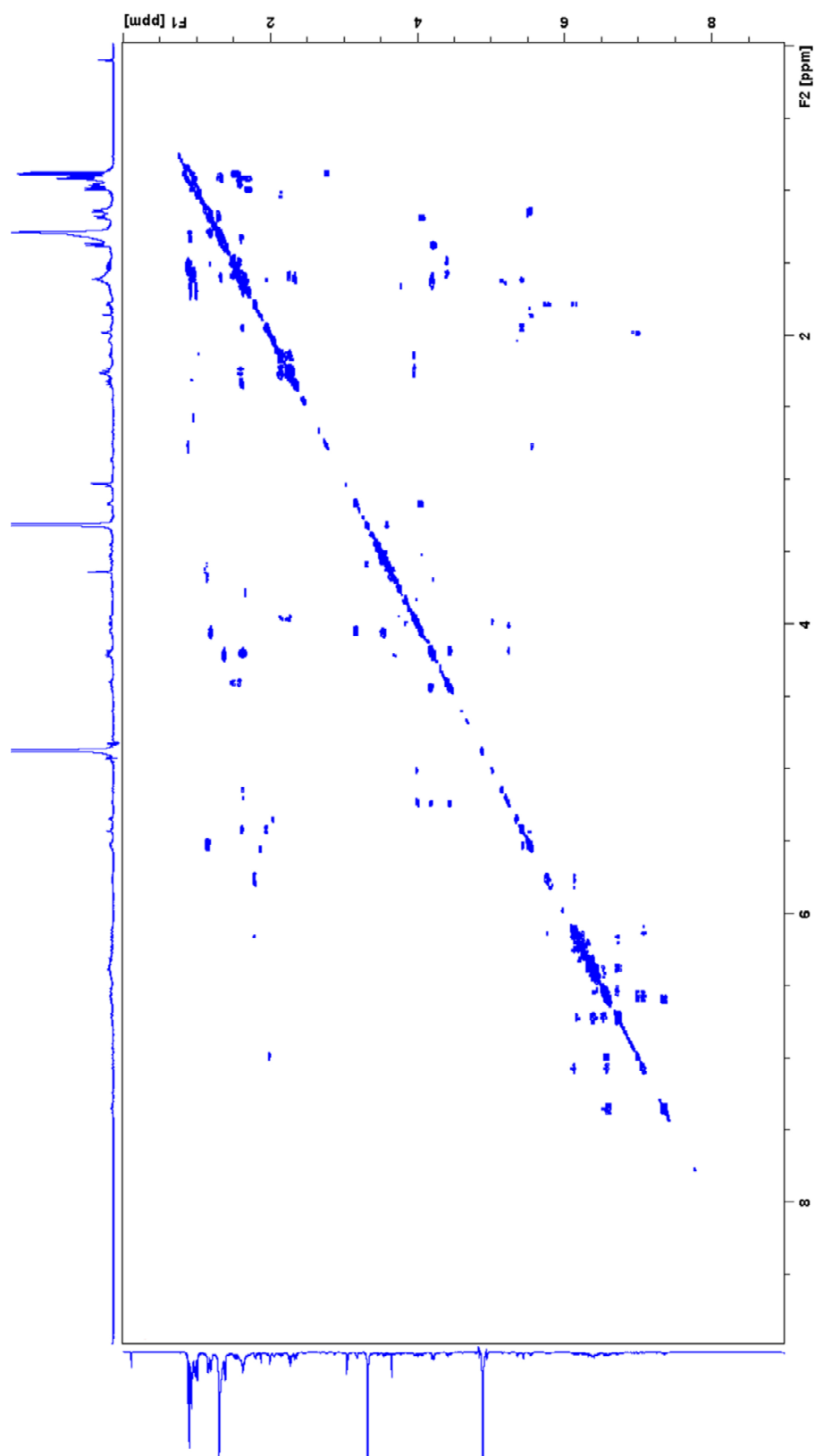
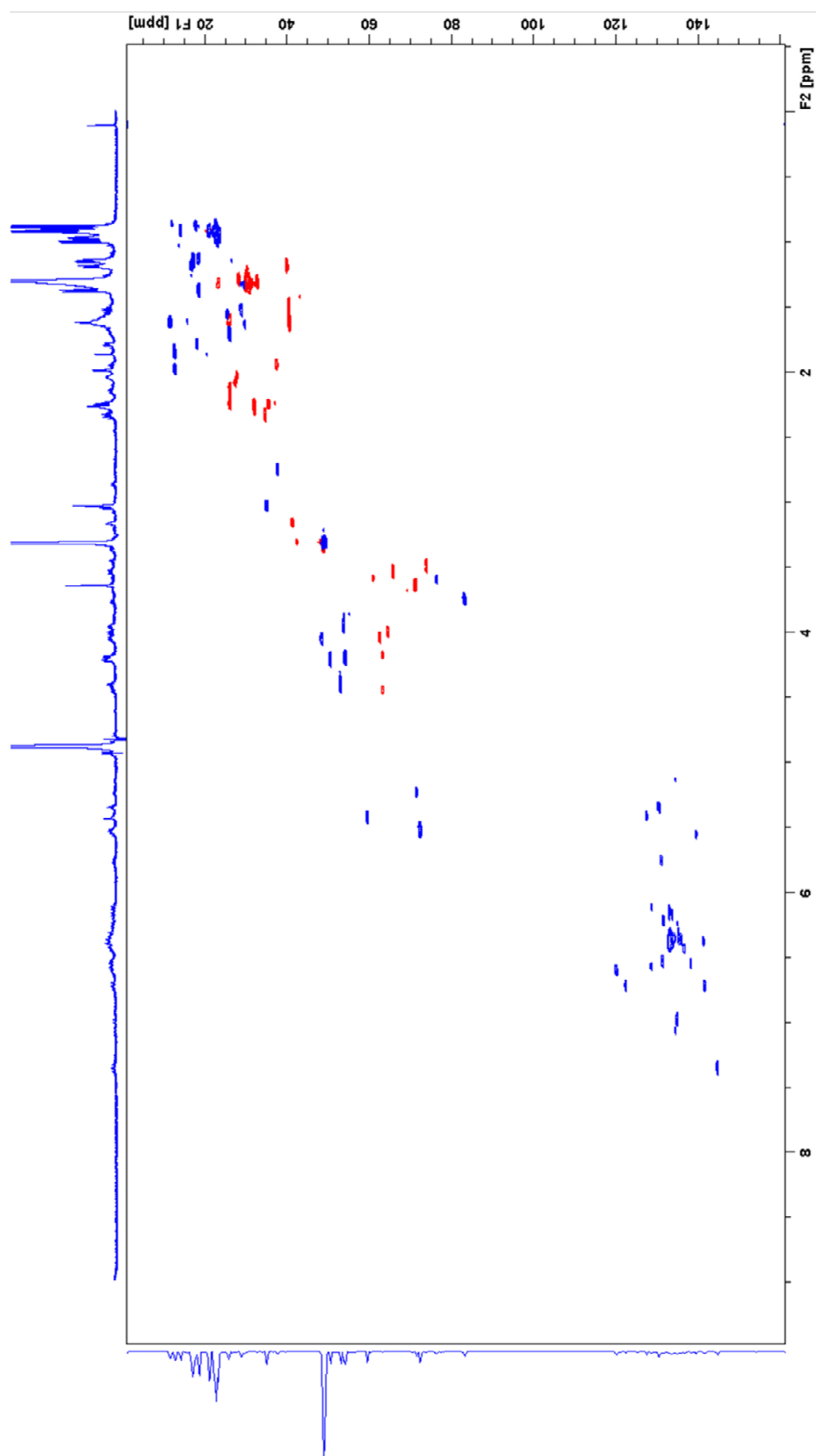


Figure S18 (continued on next page)



^1H - ^1H COSY spectrum of myxochromide SD₃ in CD₃OD (500 MHz)



HSQC spectrum of myxochromide SD₃ in CD₃OD (500 MHz)

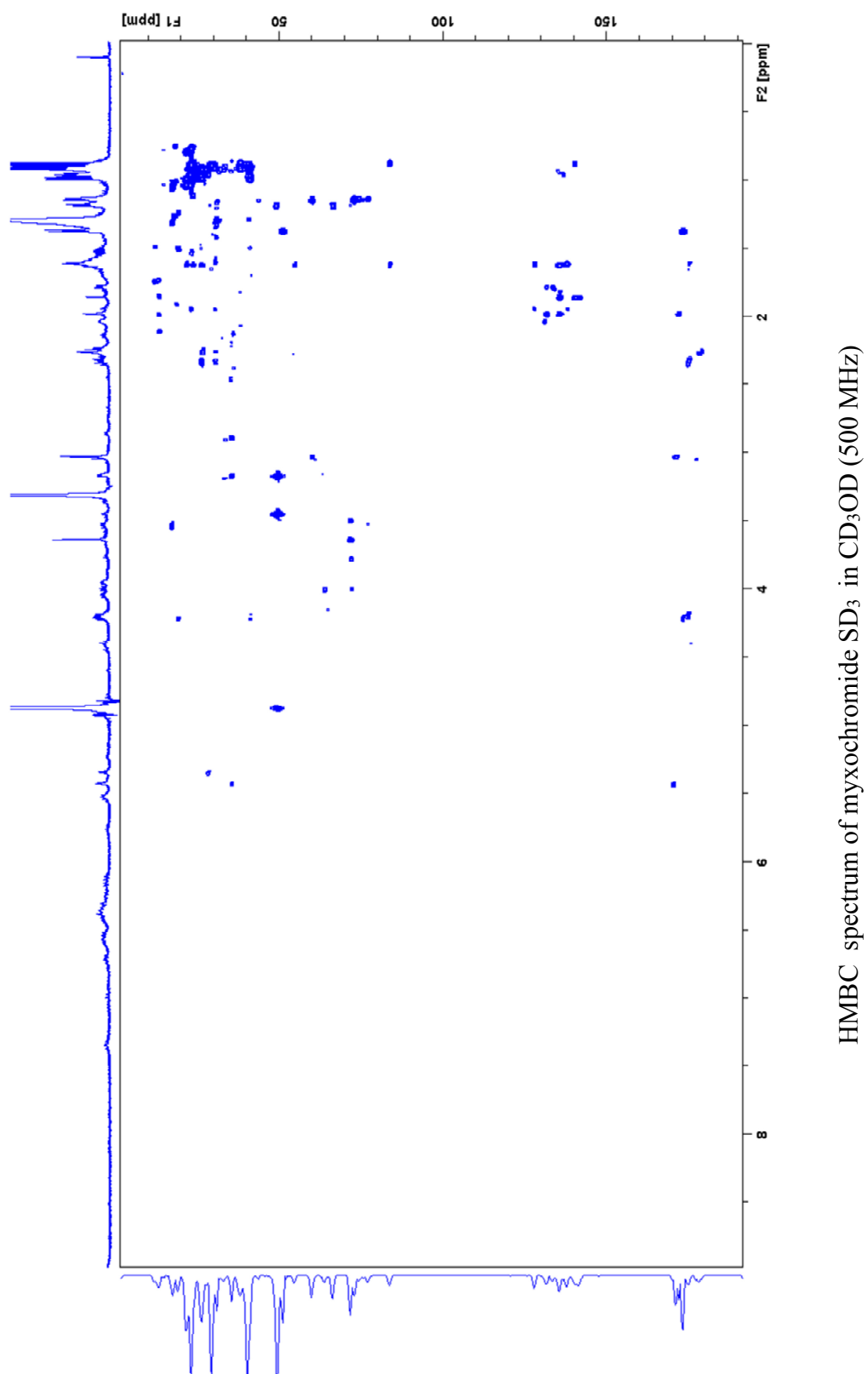


Figure S18. NMR spectra of myxochromide SD₃.

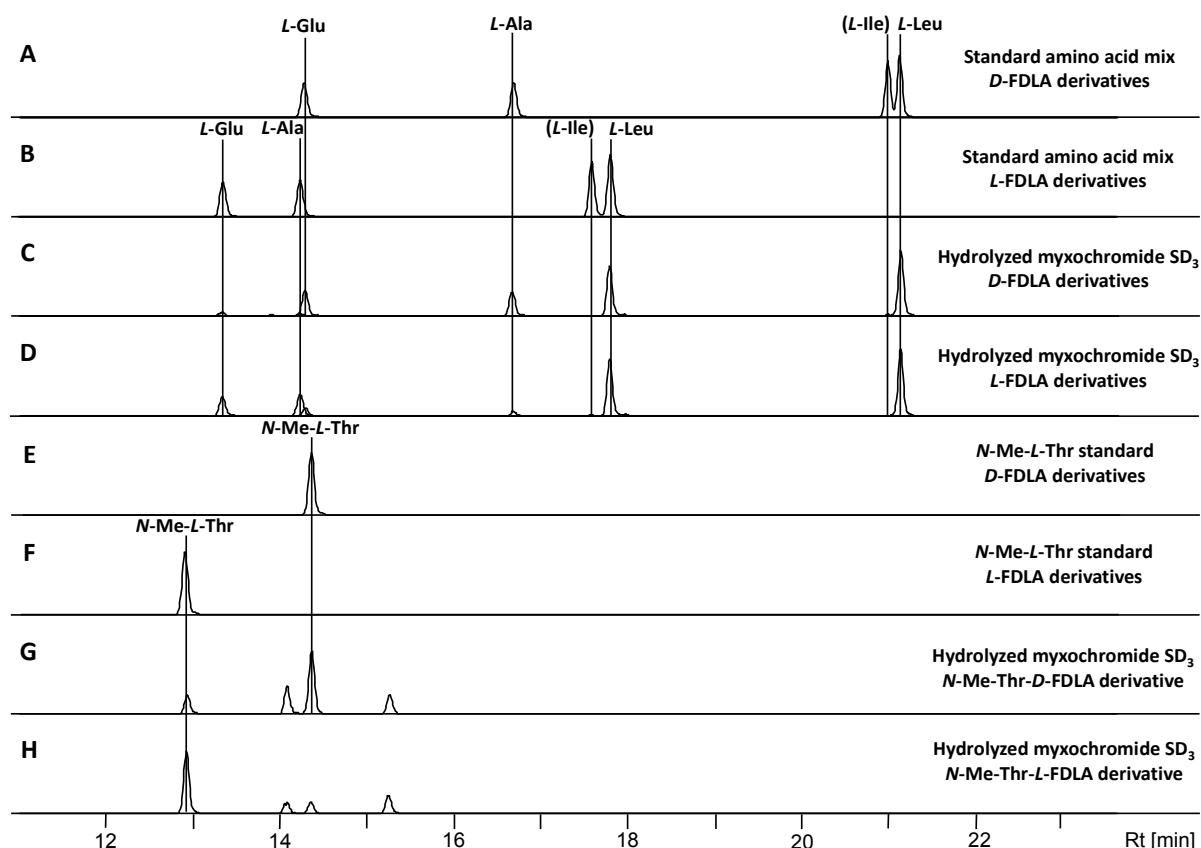


Figure S19. Results of the analysis of the absolute configuration by advanced Marfey's method.³² Extracted ion chromatograms (EIC) for ± 0.05 m/z corresponding to the $[M+H]^+$ ions of derivatized amino acids, which are present in the peptide scaffold, are shown. **A:** Standard amino acid mix derivatized with *D*-FDLA reagent. **B:** Standard amino acid mix derivatized with *L*-FDLA reagent. **C:** Hydrolyzed myxochromide SD₃ derivatized with *D*-FDLA reagent. **D:** Hydrolyzed myxochromide SD₃ derivatized with *L*-FDLA reagent. **E:** Standard solution of *N*-Me-*L*-threonine derivatized with *D*-FDLA. **F:** Standard solution of *N*-Me-*L*-threonine derivatized with *L*-FDLA. **G:** Same sample as in **C** analyzed for the *N*-Me-*L*-threonine *D*-FDLA derivative. **H:** Same sample as in **D** analyzed for the *N*-Me-*L*-threonine *L*-FDLA derivative.

Table S23. Analytical data of detected amino acid derivatives and assignment of the absolute configuration of the amino acids in myxochromide SD₃.

aa-FDLA derivative	<i>L</i> -aa standards		Peptide hydrolysate		Assigned configuration
	<i>t</i> _R [min]	<i>m/z</i> [M+H] ⁺	<i>t</i> _R [min]	<i>m/z</i> [M+H] ⁺	
Glu- <i>D</i> -FDLA	14.3	442.1578	14.3	442.1575	L
Glu- <i>L</i> -FDLA	13.3	442.1579	13.3	442.1564	
Ala- <i>D</i> -FDLA	16.7	384.1520	16.7	384.1512	L
Ala- <i>L</i> -FDLA	14.3	384.1524	14.3	384.1519	
Leu- <i>D</i> -FDLA	21.1	426.1989	21.1	426.1986	L
Leu- <i>L</i> -FDLA	17.8	426.1988	17.8	426.1986	
Leu- <i>D</i> -FDLA	21.1	426.1989	17.8	426.1982	D
Leu- <i>L</i> -FDLA	17.8	426.1988	21.1	426.1984	
<i>N</i> -Me-Thr- <i>D</i> -FDLA	14.4	428.1782	14.4	428.1773	L
<i>N</i> -Me-Thr- <i>L</i> -FDLA	12.9	428.1786	12.9	428.1784	

3.5 References

- (1) Weissman, K. J. *Nat. Prod. Rep.* **2016**, *33*, 203–230.
- (2) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. *Nat. Prod. Rep.* **2016**, *33*, 317–347.
- (3) Hertweck, C. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688–4716.
- (4) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275–287.
- (5) Weissman, K. J. *Nat. Chem. Biol.* **2015**, *11*, 660–670.
- (6) Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L. *et al. Nature* **2014**, *510*, 560–564.
- (7) Liu, Q.; Shen, Q.; Bian, X.; Chen, H.; Fu, J.; Wang, H.; Lei, P.; Guo, Z.; Chen, W.; Li, D. *et al. Sci. Rep.* **2016**, *6*, 34623.
- (8) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1846–1851.
- (9) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546–1550.
- (10) Nguyen, K. T.; Ritz, D.; Gu, J. Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17462–17467.
- (11) Baltz, R. H. *ACS Synth. Biol.* **2014**, *3*, 748–758.
- (12) Hopwood, D. A.; Malpartida, F.; Kieser, H. M.; Ikeda, H.; Duncan, J.; Fujii, I.; Rudd, B. A.; Floss, H. G.; Omura, S. *Nature* **1985**, *314*, 642–644.
- (13) Oßwald, C.; Zipf, G.; Schmidt, G.; Maier, J.; Bernauer, H. S.; Müller, R.; Wenzel, S. C. *ACS Synth. Biol.* **2014**, *3*, 759–772.
- (14) Kodumal, S. J.; Patel, K. G.; Reid, R.; Menzella, H. G.; Welch, M.; Santi, D. V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15573–15578.
- (15) Menzella, H. G.; Reisinger, S. J.; Welch, M.; Kealey, J. T.; Kennedy, J.; Reid, R.; Tran, C. Q.; Santi, D. V. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 22–28.
- (16) Mutka, S. C.; Carney, J. R.; Liu, Y.; Kennedy, J. *Biochemistry* **2006**, *45*, 1321–1330.
- (17) Burgard, C.; Zaburannyi, N.; Nadmid, S.; Maier, J.; Jenke-Kodama, H.; Luxenburger, E.; Bernauer, H. S.; Wenzel, S. C. *ACS Chem. Biol.* **2017**, *12*, 779–786.
- (18) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
- (19) Ongley, S.; Bian, X.; Neilan, B. A.; Müller, R. *Nat. Prod. Rep.* **2013**, *30*, 1121–1138.
- (20) Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. *Nucleic Acids Res.* **2008**, *36*, e113.
- (21) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. *ChemBioChem* **2005**, *6*, 375–385.
- (22) Pingoud, A.; Wilson, G. G.; Wende, W. *Nucleic Acids Res.* **2014**, *42*, 7489–7527.
- (23) Grigaite, R.; Maneliene, Z.; Janulaitis, A. *Nucleic Acids Res.* **2002**, *30*, e123.
- (24) Engler, C.; Gruetznier, R.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2009**, *4*, e5553.
- (25) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2008**, *3*, e3647.
- (26) Sanjana, N. E.; Le Cong; Zhou, Y.; Cunniff, M. M.; Feng, G.; Zhang, F. *Nat. Protoc.* **2012**, *7*, 171–192.
- (27) Cermak, T.; Starker, C. G.; Voytas, D. F. *Methods Mol. Biol.* **2015**, *1239*, 133–159.
- (28) Ohlendorf, B.; Kehraus, S.; König, G. M. *J. Nat. Prod.* **2008**, *71*, 1708–1713.
- (29) Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, *39*, 5775–5787.
- (30) Samel, S. A.; Czodrowski, P.; Essen, L.-O. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 1442–1452.
- (31) Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78–92.
- (32) Bhushan, R.; Bruckner, H. *Amino Acids* **2004**, *27*, 231–247.
- (33) Schlumbohm, W.; Stein, T.; Ullrich, C.; Vater, J.; Krause, M.; Marahiel, M. A.; Kruff, V.; Wittmann Liebold, B. *J. Biol. Chem.* **1991**, *266*, 23135–23141.
- (34) Stachelhaus, T.; Schneider, A.; Marahiel, M. A. *Biochem Pharmacol* **1996**, *52*, 177–186.
- (35) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. *Chem. Biol.* **1996**, *3*, 923–936.
- (36) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651–2674.
- (37) Koglin, A.; Walsh, C. T. *Nat. Prod. Rep.* **2009**, *26*, 987–1000.

- (38) Tanovic, A.; Samel, S. A.; Essen, L. O.; Marahiel, M. A. *Science* **2008**, *321*, 659–663.
- (39) Kraas, F. I.; Giessen, T. W.; Marahiel, M. A. *FEBS Lett.* **2012**, *586*, 283–288.
- (40) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15585–15590.
- (41) Lohman, J. R.; Ma, M.; Cuff, M. E.; Bigelow, L.; Bearden, J.; Babnigg, G.; Joachimiak, A.; Phillips, G. N., JR; Shen, B. *Proteins* **2014**, *82*, 1210–1218.
- (42) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Allen, C. L.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. *Nature* **2016**, *529*, 235–238.
- (43) Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G. *Nature* **2008**, *454*, 903–906.
- (44) Kears, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. *et al. Bioinformatics* **2012**, *28*, 1647–1649, DOI: 10.1093/bioinformatics/bts199.
- (45) Green, M. R.; Sambrook, J. *Molecular cloning: A laboratory manual*, 4th ed. / Michael R. Green, Joseph Sambrook; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2012.
- (46) Meyer, R.; Figurski, D.; Helinski, D. *Science* **1975**, *190*, 1226–1228, DOI: 10.1126/science.1060178.
- (47) Durland, R. H.; Toukdarian, A.; Fang, F.; Helinski, D. R. *J. Bacteriol.* **1990**, *172*, 3859–3867.
- (48) Thomas, C. M.; Meyer, R.; Helinski, D. R. *J. Bacteriol.* **1980**, *141*, 213–222.
- (49) Meyer, R. J.; Helinski, D. R. *Biochim. Biophys. Acta* **1977**, *478*, 109–113.

4 Discussion & Outlook

4.1 General Scope of the Present Work

The present studies described in this thesis deal with various aspects of myxobacterial natural product research ranging from secondary metabolite pathway identification, evolution and diversification to synthetic biology approaches in order to heterologously express artificial biosynthetic gene clusters based on synthetic DNA.

It could be demonstrated how genome-mining on a large scale can contribute to the in-depth analysis of the distribution and evolution of PKS/NRPS-derived lipopeptide pathways leading to numerous diversified gene cluster types responsible for the production of structurally different lipopeptide core structures. Furthermore, the observed structural differences could be rationalized on the basis of detailed comparative *in silico* sequence analyses, which additionally provided valuable insights into the evolutionary scenarios that might have led to the emergence of different but closely related lipopeptide pathways in various strains covering different genera of myxobacteria.

In addition, synthetic DNA platforms for the heterologous expression of myxobacterial lipopeptide pathways based on a dedicated gene library consisting of different lipopeptide biosynthetic genes from the identified lipopeptide gene clusters were established exhibiting a broad applicability. They provide the basis for the rational engineering of the underlying megasynthetases to produce hybrid lipopeptides, which do not occur naturally, thereby further increasing the structural diversity of this compound class. In the course of this study, a highly flexible assembly strategy for the construction of large artificial expression constructs harboring the synthetic gene clusters was established allowing the directed modification of the artificial pathways. Moreover, it was demonstrated that the established synthetic DNA platforms can also be used for the investigation of biosynthetic mechanisms and to evaluate *in silico* predictions regarding substrate specificities. Finally, synthetic biology approaches were used to mimic the observed mechanisms of pathway diversification, exemplarily demonstrated for the previously described ‘module-skipping’ process.

4.2 Bacterial Secondary Metabolite Pathways – Evolution and Diversification

Bacterial biosynthetic gene clusters are ideal genetic elements used as model systems to study gene evolution. The major prerequisite for comprehensive studies on pathway evolution and diversification is the availability of a robust genome database that can be screened for related natural product pathways using genome-mining tools.¹ As the costs for whole genome sequencing dramatically dropped over the past decades, it is now possible to decipher and

analyze bacterial genome sequences in a high-throughput manner using steadily improved state-of-the-art next generation sequencing platforms,² thereby revealing an outstanding degree of natural product pathway diversification and their widespread distribution and frequency among different bacterial taxa. This particularly accounts for the tendency of bacterial biosynthetic gene clusters to get transferred to other host microbes via horizontal gene transfer,^{3–5} thereby providing the possibility to further evolve in a different genomic context. In addition, microbial biosynthetic gene clusters evolve in relatively short time frames compared to genes from higher organisms, which results from shorter replication times of their host strains.⁶

Among the prokaryotic genomes sequenced so far, myxobacterial genomes still represent a minor fraction in common sequence repositories. However, the number of myxobacterial genomes exploded over the past years as a result of still reducing sequencing costs and the isolation of hundreds of novel strains. With the increasing number of available genome sequences, comprehensive studies on the evolutionary relationships of interesting myxobacterial natural product pathways, particularly polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) pathways, just begin to contribute to the still limited knowledge that currently exists on the evolution of microbial biosynthetic gene clusters.^{7–11} Moreover, the structural diversity of the produced natural products can be directly assessed and rationally explained on a genetic basis as biosynthetic gene clusters provide a clear link between genotype (biosynthetic gene cluster) and phenotype (natural product(s)). Common mechanisms by which biosynthetic gene clusters diverge into new pathway types such as point mutations, rearrangements, replacements, insertions, and deletions are generally known and have been previously described for very few examples based on phylogenetic data.^{10,12} However, these reports rarely gave detailed insights into how these evolutionary processes explicitly change the pathways on the sequence level, thereby lacking practical implications for biotechnological applications.

Studying natural products pathway evolution is not just a fascinating aspect for theoretical biologists as it has tremendous implications for the modification of existing pathways and, more ambitiously, for the tailor-made recreation of entirely new pathways from the scratch. Understanding the ‘evolutionary rules’ that direct natural pathway diversification would ultimately contribute to the realization of this long-term goal, thereby improving the chances to synthesize novel products with useful biological functions.¹³

In the following, the emergence of numerous myxobacterial lipopeptide pathways through diversification via various recombination events, such as module duplication and (partial)

module deletion, as well as via point mutations is discussed from an evolutionary point of view providing valuable information for the future engineering of NRPS megasynthetases.

4.2.1 Recombination Events Lead to Myxochromide Pathway Diversification

Screening of the available 122 myxobacterial genome sequences revealed 14 putative myxochromide gene clusters (*mch* clusters) in addition to the previously described A-type *mch* cluster from *M. xanthus* DK1622 and S-type *mch* cluster from *S. aurantiaca* DW4/3-1.^{14,15} Among those, four additional A-type and two S-type *mch* clusters were identified as well as the putative *mch* cluster from *Myxococcus* sp. 171, which was recently demonstrated to be a myxochromide B producer,¹⁶ was completely deciphered for the first time by a combination of genome sequencing and screening of a cosmid library of this strain, which was constructed during this thesis. The remaining 7 *mch* clusters represent three entirely new *mch* pathways, designated C-type, D-subtype 1 and D-subtype 2 *mch* clusters, which were predicted to encode the megasynthetases responsible for the biosynthesis of novel myxochromide cores using bioinformatics tools. *In silico* prediction of the expected lipopeptide products based on the acquired sequence data in conjunction with results from previous studies on myxochromide A and S biosynthesis suggested the production of novel lipopentapeptides, accordingly designated myxochromides C and D, which was experimentally verified with success. Sequence analysis of the identified B-type, C-type and D-type *mch* clusters revealed striking differences regarding the corresponding assembly line organizations, which resulted from homologous recombination events.¹⁷

The lipopeptide myxochromide B₃ has been previously detected and structurally characterized in a secondary metabolomics approach. Myxochromide B₃ harbors an additional leucine residue adjacent to the leucine residue present in myxochromides A. Whether this additional leucine residue in myxochromide B₃ is introduced as a result of a module duplication event or an iteratively acting leucine-specific NRPS module was not clear.¹⁶ The underlying biosynthetic pathway was not identified as the genome sequence of the producer strain *Myxococcus* sp. 171 was not available at that time. In this thesis, the biosynthetic pathway responsible for the biosynthesis of myxochromides B was completely established and indeed revealed a duplicated module in the MchC NRPS subunit, which was acquired via homologous recombination.¹⁷ Phylogenetic analysis revealed that the heptamodular myxochromide B megasynthetase contains a duplicate of the catalytic domains A₃, CP₃ and C₄ compared to other myxochromide assembly lines. Additionally, phylogenetic studies in conjunction with the analysis of local codon usage adaption along the catalytic domains

supports the hypothesis that the B-type *mch* cluster evolved from an ancestral A-type cluster by duplication of the A₃-CP₃-C₄ region. Intriguingly, detailed sequence analysis shed light on the exact recombination sites, which are located in the regions encoding the N-termini of A domains of module 3 and 4 of the MchC subunit near the regions encoding C-A domain interfaces. Consequently, an 'A-T-C' unit was duplicated instead of a dedicated C-A-T module.¹⁷

Since the distinct C, A and T domains, which make up an entire functional module, are usually highly homologous and thus share high sequence similarities, recombination events are assumed to take place regularly in the course of pathway evolution. In light of these results, homologous recombination in NRPS pathways might be a common strategy by which Nature employs pathway diversification. The first phylogenetic studies on selected PKS and NRPS systems revealed significant rates of homologous recombination and gene duplication.^{18–21} In particular, duplication events are assumed to play an essential role in the overall genesis of NRPS biosynthetic pathways. In many cases, individual 'A-T-C' units from a certain NRPS subunit group together as a monophyletic clade indicating that NRPSs could hypothetically arise from tandem duplication of a single ancestral module.¹¹ A fascinating example that supports this hypothesis is the family of related ferrichrome synthetases from various fungal species ranging from fission yeast, filamentous ascomycetes and basidiomycetes. Phylogenetic analysis of the ferrichrome NRPSs suggested that these biosynthetic machineries derive from an ancestral gene encoding a hexamodular NRPS, which eventually evolved independently via additional recombination events and that the hexamodular NRPS is most likely created by tandem duplication of 'A-T-C' units.¹¹ A similar scenario seen in ferrichrome synthetases is described for the 49 kb gene encoding the yet uncharacterized NRPS Plu2670 from *Phototrhobdus luminescens* from which module encoding regions group into five clades exhibiting high sequence similarity of > 85%.²² The same conclusion is drawn from phylogenetic analyses of the multimodular PKS responsible for the production of mycolactone in *Mycobacterium ulcerans*. This PKS machinery consists of the three individual subunits MLSA1, MLSA2 and MLSB and modules of these subunits also group into monophyletic clades with sequence identity of > 98% suggesting the role of an ancestral gene for the generation of this large biosynthetic complex via intragenic module duplication.²³ A recent phylogenetic study on the evolution of polyketide structural diversity in the genus *Streptomyces* by Jenke-Kodama *et al.* also underpins the role of module duplications in natural product pathway evolution. They showed that locations of the corresponding recombination sites are not restricted to interdomain linkers, but can be also

located in homologous stretches of the corresponding domains.¹⁹ Nevertheless, recombination sites may be confined to regions that account for the structural diversity of the produced compounds. Although the duplication of 'A-T-C' units in NRPS megasynthetases seems to be quite common in the '*de novo*' generation of NRPS pathways, an intragenic module duplication event in an already 'established' functional bacterial NRPS assembly line was never described before in such detail. Thus, the observed module duplication event leading to the myxochromide B pathway builds the basis for developing concepts for NRPS pathway diversification via intragenic module duplication and for the rational modification of the underlying pathways via engineered module duplications. It might be possible to engineer artificial C-A interdomain linker sequences at different positions of a gene cluster in a way that they contain the corresponding regions (N-termini of A_x/A_{x+1} domain interfaces, x = module number) of the C-A domain interface as observed in the B-type *mch* cluster, thereby facilitating functional coupling of noncognate modules.

In addition to recombination-based duplication of biosynthetic gene cluster regions, deletions of domain or module encoding regions were also detected in a few bacterial pathways.^{24–26} Recent studies on the evolutionary roots and genetic distribution of cyanobacterial toxin pathways revealed that many toxin producer strains acquired the corresponding pathways responsible for the production of several hepatotoxins via horizontal gene transfer, mostly mediated by transposases associated with these pathways. This automatically led to the emergence of non-toxic strains, in which large portions (up to 90%) of the underlying pathways were deleted leading to the inability to produce these toxins through inactivation of the gene clusters.⁷ In analogy to the duplication events discussed above, deletion of defined parts of a biosynthetic gene cluster, like functional modules, might result from the highly homologous nucleotide sequences the distinct catalytic NRPS domains share. One of the most popular and very few examples, for which NRPS gene cluster evolution was demonstrated on a broad scale, are these cyanobacterial toxins, particularly the pathways responsible for the biosynthesis of the microcystins from *Microcystis* spp. and related compounds.⁷ It was specifically shown that a recombination-based deletion event that even comprises two NRPS modules across the microcystin synthetase subunits McyA and McyB, led to the emergence of the closely related and functional nodularin biosynthetic pathway occurring in *Nodularia spumigena*.²⁴ Consequently, the cyclic nodularin peptide core exhibits a reduced ring size, thereby lacking the two amino acid residues, which would have been introduced in the core structure by the two modules present in the microcystin assembly line (Figure 1). In this case, the recombination sites were found to be located within the C domains of module 2 of McyA

and module 4 of McyB, regions that exhibit a high degree of sequence homology. Phylogenetic analysis involving the biosynthetic genes as well as selected genes associated with characterized microcystin and nodularin pathways revealed that the nodularin biosynthetic gene cluster might indeed originate from the microcystin pathway that might be the ancient ancestor from which the known related cyanobacterial toxin pathways have been evolved.²⁴

Additionally, module deletion also seems to play a role in the diversification of bacterial multimodular PKS pathways. The pathways responsible for the production of the polyketides spinosyn and butenyl-spinosyn in different *Saccharopolyspora* spp. also differ in the presence or absence of a module containing the catalytic domains KS-AT-DH-KR-ACP. Regarding the present enzymatic functions, this module might be responsible for this unique addition of the butenyl moiety. The authors stated that the butenyl-spinosyn pathway might be the common ancestor from which the spinosyn pathway may have evolved, although they did not give any additional details on phylogenetic studies and the exact recombination sites.²⁶

In the myxochromide C assembly line, module deletion occurred during pathway evolution leading to the loss of an 'A-T-C' unit between modules 5 and 6 (A₅-T₅-C₆), thereby generating a recombined and functional C₅-A₆-T₆-TE termination module. Similar to the module duplication process observed in myxochromide B biosynthesis, the recombination sites are located near regions encoding C-A domain interfaces, more precisely at the C-termini of the corresponding C₅/C₆ domains. Phylogenetic studies show overall congruence between strain and gene cluster phylogeny, which suggests that the C-type *mch* cluster also evolved from the ancestral A-type cluster by deletion of the A₅-CP₅-C₆ region.¹⁷

Partial module deletion of an 'A₄-CP₄' unit was observed in the myxochromide D subtype 2 assembly line leading to the production of a pentapeptide core lacking the proline residue, which was 'encoded' by the corresponding deleted region of module 4. Interestingly, the resulting assembly line harbors a C₄/C₅ interface after the deletion event, which does obviously not interfere with its functionality. The recombination sites were found to be located within the C₄-A₄ and CP₄-C₅ interdomain linkers, respectively, thereby conserving only parts of the original C₄-A₄ domain interface in the naturally occurring hybrid assembly line (Figure 1). Analysis of the local codon usage in all *mch* clusters revealed less codon usage adaptation in this inactive A₄-CP₄ region in all *Cystobacterineae* strains compared with the following A₅-CP₅ or preceding A₃-CP₃ region, which suggests that this region might be a kind of hotspot for assembly line modifications.

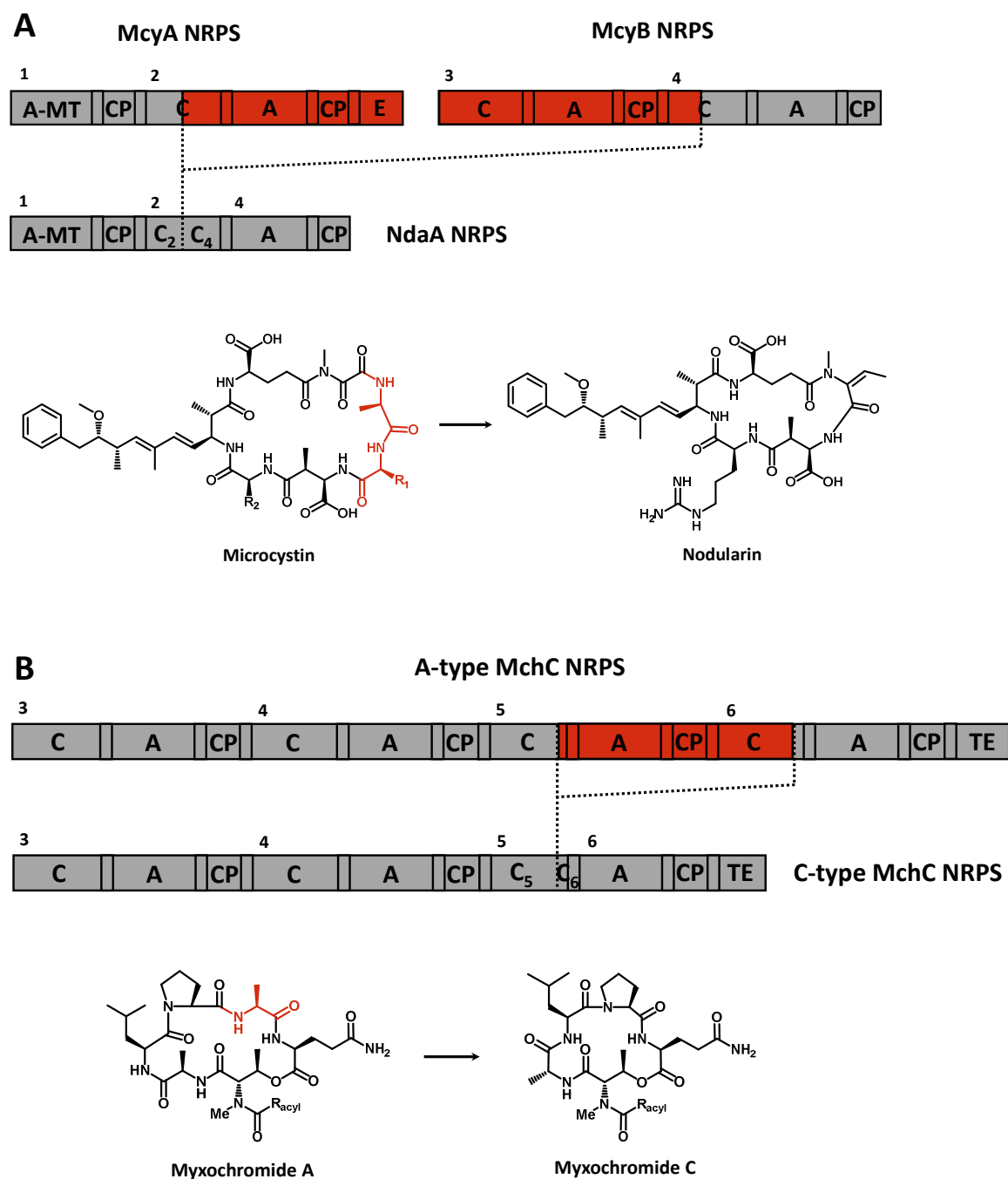


Figure 1. Module duplication events in *M. virescens* ST200611 and *N. spumigena* NSOR10. Emergence of the myxochromide C and nodularin (PKS)/NRPS assembly lines, which evolved from their ancestors myxochromide A and microcystin synthetase, respectively, resulted in structural diversification of the produced natural product families. R₁/R₂, variable amino acid residues; R_{acyl}, acyl side chains. **A:** Multiple module deletion in nodularin biosynthesis. **B:** Module deletion in myxochromide C biosynthesis.

This is further underpinned by additional mutational activities around the recombination sites, which possibly point to ongoing diversification of the resulting *mch* pathway.¹⁷ In light of the results obtained from the module deletion event in the myxochromide C pathway, in which a complete ‘A-T-C’ unit was deleted, the deletion of an ‘A-T’ unit could possibly indicate an evolutionary transition state, from which the remaining C domain region may be deleted in a

following second recombination event. Single domain deletions were also recently described for microcystin pathways, in which the *N*-methyltransferase (NMT) domain of the McyA subunit was removed yielding functional NRPS assembly lines capable of producing non-methylated microcystins.²⁷

In the context of NRPS assembly line enzymology, the observed recombination sites in the *mch* clusters near regions encoding C-A domain interfaces or within C-A linker regions can be interpreted on the basis of the general role of interdomain and intermodule linkers in NRPS megasynthetases. These linkers are assumed to function in the communication between catalytic domains within and between adjacent modules by direct interaction with the interconnected domains. Several studies on the rational engineering of NRPS systems impressively showed that most targeted module and domain exchanges, in which the native interdomain/intermodule linkers are not maintained, lead to non-functional hybrid pathways, thereby highlighting the general importance of these linker regions.^{28–30} In addition, a recent crystal structure of an entire C-A-T-TE termination module from the surfactin NRPS megasynthetase published by Tanovic *et al.*,³¹ revealed that the C-A interdomain linker seemed to have a particularly important function by serving as a structurally rigid platform that remains invariant during chain elongation. This rigidity arises from extensive interactions of the linker region with both C and A domains of this module. In contrast, the linker regions between the other catalytic domains (A-PCP and PCP-TE) are much shorter, thereby making fewer interactions with the connected domains, which is attributed to their high flexibility to allow for the required movements of the rotating PCP and TE domains during catalysis.³² In the present cases, in which module duplication (B-type *mch* cluster) and deletion (C-type and D-type *mch* clusters) arise from recombination sites around these regions, the C-A domain interfaces remain largely conserved, which might be an evolutionary strategy in order to maintain the native linker architecture and to increase the chances to yield a functional evolved pathway in the course of gene cluster diversification. However, structural details on the underlying NRPS enzymatic machineries are still limited and even less knowledge is available on the dynamical interplay between modules and domains within the megasynthetase complexes. In addition, there are only very few examples investigating biosynthetic pathway evolution, which are based on a sufficiently high number of gene cluster sequences, thereby hampering the generalization of the obtained results. Whether the C-A linker regions generally play a protruding role as dedicated recombination spots in NRPS pathway diversification remains uncertain and awaits comparable analyses of other natural product pathway families. It is more likely that biosynthetic pathways diversify by following

specific recombination rules restricted to a family of secondary metabolites as suggested by a recent study on megasynthetase evolution.¹²

4.2.2 ‘Module-Skipping’ Lead to Myxochromide Pathway Diversification

Another strategy in secondary metabolite pathway diversification is ‘module-skipping’, a process that was already demonstrated to happen in some PKS systems.^{33–36} This known deviation from textbook biosynthetic logic eventually results in the formation of polyketides lacking an extender unit. ‘Module-skipping’ in NRPS assembly lines is much rarer. In fact, only one ‘non-linear’ NRPS system has been described so far.¹⁴ Although the myxochromide S pathway from *S. aurantica* DW4/3-1 represents a hexamodular assembly line, the proline activating module is skipped during biosynthesis leading to the production of a lipopentapeptide core lacking the proline residue. In the reported case, it could be concluded that ‘inactivation’ of this module arises from point mutations in the GGHSL core motif of the corresponding PCP, in which the conserved serine residue that is required for posttranslational activation of the PCP was replaced by a proline, among additional mutations (GGHSL → GGNPS).¹⁴ However, as additional examples for ‘module-skipping’ in other NRPS megasynthetases did not appear, it was not clear whether this process represented a random event leading to a functional variant of the ancestral pathway by chance, or it is a mutation-driven diversification path. As the genome-mining approach used during this thesis enabled the identification of numerous additional myxochromide biosynthetic gene clusters, it was possible to successfully assess the rational basis of this process by comparative sequence analyses.¹⁷ In addition to the myxochromide S pathway from *S. aurantica* DW4/3-1, S-type *mch* clusters were also found in several other *S. aurantica* strains, which exhibit the reported serine to proline ‘loss of function’ mutation in the same (proline incorporating) module suggesting that ‘module-skipping’ leads to structural diversity in a programmed manner. Interestingly, the same mutation was also detected in PCP domains of the orthologous modules in the identified D-type *mch* clusters from distant *Hyalangium* sp. as well as from closely related *Stigmatella erecta* Pde77 and results in ‘module-skipping’ as well. As these gene cluster variants independently evolved and were maintained in the course of pathway diversification, these results indicate that ‘module-skipping’ indeed directly contributes to the expansion of myxochromide structural diversity.¹⁷ Since this kind of mutation-induced diversification mechanism was not found in completely different NRPS systems from other bacteria so far, it can be assumed that it may reflect a family-specific mechanism in the evolutionary history of myxochromide pathways. However, Wenzel *et al.* hypothesized that

changing the serine residue from PCPs via site-directed mutagenesis might serve as a tool to induce ‘module-skipping’ at other positions in the *mch* clusters or possibly also in other NRPS systems in order to further expand structural diversity.¹⁴ In case of the S-type *mch* pathways, they assumed that restoration of the conserved serine residue in the PCP core motif might possibly reactivate the affected module leading to incorporation of the corresponding proline residue in the peptide core.¹⁴ These hypotheses were specifically addressed in this thesis using synthetic biology approaches and are further discussed in the last section of the discussion.

4.2.3 Concluding Remarks

As the cost of high-throughput DNA sequencing continues to decrease, more and more microbial genomes and metagenomes will be publically available for the identification of PKS/NRPS-derived secondary metabolite pathways and their corresponding natural products via genome-mining. However, comprehensive *in silico* screening specific for functionally similar or closely related families of biosynthetic gene clusters might accelerate the translation of DNA sequences into biotechnological applications rather than searching for a certain secondary metabolite that is linked to its corresponding pathway. In light of the diverse mechanisms by which biosynthetic gene clusters extensively undergo pathway evolution and diversification in bacteria,^{10,12,13} one can assume that to most ancient gene clusters several related pathway variants might exist, leading to the production of structurally different molecules even among different bacterial genera. This especially accounts for the unique biological functions the produced secondary metabolite derivatives have in their natural environment as a result of constant selective pressure that is on the underlying pathways.³⁷ Biosynthetic gene clusters spread laterally via horizontal gene transfer resulting in novel microbial strains that are capable of producing the corresponding compounds, thereby subjecting the natural product pathways to further evolutionary pressure in a different genomic context.³⁸ By implementation of sophisticated sequence analysis methods based on phylogenies on the domain, module, subunit and gene cluster level as well as on local codon usage and adaption in conjunction with the modern analytical techniques to the general genome-mining workflow, the importance of genome-mining is no longer restricted to the discovery of novel natural products. It also provides a broad picture regarding the interrelationships and distribution of biosynthetic gene clusters, which evolved independently from each other. The present work provides a guideline how these approaches can be combined in order to understand the rules of natural evolution, which might have a direct

impact on the rational engineering of PKS/NRPS systems. The recombination sites, which have been detected in different *mch* clusters around the C-A interdomain linkers, can now be taken into account when engineering module duplications and deletions at other positions, e.g. in the A-type *mch* cluster to generate novel hybrid assembly lines based on protein sequence alignments. However, interdomain linker-driven recombinations may be a family-specific recombination rule and may not be applicable to other NRPS assembly lines. Since there are not that many related bacterial NRPS pathways published, which were characterized on such a broad basis, special emphasis should be on the identification and analysis of already known as well as completely novel pathway families to reevaluate the observed recombination events regarding their general implications for NRPS pathway diversification. Eventually, the present genome-mining approach builds the basis for the identification and detailed analysis of related natural product families to assess their evolutionary relationships and based on that, to deduce common rules for pathway evolution in order to significantly improve future engineering efforts and to make new products via synthetic biology approaches.

4.3 Synthetic Expression Platforms to Produce Myxobacterial Natural Products

Significant advances in the field of DNA synthesis have recently led to new biotechnological applications, which were largely proven to be difficult to perform on native DNA. One important example with respect to natural product research is the transfer and functional expression of secondary metabolite pathways in heterologous hosts. As heterologous expression of biosynthetic gene clusters is *per se* a challenging task encompassing numerous requirements to be considered, synthetic biology approaches are expected to significantly contribute to address the challenges a natural product researcher is confronted with. In the present thesis, synthetic expression platforms have been established allowing the heterologous expression of myxochromide pathways based on synthetic DNA in a myxobacterial host strain. The presented strategy is characterized by a high flexibility and broad applicability towards pathway engineering and has been further exploited to produce ‘unnatural’ myxochromide cores via combinatorial biosynthesis and to study basic biosynthetic processes in myxochromide pathways. The work described here demonstrates the power of synthetic biology approaches for natural product research and builds the basis for far-reaching and systematic investigations on the general DNA sequence design to optimize production yields and to provide simplified, ready-to-use cell factories for the production of novel molecules.

4.3.1 Heterologous Expression of Myxobacterial Biosynthetic Gene Clusters

Myxobacteria have been recognized as proficient producers of bioactive natural products, which predominantly derive from PKS and NRPS pathways and a significant number of promising lead compounds recently entered clinic trials.³⁹ The high relevance of myxobacteria as a rich source of potential drug leads is however faced with several drawbacks as most myxobacteria are slow growing and difficult to cultivate or even uncultivable under standard laboratory conditions. In addition, the number of genetic tools for the manipulation of the native producer strains is quite limited compared e.g. to the intensively studied streptomycetes. Thus, the development of genetic tools to transfer and efficiently express interesting myxobacterial pathways in alternative hosts, which are easy to handle and manipulate (e.g. streptomycetes, pseudomonads, bacilli and *Escherichia coli*) was one of the main achievements in the myxobacterial field over the past two decades and many approaches for the heterologous expression of PKS/NRPS biosynthetic pathways have been established.^{40,41} These include the direct transfer of relatively small gene clusters mobilized in cosmids or BACs into related hosts, pathway modification and expression in unrelated bacteria and the co-expression of several gene cluster containing vector systems. The latter approach is referred to as multiplasmid approach, which was particularly used for huge biosynthetic gene clusters in the past considering the fact that only a small number of methods existed at that time for the construction of large DNA constructs on one physical entity such as the Red/ET approach.^{42,43} Direct transfer and expression of cosmids or BAC-derived constructs harboring the biosynthetic pathways requires suitable vector systems such as replicative plasmids. Unfortunately, replicative plasmids for any myxobacterial species are still not available leaving integration into the chromosome via homologous recombination, transposition or phage-derived systems the only reliable way to transfer foreign genes into a myxobacterial heterologous host.^{40,41} Most of the strategies are restricted to small biosynthetic gene clusters and are characterized by significant time-consuming cloning efforts, especially if heterologous expression focuses on large biosynthetic pathways. Recent advances in recombinant DNA technologies and assembly strategies led to the addition of valuable cloning methods either based on homologous recombination,^{44–46} or restriction/ligation procedures.^{47,48} The methods generally allow the assembly of sizable gene cluster constructs and are applicable to high-throughput set-ups.

However, the assembly and heterologous expression of natural product biosynthetic pathways is not as simple and straightforward as it seems. Efficient expression of target biosynthetic pathways in a surrogate host requires in depth knowledge on several factors that influence functionality of the heterologous system such as the ability of the surrogate host to

posttranslationally activate the PKS and NRPS megasynthetases, the time-coordinated supply of biosynthetic precursors in sufficient amounts, the functionality of regulatory elements and native promoters, the stability of the transcribed mRNA, and the self-resistance of the host strain to the expressed secondary metabolites.^{40,41} Moreover, it should be pointed out that expression of huge biosynthetic pathways is a strong metabolic burden for the producer strain considering the complexity of these biosynthetic machineries. Technically, the selected heterologous host should exhibit excellent growth characteristics (short doubling times, high cell density) and a plethora of experimental tools to genetically modify the host strain should exist.^{40,41} Over the past decades, it was recognized that the efficient heterologous expression of PKS/NRPS biosynthetic pathways is often beneficial when transferring the pathways into closely related host organisms, e.g. from one myxobacterial species into another, which is attributed to a similar codon usage of the target pathway and the genome of the host strain.⁴⁹ However, many of the mentioned factors remain largely unknown, so it is not surprising that many efforts to heterologously express biosynthetic pathways often yield insufficient amounts of the target compounds or even fail completely.

In summary, many new technological innovations in the fields of DNA cloning and engineering contributed to the increasing impact of heterologous expression platforms in natural products research over the past decades.^{40,41} However, as only a limited set of suitable, well-characterized heterologous hosts are available today and most of the described classical methods still require laborious procedures relying on the mobilization of the genetic blueprint responsible for the biosynthesis of a certain secondary metabolite, it is highly desirable to develop alternative approaches for heterologous expression (e.g. synthetic biology), which address and overcome the discussed limitations.

Synthetic biology approaches are assumed to have an unprecedented impact on classical biology in general, but in particular on natural products research. This emerging discipline aims at the *de novo* recreation of biological systems using defined modular parts that can be arbitrarily recombined to yield artificial systems exhibiting novel unnatural features. Due to the *per se* modular architecture of PKS/NRPS systems, this concept might revolutionize our view on the rational reprogramming of hybrid megasynthetases producing any desired molecule.⁵⁰ In the context of PKS/NRPS derived biosynthetic pathways, synthetic biology currently largely focusses on the application of DNA synthesis which by now came of age.⁵¹ Today, large biosynthetic genes and even complete gene clusters can be designed and synthesized *de novo*, thereby allowing the addition and/or elimination of restriction sites without changing the encoded protein sequence as well as the modulation of the codon usage

bias. Implementation of ‘standardized’ regulatory elements and synthetic promoter sequences known to function in an optimized host supports heterologous expression of biosynthetic pathways based on synthetic DNA beyond the native regulatory networks.^{40,50} Today, only a limited number of microbial PKS/NRPS biosynthetic pathways have been ‘refactored’ to activate orphan gene clusters, to engineer the pathways for the production of natural product analogs and to study the effects of codon optimization.^{52–56} Interestingly, the latter approach does not necessarily lead to improvement of the production yields and even lower product amounts are often detected compared to the production titers seen in the native producers. A synthetic version of the PKS/NRPS hybrid epothilone pathway from *Sorangium cellulosum*, which was subjected to codon optimization for heterologous expression in the related host *M. xanthus*, merely yielded approximately 0.1 mg/L epothilones A-D, whereas epothilones A and B are produced at 20 and 10 mg/L, respectively, in the native producer *S. cellulosum* So ce90.⁵⁵ However, in some cases codon adaption remains the only way to achieve functional expression in a phylogenetically distant host strain at all. Gemperlein *et al.* demonstrated the successful heterologous expression of a synthetic myxobacterial gene cluster encoding for a polyunsaturated fatty acid (PUFA) synthase from *Aetherobacter fasciculatus* in *Pseudomonas putida*. Formal codon optimization was applied to the synthetic PUFA gene cluster, thereby changing the gene cluster sequence significantly compared to the native sequence. Besides further strain improvement, the synthetic expression system provided at least slightly optimized production yields of the PUFA docosaheptaenoic acid (DHA).⁵⁶

However, considering the fact that synthetic biology principles just begin to be applied to natural product biosynthetic pathways, the number of synthetic gene clusters reported in the literature is conceivably limited and it is still uncharted how to optimize a gene (cluster) sequence in an optimal way. More systematic studies, which include data acquired from interdisciplinary fields, need to be carried out to gain deeper insights into the numerous factors that contribute to the formation of a certain natural product.

Thus, the SynBioDesign project presented in the introduction of this thesis aimed at a systematic view on the design of synthetic expression platforms for the production of myxobacterial secondary metabolites by implementing qualitative and quantitative data obtained from transcriptomics, proteomics and metabolomics approaches in conjunction with high-throughput DNA synthesis and assembly techniques. The development and application of a fast and efficient cloning strategy for large biosynthetic gene cluster constructs would provide the opportunity to synthesize, assemble and heterologously express many artificial versions of a target gene cluster in parallel, e.g. in terms of gene compositions to generate

hybrid gene clusters in order to produce novel compounds or in terms of testing different codon optimization protocols to improve production yields. Consequently, integration of the acquired analytical datasets might provide guidelines on how to optimally design functional and improved gene cluster sequences for heterologous expression, which could then be subsequently analyzed in a following round of DNA synthesis, assembly, expression and data analysis (Figure 2).

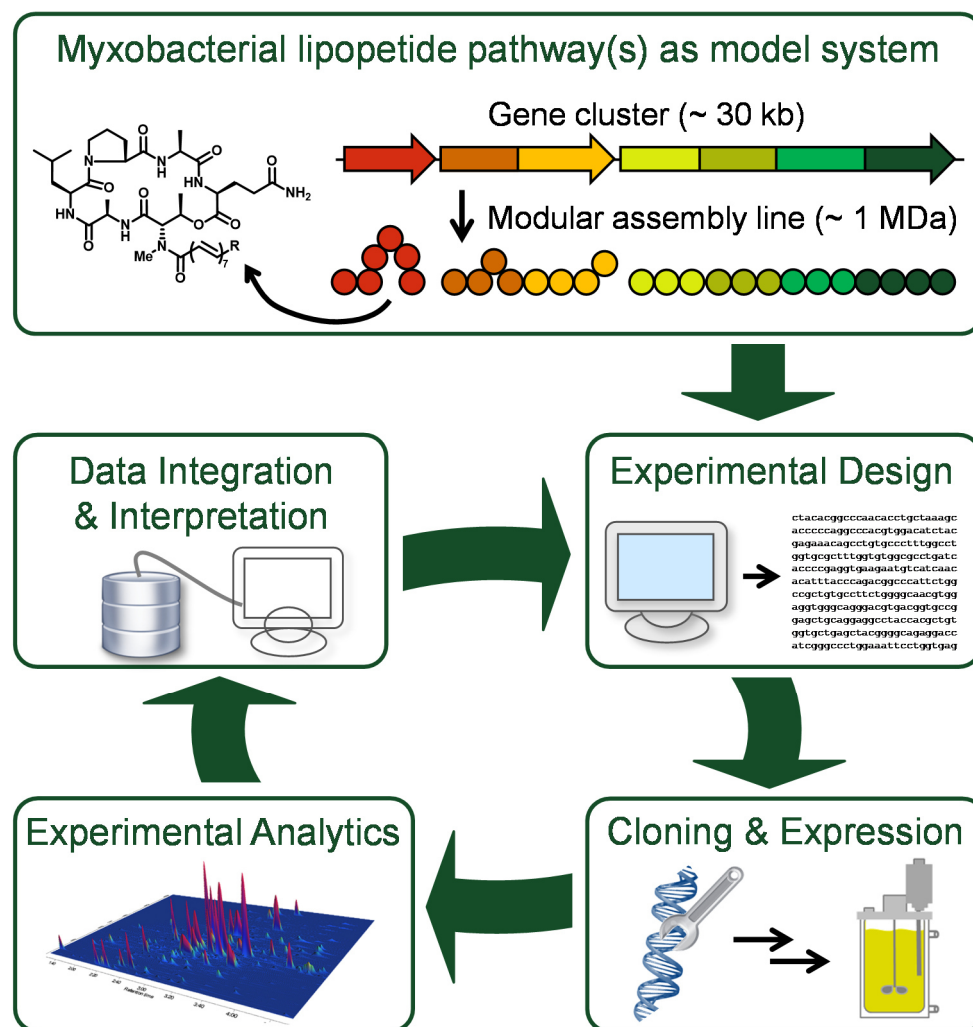


Figure 2. General high-throughput workflow for the continuous optimization of synthetic DNA platforms for the heterologous expression of biosynthetic gene clusters. Iterative cycles of DNA sequence design, assembly and heterologous expression as well as integration of the acquired analytical data might facilitate the identification of critical factors relevant for the productivity of the encoded assembly lines.

In this thesis, we specifically focused on the redesign of myxochromide biosynthetic pathways, on the development of a flexible and efficient DNA assembly strategy and on the establishment of synthetic DNA platforms for the heterologous expression and engineering of the Mch PKS/NRPS megasynthetases to produce novel myxochromide derivatives via

combinatorial biosynthesis in a suitable host organism. The choice of the host strain was one of the major issues to address as several microbes have been previously described as useful heterologous producers of myxobacterial secondary metabolites.^{40,41} One of the most promising host strains is certainly *M. xanthus* DK1622, which was already proven to functionally express foreign myxobacterial gene clusters such as PUFA pathways from *Sorangium cellulosum* and *Aetherobacter* sp.,⁵⁷ the epothilone and disorazol machineries from *S. cellulosum*,^{55,58} the tubulysin pathway from *Cystobacter* sp.,⁵⁹ as well as the myxochromide S assembly line from *S. aurantiaca*,⁶⁰ (Figure 3).

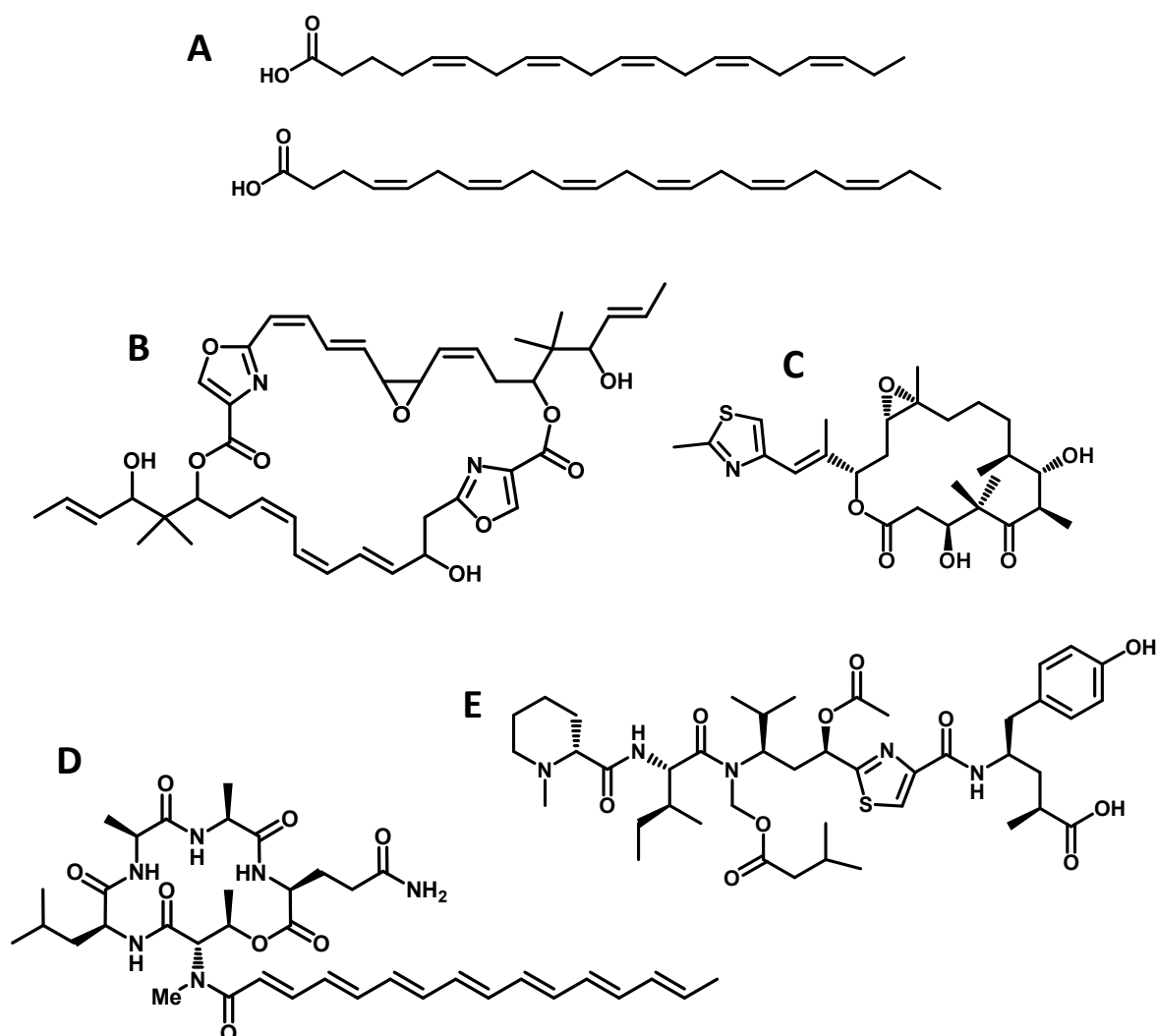


Figure 3. Structures and origin of selected myxobacterial secondary metabolites, which have been heterologously expressed in *M. xanthus* DK1622. **A:** Eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) from *S. cellulosum* and *Aetherobacter* sp. **B:** Disorazol A₂ from *S. cellulosum*. **C:** Epothilone B from *S. cellulosum*. **D:** Myxochromide S₁ produced by *S. aurantiaca*. **E:** Tubulysin A from *Cystobacter* sp.

In addition, keeping in mind that the transfer of biosynthetic pathways into closely related host strains might be favorable and that the host strain must be able to functionally express PKS/NRPS pathways, *M. xanthus* has several advantages over different microbes, even compared to other myxobacterial strains as potential heterologous hosts. *M. xanthus* is characterized by higher growth rates and genetic manipulation is much more convenient compared to e.g. *Sorangium* species as several genetic tools have been established.^{61,62} Especially the fact that a foreign myxochromide pathway was already heterologously expressed in *M. xanthus* without exhibiting toxic activities made this strain the ideal expression host. In contrast, cloning, assembly and engineering of large synthetic DNA constructs harboring the artificial myxochromide biosynthetic gene clusters was accomplished in *E. coli* and final expression constructs were transferred into *M. xanthus* DK1622 via electroporation. Taken into account that there are no replicative plasmids available for myxobacteria, the expression plasmids were equipped with genetic elements allowing for the homologous integration of the artificial pathways into the host chromosome.

In the course of this thesis, the establishment of synthetic DNA platforms for the functional heterologous expression of synthetic *mch* clusters in *M. xanthus* was initially demonstrated with success on the basis of five artificially generated pathway versions, which were shown to produce the naturally occurring myxochromides A, B, C, D and S. In all cases, production levels were comparable to those observed for the native A-type *mch* pathway in the natural producer strain (work by Dr. Fu Yan, unpublished). Thus, *M. xanthus* seems to be the ideal host for heterologous expression of different myxochromide families. However, upscaling of fermentation processes using *M. xanthus* as a production host has been proven difficult, which might be problematic when applying our strategy from model pathways to pharmaceutically interesting myxobacterial pathways.⁶³ As an alternative, other expression hosts could also be used, which are able to express and activate PKS/NRPS pathways, to sufficiently supply the precursors needed for the biosynthesis of PKs and NRPs and which show similar codon preferences. *P. putida* would be such an alternative, since it additionally exhibits even better growth characteristics, is genetically well established and has been shown to be a versatile host strain.⁶⁴ Furthermore, previous studies demonstrated significant improvement of the production yields of myxochromides S in *P. putida*, when the native S-type *mch* pathway from *S. aurantiaca* was heterologously expressed.⁶⁵ As mentioned earlier in this section, fast and efficient DNA assembly technologies are required to truly enable us to deduce common rules for the optimal DNA sequence design for improved heterologous expression by

analyzing numerous gene cluster variants in parallel, which can be rapidly assembled. This issue is discussed in the following section.

4.3.2 Establishment of an Innovative Assembly Strategy for Synthetic Gene Clusters

Besides the improvement of heterologous expression in general, synthetic biology has the potential to essentially contribute to the rational reprogramming and optimization of biosynthetic assembly lines.^{50,66,67} One major drawback of conventional engineering efforts is certainly the fact that with the standard cloning methods only a limited number of gene clusters could be assembled in parallel and in reasonable time frames. With focus on PKS/NRPS engineering, fast and efficient DNA assembly strategies for the construction of artificial biosynthetic gene clusters from smaller fragments, which can be rapidly provided by DNA synthesis, are now beginning to be applied to circumvent laborious library constructions followed by screening and stepwise assembly of target gene clusters via conventional cloning techniques.^{44,46–48,68} DNA synthesis in conjunction with recent advances in recombinant DNA technology allow us to create new entire artificial gene clusters much faster, thereby enabling the high-throughput assembly of numerous gene cluster versions to be tested and optimized for their functionality. In general, modern DNA assembly techniques include homology-based methods such as Gibson isothermal assembly, the related sequence- and ligation-independent cloning (SLIC) and transformation-associated recombination (TAR). Especially the TAR-based assembly methods have been previously used to assemble entire NRPS (and PKS) gene clusters or to capture gene clusters from (environmental) genomic DNA.⁴⁵ Whilst the Gibson assembly represents an *in vitro* tool making use of different enzymes required for the recombination of overlapping DNA fragments, TAR cloning is based on an *in vivo* recombination system in *Saccharomyces cerevisiae*. The major advantage of such homology-based approaches is that these methods are largely sequence-independent. However, high sequence homology of DNA fragments to be assembled or the occurrence of repetitive sequence elements (as it is the case for NRPS domains/modules) may lead to mispairing of the DNA fragments during the assembly process.^{66,67} Over the past few years, targeted genome editing techniques *in vivo* such as transcription activator-like effector nuclease (TALEN) and clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 systems have been implemented into the genetic toolbox of streptomyces to delete entire NRPS pathways.^{69,70} Especially the CRISPR-Cas technique is assumed to become an indispensable tool in order to modify or optimize biosynthetic gene clusters on the genome scale with unprecedented simplicity, accuracy and efficiency, thereby also allowing the

targeting of multiple engineering sites at once.⁷¹ However, the CRISPR-Cas9 technique needs has not been established for myxobacteria so far. To complement the toolbox, *in vitro* restriction/ligation-based assembly techniques have been developed such as ligase cycling reaction (LCR) and Golden Gate cloning.^{47,48} The latter technique relies on type IIS restriction enzymes, which hydrolyze the DNA double strand outside of their recognition sequence, thereby providing sequence-specific overhangs used for directed ligation. This method has been widely used for the assembly of TALEN libraries.^{72,73} Whilst *in vitro* assembly of large gene constructs usually takes several hours, *in vivo* methods may require several days to reconstitute entire NRPS pathways. On the other hand, using *in vivo* DNA assembly is usually much more efficient in terms of reconstituting large gene constructs.

In light of the recent progress in recombinant DNA technology, we were motivated to establish a fast and efficient DNA assembly strategy for the generation of synthetic DNA platforms to provide a generic platform technology for both sequence optimization to improve production yields and pathway engineering to produce novel analogues. The choice of a suitable assembly strategy for the generation of entire gene cluster constructs was a critical issue that needed to be addressed. In this thesis, we considered the Golden Gate cloning method for the generation of a gene library, which built the basis for combinatorial experiments. Gene fragments encoding for parts of the biosynthesis genes of the different *mch* pathways (A-, B-, C-, D- and S-type) were designed and subjected to restriction sites (R-site) engineering to introduce unique R-sites at specific positions for pathway assembly and to remove the corresponding recognition sequences at other positions along the whole gene cluster sequences. In addition, we established for the first time so-called splitter elements (SE) harboring R-sites for a type IIS restriction enzyme exploiting their special feature of cutting outside of their recognition sequence as well as unique R-sites for defined conventional restriction enzymes between each and every catalytic domain encoding fragment of the *mch* biosynthetic genes. By defining unique 4 bp overhangs generated via type IIS mediated hydrolysis, a directed assembly of the single DNA fragments was accomplished to assemble the full-length biosynthetic genes, which make up the *mch* gene library, thereby eliminating the SE sequences ('desplitting' procedure). The R-sites located within the splitter elements allow the straightforward replacement of either catalytic domains or modules and engineered R-sites within the coding sequence at the 5' and 3' ends of each gene cluster fragment can be used to exchange the PKS/NRPS subunits, intergenic linker fragments as well as promoter and terminator fragments. Furthermore, SEs at the 5'/3' ends of each synthetic gene cluster fragment were introduced and initially designed in a way that the resulting 4 bp overhangs

(after DNA hydrolysis) are part of the ‘conventional’ R-site recognition sequences. This strategy should allow not only the type IIS-mediated assembly of the biosynthesis genes but also the assembly of entire gene cluster constructs in a one-pot fashion, thereby making this method amenable to high-throughput cloning procedures. Neither the Golding Gate assembly nor this special SE-based strategy has ever been applied to biosynthetic genes and thus represents the first example how the established assembly method can be used for the generation of a gene library in conjunction with the innovative splitter technology for further engineering of the completely assembled pathways.

In our first attempt to construct a synthetic A-type *mch* pathway from *M. xanthus* DK1622 using the modified Golden Gate-based splitter strategy, we used the type IIS restriction enzyme *AarI*, which was previously reported to be used in conventional Golden Gate cloning protocols.⁷⁴ However, it turned out that cloning efficiencies observed in the ‘desplitting’ of biosynthetic genes using this enzyme were extremely low and sometimes led to incompletely ‘desplitted’ gene constructs or to shortened constructs lacking one or more domain encoding fragments, which is a result of ‘false ligations’ of non-complementary overhangs, which was also reported in some former studies, which described the Golden Gate approach.⁷⁵ This can happen if three out of the four nucleotides making up the fusion sites are complementary to each other, so that they can anneal and subsequently be ligated. Overall, the relatively high number of SEs or domain fragments to be religated, respectively, strongly influences the success of the ‘desplitting’ approach and significantly decreases cloning efficiencies.

However, successful ‘desplitting’ of the synthetic gene constructs was achieved using *AarI*. Unfortunately, one-pot assemblies of an artificial version of the A-type *mch* pathway was not met with success, which is most likely due to the diverse size distribution of the synthetic fragments (150 bp up to 13.4 kb). The assembly of ~ 35 kb constructs from non-standardized DNA fragments (in terms of fragment size) might thus be hard to achieve in a highly efficient and flexible way. In contrast, in the aforementioned TALEN library constructions, modules exhibiting similar sequence lengths in equimolar amounts were used for one-pot restriction/ligation assemblies and relatively small constructs were assembled (up to 10 kb). Taken this into account, fragment and final construct sizes might be the relevant factors limiting the success of this assembly strategy. However, the synthetic A-type *mch* cluster was successfully assembled by conventional means using the unique R-sites located at the 5’/3’ ends and was subsequently heterologously expressed in *M. xanthus* yielding comparable amounts of myxochromides A to those observed with the native A-type *mch* gene cluster.

This work demonstrated the general applicability and functionality of the presented synthetic DNA platform and built a promising basis for further improvement.

To further optimize the ‘desplitting’ procedure, the DNA sequence design was completely adapted to the alternative type IIS restriction enzyme *BsaI* and was extended to all available *mch* cluster types. Unfortunately, the *BsaI* enzyme only skips one nucleotide until it hydrolyzes the DNA double strand compared to four nucleotides, which are skipped by *AarI*. Thus, it was not possible anymore to design the SE-derived 4 bp overhangs in a way that they are part of the ‘conventional’ R-site recognition sequences at the 5’/3’ ends of the synthetic DNA fragments. Alternatively, the unique R-sites at the 5’/3’ ends needed to be engineered at other positions, which would be elaborating or even not possible without changing the protein sequence of the biosynthetic genes. In light of the results obtained from the *AarI*-based one-pot assemblies and due to the fact that the described limitations are also true for the *BsaI* design, we placed the outer SEs only at the 5’/3’ ends of the biosynthetic genes and in spatial separation (7-10 nucleotides) to the unique R-sites, which were engineered within the coding sequence at the same positions as done in the *AarI* design. Consequently, our assembly strategy can still be used for the generation of a gene library and for the engineering of the biosynthetic genes, but the assembly of entire synthetic gene cluster constructs can only be accomplished via stepwise stitching of the synthetic building blocks using the unique R-sites at the 5’/3’ ends.

Changing the sequence/splitter design from *AarI* to *BsaI* in fact resulted in significantly improved cloning efficiencies, especially when applied to the smaller biosynthetic gene constructs. However, ‘desplitting’ of the large *mchC*-based containing gene constructs remained difficult to achieve and laborious screenings for correct clones harboring fully ‘desplitted’ and correctly religated gene constructs needed to be carried out. However, a gene library consisting of different *mchA*, *mchB* and *mchC* gene constructs was successfully generated by using the *BsaI*-based splitter technology and subsequently used for combinatorial biosynthesis. In light of these results, it might be favorable to combine the innovative splitter technology (with a maximum number of SEs of 10 per gene construct) for the generation of dedicated gene libraries with homology-based assembly strategies such as TAR or Gibson assembly, which have already been proven to be suitable assembly strategies for the construction of large biosynthetic pathways exceeding the size of 60 kb.^{45,76} The synthetic gene fragments could be equipped with flanking homology arms the 5’/3’ ends for a directed assembly of the gene cluster constructs and could be directly provided by gene synthesis without performing any additional PCR steps (Figure 4). The assembly vector could

be designed to contain genetic elements, which are functional in *E. coli*, *S. cerevisiae* and e.g. *M. xanthus* to ensure construct assembly, propagation and functional heterologous expression in a myxobacterial host. In doing so, it might be possible to establish a highly efficient generic synthetic DNA platform based on the presented flexible splitter technology allowing the rapid assembly of numerous gene cluster variants to be tested in parallel. In addition to our gene library, other libraries containing e.g. intergenic linkers, inducible synthetic promoters and terminators, can also be included to further increase the scope of this versatile strategy.

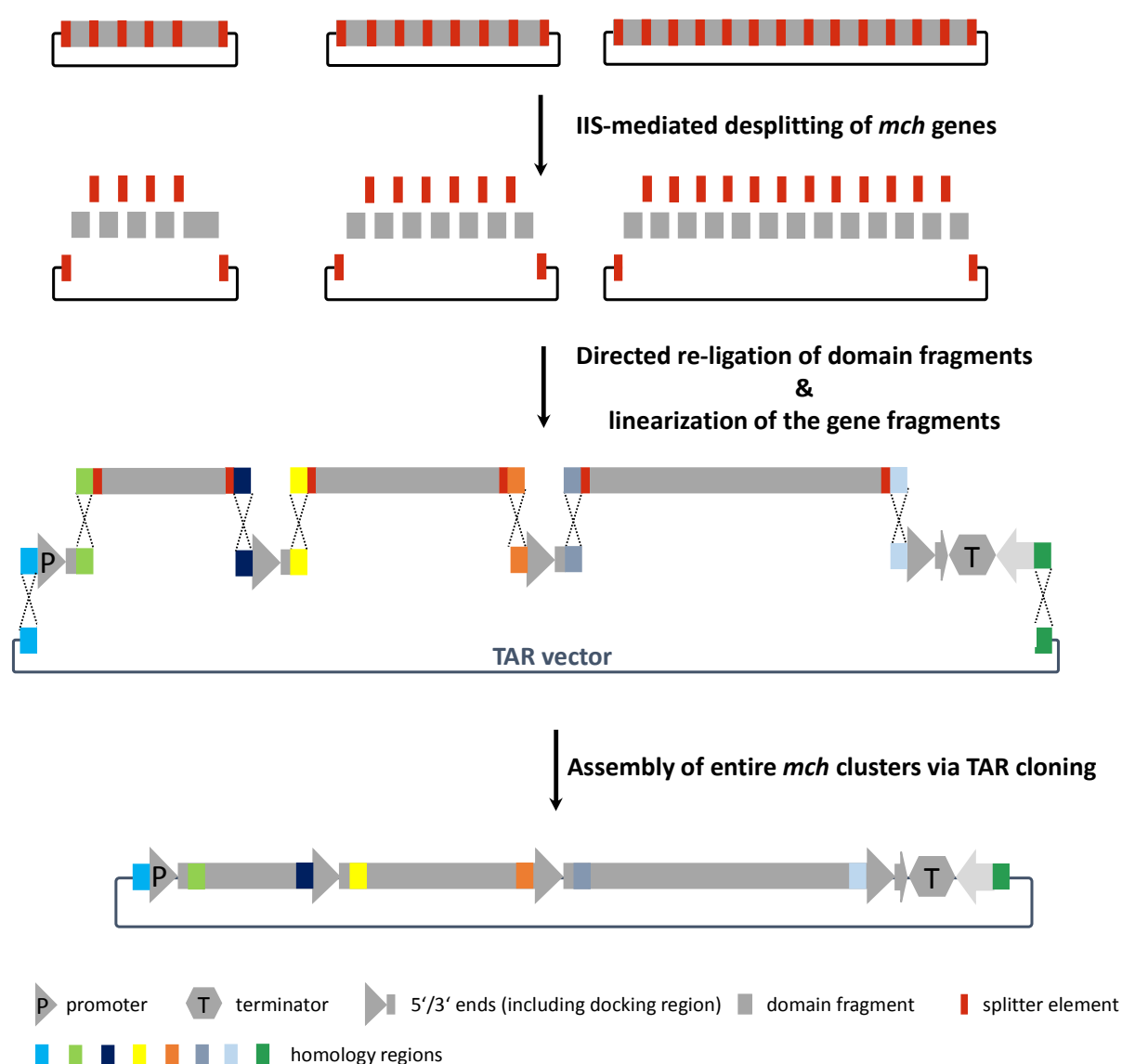


Figure 4. Possible future assembly strategy for the rapid construction of gene libraries and entire artificial biosynthetic pathways. Combination of Golden Gate based ‘desplitting’ for library construction and homology-based assembly strategies such as TAR cloning.

Our work demonstrates the general applicability of the developed expression platforms and sets the stage for further sequence optimization using high-throughput techniques. Moreover,

it gives rise to the rational engineering of artificial hybrid *mch* pathways to produce entirely new myxochromides via combinatorial approaches, which is discussed in the following section.

4.3.3 Synthetic Biotechnology to Engineer Novel Nonribosomal Peptides

Revealing and understanding the structure-activity relationships of different derivatives of a natural product family is one of the key steps towards the development of natural product derived drugs. Due to the inherent structural complexity of most secondary metabolites, their total synthesis or modification by chemical means is often limited or even impossible to achieve. Thus, genetic engineering of the underlying biosynthetic pathways was anticipated to allow alterations of the product structures yielding libraries of structurally diverse natural products, which can be subsequently screened for improved biological activity. The modular architecture of NRPS biosynthetic machineries together with their exceptional biosynthetic logic render them predestinated for such approaches, which address the alteration of single functional groups, the regiochemistry or the NRP backbone scaffold itself. In the past, early developments mainly included approaches to alter the nature of the incorporated precursor molecules such as precursor directed biosynthesis (PDB) and mutasynthesis. These methods both aim at the incorporation of novel unnatural building blocks by feeding these precursors to the culture broth, in which the respective producer strain is grown.^{77,78} Additionally, in the mutasynthesis approach, the genes encoding for enzymes that produce a natural precursor are deleted or inactivated, which enables exclusive production of the engineered compounds. Besides the engineering of the precursor molecules, the structural diversity of NRPs can be further increased via direct modification of the side chains of the peptide core. The utilization of exogenous tailoring enzymes such as halogenases, oxidases, glycosyltransferases, acylases or sulfatases from foreign NRPS pathways has been previously demonstrated to be a useful tool in the modification of the NRP scaffold, largely based on complementation mutants, in which the tailoring enzyme encoding genes were introduced.⁶⁶ Almost 30 years ago, the era of combinatorial biosynthesis began to strongly influence the field of pathway engineering and, more ambitiously, aimed at generating entirely new pathways by ‘mixing and matching’ the domains, modules or subunits of existing NRPS assembly lines.⁷⁹ Unlike the engineering of precursor supply and tailoring enzymes, this approach directly address the modification of the enzymatic assembly line. In the past, combinatorial biosynthesis included a significant number of different strategies ranging from gene fusions, inactivations and replacements to domain, subunit and module swaps. However, only a small number of NRPSs has been

reported to be subjected to combinatorial biosynthesis approaches and there is even only one example for which it was extensively performed on a broad scale. Engineering of the closely related NRPS pathways responsible for the production of the lipopeptides daptomycin, the calcium-dependent antibiotics (CDA) and the compound A54145 using the aforementioned strategies led to the formation of more than 120 different derivatives and some of these ‘unnatural’ lipopeptides even exhibited improved pharmacological properties.^{80,81} However, most of the combinatorial experiments were carried out by chance and did not have any rational basis, which often led to the generation of nonfunctional assembly lines. It should be pointed out that the slow progress made is essentially a result of our currently very limited understanding of the dynamical inter- and intramolecular interactions of the megasynthetases as well as of the functional interplay of the single modules and catalytic domains within the biosynthetic complex. Recent structural studies on higher-order architectures in NRPSs suggest that a well-defined, rigid ‘pearls on a string’ organization of the megasynthetases does not exist. In fact, the structural data shows an unexpected high degree of conformational variability between single NRPS modules, thereby proposing significant interactions between catalytic domains within a module (e.g. C-A domain interactions) but limited interactions between domains of adjacent modules of the assembly line.⁸² However, the overall organization of NRPS machineries is probably not totally unstructured, as transient interactions between individual modules might exist and potentially contribute to the functionality of the assembly line. This may also explain that even slight changes to the modular assembly line may affect proper protein-protein interactions and/or protein folding, thereby negatively influencing its functional integrity. In addition, the described approaches strongly rely on the defined substrate specificities of the biosynthetic enzymes, thereby restricting further structural diversification. A more rational approach for changing the NRP backbones is the engineering of the adenylation (A) domain specificities by exploiting their specificity-conferring code. Several examples have been described, for which a few mutations covering the amino acid residues that confer A domain specificity were sufficient to change the preference for a different amino acid precursor.^{83,84} However, and despite of the advances that have been made over the past decades, the promise of combinatorial biosynthesis to generate novel assembly lines for the production of any desired natural product at will is far from being realized. This can be partly attributed to technical reasons, e.g. most of the classical genetic tools including mobilization of the target gene clusters via library construction and screening as well as classical tools for the downstream engineering of the underlying NRPS systems are time-consuming and low-throughput.

With the established synthetic DNA platforms in hand, which circumvent the described limitations in an elegant way, ‘unnatural’ hybrid myxochromide assembly lines were created based on the *mch* gene library that was generated in the first step of the assembly process. By rationally recombining the NRPS subunits MchB and MchC from the different *mch* pathways, it was possible to generate five hybrid megasynthetases, which were successfully shown to heterologously produce five novel myxochromide families with altered structures of the peptide cores compared to the structures of the naturally occurring myxochromides (Chapter 3, Figure 8). Isolation and structure elucidation of the novel lipopeptides unambiguously demonstrated the functionality of the artificially recombined hybrid *mch* clusters. The non-cognate NRPS subunits from different *mch* pathways were obviously able to successfully interact with each other. In addition, non-native biosynthetic intermediates, which are not biosynthesized by the native *mch* biosynthetic machineries, were transferred to and processed by the downstream domains of the hybrid *mch* pathways, thereby providing fully functional synthetic DNA platforms for the production of entirely novel myxochromide lipopeptide cores. The successful heterologous production of the expected hybrid myxochromides confirms that our constructional sequence design generally works in terms of creating novel functional assembly lines. In light of the important role intergenic linker regions are supposed to play, as they contain the communication-mediating (COM) domains at the C-termini of donor proteins and at the N-termini of recipient proteins facilitating proper interaction between NRPS subunits,⁸⁵ our constructional sequence design conserves the native linker regions, thereby eliminating the risk of disruption of assembly line integrity.

Subunit swaps in NRPS systems have been previously reported only for the closely related lipopeptide pathways responsible for the production of daptomycin, CDA and A54145. The daptomycin megasynthetase consists of the three NRPS subunits DptA, DptBC and DptD, which catalyze the formation of the 13- amino acid lipopeptide. In this study, the *dptD* gene was deleted and the heterologous *cdaPS3* and *lptD* genes from CDA and A54145 biosynthesis were complemented *in trans* leading to the production of novel lipopeptides.⁸⁶ In light of these results, our work describes the first functional hybrid NRPS system, in which subunit exchanges have been successfully engineered within an intact assembly line.

Beyond the subunit-level NRPS engineering, we tried to apply the lessons learned from *mch* pathway evolution using the established synthetic DNA platforms. The unique ‘module-skipping’ process, which was observed in myxochromide S and D biosynthesis, is assumed to be the result of the mutation of the conserved serine by a proline residue in the PCP₄ core motif.¹⁷ By using synthetic DNA fragments encoding mutated PCP regions, we tried to induce

‘module-skipping’ at every position of the synthetic A-type *mch* cluster as well as to reactivate module 4 of the S-type *mch* cluster. Unfortunately, several expression constructs harbored frameshift mutations, so that only four heterologous mutants could be analyzed for production of the expected lipopeptides (mutants harboring inactive modules 1, 3 or 6 of the A-type *mch* pathway plus reactivated module 4 of the S-type *mch* pathway). Induction of ‘module-skipping’ by applying point mutations to PCP domains does not seem to be a tool that can be generally used in NRPS engineering, as none of these engineered assembly lines was shown to produce the expected lipopeptide cores lacking the corresponding amino acid, which would normally be incorporated by the skipped module. These findings indicate that the functionality of the engineered myxochromide assembly lines might somehow be impaired. In the case of the envisioned induction of ‘module-skipping’ at modules 1 and 3 in the A-type *mch* pathway, the observed results can be explained by the fact that adenylation (A) domains are not the only specificity determinants in NRPS assembly lines. Moreover, the condensation (C) domains serve as gatekeepers and exhibit distinct substrate specificities, especially at their acceptor sites. Consequently, unnatural biosynthetic intermediates are not properly recognized and processed leading to the premature release of these intermediates from the assembly line. Skipping of module 6 likely yields a nonfunctional assembly line due to the spatial separation between the upstream module and the terminal TE domain, which might impair correct termination of the NRP biosynthesis. However, as no linear biosynthetic intermediates could be identified in the culture extracts, more mutants should be obtained and analyzed to investigate whether ‘module-skipping’ may be induced via directed mutations in the PCP core motifs. If the induction of ‘module-skipping’ by applying point mutations indeed reflect total abolishment of hybrid myxochromide production, additional factors might be involved in the naturally occurring skipping processes including e.g. mutational changes in the surrounding linker regions of the PCP₄ domain and/or a specific function of the proline residue in the PCP₄ core motif.

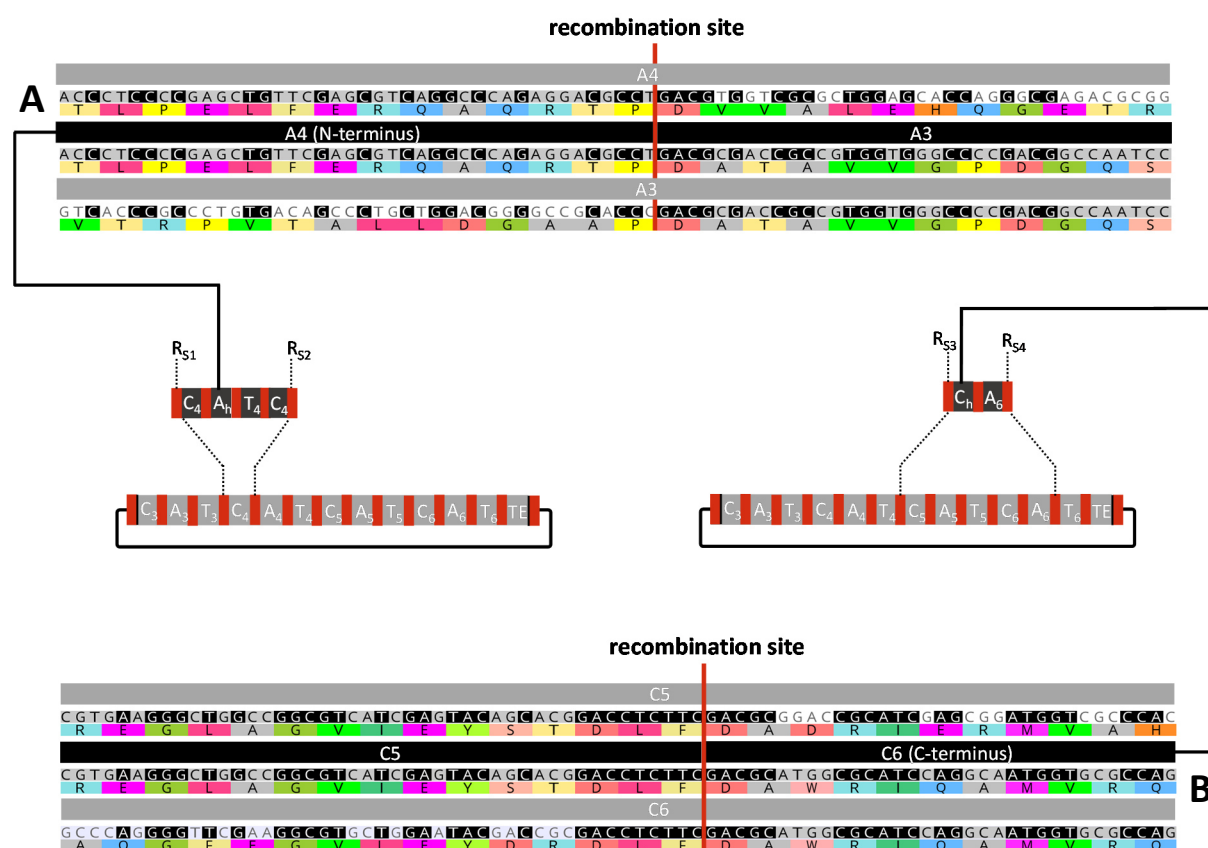


Figure 5. Cloning strategy for the engineering of module duplications and deletions. **A:** Based on protein sequence alignments, hybrid A domains can be designed, in which the N-terminus of an A domain is replaced by the N-terminus of the downstream A domain. **B:** Based on protein sequence alignments, hybrid C domains can be designed, in which the C-terminus of a C domain is replaced by the C-terminus of the downstream C domain. R_S = restriction site.

4.3.4 Concluding Remarks

Although significant progress has been made in engineering NRPS assembly lines, a true understanding of the rules for the rational reprogramming of NRPS systems to produce novel peptides is still missing. Synthetic biology is expected to essentially boost rational engineering efforts by providing ever decreasing costs for DNA synthesis, excellent DNA assembly strategies and well established heterologous hosts.^{50,66,67} The established methods and future developments will allow for the rapid construction and heterologous expression of numerous gene cluster variants, which provides the opportunity to test these variants in parallel, e.g. for improved production or for the generation of novel analogues. In this context, it might be highly desirable to establish libraries by making use of modern DNA assembly techniques. Our developed ‘desplitting’ approach provides an ideal starting point for the generation of gene libraries consisting of various genes encoding NRPS megasynthetases from different pathways. In combination with a more efficient assembly strategy for entire

gene cluster constructs, e.g. based on TAR cloning, our assembly strategy including the constructional sequence design can generally be applied to any desired biosynthetic gene cluster in order to alter product structures or to improve production yields on a broad scale. This might also be of interest for the development of natural product derived drugs with improved pharmacological properties driven by structure-activity relationship studies. In addition, as our synthetic DNA platforms can be exploited in manifold ways due to their remarkable flexibility towards domain and module exchanges, the processes leading to pathway diversification can now be mimicked to gain insights into the rules that govern NRPS evolution. These insights could potentially fuel the rational engineering of other NRPS systems, and await future exploitation. The established ‘desplitting’ procedure also enables far-reaching macroevolution approaches by e.g. parallel one-pot ‘desplitting’ and religation of several different genes harboring the identical set of splitter elements with compatible fusion sites between the different fragments that encode the catalytic domains. In this way, the generation of huge gene libraries consisting of numerous hybrid genes is no longer utopic regarding time constraints, but opens the door for high-throughput combinatorial biosynthesis followed by the analysis of countless unnatural domain, module and subunit combinations. Such applications certainly require sophisticated analytical set-ups as well as the management of the increasing appearance of samples obtained from high-throughput heterologous expression systems, e.g. in 96 well plates. As more and more NRPS pathways are being identified and analyzed, libraries of characterized linkers, which facilitate interdomain, intermodule or intersubunit communication, will be identified and could potentially one day open the door for the envisioned ‘plug-and-play’ approach. This approach aims at fusing standardized parts, e.g. modules from different NRPS pathways, to build entirely new assembly lines exhibiting novel unnatural features. Moreover, having high-throughput assembly and directed evolution methods available, the development of suitable heterologous hosts need to keep pace with this progress as well. One promising approach will be the development of minimal ‘chassis’, in which the majority of genes that are not essential for the host strain’s survival (e.g. all secondary metabolite gene clusters) are deleted. The present work sets the stage for future initiatives aiming at the understanding of general principles of the sequence modulation process for the design of complex biosynthetic gene clusters for improved heterologous expression and for the tailor-made engineering of novel natural products.

4.4 References

- (1) Ziemert, N.; Alanjary, M.; Weber, T. *Nat. Prod. Rep.* **2016**, *33*, 988–1005.
- (2) Goodwin, S.; McPherson, J. D.; McCombie, W. R. *Nat. Rev. Genet.* **2016**, *17*, 333–351.
- (3) Lawrence, J. G.; Roth, J. R. *Genetics* **1996**, *143*, 1843–1860.
- (4) Ochman, H.; Lawrence, J. G.; Groisman, E. A. *Nature* **2000**, *405*, 299–304.
- (5) Koonin, E. V.; Makarova, K. S.; Aravind, L. *Annu. Rev. Microbiol.* **2001**, *55*, 709–742.
- (6) Lenski, R. E.; Travisano, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6808–6814.
- (7) Dittmann, E.; Fewer, D. P.; Neilan, B. A. *FEMS Microbiol. Rev.* **2013**, *37*, 23–43.
- (8) Yue, Q.; Chen, L.; Zhang, X.; Li, K.; Sun, J.; Liu, X.; An, Z.; Bills, G. F. *Eukaryot. Cell* **2015**, *14*, 698–718.
- (9) Roongsawang, N.; Washio, K.; Morikawa, M. *Int. J. Mol. Sci.* **2010**, *12*, 141–172.
- (10) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. *PLoS Comput. Biol.* **2014**, *10*, e1004016.
- (11) Bushley, K. E.; Ripoll, D. R.; Turgeon, B. G. *BMC Evol. Biol.* **2008**, *8*, 328.
- (12) Wang, H.; Sivonen, K.; Fewer, D. P. *Curr. Opin. Genet. Dev.* **2015**, *35*, 79–85.
- (13) Fischbach, M. A.; Walsh, C. T.; Clardy, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4601–4608.
- (14) Wenzel, S. C.; Meiser, P.; Binz, T. M.; Mahmud, T.; Müller, R. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
- (15) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. *ChemBioChem* **2005**, *6*, 375–385.
- (16) Ohlendorf, B.; Kehraus, S.; König, G. M. *J. Nat. Prod.* **2008**, *71*, 1708–1713.
- (17) Burgard, C.; Zaburannyi, N.; Nadmid, S.; Maier, J.; Jenke-Kodama, H.; Luxenburger, E.; Bernauer, H. S.; Wenzel, S. C. *ACS Chem. Biol.* **2017**, *12*, 779–786.
- (18) Moffitt, M. C.; Neilan, B. A. *J. Mol. Evol.* **2003**, *56*, 446–457.
- (19) Jenke-Kodama, H.; Borner, T.; Dittmann, E. *PLoS Comput. Biol.* **2006**, *2*, e132.
- (20) Jenke-Kodama, H.; Dittmann, E. *Phytochemistry* **2009**, *70*, 1858–1866.
- (21) Zucko, J.; Long, P. F.; Hranueli, D.; Cullum, J. J. *Ind. Microbiol. Biotechnol.* **2012**, *39*, 1541–1547.
- (22) Duchaud, E.; Rusniok, C.; Frangeul, L.; Buchrieser, C.; Givaudan, A.; Taourit, S.; Bocs, S.; Boursaux-Eude, C.; Chandler, M.; Charles, J. F. *et al. Nat. Biotechnol.* **2003**, *21*, 1307–1313.
- (23) Stinear, T. P.; Mve-Obiang, A.; Small, P. L.; Frigui, W.; Pryor, M. J.; Brosch, R.; Jenkin, G. A.; Johnson, P. D.; Davies, J. K.; Lee, R. E. *et al. Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1345–1349.
- (24) Moffitt, M. C.; Neilan, B. A. *Appl. Environ. Microbiol.* **2004**, *70*, 6353–6362.
- (25) Rantala, A.; Fewer, D. P.; Hisbergues, M.; Rouhiainen, L.; Vaitomaa, J.; Borner, T.; Sivonen, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 568–573.
- (26) Hahn, D. R.; Gustafson, G.; Waldron, C.; Bullard, B.; Jackson, J. D.; Mitchell, J. J. *Ind. Microbiol. Biotechnol.* **2006**, *33*, 94–104.
- (27) Fewer, D. P.; Tooming-Klunderud, A.; Jokela, J.; Wahlsten, M.; Rouhiainen, L.; Kristensen, T.; Rohrlack, T.; Jakobsen, K. S.; Sivonen, K. *Microbiology (Reading, England)* **2008**, *154*, 1007–1014.
- (28) Doekel, S.; Coeffet-Le Gal, M. F.; Gu, J. Q.; Chu, M.; Baltz, R. H.; Brian, P. *Microbiology* **2008**, *154*, 2872–2880.
- (29) Yu, D.; Xu, F.; Gage, D.; Zhan, J. *Chem. Commun.* **2013**, *49*, 6176–6178.
- (30) Beer, R.; Herbst, K.; Ignatiadis, N.; Kats, I.; Adlung, L.; Meyer, H.; Niopek, D.; Christiansen, T.; Georgi, F.; Kurzawa, N. *et al. Mol. Biosyst.* **2014**, *10*, 1709–1718.
- (31) Tanovic, A.; Samel, S. A.; Essen, L. O.; Marahiel, M. A. *Science* **2008**, *321*, 659–663.
- (32) Koglin, A.; Walsh, C. T. *Nat. Prod. Rep.* **2009**, *26*, 987–1000.
- (33) Wenzel, S. C.; Müller, R. *Curr. Opin. Chem. Biol.* **2005**, *9*, 447–458.
- (34) Moss, S. J.; Martin, C. J.; Wilkinson, B. *Nat. Prod. Rep.* **2004**, *21*, 575–593.
- (35) Hardt, I. H.; Steinmetz, H.; Gerth, K.; Sasse, F.; Reichenbach, H.; Höfle, G. *J. Nat. Prod.* **2001**, *64*, 847–856.
- (36) Thomas, I.; Martin, C. J.; Wilkinson, C. J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2002**, *9*, 781–787.
- (37) Demain, A. L.; Fang, A. *Adv. Biochem. Eng. Biotechnol.* **2000**, *69*, 1–39.
- (38) Beiko, R. G.; Harlow, T. J.; Ragan, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14332–14337.
- (39) Herrmann, J.; Fayad, A. A.; Müller, R. *Nat. Prod. Rep.* **2017**, *34*, 135–160.
- (40) Wenzel, S. C.; Müller, R. *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606.

- (41) Ongley, S.; Bian, X.; Neilan, B. A.; Müller, R. *Nat. Prod. Rep.* **2013**, *30*, 1121–1138.
- (42) Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F. *et al. Nat. Biotechnol.* **2012**, *30*, 440–446.
- (43) Zhang, Y.; Muylers, J. P. P.; Testa, G.; Stewart, A. F. *Nat. Biotechnol.* **2000**, *18*, 1314–1317.
- (44) Noskov, V. N.; Kouprina, N.; Leem, S.-H.; Ouspenski, I.; Barrett, J. C.; Larionov, V. *BMC Genomics* **2003**, *4*, 16.
- (45) Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. *Biopolymers* **2010**, *93*, 833–844.
- (46) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A. ,III; Smith, H. O. *Nat. Methods* **2009**, *6*, 343–345.
- (47) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2008**, *3*, e3647.
- (48) Kok, S. de; Stanton, L. H.; Slaby, T.; Durot, M.; Holmes, V. F.; Patel, K. G.; Platt, D.; Shapland, E. B.; Serber, Z.; Dean, J. *et al. ACS Synth. Biol.* **2014**, *3*, 97–106.
- (49) Gustafsson, C.; Govindarajan, S.; Minshull, J. *Trends Biotechnol.* **2004**, *22*, 346–353.
- (50) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. *Nat. Rev. Microbiol.* **2016**, *14*, 135–149.
- (51) Kosuri, S.; Church, G. M. *Nat. Methods* **2014**, *11*, 499–507.
- (52) Menzella, H. G.; Reisinger, S. J.; Welch, M.; Kealey, J. T.; Kennedy, J.; Reid, R.; Tran, C. Q.; Santi, D. V. *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 22–28.
- (53) Shao, Z.; Rao, G.; Li, C.; Abil, Z.; Luo, Y.; Zhao, H. *ACS Synth. Biol.* **2013**, *2*, 662–669.
- (54) Luo, Y.; Huang, H.; Liang, J.; Wang, M.; Lu, L.; Shao, Z.; Cobb, R. E.; Zhao, H. *Nat. Commun.* **2013**, *4*, 2894.
- (55) Oßwald, C.; Zipf, G.; Schmidt, G.; Maier, J.; Bernauer, H. S.; Müller, R.; Wenzel, S. C. *ACS Synth. Biol.* **2014**, *3*, 759–772.
- (56) Gemperlein, K.; Zipf, G.; Bernauer, H. S.; Müller, R.; Wenzel, S. C. *Metab. Eng.* **2016**, *33*, 98–108.
- (57) Gemperlein, K.; Rachid, S.; Garcia, R. O.; Wenzel, S. C.; Müller, R. *Chem. Sci.* **2014**, *5*, 1733–1741.
- (58) Tu, Q.; Herrmann, J.; Hu, S.; Raju, R.; Bian, X.; Zhang, Y.; Müller, R. *Sci. Rep.* **2016**, *6*, 21066.
- (59) Chai, Y.; Shan, S.; Weissman, K. J.; Hu, S.; Zhang, Y.; Müller, R. *Chem. Biol.* **2012**, *19*, 361–371.
- (60) Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. *Nucleic Acids Res.* **2008**, *36*, e113.
- (61) Magrini, V.; Creighton, C.; Youderian, P. *J. Bacteriol.* **1999**, *181*, 4050–4061.
- (62) Julien, B. *J. Bacteriol.* **2003**, *185*, 6325–6330, DOI: 10.1128/JB.185.21.6325–6330.2003.
- (63) Gerth, K.; Pradella, S.; Perlova, O.; Beyer, S.; Müller, R. *J. Biotechnol.* **2003**, *106*, 233–253.
- (64) Loeschcke, A.; Thies, S. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 6197–6214.
- (65) Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, F. A.; Müller, R. *Chem. Biol.* **2005**, *12*, 349–356.
- (66) Winn, M.; Fyans, J. K.; Zhuo, Y.; Micklefield, J. *Nat. Prod. Rep.* **2016**, *33*, 317–347.
- (67) Kim, E.; Moore, B. S.; Yoon, Y. *J. Nat. Chem. Biol.* **2015**, *11*, 649–659.
- (68) Li, M. Z.; Elledge, S. J. *Nat. Methods* **2007**, *4*, 251–256.
- (69) Cobb, R. E.; Wang, Y.; Zhao, H. *ACS Synth. Biol.* **2015**, *4*, 723–728.
- (70) Joung, J. K.; Sander, J. D. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 49–55.
- (71) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. *Nature* **2012**, *482*, 331–338.
- (72) Cermak, T.; Starker, C. G.; Voytas, D. F. *Methods Mol. Biol.* **2015**, *1239*, 133–159.
- (73) Sanjana, N. E.; Le Cong; Zhou, Y.; Cunniff, M. M.; Feng, G.; Zhang, F. *Nat. Protoc.* **2012**, *7*, 171–192.
- (74) Grigaite, R.; Maneliene, Z.; Janulaitis, A. *Nucleic Acids Res.* **2002**, *30*, e123.
- (75) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. *PLoS ONE* **2009**, *4*, e5553.
- (76) Li, L.; Zhao, Y.; Ruan, L.; Yang, S.; Ge, M.; Jiang, W.; Lu, Y. *Metab. Eng.* **2015**, *29*, 12–25.
- (77) Thiericke, R.; Rohr, J. *Nat. Prod. Rep.* **1993**, *10*, 265–289.
- (78) Rinehart, K. L. *Pure Appl. Chem.* **1977**, *49*, 1361–1384.
- (79) Sun, H.; Liu, Z.; Zhao, H.; Ang, E. L. *Drug Des. Devel. Ther.* **2015**, *9*, 823–833.
- (80) Baltz, R. H. *ACS Synth. Biol.* **2014**, *3*, 748–758.
- (81) Nguyen, K. T.; Ritz, D.; Gu, J. Q.; Alexander, D.; Chu, M.; Miao, V.; Brian, P.; Baltz, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17462–17467.
- (82) Tarry, M. J.; Haque, A. S.; Bui, K. H.; Schmeing, T. M. *Structure (London, England : 1993)* **2017**, *25*, 783–793.

- (83) Kries, H.; Wachtel, R.; Pabst, A.; Wanner, B.; Niquille, D.; Hilvert, D. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10105–10108.
- (84) Thirlway, J.; Lewis, R.; Nunns, L.; Al Nakeeb, M.; Styles, M.; Struck, A. W.; Smith, C. P.; Micklefield, J. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 7181–7184.
- (85) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15585–15590.
- (86) Miao, V.; Coeffet-Le Gal, M. F.; Nguyen, K.; Brian, P.; Penn, J.; Whiting, A.; Steele, J.; Kau, D.; Martin, S.; Ford, R. *et al. Chem. Biol.* **2006**, *13*, 269–276.