Genome Mining in the Myxobacterium *Chondromyces crocatus* Cm c5 for the Discovery of Novel Secondary Metabolites

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

eingereicht von
Konrad Viehrig

November 2015

Tag des Kolloqiums	12. Februar 2016
Dekan:	Prof. Dr. Dirk Bahre
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Rita Bernhardt
Vorsitz:	Prof. Dr. Uli Kazmaier
akad. Mitarbeiter:	Dr. Joseph Zapp

Erklärung

Hiermit versichere ich an Eides Statt, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Der Anteil weiterer Autoren an den vorab publizierten Kapiteln ist jeweils auf der Rückseite des entsprechenden Titelblatts angegeben.

Konrad Viehrig, Saarbrücken, November 2015

Summary

This thesis covers studies on three myxobacterial strains, which were investigated with bioinformatic, biochemical and biotechnological methods to obtain insights into the production of novel bioactive secondary metabolites with relevance for use as clinical therapeutics. The work focused mainly on the myxobacterium *Chondromyces crocatus* Cm c5, where the genetic basis for the biosynthesis of three novel compounds, the crocapeptins, crocadepsins and the crocagins was assigned. Insights into the compounds' biosynthesis were obtained through site-directed mutagenesis, labeling studies and *in vitro* experiments. The genomes of two further myxobacterial strains were investigated by *in silico* sequence analysis, *Jahnella* sp. SBSr007 and *Pyxidococcus fallax* AndGT8. Both strains produce novel bioactive compounds, the antifungal jahnellamides and the antibacterial disciformycins. For both compounds, the corresponding pathway genes were identified and a biosynthesis scheme was developed. The advanced molecular cloning strategies Red/ET recombineering and TAR were applied to clone the crocapeptin and the disciformycin gene clusters for their expression in a heterologous host. These efforts resulted in the successful production of a disciformycin precursor in the host *Myxococcus xanthus* DK1622.

Zusammenfassung

In dieser Dissertation sind die Forschungsarbeiten zusammengefasst, die an drei Myxobakterienstämmen zum Zweck der Erforschung der von ihnen produzierten, neuartigen Naturstoffe mittels bioinformatischer, biochemischer und biotechnologischer Methoden durchgeführt wurden. Ziel war dabei die Aufklärung der Biosynthese dieser Naturstoffe, welcher teilweise hohes Potential für pharmazeutische Anwendung haben. Die Forschungsarbeit befasste sich hauptsächlich mit dem Produzenten *Chondromyces crocatus* Cm c5 und resultierte in der Identifikation der genetischen Grundlagen für die Biosynthese der Crocapeptine, Crocadepsine und Crocagine. Desweiteren wurden einzelne Aspekte der Biosynthese durch gezielte Mutationen in Cm c5, Isotopenmarkierung und *in vitro*-Studien aufgeklärt. Zusätzlich wurden die Genomsequenzen zweier weiterer Myxobakterienstämme, *Jahnella* sp. SBSr007 und *Pyxidococcus fallax* AndGT8 mittels bioinformatischer Methoden untersucht. Diese beiden Stämme produzieren ebenfalls neuartige Naturstoffe, die antifungal wirkenden Jahnellamide und die antibakteriellen Disciformycine. Beiden Substanzfamilien wurden die entsprechenden Biosynthesegene zugeordnet und jeweils ein Biosyntheseschema erstellt. Die Gencluster für die Biosynthese von Crocapeptin und Disciformycin konnten mittels Red/ET DNA-Rekombination und TAR in ein Expressionskonstrukt kloniert werden, wobei im Fall des Disciformycinclusters die erfolgreiche heterologe Produktion einer Vorstufe in dem Stamm *Myxococcus xanthus* DK1622 erzielt wurde.

Danksagung

Mein besonderer Dank gilt Prof. Rolf Müller für die Aufnahme in die Arbeitsgruppe, die Vergabe der interessanten Forschungsthemen und die Betreuung und Unterstützung während der Forschungsarbeiten und Publikationen.

Danken möchte ich weiterhin Prof. Rita Bernhardt für Hilfe und Kooperation bei Cytochrom-bezogenen Fragestellungen.

Bedanken möchte ich mich zudem bei Silke Wenzel für die Unterstützung beim Entwerfen von Klonierungstrategien und vor allem für das kritische Korrekturlesen der Dissertation.
Ich danke zudem Frank Surup, Jennifer Herrmann, Alberto Plaza und Kirsten Harmrolfs für die angenehme und erfolgreiche Zusammenarbeit in den gemeinsamen Projekten, sowie Ronald Garcia für das Zurverfügungstellen seiner Myxophotos zur Verwendung in dieser Dissertation.
Weiterhin möchte ich der gesamten Arbeitsgruppe, einschließlich ehemaliger Kolleginnen und Kollegen für die schöne und interessante Zeit danken.

Publications, Presentations and Patents

The following results from this thesis have been previously published in scientific journals:

Konrad Viehrig, Frank Surup, Kirsten Harmrolfs, Rolf Jansen, Brigitte Kunze and Rolf Müller: **"Concerted Action of P450 Plus Helper Protein To Form the Amino-hydroxy-piperidone Moiety of the Potent Protease Inhibitor Crocapeptin"** published in *Journal of the American Chemical Society* on October 8th, 2013.

Alberto Plaza, Konrad Viehrig, Ronald Garcia, Rolf Müller: **"Jahnellamides, α-Keto-β-Methionine-Containing Peptides from the Terrestrial Myxobacterium *Jahnella* sp: Structure and Biosynthesis".** published in *ORGANIC LETTERS* on October 15th, 2013.

Frank Surup, Konrad Viehrig, Kathrin I. Mohr, Jennifer Herrmann, Rolf Jansen, Rolf Müller: **Disciformycins A and B: 12-Membered Macrolide Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* Active against Multiresistant Staphylococci** published in *Angewandte Chemie International Edition* on October 7th, 2014.

Conference participations:

Biosynthesis of the Crocapeptins, oral presentation. VAAM workshop: Biology and Chemistry of Antibiotic-Producing Bacteria and Fungi, Braunschweig, September 2012.

Biosynthesis of the Crocapeptins from *Chondromyces crocatus*, Poster presentation. VAAM workshop: Biology of Bacteria Producing Natural Products, Frankfurt, September 2013.

Heterologous Production of a Disciformycin Precursor in *Myxococcus xanthus* DK1622, Poster presentation. 2nd European Conference on Natural Products, Frankfurt, September 2015.

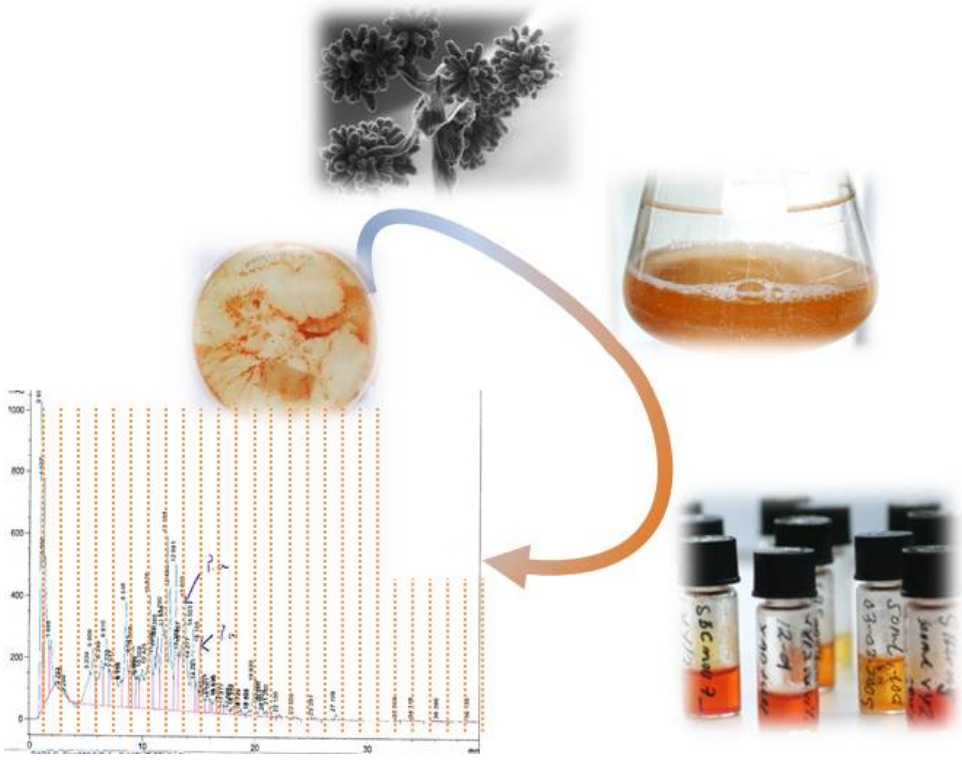
Patent application:

An international patent is pending for disciformycins and the expression construct for heterologous overproduction of the precursor P2 at the European Patent Office as PCT/EP2015/001384. The expression construct has been deposited at the "Deutsche Sammlung von Mikroorganismen und Zellkulturen" as DSM 32082.

Contents

Chapter 1

Introduction



1.1 Natural product discovery

1.1.1 Natural products - secondary metabolites

Natural products are discussed here as compounds, isolated from organisms such as plants or microbes for the use as a clinical drug or therapeuthic agent. Next to non-medicinal utilization as colorants or fragrances, natural products have long history of human use in folk medicine, usually as preparations from plants. Toxic plants, mushrooms or animals were used by prehistoric cultures to aid in hunting, for religious ceremonies and to poison enemies. Isolation of the active compound from the plant material led to the drugs used in modern medicine, which nowadays are usually synthetic derivatives of the original natural product. Early examples are the isolation of morphine from *Papaver somniferum* or salicylic acid from trees of the *Salicaceae* family.[1] Screenings for novel drugs in tradional folk medicine are still undertaken today, and have for example resulted in the discovery of artemisinin, an antimalarial compound from the shrub *Artemisia annua*, a plant traditionally used in chinese folk medicine to treat fever.

Natural products from plants, but also fungal and bacterial toxins are termed secondary metabolites, because unlike primary metabolites, their production is not essential for the survival of the producing organism - at least not in a domestic or laboratory environment. They often do, however, play an important role in the producer's ecological niche.[2] Next to their function as colorants, antioxidants or siderophores, the pharmaceutically most interesting substances fulfil their function by binding to enzymes, usually acting as an enzyme inhibitor. Many of the compounds that are highly poisonous for humans target receptors in the nervous system, leading to apnoea or coronary failure, others have strong cytotoxic effects and ultimately lead to organ failure. Natural secondary metabolites have undergone countless iterations of refinement and optimization by natural selection and as a result, they bind their target molecule with outstanding affinity and specificity. It must however be considered that these structures are not optimized for use in human clinics. For that purpose, a natural product is termed a "lead structure" on the basis of which then a drug is developed by chemical derivatization and formulation to fit the requirements of application in human patients.

Natural products are often described as "small molecules" although no exact size limitations are defined. Figure 1.1 lists a selection of natural products. A salicylic acid molecule has a mass of 160 Da, a lovastatin molecule is around 400 Da and the antibiotic vancomycin has a mass of 1450 Da, which makes it an exceptionally large "small molecule" representative.

1.1.2 Antibiotics

Antibiotics used in therapy are one important example for natural products or natural product derivatives. Although also purely synthetic antibiotics such as sulfonamides exist, the majority is derived from natural products. Antibiotics target structures in sensitive bacteria, such as enzymes involved in cell wall synthesis, cell division, transcription and translation. Many of them can be used to treat bacterial infections, because they target bacterial enzymes that are not present or structurally different in the eukaryotic cells, so that the patient's cells are not affected. Antibiotic treatment exerts selection pressure towards the sensitive bacterial population, therefore development of resistance is an inevitable result. Resistant clones are actively selected for and will repopulate if they are not eliminated by another round of treatment. Resistance is usually conferred as a genetic mutation of the pathogen or uptake of mobile genetic elements - by dividing, the infectious pathogens thus pass the genetic basis for resistance to every daughter cell they generate. This results in a quasi-infinite demand for novel antibiotics as one driving force behind the screening efforts to discover new natural products wherever possible. Some natural products that are not antibiotics are expected to provide alternative treatment strategies

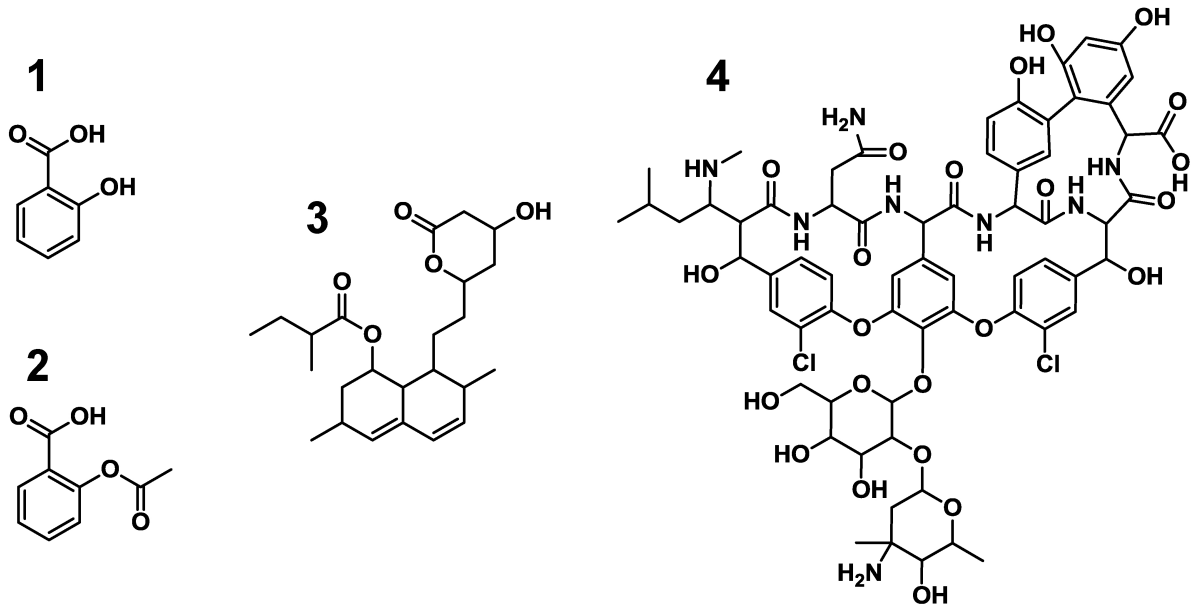


Figure 1.1: Chemical structures of selected pharmaceutically valuable natural products and derivatives. **1:** Salicylic acid. **2:** Acetyl-salicylic acid, a chemical derivative of **1**. **3:** Lovastatin, a natural hypolipidemic agent. **4:** Vancomycin, an antibiotic.

for bacterial infections, such as the inhibition of bacterial toxin production, biofilm formation or quorum-sensing inhibition.
Microbes, including bacteria and fungi are by far the richest source for antibiotics. The starting point for screening of microbes for bioactive natural products was the discovery of penicillin from the fungus *Penicillium notatum* by *Alexander Fleming* in 1928, which was the first microbial antibiotic to treat bacterial infections in clinics. Penicillin G could be used without further chemical modifications, but because of fast resistance development, chemical alterations were introduced soon after. In parallel, the bulk of the antibiotics used today were discovered in the 1950's and 1960's, an era often referred to as the *Golden Age* of antibiotic discovery. As mentioned above, most of the approved antibiotics in clinical use have been discovered over 50 years ago, and most of them originate from the bacterial order *Actinomycetes*. However, classical screening approaches employing resources used for decades cease to yield novel compounds, and other microbial sources need to be investigated.[3]

1.1.3 Screening of Microbes for Natural products

Nowadays, about 30% of human therapeutics sales worldwide account to compounds of plant or microbial origin and about 60% of all approved therapeutics and drug candidates for treatment of cancer and acute infections originate from natural product lead structures.[1] In view of the growing and ageing human population, a need for novel compounds to treat cancer and infections is obvious, the latter especially because of the increasing resistance development of human pathogens to clinically used anti-infectives.

In order to obtain novel compounds from microorganisms, screening programs are applied that combine isolation and cultivation of such organisms under laboratory conditions with subsequent extraction and profiling of the extracted compounds by their bioactivity and their chemical features. This approach, termed as "grind & find" strategy has yielded most of the microbial natural products that are known to date. Bioactive compounds can be identified by screening extract libraries for antibacterial, antifungal or cytotoxic activity using a panel of relevant test organisms and cell lines. In addition, ongoing progress in molecular medicine has revealed many

new targets in human cells for various diseases that can in some cases be conveniently screened by reporter assays *in vitro*. This way, potent and specific inhibitors of proteases, kinases and other factors involved in infections, chronic diseases or cancer development have been found. In parallel, also entirely synthetic antibiotics have been developed, e.g. the sulfonamides or the chinolones. To find novel synthetic antibiotics, libraries of chemical compounds are screened with reporter systems for antibacterial activity just like microbial extracts. The "hit rate" obtained from synthetic compound libraries has however reported to be much lower than screenings employing natural product libraries.[4]

The "grind & find" approach however does have limitations: most importantly, only compounds that are produced in sufficient amounts for activity testing and extracted in active form are detected. Enough material must be obtained for isolation and structure elucidation of the putative lead compound. This is especially challenging for organisms that can not or only poorly be cultivated, such as symbionts, cyanobacteria or sponge-associated bacteria. Culturable bacteria often produce only a fraction of their metabolite potential in laboratory conditions, as has been revealed when the first genome sequences of metabolite producing bacteria became available. This showed striking discrepancy between the genomic potential, estimated from the number of putative secondary metabolite pathways annotated in the genome and the much lower number of compounds that were identified from them.[5] Therefore, methods have been developed to access these "cryptic" products by tapping the hidden potential of their producers. This is being approached either by variations of culture conditions (i.e. example application of stress to trigger metabolite production) or mutagenesis approaches, that yield mutant organisms with a different production profile. The use of genomic information to identify novel compounds is termed "genome mining".[6]

1.1.4 Myxobacteria as source for new natural products

Myxobacteria are gliding, Gram-negative bacteria with unusual features, such as coordinated swarming behavior and formation of multicellular fruiting bodies, that require sophisticated cell-cell communication.[7,8] Consequently, their rather complex life-cycle is reflected in relatively large genomes. Especially members of the *Sorangiineae* suborder with genomes sizes above 13 Mbp are among the prokaryotes with the largest genomes.[9] Most myxobacterial species live as microbial predators in microbe-rich environments such as soil and sediment, and prey on other bacteria and fungi. The antibacterial compounds produced by myxobacteria are probably produced to kill their prey cells, which has so far been shown for myxovirescin and corallopyronin.[10]

Some members of the *Sorangiineae* suborder are saprophytic cellulose degraders who nevertheless produce higly bioactive compounds. Most myxobacterial species are mesothermophilic, but also psychrophilic[11] and moderately thermophilic[12] species have been isolated. In the past, dozens of novel compound classes have been isolated from myxobacteria which were unprecedented and appear to be exclusively produced by them.[13] Some of the antibacterial, antifungal and cytotoxic compounds exhibit novel modes-of-action and represent very promising candidates for further develpment.[14] The full extend of their secondary metabolite biosynthetic potential has emerged after the first strains have been genome-sequenced. As previously observed with the actinomycetes order, the myxobacteria also have a genomic potential that far exceeds the currently observed compound production.[15] Efforts to obtain novel bioactive secondary metabolites should therefore include sequence-based methods to overcome the limitations of classical "grind and find" approaches.

Myxobacteria have characteristics that make work with them challenging: The rather slow growth of some species increases the risk of contaminations and makes cultivations time-consuming. Most of them require complex media which contain lots of compounds that are co-extracted with the secondary metabolites. When analyzing crude extracts, detection of novel masses is

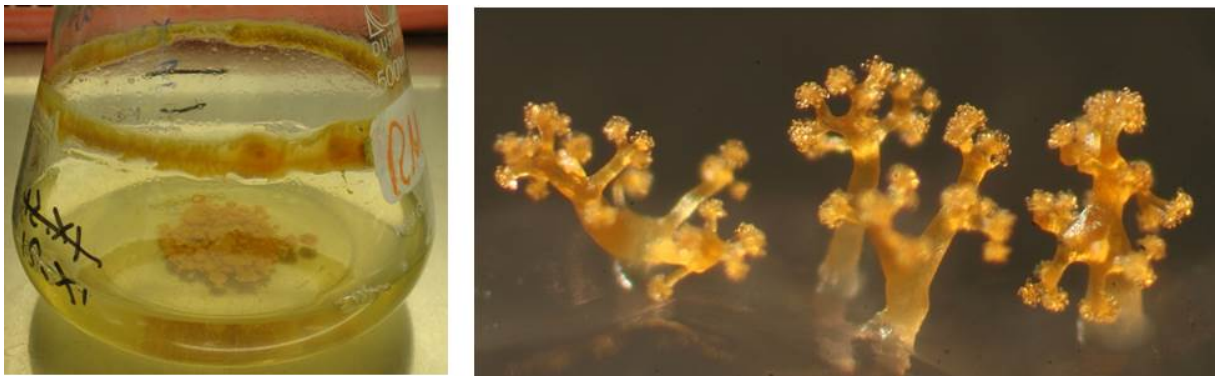


Figure 1.2: *Chondromyces crocatus* Cm c5. Left: Vegetative cells growing as orange-brown clumps in liquid culture. Right: Characterisic tree-shaped fruiting bodies, which are about 200 μm in height.

then impaired, because of the high signal background. Their comparably large genomes and high GC-content make sequencing and molecular biology applications such as PCR difficult, but mutagenesis is possible for a small number of strains either by electroporation or conjugation.

The myxobacteria investigated in this thesis for genome mining purposes are *Chondromyces crocatus* Cm c5, *Pyxidococcus fallax* AndGT8 and *Jahnella* sp. SBSr007.

Chondromyces crocatus, strain Cm c5 is an outstanding multiproducer of secondary metabolites, from which several novel bioactive compounds had been previously isolated, such as the crocacins,[16] chondramides,[17] chondrochlorenes,[18] ajudazol[19] and thuggacin.[20] As observed for other members of the *Sorangiineae* suborder, it comprises one of the largest prokaryotic genomes with over 13 Mbp, and the sequence has been recently completed.[21] Automatic annotation with antiSMASH predicts over 30 further putative secondary metabolite gene clusters to be encoded in the Cm c5 genome, which not correlated to a known product yet. Since mutagenesis tools for this strain have been developed,[22] a targeted pathway inactivation strategy in combination with phenotypic analysis of the resulting clones by extraction and subsequent HR-LCMS analysis was applied to find novel compounds. Targeted mutagenesis in Cm c5 can be done by single-cross insertion of a suicide plasmid conferring hygromycin resistance into the Cm c5 genome. The target gene can be disrupted by cloning a fragment of homologous sequence into the suicide vector. Transformation of Cm c5 is possible through biparental conjugation with the *E. coli* strain ET12567.

Pyxidococcus fallax AndGT8 was selected and prioritized for genome sequencing and subsequent compound mining because of its outstanding antibacterial activity in an activity-based screening of an myxobacterial extract library. Here, the active compound disciformycin was isolated by activity-guided fractionation. Unlike other members of the *Cystobacteriineae* suborder, *P. fallax* AndGT8 can currently not be genetically modified.

Jahnella sp. SBSr007 represents a rather underexplored genus of myxobacteria in the *Sorangiineae* suborder which have been subjected to activity-based screenings. *Jahnella* was found to produce microsclerodermin,[23] and pedein,[24] previously known from *Sorangium* strains and was investigated for further, novel metabolites.

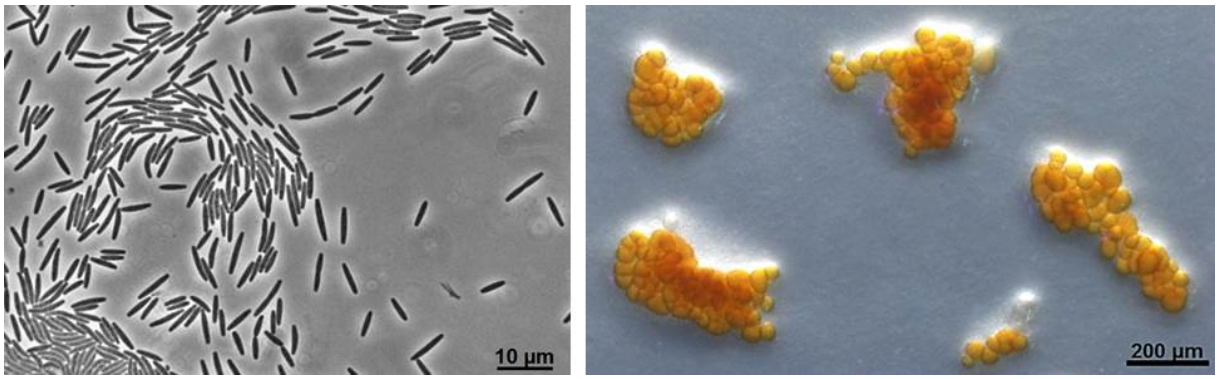


Figure 1.3: *Pyxidococcus fallax* AndGT8. Left: Vegetative cells. Right: Fruiting bodies.

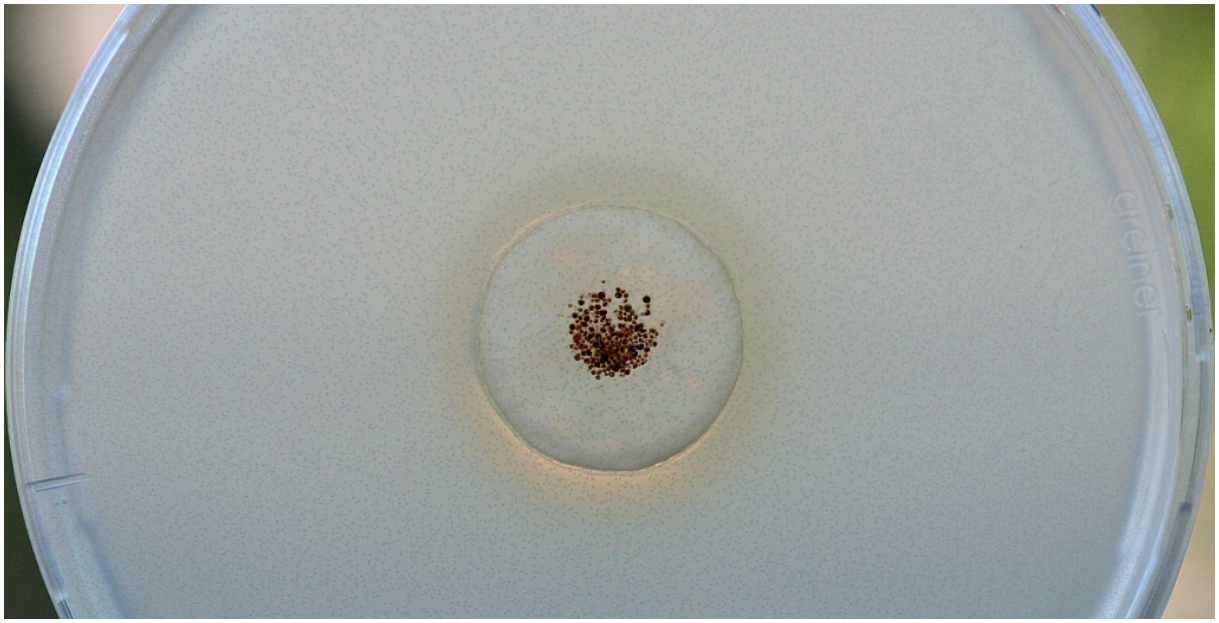


Figure 1.4: *Jahnella* sp. SBSr007 culture on an agar plate. Cell clumps from liquid culture were placed in the center, swarming cells are seen as light orange colouring around the center.

1.2 Genome Mining in Microorganisms

The awareness of the large proportion of predicted "orphan" natural product biosynthesis gene clusters in microbial genome sequences, which far exceed the amount of known products has resulted in strategies to access these products, by rational and sequence-based methods, generally labeled "genome mining".[25] It is at first a data-mining approach, that extracts hints for secondary metabolite pathways from sequence data. Putative biosynthesis genes are detected via conserved sequence motifs, which can be compared to sequence databases of known catalytic motifs, proteins and recently also entire metabolic pathways. That way, candidate natural product gene clusters can be identified, compared to clusters with known products and properties of resulting compounds may be predicted.
In the past, such genome-mining appoaches were already conducted without prior knowledge of the producer organisms' genome sequence. With hybridizing probes, physical DNA-libraries can be screened to find fragments of biosynthetic pathways through sequence homology. The recent developments of genome sequencing have however made the *in silico* approaches much more attractive, as they immediatly offer an overview over the organism's potential and eliminate the need for time-consuming hybridization expermients. However, screening either physical DNA-libraries or virtual sequence data *in silico* always relies on conserved DNA sequence motifs and homology.

To verify the predictions made *in silico* or the result of physical DNA-screening, genetic manipulation of the producer organism is applied, if possible. Inactivation of candidate biosynthetic genes is performed in order to correlate genes to a product or obtain pathway intermediates that give information about the biosynthesis of the compound. Advances in cloning of larger DNA fragments have also allowed to transfer entire biosynthetic pathways into a heterologous host and enable production of the final compound or a precursor.[26] This allows verification of pathways from organisms, which are not amenable to mutagenesis, and also opens up possibilities for efficient pathway engineering. Recently, heterologous expression of secondary metabolites has been achieved with DNA from unculturable organisms, resulting in unprecented compounds such as the fluostatins[27] and tetarimycin.[28] Rational genome mining in combination with heterologous expression is therefore an important strategy to overcome the limitations of traditional cultivation-based screening and greatly expands the possibilities of exploitation of the genetic potential of microbes.

In the following section, different genome mining strategies are presented, which have developed over time and along with the advances in high-throughput genome sequencing and cloning techniques. The genome mining strategies that are applied in each case depend on whether the genome sequence of the producer is available, whether the biosynthesis of a known compound is investigated or whether genome sequence is screened for genes that may encode pathways for novel compounds.

1.2.1 Identification of Secondary Metabolite Gene Clusters and their Products

In prokaryotes, the genes for secondary metabolite pathways are usually organized in close vicinity on the chromosome, often as one or few polycistronic operons. They are referred to as "gene clusters". A large fraction of these gene clusters encode non-ribosomal peptide synthetase (NRPS) or polyketide synthase (PKS) genes, that contain highly conserved sequence stretches, correlating to catalytic motifs in their gene products.[6] The diverse biosynthesis strategies for secondary metabolites are outlined in section 1.3. The bacterial genome can be screened for such sequences, but also for other conserved motifs via different approaches.

Before the genome sequence of microorganisms could be sequenced and deciphered in a convenient time span, genome mining strategies involved the use of physical genome libraries, usually cosmid or BAC libraries. These DNA-libraries can be screened by physical hybridizing DNA probes or PCR-based methods, to identify candidate clones containing a part of a biosynthesis pathway. Cosmid or BAC-clones that were found to contain such sequences were then sequenced manually by primer-walking or in a shotgun approach, to obtain more sequence information, which in turn could be used for site-directed mutagenesis in the producer strain. By comparison of the chemical profile of knockout mutants of the identified candidate genes with that of a wild-type culture, the inactivated genes can possibly be correlated to a product. Numerous natural product biosynthetic pathways have been elucidated by this method.

Since the availability of high-troughput sequencing technology, the screening for biosynthesis pathways is done *in silico*. This can be done by algorithms that automatically screen the six theoretically possible reading-frames in a DNA sequence for encoded conserved catalytical motifs. To date, entire microbial strain collections and microbial communities - usually referred to as environmental DNA - can be sequenced and the resulting data assembled and prepared for comprehensive analysis with databases and automated annotation tools. Key features of NRPS and PKS biosynthesis are conserved among microorganisms, on the level of biochemistry as well as on the level conserved key residues in the involved enzymes. Therefore, large amounts of genomic and metagenomic sequence data can be automatically annotated and screened for natural product gene clusters. Next to algorithms and *in silico* tools that generally predict open-reading frames or genes from crude sequence data, specialized sequence-mining tools like the NRPS-PKS predictor[29] and the NRPS-predictor2[30] detect catalytic NRPS and PKS domains from a linear amino acid sequence and display the detected domain architecture of the detected modules along with substrate specificity predictions. The antiSMASH-Tool[31–33] provides similar output from annotated nucleotide sequence data and will detect clustered secondary metabolite genes. Recently, also tools that detect gene clusters for ribosomally derived bioactive peptides (RiPPs) have been developed, for example RiPPquest.[34] However, all of these tools rely on known catalytic motifs and mechanisms - entirely novel classes of biosynthesis pathways will possibly evade detection.

1.2.2 Cloning and Heterologous Expression of Gene Clusters

The heterologous production of a natural product requires cloning and functional expression of the genes of an entire gene cluster in a suitable host. This is an option to overcome limitations of the original producer which may be difficult to cultivate or economically unfavorable in terms of production. Another reason for a heterologous expression system is the possibility to elucidate the biosynthesis and to generate novel derivatives by pathway engineering, especially if the original producer can not be mutagenzied. Recently, heterologous expression has also been applied for genome mining, where DNA libraries containing novel putative natural product gene clusters were cloned and expressed to yield novel compounds, some of which did have antibacterial or cytostatic activity. The main challenge in cloning such gene clusters is their size, which may exceed the capacity of normal bacterial plasmids. Although some gene clusters encoding small NRPS/PKS pathways can be below 10 kb in length, most natural product megasynthetases are encoded by DNA stretches between 20 kB and 100 kB. An example for a particularly large biosynthesis gene cluster is the stambomycin cluster, spanning 150 kb including the genes for tailoring enzymes and export. It was nevertheless cloned into a bacterial artificial chromosome (BAC), and heterologously expression.[35] Because of their size, PCR-based strategies including conventional restriction-ligation cloning into a suitable expression vector are usually not applicable. Instead, other strategies have been developed and applied as illustrated in figure 1.5.

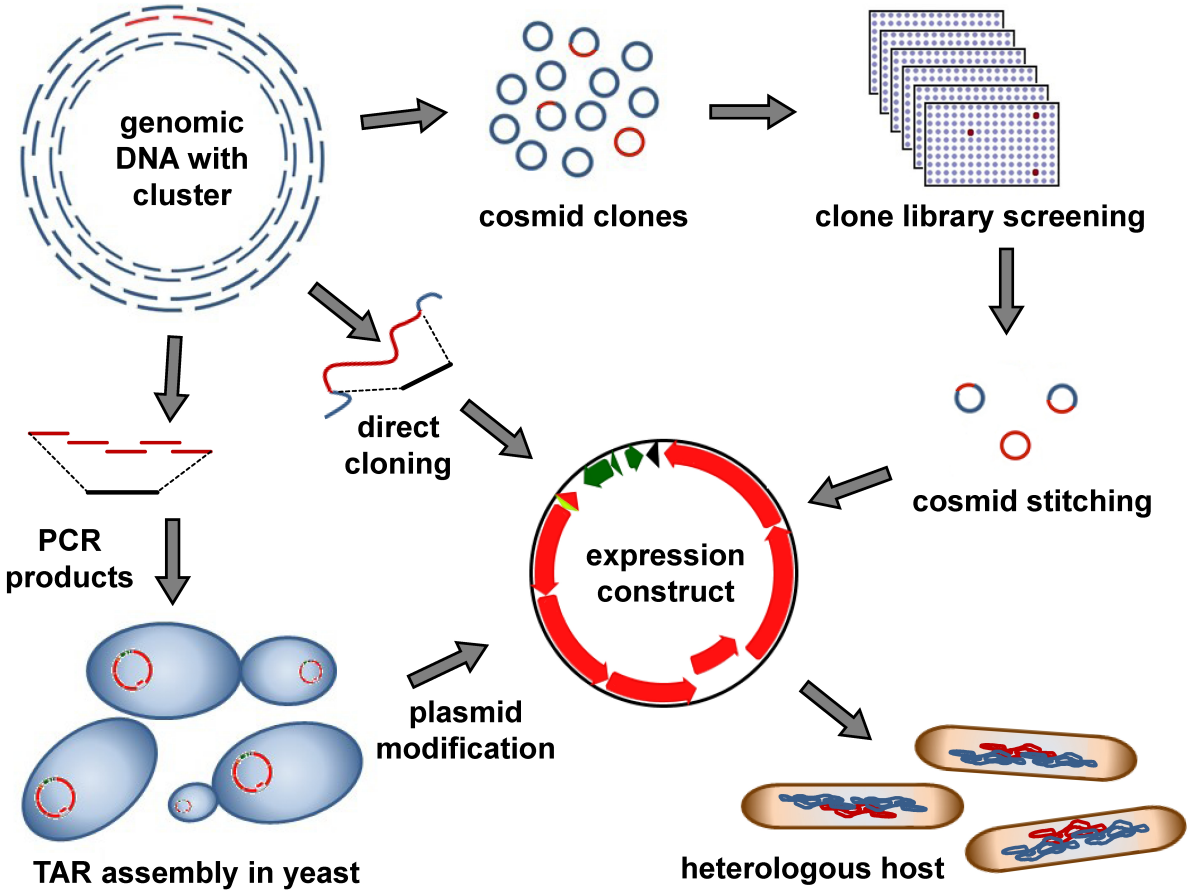


Figure 1.5: Overview over cloning strategies to obtain an expression system for a natural product. Left: The biosynthetic genes are amplified as overlapping PCR products from genomic DNA of the original producer, assembled via TAR cloning to a shuttle plasmid and recovered from the yeast cells. Middle: The gene cluster is captured from fragmented genomic DNA via "direct cloning", which can be done in *E. coli* or *S. cerevisiae* cloning strains. Top Right: A genome library is screened with hybridizing probes or *in silico* methods to identify clones that carry parts of the gene cluster. The DNA molecules of several cosmid- or BAC-clones are assembled via TAR or "Red-ET cloning". The final step is the introduction of the expression construct into the heterologous host, either as extrachromosomal plasmid or by integration into the genome.

1. Stitching of the gene cluster from cosmid or BAC clones, which were previoulsy identified from a clone library. This can be done via recombination based cloning techniques such as Red-ET or TAR-cloning.

2. Direct cloning of the pathway from genomic DNA fragments. Genomic DNA is hydrolyzed with restriction enzymes that do not cut within the target sequence. The gene cluster is then subcloned into a linear capture vector, flanked by homology regions using Red-ET or TAR-cloning.

3. PCR-based assembly by TAR-cloning. Overlapping gene cluster fragments are amplified by PCR and assembled into a capture vector in yeast.

After the successful assembly of the cluster, further modifications are usually required, such as insertion of promotors, transfer cassettes and suitable selection markers to allow propagation and functional gene expression in the desired host. These modifications can be done via recombination based cloning techniques or conventional restriction cloning, if suitable restriction sites are available.

Cloning Tools for Gene Clusters

As mentioned before, several difficulites are associated with cloning of larger gene clusters. Their size usually strongly limits the availability of unique restriction sites. Additional problems arise with typical genes encoding multimodular enzymes such as type-I PKS and NRPS. These often originate from gene duplication and contain repetetive elements of high sequence homology, which can result in problems with PCR-based methods and cloning approaches that rely on sequence homology.

Red-ET cloning: Red-ET cloning uses phage-derived proteins, either RecE and RecT or Redα and Redβ, which are expressed in a designated *E. coli* cloning strain. Two DNA molecules are joined by homologous recombination which relies on a homologous sequence stretch. RecE and Redα are DNA 5'-3'-exonucleases which create a single-stranded 3'-overhang at the end of a linear dsDNA molecule. RecT and Redβ are ssDNA binding proteins that mediate annealing and binding of homologous ssDNA regions. Red-ET cloning is performed by introducing two target DNA molecules into the cloning strain, which expresses the Rec or Red protein pair. Recombination of a circular DNA molecule with a linear molecule (LCHR) and recombination of two linear DNA molecules, resulting in one circular plasmid (LLHR) are established scenarios. The homologous regions are either identical sequences, which are present on overlapping BAC or cosmid clones, or they are introduced artificially via PCR with overlapping regions on the oligonucleotide primers. The recombination event is usually designed to result in a plasmid that contains a selectable marker, as it is done in conventional cloning. Detailed descriptions and applications of the Red-ET technology have been published by Zhang et al.[36,37] Systems for markerless recombination are also available, which involve a 2-step process and make use of toxic genes that are eliminated along with the selection marker in a second recombination step.[38]

TAR cloning: TAR stands for "Transformation Associated Recombination", a process in which linear DNA fragments are introduced into *S. cerevisiae* cells and assembled *in vivo* to an extrachromosomal YAC or plasmid.[39] Multiple DNA molecules with overlapping sequences can be assembled from fragmented gDNA, linearized BACs or cosmids, or PCR products. The desired fragments are then isolated with a linearized capture vector, which contains a selection marker and an orgin of replication for yeast. Typically, a shuttle vector that also allows selection and proliferation in *E. coli* is used. This technology has recently gained frequent use to screen metagenomic libraries from soil samples to express novel gene clusters identified in eDNA sequence libraries.[28,40] The advantage of the TAR method is the convenient assembly of more than two fragments in one step, which also provides more freedom in rearranging gene clusters for more efficient transcriptional control. Furthermore, artificially synthezised DNA fragments may be assembled by this technology if they contain enough overlapping sequence, typically 40-60 bp.

For completeness, also two other recently developed and already commercially available cloning techniques are mentioned here: Gibson assembly,[41] which works similarly to TAR but *in vitro* and the CRISPR-based cloning[42] that uses RNA-guided endonucleases, which can be directed to the desired target site by incubating it with an RNA-oligonucleotide. Both methods further expand the options for gene cluster cloning.

Heterologous Expression of Secondary Metabolite Gene Clusters

Once the biosynthetic gene cluster of interest is assembled, the expression construct must be modified to allow integration into the heterologous host of choice. For transient expression in the heterologous host, a compatible origin of replication must be present together with a selection

marker that allows isolation of recombinant clones. If the host is only amenable to genetic modification by conjugation, the respective transfer genes and additional replications of origin for the donor organism must be included along with separate selection markers. For stable expression via integration of the pathway into the host genome, four strategies exist:

1. Integration via a transposon mechanism will integrate the pathway into a random position in the host genome. For this strategy, the gene cluster has to be flanked by inverted repeats (IR) and a corresponding transposase has to be expressed in the host. The transposase can be encoded on the plasmid backbone, while the selection marker for the host must be encoded within the IR elements.

2. Integration via phage attachment sites will integrate the cluster at a spectific attB site in the host genome. The prerequesite for this strategy is the presence of such an attB site and a plasmid system containing the respective *attP* element and integrase gene.

3. Integration of the plasmid as a suicide vector via single crossing-over. A stretch of DNA-sequence homologous to a target sequence in the host genome is integrated into the expression plasmid. The plasmid will integrate at this specific site.

4. Integration of the plasmid via double crossing-over. Two homologous sequence stretches flank the gene cluster and the selecetion marker. Integration at the specific site leads to excision of the host DNA in between the two homology regions.

The choice of the expression strategy mostly depends on the availablility of the abovementioned systems. Additionally, these strategies may also be combined. A transposon clone may for example be further modified via single-crossing over, given the availability of another selection marker. Using double crossing-over strategies, multiple genetic modifications of a heterologous host may be applied by removing the selection marker to enable its reuse in another engineering step. The resulting clones can be genetically verified by PCR or southern-blotting.

In addition to pathway implementation into the host, further aspects have to be considered for successful heterologous pathway expression to produce desired metabolites. If an antibacterial compound is to be produced by the host, genes for self-resistance may have to be included as well. To finally obtain the expected product, the host strain must not only functionally express the genes but also perform the necessary post-translational modification of the enzymes and be able to provide the building blocks required for biosynthesis compound. This again emphasizes that as much information as possible about the compound to be expressed should be known. If the compound itself is not known yet, the primary sequence of the biosynthesis enzymes should be investigated to predict the product compound structure and its building blocks as precisely as possible. For this, understanding the mechanisms of natural product biosynthesis is essential, which are outlined in section 1.3. One known issue is the post-translational attachment of the 4'-phosphopantetheinyl-arm (PPant) onto carrier-protein domains of NRPS and PKS enzymes which is essential for their catalytic activity.[43] If the host of choice is not naturally able to activate heterologous enzymes, coexpression of a promiscous PPant-transferase gene is a possible solution.[44] If building blocks are required that are not provided by the heterologous host, the genes for their biosynthesis have to be included as well. Alternatively, the expression culture can be amended with the building block or a precursor thereof. Another problem is the compatibility of tailoring enzymes that require interaction with other proteins in the cell, such as cytochrome P450 enzymes, which usually depend on electron carrier proteins, either ferredoxin or a cytochrome-reductase.[45] For this problem, broad-range solutions exist as well, as for example recombinant Adrenodoxin Adx4-108, which has shown to interact with bacterial P450 enzymes.[46]

In the case of myxobacteria and cyanobacteria, the slow growth of many strains makes production in fast growing hosts desirable. The possibilites of this approach have strongly expanded, because of recent developments of cloning techniques for large gene clusters as well as facilitation of suitable expression hosts which may be used for highly efficient production of pharmaceutically valuable compounds. *Myxococcus xanthus* DK1622 has desirable growth characteristics and can be modified with all four integration strategies mentioned above. Generally, hosts that are closer related to the original producer should be preferred, but myxobacterial pathways have also been successfully expressed in phylogenetically more distant hosts such as *Pseudomonas putida*[47,48] and *E. coli*,[49] which demonstrates that also large NRPS/PKS pathways are compatible across bacterial classes and orders.

1.3 Biosynthesis of microbial secondary metabolites

The plethora of microbial natural products can be classified by different features, such as their chemical composition, their biological activity or their biosynthetic origin. Penicillin for example, can be regarded either as a beta-lactame, antibiotic or non-ribosomal peptide. In terms of their biosynthetic origin, most bacterial natural products represent polyketides, non-ribosomal peptides, hybrids thereof, and ribosomal peptides. Other groups such as terpenoids and bioactive fatty-acid derivatives are not featured here. The biosynthesis of polyketides is described in section 1.3.1, the biosynthesis of non-ribosomal peptides and PK-NRP hybrids is described in section 1.3.2 and biosynthesis of ribosomally derived peptides is described in section 1.3.3.
A common feature of the abovementioned compound classes is that they are generated by stepwise condensation of simple building blocks that are available in the cytosolic pool of the producer cell. NRPs are made from free amino-acids and amino-acid metabolism intermediates, while RiPPs are formed from propeptides originating from the aminoacyl-tRNA pool, condensed during the translation process at the ribosome. Polyketides are almost exclusively made up from activated dicarboxy-acids malonyl-CoA and methylmalonyl-CoA, which are accepted by the polyketide synthases and condensed via a decarboxylative Claisen-condensation.
Usually, natural products are modified during or after assembly, which in many cases is essential for their bioactivity. These so-called tailoring modifications are featured in section 1.3.4 and include decorations of the molecule with sugars or fatty acids, halogenation, intramolecular crosslinking and other molecular rearrangements or chemical modifications.
Understanding the basic mechanism of natural product assembly and modification is essential for genome mining approaches, because many of the enzymatic steps that are involved in the assembly of such a compound can be readily predicted from their biosynthetic genes via conserved structural motifs that are reflectd in the primary amino acid sequence of the responsible enzyme.

1.3.1 Polyketides

Polyketides are assembled in a stepwise process from simple dicarboxylic acid building blocks in sequential thioester claisen condensations. The process is similar to fatty acid synthesis and a common ancestral origin of both biosynthetic pathways is assumed. Like in FAS, both substrate molecules are condensated while they are covalently bound to an enzyme. During condensation, the donor substrate is bound to a ketosynthase and the acceptor substrate is bound and an acyl-carrier-protein. In most cases, a PKS produces only one major compound with defined incorporated building blocks and reduction stages after each elongation step, but in some cases, one PKS generates different derivatives of one compound, either through catalytic domains with relaxed substrate specificity or by skipping of certain modifications steps or elongation cycles in iterative systems. For a detailed review on PKS function, see the publication from Hertweck[50] and for comparison to FAS biosynthesis, see the review from Smith and Tsai.[51]

PKS Classification and Modular Megasynthetases

Polyketide synthases can be classified by their architecture as type I, II or III. Type I PKS systems represent multifunctional polypeptides consisting of several catalytic domains bundeled to modules, each usually incorporating one building block into the growing product. They can be further divided into *cis*-AT or *trans*-AT type-I PKS, depending on whether the acyltransferase (AT) functionality is integrated into the module as an AT-domain or the AT is provided *in trans* by a separate polypeptide, respectively. In type II PKS systems, all enzymatic functionalities are localized on individual proteins that form a transient complex during polyketide chain assembly, which is formed in an iterative fashion. Type III PKSs show similarities to chalcone synthases of plants which are able to perform iterative condensation reactions, usually generating small aromatic compounds. PKS systems can be iterative or non-iterative. Iterative modules repeat extension and beta-processing several times, like the mammalian FAS or fungal itertaive type-I PKSs. In non-iterative systems, the growing polyketide chain is transferred along the involved modules.
Among these different PKS systems, non-iterative type-I PKSs are the most important representatives with respect to sequence-based mining efforts, since the number of incorporated building blocks and their respective beta-processing stage is directly represented in the number of modules and their domain organization. Because of this *colinearity rule*, the resulting polyketide product can be predicted to some degree directly from the PKS amino acid sequence. In iterative systems, the number of incorporated building blocks is not represented by the number of modules and product prediction is not possible.

Substrate Loading and Extension

The minimal PKS module for extension of any given building block with a dicarboxylic acid contains a ketosynthase- (KS), an acyltransferase (AT) and an acyl-carrier-protein (ACP) domain. The acyltransferase domain selects the extender unit, either malonyl- or methylmalonyl-CoA and transfers it onto the ACP of its module. Based on several conserved residues in the primary amino acid sequence of AT-domains, their specificity towards malonyl or methylmalonyl-ACP can be predicted in many cases. AT-domains incorporating both precursors are also known, while alternative substrates such as ethylmalonyl-CoA are very rare. The KS domain is responsible for the elongation reaction and covalently binds the donor substrate as a thioesther. In an extension module, the donor substrate is the growing polyketide chain, which is extended with the ACP-bound acceptor substrate as illustrated in fig. 1.6.
Chain extension during polyketide assembly is carried out via thioester-claisen-condensation. The nucleophilic alpha-carbon C2 of the acceptor molecule, an ACP-bound 1,3-dicarboxylic acid thioester, forms a carbon-carbon bond with the thioester carbon of the KS-bound donor substrate. The C3-carbon of the acceptor molecule is released as carbon dioxide, and the thioester bond to which the donor substrate was attached is reduced to a free thiol group. The resulting polyketide chain intermediate is elongated by two carbons and remains bound to the ACP for further processing.

The polyketide extension of linear (noniterative) PKS assembly lines is essentially a coordinated directional process. Unlike other enzymatic reactions in which dissociable substrates are converted to a product, and where the reverse reaction is also possible, polyketide synthesis is a directed process and a reverse reaction is not possible. To achieve this, the ketosynthase domain must undergo different conformational states, which define its substrate accessibilty. In analogy to the established model of the ATP-synthase, a PKS model was proposed by Khosla et al.,[52] in which the KS can switch between three different states,: KS, KS* and KS**. In the KS state, the KS does not accept a substrate from an interacting ACP domain and is catalytically inactive. In this state, the optional beta-processing steps are performed as well as the optional

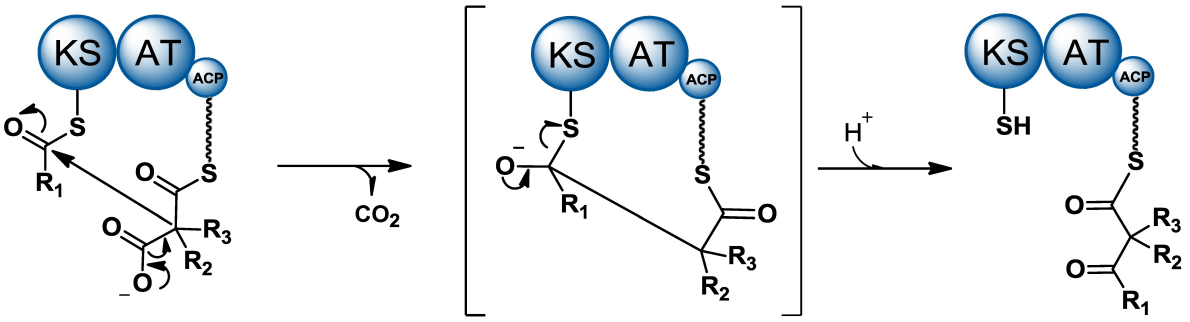


Figure 1.6: Polyketide assembly on a minimal PKS module. The ACP-bound acceptor molecule attacks the KS-bound donor at the nucleophilic carbon of the thioester bond, which leads to formation of a C-C bond and release of carbon dioxide, resulting in an ACP-bound elongated polyketide.

alpha-methylation of the substrate bound to the ACP of the same module. In the KS* state, the KS substrate binding pocket for the ACP of the upstream module is accessible. This substrate is covalently bound to the KS. A non-iterative KS-domain will only enter this state, when the ACP of the same module is unoccupied. In the KS** state, the KS has a substrate covalently bound and is now accepting the extender substrate from the downstram ACP domain. In this state, the claisen condensation is performed, which will release the substrate that is bound to the KS. Thus, during each elongation cycle, the KS goes through the KS* state to bind the donor substrate, then the KS state in which the ACP is loaded with the extender unit, followed by KS** state and the condensation. After release of the donor substrate from the KS, the KS reverts to KS state, during which the newly formed ACP-bound intermediate is modified. Once it is passed on to the downstream KS, the KS domain can enter KS* state again. By this restriction, transthiolation of the KS domain by the newly formed intermediate is supressed, because the KS will only enter KS* when the ACP of the same module is unoccuppied. In iterative PKS and the fatty acid synthase however, the KS will accept the substrate from the ACP of the same module after the processing steps, and bind it covalently. Then, the ACP is loaded with the extender unit and the next elongation cycle is possible.

Beta-keto Modifications and Release

In contrast to fatty acid biosynthesis by FAS, where all incorporated malonate units are fully reduced prior to the next round of elongation, the growing chain can be transferred to the next module as a β-ketide-, β-hydroxy-, α-β-enoyl-, or fully-reduced alkyl-thioester. The four possible modular architectures and their respective products are shown in figure 1.7.

Ketoreductase (KR) domains reduce the beta-keto group to a beta-hydroxy group and thereby generate a chiral carbon. The electrons and hydrogen atoms for every reduction reaction are provided from NADPH as a hydride ion. KR domains can be grouped into A, B and C-type, where A-type forms a *S*-configured alcohol, B-type a *R*-configured alcohol and C-type is catalytically inactive.[53,54] Dehydratase (DH) domains eliminate the β-hydroxy-group by *syn*-elimination and release a water molecule. The other proton is abstracted from the α-carbon, which usually leads to an α-β *trans*-configured double bond. *Cis*-double bonds and shifted β-γ double bonds have also been observed and are expected to result from intrinsic isomerase activity of the DH domain or separate *trans*-acting enzymes.[55] A complete reduction stage is introduced by an enoyl reductase (ER) domain, which catalyzes saturation of the α-β *trans*-double bond by NADPH consumption. This step again introduces stereochemistry, and ER domains can thus be grouped into D-type and L-type, yielding either D- or L-configured alpha-substituents.[56]

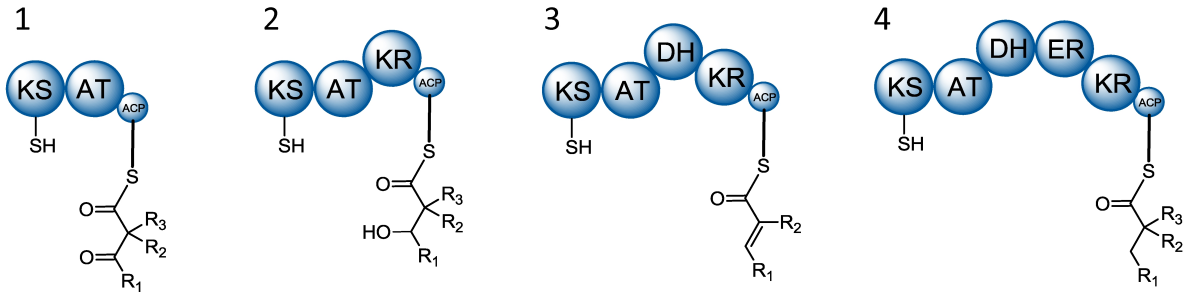


Figure 1.7: Different PKS modules with increasing number of beta-keto modifying domains with their corresponding ACP-bound product after chain extension and beta-processing. **1:** Non-reducing module with ketide product. **2:** Module with ketoreductase domain and resulting secondary alcohol. **3:** Module with ketoreductase and dehydratase domain, resulting in a *trans* double bond. **4:** Fully reducing module, resulting in a saturated C-C bond.

The polyketide product is usually released by nucleophilic attack on the thioester bond of the last module, which is usually accomplished by a thioesterase domain. The nucleophile can be part of the polyketide product itself, often a hydroxy-group, which leads to formation of a macrolactone. Release by hydrolysis generates a linear polyketide chain as a free acid. Other forms of release may involve additional enzymes such as reductases. An overview over the known release mechanisms of pokyketides and non-ribosomal peptides is provided in a review from Du and Lu.[57] In many cases, the polyketide released from the assembly line is further modified by so-called tailoring enzymes, which add functionalities or introduce molecular rearrangements leading to the final, bioactive product. These reactions are outlined in section 1.3.4.

1.3.2 Nonribosomal Peptides

Non-ribosomally made peptides are assembled from carrier-free building blocks such as proteinogenic and non-proteinogenic amino acids, hydroxy acids and intermediates of the amino acid metabolism. The vast majority of these compounds is assembled in a directional process on large multifunctional enzymes, the non-ribosomal peptide synthetases (NRPS).[58] Similarly to PKS assembly lines, they are organized in modules which consist of a defined set of catalytic domains and a carrier protein domain to which the growing peptide chain is covalently attached until product release. A minimal NRPS module consists of a condensation (C) domain, an adenylation (A) domain and a peptidyl-carrier-protein (PCP) domain. NRPS enzymes can be linear (type A), iterative (type B), or non-linear (type C), with type A being the most common type.[59]

There are also alternative routes for ribosomal-independent peptide assembly, as for example in the biosynthesis of some bacterial siderophores such as staphyloferrin,[60] called NRPS-independent siderophore (NIS) pathways. During their biosynthesis, the building blocks for these small peptides are not bound covalently to the condensing enzyme. Instead, they are adenylated at their carboxy-terminus and then condensated via nucleophilic attack of another amino acid or alcohol.

Assembly of non-ribosomal peptides

Upon recognition by the A domain, the free substrate is adenylated under consumption of ATP and then transferred to the ppant arm of the peptide-carrier-protein (PCP) domain of the module, where it is covalently bound as a thioester. Prior to peptide bond formation with the donor substrate bound to the PCP domain of the upstream module, the extender unit may be chemically modified. Substrate recongnition and prediction of an A domains' substrate specificity are featured in subsection 1.3.2.

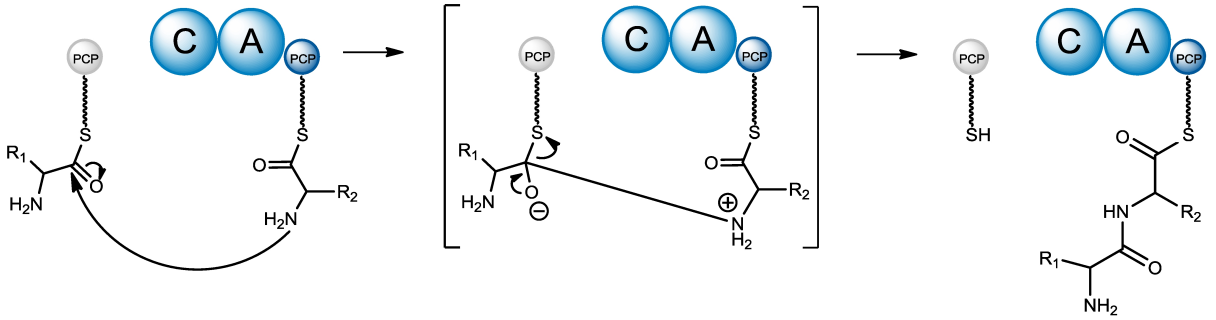


Figure 1.8: Non-ribosomal peptide assembly. A peptide bond is formed by nucleophilic attack of the amino-group of the PCP-bound extender unit towards the carbon of the thioester bond of the donor substrate, which is bound to the PCP of the previous module.

The peptide bond is then formed by the condensation (C) domain, which has a binding pocket for donor and acceptor substrate and can thus act as a gatekeeper and contribute to substrate specificity as well. Unlike in PKS assembly, the C domain does not covalently bind the donor substrate. The elongation reaction is usually a nucleophilic attack of the primary amino group of the extender unit to the thioester carbon of the donor substrate. The donor substrate is then released from its PCP and the extended product can undergo further modification (such as epimerisation) or be attacked by the next extender from the downstream module. In some cases, the donor substrate is attacked by another nucleophile, such as a distal amino group or a hydroxy group, yielding an isopeptide or ester bond, respectively.[61] The latter case can be catalyzed by some C domains as well, and is also a frequent way for product release, catalyzed by a thioesterase (TE) domain. Like in PKS systems, NRPS products can be released as linear compounds by hydrolysis or as macrocyclic products via nucleophilic attack of an internal amino or hydroxy group of the peptide chain itself, yielding macrolactone or macrolactame scaffolds. Strategies for NRP release have been reviewed by Sieber et al.[62]

Adenylation Domains and their Substrate Specificity

Adenylation (A) domains bind their free substrates, usually α-amino-acids, in a binding pocket that is formed via conserved structural motifs, as first shown by structural characterization of the phenylalanine-activating A domain of the gramcidin synthethase.[63] Next to the proteinogenic amino acids in the cytosolic pool, any suitable carboxy-group containing substrate molecule can be incorporated, such as α-hydroxy-acids or α-keto-acids. Upon binding, the substrate is adenylated at the carboxy-terminus and the acyl-adenylate is then covalently bound to the PCP domain of the same module. The binding pocket of each A domain usually exhibits a strict selectivity towards the incorporated building block, although A domains with relaxed substrate specificity are also known. The substrate specificity is defined by critical residues within the binding pocket of A domains, which coordinate the correct substrate via hydrogen bonds. From alignments of multiple A domain amino-acid sequences, experimentally obtained substrate specificities and crystal structures of substrate-saturated A domain proteins, the key residues that define an A domain's substrate specificity have been identified and subsequently, a specificity conferring "non-ribosomal" code has been deduced.[64,65] This code consists of 8 or 10 variable amino acids that occupy a conserved relative position within the A domain sequence. They are extracted by their distance towards highly conserved residues or motifs within the sequence, which is invariable in all A domains. In a simplified model, the critical amino acids shape the binding pocket to fit the substrate exactly, as in most textbook keylock-key models for induced-fit enzyme-substrate interactions. Consequently, genome mining tools have been developed that predict the substrate specificity of a given A domain sequence *in silico*, e.g. the NRPS predictor.[30] The possibility to predict substrates of A domains *in silico* greatly

advances sequence-based genome mining efforts, although the output of these algorithms is rarely unambiguous. Corroboration of the predicted substrates that make up a target compound can be achieved by feeding isotope-labeled precursors to the producing host.[66] It must however be considered, that the predicted substrate, even if the prediction was correct, may undergo modification after peptide assembly or is even cleaved from the final product during compound maturation.[67]

1.3.3 Synthesis of Ribosomal Peptides

Ribosomally synthesized peptides (RiPPs) originate from ribosomally assembled precursor peptides, which are encoded as a gene in the producer organism. The precursor peptides' linear amino acid sequence can be divided into different parts, such as the core peptide, leader peptide and recognition sequences, as illustrated in fig. 1.9 for trichamide biosynthesis.[68]

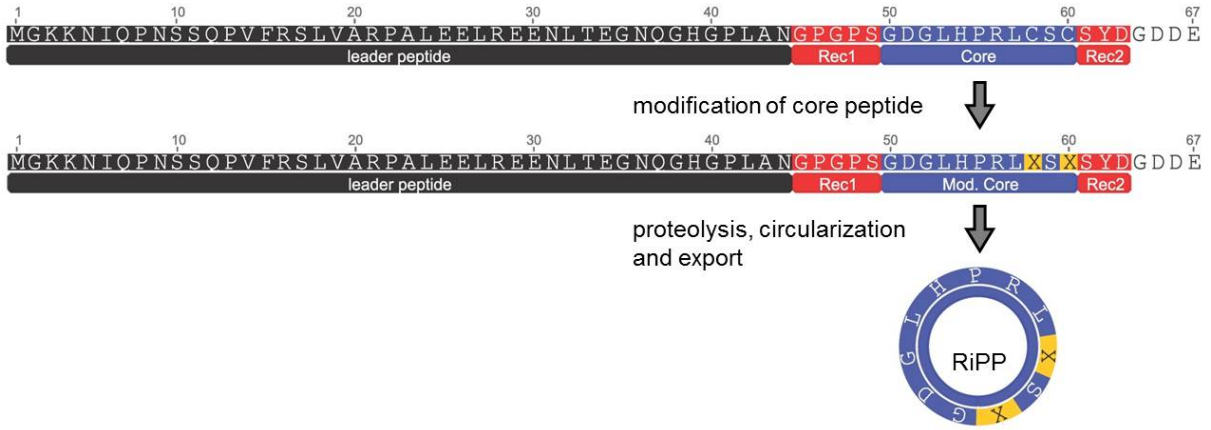


Figure 1.9: Biosynthesis of a macrocyclic RiPP, exemplified here with trichamide formation: The precursor peptide, consisting of the leader peptide (black), recognition sites for modifying enzymes (red) and core peptide sequence are shown. During core peptide modification of the trichamide precursor peptide, two cysteine residues (C) are changed to thiazole (X) heterocycles. The modified core is then cleaved by proteolysis and released via head-to-tail macrocyclization. Generally, RiPPs can have a linear, circular or complex, knot-like structure.

The core peptide is the part that is enzymatically modified after translation and then excised from the precursor peptide to yield the bioactive molecule. Leader peptide and recognition sequences mediate recognition of the modifiying enzymes. The core peptide is commonly located in the middle of the precursor peptide and flanked by leader and recognition sequences, but it may also be located at the terminal part of the precursor sequence. Because of the large variety of ribosomally derived peptides, an overview about these structures is not given here. A recent review from Arnison et al.[69] covers all known RiPP classes and also features the nomenclature for this class of compounds, which is applied in this thesis as well.

It should also be mentioned that, although a ribosomal origin of these peptides implies the resulting compounds to be linear, a large fraction of RiPPs is actually circular.[70] Next to a simple macrocyclization, some RiPPs like the microviridins[71] and the Lasso peptides[72] exhibit complex, knot-like tertiary structures, resulting from post-translational modifications. Most of the toxins known from higher organisms such as fungi, snails or snakes can be classified as RiPPs. A common feature of all RiPPs is that their structure is dictated by the DNA template which encodes the precursor peptide. Non-proteinogenic amino acids or other structural parts found in RiPPs result from post-translational modifications, which are featured in subsection 1.3.4.

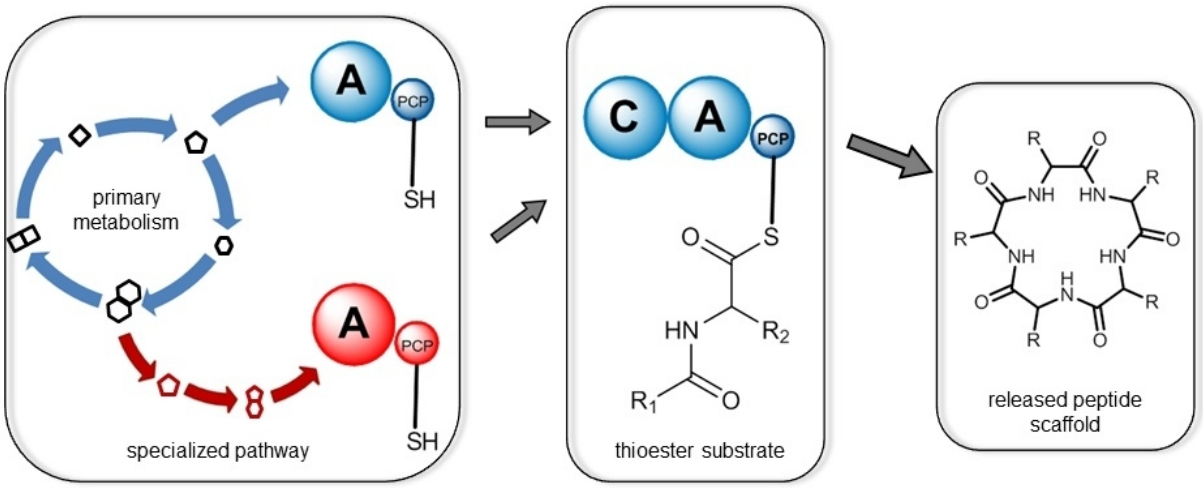


Figure 1.10: Timing of NRPS- and PKS-building block modifications, exemplified for an NRPS system. Unusual building blocks may be delivered by specialized pathways shown in red and are activated by their specific A domain which mediate loading onto the PCP domain. Once loaded, the covalently linked substrate may undergo further modifications. After release from the assembly line, the generated peptide scaffold can be the substrate for tailoring enzymes, which introduce further modifications.

1.3.4 Tailoring and Modification of Natural Products

The high degree of structural diversity in natural products is achieved by a wide variety of building blocks in NRPS biosynthesis and varying degrees of building block processing in PKS biosynthesis, the latter case having a much narrower substrate diversity. In addition to these modifiers, all types of natural products often undergo further modifications after scaffold assembly. These post-assembly line modifications are termed "tailoring" steps and include modifications such as oxidations, halogenations, glycosylations and many more, which are usually carried out with high specificity and selectivity. As mentioned before, RiPP assembly is limited to proteinogenic amino acids, but their respective precursor peptide usually undergoes several tailoring steps after assembly.
During PKS and NRPS biosynthesis, unusual building blocks may be formed either prior to incorporation, or are formed after incorporation, either while the substrate is bound to the carrier protein or after release of a readily assembled precursor from the assembly line, as illustrated in fig. 1.10. In some cases, predictions of the timing of such modification steps can be made by sequence analysis of the respective A domain. The three cases are explained in detail below:

Unusual substrate formation prior to loading: Specialized enzymes form an unusual building block from precursor molecules in the cell, which is specifically recognized by an A domain of the assembly line, adenylated, loaded onto the carrier protein and incorporated into the growing product. Incorporation of unusual building blocks which may not normally be present in the cell is a key feature of NRPS assembly.
One example is vancomycin biosynthesis, where the unusual building block 3,5-dihydroxyphenylglycine (L-Dpg) is first generated by a separate PKS pathway from malonyl-CoA, forming phenylacetic acid.[73] This precursor is then transaminated to L-Dpg and specifically recognized and loaded onto the vancomycin NRPS by the respective A domain.[74] The biosynthesis genes for L-Dpg are part of the vancomycin gene cluster.
In PKS systems, incorporation of building blocks other than malonate and methylmalonate is rarely observed. Remarkable examples are the formation and incorporation of ethylmalonyl-CoA and propylmalonyl-CoA in salinosporamid biosynthesis[75] and hexenoyl-CoA in cinna baramide biosynthesis.[76] These unusual building blocks are generated from α-β unsaturated fatty acids and carbon dioxide by crotonyl-CoA-reductase carboxylase (CCR) class enzymes[77] via reductive

carboxylation prior to loading onto the PKS assembly line.

Building block modification during assembly: A "standard" building block from the cellular metabolite pool, e.g. a proteinogenic amino acid in case of NRPS systems, is recognized by an A domain of the assembly line, adenylated and loaded onto the carrier protein. There, it is converted to form an unusual building block that differs from the substrate of the A domain. The modification is carried out either *in cis* by a specialized enzymatic domain within the module or *in trans* by a tailoring enzyme that transiently interacts with the assembly line module.
In NRPS systems, N-methylations, epimerizations and most halogenations are introduced during assembly. The downstream C domain may contribute to this modification by accepting only the modified building block as donor substrate for the following extension step. In that case, inactivation of the corresponding gene for this modification may result in no product formation at all. In PKS systems, the "standard" modifications featured in section 1.3.1 fall into that category. Next to these, modifications such as double-bond shifts[78] and β-chain branching[79,80] have been observed. Fungal PKS systems frequently harbor methyltransferase domains, that introduce S-adenosylmethionine (SAM) derived α-methyl groups during polyketide assembly.[81]

Modification after assembly: A building block from the cellular metabolite pool is loaded by a specific A domain, incorporated and extended without modifications. After product release, a specific tailoring enzyme modifies the natural product precursor at the position of this building block. Inactivation of the corresponding gene for the tailoring enzyme usually leads to an unmodified precursor molecule. A prominent example is the glycosylation in erythromycin biosynthesis.[82]

None of the modification types discussed above exclude each other. An unusual building block formed prior to loading onto the carrier protein may be further modified during assembly, and modified yet again after product release. The timing of the modification steps is in some cases dictated by chemical restrictions. Mandatory modifications before incorporation may constitute rate-limiting steps for the formation of the product. They are typical for NRPS building blocks, rare in PKS building blocks and not observed in RiPPs. Typical modifications during assembly are only possible with sufficient specificity when the substrate is carrier-protein bound, i.e. N-methylation and epimerization of amino acids. This is also the case for some oxidative cross-linking reactions and hydroxylations. On the other hand, tailoring steps after product release may be preferred when the tailoring enzymes must exhibit a strict substrate specificity towards the precursor. Glycosylation or hydroxylation of a natural product scaffold reduces the amount of unwanted side products, compared to glycosylation or hydroxylation of a free building block from cellular pool.

1.4 General Aims and Scope of this Thesis

The primary aim of this thesis was the correlation of secondary metabolites produced by the myxobacterium *Chondromyces crocatus* Cm c5 to their respective biosynthesis genes by sequence-based genome mining approaches, followed by targeted inactivation of candidate secondary metabolite pathways. Once this correlation was achieved, further studies were carried out to understand the details of the compounds' biosynthesis, involving Cm c5 mutagenesis, feeding experiments, *in-vitro* experiments and heterologous expression.

Three compounds from *Chondromyces crocatus* Cm c5 were investigated: The crocapeptins, the crocadepsins and the crocagins. The crocapeptins had already been isolated and their pla-

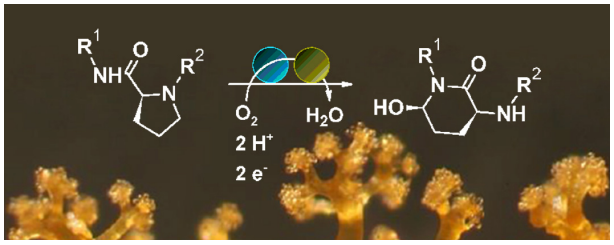
nar structures were elucidated at the HZI in Braunschweig. They were prioritized because of the general promising pharmaceutical potential of the compound class, including antiviral and anti-inflammatorial activity. The crocadepsins were identified after comparing a set of Cm c5 mutant cultures with inactivated PKS, NRPS and hybrid PKS/NRPS gene clusters and correlated to the respective mutants. Structure elucidation revealed that they form a novel compound class. The crocagins also belong to a novel compound class,which could not be correlated to any of the NRPS/PKS clusters in the Cm c5 genome. Their genetic basis was investigated with inactivation studies, followed by cloning of the gene cluster. Additionally, the biosynthesis of two other myxobacterial compounds with promising pharmazeutical potential was investigated by *in silico* sequence analysis to corroborate their chemical structures and obtain insights into their biosynthetic routes: The Jahnellamides from *Jahnella* sp. SBSr007 and the Disciformycins from *Pyxidococcus fallax* AndGT8.

To expand the possibilities of genome mining in myxobacteria, techniques to assemble, modify and transfer secondary metabolite gene cluster into heterologous hosts for functional expression were applied to the crocapeptin, crocagin and disciformyin gene clusters. Here, the aim was to assemble the gene cluster and test their expression in the related host *Myxococcus xanthus* DK1622, which has previously shown to enable overproduction and also expands the possiblities to study the compounds biosynthesis and create novel derivatives through pathway manipulation.

Chapter 2

Concerted Action of P450 Plus Helper Protein To Form the Aminohydroxy- piperidone Moiety of the Potent Protease Inhibitor Crocapeptin

Konrad Viehrig, Frank Surup, Kirsten Harmrolfs, Rolf Jansen, Brigitte Kunze, Rolf Müller*



This chapter was published as an article in the *Journal of the American Chemical Society* on October 8th, 2013. Adapted with permission from "Concerted Action of P450 Plus Helper Protein To Form the Amino-hydroxy- piperidone Moiety of the Potent Protease Inhibitor Crocapeptin".

*To whom correspondence should be adressed.

Author's efforts

The author's efforts regarding this publication comprise the sequence assembly and annotation of the crocapeptin gene cluster, design and cloning of the inactivation constructs for the mutagenesis of *Chondromyces crocatus* Cm c5, cultivation, mutagenesis and labeling experiments as well es extraction and MS analysis of the mutant cultures and precursor isolation. Further efforts are the cloning and heterologous expression of selected genes in *E. coli*, followed by protein purification and *in vitro* assays. The author contributed the main part of the manuscript.

Contributions by others

The compound was initially isolated and the structure elucidated by Rolf Jansen and Brigitte Kunze. Absolute configurations of crocapeptin A and B were elucidated by Frank Surup, who contributed all NMR data to the manuskript. NMR measurements of pre-crocapeptin, isolation of hydroxylation intermediates and investigation of Ahp-formation via NMR were conducted by Kirsten Harmrolfs. Frank Surup and Kirsten Harmrolfs contributed to the manuscript. The project was supervised by Prof. Rolf Müller, who also contributed to conceiving and writing of the manuscript. Further aknowledgements are listed at the end of this chapter.

Abstract

The crocapeptins are described here as cyclic depsipeptides, isolated from cultures of the myxobacterium *Chondromyces crocatus*. Structure elucidation of the compounds revealed a cyanopeptolin-like skeleton, containing the characteristic amino-hydroxy-piperidone (Ahp)- heterocycle. Like the cyanopeptolins, the myxobacterial crocapeptins proved to be serine protease inhibitors. The nonribosomal origin of the peptide was confirmed by mutagenesis experiments, and the biosynthesis gene cluster was sequenced. It could be shown that the Ahp-heterocycle originates from a proline residue in the precursor molecule precrocapeptin, which is converted to crocapeptin by the tailoring enzymes CpnE and CpnF. Conversion of precrocapeptin isolated from a *cpnF*$^-$ mutant into crocapeptin was achieved using recombinant CpnF, a cytochrome P450 enzyme responsible for hydroxylation of the proline residue in precrocapeptin. Addition of protein CpnE resulted in strongly increased conversion rates toward Ahp containing product. A mutant with 10-fold increased production of crocapeptin A was created through insertion of the Pnpt-promotor in front of the NRPS gene.

2.1 Introduction

Natural products from microorganisms continue to be important sources for bioactive compounds with potential for clinical, agricultural, or biotechnological use. Compounds with antimicrobial, antiviral, cytotoxic, immunosuppressive, and other biological activities are produced by bacteria and fungi. Myxobacteria are an established source of secondary metabolites,[83] and a large variety of novel chemical structures have been isolated, many of which possess novel modes of action.[14] Myxobacteria typically produce polyketides and nonribosomal peptides or hybrid molecules thereof, which are assembled in the cells by large multienzyme complexes, polyketide synthetases (PKS), or nonribosomal peptide synthethases (NRPS).[84] As in other prokaryotic producers, the genes encoding the enzymes necessary for biosynthesis of a compound family are usually found clustered in close vicinity in the genome of the producing organism, often together with genes necessary for product export, regulation, and self-resistance.

The myxobacterium *Chondromyces crocatus* Cm c5 has been found to produce an outstanding compound spectrum, comprising the ajudazols,[19] chondramids,[17] crocacins,[16] chondrochlorens,[18] and thuggacins,[85] which have been detected in activity-based screenings because of their antifungal, antibacterial, or cytotoxic qualities. Except for the chondramides, which are highly similar to the jaspamides isolated from a *Jaspis* sponge, they represent unique compound families. Partial genome sequencing of strain Cm c5 revealed the presence of more putative secondary metabolite gene clusters, which were not yet associated with a product. Thus, further bioactive compounds are expected to be produced by this strain under currently unknown conditions. In a combined genome and metabolite mining approach, noticeable UV peaks of Cm c5 extracts were investigated and a number of putative NRPS and PKS gene clusters were inactivated, to correlate compound peaks to genes and isolate novel bioactive metabolites.[86]
The crocapeptins presented in this Article belong to the cyanopeptolin-like compound family, cyanobacterial cyclic depsipeptides with over 100 congeners already isolated from various cyanobacteria.[87] Their unique structural feature is an amino-hydroxy-piperidone (Ahp)-heterocycle, which is crucial for the bioactivity as potent serine protease inhibitors. Different congeners show preference to different target enzymes, based on the amino acids incorporated at key positions within the peptide.[88] Cocrystallization of the elastase with scyptolin revealed that the compound fits into the active site of the enzyme and directly blocks the serine of the catalytic triad.[89] The rigidity of the compound, crucial for blocking the enzyme activity, is provided by the Ahp-heterocycle. Promising pharmaceutical potential may arise from this compound family as highly specific inhibitors for serine proteases, obtained either from a natural source or through modification of an existing scaffold. Because of the absence of toxicity, the cyanopeptolins are

seen as lead structures with high potential in therapy, yet their development is hampered by the low amounts typically obtained from cyanobacteria. As an alternative approach, total synthesis of some congeners has been pursued.[90]

Since their first isolation,[91] the pharmaceutical potential of various cyanopeptolins has been investigated, because serine proteases are involved in a large variety of diseases, such as virus infections, chronic inflammations, and cancer. Scyptolin is a specific elastase inhibitor,[89] while the ichthyopeptins were found to block influenza virus replication in infected cells and to increase survival rates in infected mice.[92] Intriguingly, several cyanopeptolin and crocapeptin congeners were recently patented for specific inhibition of human kallikreins to treat chronic skin diseases such as psoriasis and atopic dermatitis.[93]

Cyanopeptolins are produced by NRPS enzymes in cyanobacteria, and four cyanopeptolin NRPS clusters from different cyanobacterial strains have been sequenced and published to date.[94] Despite the numerous variants isolated and studied, the origin and biosynthesis of the characteristic Ahp-heterocycle have not yet been elucidated. Because genetic tools for Cm c5 had been established previously, overproduction of the crocapeptins and the elucidation of the Ahp formation in *Chondromyces* were of particular interest. Detailed knowledge of the biosynthetic pathway and the responsible enzymes is often crucial for overproduction and further development of a promising compound and may provide options to generate non-natural derivatives by means of biotechnology. In this Article, the isolation and structure elucidation of the crocapeptins, the identification of the corresponding biosynthetic gene cluster, the isolation and structure elucidation of a precrocapeptin from a P450 mutant, and finally Ahp formation within this precursor molecule in vitro using purified P450 plus an unprecedented helper protein are presented.

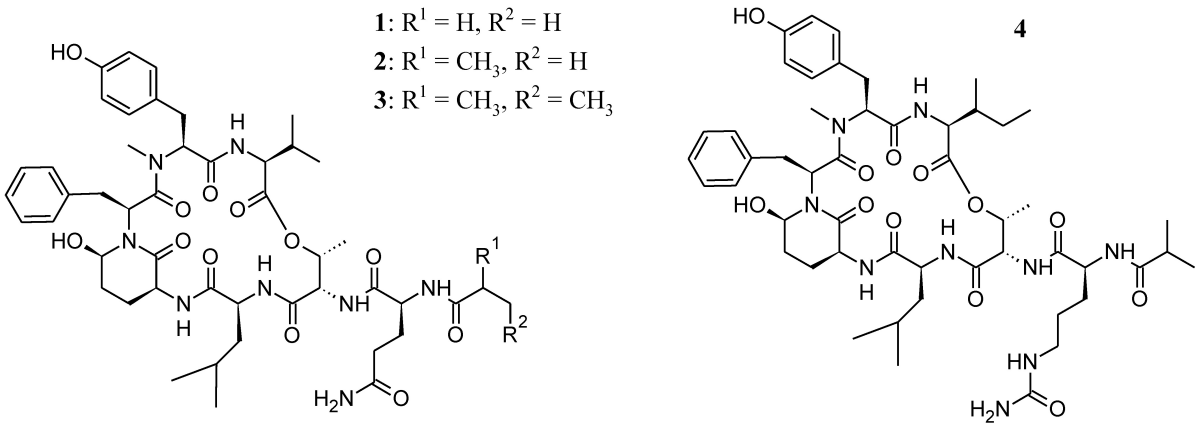


Figure 2.1: Crocapeptins A1-A3 **(1-3)**, isolated from strain Cmc5 and crocapeptin B **(4)** from Cm c2.

2.2 Results

Analysis and Isolation of Crocapeptins.
The crocapeptins **1-3** were detected in extracts of *C. crocatus* Cm c5 by HPLC-UV-HR-ESI-MS. The molecular formulas of $C_{47}H_{66}N_8O_{12}$ for crocapeptin A1 **(1)**, $C_{48}H_{68}N_8O_{12}$ for crocapeptin A2 **(2)**, and $C_{49}H_{70}N_8O_{12}$ for crocapeptin A3 **(3)** were calculated from the $[M+H]^+$ ion clusters m/z 935.4868, 949.5027, and 963.5186. Thus, **1** and **3** differ from **2** by the formal abstraction and addition of a methyl group, respectively.
The molecular formula of crocapeptin B **(4)**, which is produced by *C. crocatus* Cm c2, was determined as $C_{50}H_{73}N_9O_{12}$ by $[M+H]^+$ ion cluster at m/z 992.5450. Additionally, in the extract of strain Cm c2, minor compounds with molecular formulas $C_{49}H_{71}N_9O_{12}$ and $C_{51}H_{75}N_9O_{12}$ were observed.

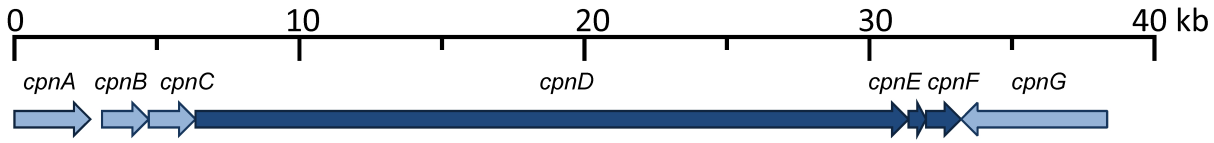
Structure Elucidation of Crocapeptins.
The basic structure of the crocapeptins (Figure 2.1) was primarily elucidated with main component **2**. Using HMQC and DEPT NMR spectra, all protons were assigned to their respective carbon atoms. The remaining 8 protons were recognized from their chemical shifts as 7 amide protons and one hydroxyl proton. The ^{13}C NMR spectrum revealed nine carbon atoms within the amide/ester range. Following the interpretation of COSY, TOCSY, and HMBC data, eight partial structures accounting for all atoms of the molecule were detected, that is, the amino acids glutamine, threonine, leucine, phenylalanine, N-methyltyrosine, and valine together with 3- amino-6-hydroxy-2-piperidone (Ahp) and isobutyric acid (Ibu) moieties. Their linear sequence in **2** was determined from HMBC correlations as Ibu-Gln-Thr-Leu-Ahp-Phe-NMeTyr-Val (Supporting Information Table S1, Figure S1). The high field shift of the oxymethine H3 of Thr (δ 5.39) indicated an ester linkage at this position, which was confirmed by the HMBC correlation between H3 of Thr (δ 5.39) and C1 of Val (δ 171.9), thus completing the planar depsipeptide structure of **2**. Comparison of the NMR data of **1**, **2**, and **3** revealed the presence of different starter units: the isobutyrate moiety of **2** is replaced by propionate in the case of **1** and by 2-methylbutyrate in **3**. The absolute configurations of L-Gln, L-Thr, L-Leu, L-Phe, L-NMeTyr, and L-Val residues were assigned by Marfeys method.17 However, initially an oxidation with CrO_3 was performed liberating glutamate from the unusual 3-amino-6-hydroxy-2-piperidone (Ahp) moiety.18 Because no *D*-glutamate was detected, the 3*S*-configuration was deduced for Ahp. Furthermore, ROESY correlations within the Ahp residue between NH and H4a, H4a, and OH suggested a 6*R*-Ahp configuration (Supporting Information Figure S3). A strong NOE between α-protons of Phe and *N*MeTyr showed a cisamide bond between these amino acids characteristic for structurally related compounds. Because 2-methylbutyrate biosynthetically

derives from isoleucine, an *S*-configuration can be proposed for the branched-chain starter unit in **3**. The NMR spectra of **4** were very similar to those of **1-3**. They revealed the presence of isoleucine, *N*-methyltyrosine, phenylalanine, 3-amino-6-hydroxy-2-piperidone, leucin, threonin, citrullin (Cit), and isobutyric acid moieties. Their linear sequence in **4** was also determined by key HMBC correlations as Ibu-Cit-Thr-Leu-Ahp-Phe-NMeTyr-Ile (Supporting Information Table S2, Figure S2). Using Marfeys method, *L*-configurations were determined for all amino acids. The Ahp-configuration of **4** was deduced as 3*S*,6*R* similarly to **2**. Database searches resulted in one match for **4**, which refers to a member of a depsipeptide patented because of submicromolar Kallikrein 7 inhibitory activity,[93,95] whereas **1-3** comprise new metabolites.

Protease Inhibition Assays.
Ahp-containing cyclic depsipeptides frequently possess serine protease inhibitory activity; the amino acids adjunct to Ahp determine the enzyme specificity.[89] With Phe and Leu in these positions, the crocapeptins belong to the class of chymotrypsin inhibitors. Consequently, their inhibitory activity against Chymotrypsin was determined and resulted in IC_{50} values of 0.1 μM for **1-3** and 0.2 μM for **4**.

Analysis of the Crocapeptin A Gene Cluster and Proposed Biosynthesis.
Because of the apparent similarity to the nonribosomally synthesized cyanopeptolins, partial genome data of crocapeptin A producing strain Cm c5 were screened for the presence of NRPS genes to identify the corresponding biosynthetic gene cluster. Using prediction of substrate specificity of adenylation domains from sequence data[29] allowed one to reduce the number of target sequences to two contigs, containing an NRPS gene fragment with expected adenylation domain specificity for Thr-Leu-Pro-Phe and Tyr. Inactivation mutants for each contig were found unable to produce crocapeptin. Thus, both identified NRPS gene fragments were proven to be part of the crocapeptin biosynthetic gene cluster. To complete the cluster sequence, a cosmid gene bank covering the Cm c5 genome[22] was screened with DNA probes directed against parts of the two crocapeptin NRPS gene fragments, which are present on cosmids B:E7, C:O19, and D:J18. Cosmid end-sequencing, primer walking, and inverse PCR were applied to complete the gene cluster sequence. In Figure 2.2, a schematic drawing of the gene cluster is shown, with the closest homologues to the identified ORFs listed below. The NRPS gene *cpnD* is 24.5 kb in size and forms an operon with the hypothetical protein gene *cpnE* and cytochrome P450 encoding gene *cpnF*. Upstream of the NRPS gene, a putative serine protease gene *cpnA* and two ABC transporter genes *cpnB* and *cpnC* are located. Downstream of *cpnF*, a protein kinase is encoded by gene *cpnG* on the opposite strand. Because of their arrangement and predicted gene product function, ORFs *cpnA* to *cpnG* were chosen for knockout experiments to test their involvement in crocapeptin biosynthesis, transport, or regulation.

The catalytic domain sequence of CpnD and the substrate specificity of its adenylation domains were obtained from the prediction based on the amino acid sequence, according to the NRPS predictor.[29] As depicted in Figure 2.3, biosynthesis is expected to start with incorporation of glutamine by domain A1, to which one of the observed short chain fatty acids is bound by domain C1, creating the crocapeptins A1-A3. The lipopeptide chain is then elongated by incorporation of threonine, leucine, proline, phenylalanine, tyrosine (which becomes N-methylated in trans), and valine. The TE domain finally releases precrocapeptin by intramolecular lactonization via the hydroxyl group of threonine and the carboxy-terminus of valine. Free precrocapeptin is hydroxylated by CpnF at the proline residue, which is then transformed to the bioactive Ahp-containing crocapeptin. This biosynthesis scheme implies that each crocapeptin derivative must have its distinct precursor, as the variable short fatty acid chain is introduced during the start of biosynthesis. No function could be assigned to the gene product of *cpnE*, which has high sequence similarity to ORFs annotated as hypothetical proteins in various prokaryotic genomes. The closest sequence homologue with an assigned protein function is an Aha1 activator protein



ORF	closest homologue	source organism
cpnA	peptidase S8/S53	[uncultured candidate division OP1 bacterium]
cpnB	ABC-type siderophore export system	[*Plesiocystis pacifica* SIR-1]
cpnC	ABC-type siderophore export system	[*Plesiocystis pacifica* SIR-1]
cpnD	NRPS	[*Streptomyces acidiscabies* 84-104]
cpnE	hypothetical protein	[*Frankia* sp. CcI3]
cpnF	cytochrome P450	[*Streptomyces acidiscabies* 84-104]
cpnG	protein kinase	[*Sorangium cellulosum* 'So ce 56']

Figure 2.2: The crocapeptin A biosynthetic gene cluster from *Chondromyces crocatus* Cm c5. The genes responsible for biosynthesis are indicated as dark blue arrows. The closest homologue found by BLAST sequence search against the nr database of the encoded protein sequence is shown below for each ORF.

of the Hsp90 ATPase.[96] The function of CnpE in crocapeptin biosynthesis is described below. The adenylation domain A4 of the crocapeptin assembly line, responsible for incorporation of the amino acid that would be converted to Ahp, is predicted to incorporate proline, as deduced from the motif "DVQFAAHVAK" by using the nonribosomal code.,[64] To verify this prediction, $^{13}C_5{}^{15}N$-proline was amended to Cm c5 wild-type liquid culture, and cultivation was continued for 2 days. After extraction of cells and supernatant, crocapeptins were detected by HR-MS. $[M + H]^+$, $[M + NH_4]^+$, and $[M-H_2O + H]^+$ molecular ion clusters of crocapeptins A1-A3 exhibiting a m/z that increased by 6 units were clearly detected, however only at a low percentage, making quantification hardly possible. In the negative control experiment, these ions were not detected. Thus, proline is indeed incorporated into the crocapeptins, and all proline N and C atoms remain in the metabolite after conversion of this residue to Ahp.

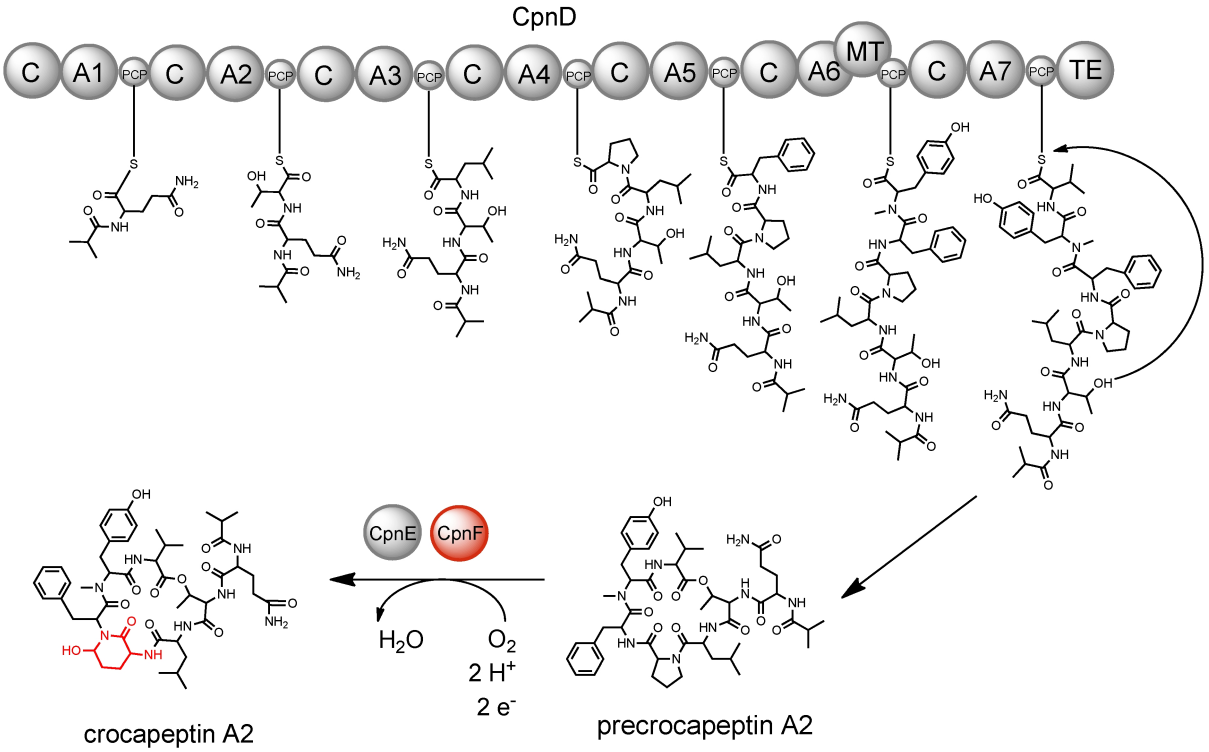


Figure 2.3: Biosynthesis of crocapeptin A2. The first condensation reaction defines the crocapeptin A derivative. The cyclic depsipeptide is released as precrocapeptin, which is then hydroxylated at the proline residue by the cytochrome P450 enzyme CpnF. Rearrangement of hydroxy-proline to Ahp is catalyzed by CpnE.

Gene Inactivation Studies in Cm c5.

To investigate the function of the genes identified, targeted insertional mutagenesis was applied. Mutant Cm c5 clones exhibiting hygromycin B resistance were grown in liquid culture for isolation of genomic DNA, and insertion was verified by PCR experiments (for genetic verification, see Supporting Information section 3). At least three independent verified mutants of genes *cpnA*, *cpnB*, *cpnC*, *cpnD*, and *cpnF* were obtained and grown in liquid culture to enable secondary metabolite analysis. For *cpnE* and *cpnG*, no mutants could be obtained. As two of the genes that were inactivated are predicted to encode ABC transporters, mutant cells and culture supernatant were extracted separately, to observe differences in crocapeptin content between cells and growth medium. Quantification of crocapeptin production was performed by dividing the summarized $[M + H]^+$ signal peak areas from MS analysis regarding crocapeptins A1-A3 by the chondramide B $[M + H]^+$ signal peak area in each chromatogram, because chondramid B production was known to be very stable in Cm c5 by experience.[20] UV data were not used because of overlapping peaks with other compounds. The average ratio of crocapeptins to chondramide was normalized to wild-type levels and is shown in Figure 2.4A.

As mentioned above, inactivation of the NRPS gene *cpnD* was found to abolish crocapeptin production. Inactivation of the putative serine protease encoding gene *cpnA* led to a 2-fold increase of crocapeptin levels in extracts from cells and medium, but high variation was observed within the individual cultures. Inactivation of the ABC transporter genes *cpnB* and *cpnC* led to cell extracts devoid of crocapeptins, while these metabolites were still present in the medium in similar amounts as compared to the wild type. Inactivation of *cpnF* did not abolish crocapeptin production completely, but reduced abundance to less than one-half of wild-type levels in the respective mutants. The effects of the inactivation experiments are discussed below.

Overproduction of Crocapeptins in Cm c5 by Promotor Insertion.

The tn5-derived *npt*-promotor *Pnpt* had been shown to enable overexpression of genes for sec-

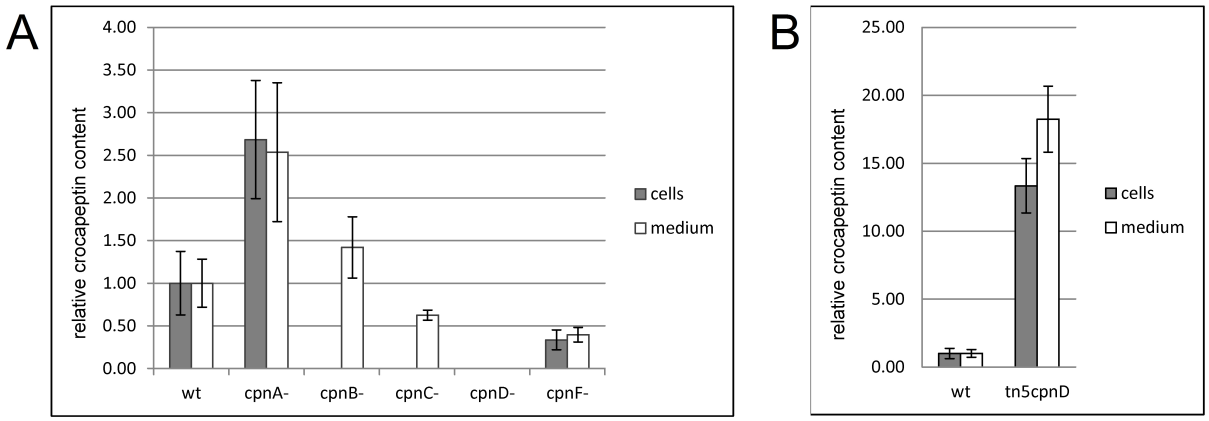


Figure 2.4: (A) Effect of gene inactivation in Cm c5 by insertional mutagenesis. For consistency, relative crocapeptin A content of extracts from Cm c5 wild type and mutants is given relative to the chondramide content in extracts of cells and growth medium. Crocapeptin A content was related to chondramid B content in each extract by division of the integrated MS-signal intensities of crocapeptins A by chondramid intensities, respectively. (B) Comparison of crocapeptin production in Cm c5 wild type and P*npt-cpnD* mutant cells. Extracts of cells and medium showed a 10-fold increased crocapeptin content in the mutant.

ondary metabolite production of strain Cm c5 previously, where replacing the native thuggacin promotor by *Pnpt* led to significantly increased levels of thuggacins.[97] To test whether a similar effect could be induced with crocapeptins, the promotor was placed in front of the NRPS gene *cpnD* by insertional mutagenesis, leading to a disrupted short copy of *cpnD* and a *Pnpt*-driven intact *cpnD* in the mutant. Extracts of mutant cells and medium were prepared and analyzed as described previously, and crocapeptin content was measured as described above. An increase of approximately 10-fold crocapeptin content was found in the cells and 15-fold in the medium, as shown in Figure 2.4B. As compared to the wild type, the overproducer mutants did not show altered growth or swarming behavior in liquid culture or on agar plate.

Detection of Precrocapeptins in *cpnF*$^-$ Mutants of Cm c5.

The proposed crocapeptin biosynthetis scheme implies that the Ahp-heterocycle is derived from a proline residue, which is converted to Ahp either *in trans* or after release of depsipeptide precursor molecules by the tailoring cytochrome P450 oxidase CpnF. These crocapeptin precursors would be expected to have a L-proline residue instead of Ahp in the respective position. Extracts of Cm c5 *cpnF*$^-$ mutants, in which the tailoring P450 was inactivated, were therefore screened for masses that would correspond to the postulated precursors.
Indeed, new $[M + H]^+$ mass signals were detected in the extracts of the growth medium from *cpnF*$^-$ mutants. These masses were later found to be detectable in the supernatant of wild-type cultures, but only in trace amounts. Because of their retention times and m/z values, the new compounds were designated precrocapeptins pA1 m/z = 919.49, pA2 m/z = 933.52, and pA3 m/z = 947.52. To verify the identity of the newly observed mass signals, Cm c5 *cpnF*$^-$ was grown in liquid culture, and the putative precursors were isolated from culture supernatant. The purified precrocapeptin A2 (pA2) with a mass of 932.3 Da was analyzed by NMR, which confirmed the expected lipopeptide lactone, corresponding to crocapeptin A2 with L-proline replacing the Ahp heterocycle (Figure 2.5).

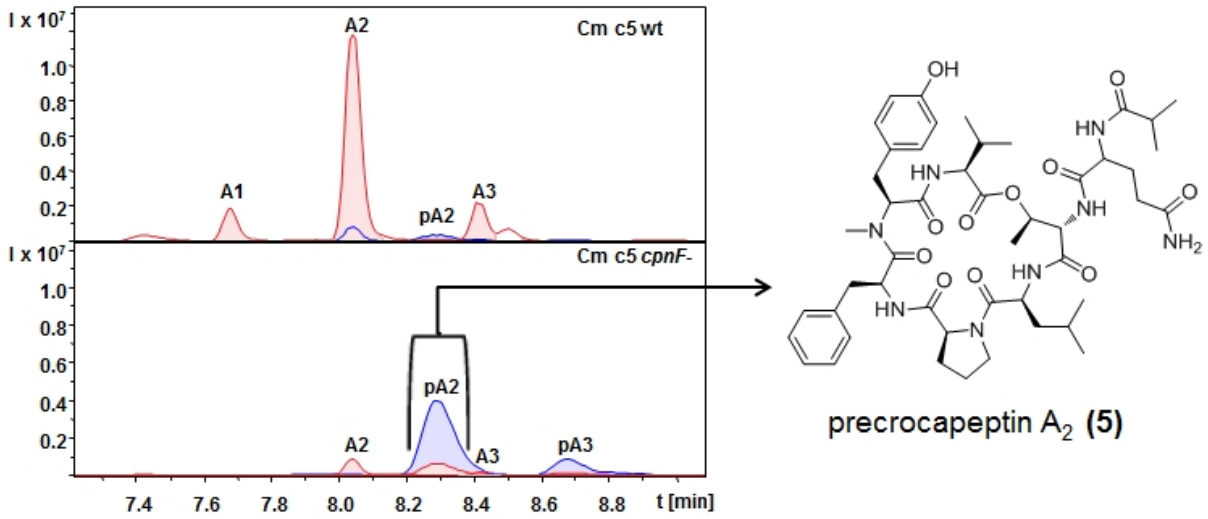


Figure 2.5: Detection and structure elucidation of precrocapeptin A2 **(5)**. Extracted ion chromatogram for crocapeptin masses (red areas) and precrocapeptin masses (blue areas). The putative crocapeptin precursors were found in extracts of *cpnF* mutants of strain Cm c5. Subsequent isolation and NMR analysis of 5 confirmed the structure shown on the right, a cyclic depsipeptide exhibiting L-proline incorporated instead of Ahp.

Establishing the Role of CpnE and CpnF in the Conversion of Precrocapeptin to Crocapeptin *in Vitro*.

To elucidate the reaction mechanism and the role of the gene product of *cpnE* and *cpnF*, corresponding proteins CpnE and CpnF were overexpressed in *E. coli* and purified, as shown in Figure 2.6A. CO-difference spectrometry was applied to verify that CpnF is expressed as active Cytochrome P450, but revealed that only a fraction of the purified CpnF is active holoenzyme, while a larger fraction is apoenzyme, as indicated by the peak at 420 nm (Figure 2.6B). Difference spectrometry of CpnF coincubated with precrocapeptin A2 led to concentration-dependent type I binding, as indicated by the reduction of the Soret absorption band at 420 nm, as shown in Figure 2.6C. This confirms that that free pA2 is indeed a substrate for CpnF.
Cytochrome P450 enzymes catalyze electron-dependent redox reactions, typically hydroxylations and epoxidations, and thus depend on electron carrier proteins to deliver the necessary electrons. Because the corresponding native electron carriers for CpnF are not known, the versatile recombinant ferredoxin Adx 4-108 was utilized instead, as it had been shown to be able to interact with several myxobacterial CYPs.[97] [98] To transfer electrons to CpnF, Adx was used together with its native electron donor, adrenodoxin reductase, a flavoenzyme that reduces oxidized adreno-

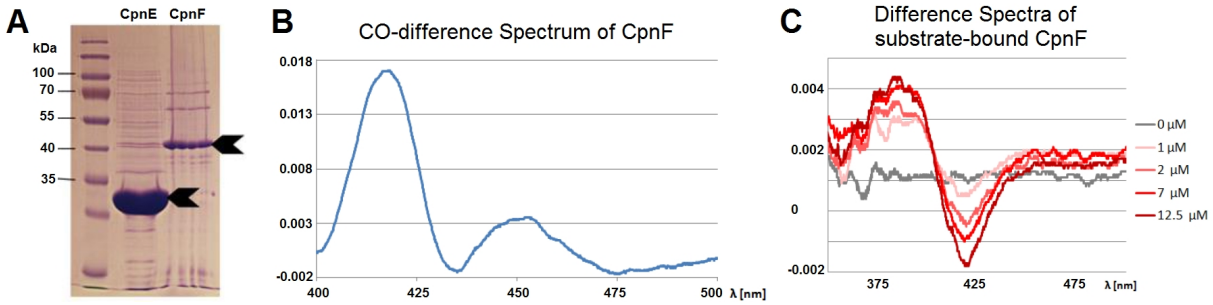


Figure 2.6: (A) SDS-PAGE of purified proteins CpnE and CpnF. The target proteins are indicated by black arrows. (B) CO difference spectrum of purified CpnF. The peak characteristic for cytochrome P450 enzymes at 450 nm can be observed. (C) Difference spectra of CpnF with increasing concentrations of precrocapeptin A2 show type I binding.

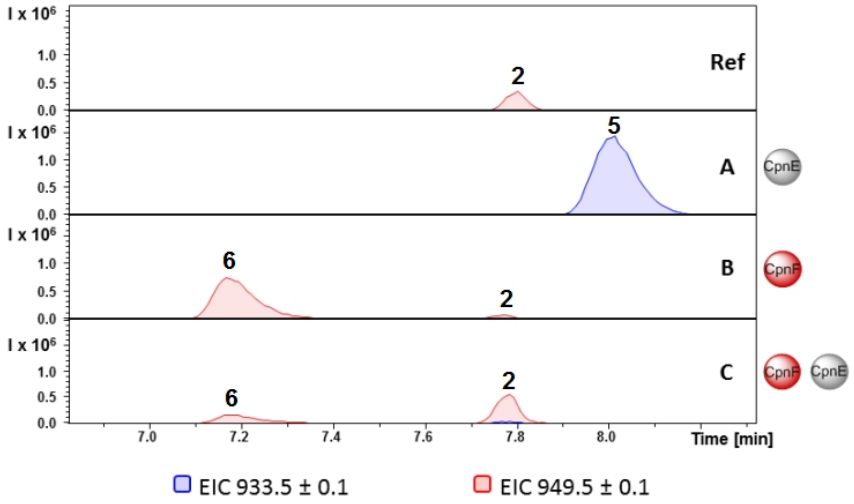


Figure 2.7: In vitro conversion of precrocapeptin A2 (peak 5) to crocapeptin A2 (peak 2). The result of incubation of precrocapeptin A2 (933.5 Da) with different enzyme combinations A, B, and C is compared to a reference of crocapeptin A2 (949.5 Da). Extracted ion chromatograms for the masses 933.5 Da (blue) and 949.5 Da (red) are shown. Peak 6 is a crocapeptin isomer that forms mostly in the absence of CpnE.

doxin by oxidizing NADPH to $NADP^+$. To keep NADPH levels elevated during the reaction, a NADPH regeneration system consisting of glucose-6-phosphate and glucose-6-phosphate dehydrogenase was used, which restores NADPH from $NADP^+$ by conversion of glucose-6- phosphate to 6-phosphoglucono-δ-lactone.[99] Purified CpnE was also included in this assay, to test its involvement in this conversion reaction. Enzymes CpnE and CpnF were tested separately and in combination in three separate reactions A, B, and C, each containing precrocapeptin A2 as substrate, as shown in Figure 2.7. Reaction A contained only CpnE, B contained only CpnF, and in C both proteins were present. After incubation for 30 min, the extracts were subjected to MS analysis and compared to a reference sample of crocapeptin A2 **(2)**. Unaltered substrate precrocapeptin A2 was detected in reaction A. Complete hydroxylation of precrocapeptin A2 took place in reactions B and C where two isomeric products were detected with m/z 949.5 by ESI-MS with retention times of 7.2 and 7.8 min. Comparison with the reference confirmed the identity of the peak at 7.8 min as crocapeptin A2. The role of CpnE in the conversion is shown by comparing the ratio of crocapeptin A2 (peak 2) and the presumed intermediate (peak 6) in reactions B and C. Without CpnE, mainly the isomer with a retention time of 7.2 min was formed, while in the presence of CpnE a much more efficient conversion into crocapeptin A2 was observed. Nevertheless, in the course of several days, a slow rearrangement of the supposed intermediate into crocapeptin A2 could also be observed.
Using pA3 as substrate for in vitro hydroxylation leads to the same effects, that is, one intermediate peak corresponding to peak 6 in Figure 2.7 and one product peak. For corroboration of the identity of the rather instable intermediate, the in vitro hydroxylation reaction was repeated in a larger scale, using 400 μg of pA3 as substrate. The reaction products were separated chromatographically (HPLC) and analyzed by NMR individually. From this analysis, formation of a 5-OH proline moiety is strongly indicated, as discussed below. Using the same type of analysis, also nonenzymatic conversion from 5-OH-proline to Ahp could be proven, as shown in Supporting Information Figure S7-4.

2.3 Discussion

The Myxobacterium *Chondromyces crocatus* Produces Potent Protease Inhibitors: The Crocapeptins.

The crocapeptins, isolated from the myxobacterium *Chondromyces crocatus*, were found to belong to the diverse compound class of cyanopeptolin-like peptides, and mutagenesis experiments in strain Cm c5 confirmed their nonribosomal origin. The Ahp-heterocycle is a structural feature crucial for protease inhibiting bioactivity of the cyanopeptolins. Here, the previously unknown biosynthesis of this structural part was elucidated by mutagenesis and biochemical studies. These investigations led to the identification of the precrocapeptins as precursor molecules released from the NRPS assembly line, featuring an L-proline instead of the Ahp heterocycle. These precursors are then converted into the bioactive crocapeptins by the hydroxylation via the tailoring P450 CpnF and the Aha1-domain containing helper protein CpnE. Comparison of the crocapeptin gene cluster with published cyanopeptolin gene clusters from cyanobacteria shows that the cyanobacterial gene clusters contain neither a predicted P450 gene nor a gene with homology to *cpnE*. Furthermore, in the cyanopeptolin NRPS genes, the predicted substrate for the A-domain corresponding to the Ahp-position is not proline. Instead, all cyanopeptolin A-domains at the Ahp-position have the same DVENAGVV motif, of which the substrate is not known yet. Consequently, further studies are required to unravel the biosynthesis of the Ahp herterocycle in cyanobacteria.
The function of the ORFs in the crocapeptin gene cluster was investigated by knockout experiments. The putative serine protease gene *cpnA* is considered to be part of the gene cluster, possibly involved in regulatory mechanisms or otherwise linked to crocapeptin production, because of elevated crocapeptin content in the respective mutant. The results of *cpnB* and *cpnC* inactivation are contrary to the initially expected function as crocapeptin export proteins and point toward a function as importer proteins. Another possibility is a rather complex export process involving the bacterial secretory pathway, which cannot be investigated by simple separation of cell clumps and media. We suggest that crocapeptins are actively secreted into the extracellular slime of the *Chondromyces* biofilm by CpnB and CpnC to serve as feeding deterrent, whereas crocapeptin dissolves in the medium by diffusion. CpnF mutants are still able to produce crocapeptins, so either the disrupted *cpnF* gene is still translated into a protein with residual activity, or another P450 in the genome is complementing its function.

CpnE Assists in Ahp Formation after Hydroxylation by Cytochrome P450 Enzyme CpnF.

Because of the small size of the gene and its location, inactivation of *cpnE* via insertional mutagenesis was not possible, and therefore the function of a hypothetical CpnE could not be elucidated by mutagenesis. A *cpnE* mutant is expected to exhibit a phenotype similar to a *cpnF*-mutant, because transcription of *cpnF* would be impaired due to the pSUP_hyg insertion. However, because *cpnE* was found encoded between *cpnD* and *cpnF*, it was supposed to play a role in crocapeptin biosynthesis. To shed light on this hypothesis, *cpnE* and *cpnF* were heterologously expressed in *E. coli* and purified.
Ahp formation is expected to occur as shown in Scheme 1. Hydroxylation at C5 of the proline residue in precrocapeptins gives rise to a cyclic hemiaminal corresponding to peak 6 in Figure 2.7. This hydroxy-proline derivative is then converted to Ahp via a glutamate-semialdehyde intermediate, which is highly reactive and forms the more stable six-membered heterocycle via the amide bond of the adjacent phenylalanine residue. From synthetic efforts of the cyanobacterial congener symplocamide, spontaneous Ahp formation has been reported starting from a L-glutamic acid semialdehyde residue,[90] which can be obtained by hydroxylation of proline at C5 and subsequent ring-opening. Indeed, a hydroxylated precrocapeptin with either hydroxyproline or glutamic-acid-semialdehyde does exhibit a mass identical to that of the respective

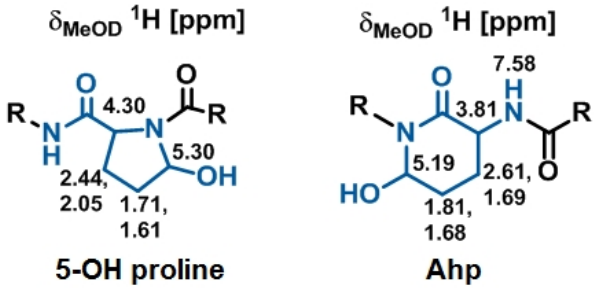


Figure 2.8: ^{1}H NMR shifts in cyclic hemiaminals after in vitro hydroxylation of proline in precrocapeptin A3: Ahp-containing crocapeptin A3 (right) and its unstable isomeric intermediate containing 5-OH proline instead (left).

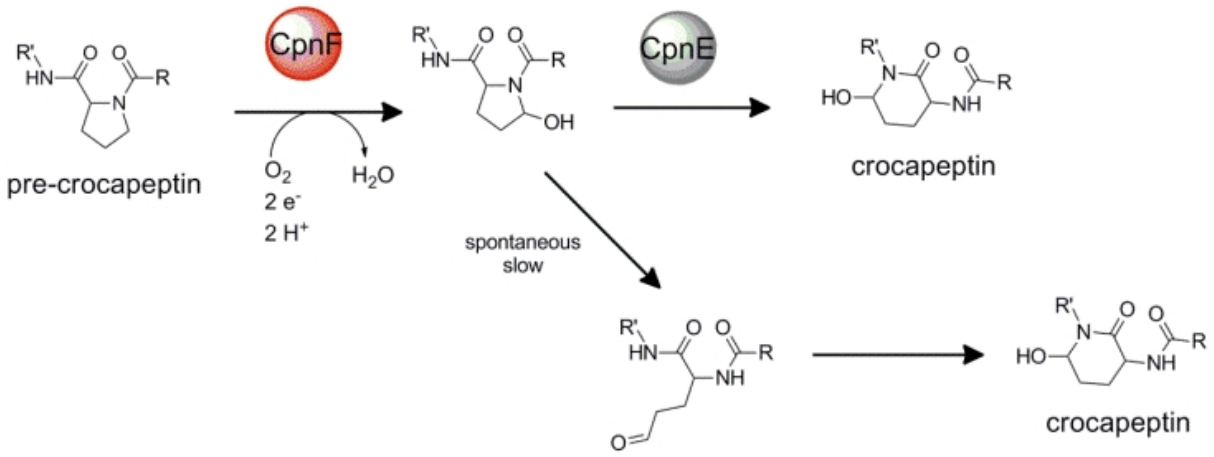


Figure 2.9: Proposed Formation of the Amino-hydroxy-piperidone (Ahp) Heterocycle

Ahp-containing crocapeptin. Hydroxylation of proline at positions other than C5 is not expected, as this would not lead to a hemiaminal able to undergo ring-opening. Proline residues with hydroxylations at C3 and C4 have been described for the lantibiotic microbiosporin, and hydroxylation at these positions does not lead to Ahp formation.[100]

Activity of Purified Recombinant Proteins.

CpnE and CpnF were tested with isolated precursor precrocapeptin A2 (5) as substrate. Hydroxylation was achieved in vitro with CpnF, as shown by HPLC-MS analyses of the reaction extracts, as shown in Figure 2.7. Incubation with CpnF alone mostly leads to an unstable isomer (peak 6) exhibiting shorter retention time, which slowly converts to crocapeptin A2 (peak 2). Incubation of precrocapeptin A2 with CpnE alone does not indicate any reaction. However, coincubation of precrocapeptin with CpnF and CpnE results mostly in crocapeptin formation, while formation of the unstable isomer (peak 6) is detected in much smaller quantities.

In vitro reactions of CpnF with precrocapeptin A3 as substrate were scaled up to isolate the presumed intermediate 5-hydroxy-proline form of crocapeptin A3 for structure elucidation. First, the mixture of both isomers was used for NMR spectroscopic analysis; afterward, the isomers were separated and analyzed individually (Supporting Information Figure S7-1). The NMR spectrum of isomeric crocapeptin A3 clearly shows shifted ^{1}H signals in the area of interest as compared to crocapeptin A3 (Ahp-ring, Figure 2.8). A chemical shift of δ 3.81 can be detected for the α-proton in Ahp. The signal of the α-proton exhibits TOCSY correlations with two CH2-groups at δ 2.61, 1.69 and δ 1.81, 1.68, respectively, additional to correlating with a CHOH-group at δ 5.19 and a NH at δ 7.58 (Supporting Information Figure S7-2). In contrast,

the α-proton of the isomer shows a chemical shift of 4.30 with TOCSY correlations to CH_2 groups at δ 1.71, 1.61 and δ 2.44, 2.05 (Supporting Information Figure S7-3). Correlation to the CHOH group at δ 5.30 cannot be detected with the applied parameters, but integrity of the spin system can be assigned via correlations of the signal of CHOH with both CH_2 groups and the α-proton. A strong indication for the presence of the hydroxy-proline is the missing correlation to an NH proton in this spin system in addition to the existence of the NH-signal belonging to the phenylalanine spinsystem at δ 8.86. Furthermore, the shifts of N-Me-Tyr aromatic protons in the intermediate (δ 7.09, 6.50) are exactly the same as for precrocapeptin (Supporting Information Table S4-4), whereas these proton signals in the product are significantly shifted to δ 7.14 and 6.85, indicating a conformational change of the structure probably due to rearrangement from a five- to a six-membered ring. Thus, Ahp formation is likely initiated by opening of the five-membered ring hemiaminal to glutamic-acid semialdehyde (Figure 2.9). Efforts to indirectly detect the glutamic-acid semialdehyde intermediate were carried out by incubating the reaction mixture with an excess of ethanolamine to capture the aldehyde. A fraction of the short-lived semialdehyde would then favor nucleophilic addition to the primary amino group of the ethanolamine over the adjacent peptide bond. A mass signal corresponding to the imine form of the ethanolamine adduct could indeed be detected, but is also formed when pure crocapeptin is treated with ethanolamine. We therefore propose that the crocapeptins exist in equilibrium between the six-membered Ahp and the five-membered 5-OH-proline heterocycles, with Ahp being the more preferred state. Incubation of precrocapeptin with ethanolamine did not result in any detectable products. Details for the aldehyde trapping experiments are provided in Supporting Information section S6.
Because the glutamic acid intermediate in the crocapeptin conversion is highly reactive, a function of CpnE, which is shown here to significantly enhance Ahp formation, could be to promote formation of Ahp and shield the reactive glutamic-acid semilaldehyde intermediate from unfavorable cross-reactions. Whether CpnE interacts directly with CpnF or with the free substrate remains to be elucidated.

The ability of CpnF to receive electrons from the broad-range ferredoxin Adx4-108 also disclosed a route for biotechnological production of Ahp-containing molecules for targeted protease inhibitor design. From the cyanobacterial cyanopeptolins, a large variety of structures with different target specificities is known, depending mostly on the amino acids incorporated at the positions adjacent to Ahp.[88] Further studies should investigate the substrate range of CpnF and CpnE to investigate whether this enzyme pair could be used to generate Ahp-containing peptides of a specific and even non-natural structure, to create highly specific/selective protease inhibitors. For heterologous expression of the compounds, coexpression of Adx4-108 in the same host might be an option to obtain a functional cytochrome in a host that naturally has no compatible ferredoxin for CpnF.

2.4 Experimental Procedures

Bacterial Strains and Growth Conditions. *C. crocatus* Cm c5 was cultivated in 70 L of Pol 0.3 medium (Probion 3 g/L, soluble starch 3 g/L, [$MgSO_4 7 H_2O$] 2 g/L, [$CaCl_2 2 H_2O$] 0.5 g/L) in a biofermenter Biostat UE100 (Braun Melsungen, Melsungen, D) at 30C for 11 days; *C. crocatus* Cm c2 was cultivated in 100 L of Pol medium in a biofermenter P150 (Bioenineering AG, Wald, CH) at 30C for 10 days. After 3 days, 1% Amberlite XAD-16 was added to the fermentation broths. For metabolite analysis of Cm c5 mutants, 1 mL of cell clumps was used to inoculate 20 mL of Pol 0.3 medium with 50 mg/L Hyg. At least three verified clones from each inactivation construct were used. Cm c5 wild-type cells were cultivated without Hyg. After 4 days, cell clumps and medium supernatant were separated. Medium supernatant was extracted with 25 mL of ethyl acetate, and cell clumps were extracted with 10 mL of ethyl acetate and 40

mL of methanol. Extracts were dried *in vacuo* and dissolved in 1 mL of methanol.

Isolation and Purification of Crocapeptins. XAD-16 and cell mass of *C. crocatus* Cm c5 were harvested by centrifugation. Cells were separated from XAD-16 by flotation and discarded. The XAD was washed with water and 50% methanol, and subsequently eluted by methanol and acetone. All fractions were analyzed by HPLC-UV-MS for the presence of target compounds. Consequently, the methanol extract was subjected to solvent partition using 85% aqueous methanol, which was extracted twice with heptane. Subsequently, the aqueous methanol was adjusted to 70% methanol and extracted twice with dichloromethane. The dichloromethane fraction containing the main amount of **1-3** was fractionated by RP MPLC [column 480 x 30 mm, ODS / AQ C18 (Kronlab), gradient 30-100% methanol in 60 min, flow 30 mL/min, UV peak detection at 210 nm]. Fractions containing 1-3 were combined and further purified by preparative RP HPLC [column 250 x 21 mm, VP Nucleodur 100-10 C18 EC, gradient 40-80% methanol in 25 min, 50 mM sodium acetate, flow 20 mL/min]. Finally, preparative RP HPLC [column 250 x 21 mm, VP Nucleodur C18 Gravity 5 μm, gradient 45-60% methanol in 25 min, 0.2% acetic acid, flow 20 mL/min] provided 6.2 mg of **1**, 8.9 mg of **2**, and 6.7 mg of **3**.

XAD-16 and cells of *C. crocatus* Cm c2 were washed stepwise with water (2 L) and 30% methanol (4 L), and extracted with methanol (8 L). The crude methanol extract was evaporated and partitioned three times between ethyl acetate and water. After evaporation of the ethyl acetate, the residue was dissolved in 85% aqueous methanol and extracted twice with heptane. The methanol extract was fractionated in two batches by RP MPLC (column 480 x 30 mm, ODS/AQ C18 (Kronlab), gradient 37-100% methanol in 60 min, flow 30 mL/min, UV detection 210 nm). The fraction containing **4** was separated by preparative RP HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5 μm, gradient 45-60% methanol / 0.2% acetic acid in 25 min, flow 20 mL/min) providing 10.8 mg of **4**.

Isolation and Purification of Precrocapeptins. Cm c5 mutant *cpnF*$^-$ was grown in shake flasks with Pol 0.3 medium containing Hyg. Medium was replaced regularly, and used medium was extracted with Amberlite XAD-16 adsorber resin under sterile conditions. Pooled XAD batches were extracted with ethyl acetate, dried *in vacuo*, and dissolved in methanol. The crude extract was separated on a Sephadex DAE20 column (80 cm length, 2.5 cm diameter) using methanol, and precrocapeptin containing fractions were identified by LC-MS. Precrocapeptin was purified by RP-HPLC [HPLC system (Dionex) Famos autosampler, P680 pump, TCC100 thermostat, and PDA100 detector; Phenomenex Luna C18 column, 250 x 4.6 mm, 5 μm dp; solvent A = H2O, solvent B = ACN, gradient from 20% B to 60% B in 30 min; flow rate 2.5 mL / min; temp 30C; UV detection 254 and 280 nm]. The sample was injected by μL-pick-up technology with a water / methanol (50:50 v/v) mixture as supporting solvent. Up to 50 μL of sample was injected for manual fraction collection. For precrocapeptin A2 NMR data, see Supporting Information Table S3. Fractions containing 0.3 mg of precrocapeptin A1, 0.9 mg of precrocapeptin A2, and 0.3 mg of precrocapeptin A3 were obtained.

General Experimental Procedures. Optical rotations were determined with a Perkin-Elmer 241 instrument; UV spectra were recorded with a Shimadzu UV-vis spectrophotometer UV-2450; IR spectra were measured with a Spectrum 100 FTIR spectrometer (Perkin-Elmer); NMR spectra were recorded with Bruker AM 300 (^{1}H 300 MHz, ^{13}C 75 MHz), Bruker ARX 600 (^{1}H 600 MHz, ^{13}C 150 MHz), and Bruker Ascend 700 with a 5 mm TXI cryoprobe (^{1}H 700 MHz, ^{13}C 175 MHz) spectrometers; HRESIMS mass spectra were obtained with an Agilent 1200 series HPLC-UV system combined with an ESI-TOF-MS (Maxis, Bruker) [column 2.1 x 50 mm, 1.7 μm, C18 Acquity UPLC BEH (Waters), solvent A, H_2O + 0.1% formic acid; solvent B, AcCN + 0.1% formic acid, gradient, 5% B for 0.5 min increasing to 100% B in 19.5 min, maintaining 100% B for 5 min, FR = 0.6 mL min-1, UV detection 200-600 nm].

Crocapeptin A1 (1). Colorless, amorphous powder, $[\alpha]_D$-34 (c 0.5, MeOH); UV (MeOH) λ_{max} (log ϵ) 230 (sh), 277 nm (3.49); IR (neat) 3380.9, 2962,2, 2934.1, 1736.0, 1655.4, 1534.4, 1518.1, 1445.4, 1384.5 cm^{-1}; ^{1}H NMR, ^{13}C NMR data (DMSO$-$d$_6$), see Supporting Information Table S3; HR-ESI-MS m/z 935.4868 $[M^+H]^+$ (calcd for $C_{47}H_{67}N_8O_{12}$, 935.4873).

Crocapeptin A2 (2). Colorless, amorphous powder, $[\alpha]_D$-31 (c 0.5, MeOH); UV (MeOH) λ_{max} (log ϵ) 229 (sh), 277 nm (3.45); IR (neat) 3401.0, 2961,5, 2931.4, 1735.0, 1659.7, 1540.4, 1517.4, 1453.1, 1384.5 cm^{-1}; ^1H NMR, ^{13}C NMR data, COSY, HMBC, ROESY data (DMSO−d_6), see Supporting Information Table S1; HR-ESI-MS m/z 949.5032 $[M^+H]^+$ (calcd for $C_{48}H_{69}N_8O_{12}$, 949.5027).
Crocapeptin A3 (3). Colorless, amorphous powder, $[\alpha]_D$-25 (c 0.3, MeOH); UV (MeOH) λ_{max} (log ϵ) 228 (sh), 275 nm (3.43); IR (neat) 3376.6, 2961.7, 2932.6, 1735.4, 1663.7, 1535.1, 1518.2, 1452.3, 1384.4 cm^{-1}; ^1H NMR, ^{13}C NMR data (DMSO−d_6), see Supporting Information Table S3; HR-ESI-MS m/z 963.5186 $[M^+]$+ (calcd for $C_{49}H_{71}N_8O_{12}$, 963.5186).
Precrocapeptin A2 (5). ^1H NMR, ^{13}C NMR, COSY, HMBC data (MeOH−d_4), see Supporting Information Table S4; HR-ESI-MS m/z 933.5070 $[M^+H]^+$ (calcd for $C_{48}H_{69}N_8O_{11}$, 933.508).
Amino Acid Stereochemistry Determination. Crocapeptin A1-A3 and B were hydrolyzed in 6 N HCl at 90C for 17 h, conditions that resulted in the conversion of Gln to Glu. The hydrolysate was evaporated to dryness and dissolved in H_2O (100 μL). 1 N $NaHCO_3$ (20 μL) and 1% 1-fluoro-2,4-dinitrophenyl-5-L-alaninamide (100 μL in acetone) were added, and the mixture was heated at 40C for 40 min. After being cooled to room temperature, the solutions were neutralized with 2 N HCl (20 μL) and evaporated to dryness. The residues were dissolved in MeOH and analyzed by LC-MS. Retention times in minutes of FDAA-derivatized amino acids were Glu 6.6; Thr 5.5; Leu 8.5; Phe 8.4; NMeTyr 5.8; Val 7.6 for 2 and Cit 5.5; Thr 5.5; Leu 8.5; Glu 6.6; Phe 8.4; NMeTyr 5.8; Ile 8.3 for 4. Retention times of the FDAA-derivatized authentic standards were L-Val 7.5, D-Val 8.5 m/z 368.123 $[M-H]^-$; L-Thr 5.4, D-Thr 6.2 m/z 370.100 $[M-H]^-$; L-allo-Thr 5.3, D-allo-Thr 5.7 m/z 370.100 $[M-H]^-$; L-NMeTyr 5.8, D-NMeTyr 8.4 m/z 446.131 $[M-H]^-$; LPhe 8.5, D-Phe 9.2 m/z 416.120 $[M-H]^-$; L-Glu 6.7, D-Glu 7.4 m/z 412.110 $[M-H]^-$; L-Leu 8.5, D-Leu 9.4 m/z 382.138 $[M-H]^-$; L-Ile 8.3, D-Ile 9.3 m/z 382.138 $[M-H]^-$; L-allo-Ile 8.4, D-allo-Ile 9.4 m/z 382.138 $[M-H]^-$; L-Cit 5.5, D-Cit 5.6 m/z 426.138 $[M-H]^-$. Oxidation of the compounds was performed as previously described.18 **2** and **4** (0.3 mg) were dissolved in glacial acetic acid (0.3 mL) and mixed with CrO_3 (2.0 mg). After being stirred at room temperature for 20 h, the reaction mixture was separated using a C18 SPE cartridge eluting with H_2O followed by MeOH. The resulting oxidized material from the MeOH fraction was treated with FDAA as described above; L-Glu was detected at t_R 6.6 min.
Mutagenesis of *Chondromyces crocatus* Cm c5. An internal fragment of the target gene was amplified by PCR and cloned into the suicide plasmid pSUP_hyg,20 yielding the respective pSUP_hyg derivative. After restriction analysis, the plasmid was introduced into the donor strain *E. coli* ET12567/pUB307. *E. coli* ET12567/pUB307 clones containing a pSUP_hyg plasmid derivative were selected on LB-Agar (low salt) at 37C supplemented with 50 mg/L Kanamycin sulfate (Kan), 30 mg/L Chloramphenicol (Cm), and 50 mg/L Hyg. Clones were grown in liquid LB medium (low salt), supplemented with the same antibiotic concentrations at 37C. Introduction of the plasmid into Cm c5 was done through biparental conjugation with *E. coli* strain ET12567/pUB307 as described elsewhere.6 Mutant clones of Cm c5 were selected on Pol03-Agar containing 100 mg/L Hygromycin B (Hyg) and 60 mg/L Spectinomycin, followed by cultivation in liquid Pol03 medium containing 50 mg/L Hyg. Hyg-resistant clones of Cm c5 were verified genetically by PCR as described in Supporting Information S1.
Heterologous Expression and Purification of CpnE and CpnF. CpnE was expressed in *E. coli* BL21 DE3 as an N-terminal his-tagged protein using the pET28b plasmid (Novagen). Cultivation, induction of expression, and protein purification were done according to standard procedures, and CpnE purity was confirmed by SDS-PAGE. Protein concentration was determined 10 mg/mL using Bradford measurement. CpnF was expressed in *E. coli* BL21 DE3 as a SUMO-Fusion protein using pSUMO plasmid (TaKaRa). Expression and purification were done according to protocols adapted for optimized heterologous expression of cytochromes.29 *E. coli* BL21 DE3/pSUMO_CYP clone was grown in 200 mL of TB medium supplemented with 50 mg/L Kan. As the culture reached OD600 = 0.6, 5-amino-levulinic acid was added to a final concentration of 1 mM, $FeCl_3$ was added to 0.5 mM, and the culture was incubated for 30 min

at 26 C before addition of IPTG to 0.2 mM. The culture was then further incubated for 63 h until cell harvest. Cleared *E. coli* lysate was separated with the ÄKTA HPLC system using a 5 mL HisTrap Nickel IMAC column. A linear gradient from 25 to 500 mM Imidazol was applied, and absorption was monitored at 280 and 430 nm. Fractions containing the absorption peak at 430 nm were subjected to SDSPAGE, and purity of SUMO-CpnF was judged by Coomassie staining. Fractions containing the 55 kDa SUMO-CpnF bands were pooled, concentrated, and cleaved with 20 u SUMO2-protease. The cleaved SUMO-tag was separated from CpnF in a second IMAC step. Buffer was exchanged four times with 10 mM potassium phosphate buffer pH 7.4, and protein was concentrated to a volume of 0.5 mL. Protein concentration was determined 7.5 mg/mL using Bradford measurement. SDS-PAGE of the protein fraction revealed a strong band at 40 kDa, and two distinct bands at 55 and 70 kDa. These bands were excised, and subjected to in-gel-digestion and MALDI-MS analysis according to standard protocols. The signal at 40 kDa is CpnF, the signal at 55 kDa is SUMO-CpnF, and the signal at 70 kDa is the *E. coli* chaperone DnaK.

In Vitro Conversion of Precrocapeptin to Crocapeptin. Precrocapeptin was isolated from *cpnF*$^-$ mutants as described above and used as substrate for purified CpnF. As electron transporters, recombinant bovine Adx 4-108 and adrenodoxin reductase were used. *In vitro* conversion was performed in 10 mM potassium phosphate buffer pH 7.4 or 50 mM Tris/HCl buffer at pH 8.0 in 0.5 mL volumes. Substrate concentration was 100 μM, and enzyme concentration was 0.25 μM for CpnF and CpnE, 2 μM for AdR, and 20 μM for Adx. To regenerate NADPH, 5 mM glucose-6-phosphate, 1 mM $MgCl_2$, and 2 u/mL glucose-6-phosphate dehydrogenase were in the reaction mixture. All ingredients were mixed on ice; the reaction was started by adding 200 μM NADPH and incubated at 30 C for 30 min. The reaction was stopped by adding 500 μL of ethyl acetate and strong mixing. After brief centrifugation, 50 μL of the organic phase was subjected to ESI-MS for detection of the reaction products.

Conclusions

The crocapeptins, cyanopeptolin-like cyclic depsipeptides with kallikrein and protease inhibiting bioactivity, have been isolated from the terrestrial myxobacterium *Chondromyces crocatus*. The previously unknown biosynthesis of the Ahp-heterocycle, the main structural feature crucial for bioactivity of this compound class, was elucidated biochemically. The formation of Ahp in crocapeptin A was achieved *in vitro* using purified precrocapeptin and recombinant tailoring enzymes CpnE and CpnF. The established ability of cytochrome P450 enzyme CpnF to receive electrons from the broad-range ferredoxin Adx4-108 opens a route for biotechnological production of Ahp-containing molecules for selective protease inhibitor design. A function could be assigned to the previously uncharacterized protein CpnE, encoding a SBPRCC-domain superfamily protein that catalyzes the rearrangement of a hydroxyproline residue to Ahp.

Acknowledgement

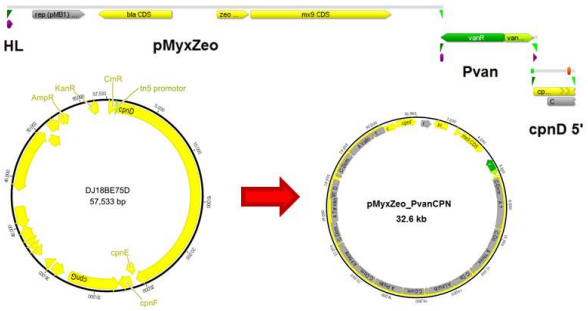
We thank Prof. Rita Bernhardt for supplying purified adrenodoxin and adrenodoxin-reductase, Prof. Uli Kazmaier for advice and discussion, Dr. Alberto Plaza for recording NMR spectra of precrocapeptin, Christel Kakoschke for NMR measurements, Dr. Giovanna Salamanca-Pinzon for help with difference spectrometry, Aileen Teichmann and Thomas Hoffmann for ESI-MS measurements. We thank Diana Telkemeyer for strain maintenance and Wolfgang Kessler and his team for cultivation. This work was supported by a return grant to F.S. from the German Academic Exchange service. Research in the laboratory of R.M. was supported by the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung and Forschung.

Supporting Information

Spectroscopic data, NMR spectra of crocapeptins, details for a chymotrypsin inhibition assay, genetic verification of Cm c5 mutants, and aldehyde-trapping experiments can be found in the Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org and attached in the appendix on page 141.

Chapter 3

Cloning of the Crocapeptin Gene Cluster for Heterologous Expression in the host *Myxococcus xanthus* DK1622



Author's efforts

The author's efforts regarding this chapter comprise all the experimental work, as well as the design of the experiments except for the intially applied cosmid fusion strategy.

Contributions by others

The cosmid fusion strategy was created with help from Silke Wenzel, who previously also constructed the pMyxZeo plasmid backbone. The project was supervised by Prof. Rolf Müller.

Abstract

The crocapeptin gene cluster from *Chondromyces crocatus* Cm c5 was assembled from two cosmid clones and the resulting cosmid further modified, aiming at expression in the host *Myxococcus xanthus* DK1622. The expression construct was transformed into the heterologous host, but no expression was observed.

3.1 Introduction

Crocapeptins are cyclic depsipeptides, produced by *Chondromyces crocatus* Cm c5 and Cm c2, and overproduction of crocapeptin A was achieved by insertion of the tn5-derived *npt*-promotor into the Cm c5 genome by insertional mutagenesis. Heterologous expression of crocapeptins is also aspired, to achieve more efficient overproduction in a fast-growing strain and to generate novel derivatives by mutasynthesis. The *Chondromyces crocatus* Cm c5 genome cosmid library was screened for sequences matching the crocapeptin biosynthetic gene cluster, which is covered by Cosmids B:E7 and D:J18, as shown in fig. 3.1. A cloning strategy was developed to combine these two cosmids via direct cloning. The resulting fused cosmid is then further modified according to the requirements of the heterologous host. As a first host, *Myxococcus xanthus* DK1622 was chosen.

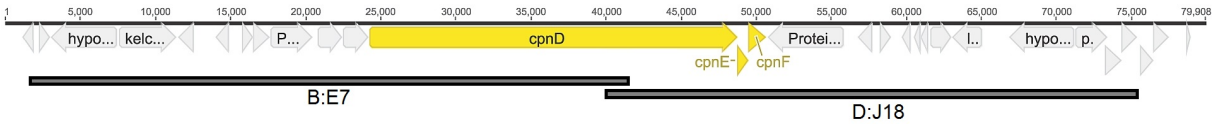


Figure 3.1: The crocapeptin biosynthesis cluster genes (yellow) and surrounding ORFs (grey) in the Cm c5 genome. The area covered by cosmids B:E7 and D:J18 is shown in boxes below.

Cloning strategy

For the cosmid stitiching, cosmid B:E7 was modified to allow later fusion with D:J18. For this modification, a linear construct was assembled by overlap PCR, consisting of the chloramphenicol resistance gene (*cmR*), the *npt*-promotor and homologous regions arms for site-directed recombination with B:E7 and D:J18. The *cmR* gene was amplified from the Plasmid pBluescript (Stratagene), the **tn5**-promotor and the first 40 bp of the crocapeptin NRPS gene *cpnD* were amplified from gDNA of Cm c5 mutant pSUP_hyg_tn5d, which has the *npt*-promotor upstream located in front of *cpnD*. The linear PCR product is shown in fig. 3.2. Homologous sequence parts HL and HR are used for recombination with B:E7.

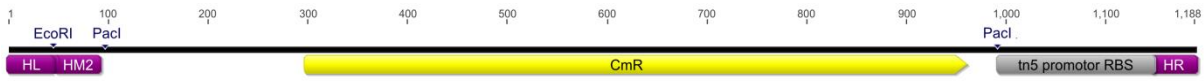


Figure 3.2: Linear fragment used for recombination with cosmid B:E7 to obtain the derivative BE75D.

After recombination, a shortened Cosmid B:E7 derivative BE75D is obtained, from which a part of the cosmid backbone, including the Kanamycin resistance gene, and the region upstream of *cpnD* is deleted and replaced by the *npt*-promotor. BE75D can restricted with *Eco*RI, to yield a linear fragment that is used for recombination with cosmid D:J18. The fragment from restriction of BE75D is shown in fig. 3.3.

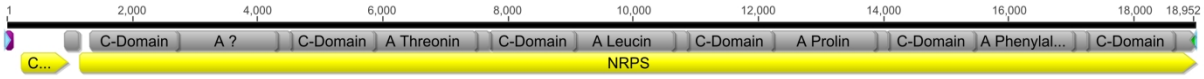


Figure 3.3: Linear fragment obtained from restriction of cosmid derivative BE75D. This fragment contains the N-terminal part of the NRPS gene *cpnD* and can be used for recombination with cosmid D:J18 using overlapping *cpnD*-sequence.

Recombination of the linear fragment with D:J18 works via sequence part HM2 (shown in fig 3.2) and the terminal part of the *cpnD* where B:E7 and D:J18 have a 1.6 kb sequence overlap. The resulting fused cosmid derivative DJ18BE75D provides resistance to Kanamycin, Ampicillin and Chloramphenicol and contains the complete *cpnD* gene followed by downstream genes. The cosmid map is shown in fig. 3.5A. To allow expression of the NRPS in *Myxococcus xanthus* DK1622, another Red-ET step was applied to further modify DJ18BE75D. By linear-to circular recombination with a linear capture vector, the cosmid backbone is replaced with the pMyxZeo backbone. Next to a zeocin resistance gene, this plasmid contains the integrase gene for stable integration of the plasmid into the the mx9 phage attachment site attB of the DK1622 genome. The linear capture vector is shown in fig 3.4 and was obtained by PCR from the plasmid pmyxozeo. Homologous sequence regions for recombination with DJ18BE75D were introduced via the oligonucleotide primers. DK1622 clones are selected by resistance to zeocin. The plasmid map of this expression construct is shown in fig. 3.5B.

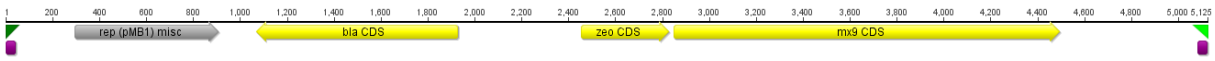


Figure 3.4: Linear capture vector for recombination with DJ18BE75D. The resulting plasmid pmyxzeo_cpn can be transferred into *M. xanthus* DK1622 via the mxp integrase gene. DK1622 mutant clones are selected by zeocin resistance.

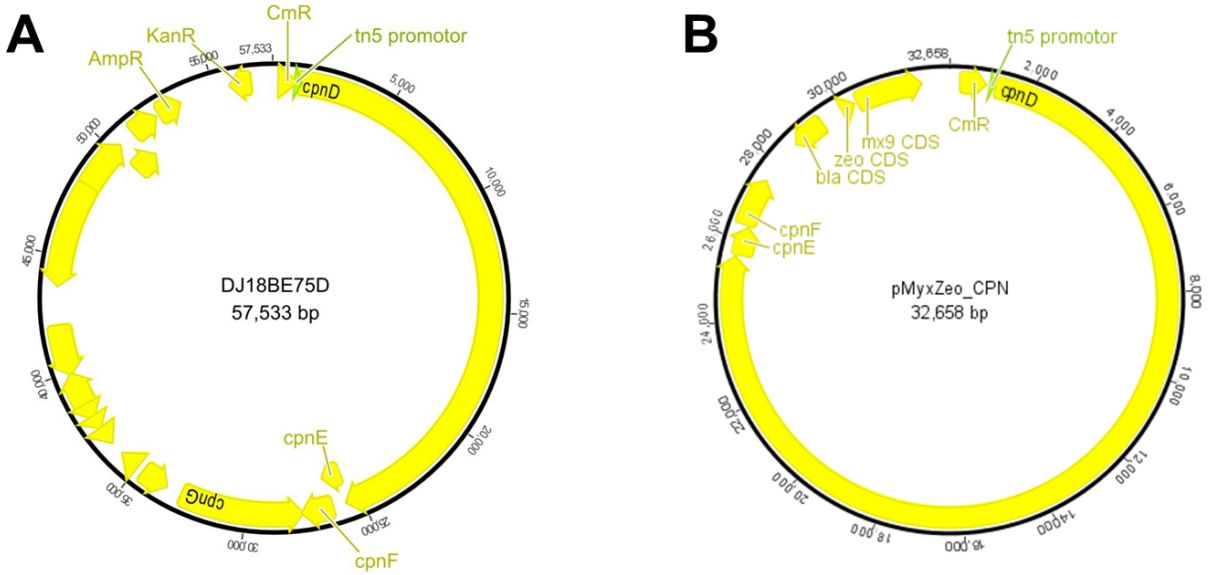


Figure 3.5: A: Plasmid map of fusion cosmid DJ18BE75D, which has the complete *cpnD*-gene, encoding the crocapeptin NRPS assembly line. **B:** Plasmid map of expression construct pmyxzeo_cpn for constitutive crocapeptin expression in *M. xanthus* DK1622.

3.2 Results and Discussion

Cosmid stiching and recombination

After the first recombination step of cosmid B:E7 with the lienar fragment shown in fig. 3.2, *E. coli* clones resistant to cm and amp were analyzed by restriction analysis. One BE75D cosmid with the correct restriction pattern was obtained, as shown in figure 3.6, left panel. The isolated cosmid DNA was restricted with *Eco*RI to yield the second linear fragment shown in fig. 3.3, which was separated from the cosmid electrophoresis and isolated by agarose gel extraction. This linear construct was then used for recombination with cosmid D:J18. Isolated cosmid DNA from *E. coli* clones resistant to cm, kan and amp was analyzed by restriction analysis. A clone showing the correct restriction pattern is shown in fig. 3.6 in the right panel.

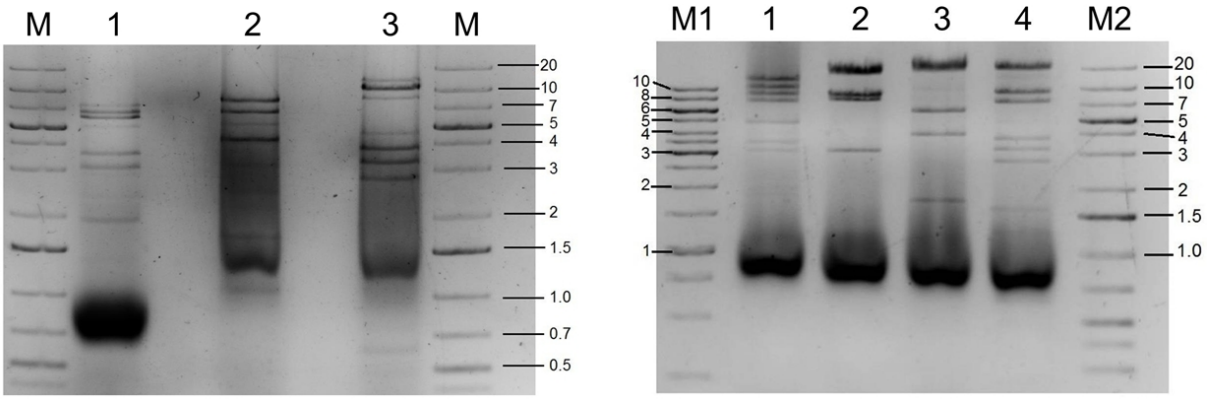


Figure 3.6: Agarose gels of restricted cosmid derivatives after recombination. DNA Fragment size is indicated in kb on the right side of each gel. **Left:** BE75D cut with: (1) *Bam*HI; (2) *Nco*I; (3) *Pvu*II; M: DNA ladder. **Right:** DJ18BE75D cut with (1) *Apa*LI; (2) *Bsp*HI; (3) *Hind*III; (4) *Pvu*II; M1, M2: DNA ladder. The cosmids show the expected restriction fragments.

Subsequently, cosmid DJ18BE75D was used for recombination with the capture vector to yield the expression plasmid pmyxzeo_cpn. After recombination, clones were selected for resitance to amp and zeocin and the plasmids analyzed by restricition. Clones with the correct restriction pattern were obtained and used for transformation into DK1622. Fig. 3.7 shows the agarose gels with the resctriction analysis.

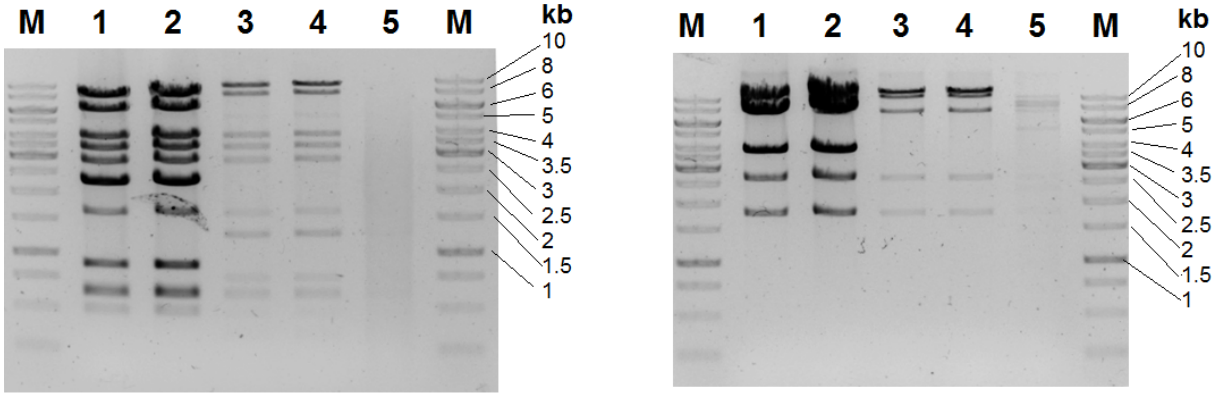


Figure 3.7: Agarose gels of restricted pmyxzeo_cpn clones. DNA Fragment size is indicated in kb on the right side of each gel. **Left:** pMyxzeo_CPN cut with PvuII; M: DNA ladder. Clones in lane 1 and 2 show the correct fragments. **Right:** pMyxzeo_CPN cut with *Sac*I; Clones in lane 1 and 2 show the correct fragments. M: DNA ladder.

Test-expression in *M. xanthus*

The two correct plasmids pmyxzeo_cpn 1 and 2 were transferred into *M. xanthus* DK1622. After 5 days, zeocin-resistant clones were obtained and inoculated in liquid culture. Integration of the *cpn*-gene cluster was verified by colony-PCR of boiled *M. xanthus* cells, using primers that bind regions on both ends of the crocapeptin NRPS gene *cpnD*. Both PCR reactions yielded the expected product F and H at 0.5 kb, indicating that *cpnD* has indeed integrated into the DK1622 genome, as shown in figure 3.8.

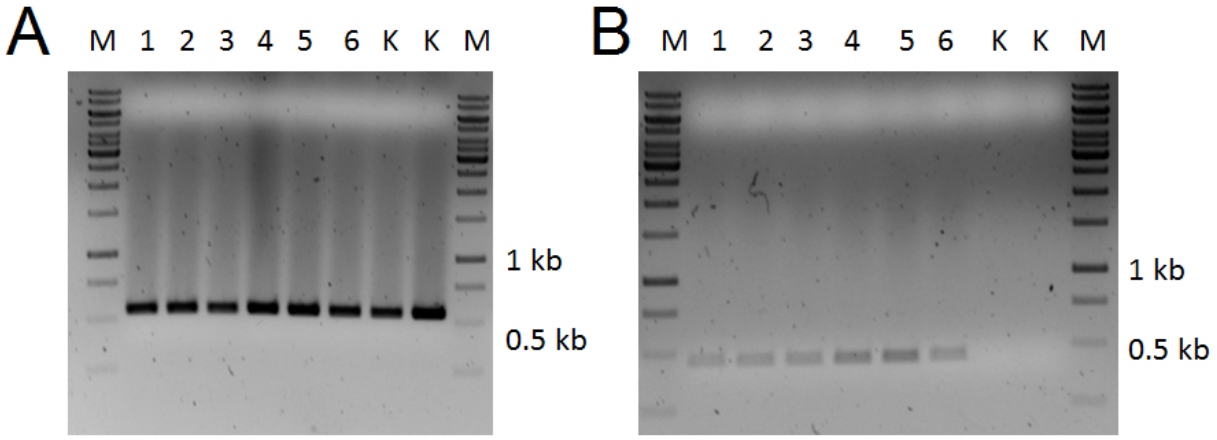


Figure 3.8: PCR to test integration of pMyxZeo_CPN into DK1622. PCR products H and F are shown. M: DNA ladder; 1-3: DK1622 clones of pMyxZeo_CPN 1; 4-6: DK1622 clones pMyxZeo_CPN 2; K: positive control reactions. For probe F, the positive control failed to yield a product.

The six genetically verified clones were then cultivated in liquid CTT medium in the presence of zeocin and XAD-16 adsorber resin. After seven days, cells and XAD were extracted and the crude extracts were analyzed by ESI-MS together with DK1622 wt crude extract for comparison. The chromatograms of all clones showed none of the mass signals of crocapeptin derivatives which are observed in the original host, *Chondromyces crocatus* Cm c5. When screening these extracts for pre-crocapeptins, peaks at 7.5 min with a m/z of 919.5, corresponding to pre-crocapeptin A1 were detected. The extracted ion chromatograms for the pA1 mass are shown in fig. 3.9. However, this peak is also present, yet less pronounced in DK1622 wild type. Comparison of the mass spectrum of a pA1 reference with the peak detected in the DK1622 clones reveals that the characterisitc $[M + NH_4]^+$ ions of pA1 with an m/z of 936.5 are only seen in the reference. The DK1622 clones therefore produced neither crocapeptin nor pre-crocapeptin and the mass peak of 919.5 belongs to another compund.

Sequence Repair and Promotor exchange of the Expression Plasmid

Because of the observed non-production of crocapeptins in the heterologous host, fusion cosmid DJ18BE75D was sequenced. As a result, a insertion mutation was found in the NRPS gene *cpnD* which was considered the most probable reason for the non-production. In DJ18BE75D, a single thymidine is inserted into *cpnD* at nucleotide position 422, resulting in a shift of the translational frame and a stop codon at nucleotide position 489. Therefore, *cpnD* translation is aborted. To repair the sequence, a part of the 5'-end of *cpnD* was amplified by PCR from Cm c5 gDNA and combined with the pmyxo_zeo-backbone by overlap PCR. To avoid transcriptional stress during the cloning and possible additional mutations, the constitutive *npt*-promotor was replaced by an inducible vannilate promotor, which was also introduced via overlap-PCR. Fig. 3.10 shows the three PCR products that were assembled to form the "rescue fragment". The overlap PCR product was then cut with *Bsp*TI and *Bsp*119I, yielding a 5.5 kb fragment, including the 5'-end

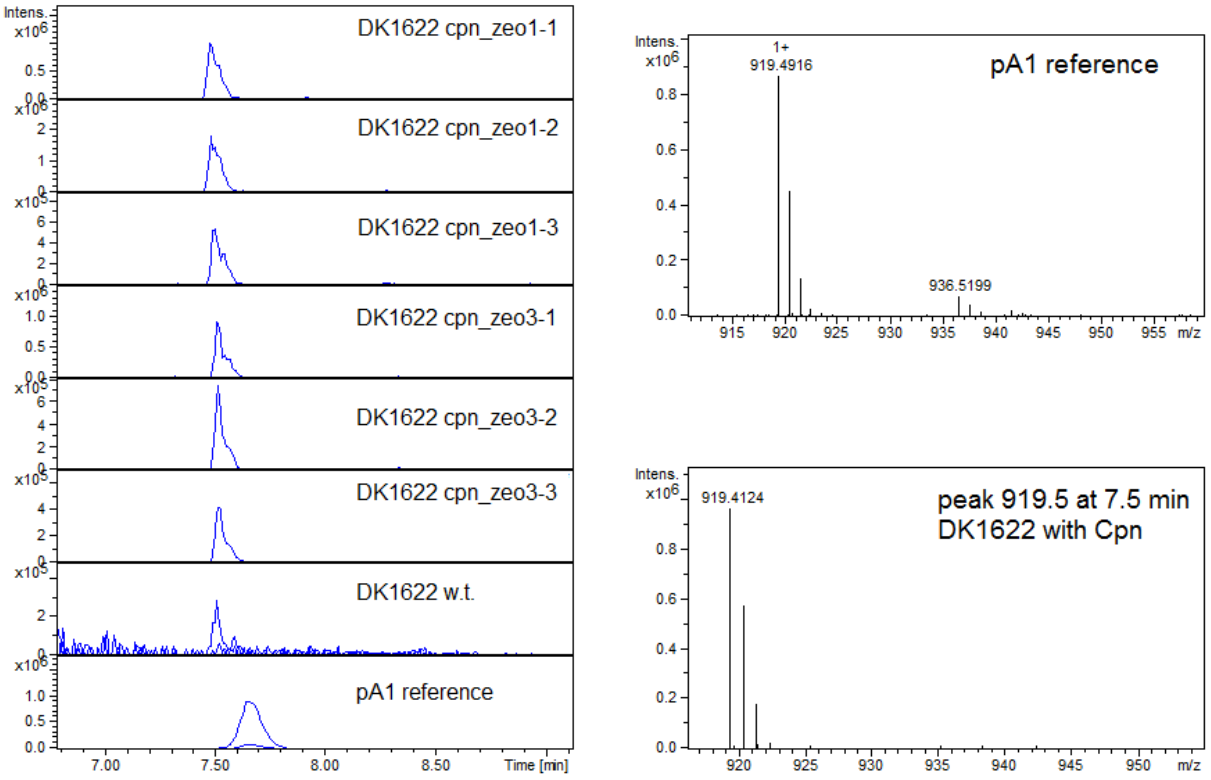


Figure 3.9: Analysis of extracts from *M. xanthus* DK1622 with pMyxZeo_CPN integrated in the genome. Left: EIC for 919.5 of six verified DK1622 clones and comparison with the pA1 reference. Comparison with the compound spectrum of the pA1 reference reveals that the peak in DK1622 is not pA1.

of uncompromised *cpnD* sequence.

Plasmid pmyxzeo_cpn was also restricted with *Bsp*TI and *Bsp*119I and the two fragments were separated by gel electrophoresis. The smaller fragment, which was discarded, contained a part of the plasmid backbone and the 5'-end of *cpnD*, including the insertional mutation. The larger fragment was excised and recovered from the gel. Fig. 3.11A shows the restricted rescue fragment and the plasmid DNA in an agarose gel. Both linear fragments were then circularized by conventional ligation to yield the inducible expression plasmid pmyxzeocpn_pvan. Clones of the ligation mixture were selected with zeocin and ampicillin, and analyzed by restriction analysis as shown in fig. 3.11B. Two selected clones with the correct restriction pattern were re-sequenced in the critical region of *cpnD* and found not to contain the insertion mutation.

The modified expression plasmid pmyxzeo_pvan was then again transformed into DK1622.

Figure 3.10: Map of PCR products assembled via overlap-PCR for repair of the mutated *cpnD* sequence and promotor exchange to the inducible vannilic acid inducible promotor.

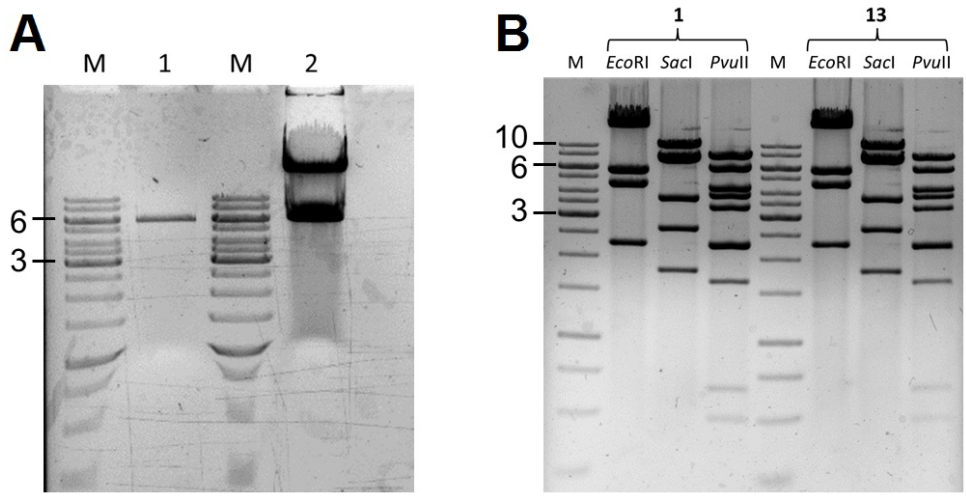


Figure 3.11: Agarose gels of restricted DNA fragments for facilitation of pmyxzeocpn_pvan. **A:** 1: restricted overlap PCR fragment for sequence rescue. 2: pmyxzeo_cpn cut with *Bsp*TI and *Bsp*119I. M: DNA ladder; **Right:** Restriction analysis of two clones of pmyxzeocpn_pvan. Both clones show the expected fragments.

DK1622 clones were then verified by PCR for probe sequences "F" and "H", as described previously. Fig. 3.12 shows the PCR result of randomly selected clones. Three clones that yielded both PCR products in the control reaction were then grown in liquid culture. Each culture was divided into two subcultures, when visible planctonic growth was observed. Expression of *cpnDEF* was was then induced with potassium-vannilate in one of each subcultures, while the other subcultures were left uninduced. After further incubation for 5 days, cultures were extracted and analyzed by LC-MS. No mass signals corresponding to crocapeptin or pre-crocapeptin derivatives were detected in any extract. Further, no difference in the mass spectra of induced and non-induced cultures was observed. It is therefore concluded that the crocapeptin biosynthesis genes are not functionally expressed in *M. xanthus*. However, an in-depth statisitical analysis of the extracts has not yet been done. This should be applied when expression of the biosynthesis genes has been proven experimentally by transcription analysis.

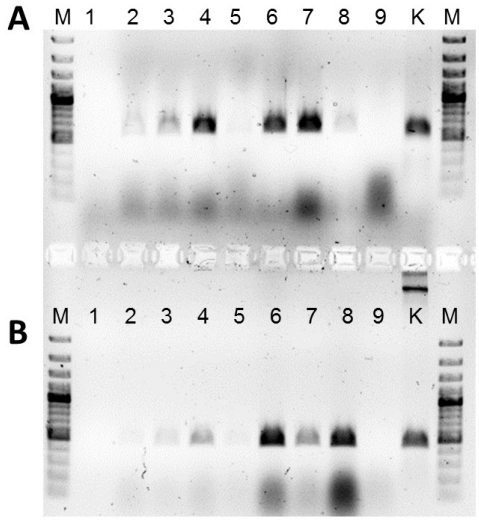


Figure 3.12: PCR to test integration of pmyxzeo_pvan into DK1622. Two sequence parts of *cpnD* are amplified in correct clones. **A:** Probe sequence "F". **B:** Probe sequence "H". Lanes 1-9: DK1622 clone gDNA template. M: DNA ladder. K: positive control. Clones 3,4,6,7 and 8 yield the expected products for both probes.

3.3 Experimental Procedures

Oligonucleotides and PCR

For Red-ET recombination to assemble the *cpn*-cluster, the following PCR products were generated using the Phusion Polymerase and the primers listed in table 3.1: PCR1 and PCR2 were fused via overlap-PCR to obtain PCR3, which is the linear fragment for recombination with cosmid B:E7. After cosmid derivatives BE75D and DJ18BE75D were obtained, plasmid modification was done with PCR product pMyxZeo.

To repair the mutation in *cpnD* and exchange the promotor from P*npt* to pvan, linear fragments were generated from separate PCR-products via overlap PCR, consisting of PCR PCRV1, PCRV3 and PCR4rescue. PCRV3 and PCR4rescue were fused to PCRV8, which was again fused with PCRV1 to yield PCRV9. PCRV9 contains the pMyxZeo backbone, the vanillic acid inducible promotor and 500 bp of original *cpnD*-sequence. Restoring the sequence of *cpnD* via direct cloning did not work, therefore, the sequence was restored via conventional restriction cloning. For this, PCR products PCRrescueSOE1 and PCR SOE2 were were fused to PCRSOE3, which was cut as described above and used for restriction cloning. The primer sequences of that part are listed in table 3.2.

Table 3.1: Oligonucleotide Primers used for PCR of *cpn*-cluster assembly

Product	**Primer Name**	**Sequence 5'-3'**
PCR 1	CmR-PacI-tn5_for	AAGGCAGTTATTGGTGCCCTTAATTAAT-GGACAGCAAGCGAACCG
	HR_D_rev	GTCTTCCGGCTTCGGTCG
PCR2	HM2-PacI-CmR_for	GATCGAAGTTAGGCTGGTAAGAGCCGCG-AGCGATCCTTGAAGTTAATTAATACCTG-TGACGGAAGATCACT
	CmR_rev	GGGCACCAATAACTGCCTT
PCR3	HL-EcoRI-HM2_for	GCTGGTTTCCTGGATGCCGACGGATTTG-CACTGCCGGTAGAAGAATTCGATCGAAG-TTAGGCTGGTAAGAG
	HR_D_rev	GTCTTCCGGCTTCGGTCG
pMyxZeo	pmyxZeo_for_HL	GATCGAAGTTAGGCTGGTAAGAGCCGCG-AGCGATCCTTGAAGGCTTTCCAGTCGGG-AAACC
	pmyxZeo_rev_HR	CACGCACGCCCTCGCTCGGGAGCAGGCA-CCCGTCAGGGCCAGGCTGCAGGGGCCCG-GGA

Table 3.2: Oligonucleotide Primers for PCR of *cgn*-cluster cloning

Product	**Primer Name**	**Sequence 5'-3'**
PCRV1	pmyxZeo_for_HL	CTACCTGACGGGTGCCTGGCCCTGACGG-GTGCCTGGCCCTGAGCTTTCCAGTCGGG-AAAC
	pMyxZeo_pVan_rev	GCATTCGCGCCGACTGAGGGATCCAATT-GGCAGTGAG
PCRV3	VanPro_for	TCAGTCGGCGCGAATGC
	pVan_HR_cpnD_rev	GTCTTCCGGCTTCGGTCGAACCGGACCT-GCGATGCTGGTCATATGCGTTTCCTCGC-ATCG
PCR4rescue	for_PCR4rescue	CGATGCGAGGAAACGCATATGACCAGCA-TCGCAGGTCC
	rev_PCR4_rescue	TGAGGCTGTGCTCCATCGAG
PCRrescueSOE1	rescue_SOE_for	AACAAATAGGGGTTCCGCGCA
	pVan_HR_cpnD_rev	GTCTTCCGGCTTCGGTCGAACCGGACCT-GCGATGCTGGTCATATGCGTTTCCTCGC-ATCG
PCRSOE2	PCR4rescue_for	CGATGCGAGGAAACGCATATGACCAGCA-TCGCAGGTCC
	rescueSOE_rev	ATGAGCGTCCTCCAGCATGAA

Plasmid Modification via Red-ET recombineering

For recombination of cosmids and cosmid derivatives via Red-ET recombineering, cloning strain *E. coli* HS996 pSC101-BAD-gbaA was used. The plasmid carries a heat-sensitive replication origin and is lost if the culture is incubated at 37°C and a tetracycline resistance gene. The recombinase genes on the plasmid are controlled by the pBAD promotor and can be induced by addition of 0.1-0.2% L-arabinose to the culture.

The circular cosmid was introduced into the cloning strain via electroporation and clones are selected with tetracyclin and and the respective antibiotics for maintaining the cosmid DNA. Cultivation temperature is 30°C unless stated otherwise. Liquid cultures of these clones are then induced at $OD_{600} = 0.2$ and further incubated until an $OD_{600} = 0.5$ is reached. Then, the cultures are washed three times with ice-cold water, resuspended in 40 μL ice-cold ceH2O and the linear fragment (PCR product or restriction fragment) is introduced via electroporation. Electroporation conditionas are: U = 1350V, R = 600 Ω, $C = 10$ μF. Depending of the size of the linear fragment, between 100 and 500 ng are used per electroporation. Cells are then resuspended in 1 mL of pre-warmed LB-medium and incubated for 90 min at 37°C at 900 rpm. Then, the cells are prelleted, resuspended in 100 μL LB-medium and plated on agar containing the antibiotics to select for the target construct. Plates are incubated at 37°C. Clones are then analyzed via colony-PCR and DNA restriction according to standard protocols. To avoid mixtures of clones carrying the original and the target plasmid, retransformation of isolated plasmid DNA followed by another round of plasmid isolation is recommended. Plasmid isolation was done by alkaline lysis.

3.4 Conclusions

The three crocapeptin biosynthesis genes *cpnD*, *cpnE* and *cpnF* were cloned via Red-ET recombineering into an expression plasmid that can integrate in the *M. xanthus* DK1622 genome. As sequencing of the cosmid derivative DJ18BE75D revealed, an insertion mutation in *cpnD* was acquired during the first cloning steps, which was had eliminated via conventional cloning later on, to restore the *cpnD* reading frame. DK1622 mutants with the expression plasmid pMyxZeoCPN integrated in the genome were obtained, but induction of the *cpn* operon with vanillic acid did not lead to novel peaks in the respective extracts' chromatogram, compared to extracts from control cultures. No crocapeptins, precursors or side products were obtained from this heterologous system.

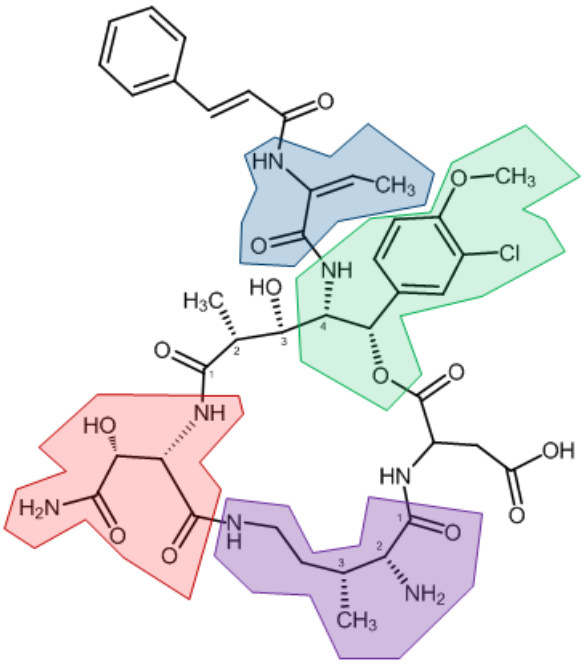
Possible reasons are either host incompatibility issues regarding gene expression, i.e. incomplete transcription, translation, incorrect protein folding or the three genes are not sufficient for crocapeptin production. Mutants of the native crocapeptin producer *C. crocatus* Cm c5 were generated prior to the heterologous expression approach, and crocapeptin production was abolished only in mutants with a disrupted NRPS gene *cpnD*. Disruption of genes upstream of *cpnD*, i.e. the ABC-transporter genes *cpnBC* and the protease gene *cpnA* did not abolish production but slightly reduce or increase it. They were therefore considered non-essential. However, their function might be compensated in Cm c5 while they are essential in a heterologous host. Downstream of *cpnF*, a protein kinase of unknown function is encoded in opposite direction, annotated as *cpnG*. No mutants with disrupted *cpnG* were obtained, which leads to suggest that this gene is essential for Cm c5, so that Cm c5 mutants with disrupted *cpnG* are not viable. Abolishment of crocapeptin production, on the other hand, is not essential for Cm c5 and does not lead to any obvious phenotypic changes. To enable heterologous expression of the crocapeptins, the DK1622 mutants should be investigated wether *cpnD* is fully transcribed, i.e. by RT-PCR and subsequent cDNA analysis. If that is the case, the remaining crocapeptin cluster genes discussed here should be integrated into the DK1622 genome as well and the mutants tested again.

Analysis of CpnD protein expression in the mutants is not considered a promising approach. DK1622 contains several other very large NRPS and PKS enzymes, making SDS-PAGE rather unlikely to yield any significant results. One option to check protein integrity could however be the expression of a CpnD-fusion enzyme in DK1622 where CpnD is C-terminally tagged with green fluorescent protein (GFP) or a simple enzyme such as beta-galactosidase. If the activity of the tag is found in the crude lysate of respective DK1622 mutant, complete translation is proven and correct folding is probable. If instead an affinity tag is used, the purified protein can be investigated for ATP-hydrolysis or probed with synthetic substrate mimics.

As a last option, a related *Chondromyces crocatus* strain could be chosen as host to test expression of the cloned construct, which first has to be adapted to the respective host. One relatively simple approach would be to clone the *cpn*-cluster in a mobile genetic element, such as the pMycoMar-transposon. This allows random integration in a wide range of hosts, given their amenablity to mutagenesis. Even hosts that are very closely related to Cm c5 and produce crocapeptin congeners, could be assessed with that method, because they might contain currently unknown additional essential genes for crocapeptin production. By HR-MS analysis, it is usually possible to distinguish their native crocapeptins from the heterologously expressed congeners.

Chapter 4

Crocadepsins A and B - Depsipeptides from the myxobacterium *Chondromyces crocatus* discovered by a genome mining approach



Author's efforts

The author's efforts regarding this project comprise the annotation of the crocadepsin gene cluster, the design and cloning of the inactivation constructs for the mutagenesis of *Chondromyces crocatus* Cm c5, followed by cultivation, mutagenesis, extraction and MS analysis. Further contributions include interpretation of the labeling experiments and creation of the biosynthesis scheme.

Contributions by others

The initial NRPS/PKS knockout mutants that led to crocadepsin discovery were obtained from a mutant library created by Shwan Rachid. Crocadepsins were initially isolated by Frank Surup, structure elucidation was carried out by Frank Surup and Alberto Plaza. The project was supervised by Prof. Rolf Müller. Further aknowledgements are listed at the end of this chapter.

Abstract

Analysis of the genome sequence of myxobacterium *Chondromyces crocatus* Cm c5 revealed the presence of a cryptic NRPS gene cluster, which we connected to two unknown chlorinated metabolites by a comparative gene inactivation approach. Subsequently, the compounds were isolated from large scale cultivations. Structure elucidation revealed a unique cyclic depsipeptide skeleton containing an unprecedented β-hydroxy-γ-amino acid building block and an ω-connected 3-methyl ornithine. The biosynthesis was elucidated by analysis of the biosynthetic gene cluster, gene knock out experiments and additional feeding experiments employing labeled precursors.

4.1 Introduction

Nonribosomal peptides (NRP), polyketides (PK) and hybrids thereof have a vast medical importance in drug development as exemplified by prominent representatives such as tacrolimus, daptomycin and cyclosporin.[101–103] Their diverse biological activities are accompanied by a mere endless structural variety, which is based mainly on the plentitude of NRP building blocks. Unlike ribosomally expressed peptides, which built-up from only 20 proteinogenic amino acids, approximately 500 naturally occurring amino acids have been identified so far as components of NRP.[104] With the ajudazoles, chondramides, crocacins, chondrochlorens, thuggacins and crocapeptins, six classes of NRP and NRP/PK products are already known from the myxobacterial strain *Chondromyces crocatus* Cm c5, making this species an exceptional secondary metabolite producer.[20,22,105–108] Preliminary whole genome sequencing of Cm c5 indicated the presence of more biosynthetic megasynthetases in charge of generating NRP and PK natural products than expected leading to a genome mining approach aiming to correlate these gene clusters to chemical products. Using targeted gene inactivation via single-cross over insertion of pSUP_*hyg* plasmid derivatives containing a fragment of each target gene cluster into the Cm c5 genome,[22] 11 non ribosomal peptide synthetase (NRPS) genes were mutated. After cultivation of the resulting mutants, their crude extracts were investigated by comparative HRMS analysis as described previously.[86]

4.2 Results and Discussion

Inactivation of two NRPS genes, which were later annotated as *cdpA* and *cdpL* led to abolishment of production of "peaks" with m/z 858.3 and 872.3 in the respective chromatograms. Sequence refinement confirmed these two NRPS genes to be part of a mixed NRPS/polyketide synthase (PKS) gene cluster, which is shown in figure 4.2A. All identified ORFs were annotated and analysed for homologies to known proteins with the aim to assign their possible function in crocadepsin biosynthesis (Figure 4.2B). Since the best matching sequences of most of the ORFs in the crocadepsin gene cluster were derived from ORFs of *Chondromyces apiculatus* strain DSM436, we investigated the published sequence of DSM436 (submitted by IMTECH-CSIR, BioProjects PRJNA192263, GenBank assembly accession GCA_000601485.1) and found an almost identical biosynthetic gene cluster in the genome. The cluster in DSM436 differs from the Cm c5 crocadepsin cluster only by the absence of the putative MFS transporter gene, which is replaced by an aminomutase gene. The ORFs up- and downstream of the cluster are different in Cm c5 and DSM436, confirming the cluster boundaries. A detailed analysis and comparison of both sequences is presented in SI-1. We then aimed to inactivate the ORFs found between *cdpA* and *cdpL* in Cm c5, to study the effects on production. Mutants with disrupted genes *cdpC*, *cdpD*, *cdpE*, *cdpF*, *cdpG*, *cdpJ*, *cdpK* were obtained, genetically verified by PCR and their extracts analysed (see Figure 4.2C and SI-3). Inactivation of the third NRPS gene *cdpF* led

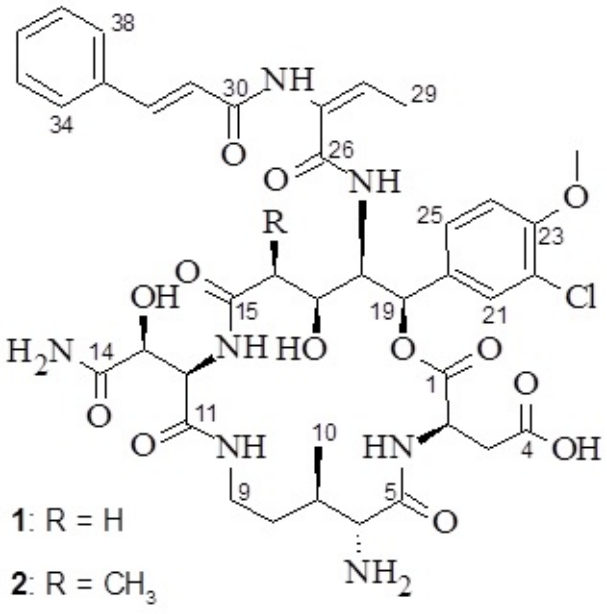
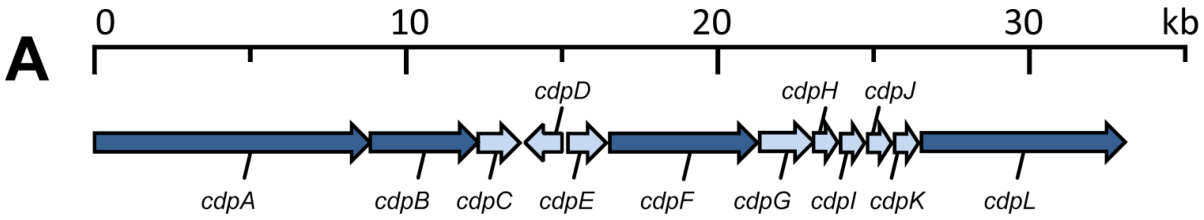


Figure 4.1: Crocadepsins A **(1)** and B **(2)**, isolated from *Chondromyces crocatus* strain Cmc5

to complete loss of crocadepsin production in the mutants, which was also the case for mutants of *cdpC*, encoding a cytochrome P450 monooxygenase. Inactivation of the halogenase gene *cdpE* reduced crodepsin signals in the respective extract, but instead a peak with m/z of 838.3 was observed corresponding to non-chlorinated crocadepsin A. Reexamination of the wild-type confirmed, that the non-chlorinated derivative crocadepsin C is generally produced as a minor metabolite. Mutants of the arginase gene *cdpK* and the methyl transferase gene *cdpJ* both showed strongly reduced crodadepsin production, but no novel derivatives were observed. Inactivation of the two putative transporter genes *cdpD* and *cdpG* led to reduced crocadepsin signals.

To elucidate the structure of the main compounds, we isolated a total of 9.4 mg of **1** and 11.2 mg of **2** from large scale cultivations. The structures were subsequently determined by extensive 2D NMR spectroscopy, derivatisation and degradation. For crocadepsin A **(1)**, HR-ESI-MS provided the molecular formula as $C_{39}H_{49}N_7O_{13}Cl$ and the presence of a chlorine atom was indicated by its characteristic isotopic pattern. Crocadepsin B **(2)** with formula $C_{40}H_{51}N_7O_{13}Cl$ comprises an additional methyl group compared to **1**. The proton NMR spectrum of 2 in H_2O/D_2O (9:1) showed the presence of 17 protons in the region between 6.5 and 10.06 ppm. A H_2O/D_2O exchange experiment in DMSO-d_6 revealed that 7 of these protons are amide protons δ_H 10.06; 8.68; 7.97; 7.96; 7.94; 7.63; 7.41), the remaining 10 protons are thereby part of aromatic systems. The analysis of ^{13}C NMR spectra in DMSO-d_6 and H_2O/D_2O gave further evidence for the peptidic nature of 2 by containing signals of 8 carbonyl carbons at δ_C 167-179 ppm. The interpretation of $^1H,^1H$ COSY and TOCSY, respectively, $^1H,^{13}C$ HMQC and HMBC spectra assigned aspartic acid besides the unusual amino acid derivatives β-methyl ornithine (βOrn), 3-hydroxyasparagine (Has), dehydrobutyrine (Dhb) and a cinnamic acid (Cin). Furthermore, a 2-amino-1-(3-chloro-4-methoxyphenyl)ethyl element was assigned in addition to 3-hydroxy- 2-methylpropanoic acid. Although only a weak TOCSY correlation was observable between these two units, HMBC correlations from H-16 to C-18 and H-17 to C-19 indicated a single cohering carbon chain and established the structure of this new β-hydroxy-γ-amino acid. The residue sequence was assigned by inter-residue $^1H,^{13}C$ HMBC correlations (Figure SX, Supporting Information). The high field shift of H-19 (δ_H 5.91) indicated an ester linkage at this position, which was confirmed by the $^1H,^{13}C$ HMBC correlation of H-19 to C-1 establishing the planar depsipeptide structure of **2**. A $^1H,^1H$ ROESY correlation between NH-19 and H-28 demon-



B

ORF	proposed function	closest homologue	source organism
cdpA	NRPS	EYF05518.1	[*Chondromyces apiculatus* DSM 436]
cdpB	PKS	EYF01031.1	[*Tolypothrix bouteillei* licb1]
cdpC	cytochrome P450	EYF01032.1	[*Chondromyces apiculatus* DSM 436]
cdpD	MFS transporter	WP_015183006.1	[*Microcoleus* sp. PCC 7113]
cdpE	halogenase	EYF01033.1	[*Chondromyces apiculatus* DSM 436]
cdpF	NRPS	EYF01034.1	[*Chondromyces apiculatus* DSM 436]
cdpG	ABC transporter	EYF01036.1	[*Chondromyces apiculatus* DSM 436]
cdpH	SyrP-like hydroxylase	EYF01037.1	[*Chondromyces apiculatus* DSM 436]
cdpI	aminotransferase	EYF01038.1	[*Chondromyces apiculatus* DSM 436]
cdpJ	methyltransferase	EYF01039.1	[*Chondromyces apiculatus* DSM 436]
cdpK	arginase	EYF01040.1	[*Chondromyces apiculatus* DSM 436]
cdpL	NRPS	EYF01041.1	[*Chondromyces apiculatus* DSM 436]

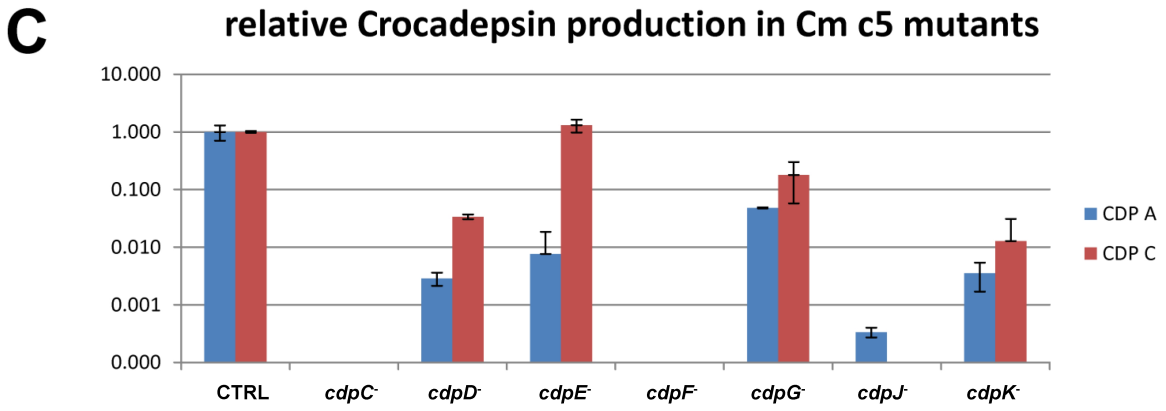


Figure 4.2: A: Crocadepsin gene cluster from Chondromyces crocatus Cm c5. NRPS and PKS genes are schown in dark blue, other ORFs are shown as light blue arrows. **B:** Annotated ORFs in the *cdp* gene cluster, with proposed functions according to closest homologues found via BLAST against NRD.**C:** Effect of gene inactivation on production of crocadepsin derivatives A and C in Cm c5 cell extracts. Production levels of crocadepsins are shown relative to chondramide B production levels, which is used as internal reference compound. Masses were normalized to average crocadepsin A and C levels in the control cultures, which were set to 1. CTRL: Control cultures of Cm c5 with pSUP_hyg insertions in other independent gene clusters. Crocadepsin B was excluded from the analysis.

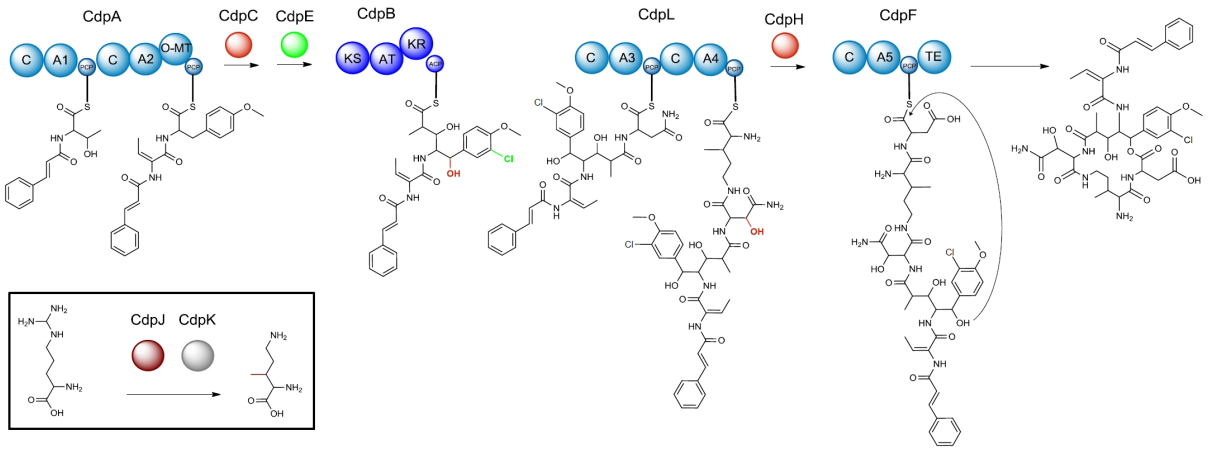


Figure 4.3: Model for crocadepsin biosynthesis. Cinnammoyl-CoA serves as carrier protein free starter molecule and is extended by CdpA, CdpB, CdpL and CdpF. Modifications of the amino acid residues by CdpC and CdpH are expected to be introduced during chain assembly but may as well occur with different timing. Beta-methyl-ornithine is prepared by catalysis of CdpJ and CdpK from arginine prior to loading onto the NRPS, as shown in the box on the lower left. Crocadepsin B is produced, when CdpB incorporates malonyl-CoA instead of methylmalonyl-CoA.

strated an unusual *trans*-configuration of the [Γ][27,28] double bond. The *trans* configuration of the [Γ][31,32] double bond was established by the large coupling constant $^{3}J_{31,32}$=15.8 Hz.

Crocadepsin B **(2)** features nine chiral centers, whose stereochemistry was determined by a combination of NMR spectroscopy, chemical degradation and derivatisation. First, the relative configuration of all substituents in Cyt, βOrn and Has moieties was deduced by the J-based method developed by Murata and Tachibana (Figure S4), which indicate *syn* configurations in all cases.[109] Subsequently, the absolute configuration of the Asp and β-hydroxy asparagine was established as L- by hydrolysis and subsequent derivatisation with FDAA, respectively FDLA.[110] Treatment with Mosher's chloride (MTPA, α-methoxy-a-(trifluoromethyl)phenylacetic acid) at room temperature led to the mono-MPTA ester at C-6.[[111] Its $\Delta\delta_{SR}$ values (Figure S5) revealed an *S* configuration of C-6. The elucidation of the absolute configuration of the remaining stereo centers C-16 to C-19 proved to be most challenging. Several attempts to derivatize the hydroxyl function at C-17 of **1** and **2** by MPTA or MPA using DMAP/EDCI as coupling agents failed; attempts for ring opening with HCl, NaOH or MeOH led to decomposition. Therefore, the reaction with Mosher's chloride was repeated at an elevated temperature (40C for a prolonged time (16h). Because the reaction was carried out in pyridine-d_4, a potentially problematic work up could be avoided by a direct measurement of NMR spectra. Despite poor spectrum quality, the methyl group at C-16 could be identified easily. Interpretation of TOCSY NMR spectra enabled the assignment of the $\Delta\delta_{SR}$ values of +0.09 for the methyl group H_3-16' and +0.12 for the adjunct proton H-16. Both values indicate an *S*-configuration of C-17, finalizing the elucidation of the absolute stereochemistry of **1** and **2** as presented in figure 4.1.

Of particular interest in the mixed NPRS/PKS biosynthesis is the origin of the uncommon β-methyl ornithine moiety and its unique ω-amino connection. Analysis of relevant genes suggested that the unit is derived from arginine, which is hydrolyzed and transmethylated by S-adenosyl methionine. To examine this hypothesis, we carried out feeding experiments employing $^{13}C_6$-arginine and ^{13}C-methionine. Compared to the unlabelled substance (Figure S6a) the molecular ion clusters of **1** and **2** are significantly altered by the feeding of ^{13}C-methionine (Figure S6b). The most prevalent pseudomolecular ion peaks m/z 860.4 and 874.4 indicate the incorporation of two molecules of methionine per molecule **1** or **2**. By MS/MS analysis, the loci of the label were proven to be the β-methyl ornithine and the crocatyrosine moieties (see SI-2). Thereby

this experiment supported the hypothesis of the biosynthesis for β–methyl ornithine predicting the incorporation of one molecule in β-methyl ornithine unit besides the O-methylation of the crocatyrosine moiety. Feeding of $^{13}C_6$-arginine resulted in the appearance of new satellites in the molecular ion clusters with m/z of 863.4/865.5 for **1** and 877.4/879.4 for **2** (Figure S6c and SI-2), which is consistent with the proposed incorporation of 5 out of the 6 arginine carbon atoms. The methyl C-16' is not derived from SAM, but from incorporation of methylmalonyl-CoA into **2** instead of malonyl-CoA into **1** catalyzed by CdpB.

With the three dimensional chemical structure of the crocadepsins at hand, the findings from sequence analysis, mutagenesis and feeding experiments were combined to provide a biosynthesis proposal, shown in fig. 4.3. Crocadepsin assembly starts at CdpA, where activated cinnamic is condensed by the NRPS with threonine and then extended by tyrosine. No carrier protein for loading of cinnamoyl-CoA was detected which is why the free CoA is expected to be chosen by the first C domain of CdpA for condensation with threonine. The hydroxy-group of threonine is eliminated in a yet unresolved step resulting in the dehydrobutyrine moiety found in **1** and **2**. Whether and how this reaction involves activation of the OH group remains to be determined. The tyrosine residue is O-methylated during assembly by the methyl-transferase domain located adjacent to the tyrosine adenylation domain, and chlorinated in ortho-position by the flavin-dependent halogenase CdpE, although also small amounts of non-chlorinated Crocadepsins C are detected. Hydroxylation of tyrosine in the β-position by CpdC, a cytochrome P450 monooxygenase is expected to occur on a nascent PCP-bound intermediate, because the hydroxy-group is required for product release via lactonisation via the aspartate residue. This hypothesis is corroborated by the loss of production of crocadepsins in *cdpC*$^-$ mutants. The peptide is extended by either malonyl-CoA or methylmalonyl-CoA by the PKS CdpB, thereby defining whether crocadepsin A or B is formed. Extension continues at the NRPS CdpL incorporating asparagine. Hydroxylation to β-hydroxy-asparagine is assumed to be catalyzed by CdpH, a non-heme hydroxylase with high similarity to SyrP from *Pseudomonas syringae*, which catalyzes a similar step in syringomycin biosynthesis.[112] The following extender unit β-methyl-ornithine (bmeOrn) is derived from arginine, the methyl group from methionine, as obtained from labeling and MS^2 experiments (see SI-3). Sequence analysis of the responsible A-domain CdpL-A4 does not allow a clear prediction of its substrate. Either bmeOrn is formed prior to loading onto the NRPS or arginine is loaded first by CdpL-A4 and subsequently modified while tethered to the PCP-domain. Methylation of ornithine or arginine at the β-carbon is expected to result from the concerted action of CdpI and CdpJ, a putative transaminase and SAM-dependent methyl-transferase, respectively. Here, a transient α-keto-acid is formed which can then be β-methylated, a mechanism known from indolmycin biosynthesis.[113] The arginase CdpK catalyzes hydrolysis of urea from arginine, either prior to or after methylation, followed by condensation with asparagine via the ϵ-amino group of βmeOrn. During transfer of the peptide to CdpF, aspartate is incorporated and a nonchlorinated crocadepsin precursor is released by the TE domain via lactonisation between the β-hydroxy-group of the tyrosine residue and the proximal carboxy-group of aspartate.

Crocadepsins A and B **(1-2)** contain dehydrobutyrine, an uncommon residue in NRPS-derived peptides but a characteristic part of ribosomally produced lantibiotics. Additionally, they contain a β-hydroxy-γ-amino acid residue known as a component of many protease inhibitors; i.e. the representative pepstatine is widely used in the field of peptidomimetics.[114] The inhibition pattern varies greatly with individual structures; metabolites containing simple elongated tyrosine units have been patented as inhibitors of the cysteine protease caplain. However, we could not detect any inhibitory activity against chymotrypsin or pepsin, so a comprehensive screening program for biological activity is required to find the actual target of **1** and **2**.

Conclusions

In conclusion, we have discovered two unique cyclic depsipeptides named crocadepsin A **(1)** and B **(2)** by a genome mining approach that is based on targeted knock out of NRPS genes in the sequenced myxobacterium *Chondromyces crocatus* strain Cm c5. The metabolites contain inter alia unusual or unique amino acids β-methyl ornithine, dehydrobutyrine and a β-hydroxy-γ-amino acid as building blocks. Analysis of the gene cluster revealed insights into the biosynthesis of the metabolites and their amino acid bulding blocks. Our study underlines the great potential of genome based screening approaches for finding novel chemical entities. We emphasize that the screening for novel secondary metabolites even from intensively studied organisms such as *C. crocatus* Cm c5 exhibits great potential for finding new types of chemistry and unusual biosynthetic pathways.

Experimental Procedures

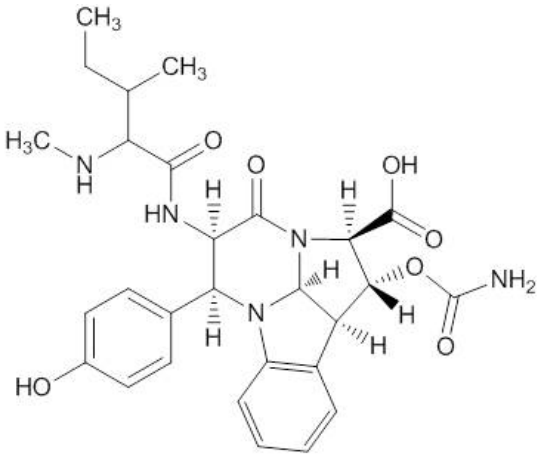
Experimental details, including bioinformatic analyses, mutagenesis experiments and MS-analysis of labeling experiments can be found in the Supporting Information attached in the appendix on page 192. Further details such as NMR spectra and structure elucidation procedures will be avaiable after publication of the manuscript.

Acknowledgement

We thank Christel Kakoschke for recording NMR spectra, Aileen Teichmann and Thomas Hoffmann for ESI-MS measurements, Diana Telkemeyer for strain maintaince, Birte Trunkwalter for help with labeling experiments and Wolfgang Kessler and his team for large scale cultivation.

Chapter 5

Crocagins - RiPPs from the myxobacterium *Chondromyces crocatus*



Exact Mass: 551,238

Author's efforts

The author's experimental efforts regarding this chapter comprise the identification and annotation of the crocagin biosynthesis gene cluster through genome analysis, the design and cloning of the inactivation constructs, Cm c5 mutagenesis and cultivation, mutant extraction and MS analysis. Further contributions include design and conduction of TAR assembly and subsequent cloning steps.

Contributions by others

Crocagins were initially isolated by Frank Surup, who also carried out structure elucidation and labeling experiments. The project was supervised by Prof. Rolf Müller.

Abstract

Crocagins A and B are novel tetracyclic peptides, isolated from the myxobacterium *Chondromyces crocatus* Cm c5. Structure analysis and feeding experiments confirmed a highly modified Ile-Tyr-Trp tripeptide which results from a ribosomally translated precursor peptide that is highly modified during maturation. By genome sequence analysis of the producer strain, the putative precursor peptide was identified, and adjacent biosynthesis genes were inactivated, resulting in abolished crocagin production in one case and production of a crocagin precursor in another mutant. The crocagin biosynthesis gene cluster was assembled by PCR-based TAR-cloning, to obtain a plasmid for crocagin expression in a heterologous host.

5.1 Introduction

The myxobacterium *Chondromyces crocatus* Cm c5 is a prolific producer of natural products, such as the ajudazols,[19] chondramides,[17] crocacins,[16] chondrochlorens,[18] thuggacins,[85] and crocapeptins.[108] Initially, most of these compounds had been detected in activity-based screenings by testing for antifungal, antibacterial, and cytotoxic qualities, followed by activity-guided isolation of the active principle. The availablity of the Cm c5 genome sequence now enables further compound mining of this organism by genome mining, to obtain compounds that are not found in activity-based screenings. As a first target, putative novel gene clusters containing NRPS or PKS genes were inactivated and the extracts of mutants and wild-type cultures compared to correlate genes to compounds. Mass peaks that are absent in a specific mutants' extract chromatogram are then isolated from a large-scale cultivation of the Cm c5 wild type, as they represent most probably a compound produced by the encoded biosynthesis genes. The crocagins presented herein were initially isolated during workup of the crude extracts from a Cm c5 large-scale cultivation, and its chemical structure was investigated by NMR. The crocagin structure is formed by the tripeptide Ile-Tyr-Trp, in which the indole ring of the tryptophane has undergone multiple crosslinking, forming a tetrycyclic structure, as shown in fig. 5.1C.

The Cm c5 genome sequence was then screened for a suitable NRPS gene cluster with a predicted domain architecture that fits observed peptide sequence, but no candidate NRPS genes exhibiting such substrate specifity were found. Therefore, alternative biosynthetic routes to obtain this secondary metabolite were considered, such as a ribosomal origin. Ribosomally derived peptides (RiPPs) originate from ribosomally assembled precursor peptides, which are encoded as a gene in the producer organism. The precursor peptide contains the amino acid sequence that forms the final RiPP, and additional sequences such as leader sequences and recognition sites for modifying enzymes, which are cleaved off during compound maturation.[69]

Herein, crocagins are presented along with the corresponding biosynthetis gene cluster. Their ribosomal origin and the involvement of two of the genes in the cluster was verified by targeted gene inactivation in the producer strain Cm c5. To test whether the ORFs annotated in the putative crocagin biosynthetic gene cluster are indeed sufficient to form crocagin, the gene cluster was assembled by TAR into the shuttle plasmid to allow heterologous expression after plasmid modification.

5.2 Results and Discussion

The crocagins are tripeptides, consisting of the amino acid sequence Ile-Tyr-Trp. The Tyr and Trp residues are covalently crosslinked and form two additional heterocycles. The nitrogen of the indole moiety is crosslinked to the β-carbon of the Tyr residue and one carbon atom of the

indole ring is crosslinked to the nitrogen atom of Tyr-Trp peptide bond. The crosslinking results in an additional proline and 1,3-isopiperazone heterocycle. The proline ring has is hydroxylated at carbon C2 and additional O-carbamoylated. Crocagins A and B differ in their N-terminus, which is methylated in case of the major metabolite crocagin B. The chemical crocagin structure is shown in fig. 5.1C.

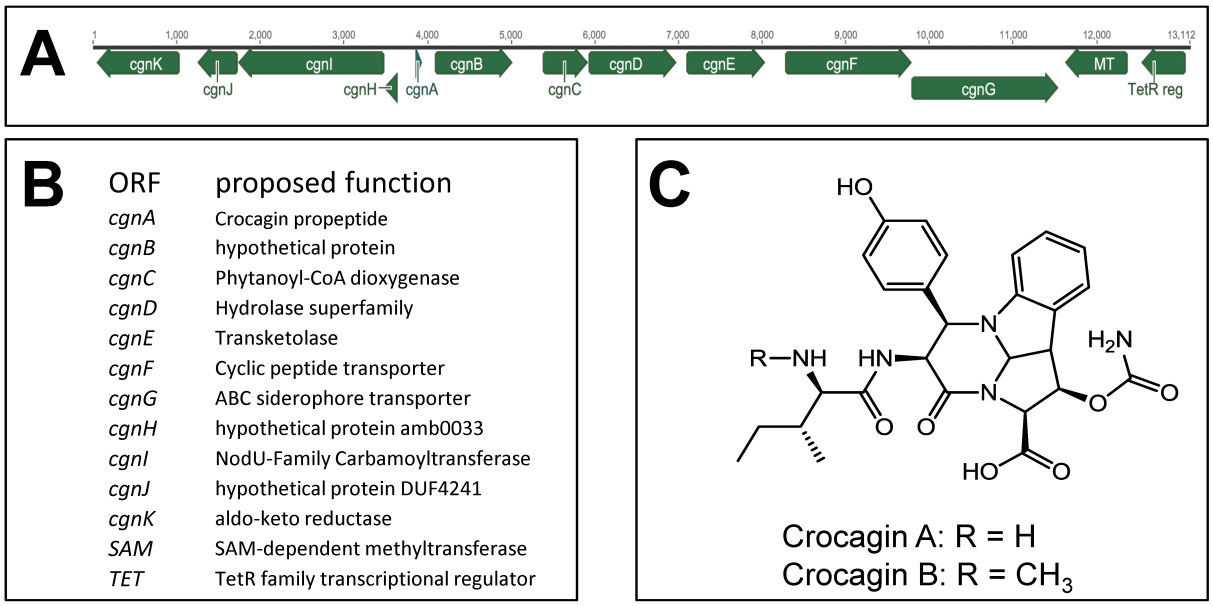


Figure 5.1: A: Crocagin biosynthetic gene cluster from *Chondromyces crocatus* Cm c5. **B:** ORFs annotated in the crocagin gene cluster and their proposed function, as obtained from BLAST analysis. **C:** Chemical structure of crocagin A and B.

Identification on the Crocagin biosynthetic gene cluster

As mentioned above, a non-ribosomal origin of this compound was initially exprected but ruled out after genome sequence analysis. Considering a ribosomal origin of the molecule, the crocagin peptide sequence Ile-Tyr-Trp theoretically back-translates to six different nucleotide sequences. Screening the genome for these six sequences resulted in 43 sequence matches, of which only only one was found to actually encode the peptide sequence Ile-Tyr-Trp within the reading frame. The sequence belongs to a gene encoding the putative precursor peptide gene, consisting of 21 amino acids, with the tripeptide Ile-Tyr-Trp at its carboxy-terminus. Consequently, the sequence up- and downstream of the propeptide gene was investigated and all identified ORFs were annotated. The annotated crocagin gene cluster is shown in fig. 5.1A, and the annotated ORFs are listed fig. 5.1B, with their putative function as obtained from BLAST sequence homology. Upstream of the propeptide gene *cgnA*, a putative carbamoyl-transferase gene *cgnI* is encoded in opposite direction. This finding further corroborates to this cluster being the crocagin biosynthesis locus, since the crocagins are carbamoylated. Downstream of *cgnA*, a hypothetical protein *cgnB*, a putative oxidoreductase enzyme *cgnC*, hydrolase *cgnD* and transketolase *cgnE* and two transporter genes *cgnF* and *cgnG* are encoded. Further downstream, a SAM-dependent methyl transferase *MT* and a transcriptional regulator *tetR* are encoded on the opposite strand. Inactivation of the ORFs was them aimed to verify their involvement in crocagin production.

Mutagenesis experiments in *Chondromyces crocatus* Cm c5

To investigate the function of the genes identified, targeted insertional mutagenesis was applied with derivatives of the hygromycin-resistance conferring suicide plasmid pSUP_*hyg*. For gene

inactivation, an internal fragment of the target gene was amplified by PCR and cloned into the pSUP_*hyg*, yielding the respective plasmid derivative. After verification by restriction analysis, the plasmid was introduced into the donor strain *E. coli* ET12567/pUB307, which was used for biparental conjugation. *C. crocatus* Cm c5 clones exhibiting hygromycin B resistance were then genetically verified by PCR, as shown in fig. 5.2 and 5.3. One genetically verified Cm c5 mutant of *cpnB* and five of *cpnI* were obtained. Verified clones were grown in liquid culture and their extracts analyzed for crocagin production by LC-MS. In the MS-method applied, crocagins A and B are detected at a retetion time of 4.5 - 4.7 min as $[M + H]^+$ ions with an m/z of 538.2 and 552.2, respectively. Because crocagin A constitutes a minor metabolite, which is sometimes not detected at all, it was excluded from the analysis.

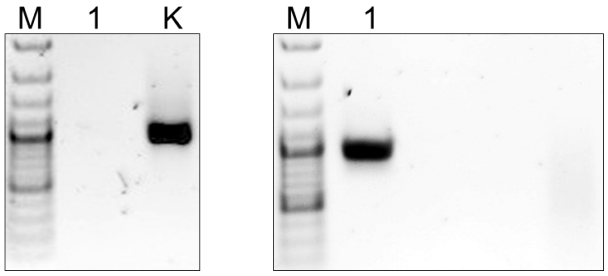


Figure 5.2: Genetic verification of Cm c5 *cgnB* mutants. **Left:** PCR to verify gene *cgnB* disruption. DNA template of the wild type yields the PCR product (lane K), while mutant DNA template yields no product (lane 1). **Right:** PCR to verify pSUP_hyg integration into the Cm c5 genome. Mutant DNA template yields the expected product (lane 1).

Figure 5.3: Genetic verification of Cm c5 *cgnI* mutants.**Left:** PCR to verify gene *cgnI* disruption. DNA template of the wild type yields the PCR product (lane K), while mutant DNA templates yield no product (lane 1-5). **Right:** PCR to verify pSUP_hyg integration into the Cm c5 genome. Mutant DNA templates yields the expected product (lanes 1-5).

The total ion chromatograms from crude extracts of mutant and wild-type cultures are shown in fig. 5.4. The peak at 4.7 min with an m/z of 552.2 (red) corresponds to the $[M + H]^+$ ion of crocagin B which is produced by the wild type. The peak at 5.1 - 5.4 min with an m/z of 872.3 (green) corresponds to the $[M+H]^+$ ion of crocadepsin, a non-ribosomal peptide from Cm c5 that is used as internal reference here. The cm c5 mutants with disrupted gene *cpnI* do not produce the crocagin mass with an m/z of 552.2, as seen in top panel. Instead, a mass peak with an m/z of 509.2 (blue)is detected, which corresponds to the non-carbamoylated crocagin B. This mass is considered a precursor and is not found in wild-type extracts. The mutants with disrupted *cpnB* produce neither crocagin nor any detectable precursors thereof, but crocadepsin production is not affected.

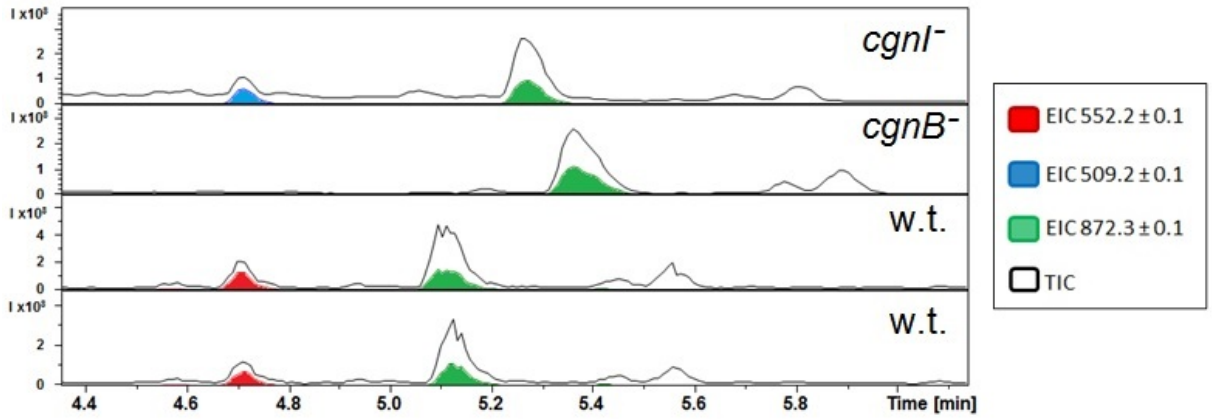


Figure 5.4: Total ion chromatograms (black line) of Cm c5 wild type and mutants of crocagin biosynthetic genes *cgnB* and *cgnI*. Crocagin B mass peak is shown in red. Inactivation of *cgnI* leads to exclusive formation of non-carbamoylated crocagin derivatives shown as blue peak. Inactivation of *cgnB* leads to complete loss of crocagin production with no detectable precursors or derivatives. The crocadepsin B mass peak (green) is used as internal reference.

Cloning of the Crocagin Gene Cluster

A detailed biosynthesis scheme invlving the gene products of the annotated ORFs of the *cgn*-cluster has not been developed yet. One important aspect would be the proof that the annotated *cgn*-cluster is able to produce crocagin in a heterologous system, thus exluding the necessity of essential genes elsewhere in the Cm c5 genome. Because the cluster appears to consist of at least two operons that are encoded on the opposite strands up- and downstream of the propeptide gene *cgnA*, the TAR-assembly strategy was chosen, because it allows recombination of linear DNA fragments, which is controlled by the homologous regions on their ends. The assembly strategy is shown in the upper panel in fig. 5.5.

The *cgn*-cluster was amplified as three PCR-products PCR1-3 that either overlap or carry homologous overhangs that are introduced via the oligonucleotide primers. The PCR products were then mixed with a linear capture vector which is a PCR product as well, consisting of a partial pCLY10 plasmid backbone and homologous fragments on each end, which were introduced via the primers. The fragments were chosen to be homologous to the beginning of the propeptide gene *cgnA* at one end and homologous to the region in front of the hypothetical protein gene *cgnH* on the other end. The four separate fragments were then assembled transformed into *S. cerevisiae* ATCC 4004247 cells, where assembly and circularization take place. Yeast clones were selected by LEU2 complementation in Leu-dropout medium and pre-screened with colony PCR for sequence motifs from the *cgn*-cluster. The plasmids of selected isolated, propagated in *E. coli* and analyzed by restriction. From the plasmids tested, 15 out of 16 yielded the correct restriction pattern, as shown in fig. 5.6.

Table 5.1: Calculated restriction patterns for pCLY_CGN cut with enzymes *Apa*LI or *Bsp*1407I.

***Apa*LI**	***Bsp*1407I**
6692	7952
4457	6511
3746	4026
3291	
303	

To allow expression in a heterologous host, the plasmid must be further modified with a

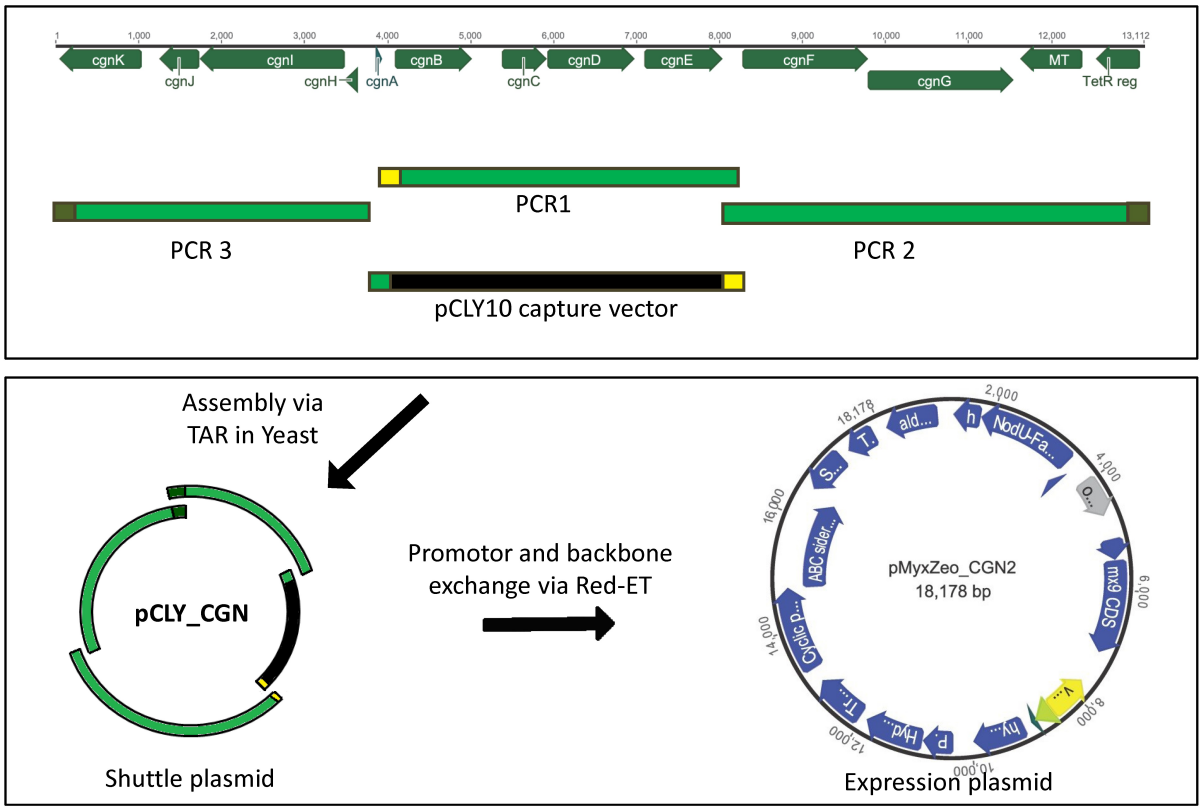


Figure 5.5: Cloning strategy for assembly of the crocagin cluster from PCR products by TAR.

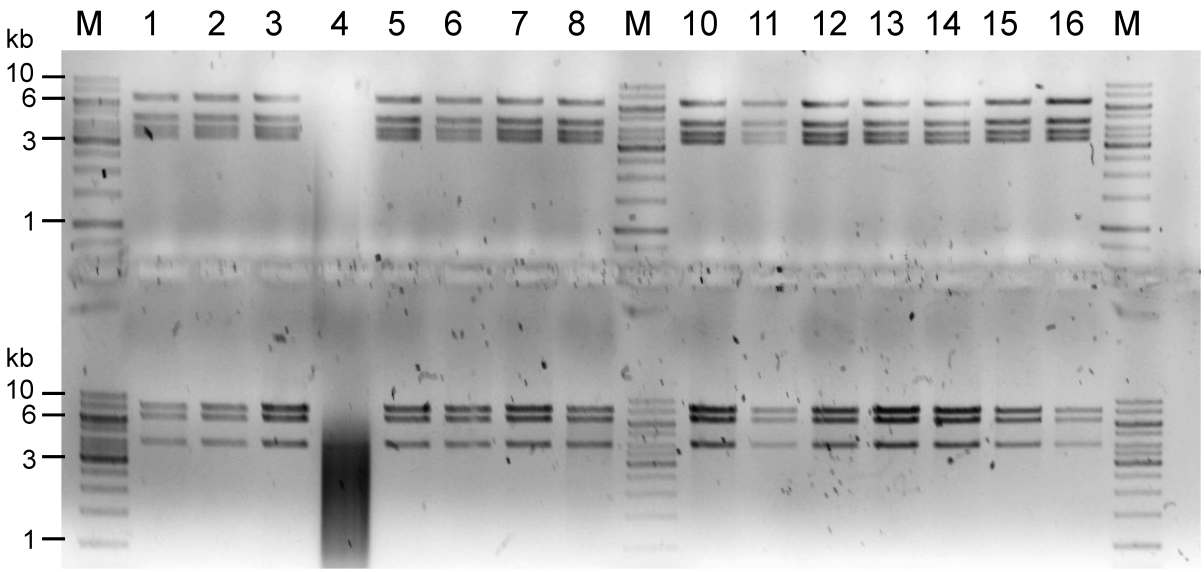


Figure 5.6: Restriction analysis of pCLY_CGN clones, cut with *Apa*LI (top) and *Bsp*1407I (bottom). M: kb-ladder. 1-16: Restriction analysis of different clones. All lanes except 4 show the expected restriction pattern, which is detailed in table 5.1.

selection marker and a promotor that is known to be functional in the host. *M. xanthus* DK1622 was chosen as an expression host, because it has been previously used for this purpose, and has desirable features such as fast growth and easy genetic manipulation. As a myxobacterium, it is most likely to be able to functionally express the biosynthetic genes from Cm c5. For inducible expression of the operon which includes genes *cgnA* - *cgnG*, the Pvan promotor was chosen. This promotor-repressor sequence allows tight control of transcription which is suppressed until the

expression is switched on with vanillic acid. For integration of the *cgn*-genes in DK1622, the mx9-integrase gene is used, which facilitates integration of the plasmid in the DK1622 genome at the mx9-attB locus. The integrase is encoded plasmid on the plasmid together with the zeocin resistance gene for clone selection in DK1622 and the vannilate promotor/repressor pair. Exchange of the plasmid backbone was performed as a one-step linear-to-circular recombination in *E. coli* GB-dir. This *E. coli* cloning strain contains inducible phage-derived recombination genes *recA* and *recB*. The plasmid pCLY_CGN was introduced into this strain as circular plasmid, along with the linear pMyxZeo backbone, which contained homologous sequences on each end which will allow recombination to the plasmid pMyxZeoCGN, which is shown in fig. 5.5 in the lower panel. Zeocin-resistant *E. coli* clones were then screened by restriction analysis, as shown in fig. 5.7. 11 out of 18 clones showed a restriction pattern that corresponds to the theoretically expected pattern, except for the second-smallest fragment at 991 bp. Instead, there is an additional band at 500 bp. This fragment corresponds to a region in the plasmid backbone. Furthermore, it is not possible to separate the three largest restriction fragments properly, as they are almost the same size. The construct will be sequenced before transformation into the heterologous host or further mutagenesis to exclude other potential mutations.

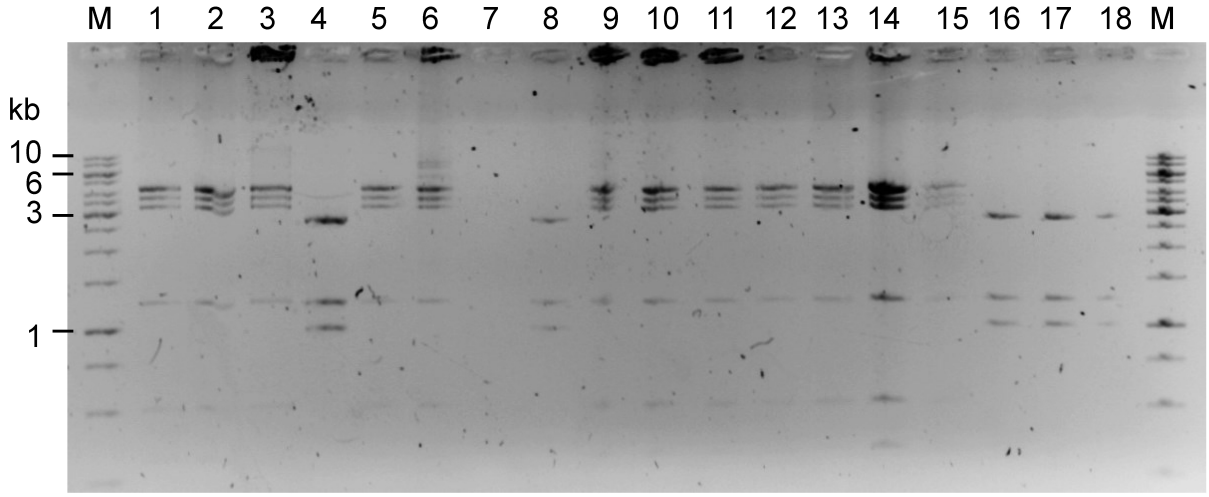


Figure 5.7: Restriction analysis of pMyxZeo_CGN clones, cut with *Apa*LI. M: kb-ladder. 1-18: Restriction analysis of different clones. The theoretically expected fragment size is shown in table 5.2. Clones 1-3, 5 and 9-14 show almost all of the expected fragments, except for the two smallest bands. This indicates an unwanted recombination event in the plasmid backbone.

Table 5.2: Calculated and observed restriction patterns for pCLY_CGN cut with restriction enzyme *Apa*LI.

calculated	**observed**
4533 bp	4.5 kb
4457 bp	4.5 kb
4402 bp	4.4 kb
3291 bp	3.3 kb
1246 bp	1.2 kb
991 bp	-
303 bp	0.5 kb

In the expression plasmid pMyxZeo_CGN, the native promotor from Cm c5 is driving transcription of genes *cgnHIJ*, and probably also *cgnK*, *tetR* and *MT*, encoded upstream of gene *cgnA* on the opposite strand. It is unclear, whether this promotor is functional in DK1622. Re-

placement of this region with a defined, inducible promotor can be achieved via a second direct cloning step. Transformation of DK1622 with pMyxZeo_CGN has not yet been pursued.

5.3 Experimental Procedures

Cm c5 Mutagenesis and Genetic Verification

Derivatives of the plasmid pSUP_*hyg* were used for mutagenesis of Cm c5, which contain a fragment that is homologous to the gene of interest. The plasmid integrates into the Cm c5 chromosome at the desired locus via a single crossing over event. The resulting mutant is then hygromycin resistant. Cm c5 was mutagenized by biparental conjugation as described previously.[108] The gene of interest is disrupted by the plasmid backbone, leading to a knockout mutant. To genetically verify the genotype of an isolated mutant, gDNA of mutant cells grown in liquid culture was extracted and plasmid integration tested with PCR reactions. The first PCR was performed with the respective verification primers that bind in the up- and downstream of the locus of plasmid insertion. This reaction must only yield a product in wild type cells or control mutants, as shown in Fig. 5.8A. In the mutant, the target gene is disrupted by the 8 kb plasmid backbone, thus no PCR product is formed. In the second PCR, one of the verification primers was combined with a primer that binds the plasmid backbone. The correct combination of these primers leads to a product only if the pSUP_hyg plasmid backbone has indeed integrated at the desired position in the Cmc5 genome as shown in Fig. 5.8B.

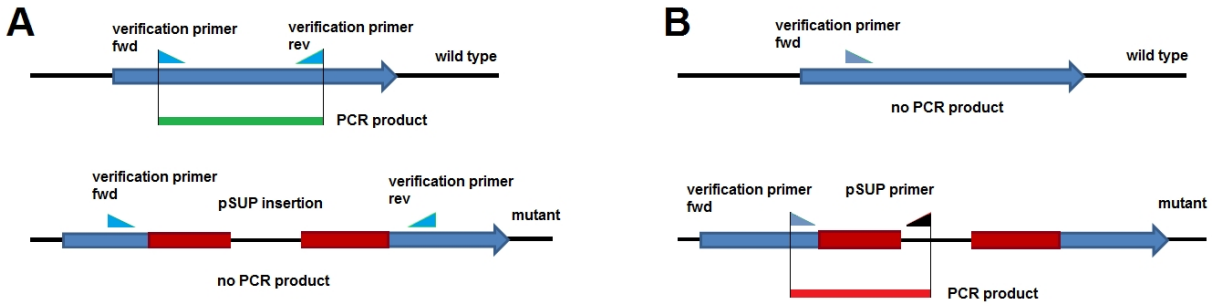


Figure 5.8: Genetic verification of Cm c5 mutants by PCR. **A:** PCR with primers that bind outside the plasmid integration site on the gene of interest. The mutants must not yield a product in this reaction. **B:** The PCR is performed with one primer binding on the plasmid sequence. Only the mutant DNA template must yield a product in this reaction.

Assembly of the *cgn*-Cluster via TAR

The overlapping DNA fragments of the *cgn*-cluster were obtained by PCR from Cm c5 gDNA using the primer combinations listed below in table 5.4. Phusion polymerase and buffers were used according to manufacturers manual, with a final DMSO concentration up to 8% in the reaction mixture. The linear capture vector was obtained by PCR from the circular pCLY10 template, using primers with an overhang that serves as homologous region for recombination with the corresponding sequence on the linear *cgn*-cluster fragments.

For TAR assembly of the *cgn*-cluster, 50 mL YPD medium were incubated with 1 mL overnight culture of yeast strain *S. cerevisiae* ATCC 4004247 and grown for 4h at 200 rpm and 30°C to an OD_{600}=2.0. In the meantime, the transfomation mix (see table 5.5) was prepared and stored on ice. About 100 ng of each PCR product, i.e. gel-purified PCR1, PCR2, PCR3 and the capture vector, were added to the mixture directly. About 4h after inoculation,

Table 5.3: Oligonucleotide Primers for PCR of homologous fragments of *cgn* genes

Purpose	**Primer name**	**Sequence 5'-3'**
Inactivation of *cgnI*	552carb_for	TGACCTTATGGCGAGCCCACCTTC
	552carb_rev	TTCCAGTCGTTACACGTCCGC
Inactivation of *cgnB*	552hypo_for	TGATTGATGCAGGCCTGGGAGAC
	552hypo_rev	TCACTCCTTGCCGACGATCAGCTC
Verification of *cgnI* mutants	552_carb_veri_for	AGTGCCTTCTGGGCGCTCTT
	552_carb_veri_rev	AAGCTGTGGACCGAGCGGAA
Verification of *cgnB* mutants	552_hypo_veri_for	TGCGTACTGCCCCCTTCGTT
	552_hypo_veri_rev	TCTTTCCCCATCCCCACGCC
Primers binding pSUP_hyg	psup_hyg_fwd	ATGTAGCACCTGAAGTCAGC
	psup_hyg_rev	GCATATAGCGCTAGCAGC

Table 5.4: Oligonucleotide Primers for PCR of *cgn* cluster

Purpose	**Primer name**	**Sequence 5'-3'**
PCR 1	cgn1_for	GGTCCGAATGACGTCCGGAAAT
	cgn1_rev	CGGATCACCTCGTTGATGCTGT
PCR 2	cgn2_for	CAGGACAGGAGTGGAGCTTTGG
	cgn2_rev	CCATTCCCGATACACACCCAGC
PCR 3	cgn3_rev	TGTGAACAATTCCTCTCGCCGT
	cgn3_for	CATGGGGGAGGATCTGCCGCTGGGTGTGTATC
		GGGAATGGGACATGGCCCGAGGTCCTG
capture vector	pCLY_CGN_for	CAAGGCCGAACGCCTCTTTCACGGCGAGAGGA
		ATTGTTCACATTCGGTGAACTTCGGGTACGC
	pCLY_CGN_rev	GTGGACACGAAAGCGTCCCGCATTTCCGGACG
		TCATTCGGACCTCGCCCTTTGACGTTGGAGTC

the yeast cells had grown to the required OD_{600}, and were pelleted at 4000 g at room temperature in a free-swinging rotor, washed twice with sterile deionized water and resuspended in 1 ml water. Aliquots of 100 μL, corresponding to 10^8 yeast cells each, were transferred to a 1.5 mL tube, one for each transformation plus two controls. The aliquoted cells were spun at top speed for 30 s in a tabletop centrifuge at room temperature, and the supernatant removed by pipetting. The transformation mix was added and mixed with the cells by careful pipetting up and down. Each tube with cells and transformation mix was placed in a tabletop incubator at 42°C for 90 min. After incubation, the tubes were centrifuged at top speed for 30 s and the supernatant removed. Cells were washed with 1 mL water, centrifuged for 30s and resuspended in 100 μL water. Each mix was spread completely on a separate YNB-Leu Agar plate and incubated at 30°C. Yeast clones appeared within 2-4 days and were first screened by colony PCR prior to inoculation in liquid YNB-medium.

Plasmid Recombination via Red-ET

For modification of plasmid pCLY_CGN to yield the expression plasmid pMyxZeo_CGN, linear-to-circular Red-ET recombination was applied, following standard protocols.[36] Briefly, pCLY_CGN was cloned into an engineered *E. coli* cloning strain designated "GB-dir", then grown to an OD_{600} of 0.2 and expression of the recombinase genes induced by addition of L-arabinose to a final con-

Table 5.5: Transformation mixture for TAR

PEG 4000 50%	240 μL
LiAc 1 M	36 μL
sheared salmon sperm DNA (2 mg/mL)	50 μL
PCR products	34 μL
DMSO	20 μL

centration of 3 mg / L, then further cultivated to an OD_{600} of 0.5, washed and electroporated with the linear fragment, in this case the pMyxZeo plasmid backbone with the homology regions at each side. Clones were selected on LB-agar containing Ampicillin and Zeocin, and screened by colony PCR and restriction analysis as described above.

Structure Elucidation

The NMR data and experimental details of crocagin structure elucidaion will be publicly available after publication of this manuscript.

5.4 Conclusions

The crocagins were found to be derived from a ribosomally assembled precursor peptide and not from an NRPS assembly line, although a the putative precursor peptide was not yet detected directly. The biosynthetic gene cluster was identified by manual search for a gene encoding the a putative precursor peptide sequence, and verified by inactivation of the adjacent biosynthesis genes, resulting in abolished crocagin production in the corresponding mutants. The exact biosynthetic pathway remains to be elucidated. Ribosomally derived bioactive peptides (RiPPs) consitute a vast and diverse group of secondary metabolites which are usually isolated from cyanobacteria, gram-positive bacteria such as *Bacillus* and eukaryotes. The crocagins presented herein belong to an unprecedented compound class, although a plethora of natural products with internally crosslinked indole moieties exist. Examples for RiPPS with such structural features are for expample the ComX-peptides from *Bacillus* species,[115] which are shown in fig. 5.9. Next to those, cross-linked indole rings are found all over nature: the numerous plant-derived indole alkaloids, bacterial indole terpene alkaloids[116] and a plethora of unique indole alkaloids from marine invertebrates, such as the variolines and securines.[117]

By genetic manipulaton of the producer strain, *C. crocatus* Cm c5, two biosynthetic genes *cgnB* and *cgnI* were confirmed to be involved in crocagin biosynthesis. These genes are adjacent to the putative propeptide gene, indirectly confirming the locus of the propeptide. The propetide gene itself can not be mutangized in Cm c5 by targeted insertion because of its small size, encoding only 21 amino acids. Crocagins are derived from the ultimate three amino acids of their propeptides' C-terminus, which unusual for RiPPs. The ComX peptides are also formed from the 6 C-terminal residues of their 58 aa propeptides.

A heterologous expression strategy for the crocagins is aspired, because the gene cluster is small compared to NRPS assembly lines and lacks their repetitive and modular structure, which often make cloning and modification of such pathways difficult. Assembly of the gene cluster from PCR-products via TAR cloning was chosen because it allows to freely organize the indiviual operons, making subsequent modifications like promotor exchanges easier. The expression construct pMyxZeo_CGN is designed for transfer into the genome of *M. xanthus* DK1622, where it integrates via the mx9-integrase. DK1622 was chosen as host for this myxobacterial pathway, because it is an easy manipulable host with good growth characterisitcs and has been used

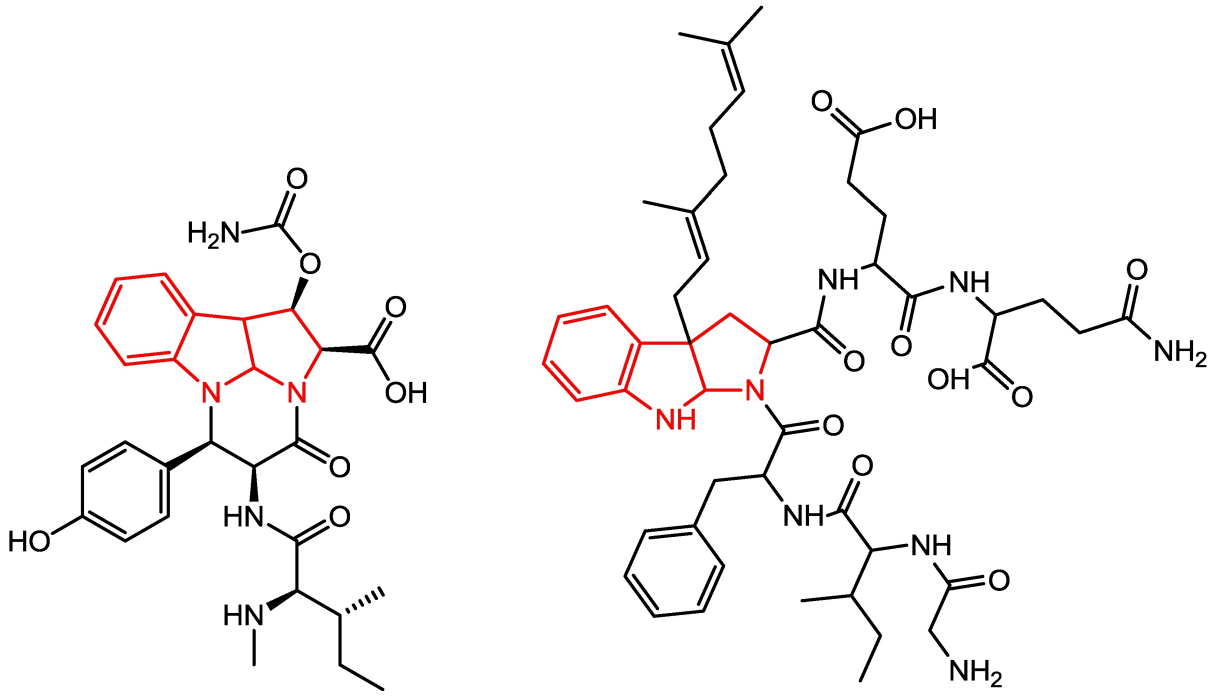


Figure 5.9: Comparison of the chemical structures of crocagin B and comX-RO-E-2. Both RiPPs share the Trp-derived indole-proline ring system highlighted in red.

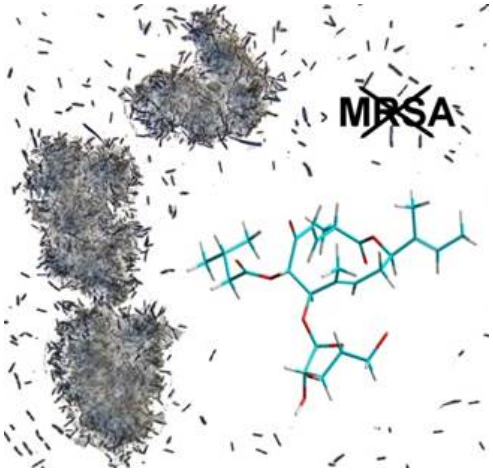
for heterolous expression previously. However, crocagin expression could also be tested in other hosts such as *E. coli* or *B. substilis*, the latter being known to produce RiPPs as quorum sensing pheromones. This would however require exchange of the plasmid backbone and promotor replacement. A heterologous system would then finally allow for markerless deletion of each ORF in the cluster to precisely study their function in crocagin biosynthesis.

The crocagins show neither antibiotic nor cytotoxic activity. Their biological function remains unclear, as the respective crocagin knockout mutant cultures of Cm c5 show no obvious changes in growth and swarming. Because of their similarity to the ComX-peptides, extracellular pheromones from *Bacillus* sp., a similar function is supposed for crocagin. ComX pheromones are quorum-sensing pheromones that enable cellular competence (uptake of DNA from the environement) at high cell density. The pheromones bind a two-component regulator on *Bacillus* cells, triggering a signal cascade that results in expression of competence genes.[118] Similar quorum-sensing (QS) systems are known for many pathogenic bacteria, such as *Staphylococcus aureus* or *Pseudomonas aerigunosa*, where the production of virulence factors is often triggered at high cell density. Crocagins should therefore be tested as inhibitors for such QS systems. QS inhibition is a promising approach to combat bacterial infections, especially from toxin-producing pathogens.[119,120] Contrary to antibiotic treatment, QS inhibition does not apply selection pressure towards resistant clones, as QS inhibition does not impair the viability of the targeted cells.

Chapter 6

Disciformycins A and B: 12-Membered Macrolide Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* Active against Multiresistant Staphylococci

Frank Surup, Konrad Viehrig, Kathrin I. Mohr, Jennifer Herrmann, Rolf Jansen, Rolf Müller*



This chapter was published as an article in *Angewandte Chemie International Edition* on October 7th, 2014. Adapted with permission from "Disciformycins A and B: 12-Membered Macrolide Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* Active against Multiresistant Staphylococci". Permission granted in both print and electronic formats.

*To whom correspondence should be adressed.

Author's efforts

The author's efforts regarding this publication comprise the identification and annotation of the disciformycin biosynthesis gene cluster, the biosynthesis scheme and subsequent in-depth analysis of the sequence for corroboration of the observed stereochemistry, as well as the writing of the respective parts of the manuscript.

Contributions by others

The compound was detected in activity screenings carried out by Katrin Mohr and Rolf Jansen. Isolation and structure elucidation of disciformycin were conducted by Frank Surup and Rolf Jansen, further bioactivity testing was carried out by Jennifer Hermann. The project was supervised by Prof. Rolf Müller, who also contributed to conceiving and writing of the publication.

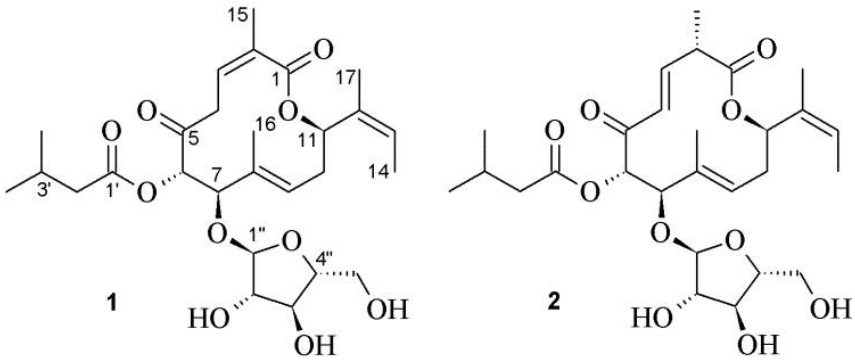


Figure 6.1: Structures of disciformycin A **(1)** and B **(2)**.

Abstract

Two macrolide glycosides with a unique scaffold were isolated from cultures of the myxobacterium *Pyxidicoccus fallax*. Their structures, including absolute configurations, were elucidated by a combination of NMR, MS, degradation, and molecular modeling techniques. Analysis of the proposed biosynthetic gene cluster led to insights into the biosynthesis of the polyketide and confirmed the structure assignment. The more active compound, disciformycin B, potently inhibits methicillin- and vancomycin-resistant *Staphylococcus aureus*.

6.1 Introduction

Staphylococcus aureus causes severe hospital- and community-acquired infections and consequently the steadily rising occurrence of multiresistant and methicillin-resistant *S. aureus* (MRSA) poses a serious public health threat.[121] Consequently, development of novel antibiotics active against such pathogens is urgently needed.[122] Myxobacteria are a prolific source of novel bioactive secondary metabolites, for example the anticancer drug epothilone,[123] the antibacterial sorangicin,[124] and the antifungal soraphen.[125] In order to access the tremendous genetic potential of yet unexplored metabolites,[83] a selection of our myxobacterial strain collection comprising 8550 species is currently being re-assessed for the production of novel antibiotics.[106,126] In this rescreening, *Pyxidicoccus fallax* strain AndGT8 stood out with a bioactivity against Gram-positive indicator bacteria. Because none of the UV-MS-characterized compounds in the bioactive fractions obtained by RP-HPLC was included in our internal database "Myxobase"[127] or in the "Dictionary of Natural Products",[128] this strain was chosen for comprehensive examination.

6.2 Results and Discussion

Bioactivity-guided isolation of an initial cultivation on a 10 L scale led to the isolation of 1.0 mg of disciformycin A **(1)**; *N. flava* was used as the indicator organism. HRESI-MS provided the elemental composition $C_{27}H_{40}O_{10}$ implying eight double-bond equivalents. Interpretation of the HMQC spectrum (CD_3OD) enabled the assignment of all protons to their respective carbon atoms, leaving only three exchangeable protons.
^{1}H,^{1}H-COSY and ^{1}H,^{1}H-TOCSY correlations showed six ^{1}H spin systems (Figure S1, Table S1, Supporting Information). These partial structures were subsequently linked by series of ^{1}H,^{13}C HMBC correlations: the correlation of methyl group H_3-15 to C-1, of methylene H_2-4 and methine H-6 to C-5, of methine H-7 and methyl H_3-16 to C-8, and of H-11 and H_3-17 to C-12 established the carbon skeleton of the aglycon. The low-field shift of H-11 (δ_H=5.61 ppm) indicated an ester linkage at this position which was verified by a HMBC correlation between H-11

and C-1, establishing the lactone ring closure of the aglycon. The configuration of the methyl-substituted double bonds was derived from ROESY correlations. A strong ROESY correlation between H_3-15 and H-3 indicated the *Z* configuration of the $\Delta^{2,3}$ double bond, while the ROESY correlation between H_3-16 and H_a-10 supported an E configuration of the $\Delta^{8,9}$ bond. Finally, a ROESY correlation between H-13 and H_3-17 indicated the $\Delta^{12,13}$ Z configuration in the side chain of the aglycon. A ^{1}H,^{13}C HMBC correlation between H-6 and C-1 proved the ester linkage of 3-methylbutyric acid to C-6, simultaneously explaining the low-field shift of H-6 (δ_H=5.05 ppm).

The ^{13}C NMR data of **1** were characteristic for an α-arabinofuranose configuration (observed δ_C=110.6, 83.8, 79.1, 86.3, 63.0 ppm; methyl furanoside δ_C=109.3, 81.9, 77.5, 84.9, 62.4 ppm).[129] The absolute D-(-)configuration of the arabinose moiety was determined by GC-MS comparison of the (-)-2-butyl glycoside derivative to authentic standards.[130]
The relative configuration of the aglycon of **1** could be derived from vicinal coupling constants and ROESY correlations (Figure S2). The large coupling constant of 9.5 Hz between H-6 and H-7 and the absence of a ROESY correlation indicated their *trans* configuration. Strong NOEs were observed for H-6 and H_b-10 with methyl H_3-16, but not with H-11. On the other side H-7 showed a strong NOE with H-9, which itself correlated to H-11 and H_a-10, indicating a cisoidal relation between H-7 and H-11.
The absolute configuration of the aglycon was derived from the previously determined absolute configuration of the D-(-)-arabinosyl residue: A strong ROESY correlation between H-1" and H-7 confirmed the typical conformation of the glycoside in solution. Consequently, a weak but unambiguous correlation between H-4" and methyl H_3-16 was used to assign the configuration of the chiral center C-7 (Figure 6.1). Thus the absolute configuration of the disciformycin A aglycon **(1)** was assigned as 6*S*,7*R*,11*R*.
After the strain and culture conditions had been optimized, a 70 L fermentation of strain AndGT8 provided 25.4 mg of **1** as well as 7.6 mg of isomer **2** with identical molecular formula $C_{27}H_{40}O_{10}$. 1D and 2D NMR data showed that the only difference between **1** and **2** was the shifting of the $\Delta^{2,3}$ double bond to position $\Delta^{3,4}$ with *E* configuration, which was indicated by the large coupling constant ($J_{3,4}$=15.3 Hz). Coupling constants and ROESY correlations of the C-6 to C-11 part remained largely unchanged compared to **1** (Table S2, Supporting Information). Therefore, a 6*S*,7*R*,11*R* configuration can be concluded for **2** as well.

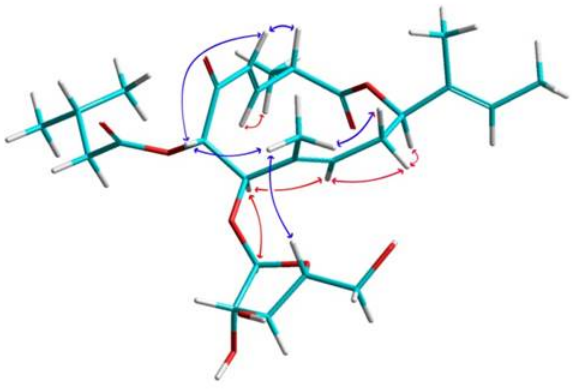


Figure 6.2: Model of **2** calculated by PM3 with HyperChem with selected ROESY correlations (blue above, red below main plain) indicating the conformation of disciformycin B.

To determine the configuration at C-2 a structure model of compound **2** was calculated by PM3 with HyperChem (Figure 6.2). Due to the rigid structure elements, that is, the double bond, the α,β-unsaturated ketone, and the ester, the core part of **2** is locked in a twisted configuration, with protons H-2, H-4, and H-6 and methyl group H_3-16 located above the main plane.

Strong ROESY correlations between H-4 and H-2, and between H-3 and the methyl group H_3-15 indicate a 2*S* configuration.

For further corroboration of the observed structure we sequenced and screened the AndGT8 genome sequence for a suitable gene cluster using antiSMASH.[32] A candidate gene cluster was found in the partial genome sequence encoding a linear type I PKS that consists of six separate PKS modules with a domain organization in line with the observed reduction stages of the polyketide. The core biosynthetic gene cluster is shown in Figure 6.3A, with identified open reading frames (ORFs) and their assumed function listed in Figure 6.3C. ORFs *difB* to *difG* encode the PKS genes, *difA* encodes the tailoring cytochrome P450. The functions of *difH* and *difI* are yet unclear. The PKS domain organization, as obtained from sequence analysis with the NRPS/PKS predictor[29] is shown in Figure 6.3B, together with a biosynthesis proposal.
Assembly of the polyketide starts at DifG, which resembles a fused starter and extender module, as observed previously in the stigmatellin synthase.[131] The polyketide is then extended at modules DifB-DifF and released by macrocyclization. DifF contains a domain of unknown function between the DH and the KR domain. This position is usually occupied by an ER domain in fully reductive modules. A Rossman fold for NADP(H) binding is found in the sequence by BLAST analysis, but other typical sequence motifs of ER domains are absent, as well as other hints for its potential function.
The substrate specificity of the seven AT domains was predicted by multiple alignments of their respective amino acid sequences and comparison of key residues to published data.[132] The predicted substrate of the AT domains from modules DifB, DifD, DifE, and DifG1 are malonyl-CoA, while AT domains from modules DifC, DifF, and DifG2 are predicted to be specific for methylmalonyl-CoA, as observed in the disciformycin molecule. A detailed alignment is presented in the Supporting Information.

To further validate our biosynthesis proposal, the *dif* KR domain sequences were screened for described motifs, which allows the prediction of the configuration of the hydroxy groups and double bonds in the molecule.[54,53] Based on conserved sequence motifs, most KR domains can be categorized as either A or B type, which reduces their ketoacyl substrate to either *S*- or *R*-configured secondary alcohols, respectively. In the case of elimination after keto reduction, the configuration of the cryptic KR-derived alcohol can be predicted with this method. This approach has been used previously to determine the stereochemistry of molecules in which the absolute configuration could not be determined by experimental methods.[133,134]
In the case of disciformycin, the *R*-configured hydroxy group at carbon atom C-7 results from keto reduction by the KR of DifD, which shows B-type sequence features. The KR of DifB, which reduces the keto group at C-11, cannot be categorized as A or B type. For the *Z*-$\Delta^{12,13}$ double bond and the *E*-$\Delta^{8,9}$ double bond, the KR domain sequences correspond to a A-type KR for DifG and a B-type KR for DifC, respectively.
The DifF KR domain is predicted to be B type; therefore an *R*-hydroxy intermediate is expected to be the substrate for the DH domain of DifF, which is then processed to either a *Z*-$\Delta^{2,3}$ double bond in **1** or an *E*-$\Delta^{3,4}$ double bond in **2**. Therefore, a double-bond shift has occurred in **2**. Such double-bond shifts in PKS products have been reported previously, e.g. in rhizoxin,[135] ansamitocin,[55] corallopyronin,[78] and bacillaene,[136] and have been shown to take place during biosynthesis while the substrate is tethered to the PKS. Since the *dif* cluster contains no separate "shifting module" as is the case for the corallopyronin cluster, we expect the double-bond migration to be introduced by the DH domain of DifF, as observed in ansamitocin biosynthesis. To our knowledge, no sequence motifs have been described yet by which a DH domain could be predicted to perform such double-bond shifts along with the dehydration step, and future work is required to clarify this question in disciformycin biosynthesis.
The OH group at C-6, to which the valerate is attached in the final compound, must result from hydroxylation by a tailoring enzyme during or after assembly, most likely a P450 enzyme

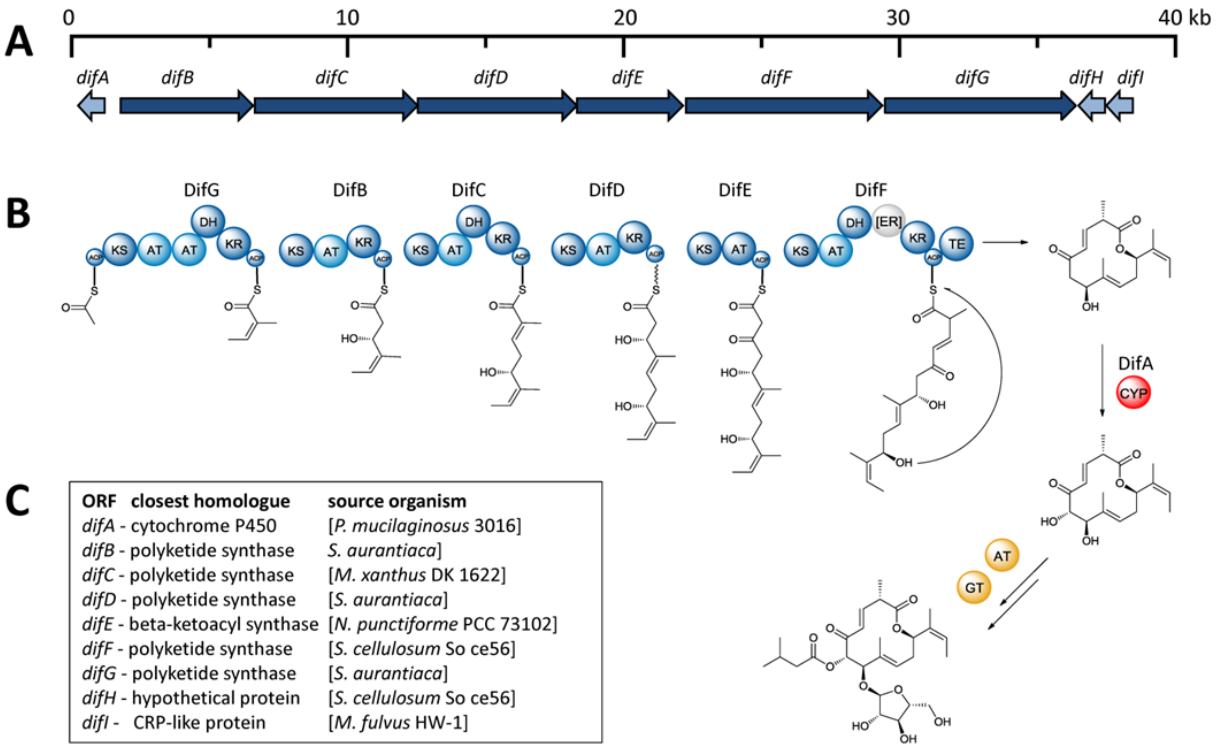


Figure 6.3: A: Disciformycin biosynthetic gene cluster. PKS genes are displayed in dark blue, other ORFs are shown in light blue. **B:** Proposed disciformycin B assembly. The hydroxy group at C6 is introduced by cytochrome P450 enzyme DifA. The tailoring glycosyltransferase and acyltransferase are yet unknown and must be located elsewhere in the genome. ACP = acyl carrier protein domain; AT = acyl transferase (domain); CYP = cytochrome P450; DH = dehydrogenase domain; ER = enoyl reductase domain; GT = glycosyl transferase; KS = ketosynthetase domain; KR = ketoreductase domain; TE = thioesterase domain. **C:** Annotated ORFs in the dif cluster and closest homologues found by BLAST-search.

encoded by *difA*. Suitable genes for the acylation and glucosylation could not be found in the vicinity of the PKS gene cluster, but numerous such functionalities are encoded in the genome. Currently, further studies are underway to identify the missing functionalities. However, despite serious efforts the producing strain currently cannot be genetically manipulated.[137,138]

The bioactivity assessment of **1** and **2** revealed strong antimicrobial activity (Tables 6.1 and S3) against Gram-positive bacteria, for example, *Bacillus subtilis* DSM-10. Disciformycins were found to be especially active against staphylococci, such as *S. carnosus* DSM-20501 and *S. aureus* Newman.[139] In addition, both tested strains of methicillin-resistant *S. aureus* (MRSA) were inhibited: *S. aureus* DSM-11822 and *S. aureus* N315.[140] Moreover, these MRSA strains show reduced susceptibility to other antibiotic classes, such as macrolides and quinolones, indicating a novel target to be addressed by disciformycins. The MIC values were in the range of the reserve antibiotic vancomycin (Table S3). Importantly, no cross-resistance was observed to vancomycin, which is currently in use for human therapy as the antibiotic of last resort; the disciformycins displayed pronounced activity against the methicillin- and vancomycin-resistant *S. aureus* (MRSA/VRSA) Mu50 (ATCC 700699). Importantly, at concentrations of up to 10 μm no cytotoxicity was found against human HCT-116 colon carcinoma cells, murine fibroblast-like L929 cells, and Chinese hamster ovary CHOK1 cells.

Table 6.1: Minimum inhibitory concentrations for selected bacteria (MIC in μg/mL) and half-inhibitory concentrations for cell lines (IC_{50} in μM) of disciformycin A **(1)**, and disciformycin B **(2)** and vancomycin **(VAN)**.

	1	**2**	**VAN**
Bacillus subtilis DSM-10	4.2	0.8	0.25
Nocardioides simplex DSM-20130	33.3	16.6	0.42
Paenibacillus polymyxa DSM-36	16.6	16.6	-
Staphylococcus carnosus DSM-20501	7.8	2.4	0.25
Staphylococcus aureus Newman	8.0	1.2	0.5
Staphylococcus aureus DSM-11822 (MRSA)	4.0	0.6	1.0
Staphylococcus aureus N315 (MRSA)	8.0	1.2	1.0
Staphylococcus aureus Mu50 (MRSA/VRSA)	2.0	0.6	16.0
Colon carcinoma cells HCT-116	>10	>10	-
Murine fibroblast-like cells L929	>10	>10	-
Chinese hamster ovary cells CHO-K1	>10	>10	-

Conclusions

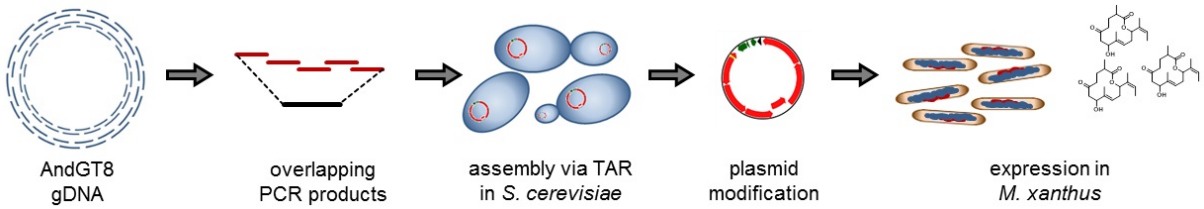
In conclusion, we present the structure elucidation of two new antibiotics from *Pyxidicoccus fallax* AndGT8 which were discovered by rescreening our myxobacterial strain collection. Identification and analysis of the fundamental biosynthetic gene cluster properly matches the stereochemistry determined by structure elucidation for the more active metabolite **2**, which is most likely the primary product of the biosynthesis. Disciformycins A **(1)** and B **(2)** might be suitable starting points for antibiotic development; ongoing work in our laboratory aims to identify the presumably novel bacterial target of these antibiotics. In any case, the compounds provide a novel scaffold for future total synthesis approaches, which will provide insight into their structure-activity relationship.

Supporting Information

Full experimental details, including structure elucidation procedures, in-depth bioinformatic analyses and NMR spectra, can be found in the Supporting Information attached in the appendix on page 203.

Chapter 7

Heterologous Expression of Disciformycin Precursors in the Mxyobacterium *Myxococcus xanthus*



Author's efforts

The author's efforts regarding this chapter comprise design and conduct of the cloning strategy and all experimental work, except for the contributions listed below.

Contributions by others

David Auerbach conducted isolation of **P1** and **P2** and structure elucidation via NMR and MS measurements. Nestor Zaburannyi analyzed and assembled the raw sequence data of the expression construct after Illumina-sequencing. *In vitro* activity assays of **P1** and **P2** with *S. aureus* RNAP were conducted by Jennifer Herrmann. The project was supervised by Prof. Rolf Müller.

Abstract

Disciformycins are unprecedented macrolide-glycoside antibiotics isolated from the myxobacterium *Pyxidicoccus fallax* AndGT8. They potently inhibit methicillin- and vancomycin-resistant MRSA-strains and show very weak cytotoxicity, making them a promising lead structure for MRSA treatment. The disciformycin molecule is assembled by a polyketide-synthase (PKS) encoded by six PKS-genes and a number of yet unidentified tailoring enzymes. To further study disciformycin biosynthesis, the PKS gene cluster was cloned and assembled via transformation-associated recombination (TAR) in *S. cerevisiae*. Integration of the gene cluster into the genome of the myxobacterium *Myxococcus xanthus* DK1622 led to production of a polyketide macrolactone, resembling a disciformycin precursor in amounts that exceed those in the original producer. The heterologous system allows to further investigate yet unresolved aspects of disciformycin biosynthesis and and opens up possibilites for mutasynthesis efforts.

7.1 Introduction

Disciformycins A **(1)**and B **(2)**, isolated from *Pyxidococcus fallax* AndGT8, are antibiotic macrocyclic polyketides which are glycosylated and acylated. Both compounds showed strong antibiotic activity against gram-positive bacteria, including methicillin- and vancomycin-resistant MRSA-strains. Their weak cytotoxicity makes them a promising lead structure for MRSA treatment. Both coumpounds differ from another by the position of a double bond in the molecule, which is Z-$\Delta^{2,3}$ in **1** and E-$\Delta^{3,4}$ in **2**. The position of this double bond leads to a different spatial structure of the compounds which has an effect of their bioactivity, with **2** having a 10-fold higher antibacterial activity against *S. aureus* than **1**.

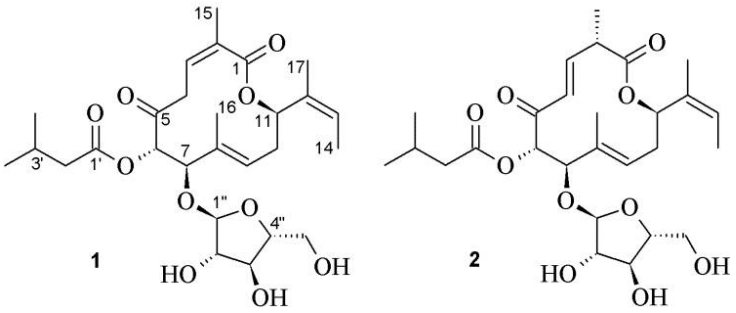


Figure 7.1: Chemical structures of Disciformycin A **(1)**and B **(2)**.

In silico sequence analysis of the *Pyxidococcus fallax* AndGT8 genome was performed, and a candidate polyketide synthase (PKS) gene gene cluster was found for disciformycin production. A biosynthesis proposal was elaborated and the domain organization of the PKS assembly line observed in the sequence is in agreement with the observed structure. Deeper sequence analysis of the PKS genes *difB-difG* further allowed prediction of the acyltransferase domains' substrate specificity and the stereoselectivity of the encoded ketoreductase domains, which was overall in accordance with the observed absolute configuration of the molecule. The biosynthesis proposal of disciformycin B is shown in fig. 7.2. After release the polyketide from the last PKS module by internal macrocyclization, post-assembly tailoring steps are expected to introduce the arabinose side-chain at C5 as well as a hydroxy-group to C6, to which isovaleric acid is attached via an ester bond. The hydroxylation at C6 is expected to be introduced by the cytochrome P450 gene *difA*, which is encoded in front of the PKS genes. Candidate genes for glycosylation at C5 and acylation at C6 could not be found near the PKS gene cluster and must be located elsewhere in the genome. The other unresolved issue in disciformycin biosynthesis is the double-bond shift

at C3 in **2**, and the reason for the production of metabolites containing either the shifted or the non-shifted double-bond at this position. The double bond shift is expected to be introduced during assembly on module DifF, which incorporates methylmalonate as the last building block. Therefore, precursors of **1** and **2** are expected to be produced by the PKS and modified later on.

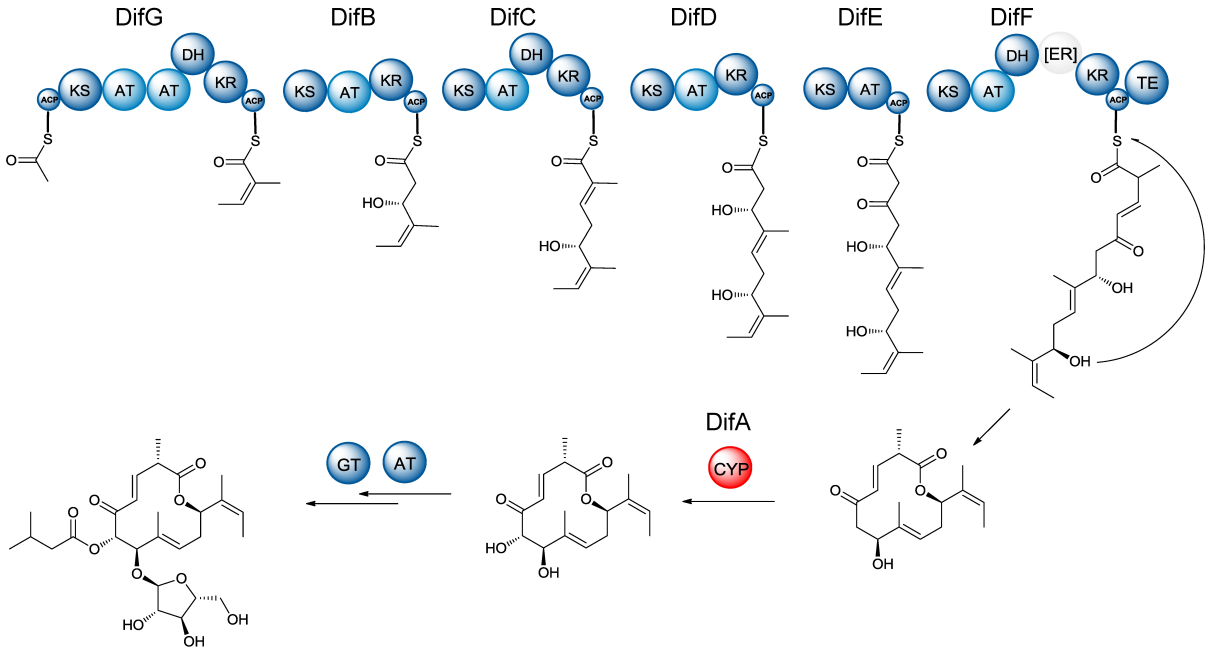


Figure 7.2: Proposed disciformycin B assembly. The hydroxy group at C6 is expected to be introduced by the tailoring cytochrome P450 encoded by *difA* after polyketide assembly. The tailoring glycosyltransferase and acyltransferase are yet unknown and must be located elsewhere in the genome.

The disciformycin producer *Pyxidococcus fallax* could not be mutagenized so far with the genetic tools available in our lab, although related strains were amenable to genetic manipulation through electroporation. As an alternative strategy to investigate disciformycin biosynthesis, a heterologous expression strategy was chosen. To circumvent generation of a genome library, a cloning strategy from PCR products via TAR was chosen, because the disciformycin PKS gene cluster is only about 30 kb in size. TAR (transformation-associated recombination) is a direct-cloning strategy that allows linear-to-linear homologous recombination of two or more DNA molecules with suitable homologous overhangs to be assembled as a replicating plasmid in suitable yeast (*S. cerevisiae*) cells. The strategy for cluster assembly is shown in fig. 7.3.

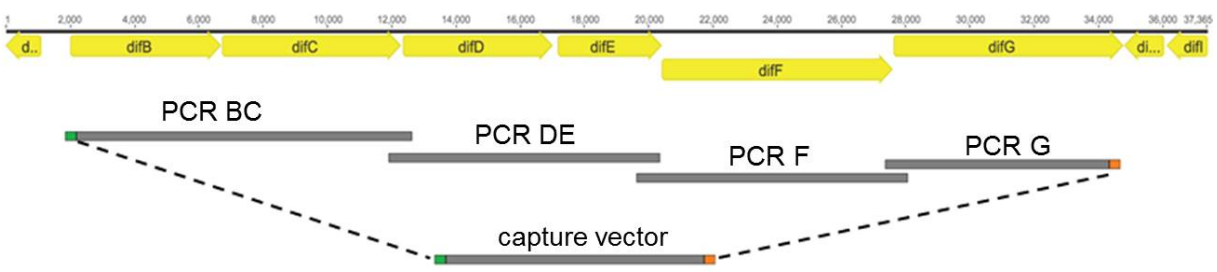


Figure 7.3: Assembly of the *dif*-PKS gene cluster by TAR. The six PKS genes *difB*-*difG* were amplified as 4 overlapping PCR products. In parallel, a part of the pCLY10 plasmid was amplified by PCR, carrying homologous sequences for assembly at each end.

7.2 Results and Discussion

Cloning of the Disciformycin PKS genes

The PKS genes were amplified as PCR products that overlap with the neighboring fragment over 30-150 bp. In addition, a part of the pCLY10 plasmid with homology arms and selection markers for propagation in *S. cerevisiae* and *E. coli* was generated by PCR, with the homology arms of 40 base pairs introduced via the oligonucleotide primers. An equimolar mixture of all purified PCR products was then used for transformation of LEU2-deficient *S. cerevisiae* ATCC4004247. Clones were selected on leucine-free selection medium, which is complemented when a replicating plasmid has been assembled from the linear DNA fragments. The map of the correctly assembled plasmid pCLY10_BG is shown in fig. 7.4.

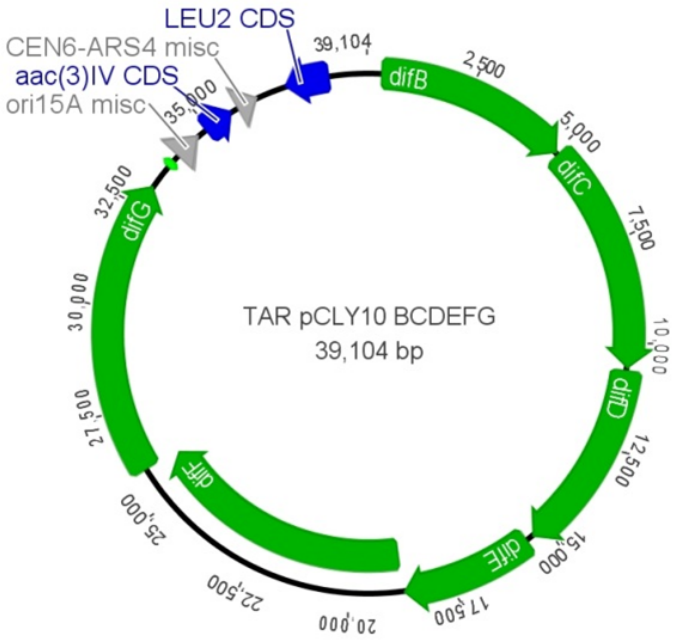


Figure 7.4: Plasmid map of pCLY10_BG shuttle vector.

The obtained *S. cerevisiae* clones were screened with colony PCR for correct assembly of the plasmid using primers that only yield a product if the capture plasmid fragment has been joined with the respective adjacent *dif*-PCR product on each side. 13 out of 16 clones yielded the expected band at around 0.6 kB, as shown in fig. 7.5. These clones were then cultivated, their plasmids isolated and transformed into *E. coli* for plasmid propagation and subsequent restriction analysis. Figure 7.6 shows the restriction analysis of pCLY10_BG plasmids. Clones 3, 7 and 8 exhibited the correct fragment patterns.

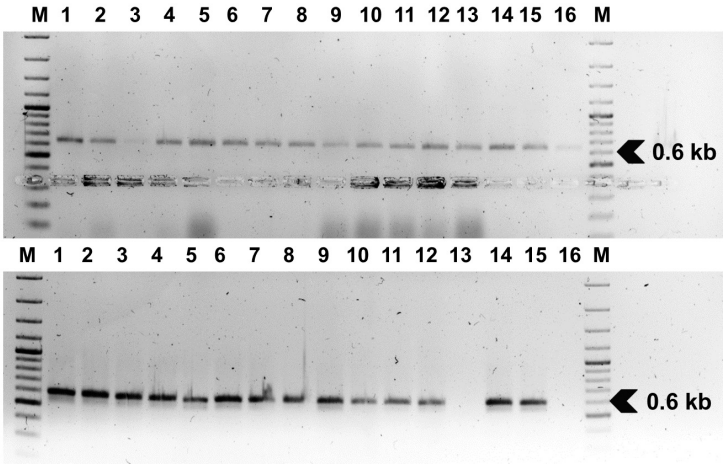


Figure 7.5: Colony PCR of pCLY10_BG yeast clones. Top: PCR for sequence "KB". bottom: PCR for sequence "KG". 13 out of 16 clones yielded the expected bands at around 0.6 kB. M: Fermentas "100bp plus" DNA Ladder

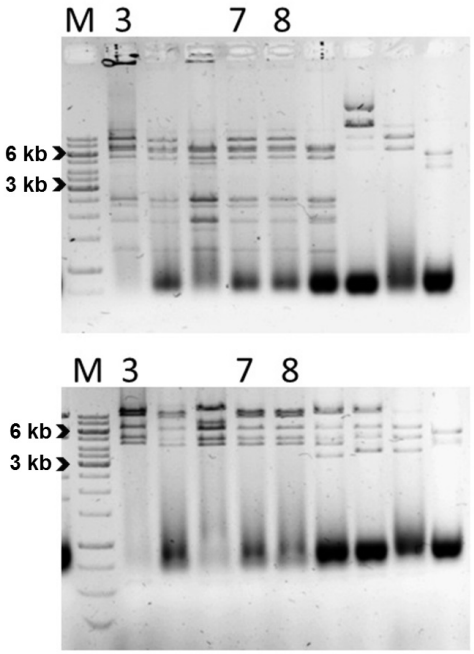


Figure 7.6: Restriction analysis of *E. coli* transformants with pCLY10_BG. A: Plasmid DNA was cut with *Apa*LI. B: Plasmid DNA was cut with *Not*I. The expected bands were observed for clones 3, 7 and 8. M: Fermentas kb-Ladder.

The pCLY10 shuttle vector backbone was then replaced with the pmyxZeo backbone, which contains the necessary resistance genes and Mx9-integrase for integration into the mx9-attB site in the genome of *M. xanthus* DK1622. To enable inducible expression of the PKS genes in the heterologous host, the pMyxZeo_Pvan backbone, containing the vanillate-inducible Pvan promotor[141] was used. The map of the resulting plasmid pMyxzeo_dif is shown in fig. 7.7. This step was carried out via Red-ET cloning as a linear-to-circular one-step strategy. The capture vector pmyxzeo_pvan was amplified by PCR with primers that contain a 42 bp homologous sequence of *difB* and *difG* for recombination with pCLY10_B. Ampicillin- and zeocin-resistant *E. coli* clones were screened by colony PCR for presence of two sequene parts of the PKS genes. The agarose gel with the colony PCR products of the tested clones is shown in fig. 7.8. 11 clones that showed products for both tested sequences were grown in liquid and their plasmids recovered for restriction analysis. The agarose gel with the restriction fragments is shown in fig. 7.9. Two

clones were found to contain the plasmid with correct restriction fragments, clones 16 and 20. These two plasmids were then selected for mutagenesis of *M. xanthus* DK1622 to test for their integration, subsequent heterologous expression and formation of the expected polyketide product.

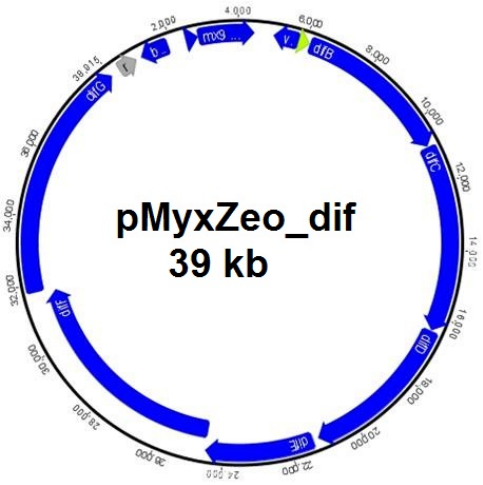


Figure 7.7: Plasmid map of pMyxZeoDIF, containing the core PKS genes of the disciformycin PKS gene cluster, driven by the vanillate-inducible Pvan promotor.

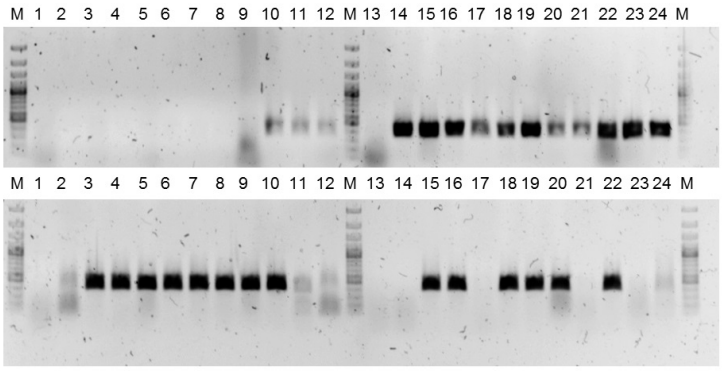


Figure 7.8: Colony-PCR of *E. coli* clones, to test for presence of PKS fragemtns on recombinant pMyxzeoDIF plasmid. Top: sequence fragment "BC". Bottom: sequence fragment "EF".

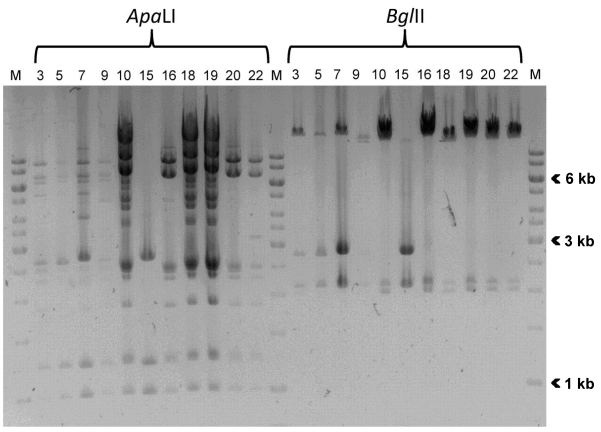


Figure 7.9: Restriction analysis of pMyxzeoDIF plasmid clones. M: Fermentas kb-Ladder. Clones 16 and 20 show the expected pattern.

Heterologous Production of P1 and P2 in *M. xanthus* DK1622

After electroporation of *M. xanthus* DK1622 with pMyxZeo_DIF 16 and 20, clones were selected on CYE-agar with zeocin. Randomly selected *M. xanthus* clones were then grown in liquid and their gDNA extracted for genetic verification, which was performed by PCR. There the presence of parts of the *dif*-genes and integration of the construct into the attB site of the DK1622 genome was assessed. One genetically verified clone of plasmid pMyxZeo_dif16 and pMyxZeo_dif20 was then grown in triplicates in a 20 ml scale in liquid culture and induced with either 1 mM or 2 mM potassium vanillate (Kvan) when the culture showed visible planctonic growth. For comparison, also a non-induced culture and control cultures carrying another heterologous gene cluster integrated in its attB site were co-cultivated. Cultivation was continued until the cultures were turning brownish, indicating the dying phase. Cells and growth medium were then extracted twice with ethyl acetate, the extracts dried, redissolved in methanol and subjected to MS-analysis. The chromatograms of extracts from three *M. xanthus* DK1622 attB::DIF20 cultures and one DK1622 attB::CPN13 culture (carrying another gene cluster for comparison) were compared, as shown in fig. 7.10.

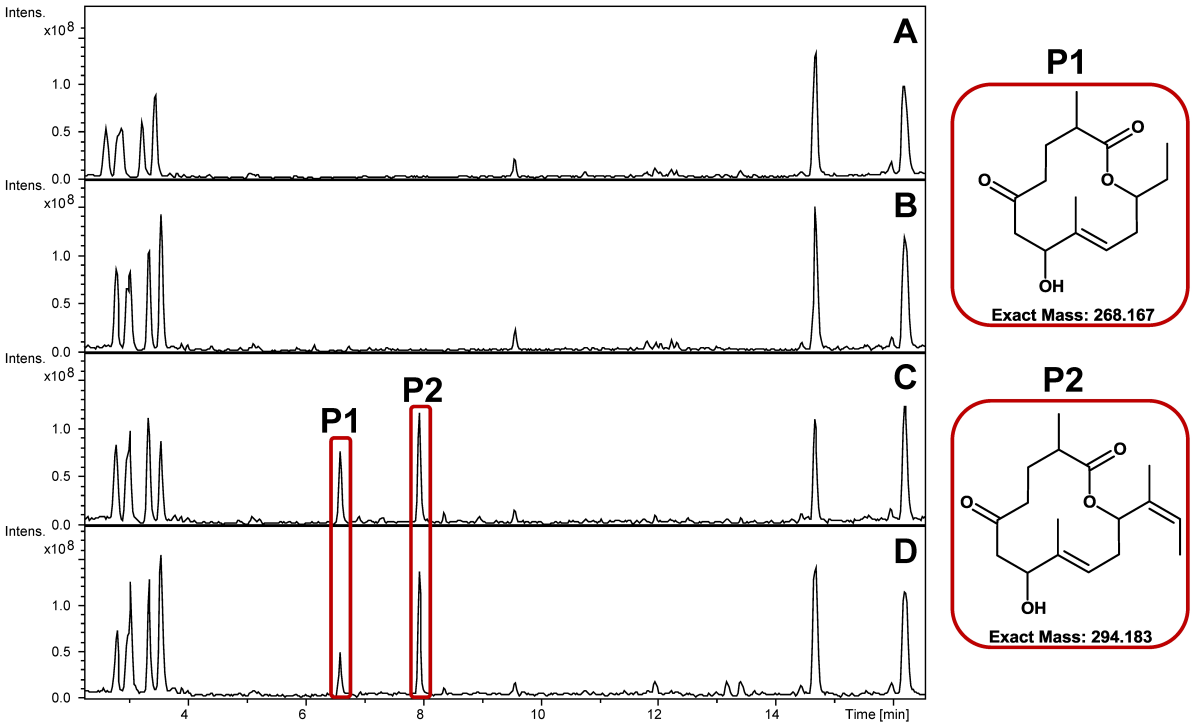


Figure 7.10: Total ion chromatograms of four cultivations of *M. xanthus* DK1622 mutants containing inducible heterologous gene cluster. A: mutant with *dif*-cluster, not induced. B: Mutant with different cluster, induced. C and D: mutants with *dif*-cluster, induced. Mass peaks P1 and P2 are only observed in extracts of the induced mutants with the *dif*-cluster. Their observed m/z and chemical structure is shown in the right panels.

The total-ion chromatogram A (fig. 7.10, top) shows an extract from non-induced DK1622 attB::DIF20, B showes attB::CPN13 (a different gene cluster used as control) induced with 1 mM Kvan. Chromatograms C and D show extracts of attB::DIF20, induced with 1 and 2 mM Kvan, respectively. The mass peaks marked in the red boxes were only observed in the extracts from induced attB::DIF20 cultures, and are therefore resulting from expression of the PKS encoded by the *dif*-genes. Mass peaks P1 with a m/z of $[M+H]^+$=269.2 and P2 with a m/z of $[M+H]^+$=295.2 do however not correspond to the calculated mass of the macrolactone that was expected to be produced from the *dif*-PKS genes. The disciformycin precursor released from the PKS according to the biosynthesis scheme shown in fig. 7.2 would have a m/z of $[M+H]^+$=293.2.

Indeed, also a mass peak for 293.2 Da was observed in the chromatograms at a retention time of 7.4 min, but with very low intensity. Because of its mass difference of 2 Da, compound P2 was considered to be the expected polyketide lactone, having a saturated C-C bond instead of a C-C double-bond at one position in the molecule. P1, being 24 Da smaller than the expected hypothetical product, was considered either a shortened shunt or degradation product of P2.

To obtain further insight into the compounds' identity, cultivation was upscaled to 400 ml CYE-medium and 800 ml CTT-medium, induced and extracted as before and the putative precursors P1 and P2 isolated by preparative LC-MS. The culture in CYE yielded 0.7 mg of each compound, whereas the culture in CTT yielded 0.3 mg of P1 and 1.0 mg of P2. Subsequently, 0.7 mg of each purified compounds P1 and P2 were analyzed by LC-MS and NMR.

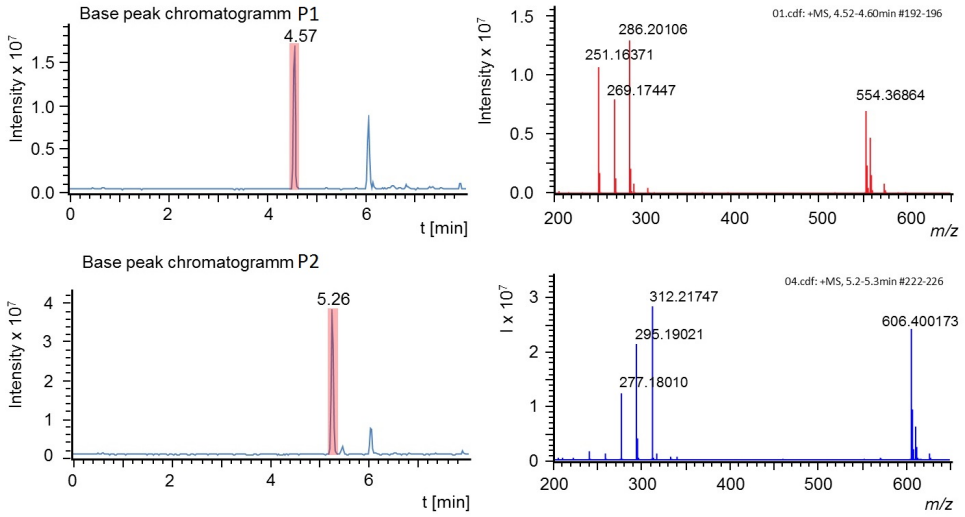


Figure 7.11: Base peak ion chromatogram of isolated P1 (top) and P2 (bottom), and corresponding mass spectra (right).

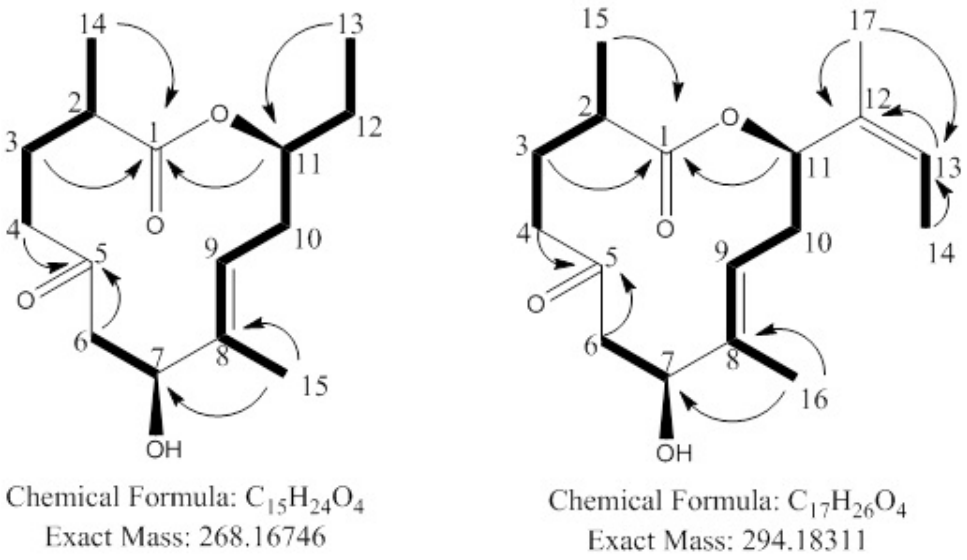


Figure 7.12: Selected COSY and HMBC correlations indicating the structure of the two congeners P1 (shown on the left) and P2 (shown on the right). Bonds shown in bold indicate correlations observed in ^{1}H,^{1}H-COSY spectra, arrows mark correlations from ^{1}H,^{13}C-HMBC spectra.

The structure of P2 was confirmed to consist of the macrocyclic disciformycin polyketide core, lacking the post-PKS modifications and posessing a fully saturated C2-C3 and C3-C4-

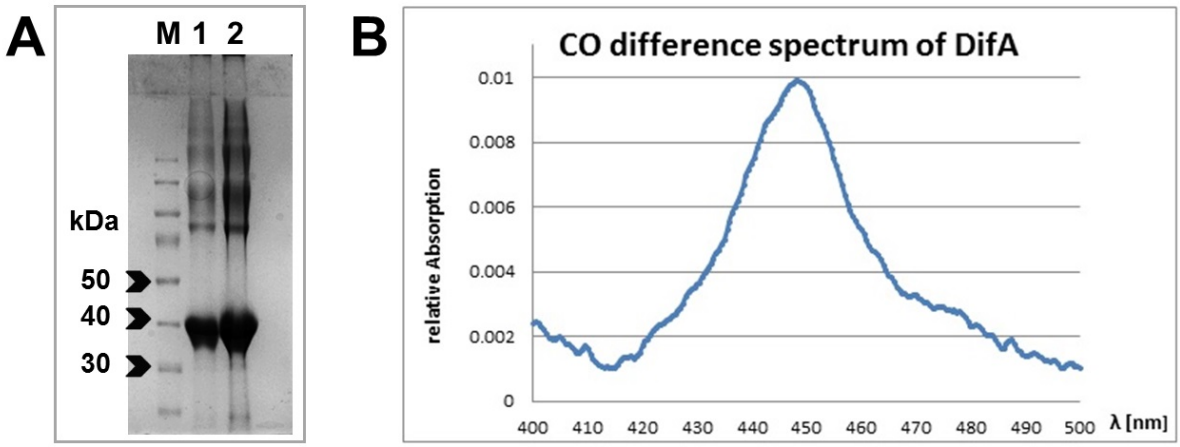


Figure 7.13: A: SDS-PAGE of purified recombinant DifA. M: prestained protein ladder. Lanes 1 and 2 show purified DifA at 40 kDa size. Protein impurities of higher molecular size are also visible, which are probably aggregates and uncleaved fusion protein. **B:** CO-difference spectrum of carbon monoxide-treated DifA. The peak at 450 nm indicates a functional cytochrome P450 enzyme.

bond, where the disciformycins A and B are unsaturated. The structure of P1 is identical to P2 except for its side chain, which is shortened by two carbon atoms. It is probably a shunt product, resulting from skipping one elongation cycle in the combined starter/extender module, encoded by *difG*. A mass corresponding to a disciformycin derivative with the shortened side chain observed in P1 was not detected in extracts of the original producer *P. fallax* AndGT8.

Table 7.1: Exact mass data aquired for isolated compound **P1** and comparison to calculated masses for the expected chemical formulae.

Ion species	$[M+H]^+$	$[M+NH_4]^+$	$[M-H_2O+H]^+$	$2[M+NH_4]^+$
Experimental *m/z*	269.174736	286.201285	251.164171	554.368744
Calculated *m/z*	269.174470	286.201060	251.163710	554.368640
Deviation [ppm]	0.988	0.79	1.84	0.188

Table 7.2: Exact mass data aquired for isolated compound **P2** and comparison to calculated masses for the expected chemical formulae.

Ion species	$[M+H]^+$	$[M+NH_4]^+$	$[M-H_2O+H]^+$	$2[M+NH_4]^+$
Experimental *m/z*	295.190210	312.217470	277.180100	606.400173
Calculated *m/z*	295.190386	312.216935	277.179821	606.400044
Deviation [ppm]	-0.596	1.71	1.01	0.213

Expression and Characterization of the Tailoring Cytochrome P450 Enzyme DifA

One post-PKS modification step to which a candidate gene could be assigned in the gene cluster is the hydroxylation of carbon atom C6 of the disciformycin macrolactone, which is a prerequisite for connecting the isovaleric acid via an ester bond to the polyketide. The putative tailoring

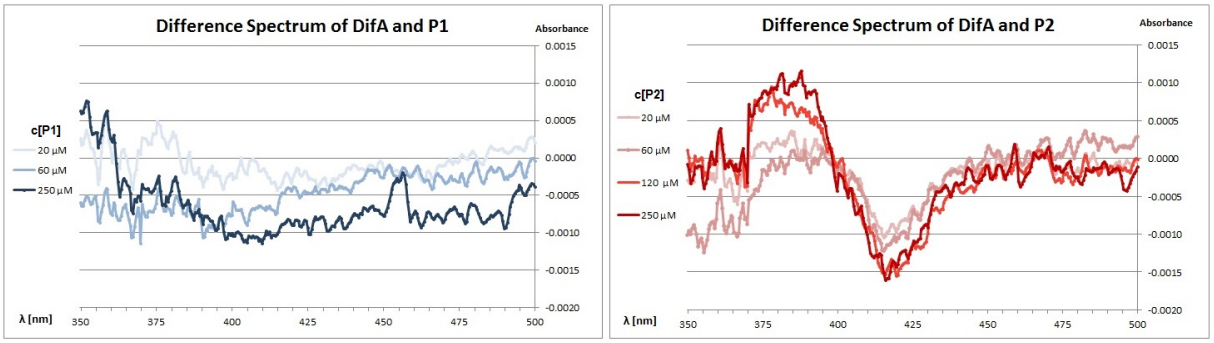


Figure 7.14: Substrate binding spectra of DifA with putative substrates P1 and P2 from heterologous expression of the *dif*-PKS genes in DK1622. P2 induces concentration-dependent type-I binding in DifA, indicating an enzyme-substrate interaction. Incubation of DifA with P1 does not produce this effect.

cytochrome P450 gene *difA* was expressed in *E. coli* BL21 as a maltose-binding-protein (MBP)-fusion gene and the protein purified via affinity chromatography. After washing and MBP-tag removal, the concentrated protein was investigated with difference spectroscopy. The difference spectrum after CO-treatment of the enzyme (fig. 7.13B) clearly shows a peak at 450 nm, indicating an intact CYP450 enzyme. To test, whether the isolated products P1 and P2 from the heterologous expression in *M. xanthus* are substrates of DifA, difference spectra of DifA were recorded with increasing concentration of the putative substrates. As shown in fig. 7.14, incubation with P2 induces a typical concentration-dependent absorption shift that corresponds to type-I binding. The increase of absorption at 390 nm and a decrease at 420 nm indicate a change of the spin state of the iron ion from low-spin to high-spin, induced by a conformational change in DifA after substrate binding. Incubation with P1 did not induce this shift, and is thus no substrate for DifA.

Consequently, hydroxylation of isolated P1 and P2 was then tested *in vitro*, using recombinant Adrenodoxin 4-108 (Adx)[97] and its native Adrenodoxin-Reductase (AdR) as electron delivery system. This system had been shown previously to be compatible with recombinant P450s from *Sorangium cellulosum* Soce56 and *Chondromyces crocatus* Cm c5. Here, hydroxylation of P2 was expected, while P1 was not expected to be hydroxylated. Substrate and enzymes were incubated as reported previously,[108] extracted and analyzed with HR-MS. Hydroxylation of the substrates would result in a new product peak with a mass increase of 16 Da and a slight decrease in retention time, compared to the substrate. MS-analysis showed no conversion of either substrate. The integrity of Adx/AdR proteins was confirmed in a control reaction. Therefore, Adx/AdR are probably no suitable electron donors for DifA, and the native electron transport proteins from *P. fallax* AndGT8 need to be reombinantly expressed and purified to further investigate the biosynthesis of disciformycin *in vivo*.

7.3 Conclusions

By cloning and heterologous expression of the disciformycin PKS in *M. xanthus* DK1622, the polyketide lactones P1 and P2 were produced. After isolation and elucidation of their chemical structures, P2 was found to resemble the disciformycin scaffold, while P1 has a shortened side chain and is probably the result of module skipping during PKS assembly. Surprisingly, the oxidation state of P1 and P2 differs from disciformycin at carbon C3, where it is fully reduced. The reason for this is currently not known. Incubation of the recombinant putative tailoring cytochrome P450 enzyme DifA with P2 *in vitro* led to P2 binding, indicating that P2 is a substrate for DifA. Incubation of P2 and DifA with the broad spectrum electron transfer system

Adx/AdR did not lead to formation of new products, probably because DifA is does not accept electrons from recombinant Adx. Incubation of P1 and P2 with crude lysate from *P. fallax* AndGT8 did not reveal any new peaks either. To achieve hydroxylation of P2 *in vitro*, heterologous expression and purification of the native electron donors for DifA may be necessary, or a compatible broad-spectrum donor other than Adx/AdR must be tested.

P1 and P2 showed no antibacterial activity *in vivo* and no *in vitro* inhibition of RNA Polymerase. The remaining modifications of the polyketide scaffold are thus considered essential for their bioactivity. Future experiments must aim at the identification of the remaining genes encoding the tailoring enzymes, i.e. a gylcosyl transferase, an acyl transferase and suitable redox partner proteins for DifA. The *in vivo* most active disciformycin derivative is disciformycin B, bearing the unusual E-$\Delta^{3,4}$ double bond, whose biosynthesic origin should therefore be elucidated as well. Although so far, the heterologous expression system established herein currently only yields an inactive disciformycin precursor, it is considered a valuable tool for future work, since the original producer can not be mutagenized currently and produces only very small amounts of disciformycin.

7.4 Materials and Methods

Cloning of DNA fragments for TAR

The overlapping DNA fragments of the *dif*-cluster were obtained by PCR from AndGT8 gDNA using the primer combinations listed below in table 7.3. Phusion polymerase and buffers were used according to manufacturers, with a final concentration of 10% DMSO in the reaction mixture. Isolated PCR products B+C and D+E were joined by overlap-PCR to yield products "BC" and "DE", respectively. The linear capture vector was obtained by PCR from the circular pCLY10 template, using primers with a 42 bp overhang that serve as homologous region for recombination with the corresponding sequence on the linear *dif*-cluster fragments. The Long-range PCR Polymerase and buffer system was used, with the PCR sequence shown in Table 7.4.

Table 7.3: Primers used for PCR of *dif*-cluster fragments for TAR

Product	Primers	Sequence 5'-3'
B	difBnew_for	GTTTTTACTTGACCCGCCGCA
	difBnew_rev	GCTTCTGCGGGGAGAGATTCT
C	difCnew_for	TTCTCGATGGGCTTCGACTCC
	difCnew_rev	ATCCAACATCGACATCGGCCT
D	difD_new2_for	AATCCTCTGGAGACGCAGGC
	difD_new2_rev	CGAAGTCGCCGTAGTCGTTGT
E	difEnew_for	GGATACGAACGCCCTCCTTGC
	difEnew_rev	GCGAGGAACCGCCAGAAGG
F	difFnew_for	TTCGAGCACCCCACGATTGAA
	difFnew_rev	TCCATGGGTGTGTTCTCGCTT
G	difGnew_for	CCCTTCCTGTTCGAAGCGAGA
	difGnew_rev	ACCACTTCGAGGTCAGCATCC
capture vector	TAR_pCLY10_difG_for	CCGCGCACGCGCACTGCGGGA- GGATGCTGACCTCGAAGTGGT- TTCGGTGAACTTCGGGTACGC
	TAR_pCLY10_difB_rev	TTCATGCTTATTTCGGGTTAA- TGCGGCGGGTCAAGTAAAAAC- TCGCCCTTTGACGTTGGAGTC

Table 7.4: Long-range PCR sequence for capture vector

step	duration	temperature	
1	3'	94C	
2	20"	96C	
3	30"	64C	
4	6'	68C	steps 2-4 were repeated 10 times
5	2'	94C	
6	3'	66C	
7	6'	98C	steps 5-7 were repeated 22 times

TAR protocol

For TAR assembly of the *dif*-cluster, 50 mL YPD medium were incubated with 1 mL overnight culture of yeast strain *S. cerevisiae* ATCC 4004247 and grown for 4h at 200 rpm and 30C to an OD_{600}=2.0. In the meantime, the transfomation mix (see table 7.5 was prepared. The sheared salmon sperm DNA serves as carrier DNA for transformation. The freeze-dried powder is first dissolved by boiling in sterile TE buffer for 5 min, then chilled on ice prior to adding to the mixture. The transformation mix was completely mixed and stored on ice. The linear fragments used for each transformation mix are shown below in table 7.6. The volume of of the linear DNA fragment mix was adjusted to 34 μL. If the volume exceeded 34 μL, the mix was dried *in vacuo* and dissolved in 34 μL. About 4h after inoculation, the yeast cells had grown to the required OD. The cells were pelleted at 4000 g at room temperature in a free-swinging rotor, washed 2x with sterile deionized water and resuspended in 1 ml water. Aliquots of 100 μL, corresponding to 10^8 yeast cells each, were transferred to a 1.5 mL tube, one for each transformation plus two controls. The aliquoted cells were spun at top speed for 30s in a tabletop centrifuge at room temperature, and the supernatant removed by pipetting. The transformation mix was added and

mixed with the cells by careful pipetting up and down. Each tube with cells and transformation mix was placed in a tabletop incubator at 42C for 90 min. After incubation, the tubes were centrifuged at top speed for 30 s and the supernatant removed. Cells were washed with 1 mL water, centrifuged for 30s and resuspended in 100 μL water. Each mix was spread completely on a separate YNB-Leu Agar plate. Yeast clones appeared within 2-4 days and were first screened by colony PCR prior to inoculation in liquid YNB-medium.

Table 7.5: Transformation mixture for TAR. Mix for one transformation of 10^8 yeast cells

PEG 4000 50%	240 μL
LiAc 1 M	36 μL
sheared salmon sperm DNA (2 mg/mL)	50 μL
linear DNA fragment mix	34 μL
DMSO	20 μL

Table 7.6: Mixture of linear DNA fragments for TAR

Mix	**Fragments**	**amount of DNA**
BG	BC	200 ng
	F and G	150 ng
	DE	200 ng
	pCLY10_BG	100 ng
K+	pCLY10 (circular)	100 ng
K-	pCLY10_BG (linear)	100 ng

Colony-PCR

S. cerevisiae and *E. coli* clones were screened with colony-PCR prior to inocultion in liquid. The primers used for colony-PCR are listed in table 7.7.

Table 7.7: Primers used for PCR to screen *S. cerevisiae* and *E. coli* clones

PCR product	**Primers**	**Sequence 5'-3'**
kB	TAR_kpCLY10for	TGTGGATTTTGATGTAATTGTTGGGA
	TAR_kB_rev	TTCCAGTCGTTACACGTCCGC
kG	TAR_kG_for	AGATAGCCCTCCAGCCGCTC
	TAR_kpCLY10rev	CAGTGAGCGAGGAAGCGGAAT
CB	difCnew_for	TTCTCGATGGGCTTCGACTCC
	difBnew_rev	GCTTCTGCGGGGAGAGATTCT
FE	difFnew_for	TTCGAGCACCCCACGATTGAA
	difEnew_rev	GCGAGGAACCGCCAGAAGG

Plasmid Recombination via Red-ET

For modification of plasmid pCLY_BG to yield the expression plasmid pMyxZeo_dif, linear-to-circular Red-ET recombination was applied, following standard protocols.[36] Briefly, pCLY_BG

was cloned into an engineered *E. coli* cloning strain designated "GB-dir", then grown to an OD_{600} of 0.2 and expression of the recombinase genes induced by addition of L-arabinose to a final concentration of 3 mg / L, then further cultivated to an OD_{600} of 0.5, washed and electroporated with the linear fragment, in this case the pMyxZeo plasmid backbone with suitable homology regions at each side. Clones were selected on LB-agar containing Ampicillin and Zeocin, and screened by colony PCR and restriction analysis as described above.

Expression and Characterization of recombinant DifA

Cytochrome P450 gene *difA* cloned in expression plasmid pETM44 and expressed as MBP-tagged protein according to protocols adapted for optimized heterologous expression of cytochromes: *E. coli* BL21 DE3/pETM44_difA was grown in 200 mL of TB medium supplemented with 50 mg/L Kan. As the culture reached OD600 = 2.0, 5-amino-levulinic acid was added to a final concentration of 1 mM, $FeCl_3$ was added to 0.5 mM, and the culture was incubated for 30 min at 25 C before addition of IPTG to 0.25 mM. The culture was then further incubated for 60 h (25 C, 200 rpm) until cell harvest. Cleared *E. coli* lysate was separated with the ÄKTA HPLC system using the MBP-Trap column. A one-step elution with maltose buffer was applied according to the manufacturers protocol. All buffers used contained 5 mM DTT to prevent oxidation of the protein. During the purification, absorption was monitored at 280 and 430 nm. Fractions containing the absorption peak at 430 nm were subjected were pooled, concentrated, and cleaved with TEV-protease. The cleaved MBP-tag was separated from DifA in a second Chromatography step. After buffer exchange, the protein was concentrated to 0.5 mL final volume. Protein concentration was determined 10 mg/mL using Bradford measurement. SDS-PAGE of the protein revealed a strong band at 40 kDa, with protein impurities at above 80 kDa. Recording of CO-spectra and difference spectra with test substrates P1 and P2 was performed according to standard protocols.

Table 7.8: Primers for cloning of *difA*

Primer name	**Sequence 5'-3'**
difA_for71	`CGAACATATGCCATGGGCATGCGTACGTTCCCGGCC`
difA_rev72	`CGTTGAATTCAGCTGGCCCGGAACG`

Cultivation and Extraction of *M. xanthus* DK1622 clones

M. xanthus DK1622 was cultivated in CTT or CYE media and transformed via electroporation according to standard protocols. Genetic verification of clones carring pMyxZeo_dif were performed on isolated genomic DNA. Verified clones were grown in liquid until the medium was turbid. Then, cultures were divided into two subscultures, and in one culture, expression of the *dif*-genes was induced by addition of potassium vanillate to a final concentration of 1 mM. Cultures were then further cultivated for 2-5 days. Cultures and cells were extracted together with ethyl acetate, dried *in vacuo* and redissolved in methanol.

Potassium vanillate stock solution was prepared by mixing solid vanillic acid in sterile H_2O. Then, 1 M KOH solution was slowly added to the suspension until vanillic acid was completely dissolved. The volume was then adjusted with H_2O to a concentration of 1 M.

Isolation and Characterization of compounds P1 and P2

Purification was carried out on a Waters Autopurifier (Eschborn, Germany) high pressure gradient system, equipped with 2545 binary gradient module, SFO system fluidics organizer, 2767 sample manager, and a 2998 photodiode array detector coupled to a 3100 single quadrupole mass spectrometer operated in positive ion mode. Source and voltage settings for the MS were as follows: mass range, m/z 200 400; scan duration, 1 s; points per Dalton, 4; capillary voltage, 3.5 kV; cone voltage, 30 V; extractor voltage 3 V; RF lens, 0.1 V; source temperature 120C, desolvation temperature, 250 C; desolvation gas flow, 400 L hr^{-1}; cone gas flow, 50 L hr^{-1}; ion counting threshold, 30. Preparative scale separation was carried out using 800 μL injections of crude extract dissolved in MeOH on a Phenomenex Gemini C18 axia packed preparative HPLC column (21.2 x 250 mm, 5 μm). Mobile phases were (A) water and (B) methanol. The separation was carried out at room temperature at a flow rate of 25 mL min^{-1}. Gradient conditions were as follows: 0 30.6 min, 5 95% B; 30.6 33.1 min, 95% B; 33.1 33.3 min, 95 5% B; 33.3 41 min, 5% B. Quality control measurements were performed on a Dionex Ultimate 3000 RSLC system using a Waters BEH C18 column (50 x 2.1 mm, 1.7 μm) by injection of five μl of each fraction. Separation was achieved by a linear gradient using (A) H2O + 0.1 % FA and (B) ACN + 0.1 % FA as eluents. The column was thermostatted to 45C and operated at a flow rate of 550 μL min^{-1}. The gradient was initiated by a 0.27 min isocratic step at 5 % B, followed by an increase to 95 %B in 18 min to end up with a 1.5 min flush step at 95 % B before re-equilibration. The LC system was coupled to a TriVersa NanoMate for flow splitting and high resolution MS data acquisition using a Thermo Scientific (Bremen, Germany) Orbitrap. Mass spectra were acquired in centroid mode ranging from 200 2000 m/z at a resolution setting of 30000. The flow was split to 500 nL min^{-1} before entering the ion source. NMR measurements were performed on a Bruker Ascend 700 spectrometer with a 5 mm TXI cryoprobe (^{1}H 700 MHz, ^{13}C 175 MHz).

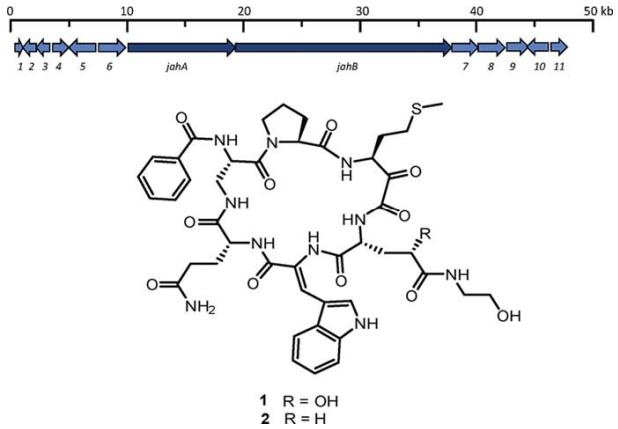
Supporting Information

Supporting Data, covering the the ^{1}H and COSY NMR spectra of P2, the observed NMR data of **P1** and **P2**, and *in vitro* activity tests performed with *Staphylococcus areus* RNA-polymerase can be found in the appendix on page 228.

Chapter 8

Jahnellamides, α-Keto-β-Methionine-Containing Peptides from the Terrestrial Myxobacterium *Jahnella* sp.: Structure and Biosynthesis

Alberto Plaza, Konrad Viehrig, Ronald Garcia, Rolf Müller*



This chapter was published as an article in *ORGANIC LETTERS* on October 15th, 2013.

*To whom correspondence should be adressed.

Author's efforts

The author's efforts regarding this publication comprise the identification and annotation of the jahnellamide biosynthesis gene cluster, the biosynthesis scheme and subsequent in-depth analysis of the sequence for corroboration of the predicted substrate specifities of the involved enzymes, as well as contributing the respective parts to the manuscript.

Contributions by others

Cultivation and isolation of the producer strain were carried out by Ronald Garcia. Compound isolation and structure elucidation were done by Alberto Plaza, who also carried out labeling experiments and contributed to the biosynthesis proposal. Alberto Plaza contributed the main part of this manuscript. The project was supervised by Prof. Rolf Müller, who also contributed to conceiving and writing of the publication.

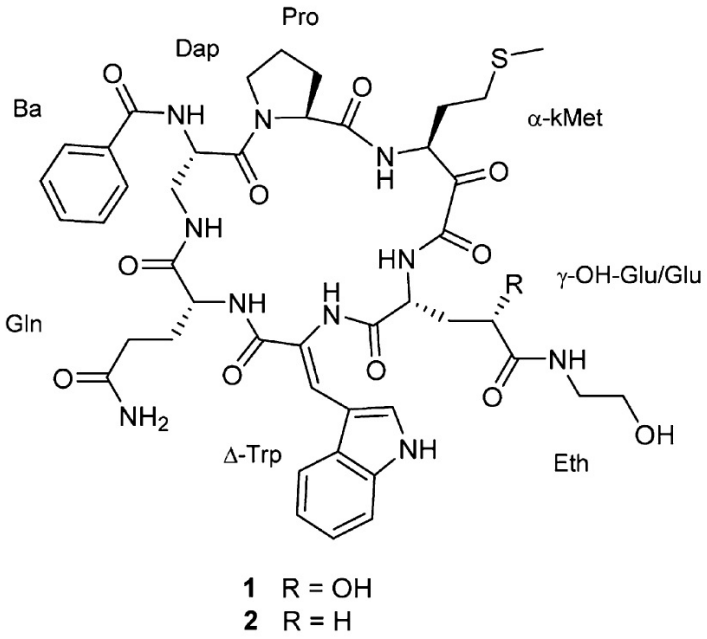


Figure 8.1: Structures of jahnellamide A **(1)** and B **(2)**.

Abstract

Two new cyclic peptides, termed jahnellamides A and B, were isolated from the myxobacterium *Jahnella* sp. Their structures were solved by NMR, ESIMS, and chemical derivatizations. Jahnellamides are a new class of α-ketoamide-containing peptides comprised of nonproteinogenic amino acids, including α-keto-β-methionine and 4-hydroxyglutamic acid. Moreover, *in silico* analysis of the genome sequence along with feeding experiments allowed us to identify and annotate a candidate nonribosomal peptide synthetase biosynthetic gene cluster containing a polyketide synthase module involved in the formation of the α-ketoamide moiety.

8.1 Introduction

Even though finding new scaffolds from natural products has become increasingly difficult, myxobacteria are still a rich source of biologically active secondary metabolites with distinctive structural features.[83] To increase chemical diversity and reduce rediscovery rates in our screening programs, we have been using NMR based technology to profile crude extracts prepared from new myxobacterial families[143] and underexplored species. Using this strategy, a methanol extract from *Jahnella* sp. (strain SBSr007) exhibiting strong antifungal activity was profiled by LC-SPE-NMR/MS. Analysis of this extract showed the presence of at least three different classes of peptides. The following structural studies led to the isolation of two new cyclic peptides containing an α-keto-β-methionine residue, termed jahnellamide A **(1)** and B **(2)** (Figure 8.1), together with the known peptides microsclerodermin D[144] and pedein A.[24]

8.2 Results and Discussion

The HR-ESI-MS spectrum of jahnellamide A **(1)** dissolved in MeOH displayed three ion peaks atm/z 947.3702 $[M + H]^+$, 965.3834 $[M + H_2O]^+$, and 979.3968 $[M + CH_3OH]^+$ supporting a molecular formula of $C_{44}H_{54}N_{10}O_{12}S$ (calcd for $C_{44}H_{55}N_{10}O_{12}S$, 947.3722) and requiring 23 degrees of unsaturation. The HSQC spectrum of 1 in CD_3OD exhibited signals characteristic

of a peptide containing aromatic and oxygenated residues, including five α-amino methines (δ_C 52.4, δ_H 4.58; δ_C 61.3, δ_H 4.33; δ_C 53.8, δ_H 4.83; δ_C 53.7, δ_H 4.86; δ_C 57.6, δ_H 4.31), one oxymethine (δ_C 69.5, δ_H 4.16), one oxymethylene (δ_C 61.6, δ_H 3.62), and one olefinic methine (δ_C 127.2, δ_H 8.13). Moreover, the ^{13}C NMR spectrum showed signals corresponding to nine carbonyls at δ 167.7-179.2 and five sp^2 quaternary carbons at δ 134.5 (C-2$_{Ba}$), 120.4 (C-2$_{\Delta\text{-Trp}}$), 110.7 (C-3'$_{\Delta\text{-Trp}}$), 129.5 (C-3a'$_{\Delta\text{-Trp}}$), and 137.5 (C-7a'$_{\Delta\text{-Trp}}$). A detailed analysis of the 2D NMR data (HSQC, HSQCTOCSY, HMBC, COSY) of **1** clearly established the presence of 2,3-diaminopropionic acid (Dap), proline, 4-hydroxyglutamic acid (γ-OH-Glu), and glutamine along with benzoic acid (Ba) and an ethanolamine (Eth) moiety (see Table S1). Key HMBC correlations from the olefinic singlet at δ 8.13 to the carbon resonances at δ 167.7 (C-1$_\Delta$-Trp), 130.1 (C-2'$_\Delta$-Trp), 120.4, and 129.5, and from the aromatic proton at δ 7.71 (H-2'$_\Delta$-Trp) to the carbon resonances at 110.7, 137.5, and 120.4, determined the presence of an $\alpha-\beta$-dehydrotryptophan (Δ-Trp) residue.
The structure of the remaining $C_6H_9NO_2S$ unit was assembled as follows. HSQC-TOCSY, COSY, and HMBC implied the presence of a modified methionine. Indeed, the aminomethine proton at δ 4.83 and the methylene proton at δ 1.81 showed long-range correlations to a quaternary carbon resonance at δ 99.4 that was diagnostic of a hemiketal functional group. An unassigned carbon resonance at δ 174.2 remained, and no HMBC correlation from the proton at δ 4.83 to the latter carbon was observed. Additionally, a ^{13}C NMR spectrum of **1** recorded in DMSO-d$_6$ clearly showed C-2$_{\alpha\text{-kMet}}$ resonating at δ 192.0 suggesting conversion from a hemiketal to a carbonyl (see Table S2).[145] Taken together, these data indicated the presence of an α-keto-β-methionine (α-kMet) residue. Further evidence supporting the presence of an α-ketoamide functionality was obtained from the ESI-MS spectrum, where an ion peak at m/z 979.3968 $[M + CH_3OH]^+$ was indicative of **1** forming a hemiketal with MeOH at C-2$_{\alpha\text{-kMet}}$.

A semiselective HMBC spectrum in CD_3OD was used to increase the resolution of the overlapped carbonyl region, and long-range correlations between α-protons and carbonyl carbons of adjacent amino acids allowed us to establish the following two partial sequences: γ-OH-Glu-Δ-Trp-Gln and Dap-Pro (see Figure S1). Connectivity between Gln and the Dap β-amine was apparent from HMBC correlations between the β-methylene protons of Dap (δ_H 4.34, 3.25) and the carbonyl resonance at δ 175.0 (C-1$_{Gln}$). An HMBC cross peak from the α-proton at δ 4.83 (H-3 $_{\alpha\text{-kMet}}$) to the carbonyl at δ 176.7 (C-1$_{Pro}$) attached the α-kMet residue to proline. In addition, an HMBC correlation from the equivalent methylene at δ 3.36 (H-1$_{Eth}$) to the carbonyl at δ 177.8 (C-5$_{\gamma\text{-OH-Glu}}$) linked the ethanolamide moiety to the δ-carboxylic acid of γ-OH-Glu, while the connectivity of the benzoic acid to the N-terminus of Dap was deduced from a long-range correlation from the proton at δ 4.58 (H-2$_{Dap}$) to the carbon resonance at δ 170.3 (C-1$_{Ba}$). Lastly, NMR data acquired in DMSO-d$_6$ allowed us to complete the planar structure of **1**. In particular, closure of the macrocyclic ring was deduced from an ROE correlation between the amide proton at δ 8.94 (NH$_{\gamma\text{-OH-Glu}}$) and the α-proton at δ 4.50 (H-3$_{alpha\text{-kMet}}$) along with an HMBC correlation from H-2$_{\gamma\text{-OH-Glu}}$ (δ 4.69) to the carbonyl resonance at δ 161.2. The Pro residue was assigned as trans on the basis of the differential value of ^{13}C chemical shifts of Cβ and Cγ ($\Delta\delta_{\beta\gamma}$) of around 3 ppm along with an observed ROE correlation between H-5a$_{Pro}$ and H-2$_{Dap}$ (see Table S2).[146] Comparison of the chemical shift of H-3$_{\Delta\text{-Trp}}$ (δ 7.85 in DMSO-d6) with that of keramamide F (δ 7.83)[147] and with those of the geometrical isomers of methyl-α-acetamido-6-methylindole-3-acrylate [δ 7.69 (Z) and 7.00 (E)][148] indicated the geometry of the α,β double bond to be Z. ROE correlations from H-2'$_{\Delta\text{-Trp}}$ (δ 7.48) to NH$_{\Delta\text{-Trp}}$ (δ 8.41) and H-2$_{\gamma\text{-OH-Glu}}$ (δ 4.70) further supported this result.

The absolute configurations of L-Dap, L-Pro, D-Gln, and 4S-hydroxy-D-Glu residues were assigned by acid hydrolysis and derivatization of **1** with L-FDLA (1-fluoro-2,4-dinitrophenyl-5-L-leucinamide) and D-FDLA followed by MS-detected chromatographic comparison of derivatives of amino acid standards.[149,150] The L configuration of α-kMet was established by treating **1**

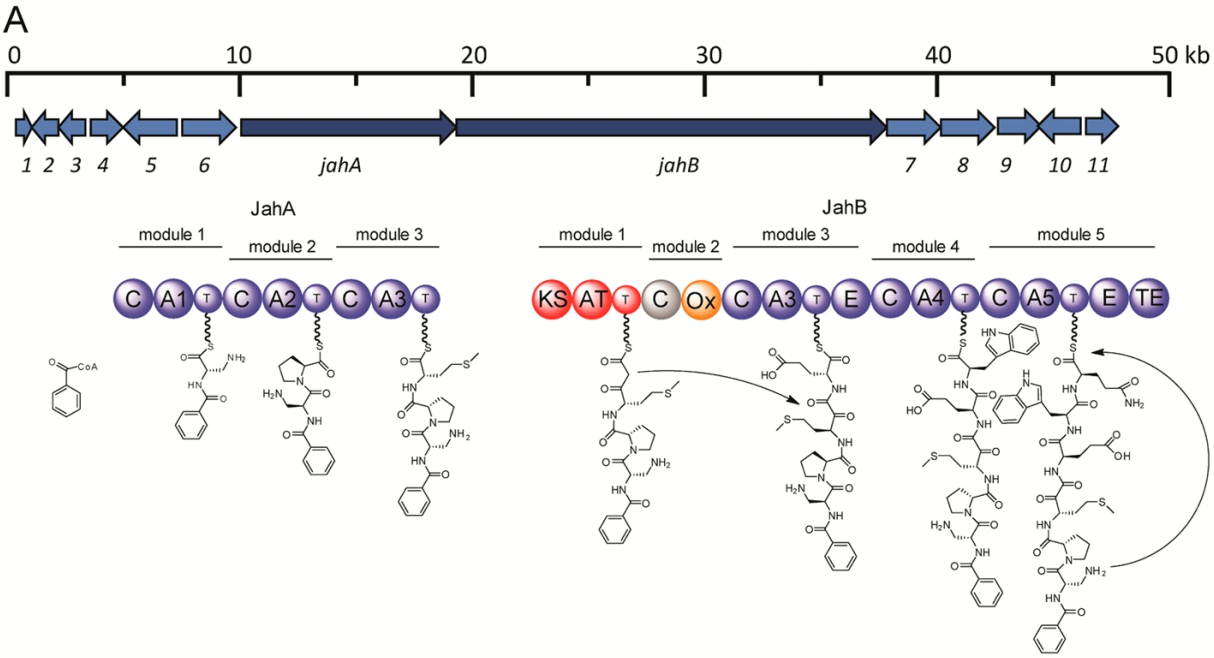
with H_2O_2/ NaOH followed by acid hydrolysis and application of the advanced Marfey method.

Jahnellamide B **(2)** showed an ion peak at m/z 931.3762 $[M+H]^+$ ($C_{44}H_{54}N_{10}O_{11}S$, calcd for $C_{44}H_{55}N_{10}O_{11}S$, 931.3772), which is 16 mass units lower than that of **1**. The 2D NMR data for **2** closely resembled those of **1** with the exception that resonances belonging to γ-OH-Glu were replaced by resonances belonging to a glutamic acid residue. The *Z* geometry of the α,β-double bond in Δ-Trp was deduced from a strong ROE correlation between H-2'$_{\Delta\text{-Trp}}$ and H-2'$_{\text{Glu}}$ (see Table S3). Marfeys analysis was not performed due to the insufficient quantity of **2**. However, we assumed identical configurations for **1** and **2** at comparable chiral centers because their structures and NMR data are very similar.

With the structures of the jahnellamides in our hands, a tentative biosynthetic scheme was proposed by using a combination of feeding experiments with labeled precursors and *in silico* analyis of the draft genome sequence of *Jahnella* sp. ^{2}H NMR analysis of **1** obtained from an L-[2,3,3-$^2H_3$]serine feeding experiment showed resonances for H-2$_{\text{Eth}}$ (δ 3.62) and H-3a$_{\text{Dap}}$ (δ 4.34) indicating that both Eth and Dap originate from serine (see Figure S2). It is worth mentioning that no deuterium was observed at δ 4.58 (H-2$_{\text{Dap}}$) which is consistent with the possible role of dehydroalanine as an intermediate during the conversion from serine to Dap.[151,152] The biosynthetic origin of the α-kMet residue was studied as follows. Intact incorporation of L-[U-^{13}C, ^{15}N]methionine (7%) was measured by HRESI-MS (see Figure S3a) while addition of [2-^{13}C] acetate led to enhancement of the ^{13}C NMR signal of C-1$_{\alpha\text{-kMet}}$ (δ 174.2) (see Figure S4). However, feeding experiments with [1-^{13}C] acetate did not result in the labeling of C-1$_{\alpha\text{-kMet}}$. Taken together, these data suggest that the α-kMet residue is formed by condensation of methionine to the carbon at position 2 of a malonate moiety, and therefore α-kMet has a hybrid NRPS/ PKS origin. Additionally, HRESI-MS spectra showed intact incorporation of L-[ring-$^{13}C_6$]phenylalanine (25%), and [$^2H_5$]benzoic acid (29%), while incorporation of [$^2H_7$]cinnamic acid was only observed for the aromatic ring (33%) (see Figure S3b-d).

On the basis of the chemical structure of the jahnellamides along with the results from the feeding experiments, we postulated **1** and **2** to be assembled by a modular mixed NRPS/PKS pathway. Analysis of the *Jahnella* sp. draft genome sequence with anti-SMASH[31] allowed us to identify a gene cluster containing one NRPS gene and one hybrid NRPS/PKS gene with modular organization. The predicted substrate specificity for each module was in good agreement with the structures of **1** and **2**, and with our feeding experiment results. Further analysis of the adenylation domains substrate specificity was performed by using NRPSpredictor2.[30] The candidate *jah* gene cluster (50 kb in size) comprises 13 putative open reading frames (ORF) including a trimodular NRPS *(jahA)* and a pentamodular NRPS/PKS *(jahB)* (Figure 8.2). However, the exact boundaries of the gene cluster have not been determined yet. In fact, gene inactivation in the producing strain has not been successful so far, in particular due to the fact that *Jahnella* sp. does not grow in suspension.

Our analysis indicates that assembly starts with the condensation of benzoyl-CoA to Dap which originates from serine (vide supra). This is in agreement with the fact that the JahA starter module has a C-A-T domain organization characteristic of N-acylated NRPS products.[153] Assembly then proceeds with sequential incorporation of Pro and Met. The growing peptide chain is transferred onto JahB, where malonyl-CoA is incorporated by the PKS part. Our incorporation experiments suggested the α-kMet residue is formed by condensation of the C-2 carbon of an acetate moiety to Met. Consequently, we theorized that the additional condensation and oxidation domains of the PKS module catalyze an oxidation at the methylene carbon of the malonyl residue with subsequent elimination of the malonyl carbon at position 1. Interestingly, this catalytic domain architecture is also present in the hybrid PKSNRPS encoded by MXAN_3779 of *Myxococcus xanthus*, which is responsible for the biosynthesis of the α-ketoserine-containing



B

Predicted adenylation domain specificity

	Active site residues	Amino acids	
		Predicted	In jahnellamides
JahA			
A1	DIWEITAD	Ser	Dap
A2	DVQFIAHV	Pro	Pro
A3	DVYYSAGI	Met	α-kMet
JahB			
A3	DASVIGCI	Gln/Glu	δ-OH-Glu/Glu
A4	DVCHIGGV	Trp/Tyr	Δ-Trp
A5	DAAHQGSV	Gln/Glu	Gln

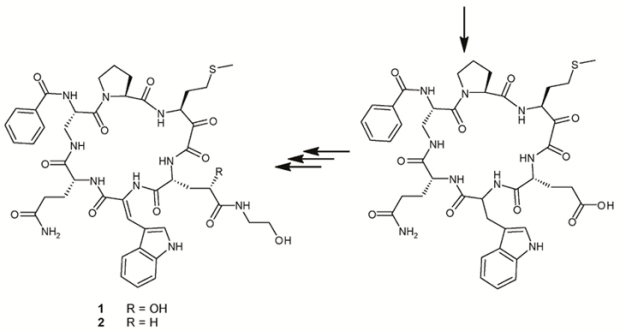


Figure 8.2: (A) Putative gene cluster of jahnellamides in *Jahnella* sp. and proposed biosynthetic pathway of **1** and **2**. **(B)** Prediction of A domain substrate specificity.

peptide named myxoprincomide.[86] Thus, we predict that formation of the α-keto functionality in the jahnellamides occurs in similar fashion to that proposed for myxoprincomide. Next, assembly continues with the consecutive incorporation of Glu, Trp, and Gln. The presence of E domains downstream of the C-A-T motifs in modules 3 and 5 of JahB correlates nicely with the presence of D-Glu and D-Gln. Finally, the C-terminal thioesterase (TE) domain of JahB would catalyze the peptide bond between the C-terminus of Gln and the β-amine of Dap and offload of the prejahnellamides as cyclic peptides. Subsequently, jahnellamides are formed by condensation of ethanolamine to Glu and desaturation of the α- and β-carbon of the tryptophane residue. Jahnellamide B is considered to be a precursor of **1**, which bears a hydroxyl group at position 4 of the glutamate residue. The roles of the remaining ORFs have yet to be established; nonetheless a function such as tailoring enzymes, facilitation of building block biosynthesis, or compound export is expected (see Table S4).

Our feeding experiments clearly determined that L-serine is the precursor for L-Dap and ethanolamine, which are both expected to be provided by enzymes encoded within the cluster. Upstream of *jahA* we located *orf2* and *orf3*. Interestingly, their sequences are similar to those of *sbnB* and *sbnA*, respectively, which are genes encoding L-Dap synthetases.[152] Therefore, conversion of L-serine to L-Dap in JahA may be catalyzed by the synergetic action of *orf2* and *orf3*. The *orf8* and *orf9* are putative desaturases and therefore are proposed to be responsible for the formation of Δ-Trp, as demonstrated for Chromobacterium violaceum.[154] The decarboxylation of serine to produce ethanolamine is most likely performed by the PLP-dependent decarboxylase *orf10*.[155] The hydroxylation of the Glu residue in **1** is expected to result from the action of a P450 oxidase or nonheme oxidoreductase enzyme encoded elsewhere in the *Jahnella* sp. genome. Unfortunately, we were not able to confirm the proposed functions of the genes in the *jah* gene cluster because mutagenesis is not established for this strain (vide supra). Additionally, precursor-directed biosynthesis was applied to generate analogues of **1**. A 4-fluorobenzoic acid derivative and a 2-fluoroethan-1-amine analogue of **1** (**3** and **4**, respectively; see Figures S4 and S5) were detected by LC-HRMS in a methanol extract of *Jahnella* sp. However their low yields prevented their isolation and confirmation of their structure by NMR.

Jahnellamide A showed neither antifungal activity nor cytotoxicity toward HCT-116 cells, and the antifungal activity of the extract was traced to the microsclerodermins. However the intriguing structural features of the jahnellamides warrant further evaluation of its biological function.

Conclusions

In summary, jahnellamides are a new class of cyclic peptides that contain unusual amino acids, including 4-hydroxy-glutamic acid, α-β-dehydrotryptophan, 2,3-diaminopropionic acid, and the previously undescribed α-keto-β-methionine. **1** and **2** are characterized by a six residue lactam ring formed by cyclization between the C-terminus and the β-amine of Dap. Exocyclic to the lactam, a benzoic acid is linked to the α-amine of Dap and an ethanolamine moiety is connected to the δ-carboxylic acid of glutamic acid. Although a few examples of α-ketoamide-containing cyclic peptides have been reported from Lithistid demosponges,[156] the jahnellamides are the first myxobacteria-derived examples of this class of natural products. In fact, the jahnellamides show a slight structural similarity to the strong thrombin inhibitors, cyclotheonamides. These peptides are characterized by the presence of an α-Keto-Arg moiety and a pentaresidue macrocycle formed by condensation of the C-terminus to the β-amine of Dap.[145,157,158] However, the jahnellamides are unprecedented in their hexapeptide macrocycle and also in the presence of the α-keto-β-methionine and ethanolamine residues. Besides, our feeding experiments and *in silico* analysis demonstrated a hybrid NRPS/PKS biosynthetic origin for **1** and **2**. These data estab-

lish an excellent stage for further investigations of jahnellamide biosynthesis, in particular in the formation of the α-keto-amide functionality. In closing, identification of two different structural classes of peptides in strain SBSr007 provides another example of the notable biosynthetic capabilities of myxobacteria.

Acknowledgement

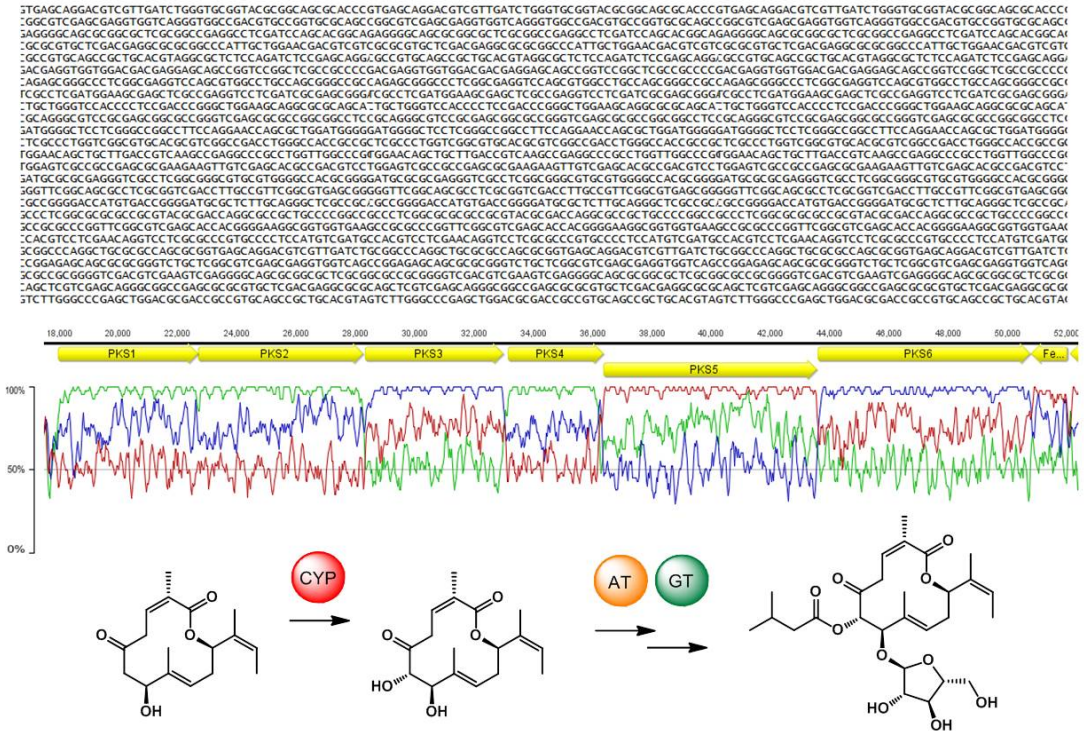
Work in the laboratory of R.M. was supported by the Bundesministerium für Bildung und Forschung (Grant No. 0315798) and the Deutsche Forschungsgemeinschaft.

Supporting Information

Supporting Information Available. Experimental details, ^{1}H and ^{13}C NMR assignments, Figures S1-S5, and 1D and 2D NMR spectra for **1** and **2**. This material is attached in the appendix on page 236.
The authors declare no competing financial interest.

Chapter 9

Discussion



9.1 Summary of Major Results from this Thesis

9.1.1 Genome mining in *Chondromyces crocatus* Cm c5

In the myxobacterium *Chondromyces crocatus* Cm c5, the following secondary metabolites were correlated to their respective biosynthetic gene cluster by targeted mutagenesis: The crocapeptins, the crocadepsins and the crocagins. For each compound class, Cm c5 mutants were generated, in which key genes for the biosynthesis of these products were disrupted. As a result, production of the target compound was abolished. By targeted inactivation of genes involved in later steps of compound biosynthesis, novel derivatives or precursors were also obtained, which support the biosynthesis hypothesis and confirm the involvement of these genes in the respective biosynthesis schemes.

Crocapeptins A and B were isolated prior to this work from large-scale fermentations of *Chondromyces crocatus* strains Cm c5 and Cm c2, and congeners of this compound class were previously described from cyanobacteria. Their biosynthesis, especially the mechanism of the formation of the amino-hydroxy-piperidone (Ahp) ring, was not yet unraveled and therefore investigated during this thesis. As a result, the crocapeptin precursor pre-crocapeptin was isolated from a Cm c5 mutant and successfully converted to bioactive crocapeptin *in vitro* using recombinant enzymes. The established bioassay enabled in depth studies on the formation of the Ahp ring, which represents a crucial structural feature of this compound class as it is responsible for the bioactivity of crocapeptins and cyanopeptolins. It was also demonstrated that crocapeptin production can be strongly increased by targeted promotor insertion in Cm c5. The obtained results were published in 2013.[108]
The crocadepsins were discovered by directed gene inactivation studies targeting numerous putative NRPS and PKS genes in Cm c5. Comparative secondary metabolite production analysis between the generated mutants and wild type cultures was applied to find novel compounds that can immediatly be related to their biosynthetic gene locus, allowing subsequent biosynthesis studies. In contrast to previously discovered natural products from Cm c5, i.e. the chondramides, crocacins, ajudazols and thuggacins, which all show either cytotoxic, antifungal or anti-parasitic bioactivity, no such activity was detected for the crocadepsins. Some features of their biosynthesis as well as their chemical structure hint towards a function as siderophore, but experiments carried out by Dr. Frank Surup did not confirm this.
The crocagins were isolated from large-scale fermentations of *Chondromyces crocatus* Cm c5 and initial insights on biosynthetic precursors were available from previous labeling experiments. By sequence-based genome analysis and subsequent targeted gene inactivations, it could be shown that crocagins originate from a ribosomally made peptide and thus constitute an unprecedented myxobacterial RiPP class. In contrast to the numerous NRPS/PKS pathways found in Cm c5, the sequences of the encoded gene products of the *cgn* cluster show very low homology to other proteins from public databases, making this biosynthetic pathway exceptionally interesting, although no bioactivity has been detected for the crocagins so far. Details and yet unresolved issues of crocagin biosynthesis are discussed in section 9.2.1.

9.1.2 *In silico* genome mining in *Jahnella* sp. SBSr007 and *Pyxidococcus fallax* AndGT8:

In addition to the established multiproducer *C. crocatus* Cm c5, the genomes of two further myxobacterial producers were investigated by *in silico* sequence analysis in the course of this work. From both producer strains, novel bioactive compounds had been isolated in bioactivity-guided screenings: The jahnellamides, produced by *Jahnella* sp. SBSr007, and the disciformycins, produced by *Pyxidococcus fallax* AndGT8.
The genome sequence analysis of *Jahnella* sp. SBSr007 was analyzed *in silico* with datamin-

ing tools including antiSMASH. Through a combination of labeling experiments and sequence analysis, the gene cluster was found and a biosynthesis scheme was proposed. The findings were published along with the Jahnellamide structure and activity data in 2013.[159]
The disciformycins were discovered via an activity guided screening approach and prioritized because of their promising anti-MRSA activity. Analysis of the *Pyxidococcus fallax* AndGT8 genome sequence was exerted as for the jahnellamides, and a gene cluster that encodes the responsible PKS genes was found. Closer sequence analysis of the individual enzymatic domains of the disciformycin PKS allowed to predict their substrate specificity and the stereospecificity introduced during the reduction reactions. The predictions were found to match the experimentally obtained disciformycin structure in almost all cases. The biosynthesis proposal and sequence analysis were published along with the disciformycin structure and bioactivity data in 2014.[160]

9.1.3 Cloning and Heterologous Expression of Selected Gene Clusters

To overcome the limitations of mutagenesis-based genome-mining approaches (as discussed in section 9.2.1), heterologous production of selected compounds was aspired. This involves cloning and assembly of the target genes in engineered *E. coli* strains, followed by modifications that allow integration and expression into a host strain. As a host for the pathways cloned herein, *M. xanthus* DK1622 was chosen. To allow for controlled expression of these genes in this host, a vanillic acid inducible promotor was chosen, which avoids leaky transcription during the cloning steps in *E. coli* and enables DK1622 clones to grow to high cell density without additional metabolic burden or toxic effects resulting from heterologous pathway expression.

The 30 kb crocapeptin gene cluster was assembled from two cosmids and modified via Red/ET recombination[36] to yield an expression construct that can be integrated into the *M. xanthus* DK1622 chromosome. Mutants harboring the cluster were obtained and cultivated under different conditions, but in-depth MS-analysis of culture extracts revealed that neither crocapeptins nor other putative pathway products were produced. Possible reasons for this are discussed in section 9.2.4.
The 32 kb disciformycin biosynthesis gene cluster was assembled via TAR cloning[161] from PCR products and subsequently modified again for integration and expression in DK1622. Induced cultures yielded substantial amounts of a polyketide scaffold which was identified as biosynthetic precursor of disciformycin. The expression plasmid was included in the patent application, which is pending for the disciformycins, as it discloses a route for overproduction of the macrolide core of disciformycins. Here, the remaining challenge is the identification of the genes for the tailoring enzymes that form the bioactive final product, and elucidate the formation of the double-bond at C3, which is not present in the heterologously produced precursor, but apparently important for the compounds' bioactivity. This is further discussed in section 9.2.4.

Next to compound overproduction and biosynthesis studies, gene rearrangements and targeted modifications of gene clusters also have important consequences for patent protection and intellectual property rights. Although isolated genes may currently be patented under European law, such claims are not possible in the United States.[162] Functional rearrangements of biosynthesis gene clusters usually add sufficient technological novelty to make the resulting sequences patent-eligible, especially when the efforts result in a previously unknown natural product or novel derivatives.

9.2 Discussion of the Applied Genome Mining Approaches

In this section, the limitations of the genome mining strategies applied in the course of this work are discussed on the basis of the obtained results. In *C. crocatus* Cm c5, site-directed mutagenesis was previoulsy established, and genome sequence information was used to inactivate putative biosynthesis genes, followed by phenotypic characterization of the mutants, i.e. analysis of their secondary metabolite production profile.
For *Jahnella* sp. SBSr007 and *Pyxidococcus fallax* AndGT8, a mutagensis protocol was not available, therefore genome mining was pursued via *in silico* sequence analysis and, in case of the disciformycins, heterologous gene expression.

9.2.1 Limitations of the Targeted Gene Inactivation Approach

In Cm c5, three additional secondary metabolite gene clusters were correlated to their respective products and intriguing aspects of their biosynthesis were elucidated by targeted inactivation of the biosynthesis genes in the host. There are however a number of potential secondary metabolite gene clusters in this genome, whose inactivation did not result in the shortfall of a peak in the respective mutant's extract. Apparently, these compounds are either not produced in the wild type under laboratory conditions or they escaped detection, e.g. due to stability issues, biased extraction/analytical procedures or production levels that are below the detection limit. This result shows a drawback of the gene inactivation-based approach: it does not lead to the identification of the corresponding natural product if the targeted pathway is inactive in the wild type or not expressed under the tested laboratory conditions. If the targeted pathway is encoded by a single transcriptional unit, insertion of an active promotor via single crossing-over represents a promising strategy, given an available mutagenesis strategy and a known promotor sequence, as demonstrated in this work for the crocapeptin. By insertion of an artificial promotor sequence in front of a transcriptional unit, also binding motifs for negative regulatory elements upstream of the transcriptional start site are replaced, which contributes to overexpression as well. However, natural product gene clusters are often organized in complex gene arrangements and only few are organized in simple operons, which lowers the applicability of this approach.

9.2.2 The crocagins: a "lucky case"-example of genome mining?

Another important limitation of the targeted inactivation approach is shown by the crocagin gene cluster, which has no similarities to any known gene cluster and would have evaded a purely sequence-based mining strategy. The cluster was identified after crocagin isolation from the wild type and subsequent feeding experiments which revealed the compounds' building blocks. Since no suitable NRPS genes predicted to assemble the crocagin tripeptide Ile-Phe-Trp were found in the Cm c5 genome, the genome sequence was then screened for genes that directly encode this peptide sequence, resulting in the candidate gene *cgnA*. Since this putative precursor peptide gene is too short for targeted insertional mutagenesis, the adjacent ORFs *cgnB* and *cgnI* were inactivated and the resulting mutants were indeed crocagin non-producers and producers of a precursor, respectively. The discovery of the crogagins is therefore not a straightforward result of targeted genome mining, but rather a chance find that was nevertheless pursued consequently and resulted in the annotation of a previously unknown myxobacterial RiPP biosynthesis gene cluster. A silent crocagin gene cluster with no detectable product would not have been discovered. A subsequent BLAST search of the CgnA, CgnB and CgnI sequences against other myxobacterial genomes from the *Sorangiineae* and *Cystobacteriineae* suborders yielded no matches at all. One can therefore argue that the crocagins present here a "lucky case" as they are produced in the wild type in sufficient amounts to allow isolation, chemical characterization and structure elucidation. Many of the ORFs in the crocagin gene cluster and also in other clusters of Cm c5

can not be inactivated by the established method because of their short sequence length. Gene inactivation by insertional mutagenesis relies on a sufficiently large stretch of identical DNA sequence, that leads to integration of the suicide plasmid into the reading frame of the particular gene, thereby disrupting it. By experience, the homologous fragment should be at least 500 base pairs in size to yield any mutants and ideally is 1 kb in size. This methods works excellent for large NRPS or PKS genes, which usually exceed 3 kb of sequence length. All the NRPS/PKS mutants generated in Cm c5 in this thesis were non-producers of the respective product, and false-positive mutants, which have the plasmid integrated elsewere in the genome were not an issue. For shorter genes, the number of false-positive clones is increased, and many genetically verified mutants of smaller genes did not exhibit the expected phenotype, as explained below.

9.2.3 Inactivation of genes encoding putative tailoring enzymes via single crossing-over

Inactivation of the shorter tailoring genes by insertional mutagenesis in Cm c5 was less efficient compared to inactivation of larger NRPS or PKS genes. In genetically verified Cm c5 mutants of such genes, the chemical phenotype of the mutants was often difficult to detect. One example is the tailoring cytochrome P450 in crocapeptin biosynthesis, which was inactivated in *cpnF*- mutants. Theoretically, the resulting mutants should exclusively produce pre-crocapeptins, but in their extracts, crocapeptins and precrocapeptins were detected at about equal levels. The reason for this may be complementation by functional homologues or other oxygenase enzymes, although such tailoring enzymes usually exhibit strict substrate specifity. Another reason may by an incomplete abolition of CpnF function in the respective mutants. Since a sufficiently large stretch of homologous sequence is necessary for targeted insertion of the suicide plasmid into the desired locus, the disrupted reading frame of the target gene may remain large enough to allow translation of a shortened peptide with residual function. In the case of the *cpnF* locus, a 0.9 kb fragment was used to inactivate the 1.2 kb gene, which probably resulted in a *cpnF* gene with more than two-thrids of intact reading frame in the mutant. Inactivation of the larger tailoring gene *cgnI* in the crocagin cluster led to exclusive production of a non-carbamylated crocagin precursor, as theoretically expected. There, the homologous sequence fragment was 0.9 kb, and the gene is 1.8 kb in length.

In the crocapeptin cluster, inactivation of the putative helper protein gene *cpnE* was not achieved by insertional mutagenesis, because of its small size and because of its location. It is cotranscribed with *cpnF*, which is encoded downstream. Inactivation of *cpnE* is unlikely to result in a visible phenotype anyway: the results of the gene inactivation would be overshadowed by polar effects from abolished expression of the downstream encoded *cpnF*. Generally, in such a situation, the insertion will lead to a phenotype that shows not only the effect of the inactivation of the target gene, but also the effects of the impaired expression of the gene(s) encoded further downstream, in the worst case resulting in false assignment of functions to genes. To circumvent these problems, the homologous fragment used for inactivation may be additionally amended with an active promotor that restores transcription of the downstream encoded genes, as demonstrated for Cm c5 in studies of ajudazol biosynthesis.[20]

Mutagenesis by double crossing-over, in which the integrated plasmid sequence is excised from the genome in a second step would avoid these effects. For Cm c5, double crossing-over is not established, and the second crossing-over step would have to be applied to multiple mutants all the genes of interest, which is highly time-consuming, especially with the slow-growing *Sorangiineae*. Complex pathway modifications are best performed in a fast-growing heterologous host, while the gene cluster is present on a relatively small genetic element, i.e. a plasmid or cosmid. This has a lot of advantages, especially with respect to sequence analysis. Recent advances in Red/ET cloning allow for fast and comparably efficient cloning in *E. coli*, using negative se-

lection markers such as the toxic ccdB-peptide.[38] The modified biosynthetic pathways can then be transformed into the heterologous host and the resulting mutants screened for the expected effects.

9.2.4 Limitations of the Heterologous Expression Approach

As mentioned above, *M. xanthus* DK1622 mutants with the crocapeptin biosynthesis genes showed no crocapeptin production, although this host had been shown previously to be suitable for functional expression of gene clusters from the *Sorangiineae* suborder.[163] The identical plasmid backbone was utilized for the expression of the discformycin PKS genes in the same host, resulting in substantial production of the corresponding polyketide product. Possible reasons for the non-production of crocapeptins in the respective DK1622 mutants are either host-related, i.e. incomplete transcription and/or translation of the NRPS gene or post-translational modifications. On the other hand, the *cpnDEF* operon may not be sufficient for crocapeptin production, and additional genes are required in the heterologous host, although their inactivation in Cm c5 did not abolish crocapeptin production. This can be tested by integrating further genes into the expression construct or via other methods into a verified DK1622 attB::CPN mutant. Future research also aims at modification of the expression system to allow either introduction into other hosts, such as members of the *Sorangiineae* suborder to which Cm c5 belongs to. Transcription of the NRPS genes could be probed by isolation of RNA from the induced host followed by reverse transcription using primers that hybridize with the 3'-terminus of the transcript. If a signal is obtained, incomplete transcription can be ruled out. Although the crocapeptin biosynthesis NRPS gene *cpnD* represents a very large prokaryotic gene with over 25 kb, the heterologous host DK1622 should be able to express it, since it harbours and expresses the even larger gene MXAN_3779, encoding the myxoprincomide NRPS.[86]

Heterologous expression can be used to investigate the biosynthesis of a compound, if the original producer can not be genetically manipulated and also offers a route to overproduction and pathway manipulation, once heterologous expression of the compound or a precursor molecule is achieved. In the last few years, heterologous expression of "orphan" gene clusters with yet unknown products has been applied for genome mining purposes. Target clusters for this strategy are either gene clusters, that are aparrently silent in their cultivated hosts or gene clusters identified by sequence screening of uncultured organisms or environmental DNA sequence data. In a number of cases, new structures were obtained from these approaches and even novel bioactive scaffolds were discovered.[40] This strategy strongly depends on clustering of the biosynthesis genes in a defined locus in the host genome, because the gene clusters are usually cloned as a whole entity without functional characterization of the individual ORFs. Even though precise predictions regarding the substrate specificity and stereospecificity of NRPS/PKS assembly lines can be made from DNA sequence analysis, prediction of the chemical properties of the resulting natural products is currently not possible. Predictions of their bioactivity on the other hand can be made by the identification of putative resistance genes in the cluster. These may however also be encoded elsewhere in the genome, which is also the case for genes involved in biosynthesis steps. In some cases, i.e. pyoverdin[164] or xenocoumacin,[67] the product results from a pro-drug mechanism and the final bioactive compound differs from the precursor that is assembled on the NRPS/PKS assembly line. In these cases, compound maturation is coupled to periplasmic export and compatibility of the gene products responsible for tailoring and export must be compatible with the host membrane system. Heterologous expression of a predicted pathway would probably produce inactive precursors rather than the actual products.

In case of the disciformycins presented in this thesis, the genes for the tailoring enzymes responsible for glycosylation and acylation of the PKS scaffold are not known and probably not encoded in the vicinity of the PKS gene cluster. Although the PKS assembly line produces a

disciformycin precursor, this precursor is not bioactive. Here, the active compound was identified in a conventional activity-based "grind & find" approach combined with strong fractionation of the crude extracts to detect compounds that are produced at very low levels or overshadowed by other, already known compounds. It can therefore be concluded that, although recent advances in high-throughput genome sequencing, automated sequence annotation and prediction as well as convenient direct cloning techniques such as TAR cloning and Red/ET recombineering allow for mining and investigating natural product pathways beyond the limitations of cultivation-based methods, classical screenings can not be neglected and must be improved further. Especially the cultivation of previously uncultured bacterial species and development of methods that activate expression of silent gene clusters in their original host under laboratory conditions remain important challenges.

9.2.5 Heterologous Expression in *M. xanthus* DK1622

M. xanthus DK1622 was chosen as a host strain for the expression of the crocapeptin and disciformycin gene clusters mainly because it is a myxobacterial strain and therefore most probably able to functionally express gene clusters from related myxobacteria, originating i.e. from the *Cystobacteriineae* and *Sorangiineae* suborders. It is further a prolific producer of a number of secondary metabolites itself, which ensures that necessary genes for precursor supply and post-translational activation of biosynthesic enzymes, e. g. ppant-transferases, are present and functional. A number of myxobacterial gene clusters have been expressed previously in *M. xanthus* DK1622, resulting in heterologous production of myxothiazol,[165] tubulysins,[166] myxochromide S[167] and epothilone.[168] *M. xanthus* DK1622 is also able to express non-myxobacterial PKS genes, as shown by heterologous oxytetracyclin production.[169] Integration of heterologous gene clusters into the DK1622 genome is possible via the the mx9 phage-attachment site, as performed in the course of this work. Other methods of integration were not tested, and to do this, the expression plasmid must be modified accordingly. An additional mx8 phage-attachment site is described in the DK1622 genome, which could be used by cloning the corresponding integrase into the expression construct of interest. Alternative expression hosts for the crocapeptin and disciformycin gene clusters have not yet been tested. Myxobacterial pathways have been expressed in non-myxobacterial hosts previously, i.e. Epothilon in *E. coli*[49] and Myxochromide in *Pseudomonas putida*.[47]

9.3 The Chemical Diversity Produced by *Chondromyces crocatus* Cm c5

9.3.1 The predicted secondary metabolite potential of Cm c5

The genome sequence of Cm c5 has been completed and is currently awaiting publication.[21] A sequence analysis using antiSMASH revealed a total of 36 putative natural product biosynthesis gene clusters, of which 19 are predicted to be NRPS/PKS genes clusters, while the remaining clusters appear to encode genes for the biosynthesis of other compound classes, such as terpenoids, lantibiotics and siderophores. Of the NRPS and PKS gene clusters, six are already related to a product and published: The ajudazol,[133] crocacin,[106] chondrochloren,[107] thuggacin,[85] chondramid[22] and crocapeptin[108] gene clusters. Excluding the yet unpublished crocadepsin and anabaenopeptin biosynthesis gene clusters, 11 further NRPS/PKS gene clusters remain to be correlated to a product in this producer. A special case is the crocagin biosynthetic gene cluster, as pointed out in section 9.2.2. This cluster appears to encode a yet unprecedented RiPP class and was not detected by predictive algorithms. This finding further demonstrates that the genomic potential of this producer and related myxobacterial producers may exceed currently predicted pathways. The immense natural product potential of Cm c5 is illustrated in fig. 9.1,

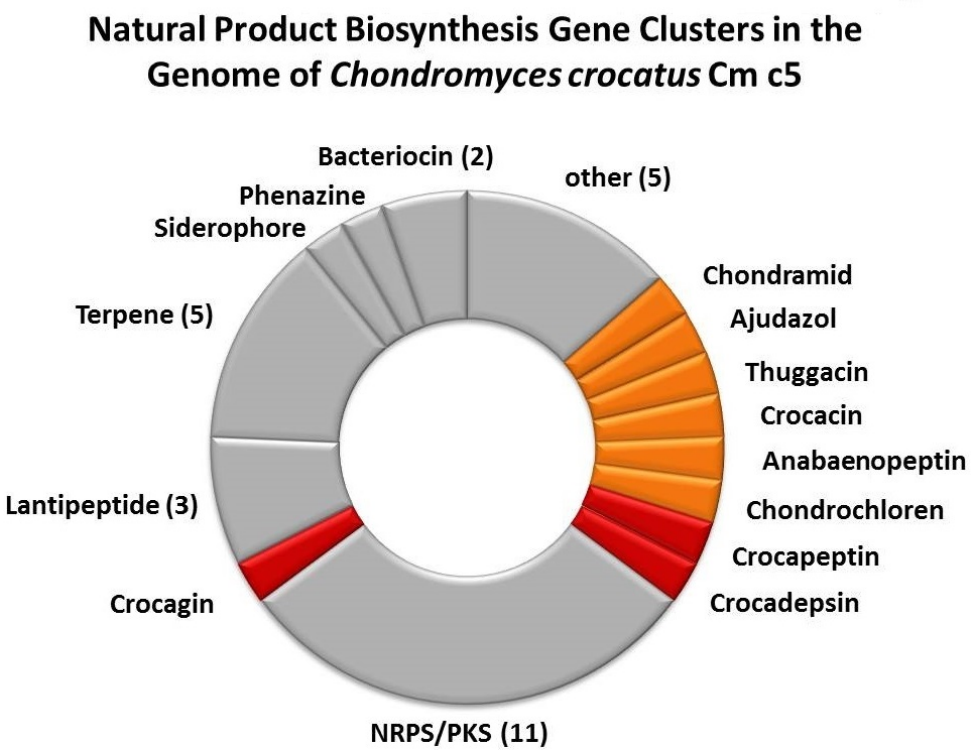


Figure 9.1: Predicted and confirmed secondary metabolite gene clusters in the Cm c5 genome, with predicted pathways shown in grey. Pathways that were correlated to a product in the course of this thesis are shown in red, other confirmed pathways are shown in orange. The ajudazol,[133] crocacin,[106] chondrochloren,[107] thuggacin,[85] chondramid[22] and crocapeptin[108] gene clusters are published, while the crocadepsin, crocagin and anabaenopeptin gene clusters are not yet published.

showing predicted and known secondary metabolite pathways in the genome.

Since a number of compounds with high structural similarity to marine compounds were isolated from related *Chondromyces* and *Sorangium* strains, further genome mining approaches in Cm c5 are expected to yield congeners of such compounds, but the chance to obtain entirely novel products is still considered quite high. As mentioned previously, the gene disruption approach does in some cases not yield clear phenotypes in terms of metabolite production, and promotor insertions as performed with a myxoprincomide producer[86] can be applied only in a few cases, where the organization of the ORFs appears promising. Given the abovementioned limitations of mutagenesis and heterologous expression, alternative approaches to tap the genetic potential of microorganisms may have to be applied to Cm c5, as outlined in the following section.

9.3.2 Non-mutagenic approaches to tap the natural product spectrum of Cm c5

The mutagenesis-based strategy has not yet been fully applied to all predicted natural product gene clusters in Cm c5, but given the abovementioned limitations of these approaches, alternative ways to extract yet unknown compounds should be applied to reveal compounds that are produced at low levels or only under certain conditions. The chance-find case of the crocagins indicates, that there may be products whose biosynthesis gene clusters are not yet described and are therefore not detected by the screening algorithms. One way to activate "silent" biosynthesis gene clusters is application of stress to the cultures, for example by adding chemical stress factors. Generally, varying cultivation conditions for one bacterial strain followed by standardized extraction and mass spectrometry analysis (MS) is an established compound mining strategy.[170]

An extensive collection of strategies involving addition of chemical stress factors is given in a review by Müller and Wink.[171] Interesting findings have also been obtained by co-cultivation experiments of the myxobacterium *Corallococcus coralloides* and the actinomycete *Streptomyces coelicolor*, which led to significantly altered production profiles in the bacteria,[172] but did not result in the discovery of novel compounds.

Other approaches make use of the advances in high resolution mass spectrometry (HRMS), employing large datasets to identify bacterial metabolites via statistical methods rather than manual investigations of chromatograms. For *Sorangium cellulosum* So ce38, this has been applied recently[173] in a data-mining approach, combining MS and MS/MS data from bacterial culture extracts and medium blank extracts. The MS data were processed by statistical methods, focussing on signals that match certain criteria, i.e. only acquired from bacterial cultures but never present in medium blank extracts. The list of relevant signals was then further grouped according to their fragmentation patterns and predicted sum formula. This approach is also able to identify derivatives of a coumpound family in the extracts, via identical MS^2 ion signals. In this study, two new metabolites from So ce38 were discovered.
A similar study was carried out with a larger sample set, comprising extracts from 98 different *M. xanthus* strains.[174] Here, relevant signals were filtered out via statistical examination, using ion mass, retention time and signal abundance as criteria. Using *M. xanthus* DK1622 extracts as a reference, principle component analysis of the extracted "relevant" signals then revealed strains that show a different product spectrum and even potentially new product families. Besides unexpected findings regarding production profiles of a global *M. xanthus* strain collection, the analysis resulted in the discovery of a new compound, cittilin. The major challenge of the statistical approach is to filter out the relevant mass signal and exclude as many background signals as possible. This is especially challenging for myxobacterial cultures, which usually require complex media that already contain a high amount of compounds with secondary metabolite-like characteristics in terms of solubility, molecular size and ionization behaviour. Such an approach has not yet been applied to Cm c5 and might reveal new compounds.

MS-based metabolite-mining experiments may also be combined with new technical possibilities such as MALDI-imaging,[175] allowing for visualization of compound distribution over a surface with bacterial growth. These methods reveal the spatial distribution of metabolites, and thereby give insight into chemical signaling between microorganisms,[176] biofilm formation[177] and other aspecs of microbial ecology, that may also reveal novel drug targets for treatment of infections with persistent, biofilm-forming bacteria.

Another possibility to harness yet unknown compounds is the investigation of extracts from Cm c5 fruiting bodies. From a biological point of view it seems plausible that the fruiting bodies and especially the spores are protected from microbial exoenzymes by natural products. Currently, compound isolation focuses almost exclusively on vegetative liquid cultures, because that way, the most biomass is obtained. However, since myxobacteria are among the very few prokaryotes that are able to form multicellular structures, this ability should be exploited much more with a focus on compound mining. The challenge will then be be to find and activate the corresponding pathways that may be inactive in vegetative cells, because Cm c5 fruiting bodies are probably not an option for isolation and purification of the compounds in quantities that allow structure elucidation or bioactivity testing. In the course of this work, it was demonstrated that retrobiosynthetic analysis allows for straightforward assignment of biosynthesis gene clusters from genome data, once basic features of the resulting natural products are known, especially when aided with isotope-labeling studies.

9.3.3 The Crocapeptins: Myxobacterial Representatives of a Widespread Cyanobacterial Compound Family

The crocapeptins were found to belong to a compound class previously known to be isolated from cyanobacteria, the cyanopeptolins. Earlier genome analysis of the producing *Mycrocystis* and *Anabaena* species had confirmed their non-ribosomal origin, but the exact formation of the amino-hydroxy-piperidone (Ahp) heterocycle, which is the critical feature for the compounds' bioactivity, had not been elucidated, mainly because cyanobacteria are difficult to cultivate and mutagenize. Since genetic tools were previously established for Cm c5, crocapeptin biosynthesis was investigated by targeted inactivation of the putative tailoring cytochrome P450 gene *cpnF*. This led to the appearance of crocapeptin precursor molecules in extracts of *cpnF* mutant cultures, labeled pre-crocapeptins. By converting these precursors to crocapeptin using purified heterologous CpnF, Ahp formation was confirmed. Additionally, the molecular rearrangement following the hydroxylation was monitored by NMR measurements, and subsequent NMR-data analysis revealed the reaction mechanism. There is however the possibility, that Ahp formation follows a different path in cyanobacteria. In Cm c5, it was shown that Ahp is formed from 5-OH proline, which then undergoes ring opening to Ahp via a reactive glutamic acid semialdehyde intermediate. 5-OH proline is unstable and not incorporated into crocapeptin as NRPS building block. Instead, the NRPS produces pre-crocapeptin, containing L-proline at this position. After assembly and release of pre-crocapeptin, the proline residue is then oxidized at carbon atom C5 by CpnF, yielding a 5-OH proline residue that can spontaneously or with aid of CpnE rearrange to Ahp. CpnE is a chaperone-like helper protein that assists in the rearrangement, which was indirectly demostrated *in vitro*. While conversion without CpnE resulted mostly in formation of 5-OH proline, which then slowly rearranges to Ahp over the course of several hours, the conversion reaction with CpnE in the reaction mix led to immediate formation of Ahp. As discussed in the crocapeptin publication, a possible function of CpnE is to to ensure rapid formation of the Ahp heterocycle, and avoid exposure of the transient glutamic acid semialdehyde intermediate, which can easily crosslink with a nearby amino or amido group. The published sequences of cyanopeptolin gene clusters[94] from *Microcystis*, *Anabaena* and *Planktothrix* species were investigated and neither of them contains a cytochrome P450 gene or a gene with sequence similarities to *cpnE*. Incorporation of proline into crocapeptins and pre-crocapeptins was verified by feeding labeled proline and by NMR. Additionally, the predicted substrate specifity of A-domain A4 of the crocapeptin NRPS gene *cpnD* indicates proline as substrate for this domain. On the other hand, the substrate of the respective cyanobacterial A-domains responsible for substrate incorporation at the position of Ahp-ring is not known and their extracted binding pocket motif does not predict incorporation of proline. Efforts from total synthesis have already shown that Ahp can also be generated from glutamic acid semialdehyde, which is an intermediate stage in the rearrangement from 5-OH proline to Ahp. Because of its reactivity, glutamic acid semialdehyde is unlikely a substrate for the A-domain. Incorporation of glutamate or glutamine and subsequent reduction to the semialdehyde are an option to obtain Ahp. However, A-domains with binding pockets specific for glutamine or glutamate are described and their binding pocket motif does not match the A-domains either. Tailoring enzymes for such rearrangement steps are sometimes also found to be encoded distant from the actual biosynthetic gene cluster, therefore it is still possible that the biosynthesis of cyanopeptolines and crocapeptins is identical, but their genes are arranged differently.

Compounds with high structural similarities to crocapeptin B are patented as inhibitors for kallikrein,[178] a regulatory serine protease that transmits inflammation signals. The compounds may therefore be used to treat chronic inflammations and related syndroms without the side effects of cortison. In total, there are over a hundred congeners of cyanopeptolins found, and they are all expected to be serine protease inhibitors, but with different target proteases. While the Ahp-heterocycle is crucial for the enzyme inhibition itself, the surrounding amino acid positions define the target specificity. Some congeners of crocapeptin exhibited promising activity against

proteases that are involved in virus infections. They do not inhibit viral proteases but human proteases on the cell surface which are needed for virus entry into the host cell. Crocapeptin B did inhibit Hepatitis C replication in a cell-based assay and ichthyopeptin A worked as influenza treatment in a mouse model, when applied as aerosol.[92] Some congeners have been generated by chemical total synthesis, which allows the generation of freely designed derivatives with tailored target specificity. By rational drug design, highly specific inhibitors of the proteases involved in influenza and HCV, and also other serine proteases that are involved in pathways related to cancer or inflammation may be obtained.

9.3.4 Several Compounds Produced by the *Sorangiineae* Suborder Have Marine Congeners

Next to unique compound classes such as crocagins, thuggacins and ajudazol, Cm c5 also produces compounds that belong to substance classes previously known only from cyanobacteria. This has been demonstrated for the crocapeptins, which can be seen as cyanopeptolin congeners. There are also (yet unpublished) indications, that Cm c5 produces anabaenopeptins [Stephan Hüttel, personal communications], another compound class commonly found in cyanobacteria.[87] Additionally, Cm c5 produces the highly cytotoxic chondramides, which belong to the jaspamides, previously described and isolated from a *Jaspis* sponge. Finding congeners of marine natural products in myxobacteria is however not that rarely observed. Other member of the *Sorangiineae* suborder, *Sorangium cellulosum* So ce38 and *Chondromyces pediculatus* Cm p3 produce pedeins[24] and microsclerodermins,[23] previously isolated from marine *Microscleroderma* sponges. The apicularens[179] are V-ATPase inhibors, isolated from a *Chondromyces* sp. and are structurally similar to salicylihalamides,[180] isolated from a marine *Haliclona* sponge. Whether these pathways are highly conserved or result from horizontal gene transfer is not clear, but the latter is commonly observed in microorgansims. Marine sponges harbour a vast prokaryotic flora that lives whithin the sponge tissue, with members of a diverse selection of prokaryotic orders, and there are indications based on 16S-rRNA sequences, that also marine myxobacteria are present in there.[181] It is therefore regarded possible that the i.e. the jaspamides and salicylihalamides are in fact produced by a marine prokaryote. Producers such as Cm c5 should be contemplated as potential hosts for heterologous expressions of such sponge-derived pathways, because sponge symbionts are usually difficult to cultivate, and Cm c5 has the biosynthetic ability to produce these scaffolds. The situation is similar for cyanobacterial pathways, which are however rather unsuitable for expression in a myxobacterial host, because of a generally lower CG content that results in a different codon preference.

9.4 TAR-Cloning and Cluster Rearrangement for Heterologous Expression Purposes

The applicability of PCR-based TAR cloning has been shown for the first time for a myxobacterial gene cluster in the course of this work. Sequencing of the resulting expression plasmid pMyxZeo_dif revealed that no mutations were acquired during this process and that the individual PKS genes were assembled properly. This finding shows that such biosynthesis gene clusters can be successfully assembled from PCR products via TAR, despite the repetitive nature of type-I NRPS and PKS genes, and the usually high GC content of myxobacterial genes, which often leads to formation of secondary DNA structures, which may inhibit the assembly.

In contrast to cluster assembly from cosmids or direct cloning from genomic DNA, the PCR based approach offers high flexibility in terms of gene cluster rearrangement, because it is assembled from multiple linear fragments. Thus, the gene cluster may be rearranged and optionally

mixed with other DNA fragments. This is important, when transcriptional regulation is altered at multiple sites, for example to obtain controlled expression of the genes through inducible promotor systems. In many cases, promotor regions must be adapted, if the genes are to be expressed in a host, that is not closely related to the cluster's source organism. Promotor sequences and ribosome-binding sites can also be introduced by other methods, but this usually requires multiple cloning steps, each bearing the risk of acquiring unwanted mutations. In the TAR assembly, it is all done in the assembly step. An important prerequisite for gene cluster rearrangements is a careful sequence annotation, as the translational start of each reading frame must be correctly annotated. Gene cluster rearrangements that aim at manipulation of the resulting biosynthetic product can be exerted during the assembly step as well, e.g. excluding genes for tailoring enzymes or introducing additional biosynthesis genes from related pathways. This assembly strategy opens up possibilities beyond genome mining and compound discovery, and enables the convenient construction of non-natural assembly lines to generate chemical diversity.

9.5 Conclusions

In this thesis, genome mining approaches based on sequence analysis, targeted mutagenesis and heterologous expression of secondary metabolite gene clusters were applied to three myxobacterial producer strains. As a result, the genetic basis for the production of three compounds, the crocapeptins, crocadepsins and crocagins was elucidated and verified by inactivation of pathway genes in the producer *C. crocatus* Cm c5. In a fourth case, where the producer organism *P. fallax* AndGT8 was not genetically amenable, cloning and heterologous expression of the core biosynthetic genes, which were identified by *in silico* sequence analysis could be achieved. Successful production of the corresponding compound confirmed the predictions made from sequence analysis. Cloning and heterologous expression of a secondary metabolite pathway was applied successfully in one of two cases and remains to be completed for a third pathway. This approach to harness prokaryotic secondary metabolites becomes increasingly important, as it is expected to enable acess to the full genetic potential of "orphan" gene clusters and unculturable or otherwise unaccessible organisms,[182] especially in light of recent advances in molecular cloning, DNA-sequencing and DNA-synthesis. Despite their sometimes ambigous output, bioinformatic tools such as NRPSpredictor or antiSMASH have become indispensable for genome analysis, since they allow much more rational strategies and help reduce the empiric workload drastically.

However, the compounds presented in this thesis that had the most promising pharmaceutical potential, i.e. antibacterial and antiviral activity, were discovered by the "grind & find" approach from culturable organisms. Although the mutagenesis and heterologous expression efforts that were applied contributed significantly to the understanding of their biosynthesis, their importance regarding compound overproduction and generation of novel derivatives remains in the range of a 5-fold overproduction for crocapeptin A by promotor insertion and the 10-fold overproduction of a biologically inactive precursor for disciformycin. Conventional approaches to obtain overproduction of a certain natural product, such as optimization of cultivation conditions and media, or random mutagenesis and clone selection produce impressive yields in industrial applications, but these are executed at a much larger scale in terms of manpower and time. To further imoprove the yields of targeted mutagenesis or heterologous production approaches, rational metabolic engineering of the producer organism may be necessary, as nowadays already performed for microorganisms in industrial production of simple compounds such as citric acid or amino acids. There are also strategies to generate high-yield expression platforms for secondary metabolite pathways from *Actinomycetes*, for example a genome-minimized *Streptomyces avermitilis*.[183] In this strain, most of the secondary metabolite pathways were deleted with the intention to generate a host for *Actinomycetes*-derived gene clusters that can direct all metabolic

capacity into the heterologous gene cluster. Expression strains harboring a heterologous gene cluster integrated in the genome can also be optimized further by the conventional methods that are applied to generate industrial-scale production microorganisms, such as random mutagenesis.

The long-term perspective of heterologous expression and pathway engineering, that goes beyond genome-mining is pointed out in comments and reviews in scientific literature: These techniques are expected to help create a biotechnology platform that is able produce all kinds of chemical molecules[184,185] that are currently produced by petrol based synthetic chemistry towards engineered and individually tailored microorganisms[186,187] - a transition that is already in progess.

The currently most obvious challenge for natural product discovery is provided by ongoing genome sequencing and automated annotation efforts, showing the large amount of predicted natural product gene clusters that remain to be correlated to the respective products. With respect to pharmaceutical applications, finding the target of these products is a further challenge. It is probable that the amount of yet unknown compounds also corresponds to a large number of yet unidentified natural product targets remaining to be discovered. With the rise of multiresistant pathogenic bacteria, research that investigates these pathogens and their resistance mechanisms, next to clinical issues such as biofilm formation, quorum-sensing, toxin production and persistence, should result in new bacterial targets, that again must be implemented into assays that are suitable for high-throughput screenings of compound or extract libraries. This applies of course to targets in human cells as well, as more and more insights in the development of chronic diseases, cancer and virus infections are obtained. Here, new drug targets will be acquired that may be addressed with small molecule drugs as well.

Bibliography

[1] Susanne Grabley and Ralf Thiericke. *Drug discovery from nature*, chapter 1, pages 3–10. Springer Science & Business Media, 1999. Cited on pages 10 and 11.

[2] Rachael A Maplestone, Martin J Stone, and Dudley H Williams. The evolutionary role of secondary metabolites - a review. *Gene*, 115(1):151–157, 1992. Cited on page 10.

[3] Christopher T Walsh and Timothy A Wencewicz. Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics*, 67(1):7–22, 2014. Cited on page 11.

[4] Jesse W-H Li and John C Vederas. Drug discovery and natural products: end of an era or an endless frontier? *Science*, 325(5937):161–165, 2009. Cited on page 12.

[5] Helge B Bode and Rolf Müller. The impact of bacterial genomics on natural product research. *Angewandte Chemie International Edition*, 44(42):6828–6846, 2005. Cited on page 12.

[6] Christopher N Boddy. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *Journal of industrial microbiology & biotechnology*, 41(2):443–450, 2014. Cited on pages 12 and 15.

[7] Martin Dworkin. Introduction to the Myxobacteria. *Prokaryotic Development*, pages 221–242, 2000. Cited on page 12.

[8] Hans Reichenbach and G Höfle. *Drug discovery from nature*, chapter 9, pages 149–179. Springer Verlag: Berlin, Germany, 1999. Cited on page 12.

[9] Susanne Schneiker, Olena Perlova, Olaf Kaiser, Klaus Gerth, Aysel Alici, Matthias O Altmeyer, Daniela Bartels, Thomas Bekel, Stefan Beyer, Edna Bode, et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nature biotechnology*, 25(11):1281–1289, 2007. Cited on page 12.

[10] Yao Xiao, Xueming Wei, Richard Ebright, and Daniel Wall. Antibiotic production by myxobacteria plays a role in predation. *Journal of bacteriology*, 193(18):4626–4633, 2011. Cited on page 12.

[11] Katrin Ravenschlag, Kerstin Sahm, Jakob Pernthaler, and Rudolf Amann. High bacterial diversity in permanently cold marine sediments. *Applied and environmental microbiology*, 65(9):3982–3989, 1999. Cited on page 12.

[12] Klaus Gerth and Rolf Müller. Moderately thermophilic Myxobacteria: novel potential for the production of natural products isolation and characterization. *Environmental microbiology*, 7(6):874–880, 2005. Cited on page 12.

[13] Till F Schäberle, Friederike Lohr, Alexander Schmitz, and Gabriele M König. Antibiotics from myxobacteria. *Natural product reports*, 31(7):953–972, 2014. Cited on page 12.

[14] Kira J Weissman and Rolf Müller. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Natural product reports*, 27(9):1276–1295, 2010. Cited on pages 12 and 31.

[15] Silke C Wenzel and Rolf Müller. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Natural product reports*, 26(11):1385–1407, 2009. Cited on page 12.

[16] B. Kunze, R. Jansen, G. Höfle, and H. J. Reichenbach. Crocacin, a new electron transport inhibitor from *Chondromyces crocatus* (Mxobacteria). production, isolation, physicochemical and biological properties. *The Journal of antibiotics*, 47(8):881–886, 1994. Cited on pages 13, 31, and 69.

[17] B. Kunze, R. Jansen, F. Sasse, G. Höfle, and H. J. Reichenbach. Chondramides A-D, new antifungal and cytostatic depsipeptides from *Chondromyces crocatus* (Myxobacteria). production, physico-chemical and biological properties. *The Journal of antibiotics*, 48(11):1262–1266, 1995. Cited on pages 13, 31, and 69.

[18] Rolf Jansen, Brigitte Kunze, Hans Reichenbach, and Gerhard Höfle. Chondrochloren A and B, new beta-amino styrenes from *Chondromyces crocatus* (Myxobacteria). *European Journal of Organic Chemistry*, 2003(14):2684–2689, 2003. Cited on pages 13, 31, and 69.

[19] Brigitte Kunze, Rolf Jansen, Gerhard Höfle, and Hans Reichenbach. Ajudazols, new inhibitors of the mitochondrial electron transport from *Chondromyces crocatus*: production, antimicrobial activity and mechanism of action. *The Journal of antibiotics*, 57(2):151–155, 2004. Cited on pages 13, 31, and 69.

[20] Kathrin Buntin, Kira J Weissman, and Rolf Müller. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *ChemBioChem*, 11(8):1137–1146, 2010. Cited on pages 13, 36, 61, and 115.

[21] Nestor Zaburannyi, Boyke Bunk, Josef Maier, Jörg Overmann, and Rolf Müller. Genome analysis of the fruiting body forming myxobacterium *Chondromyces crocatus* reveals high potential for natural product biosynthesis. *Submitted for publication*, August 2015. Cited on pages 13 and 117.

[22] Shwan Rachid, Daniel Krug, Brigitte Kunze, Irene Kochems, Maren Scharfe, T Mark Zabriskie, Helmut Blöcker, and Rolf Müller. Molecular and biochemical studies of chondramide formation - highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chemistry & biology*, 13(6):667–681, 2006. Cited on pages 13, 34, 61, 117, and 118.

[23] Thomas Hoffmann, Stefan Müller, Suvd Nadmid, Ronald Garcia, and Rolf Müller. Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. *Journal of the American Chemical Society*, 135(45):16904–16911, 2013. Cited on pages 13 and 121.

[24] Brigitte Kunze, Bettina Bohlendorf, Hans Reichenbach, and Gerhard Höfle. Pedein A and B: Production, isolation, structure elucidation and biological properties of new antifungal cyclopeptides from *Chondromyces pediculatus* (Myxobacteria). *Journal of Antibiotics*, 61(1):18–26, 2008. Cited on pages 13, 105, and 121.

[25] Steven A Haney, Lefa E Alksne, Paul M Dunman, Ellen Murphy, and Steven J Projan. Genomics in anti-infective drug discovery-getting to endgame. *Current pharmaceutical design*, 8(13):1099–1118, 2002. Cited on page 15.

[26] Sarah E Ongley, Xiaoying Bian, Youming Zhang, Rocky Chau, William H Gerwick, Rolf Müller, and Brett A Neilan. High-titer heterologous production in *E. coli* of lyngbyatoxin,

a protein kinase C activator from an uncultured marine cyanobacterium. *ACS chemical biology*, 8(9):1888–1893, 2013. Cited on page 15.

[27] Zhiyang Feng, Jeff H Kim, and Sean F Brady. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type Ii PKS gene cluster. *Journal of the American Chemical Society*, 132(34):11902–11903, 2010. Cited on page 15.

[28] Dimitris Kallifidas, Hahk-Soo Kang, and Sean F Brady. Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. *Journal of the American Chemical Society*, 134(48):19552–19555, 2012. Cited on pages 15 and 18.

[29] Brian O Bachmann and Jacques Ravel. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods in enzymology*, 458:181–217, 2009. Cited on pages 16, 34, and 83.

[30] Marc Röttig, Marnix H Medema, Kai Blin, Tilmann Weber, Christian Rausch, and Oliver Kohlbacher. NRPSpredictor2 - a web server for predicting NRPS adenylation domain specificity. *Nucleic acids research*, page gkr323, 2011. Cited on pages 16, 24, and 107.

[31] Marnix H Medema, Kai Blin, Peter Cimermancic, Victor de Jager, Piotr Zakrzewski, Michael A Fischbach, Tilmann Weber, Eriko Takano, and Rainer Breitling. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research*, 39(suppl 2):W339–W346, 2011. Cited on pages 16 and 107.

[32] Kai Blin, Marnix H Medema, Daniyal Kazempour, Michael A Fischbach, Rainer Breitling, Eriko Takano, and Tilmann Weber. antiSMASH 2.0 - a versatile platform for genome mining of secondary metabolite producers. *Nucleic acids research*, page gkt449, 2013. Cited on pages 16 and 83.

[33] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Bruccoleri, Sang Yup Lee, Michael A Fischbach, Rolf Müller, Wolfgang Wohlleben, et al. antismash 3.0a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research*, page gkv437, 2015. Cited on page 16.

[34] Hosein Mohimani, Roland D Kersten, Wei-Ting Liu, Mingxun Wang, Samuel O Purvine, Si Wu, Heather M Brewer, Ljiljana Pasa-Tolic, Nuno Bandeira, Bradley S Moore, et al. Automated genome mining of ribosomal peptide natural products. *ACS chemical biology*, 9(7):1545–1551, 2014. Cited on page 16.

[35] Lijiang Song, Luisa Laureti, Christophe Corre, Pierre Leblond, Bertrand Aigle, and Gregory L Challis. Cytochrome p450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *The Journal of antibiotics*, 67(1):71–76, 2014. Cited on page 16.

[36] Youming Zhang, Joep PP Muyrers, Giuseppe Testa, and A Francis Stewart. DNA cloning by homologous recombination in Escherichia coli. *Nature biotechnology*, 18(12):1314–1317, 2000. Cited on pages 18, 76, 100, and 113.

[37] Joep PP Muyrers, Youming Zhang, and A Francis Stewart. ET-cloning: think recombination first. In *Genetic engineering*, pages 77–98. Springer, 2000. Cited on page 18.

[38] Hailong Wang, Xiaoying Bian, Liqiu Xia, Xuezhi Ding, Rolf Müller, Youming Zhang, Jun Fu, and A Francis Stewart. Improved seamless mutagenesis by recombineering using *ccdB* for counterselection. *Nucleic acids research*, 42(5):e37–e37, 2014. Cited on pages 18 and 116.

[39] Natalay Kouprina and Vladimir Larionov. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics*, 7(10):805–812, 2006. Cited on page 18.

[40] Jeffrey H Kim, Zhiyang Feng, John D Bauer, Dimitris Kallifidas, Paula Y Calle, and Sean F Brady. Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers*, 93(9):833–844, 2010. Cited on pages 18 and 116.

[41] Daniel G Gibson. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol*, 498:349–361, 2011. Cited on page 18.

[42] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821, 2012. Cited on page 18.

[43] Ralph H Lambalot, Amy M Gehring, Roger S Flugel, Peter Zuber, Michael LaCelle, Mohamed A Marahiel, Ralph Reid, Chaitan Khosla, and Christopher T Walsh. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & biology*, 3(11):923–936, 1996. Cited on page 19.

[44] Nikolaos Gaitatzis, Astrid Hans, Rolf Müller, and Stefan Beyer. The mtaA gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases. *Journal of biochemistry*, 129(1):119–124, 2001. Cited on page 19.

[45] Rita Bernhardt. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology*, 124(1):128–145, 2006. Cited on page 19.

[46] Kerstin Maria Ewen, Michael Ringle, and Rita Bernhardt. Adrenodoxin - a versatile ferredoxin. *Iubmb Life*, 64(6):506–512, 2012. Cited on page 19.

[47] Silke C Wenzel, Frank Gross, Youming Zhang, Jun Fu, A Francis Stewart, and Rolf Müller. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chemistry & biology*, 12(3):349–356, 2005. Cited on pages 20 and 117.

[48] Frank Gross, Michael W Ring, Olena Perlova, Jun Fu, Susan Schneider, Klaus Gerth, Silvia Kuhlmann, A Francis Stewart, Youming Zhang, and Rolf Müller. Metabolic engineering of *Pseudomonas putida* for methylmalonyl-coA biosynthesis to enable complex heterologous secondary metabolite formation. *Chemistry & biology*, 13(12):1253–1264, 2006. Cited on page 20.

[49] Sarah C Mutka, John R Carney, Yaoquan Liu, and Jonathan Kennedy. Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry*, 45(4):1321–1330, 2006. Cited on pages 20 and 117.

[50] Christian Hertweck. The biosynthetic logic of polyketide diversity. *Angewandte Chemie International Edition*, 48(26):4688–4716, 2009. Cited on page 20.

[51] Stuart Smith and Shiou-Chuan Tsai. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports*, 24(5):1041–1072, 2007. Cited on page 20.

[52] Chaitan Khosla, Daniel Herschlag, David E Cane, and Christopher T Walsh. Assembly line polyketide synthases: Mechanistic insights and unsolved problems. *Biochemistry*, 53(18):2875–2883, 2014. Cited on page 21.

[53] Adrian T Keatinge-Clay. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chemistry & biology*, 14(8):898–908, 2007. Cited on pages 22 and 83.

[54] Patrick Caffrey. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem*, 4(7):654–657, 2003. Cited on pages 22 and 83.

[55] Florian Taft, Tobias Knobloch, Heinz G Floss, and Andreas Kirschning. Timing of the $\delta^{10,12}$-$\delta^{11,13}$ double bond migration during ansamitocin biosynthesis in *Actinosynnema pretiosum*. *Journal of the American Chemical Society*, 131(11):3812–3813, 2009. Cited on pages 22 and 83.

[56] David H Kwan and Peter F Leadlay. Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS chemical biology*, 5(9):829–838, 2010. Cited on page 22.

[57] Liangcheng Du and Lili Lou. PKS and NRPS release mechanisms. *Natural product reports*, 27(2):255–278, 2010. Cited on page 23.

[58] Mohammed A Marahiel and Lars-Oliver Essen. Nonribosomal peptide synthetases: mechanistic and structural aspects of essential domains. *Methods in enzymology*, 458:337–351, 2009. Cited on page 23.

[59] Henning D Mootz, Dirk Schwarzer, and Mohamed A Marahiel. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem*, 3(6):490–504, 2002. Cited on page 23.

[60] Johnson Cheung, Federico C Beasley, Suya Liu, Gilles A Lajoie, and David E Heinrichs. Molecular characterization of staphyloferrin B biosynthesis in staphylococcus aureus. *Molecular microbiology*, 74(3):594–608, 2009. Cited on page 23.

[61] Jorge H Crosa and Christopher T Walsh. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiology and Molecular Biology Reviews*, 66(2):223–249, 2002. Cited on page 24.

[62] Stephan A Sieber and Mohamed A Marahiel. Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides. *Journal of bacteriology*, 185(24):7036–7043, 2003. Cited on page 24.

[63] Elena Conti, Torsten Stachelhaus, Mohamed A Marahiel, and Peter Brick. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO Journal*, 16(14):4174–4183, 1997. Cited on page 24.

[64] Torsten Stachelhaus, Henning D Mootz, and Mohamed A Marahiel. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & biology*, 6(8):493–505, 1999. Cited on pages 24 and 35.

[65] Gregory L Challis, Jacques Ravel, and Craig A Townsend. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & biology*, 7(3):211–224, 2000. Cited on page 24.

[66] Nathan A Magarvey, Zachary Q Beck, Trimurtulu Golakoti, Yousong Ding, Udo Huber, Thomas K Hemscheidt, Dafna Abelson, Richard E Moore, and David H Sherman. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. potent anticancer agents from *Nostoc* cyanobionts. *ACS chemical biology*, 1(12):766–779, 2006. Cited on page 25.

[67] Daniela Reimer, Klaas M Pos, Marco Thines, Peter Grün, and Helge B Bode. A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nature chemical biology*, 7(12):888–890, 2011. Cited on pages 25 and 116.

[68] Mohamed S Donia, Jacques Ravel, and Eric W Schmidt. A global assembly line for cyanobactins. *Nature chemical biology*, 4(6):341–343, 2008. Cited on page 25.

[69] Paul G Arnison et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports*, 30(1):108–160, 2013. Cited on pages 25 and 69.

[70] Kyle L Dunbar and Douglas A Mitchell. Revealing Nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS chemical biology*, 8(3):473–487, 2013. Cited on page 25.

[71] Thomas K Hemscheidt. *Microviridin Biosynthesis*, volume 516. Academic Press: San Diego, CA, USA, 2012. Cited on page 25.

[72] Julian D Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, and Mohamed A Marahiel. Lasso peptides from proteobacteria: genome mining employing heterologous expression and mass spectrometry. *Peptide Science*, 100(5):527–542, 2013. Cited on page 25.

[73] Volker Pfeifer, Graeme J Nicholson, Johannes Ries, Jürgen Recktenwald, Alexandre B Schefer, Riham M Shawky, Joachim Schröder, Wolfgang Wohlleben, and Stefan Pelzer. A polyketide synthase in glycopeptide biosynthesis the biosynthesis of the non-proteinogenic amino acid (s)-3, 5-dihydroxyphenylglycine. *Journal of Biological Chemistry*, 276(42):38370–38377, 2001. Cited on page 26.

[74] Li Xu, He Huang, Wei Wei, Yi Zhong, Biao Tang, Hua Yuan, Li Zhu, Weiyi Huang, Mei Ge, Shen Yang, et al. Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC genomics*, 15(1):363, 2014. Cited on page 26.

[75] Yuan Liu, Christopher Hazzard, Alessandra S Eustáquio, Kevin A Reynolds, and Bradley S Moore. Biosynthesis of salinosporamides from α, β-unsaturated fatty acids: implications for extending polyketide synthase diversity. *Journal of the American Chemical Society*, 131(30):10376–10377, 2009. Cited on page 26.

[76] Hye-Gyeong Yoo, So-Yeon Kwon, Suji Kim, Suman Karki, Zee-Yong Park, and Hyung-Jin Kwon. Characterization of 2-octenoyl-coA carboxylase/reductase utilizing *pteB* from *Streptomyce avermitilis*. *Bioscience, biotechnology, and biochemistry*, 75(6):1191–1193, 2011. Cited on page 26.

[77] Tobias J Erb. Carboxylases in natural and synthetic microbial pathways. *Applied and environmental microbiology*, 77(24):8466–8477, 2011. Cited on page 26.

[78] Friederike Lohr, Imke Jenniches, Maxim Frizler, Michael J Meehan, Marc Sylvester, Alexander Schmitz, Michael Gütschow, Pieter C Dorrestein, Gabriele M König, and Till F Schäberle. α, β β, γ double bond migration in corallopyronin A biosynthesis. *Chemical Science*, 4(11):4175–4180, 2013. Cited on pages 27 and 83.

[79] Tom Bretschneider, Joel B Heim, Daniel Heine, Robert Winkler, Benjamin Busch, Björn Kusebauch, Thilo Stehle, Georg Zocher, and Christian Hertweck. Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature*, 502(7469):124–128, 2013. Cited on page 27.

[80] Daniel Heine, Tom Bretschneider, Srividhya Sundaram, and Christian Hertweck. Enzymatic polyketide chain branching to give substituted lactone, lactam, and glutarimide heterocycles. *Angewandte Chemie International Edition*, 53(43):11645–11649, 2014. Cited on page 27.

[81] Isao Fujii, Naho Yoshida, Shigeki Shimomaki, Hideaki Oikawa, and Yutaka Ebizuka. An iterative type I polyketide synthase PKS catalyzes synthesis of the decaketide alternapyrone with regio-specific octa-methylation. *Chemistry & biology*, 12(12):1301–1309, 2005. Cited on page 27.

[82] Ho Young Lee, Hak Suk Chung, Chao Hang, Chaitan Khosla, Christopher T Walsh, Daniel Kahne, and Suzanne Walker. Reconstitution and characterization of a new desosaminyl transferase, EryCIII, from the erythromycin biosynthetic pathway. *Journal of the American Chemical Society*, 126(32):9924–9925, 2004. Cited on page 27.

[83] Silke C Wenzel and Rolf Müller. The biosynthetic potential of myxobacteria and their impact in drug discovery. *Current opinion in drug discovery & development*, 12(2):220–230, 2009. Cited on pages 31, 81, and 105.

[84] Silke C Wenzel and Rolf Müller. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Natural product reports*, 24(6):1211–1224, 2007. Cited on page 31.

[85] Heinrich Steinmetz, Herbert Irschik, Brigitte Kunze, Hans Reichenbach, Gerhard Hoefle, and Rolf Jansen. Thuggacins, macrolide antibiotics active against *Mycobacterium tuberculosis*: isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chemistry-A European Journal*, 13(20):5822–5832, 2007. Cited on pages 31, 69, 117, and 118.

[86] Niña Socorro Cortina, Daniel Krug, Alberto Plaza, Ole Revermann, and Rolf Müller. Myxoprincomide: a natural product from myxococcus xanthus discovered by comprehensive analysis of the secondary metabolome. *Angewandte Chemie International Edition*, 51(3):811–816, 2012. Cited on pages 31, 61, 109, 116, and 118.

[87] Martin Welker and Hans Von Döhren. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS microbiology reviews*, 30(4):530–563, 2006. Cited on pages 31 and 121.

[88] Hiroshi Yamaki, Namthip Sitachitta, Tomoharu Sano, and Kunimitsu Kaya. Two new chymotrypsin inhibitors isolated from the Cyanobacterium *Microcystis aeruginosa* NIES-88. *Journal of natural products*, 68(1):14–18, 2005. Cited on pages 31 and 42.

[89] Ute Matern, Christian Schleberger, Stjepan Jelakovic, Jürgen Weckesser, and Georg E Schulz. Binding structure of elastase inhibitor scyptolin A. *Chemistry & biology*, 10(10):997–1001, 2003. Cited on pages 31, 32, and 34.

[90] Sara C Stolze, Michael Meltzer, Michael Ehrmann, and Markus Kaiser. Solid phase total synthesis of the 3-amino-6-hydroxy-2-piperidone (Ahp) cyclodepsipeptide and protease inhibitor symplocamide A. *Chemical Communications*, 46(46):8857–8859, 2010. Cited on pages 32 and 40.

[91] C. Martin, L. Oberer, T. Ino, W. A. Konig, and J. J. Busch, M.and Weckesser. Cyanopeptolins, new depsipeptides from the cyanobacterium *Microcystis* sp. PCC 7806. *The Journal of antibiotics*, 46(10):1550–1556, 1993. Cited on page 32.

[92] Elmi N Zainuddin, Renate Mentel, Victor Wray, Rolf Jansen, Manfred Nimtz, Michael Lalk, and Sabine Mundt. Cyclic depsipeptides, ichthyopeptins A and B, from *Microcystis*

ichthyoblabe. *Journal of natural products*, 70(7):1084–1088, 2007. Cited on pages 32 and 121.

[93] P. Krastel, B.-M. Liechty, J. G. Meingassner, E. Schmitt, and E. P; Schreiner. Cyclic depsipeptides. wo patent, 2009. Cited on pages 32 and 34.

[94] Trine B Rounge, Thomas Rohrlack, Tom Kristensen, and Kjetill S Jakobsen. Recombination and selectional forces in cyanopeptolin NRPS operons from highly similar, but geographically remote *Planktothrix* strains. *BMC microbiology*, 8(1):141, 2008. Cited on pages 32 and 120.

[95] P. Krastel, B.-M. Liechty, J. G. Meingassner, C. Moore, and E. Schmitt. Nucleic acid of a biosynthetic cluster encoding non ribosomal peptide synthases and uses therof. wo patent, 2010. Cited on page 34.

[96] Atanas V Koulov, Paul LaPointe, Bingwen Lu, Abbas Razvi, Judith Coppinger, Meng-Qiu Dong, Jeanne Matteson, Rob Laister, Cheryl Arrowsmith, John R Yates, et al. Biological and structural basis for aha1 regulation of Hsp90 ATPase activity in maintaining proteostasis in the human disease cystic fibrosis. *Molecular biology of the cell*, 21(6):871–884, 2010. Cited on page 35.

[97] Yogan Khatri, Frank Hannemann, Kerstin M Ewen, Dominik Pistorius, Olena Perlova, Norio Kagawa, Alexander O Brachmann, Rolf Müller, and Rita Bernhardt. The CYPome of *Sorangium cellulosum* So ce56 and identification of CYP109d1 as a new fatty acid hydroxylase. *Chemistry & biology*, 17(12):1295–1305, 2010. Cited on pages 37, 38, and 97.

[98] Kerstin Maria Ewen, Frank Hannemann, Yogan Khatri, Olena Perlova, Reinhard Kappl, Daniel Krug, Jürgen Hüttermann, Rolf Müller, and Rita Bernhardt. Genome mining in *Sorangium cellulosum* So ce56 identification and characterization of the homologous electron transfer proteins of a myxobacterial cytochrome P450. *Journal of Biological Chemistry*, 284(42):28590–28598, 2009. Cited on page 38.

[99] Haitham A Hussain and John M Ward. Enhanced heterologous expression of two *Streptomyces griseolus* cytochrome P450s and *Streptomyces coelicolor* ferredoxin reductase as potentially efficient hydroxylation catalysts. *Applied and environmental microbiology*, 69(1):373–382, 2003. Cited on page 39.

[100] Lucy C Foulston and Mervyn J Bibb. Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proceedings of the National Academy of Sciences*, 107(30):13461–13466, 2010. Cited on page 41.

[101] Toru Kino, Hiroshi Hatanaka, Michisane Hashimoto, Michihisa Nishiyama, Toshio Goto, Masakuni Okuhara, Masanobu Kohsaka, Hatsuo Aoki, and Hiroshi Imanaka. Fk-506, a novel immunosuppressant isolated from a streptomyces. i. fermentation, isolation, and physico-chemical and biological characteristics. *The Journal of antibiotics*, 40(9):1249–1255, 1987. Cited on page 61.

[102] M. Debono, B. J. Abbott, R. M. Molloy, D. S. Fukuda, A. H. Hunt, V. M. Daupert, F. T. Counter, J. L. Ott, C. B. Carrell, and L. C. Howard. Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J Antibiot (Tokyo)*, 41(8):1093–1105, Aug 1988. Cited on page 61.

[103] G. Kolata. Drug transforms transplant medicine. *Science*, 221(4605):40–42, Jul 1983. Cited on page 61.

[104] Christopher T Walsh, Robert V O'Brien, and Chaitan Khosla. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angewandte Chemie International Edition*, 52(28):7098–7124, 2013. Cited on page 61.

[105] Kathrin Buntin, Shwan Rachid, Maren Scharfe, Helmut Blöcker, Kira J Weissman, and Rolf Müller. Production of the antifungal isochromanone ajudazols a and b in chondromyces crocatus cm c5: biosynthetic machinery and cytochrome p450 modifications. *Angewandte Chemie International Edition*, 47(24):4595–4599, 2008. Cited on page 61.

[106] Stefan Müller, Shwan Rachid, Thomas Hoffmann, Frank Surup, Carsten Volz, Nestor Zaburannyi, and Rolf Müller. Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chemistry & biology*, 21(7):855–865, 2014. Cited on pages 61, 81, 117, and 118.

[107] Shwan Rachid, Maren Scharfe, Helmut Blöcker, Kira J Weissman, and Rolf Müller. Unusual chemistry in the biosynthesis of the antibiotic chondrochlorens. *Chemistry & biology*, 16(1):70–81, 2009. Cited on pages 61, 117, and 118.

[108] Konrad Viehrig, Frank Surup, Kirsten Harmrolfs, Rolf Jansen, Brigitte Kunze, and Rolf Mu ller. Concerted action of p450 plus helper protein to form the amino-hydroxypiperidone moiety of the potent protease inhibitor crocapeptin. *Journal of the American Chemical Society*, 135(45):16885–16894, 2013. Cited on pages 61, 69, 75, 97, 112, 117, and 118.

[109] Nobuaki Matsumori, Daisuke Kaneno, Michio Murata, Hideshi Nakamura, and Kazuo Tachibana. Stereochemical determination of acyclic structures based on carbon-proton spin-coupling constants. a method of configuration analysis for natural products. *The Journal of organic chemistry*, 64(3):866–876, 1999. Cited on page 64.

[110] Peter Marfey. Determination of D-amino acids. II. *Carlsberg Research Communications*, 49(6):591–596, 1984. Cited on page 64.

[111] Thomas R Hoye, Christopher S Jeffrey, and Feng Shao. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nature Protocols*, 2(10):2451–2458, 2007. Cited on page 64.

[112] Gitanjali M Singh, Pascal D Fortin, Alexander Koglin, and Christopher T Walsh. β-hydroxylation of the aspartyl residue in the phytotoxin syringomycin e: Characterization of two candidate hydroxylases asph and syrp in pseudomonas syringae. *Biochemistry*, 47(43):11310–11320, 2008. Cited on page 65.

[113] U. Hornemann, L. H. Hurley, M. K. Speedie, and H. G. Floss. Biosynthesis of indolmycin. *Journal of the American Chemical Society*, 93(12):3028–3035, 1971. Cited on page 65.

[114] Hamao Umezawa, Takaaki Aoyagi, Hazime Morishima, Meiki Matsuzaki, Masa Hamada, and Tomio Takeuchi. Pepstatin, a new pepsin inhibitor produced by agtinomygetes. *The Journal of antibiotics*, 23(5):259–262, 1970. Cited on page 65.

[115] Masahiro Okada, Isao Sato, Soo Jeong Cho, Hidehisa Iwata, Toshihiko Nishio, David Dubnau, and Youji Sakagami. Structure of the *Bacillus subtilis* quorum-sensing peptide pheromone ComX. *Nature chemical biology*, 1(1):23–24, 2005. Cited on page 77.

[116] Zhongli Xu, Martin Baunach, Ling Ding, and Christian Hertweck. Bacterial synthesis of diverse indole terpene alkaloids by an unparalleled cyclization sequence. *Angewandte Chemie International Edition*, 51(41):10293–10297, 2012. Cited on page 77.

[117] P. Pauletti, L. Cintra, C.s Braguine, W. Cunha, and A. Januario. Halogenated indole alkaloids from marine invertebrates. *Marine drugs*, 8(5):1526–1549, 2010. Cited on page 77.

[118] Mireille Ansaldi, Darja Marolt, Tina Stebe, Ines Mandic-Mulec, and David Dubnau. Specific activation of the bacillus quorum-sensing systems by isoprenylated pheromone variants. *Molecular microbiology*, 44(6):1561–1573, 2002. Cited on page 78.

[119] Erin K Sully, Natalia Malachowa, Bradley O Elmore, Susan M Alexander, Jon K Femling, Brian M Gray, Frank R DeLeo, Michael Otto, Ambrose L Cheung, Bruce S Edwards, et al. Selective chemical inhibition of agr quorum sensing in staphylococcus aureus promotes host defense with minimal impact on resistance. 2014. Cited on page 78.

[120] Michael P Storz, Christian Brengel, Elisabeth Weidel, Michael Hoffmann, Klaus Hollemeyer, Anke Steinbach, Rolf Mu ller, Martin Empting, and Rolf W Hartmann. Biochemical and biophysical analysis of a chiral pqsd inhibitor revealing tight-binding behavior and enantiomers with contrary thermodynamic signatures. *ACS chemical biology*, 8(12):2794–2801, 2013. Cited on page 78.

[121] Robert C Moellering, Jr. MRSA: the first half century. *J Antimicrob Chemother*, 67(1):4–11, Jan 2012. Cited on page 81.

[122] Michael A. Fischbach and Christopher T. Walsh. Antibiotics for emerging pathogens. *Science*, 325(5944):1089–1093, Aug 2009. Cited on page 81.

[123] Gerhard Höfle, Norbert Bedorf, Heinrich Steinmetz, Dietmar Schomburg, Klaus Gerth, and Hans Reichenbach. Epothilone A and B - novel 16-membered macrolides with cytotoxic activity: Isolation, crystal structure, and conformation in solution. *Angewandte Chemie International Edition in English*, 35(13-14):1567–1569, 1996. Cited on page 81.

[124] Herbert Irschik, Rolf Jansen, Klaus Gerth, Gerhart Höfle, and Hans Reichenbach. The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria. *The Journal of antibiotics*, 40(1):7–13, 1987. Cited on page 81.

[125] H. F. Vahlensieck, L. Pridzun, H. Reichenbach, and A. Hinnen. Identification of the yeast ACC1 gene product (acetyl-coA carboxylase) as the target of the polyketide fungicide soraphen A. *Current genetics*, 25(2):95–100, 1994. Cited on page 81.

[126] Anne Osbourn, Rebecca Goss, and Guy T Carter. *Natural Products: Discourse, Diversity, and Design.* John Wiley & Sons, 2014. Cited on page 81.

[127] Daniel Krug and Rolf Müller. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Natural product reports*, 31(6):768–783, 2014. Cited on page 81.

[128] John Buckingham. *Dictionary of natural products.* CRC press, 1993. Cited on page 81.

[129] Pawan K Agrawal. NMR spectroscopy in the structural elucidation of oligosaccharides and glycosides. *Phytochemistry*, 31(10):3307–3330, 1992. Cited on page 82.

[130] Gerit J Gerwig, Johannis P Kamerling, and Johannes FG Vliegenthart. Determination of the absolute configuration of monosaccharides in complex carbohydrates by capillary GLC. *Carbohydrate research*, 77(1):1–7, 1979. Cited on page 82.

[131] S. Beyer, B. Kunze, B. Silakowski, and R. Müller. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim Biophys Acta*, 1445(2):185–195, May 1999. Cited on page 83.

[132] Gitanjali Yadav, Rajesh S Gokhale, and Debasisa Mohanty. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *Journal of molecular biology*, 328(2):335–363, 2003. Cited on page 83.

[133] Sebastian Essig, Sebastian Bretzke, Rolf Müller, and Dirk Menche. Full stereochemical determination of ajudazols A and B by bioinformatics gene cluster analysis and total synthesis of ajudazol B by an asymmetric ortholithiation strategy. *Journal of the American Chemical Society*, 134(47):19362–19365, 2012. Cited on pages 83, 117, and 118.

[134] Dominic Janssen, Dieter Albert, Rolf Jansen, Rolf Müller, and Markus Kalesse. Chivosazole A - elucidation of the absolute and relative configuration. *Angewandte Chemie International Edition*, 46(26):4898–4901, 2007. Cited on page 83.

[135] Björn Kusebauch, Benjamin Busch, Kirstin Scherlach, Martin Roth, and Christian Hertweck. Functionally distinct modules operate two consecutive α, β β, γ double-bond shifts in the rhizoxin polyketide assembly line. *Angewandte Chemie International Edition*, 49(8):1460–1464, 2010. Cited on page 83.

[136] Jana Moldenhauer, Daniel CG Götz, Christian R Albert, Sebastian K Bischof, Kathrin Schneider, Roderich D Süssmuth, Marianne Engeser, Harald Gross, Gerhard Bringmann, and Jörn Piel. The final steps of bacillaene biosynthesis in *Bacillus amyloliquefaciens* fzb42: Direct evidence for β, γ dehydration by a trans-acyltransferase polyketide synthase. *Angewandte Chemie*, 122(8):1507–1509, 2010. Cited on page 83.

[137] Maren Kopp, Herbert Irschik, Frank Gross, Olena Perlova, Axel Sandmann, Klaus Gerth, and Rolf Müller. Critical variations of conjugational DNA transfer into secondary metabolite multiproducing *Sorangium cellulosum* strains So ce12 and So ce56: development of a mariner-based transposon mutagenesis system. *Journal of biotechnology*, 107(1):29–40, 2004. Cited on page 84.

[138] Yi Chai, Dominik Pistorius, Angelika Ullrich, Kira J Weissman, Uli Kazmaier, and Rolf Müller. Discovery of 23 natural tubulysins from *Angiococcus disciformis* And48 and *Cystobacter* SBCb004. *Chemistry & biology*, 17(3):296–309, 2010. Cited on page 84.

[139] E. St. Duthie and Lisa L. Lorenz. Staphylococcal coagulase: mode of action and antigenicity. *Journal of general microbiology*, 6(1-2):95–107, 1952. Cited on page 84.

[140] Kenji Okonogi, Yumiko Noji, Masahiro Kondo, Akira Imada, and Takeshi Yokota. Emergence of methicillin-resistant clones from cephamycin-resistant *Staphylococcus aureus*. *Journal of Antimicrobial Chemotherapy*, 24(5):637–645, 1989. Cited on page 84.

[141] Antonio A Iniesta, Francisco García-Heras, Javier Abellón-Ruiz, Aránzazu Gallego-García, and Montserrat Elías-Arnanz. Two systems for conditional gene expression in myxococcus xanthus inducible by isopropyl-β-d-thiogalactopyranoside or vanillate. *Journal of bacteriology*, 194(21):5875–5885, 2012. Cited on page 92.

[142] Bertrand Aigle and Christophe Corre. Waking up streptomyces secondary metabolism by constitutive expression of activators or genetic disruption of repressors. *Methods in enzymology*, 517:343–366, 2011. No cited.

[143] Alberto Plaza, Ronald Garcia, Giuseppe Bifulco, Javier Pablo Martinez, Stephan Hüttel, Florenz Sasse, Andreas Meyerhans, Marc Stadler, and Rolf Müller. Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus Aetherobacter. *Organic letters*, 14(11):2854–2857, 2012. Cited on page 105.

[144] Eric W Schmidt and D John Faulkner. Microsclerodermins C-E, antifungal cyclic peptides from the lithistid marine sponges *Theonella* sp. and *Microscleroderma* sp. *Tetrahedron*, 54(13):3043–3056, 1998. Cited on page 105.

[145] Nobuhiro Fusetani, Shigeki Matsunaga, Hisao Matsumoto, and Yukihiro Takebayashi. Bioactive marine metabolites. 33. Cyclotheonamides, potent thrombin inhibitors, from a marine sponge *Theonella sp. Journal of the American Chemical Society*, 112(19):7053–7054, 1990. Cited on pages 106 and 109.

[146] Ignacy Z Siemion, Theodor Wieland, and Karl-Heinz Pook. Influence of the distance of the proline carbonyl from the β and γ carbon on the ^{13}C chemical shifts. *Angewandte Chemie International Edition in English*, 14(10):702–703, 1975. Cited on page 106.

[147] Fumio Itagaki, Hideyuki Shigemori, Masami Ishibashi, Takemichi Nakamura, Takuma Sasaki, and Junichi Kobayashi. Keramamide F, a new thiazole-containing peptide from the okinawan marine sponge *Theonella* sp. *The Journal of Organic Chemistry*, 57(20):5540–5542, 1992. Cited on page 106.

[148] Zohreh Hojati, Claire Milne, Barbara Harvey, Lyndsey Gordon, Matthew Borg, Fiona Flett, Barrie Wilkinson, Philip J Sidebottom, Brian AM Rudd, Martin A Hayes, et al. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chemistry & biology*, 9(11):1175–1187, 2002. Cited on page 106.

[149] Kiyonaga Fujii, Yoshitomo Ikai, Hisao Oka, Makoto Suzuki, and Ken-ichi Harada. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: combination of Marfey's method with mass spectrometry and its practical application. *Analytical Chemistry*, 69(24):5146–5151, 1997. Cited on page 106.

[150] Kiyonaga Fujii, Yoshitomo Ikai, Tsuyoshi Mayumi, Hisao Oka, Makoto Suzuki, and Ken-ichi Harada. A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: elucidation of limitations of Marfey's method and of its separation mechanism. *Analytical Chemistry*, 69(16):3346–3352, 1997. Cited on page 106.

[151] Mu Wang and Steven J Gould. Biosynthesis of capreomycin. 2. incorporation of L-serine, L-alanine, and L-2, 3-diaminopropionic acid. *The Journal of Organic Chemistry*, 58(19):5176–5180, 1993. Cited on page 107.

[152] Federico C Beasley, Johnson Cheung, and David E Heinrichs. Mutation of L-2, 3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*. *BMC microbiology*, 11(1):199, 2011. Cited on pages 107 and 109.

[153] Jason M Crawford, Cyril Portmann, Renee Kontnik, Christopher T Walsh, and Jon Clardy. NRPS substrate promiscuity diversifies the xenematides. *Organic letters*, 13(19):5144–5147, 2011. Cited on page 107.

[154] Roger Genet, Pierre-Henri Bénetti, Akli Hammadi, and André Ménez. L-tryptophan 2 , 3 -oxidase from *Chromobacterium violaceum* substrate specificity and mechanistic implications. *Journal of Biological Chemistry*, 270(40):23540–23545, 1995. Cited on page 109.

[155] Gregory L Challis. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem*, 6(4):601–611, 2005. Cited on page 109.

[156] Miki Kimura, Toshiyuki Wakimoto, Yoko Egami, Karen Co Tan, Yuji Ise, and Ikuro Abe. Calyxamides A and B, cytotoxic cyclic peptides from the marine sponge *Discodermia calyx*. *Journal of natural products*, 75(2):290–294, 2012. Cited on page 109.

[157] Youichi Nakao, Shigeki Matsunaga, and Nobuhiro Fusetani. Three more cyclotheonamides, C, D, and E, potent thrombin inhibitors from the marine sponge *Theonella swinhoei*. *Bioorganic & medicinal chemistry*, 3(8):1115–1122, 1995. Cited on page 109.

[158] Yoichi Nakao, Naoya Oku, Shigeki Matsunaga, and Nobuhiro Fusetani. Cyclotheonamides E2 and E3, new potent serine protease inhibitors from the marine sponge of the genus *Theonella*. *Journal of natural products*, 61(5):667–670, 1998. Cited on page 109.

[159] Alberto Plaza, Konrad Viehrig, Ronald Garcia, and Rolf Müller. Jahnellamides, α-keto-β-methionine-containing peptides from the terrestrial myxobacterium *Jahnella* sp.: structure and biosynthesis. *Organic letters*, 15(22):5882–5885, 2013. Cited on page 113.

[160] Frank Surup, Konrad Viehrig, Kathrin I Mohr, Jennifer Herrmann, Rolf Jansen, and Rolf Müller. Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angewandte Chemie International Edition*, 53(49):13588–13591, 2014. Cited on page 113.

[161] Kazuya Yamanaka, Kirk A Reynolds, Roland D Kersten, Katherine S Ryan, David J Gonzalez, Victor Nizet, Pieter C Dorrestein, and Bradley S Moore. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proceedings of the National Academy of Sciences*, 111(5):1957–1962, 2014. Cited on page 113.

[162] Isabelle Huys, Geertrui Van Overwalle, and Gert Matthijs. Gene and genetic diagnostic method patent claims: a comparison under current european and us patent law. *European Journal of Human Genetics*, 19(10):1104–1107, 2011. Cited on page 113.

[163] Corina Oßwald, Gregor Zipf, Gisela Schmidt, Josef Maier, Hubert S Bernauer, Rolf Müller, and Silke C Wenzel. Modular construction of a functional artificial epothilone polyketide pathway. *ACS synthetic biology*, 3(10):759–772, 2012. Cited on page 116.

[164] Emilie Yeterian, Lois W Martin, Laurent Guillon, Laure Journet, Iain L Lamont, and Isabelle J Schalk. Synthesis of the siderophore pyoverdine in *Pseudomonas aeruginosa* involves a periplasmic maturation. *Amino acids*, 38(5):1447–1459, 2010. Cited on page 116.

[165] Olena Perlova, Jun Fu, Silvia Kuhlmann, Daniel Krug, A Francis Stewart, Youming Zhang, and Rolf Müller. Reconstitution of the myxothiazol biosynthetic gene cluster by red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Applied and environmental microbiology*, 72(12):7485–7494, 2006. Cited on page 117.

[166] Yi Chai, Shiping Shan, Kira J Weissman, Shengbiao Hu, Youming Zhang, and Rolf Müller. Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using red/et recombineering and inactivation mutagenesis. *Chemistry & biology*, 19(3):361–371, 2012. Cited on page 117.

[167] Jun Fu, Silke C Wenzel, Olena Perlova, Junping Wang, Frank Gross, Zhiru Tang, Yulong Yin, A Francis Stewart, Rolf Müller, and Youming Zhang. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic acids research*, 36(17):e113–e113, 2008. Cited on page 117.

[168] Bryan Julien and Sanjay Shah. Heterologous expression of epothilone biosynthetic genes in myxococcus xanthus. *Antimicrobial agents and chemotherapy*, 46(9):2772–2778, 2002. Cited on page 117.

[169] D Cole Stevens, Michael R Henry, Kimberly A Murphy, and Christopher N Boddy. Heterologous expression of the oxytetracycline biosynthetic pathway in myxococcus xanthus. *Applied and environmental microbiology*, 76(8):2681–2683, 2010. Cited on page 117.

[170] Helge Björn Bode, Barbara Bethe, Regina Höfs, and Axel Zeeck. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem*, 3(7):619–627, 2002. Cited on page 118.

[171] Rolf Müller and Joachim Wink. Future potential for anti-infectives from bacteria–how to exploit biodiversity and genomic potential. *International Journal of Medical Microbiology*, 304(1):3–13, 2014. Cited on page 119.

[172] Till F Schäberle, Annika Orland, and Gabriele M König. Enhanced production of undecylprodigiosin in *Streptomyces coelicolor* by co-cultivation with the corallopyronin A-producing myxobacterium, textitCorallococcus coralloides. *Biotechnology letters*, 36(3):641–648, 2014. Cited on page 119.

[173] Thomas Hoffmann, Daniel Krug, Stephan Hüttel, and Rolf Müller. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Analytical chemistry*, 86(21):10780–10788, 2014. Cited on page 119.

[174] Daniel Krug, Gabriela Zurek, Ole Revermann, Michiel Vos, Gregory J Velicer, and Rolf Müller. Discovering the hidden secondary metabolome of myxococcus xanthus: a study of intraspecific diversity. *Applied and environmental microbiology*, 74(10):3058–3068, 2008. Cited on page 119.

[175] Simona Francese, Francesca R Dani, Pietro Traldi, Guido Mastrobuoni, Giuseppe Pieraccini, and Gloriano Moneti. Maldi mass spectrometry imaging, from its origins up to today: the state of the art. *Combinatorial chemistry & high throughput screening*, 12(2):156–174, 2009. Cited on page 119.

[176] Chunxu Song, Mark Mazzola, Xu Cheng, Janina Oetjen, Theodore Alexandrov, Pieter Dorrestein, Jeramie Watrous, Menno van der Voort, and Jos M Raaijmakers. Molecular and chemical dialogues in bacteria-protozoa interactions. *Scientific Reports*, 5, 2015. Cited on page 119.

[177] Delphine Debois, Emmanuel Jourdan, Nicolas Smargiasso, Philippe Thonart, Edwin De Pauw, and Marc Ongena. Spatiotemporal monitoring of the antibiome secreted by bacillus biofilms on plant roots using maldi mass spectrometry imaging. *Analytical chemistry*, 86(9):4431–4438, 2014. Cited on page 119.

[178] Philipp Krastel, Brigitta-Maria Liechty, Esther Schmitt, and Erwin Paul Schreiner. Use of cyclic depsipeptides to inhibit kallikrein 7, January 17 2013. US Patent App. 13/743,611. Cited on page 120.

[179] Brigitte Kunze, Rolf Jansen, Florenz Sasse, Gerhard Höfle, and Hans Reichenbach. Apicularens A and B, new cytostatic macrolides from *Chondromyces* species (myxobacteria): production, physico-chemical and biological properties. *The Journal of antibiotics*, 51(12):1075–1080, 1998. Cited on page 121.

[180] Karen L Erickson, John A Beutler, John H Cardellina, and Michael R Boyd. Salicylihalamides A and B, novel cytotoxic macrolides from the marine sponge *Haliclona* sp. *The Journal of organic chemistry*, 62(23):8188–8192, 1997. Cited on page 121.

[181] Ute Hentschel, Jörn Piel, Sandie M Degnan, and Michael W Taylor. Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*, 10(9):641–654, 2012. Cited on page 121.

[182] William H Gerwick and Bradley S Moore. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chemistry & biology*, 19(1):85–98, 2012. Cited on page 122.

[183] Mamoru Komatsu, Takuma Uchiyama, Satoshi Ōmura, David E Cane, and Haruo Ikeda. Genome-minimized streptomyces host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences*, 107(6):2646–2651, 2010. Cited on page 122.

[184] George H McArthur and Stephen S Fong. Toward engineering synthetic microbial metabolism. *BioMed Research International*, 2010, 2009. Cited on page 123.

[185] Cheol-Min Ghim, Taesung Kim, Robert J Mitchell, and Sung Kuk Lee. Synthetic biology for biofuels: building designer microbes from the scratch. *Biotechnology and Bioprocess Engineering*, 15(1):11–21, 2010. Cited on page 123.

[186] Maureen B Quin and Claudia Schmidt-Dannert. Designer microbes for biosynthesis. *Current opinion in biotechnology*, 29:55–61, 2014. Cited on page 123.

[187] Kristala L Jones Prather and Collin H Martin. *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Current Opinion in Biotechnology*, 19(5):468–474, 2008. Cited on page 123.

Chapter 10

Appendix

10.1 Supporting Information: Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin

Supporting Information

Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin

Konrad Viehrig[†], Frank Surup[‡], Kirsten Harmrolfs[†], Rolf Jansen[‡], Brigitte Kunze[§], and Rolf Müller*[,†]

[†]*Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research and Department of Pharmaceutical Biotechnology, Saarland University, Building C 2.3, D-66123 Saarbru ¨cken, Germany*

[‡]*Helmholtz Centre for Infection Research, Department of Microbial Drugs, Inhoffenstraße 7, D-38124 Braunschweig, Germany*

[§]*Helmholtz Centre for Infection Research, Inhoffenstraße 7, D-38124 Braunschweig, Germany*

TABLE OF CONTENTS

S1: HR-MS analysis of crocapeptins and precrocapeptins

Chondromyces crocatus Cm c5 cells and culture supernatant were extracted as described in the General Experimental Procedures section. Peaks were separated using a 20 min gradient from 5 to 95% ACN. Crocapeptin A peaks were detected as $[M\text{-}OH]^+$, $[M+H]^+$ and $[M+NH_4]^+$ ions with the following retention times and m/z:

	RT	$[M\text{-}OH]^+$	$[M+H]^+$	$[M+NH_4]$
crocapeptin A1	7.7'	917.5	935.5	952.5
crocapeptin A2	8.0'	931.5	949.5	963.5
crocapeptin A3	8.4'	952.5	966.5	980.5

In culture supernatant extracts of *cpnF*⁻ Cm c5 mutants, the non-hydroxylated compounds were detected as $[M+H]^+$ and $[M+NH_4]$+ with the following retention times and m/z:

	RT	$[M+H]^+$	$[M+NH4]^+$
pre-crocapeptin A1	7.9'	919.49	936.52
pre-crocapeptin A2	8.3'	933.52	950.54
pre-crocapeptin A3	8.7'	947.52	964.55

S2: $^{13}C_5{}^{14}N$-Proline feeding of Cm c5

Isotope labeled proline was added to Cm c5 liquid culture, cultivated for 2 more days and extracted as described in the methods section. HR-MS was performed as described, and crocapeptin A ions with m/z increased by 6 units were detected within the crocapeptin A ion peaks only in the fed culture, but not in the non-fed culture.
As an example, figure S2-1B shows part of the ion spectrum of the crocapeptin A2 peak in the range from 952 to 958 units. Additional $[M+H]^+$ ions of crocapeptin A2 can be detected only in the lower panel, corresponding to the extract of supplemented culture. These correspond to the the isotope distribution of the crocapeptin A2 $[M+H]^+$ ions with an increased mass of 6 Da due to the incorporation of $^{13}C_5{}^{14}N$-proline.

Ions with the following m/z values were detected within the respective crocapeptin peaks in extracts from $^{13}C_5{}^{14}N$-proline-fed cultures:

	$[M\text{-}OH]^+$	$[M+H]^+$	$[M+NH_4]^+$
$^{13}C_5{}^{14}N$-crocapeptin A1 ions:	923.49	941.50	957.47
$^{13}C_5{}^{14}N$-crocapeptin A2 ions:	937.51	955.52	971.48
$^{13}C_5{}^{14}N$-crocapeptin A3 ions:	951.52	969.52	986.50

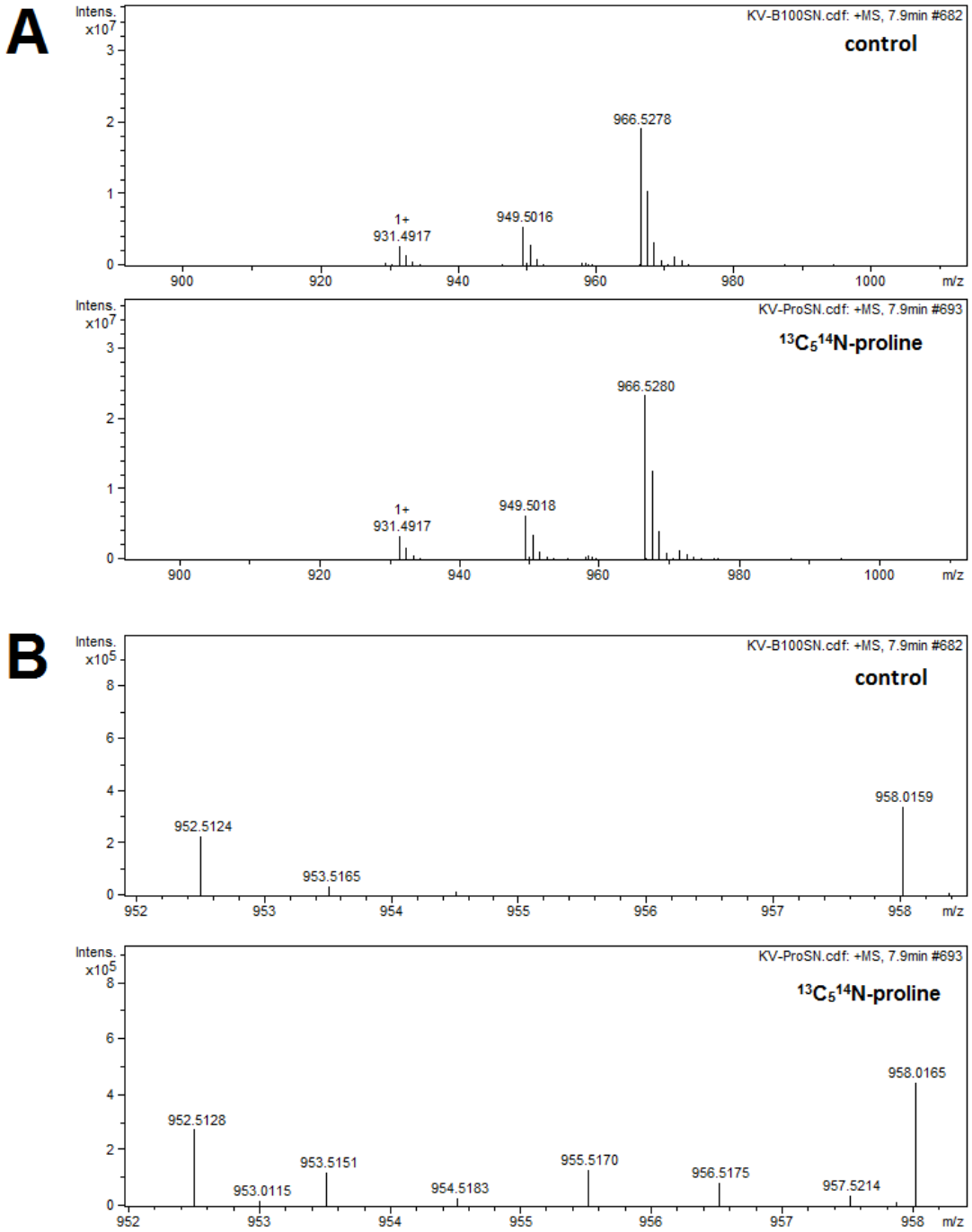


Figure S2-1 A: Mass spectra of crocapeptin A2 ions from Cm c5 supernatant extracts. The lower panel is from Cm c5 liquid culture that had been amended with $^{13}C_5{}^{14}N$-proline, while the upper panel shows a spectrum of a control extract. **B:** Mass spectra of crocapeptin A2 $[M+H]^+$ ions, zoomed into the mass region where additional crocapeptin ions with a mass increase of 6 Da are detected, i.e. $[M+H]^+$ = 955.52 for crocapeptin A2.

S3: Cm c5 mutagenesis

Derivatives of the plasmid pSUP_*hyg* are used for mutagenesis of Cm c5. When a suffiently long homologous region is inserted into pSUP_hyg, the plasmid can integrate into the Cm c5 chromosome at the desired locus via a single crossing over event. The resulting mutant is then hygromycin resistant. By cloning an internal fragment of a gene of interest into the pSUP_*hyg* cloning site, the gene is disrupted by the plasmid backbone as shown in fig. S3-1 A, leading to a knockout mutant. Cloning the 5' region of a gene of interest fused to a promotor leads to a mutant in which the target gene is driven by the introduced promotor, to study overexpression of a gene of interest, as shown in fig. S3-1 B.

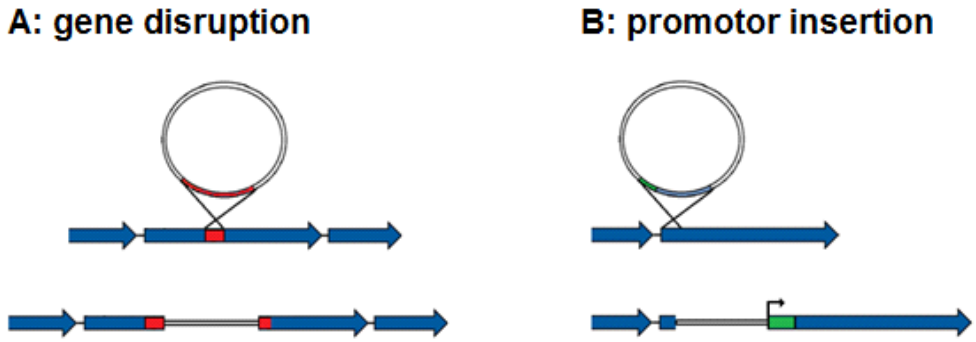


Figure S3-1: Gene inactivation by single-crossing over mutagenesis in Cm c5 with the integrative pSUP_*hyg* plasmid. **A:** The target gene is disrupted by targeted integration of pSUP_*hyg* into the coding region via the homologous fragment shown in red. **B:** A homologous fragment containing the constitutive npt-promotor (green) is used to create a Cm c5 mutant with the target gene under control of the inserted promotor.

Primer lists

Primers used for Inactivation constructs

SerP_for	ACAGGGGAACGTGATCGTCG	Insert for *cpnA*
SerP_rev	TGCGGATCGGTGAGCTTGAA	Insert for *cpnA*
Perm1_for	TTCACGTGCCATCGCGTTTC	Insert for *cpnB*
Perm1_rev	GATCATCGCGAGGCGCTTTC	Insert for *cpnB*
perm2_for	GCACCGGGTCAAGACGAGGA	Insert for *cpnC*
perm2_rev	CGCACTTTCTCGCCGTCCAC	Insert for *cpnC*
crpNRPS2_for	GATCGCCTTCATGCTTTCGGAC	Insert for *cpnD*
crpNRPS2_rev	CTGACCGTCGTAGCTGCTGTTC	Insert for *cpnD*
CypCP2_for	CCAGAAGAACGACACCCAG	Insert for *cpnF*
CypCP_rev	GGACCAGGCGACGTTGAGAT	Insert for *cpnF*
InsertH_for	CCGGTGAACGAAGAAGGGGT	Insert for *cpnG*
InsertH_rev	AGATGCTGACCGGGAGCTTG	Insert for *cpnG*

Primers used for verification of Cm c5 mutants

psup_hyg_fwd	ATGTAGCACCTGAAGTCAGC	pSUP_hyg backbone
psup_hyg_rev	GCATATAGCGCTAGCAGC	pSUP_hyg backbone
serp_veri_for	GCTGCCGCTTTCGTTCAC	verification of *cpnA* mutants
serp_veri_rev	CTCACGTGATCCGGGAAAT	verification of *cpnA* mutants

perm_veri_for	TCACGTTCCTGCTCAATGAA	verification of *cpnB* mutants
perm_veri_rev	CTTCCCGTACTCCATGGTC	verification of *cpnB* mutants
Cveri_for	AGCTCGTCACCCAGGAAGCC	verification of *cpnC* mutants
Cveri_rev	ATGGGCCGGTCTTCGAGCAA	verification of *cpnC* mutants
nrps2_veri_for	GACCTATGGCGAGCTGAAC	verification of *cpnD* mutants
nrps2_veri_rev	AGTAGCGTTCGCAGTGAG	verification of *cpnD* mutants
Gveri_for	GTCGACGAGGCGAAGACCGA	verification of *cpnF* mutants
Gveri_rev	CAGGTGGTGCTCACCGGGAT	verification of *cpnF* mutants
PKcrp2_for	TTCCTCGTACTTGCCGACCG	verification of *cpnG* mutants

Primers for heterologous protein expression in E. coli

exp_francci_ndeI_for	CGATCATATGACGCAGACGCTCGAC	CpnE expression in pET28b
Exp_francci_hindIII_rev	CGATAAGCTTCATGGTCGGGTCTCCG	CpnE expression in pET28b
ExpCyp_for	AGCAGCCCCACCATCGCCGAG	CpnF expression in pSUMO
ExpCyp_rev	CGATGAATTCTCAGGACCAGGCGACGTTGAG	CpnF expression in pSUMO

Primers for promotor insertion experiments

mch71	TGGACAGCAAGCGAACCGG	tn5 promotor from pTOPO *npt* gene
mch72	CATAATCTGTACCTCCTTATCCTGTCTCTTGATCAG	tn5 promotor from pTOPO *npt* gene
crocstart	CAGGATAAGGAGGTACAGATTATGACCAGCATCGCAGGTC	*cpnD* 5' fragment
crocrev	AGTTCGAGGAACGTCAGG	*cpnD* 5' fragment

Verification of Cm c5 mutants by PCR

To genetically verify the genotype of the isolated mutants, gDNA of mutant cells grown in liquid culture was extracted and tested with PCR reactions. The first PCR was performed with the respective verification primers that bind in the vicinity of the region of plasmid insertion. This reaction will only yield a product if no insertion has taken place, as shown in Fig. S3-2. In the mutant, the target gene is disrupted by the 8 kb plasmid backbone, thus no PCR product is formed.

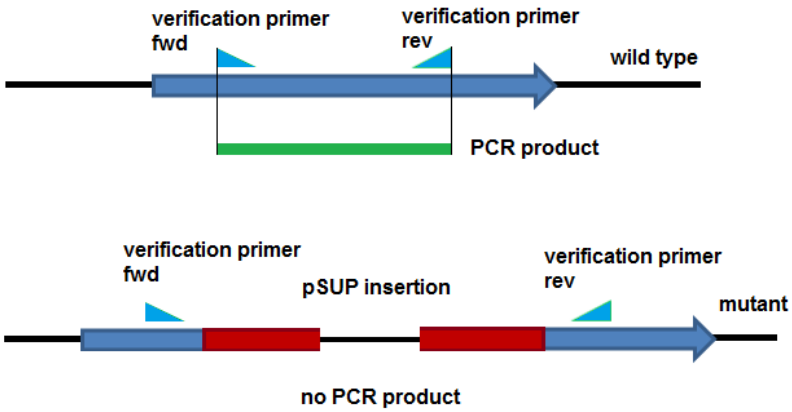


Figure S3-2: Genomic organization of target genes in the Cmc5 inactivation mutants. Blue bar/arrow: 5' and 3' regions of target gene inactivated by insertional mutagenesis using the respective pSUP_*hyg* derivatives. Light blue arrowheads: Position of primers used in a control amplification in order to verify disruption of the wild type allele of the target gene in the mutants.

The second PCR was done using one of the respective verification primers combined with another primer that binds the plasmid backbone. The correct combination of these primers leads to a product only if the pSUP_hyg plasmid backbone has indeed integrated at the desired position in the Cmc5 genome, as shown in Fig. S3-3.

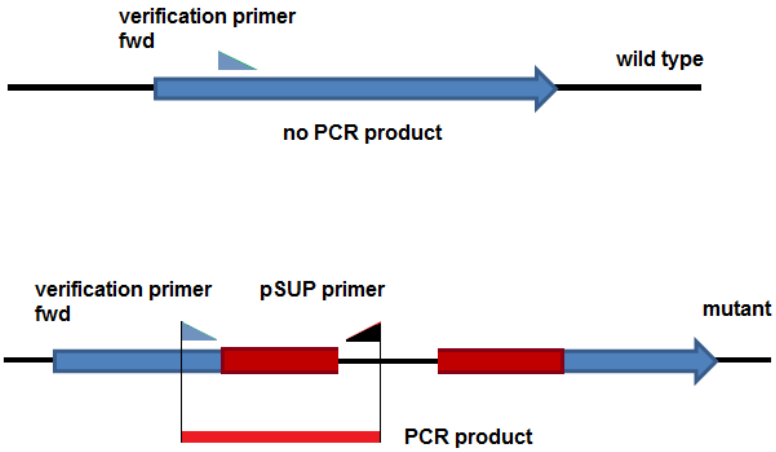


Figure S3-3: Genetic verification of Cmc5 mutants by PCR Arrowheads: Position of primers used in a control amplification (blue: primers in the genomic up- and downstream regions; black: primer in the pSUP_hyg backbone) in order to verify disruption of the wild type allele of the target gene in the mutants.

Mutants matching these two requirements were thus genetically verified and subjected to extraction. As shown in Fig. S3-4, the gDNA isolated from different mutant cells are not yielding the PCR product that corresponds to the intact gene anymore. As a positive control for this reaction, gDNA from cm c5 wild type and mutant was used as template, yielding a product because there is either no pSUP_hyg insertion or an insertion at a different locus.

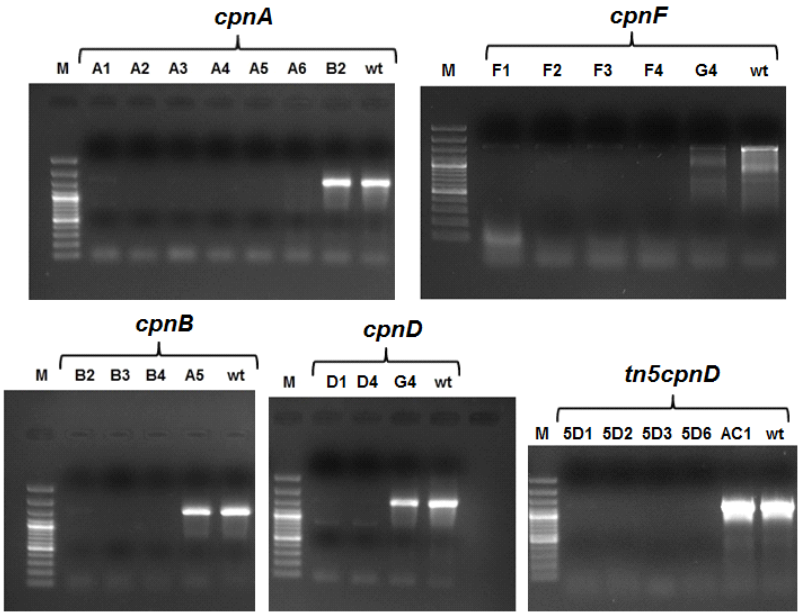


Figure S3-4: Testing for absence of the wild type allele of target genes in inactivation mutants. Photographs of agarose gels of PCR reactions to test for disruption of the target genes *cpnA*, *cpnB*, *cpnD*, *cpnF* and the promotor insertion of tn5 upstream of *cpnD* (*tn5cpnD*) in Cm c5 are shown. The respective intact gene will give a productof defined size, whereas a disrupted gene harboring the 8 kbp pSUP_hyg-insertion will not yield a PCR product. M: 100 bp+ DNA ladder; A1-G5: gDNA template from Cm c5 mutants; wt: gDNA template from Cmc5 wild type. All mutants of genes *cpnA*, *cpnB*, *cpnD*, *cpnF* and *tn5cpnD* that were tested yield no PCR product, indicating that the gene is disrupted. As positive control, wt gDNA and gDNA from a different mutant were used in each reaction.

Fig S3-5 shows the PCR to prove insertion of the pSUP_hyg backbone at the desired locus. In this reaction, only the mutants shall yield a product as one of the primers binds to a locus that is only present on the pSUP_hyg plasmid backbone and does not hybridize with wild type gDNA.

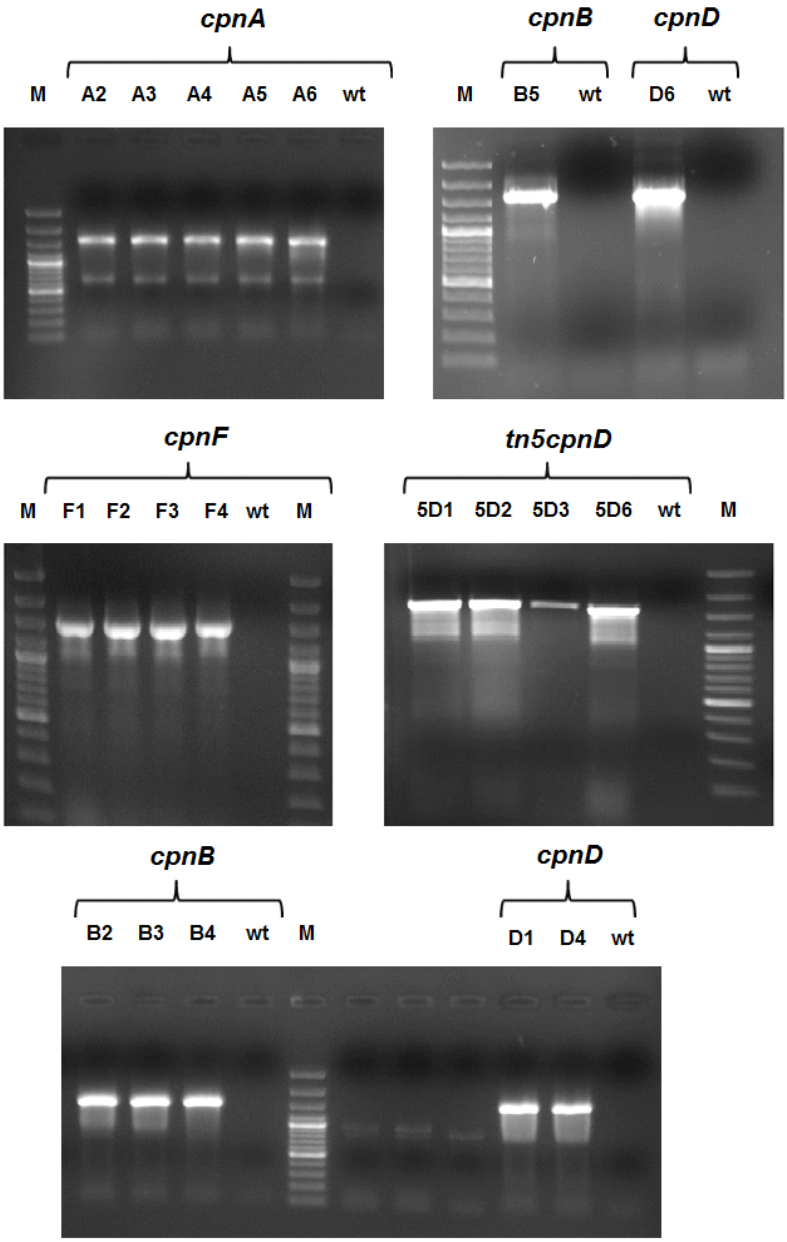


Figure S3-5: Testing for correct insertion of the pSUP_hyg derivatives into target genes in inaction mutants. Photographs of agarose gels of PCR reactions to test for insertion of pSUP_*hyg* derivatives into the target genes of Cm c5 are shown. Depending on the orientation of the homologous DNA region in the pSUP_hyg construct, primer combination fwd or rev will yield a product. Wild type gDNA was used as a negative control. M: 100 bp+ DNA ladder; All mutants of genes that were tested yield the expected PCR product for one primer combination, indicating that the gene of interest has indeed been disrupted by insertion the pSUP_hyg construct.

S3.3 MS-Analysis of Cm c5 mutants

All genetically verified mutants were cultivated and extracted as described in the methods section. Extracts were made from cells and medium supernatant separately, and the peak area of the three crocapeptin A derivatives and chondramide B (the chondramide derivative produced by Cm c5) was estimated using the peak integration function of the "Bruker Data Analysis" software. To define relative crocapeptin production in each culture, the ratio of crocapeptins to chondramide was calculated.

cells	crocapeptin A1	crocapeptin A2	crocapeptin A3	chondramide B	total crocapeptin	ratio
wt 1	1176008	6917388	1049817	25800651	9143213	0.35
wt 2	1185164	4102309	659598	32734554	5947071	0.18
wt 3	1528591	4935460	1054948	26426109	7518999	0.28
wt 4	0	58181591	11513076	69694667	116422521	0.6
wt 5	4055024	43284918	7122786	54462728	109381295	0.5
wt 6	1656742	8593171	1449611	200150084	11699524	0.06
wt 7	1168430	4837227	879491	222780968	6885148	0.03
cpnA 1	0	9385303	2842886	36764657	12228189	0.33
cpnA 2	0	2681036	168872	3003858	2849908	0.95
cpnA 3	0	2216397	482971	11943172	2699368	0.23
cpnA 4	5535711	106438701	25458054	137432466	125626608	1.09
cpnA 5	2740846	114328947	12095080	129164873	110700853	1.17
cpnA 6	5607378	105264340	28804671	139676389	164756143	0.85
cpnB 1	0	0	0	7206406	0	0
cpnB 2	0	0	0	30849669	0	0
cpnB 3	0	0	0	159048223	0	0
cpnB 4	0	0	0	173923847	0	0
cpnB 5	0	0	0	112307657	0	0
cpnB 6	0	0	0	108427764	0	0
cpnC 1	0	0	0	31880059	0	0
cpnC 2	0	0	0	66250865	0	0
cpnC 3	0	0	0	110427338	0	0
cpnD 1	0	0	0	12654816	0	0
cpnD 2	0	0	0	2195804	0	0
cpnD 3	0	0	0	6151685	0	0
cpnF 1	0	8892938	1883591	115882370	10776529	0.09
cpnF 2	0	2020769	459811	80466091	2480580	0.03
cpnF 3	0	14479782	2974752	106291302	17454534	0.16
tn5D 1	139689124	291775458	39305462	156346772	470770044	3.01
tn5D 2	180984061	357698203	50569708	127152319	589251972	4.63
medium	**crocapeptin A1**	**crocapeptin A2**	**crocapeptin A3**	**chondramide B**	**total crocapeptin**	**ratio**
wt 1	6362466	36752288	6741558	7046709	49856312	7.08
wt 2	3420604	14830502	3013166	6883095	21264272	3.09
wt 3	7066827	31022947	7268432	8383996	45358206	5.41
wt 4	1518878	31369365	6895529	39783772	11305575	3.52
wt 5	1150234	20677051	5163674	26990959	8846484	3.05

wt 6	6146706	39785210	9961776	36036359	55893692	1.55
wt 7	1480359	8458656	1662484	14258353	11601499	0.81
cpnA 1	356705	17016139	5406714	2636350	22779558	8.64
cpnA 2	0	20519191	2768159	21853065	23287350	1.07
cpnA 3	837072	7687928	2318865	3363743	10843865	3.22
cpnA 4	5796610	117835561	31331847	154964018	11499663	13.48
cpnA 5	5227761	29418990	10249984	44896735	3847294	11.67
cpnA 6	5321344	122211981	34015617	161548942	10618287	15.21
cpnB 1	261784	7475274	2856842	3916348	10593900	2.71
cpnB 2	157823	9039195	2978824	1265602	12175842	9.62
cpnB 3	3010313	45139282	11565697	59715292	12697673	4.7
cpnB 4	4728448	65296389	16842106	86866943	15220120	5.71
cpnB 5	2324142	56808586	13444426	72577154	19085208	3.8
cpnB 6	1616669	39767907	8942617	50327193	15309691	3.29
cpnC 1	3713628	5954562	15232630	9571691	24900820	2.6
cpnC 2	6446015	9196237	13654726	13349584	29296978	2.2
cpnC 3	11949610	14122702	15574499	23512599	41646811	1.77
cpnD 1	0	0	0	6340237	0	0
cpnD 2	0	0	0	8572431	0	0
cpnD 3	0	0	0	6596406	0	0
cpnF 1	1194655	12625346	3452523	10530998	17272524	1.64
cpnF 2	874477	9178559	2404861	6893920	12457897	1.81
cpnF 3	320350	3144147	715559	5912588	4180056	0.71
tn5D 1	529732434	921450006	297688871	33748625	1748871311	51.82
tn5D 2	513994595	874193881	270175487	21846031	1658363963	75.91

S4: NMR Experimental Data

Table S4-1. Spectroscopic Data of Crocapeptin A_2 (**2**), (600 MHz, DMSO-d_6)

unit	pos	C, mult.	H, (J in Hz)	COSY	HMBC	ROESY
Val	1	171.9, qC				
	2	55.5, CH	4.71, dd (9.5;4.4)	NH	1,3,4,5,*N*MeTyr 1	3,4,5,NH
	3	30.9, CH	2.06, m	2,4,5	2,4,5	2,4,5
	4	19.3, CH_3	0.86, d (6.6)	3	2,3,5	3
	5	17.1, CH_3	0.72, d (6.6)	3	2,3,4	3
	NH		7.39, d (9.5)	2	2,*N*MeTyr1	2,3,5, *N*MeTyrNMe,*N*MeTyr2
*N*MeTyr	1	169.1, qC				
	2	60.9, CH	4,92, dd (11.7;2.2)	3a,3b	1, 3	3a,5/9,Phe2,ValNH
	3a	33.6, CH_2	3,11, dd (14.7;2.2)	2,3b	2,4,5/9	2,3b,5/9
	3b		2,72, dd (14.7;11.7)	2,3a	2,4,5/9	3a,5/9
	4	127.4, qC				
	5/9	130.4, CH	7.00, d (8.8)	6/8	3,5/9,6/8 ,7	2,3a,3b,6/8,Phe2
	6/8	115.3, CH	6,78, d (8.8)	5/9	4,6/8,7	5/9
	7	156.2, qC				
	N-Me	30.3, CH_3	2,76, s		2,Phe1	5/9,Val4,Val5,ValNH
	7-OH		9,36, brs			
Phe	1	170.3, qC				
	2	50.2, CH	4,76, dd (11.7; 4.4)	3a,3b	1, 3,Ahp2,Ahp6	3b,5/9,*N*MeTyr2,*N*MeTyr5/9
	3a	35.3, CH_2	2,86, dd (13.9, 11.7)	2,3b	2,4,5/9,	3b 5/9,Ahp6
	3b		1,80, dd (13.9, 4.4)	2,3a	4,5/9	2,3a,5/9
	4	136.7, qC				
	5/9	129.4, CH	6,83, d (7.2)	6/8	3,5/9, 7	2,3a,3b,6/8, Ahp5b,Ahp6
	6/8	127.7, CH	7,18, dd (7.2, 7.2)	5/9, 7	4,6/8	5/9,7
	7	126.2, CH	7.13, t (7.2)	6/8	5/9	6/8
Ahp	2	168.9, qC				
	3	48.5, CH	3,63, m	NH,4a,4b	2,4	NH,4b,Phe5/9
	4a	21.3, CH_2	2,41, m	3,4b, 5a,5b		NH,4b
	4b		1.57, m	3,4a,5a,5b	5	3,4a
	5a	29.3, CH_2	1,69, m	4a,4b,5b,6	4	5b,6
	5b		1,55, m	4a,4b,5a,6	4,6	5a,6
	6	74.1, CH	5,06, brs	6-OH, 4,5b	2,4	5a,5b,6-OH, Phe2,Phe3a,Phe5/9
	NH		7,06, d (9.5)	3	3,Leu1	3,4a,LeuNH,Thr3,Leu2
	6-OH		6.00, brs	6		4a,5a,6,Leu6
Leu	1	170.1, qC				
	2	50.2, CH	4,21, dd (9.5,8.8)	NH,3a,3b		3a, 6,AhpNH
	3a	39.0, CH_2	1,70, m	2,3b	2,4,6	3b
	3b		1,30, m	2,3a	2,6	3a
	4	24.1, CH	1,43, m	3a,5,6		5,6
	5	23.2, CH_3	0,83, d (6.6)	4	3,4,6	4,6
	6	20.9, CH_3	0,73, d (6.6)	4	3,4,5	4,5
	NH		8,37, d (8.8)	2	Thr1,Leu2	3b,4,Thr2,Thr3
Thr	1	169.1, qC				
	2	54.5, CH	4,56, d (9.5)	NH,3	Thr1,Gln1	NH,3,4,LeuNH
	3	71.8, CH	5,39, m	2,4	1,4,Val1	2,4,AphNH,LeuNH
	4	17.6, CH_3	1,17, d (6.6)	3	2,3	2,3
	NH		7,74, d (9.5)	2	Gln1	2,4,Gln2
Gln	1	172.4, qC				
	2	52.0, CH	4,34, td (8.1, 5.1)	NH,3a,3b	1,3,Ibu1	NH,3a,3b,4,ThrNH
	3a	27.3, CH_2	1,91, m	2,3b,4	1,2,4,5	2,3b,4
	3b		1,73, m	2,3a,4	1,2,4,5	2,3a,4
	4	30.8, CH_2	2,11, td (9.7, 6.2)	3a,3b	2,3,5	2
	5	173.8, qC				
	NH		7.99, d (8.1)	2	2,Ibu1	2,3b,4,Ibu2
	NH2a		7,23, brs	NH2b	5	4
	NH2b		6,72, brs	NH2a	4,5	4
Ibu	1	176.3, qC				
	2	33.7, CH	2,46, m	3/4	1,3/4	3/4
	3/4	19.5, CH_3	1,03, t (6.6)	2	1,2,3/4	2

Table S4-2. Spectroscopic Data for Crocapeptin B (**4**), (600 MHz, DMSO-d_6)

unit	pos	δC, mult.	δH, (*J* in Hz)	COSY	HMBC	ROESY
Ile	1	172.1, qC				
	2	55.1, CH	4.76, m	NH,3	1,3,4	NH,4a,4b,5
	3	37.6, CH	1,81, m	2,3,4b		NH, 4a,6
	4a	24.3, CH_2	1.21, m	3,4b	Ile5,Ile6	3,4b,5
	4b		0.99, m	3,4a,5	Ile5,Ile6	4a,5
	5	11.4, CH_3	0.82, t (7.3)	4a,4b	3,4	3,4a,4b
	6	16.1, CH_3	0.84, t (7.0)	3	2,3,4	2,3,NMeTyrNMe
	NH		7.44, d (9.5)	2	NMeTyr1	2,3,4a,4b,6,NMeTyr2,NMeTyr3b
*N*MeTyr	1	169.0, qC				
	2	60.8, CH	4.91, dd (11.0, 2.2)	3a,3b	1,NMe	3a,3b,5/9,Phe2,IleNH
	3a	32.7, CH_2	3.11, dd (14.0, 2.2)	2,3b	2,4,5/9	2,3b,5/9
			2.71, dd (14.0, 11.0)	2,3a	4,5/9	3a,5/9
	4	127.4, qC				
	5/9	130.3, CH	7.00, d (8.1)	6/8	3,5/9,7	2,3a,3b,6/8,Phe2
	6/8	115.3, CH	6.78, d (8.1)	5/9	4,6/8,7	5/9
	7	156.2, qC				
	N-Me	30.2, CH_3	2.75, s		2,Phe1	5/9,IleNH,Ile4b,Ile5
	OH		9.33, brs			
Phe	1	170.3, qC				
	2	50.3, CH	4.76, m	3a,3b	1,Ahp2,Ahp6	3a,5/9,NMeTyr2,NMeTyr5/9
	3a	35.3, CH_2	2.86, dd (14.3, 12.1)	2,3b	4,5/9	2,3b,5/9,Ahp6
	3b		1.81, dd (14.3, 4.2)	2,3a	4,5/9	3a,5/9
	4	136.7, qC				
	5/9	129.3, CH	6.83, d (7.3)	6/8,7	3,5/9,7	2,3a,3b,6/8,Ahp5b,Ahp6
	6/8	127.7, CH	7.18, dd (7.3, 7.3)	5/9,7	4,6/8	5/9
	7	126.1, CH	7.14, t (7.3)	6/8	5/9	6/8
Ahp	2	168.9, CH				
	3	48.6, CH	3.64, m	NH,4a,4b	2,4	NH,4b,6,Phe5/9
	4a	21.6, CH_2	2.41, m	3,4b,5a,5b		NH,4b,6
	4b		1.57, m	3,4a,5a		3,4a
	5a	29.1, CH_2	1.68, m	4a,5b,6		4a,6,OH
	5b		1.55, m	4a,5a,6		5a,Phe3a,Phe5/9
	6	73.7, CH	5.05, brs	OH,5a	2	3,4a,5a,OH,Phe3a,Phe5/9
	NH		7.07, d (8.8)	3	Leu1	3,4a,LeuNH,Leu2,Thr3
	OH		6.00, d (2.9)	6		4a,5a,6,IleNH,Ile4a,Ile4b,Ile5
Leu	1	170.1, qC				
	2	50.2, CH	4.20, m	NH,3a,3b		3a,6,AhpNH
	3a	39.1, CH_2	1.71, m	2,3b,4		2,3b,6
	3b		1.30, m	2,3a	5,6	3a,4
	4	24.1, CH	1.44, m	3a,5,6		5,6
	5	23.2, CH_3	0.84, t (7.0)	4	3,4,6	4,6
	6	20.8, CH_3	0.72, d (6.6)	4	3,4,5	2,3a,4,5
	NH		8.33, d (8.8)	2	2,Thr1	3b,4,Thr2,Thr3
Thr	1	169.0, qC				
	2	54.4, CH	4.57, d (9.5)	NH,3	1,Cit1	NH,3,4,LeuNH,Leu4,Cit5
	3	71.9, CH	5.37, m	2,4	4,Ile1	2,4,AhpNH,Citγ NH,LeuNH
	4	17.6, CH_3	1.16, d (6.6)	3	2,3	2,3
	NH		7.73, d (9.5)	2	Cit1	2,4,Cit2,Cit3a
Cit	1	172.7, qC				
	2	52.0, CH	4.41, td (8.8, 5.1)	NH,3a,3b	1,3,Ibu1	NH,3a,3b,4,5,ThrNH,Ibu3/4
	3a	28.8, CH_2	1.69, m	2,3b,4	1,2,4,5	2,3b
	3b		1.50, m	2,3a,4	1,2,4,5	NH,2,3a
	4	26.6, CH_2	1.35, m	3a,3b,5	2,3,5	2,3a,5
	5	38.8, CH_2	3.02-2.89, m	4,γNH	3,4,6	γNH,2,4
	6	158.8, qC				
	NH		7.95, d (8.1)	2	2,Ibu1	2,3b,4,Ibu2,Ibu3/4
	γNH		5.89, t (5.9)	5	6	4,5,Thr3
Ibu	1	176.2, qC				
	2	33.6, CH	2.47, m	3/4	1,3/4	3/4,Cit2
	3/4	19.4, CH_3	1.01, dd (6.6, 5.1)	2	1,2,3/4	2

Table S4-3. Spectroscopic Data of Crocapeptin A_1 (**1**) and Crocapeptin A_3 (**3**) (600 MHz, DMSO-d_6)

unit	pos	δC, mult. (**1**)	δH, (*J* in Hz) (**1**)	δC, mult.(**3**)	δH, (*J* in Hz) (**3**)
Val	1	172.6		171.9	
	2	55.7	4.72, dd (9.5, 4.4)	55.6	4.71, dd (9.5, 4.4)
	3	30.8	2.07, m	30.8	2.08, m
	4	19.2	0.87, d (6.6)	19.2	0.86, d (6.6)
	5	14.0	0.73, d (6.6)	17.1	0.72, d (7.3)
	NH		7.40, d (9.5)		7.39, d (9.5)
*N*MeTyr	1	169.1		169.1	
	2	60.9	4.92, dd (11.0, 2.2)	60.9	4.92, dd (11.7, 2.2)
	3a	32.8	3.10, dd (14.7, 2.2)	32.8	3.11, dd (14.4, 2.2)
	3b		2.72, dd (14.7, 11.0)		2.72, dd (14.4, 11.7)
	4	127.4		127.4	
	5/9	130.4	7.00, d (8.1)	130.4	7.00, d (8.1)
	6/8	115.3	6.78, d (8.1)	115.3	6.78, d (8.1)
	7	156.6		156.2	
	N-Me		2.76, s		2.76, s
Phe	1	170.4		170.3	
	2	50.2	4.76, dd (11.7, 4.0)	50.2	4.76, dd 11.7, 4.4)
	3a	35.3	2.86, dd (13.2, 11.7)	35.3	2.86, dd (13.9, 11.7)
	3b		1.80, dd (13.2, 4.0)		1.79, dd (13.9, 4.4)
	4	136.7		136.7	
	5/9	129.4	6.83, d (6.6)	129.4	6.83, d (7.3)
	6/8	127.7	7.18, dd (7.0, 6.6)	127.7	7.18, dd (7.3, 7.2)
	7	126.1	7.13, t (7.0)	126.1	7.13, t (7.2)
Ahp	2	168.9		168.9	
	3	48.5	3.63, m	48.6	3.62, m
	4a	30.4	1.70, m	29.2	1.70, m
	4b		1.55, m		1.56, m
	5a	21.6	2.41, m	21.6	2.40, m
	5b		1.56, m		1.56, m
	6	73.7	5.06, brs	73.7	5.06, brs
	NH		7.07, d (9.5)		7.06, d (8.8)
	OH		6.00, brs		6.00, brs
Leu	1	169.8		170.1	
	2	50.2	4.21, dd (8.8, 8.8)	50.2	4.21, m
	3a	40.0	1.70, m	39.0	1.70, m
	3b		1.30, m		1.30, m
	4	24.0	1.45, m	24.1	1.44, m
	5	23.2	0.84, d (6.6)	23.2	0.84, d (6.6)
	6	20.8	0.73, d (6.6)	20.9	0.73, d (6.6)
	NH		8.36, d (8.8)		8.38, d (8.8)
Thr	1	169.1		169.1	
	2	54.6	4.55, d (9.5)	54.5	4.56, d (9.5)
	3	71.8	5.39, q (6.6)	71.8	5.39, q (6.6)
	4	17.6	1.17, m	17.6	1.17, d (6.6)
	NH		7.86, d (9.5)		7.75, d (9.5)
Gln	1	173.0		172.3	
	2	52.0	4.37, td (8.1, 5.5)	52.0	4.36, td (8.4, 5.1)
	3a	27.6	1.87, m	26.7	1.89, m
	3b		1.70, m		1.71, m
	4	31.4	2.14, m	31.5	2.12, m
	5	173.8		173.8	
	NH		7.99, d (8.1)		8.03, d (8.1)
	γNHa		7.22, brs		7.23, brs
	γNHb		6.71, brs		6.71, brs
Starter	1	175.9		175.7	
	2	28.2	2.11, m	41.0	2.26, tq (7.3, 6.6)
	3	9.8	1.01, t (7.6)	27.3	1.52, m
					1.32, m
	4			11.8	0.82, t (7.3)
	5			17.5	0.99, d (6.6)

Table S4-4. Spectroscopic Data for Pre-Crocapeptin A_2 (**5**), (700 MHz, MeOH-d_4)

unit	pos	δC, mult.	δH, (*J* in Hz)	COSY	HMBC
Val	1	168.6			
	2	55.4	4.78, m	3, NH	1, 3, 4, 5
	3	28.1	2.43, m	2, 4, 5	4, 5
	4	17.4	0.96, d (7.0)	3	2, 3, 5
	5	15.4	0.90, d (6.8)	3	2, 3, 4
	NH		8.97, d (9.9)	2	Phe1
NMeTyr	1	169.6			
	2	62.1	5.29, dd (11.5, 3.5)	3a, 3b	1, 3, 4, 8, Phe1
	3	31.5	3.50, m	2, 3b	2, 4, 5/9
			2.93, m	2, 3a	1, 2, 4, 5/9
	4	127.5			
	5/9	129.2	7.09, d (8.4)	3a, 6/8	5/9, 7
	6/8	114.2	6.50, d (8.4)	5/9	4, 6/8, 7
	7	154.9			
	8	28.0	2.94, s		2, Phe1
Phe	1	173.2			
	2	47.2	5.21, dd (12.0, 3.0)	3a, 3b	1, 3, 4
	3	32.9	2.40, m	2, 3b	1, 2, 4, 5/9
			1.32, m	2, 3a	4, 5/9
	4	135.3			
	5/9	127.1	7.29, d (7.5)	6/8, 3a, 3b	3, 5/9, 7
	6/8	126.8	7.23, dd (7.5, 7.5)	5/9, 7	4, 6/8
	7	124.7	7.15, t (7.5)	6/8	5/9
Pro	1	170.8			
	2	57.9	4.28, dd (8.5,3.0)	3a, 3b	3, 4, 5
	3	30.2	2.25, m	2, 3b, 4	1
				2, 3a, 4	
	4	20.9	1.80, m	3a, 3b, 5a, 5b	2, 3
	5	45.2	3.49 ,m	4, 5b	2, 3, 4
			3.26, dt (11.3,8.0)	4, 5a	2, 3, 4
Leu	1	171.8			
	2	47.3	4.55, d (10.6)	NH, 3a, 3b	1, 3, 4, Thr1
	3	41.0	1.53, m	2, 3b, 4	2 ,4 , 5, 6
			1.20, m	2, 3a, 4	4, 5, 6
	4	23.1	1.50 (m)	3a, 3b, 5, 6	5, 6
	5	21.2	0.86, d (6.4)	(3b), 4	3, 4, 6
	6	19.9	0.69, d (6.4)	(3b), 4	3, 4, 5
	NH		7.96, d (9.4)	2	Thr1
Thr	1				
	2	53.9	4.78, m	3	1, Gln1
	3	71.4	5.34, dd (6.2,2.8)	2, 4	4, Val1
	4	15.3	1.32, m	3	2, 3
Gln	1	172.2			
	2	50.9	4.67, dd (9.4,4.9)	3a, 3b	1, 3, 4, Ibu1
	3	25.1	2.32, m	2, 3b, 4	2, 4
			2.05, m	2, 3a, 4	2, 5
	4	29.7	2.40, m	3a, 3b	2, 3, 5
	5	175.1			
Ibu	1	177.8			
	2	33.6	2.58, sept. (7.0)	3/4	1, 3/4
	3/4	17.5	1.18, m	2	1, 2, 3/4

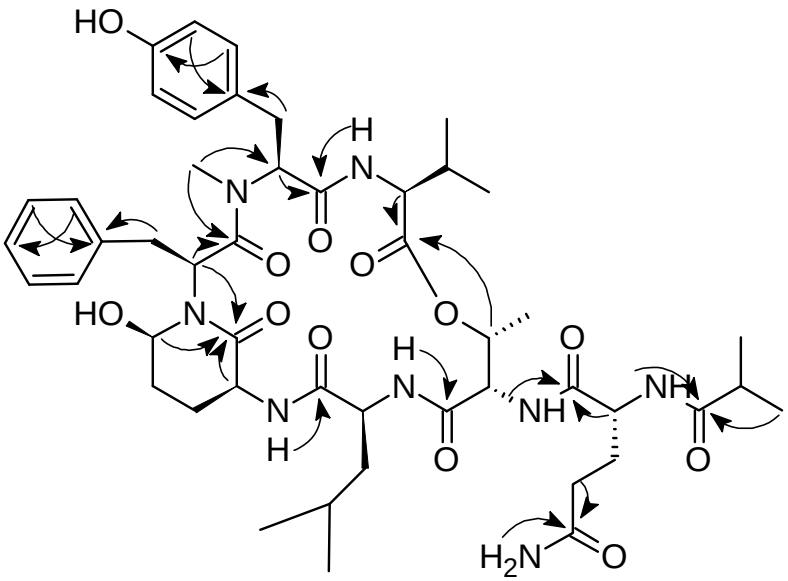


Figure S4-1. Selected HMBC correlations establishing the structure of crocapeptin A_2 (**2**).

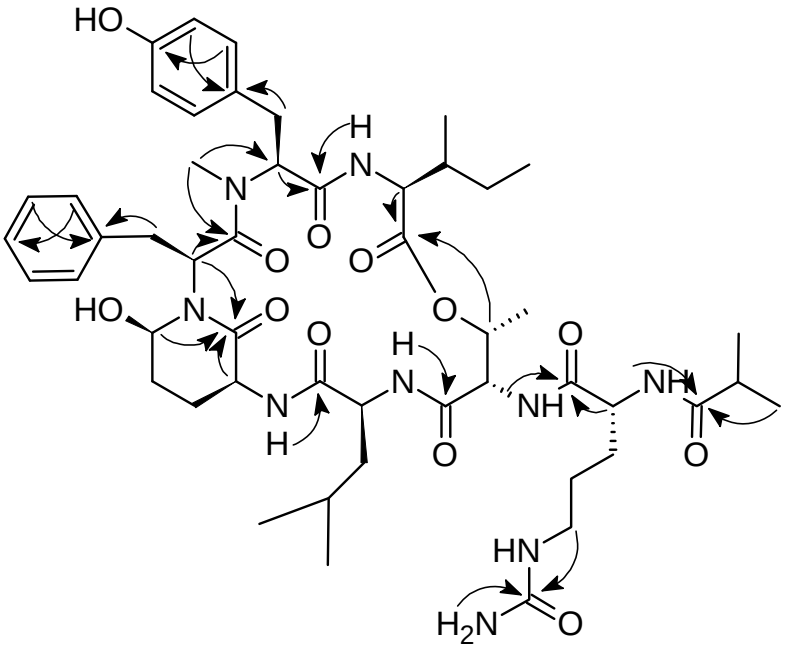


Figure S4-2. Selected HMBC correlations establishing the structure of crocapeptin B (**4**).

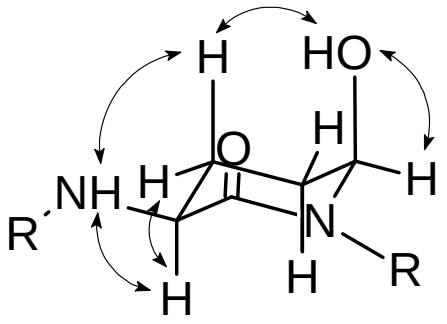
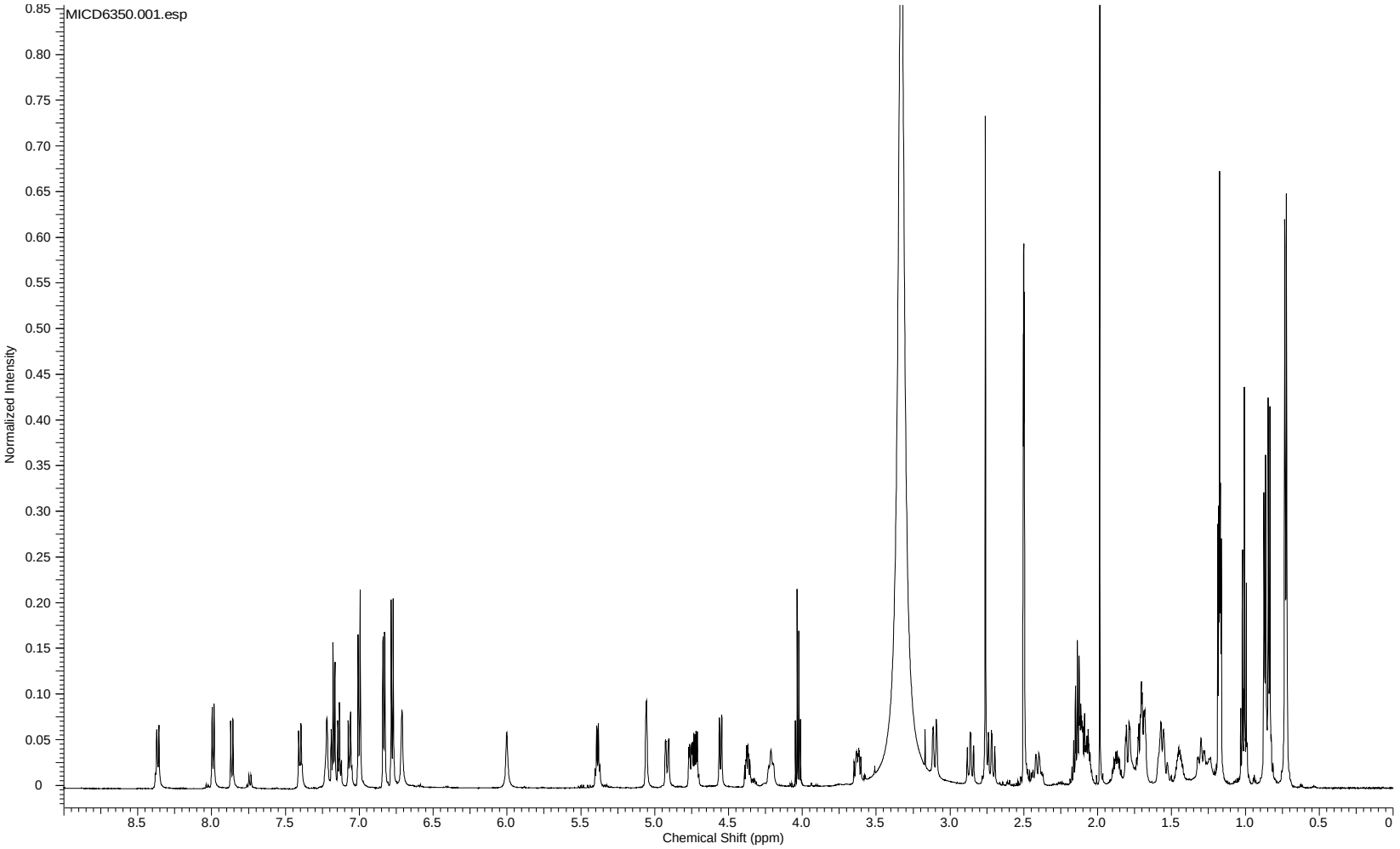
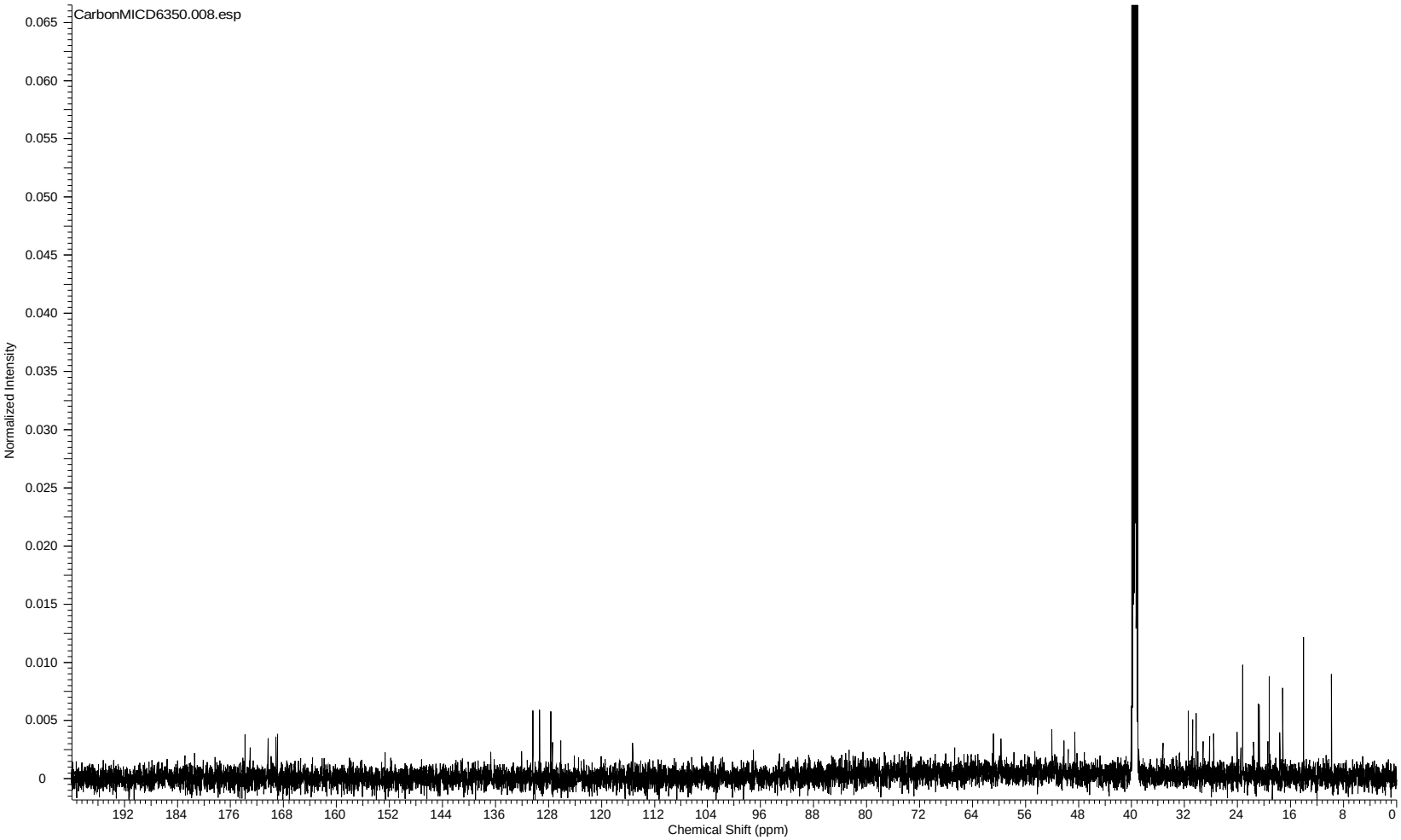


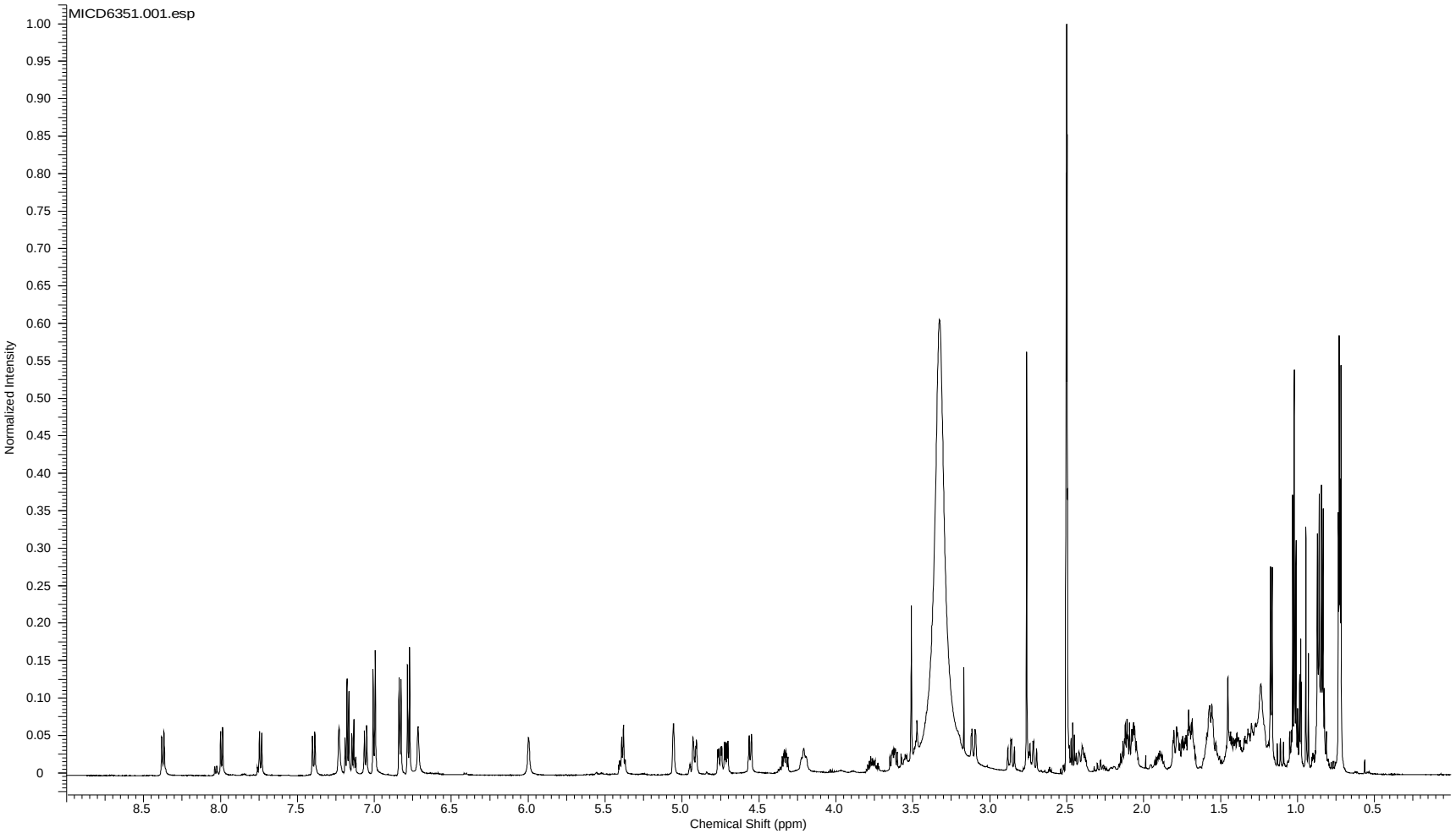
Figure S4-3. Selected ROESY correlations within Ahp moiety .



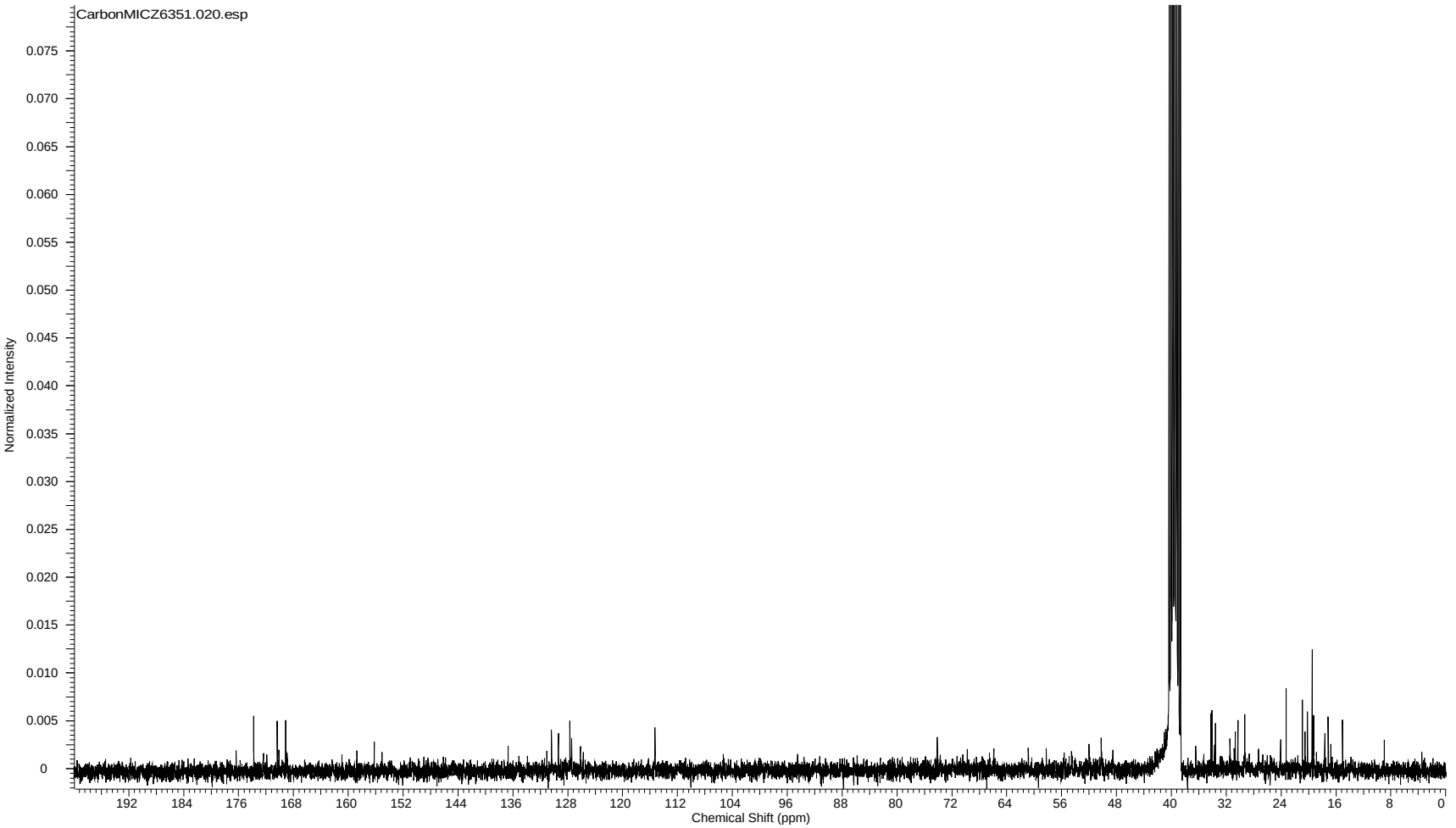
^{1}H NMR spectrum of crocapeptin A_1 (**1**) in DMSO-d_6



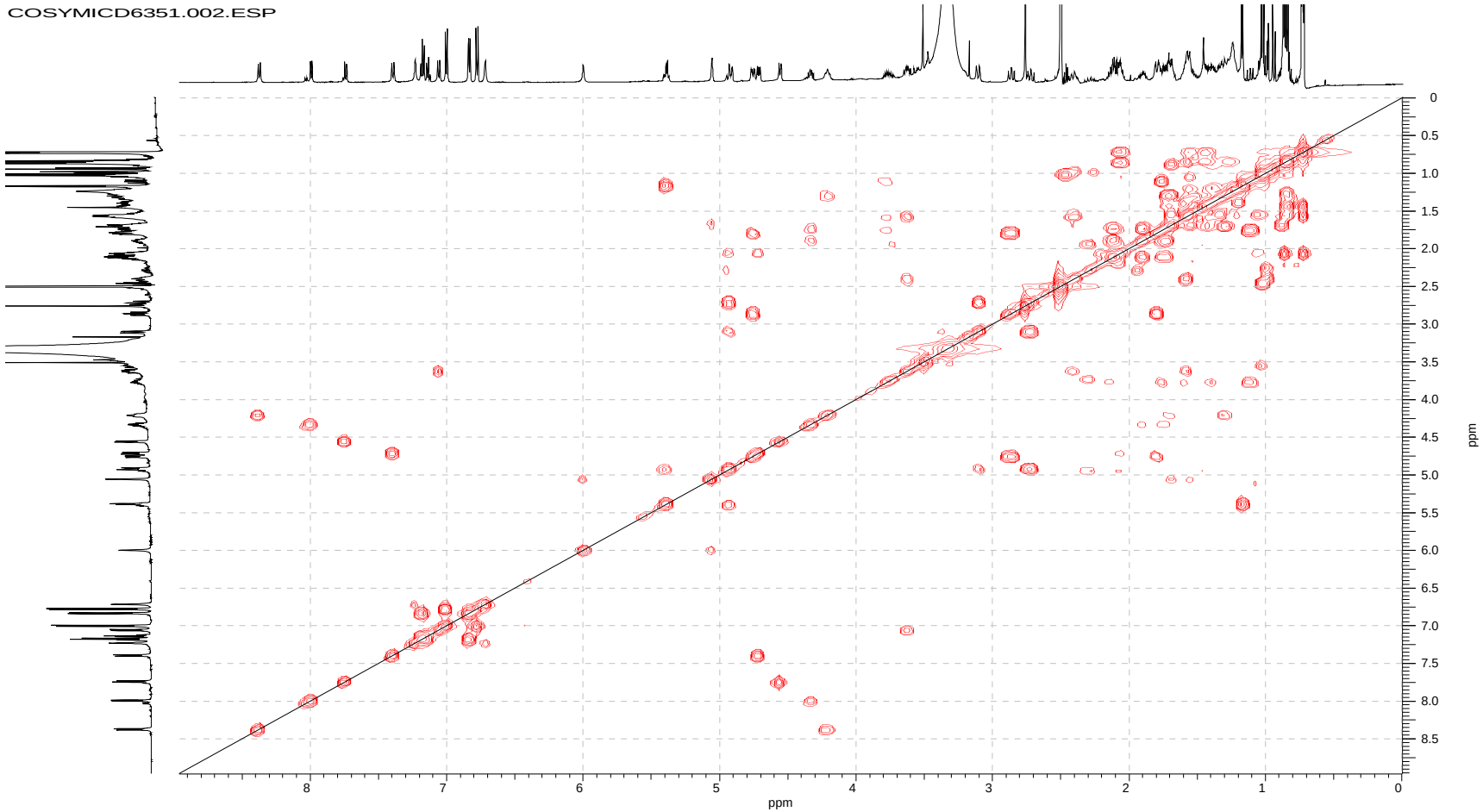
^{13}C NMR spectrum of crocapeptin A_1 (**1**) in DMSO-d_6



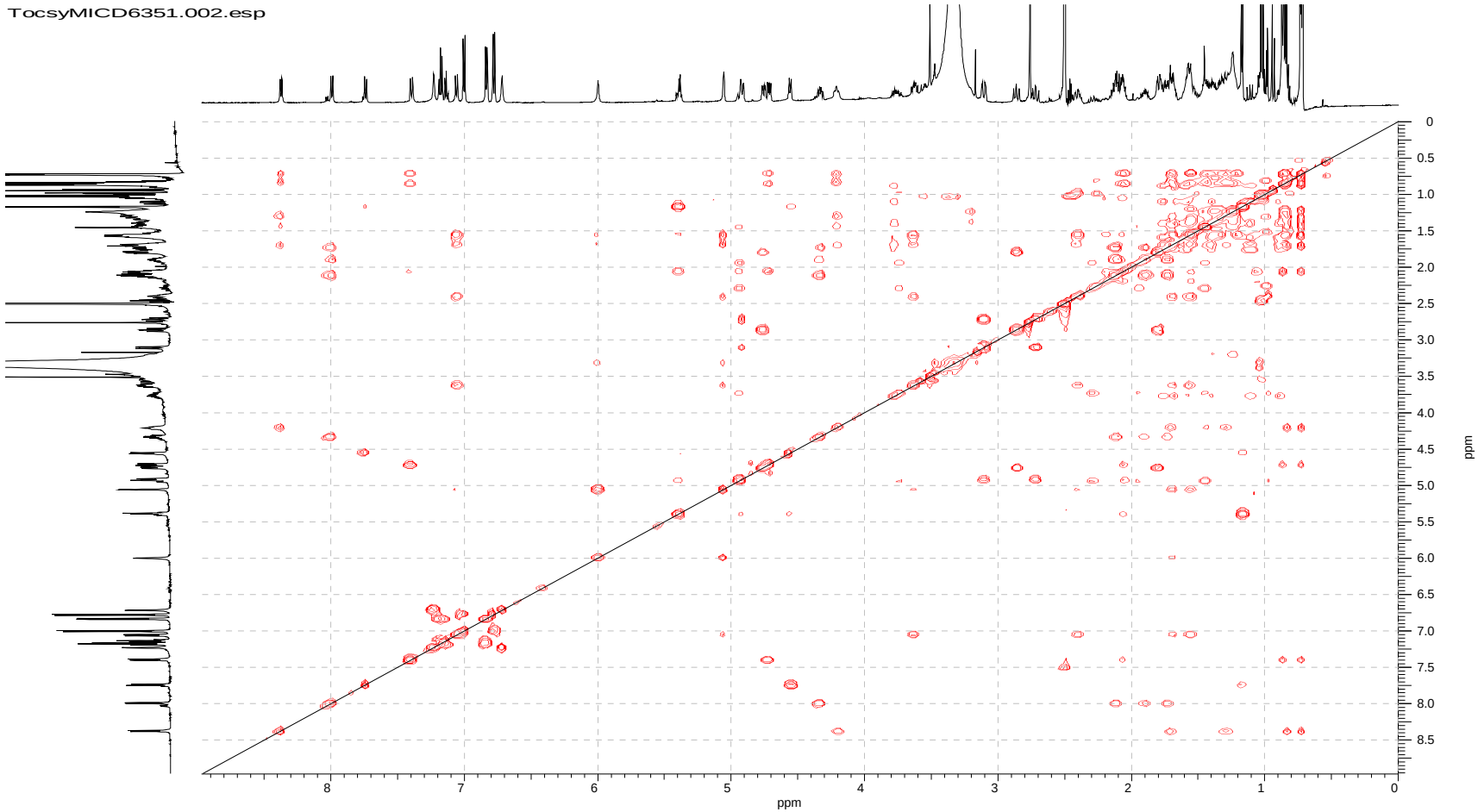
^{1}H NMR spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



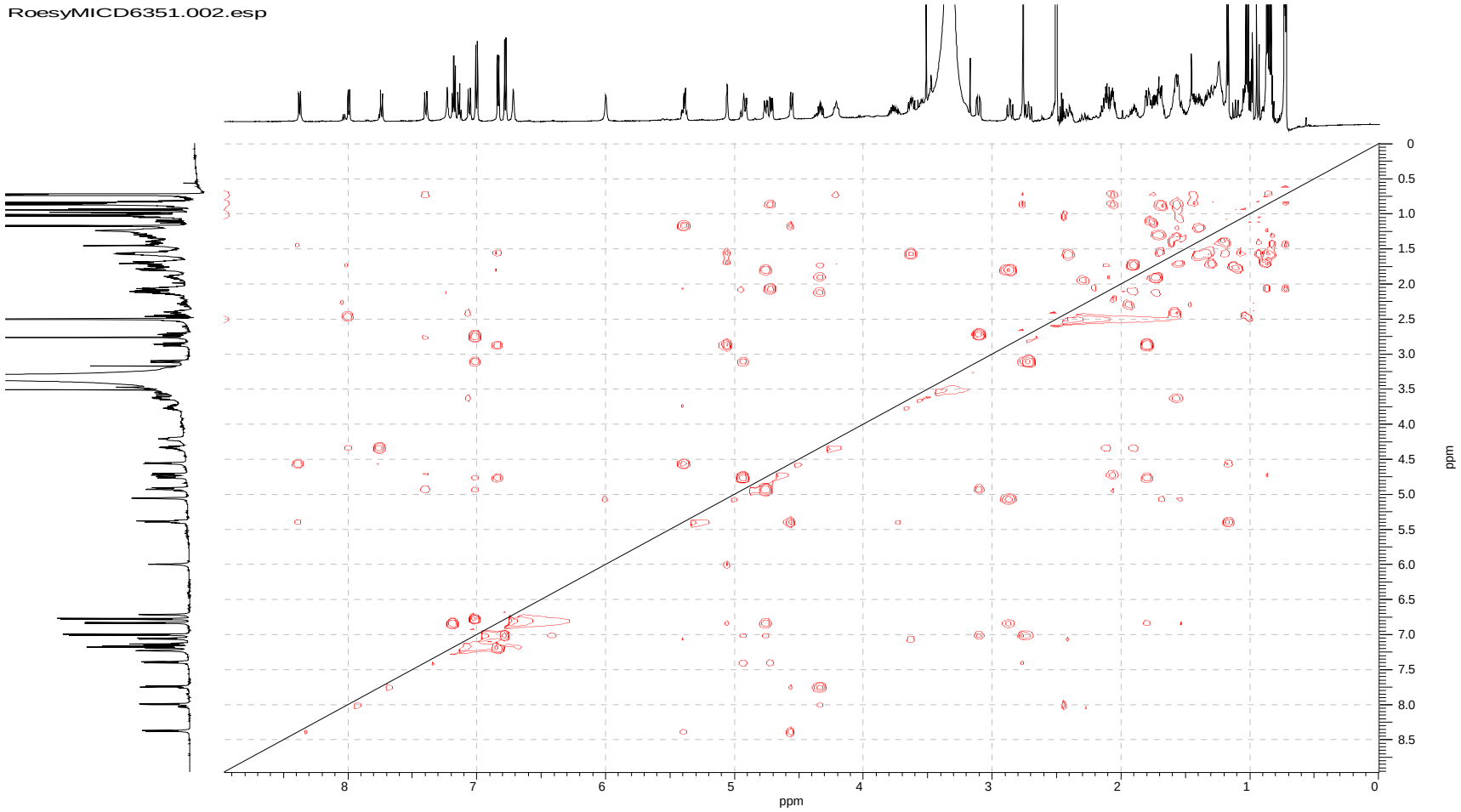
^{13}C NMR spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



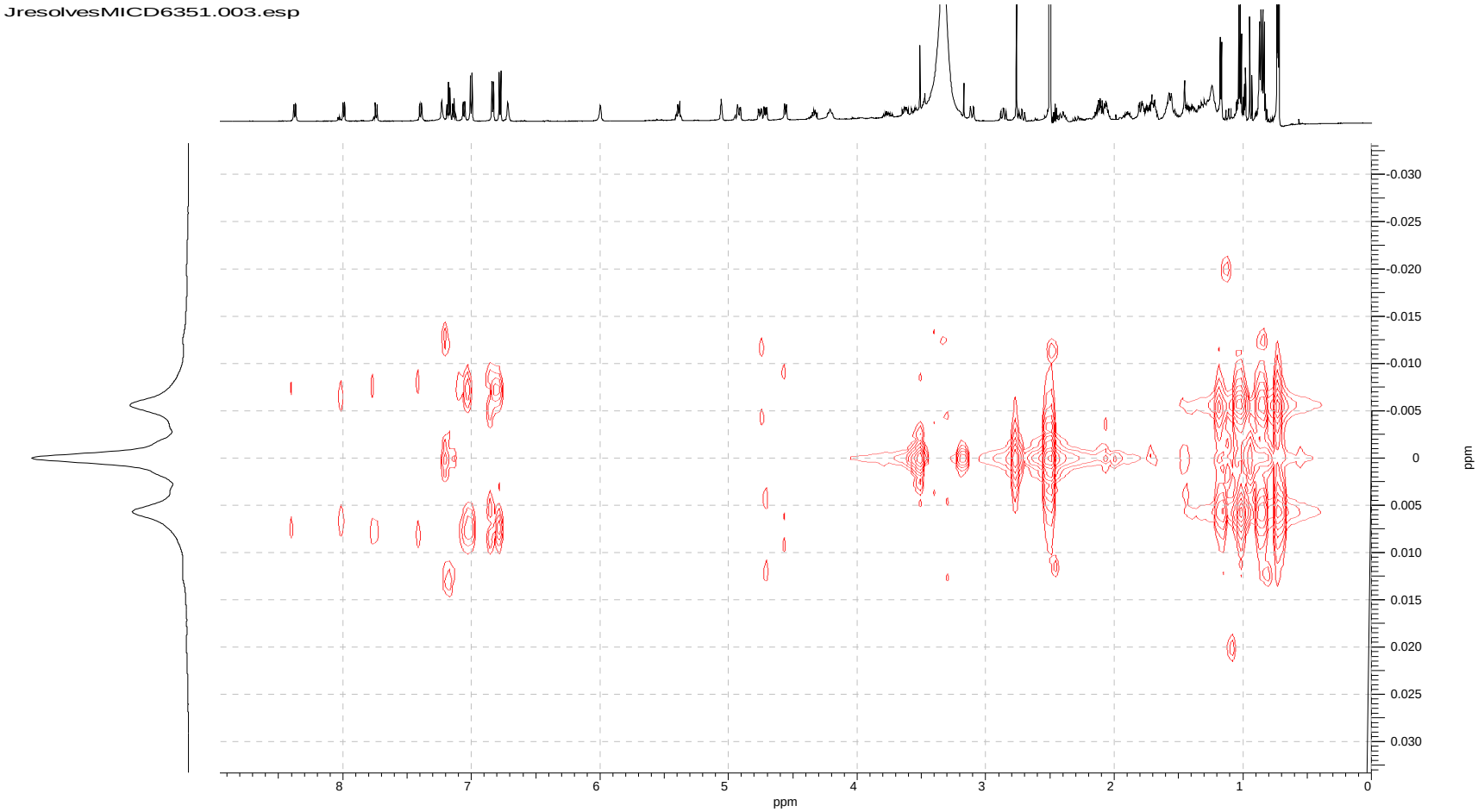
COSY spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



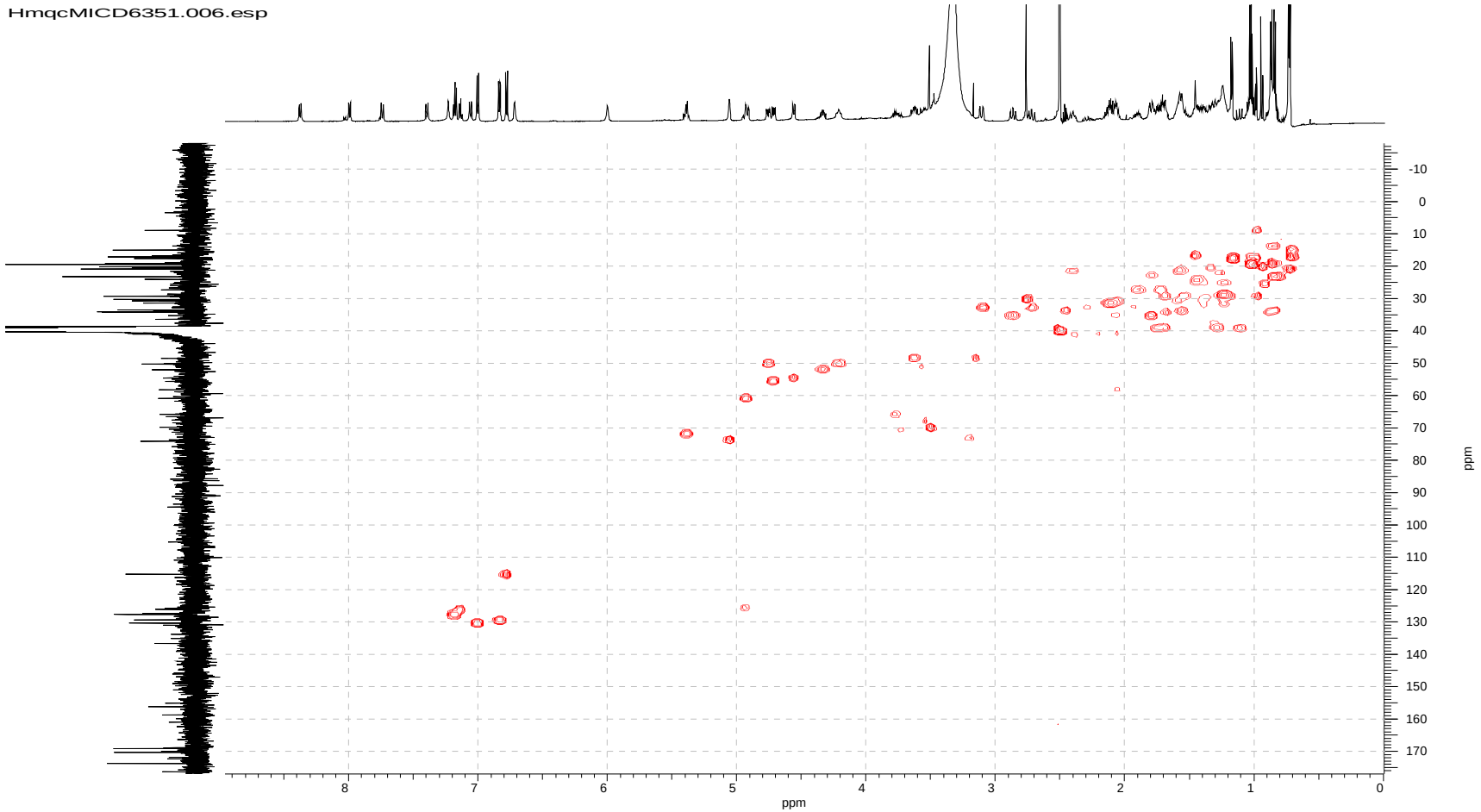
TOCSY spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



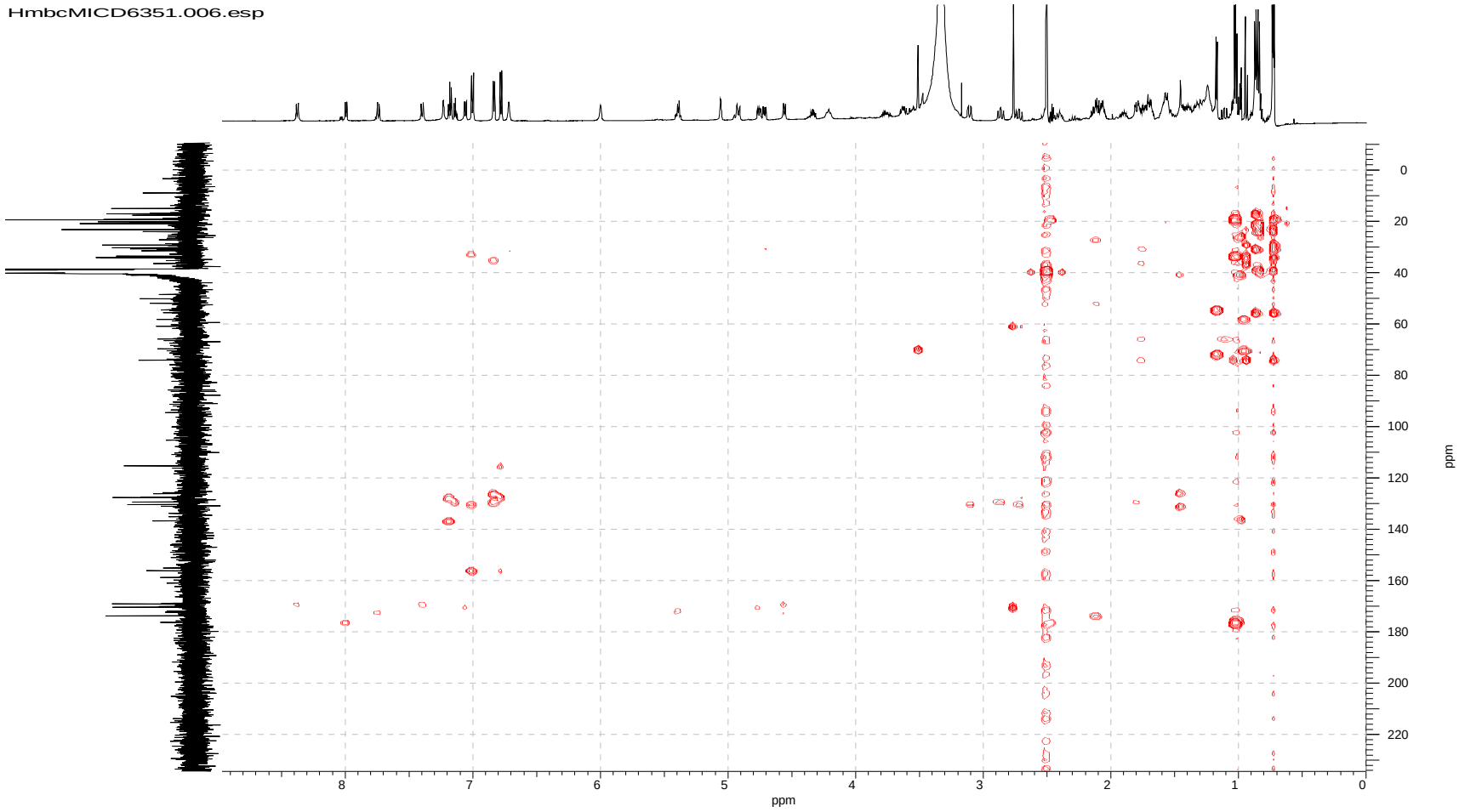
ROESY spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



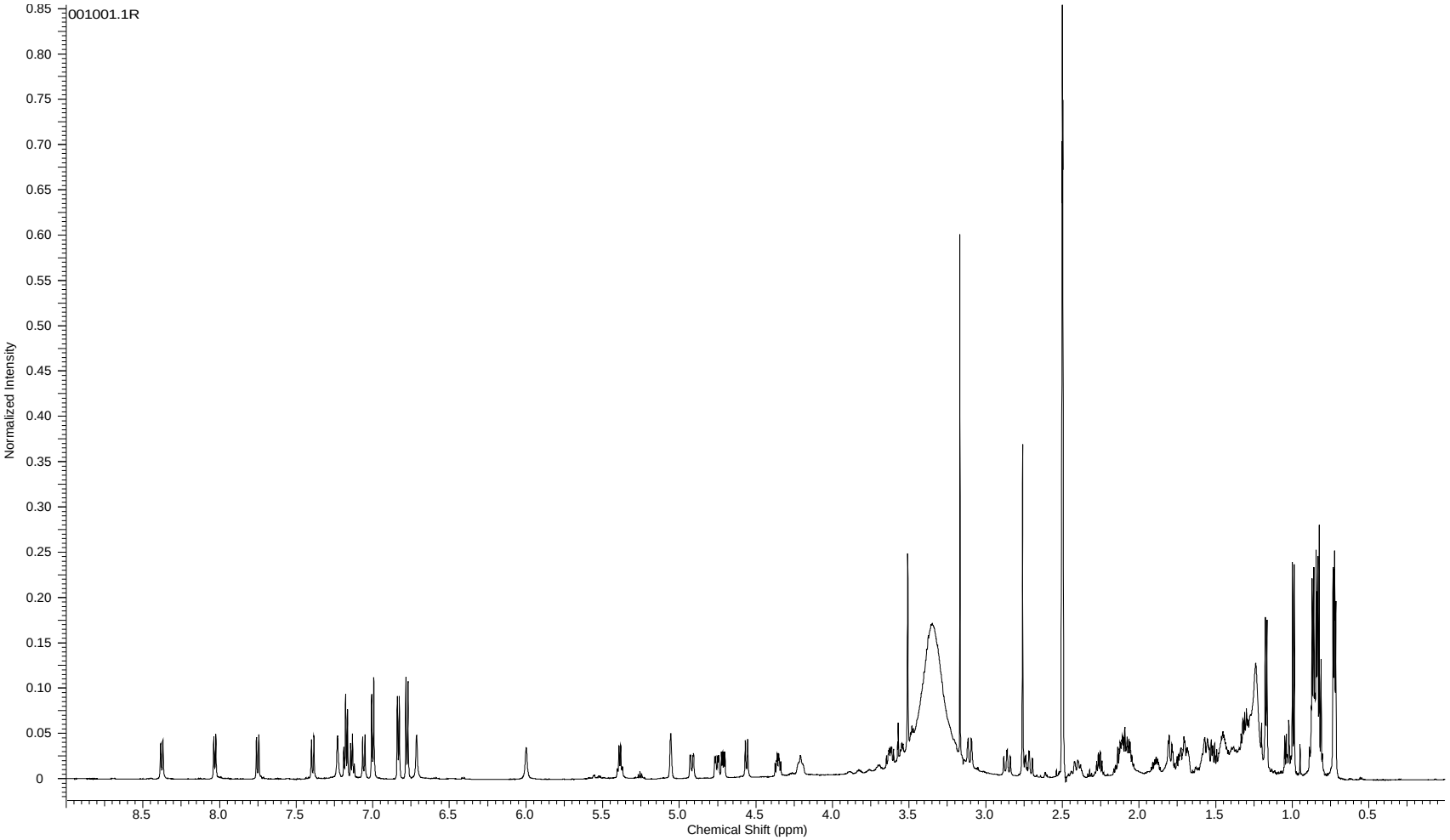
J-resolved spectrum of crocapeptin A_2 (**2**) in DMSO-d_6



HMQC spectrum of crocapeptin A_2 (**2**) in DMSO-d_6

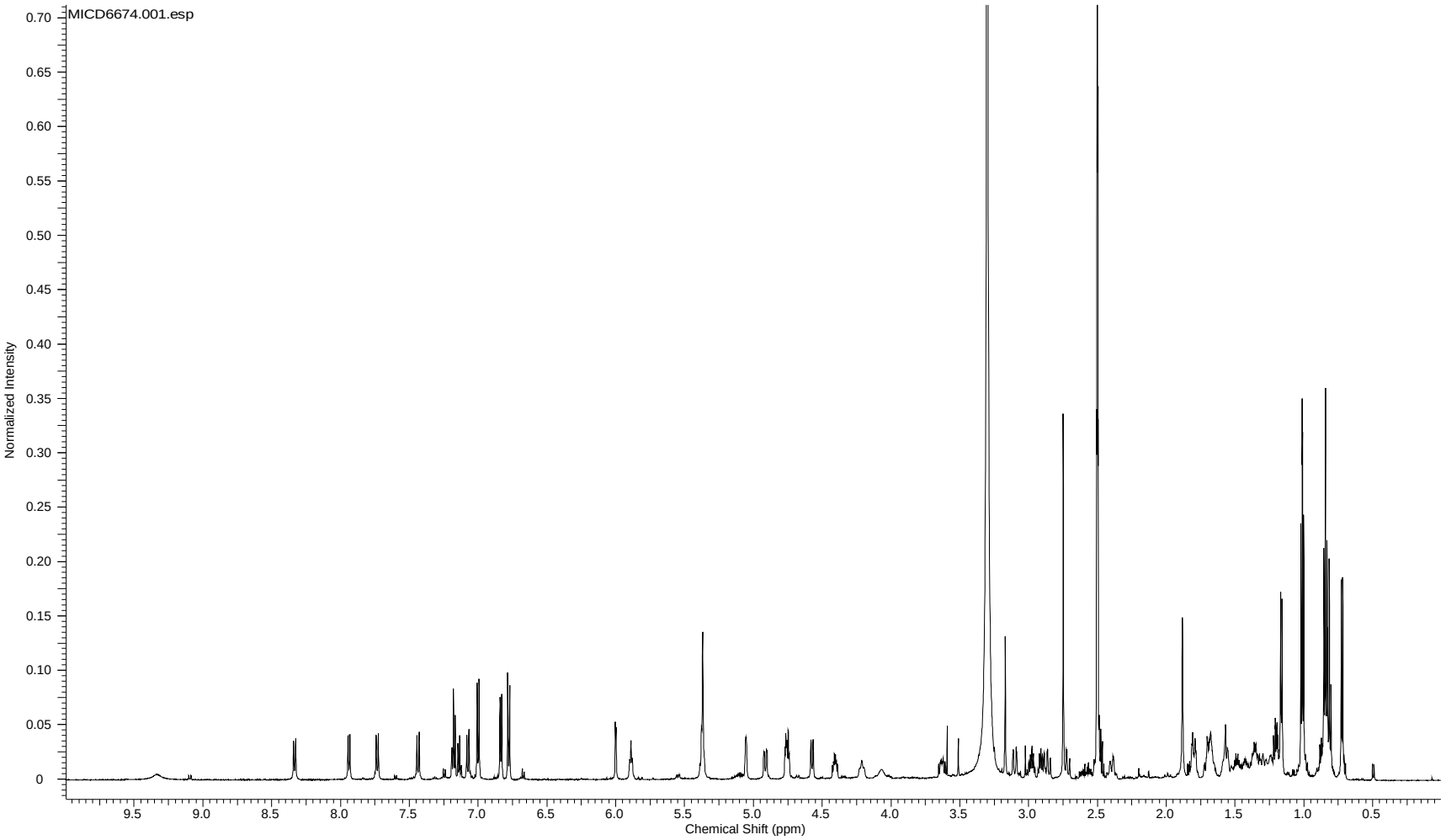


HMBC spectrum of crocapeptin A_2 (**2**) in DMSO-d_6

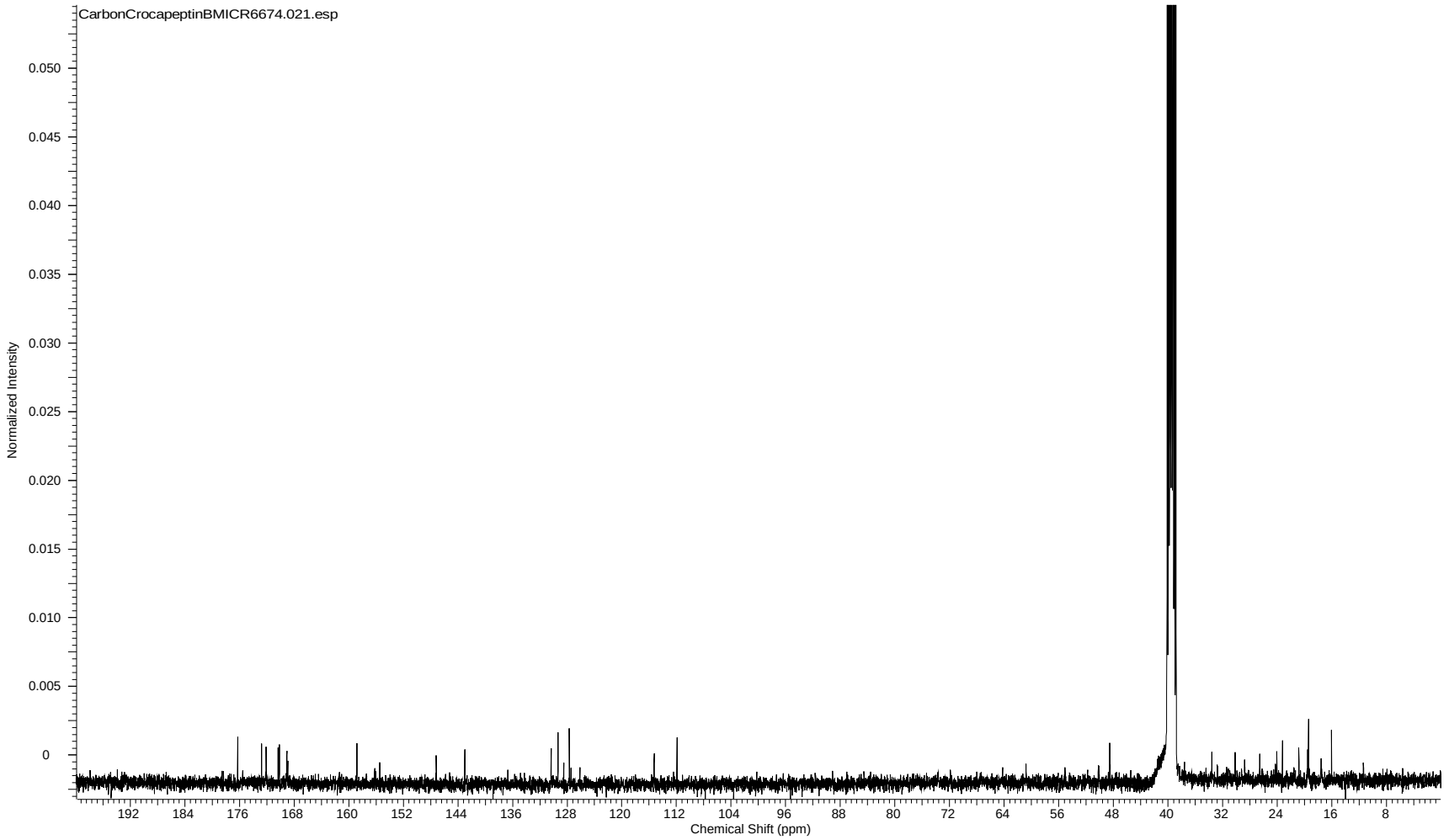


^{1}H NMR spectrum of crocapeptin A_3 (**3**) in DMSO-d_6

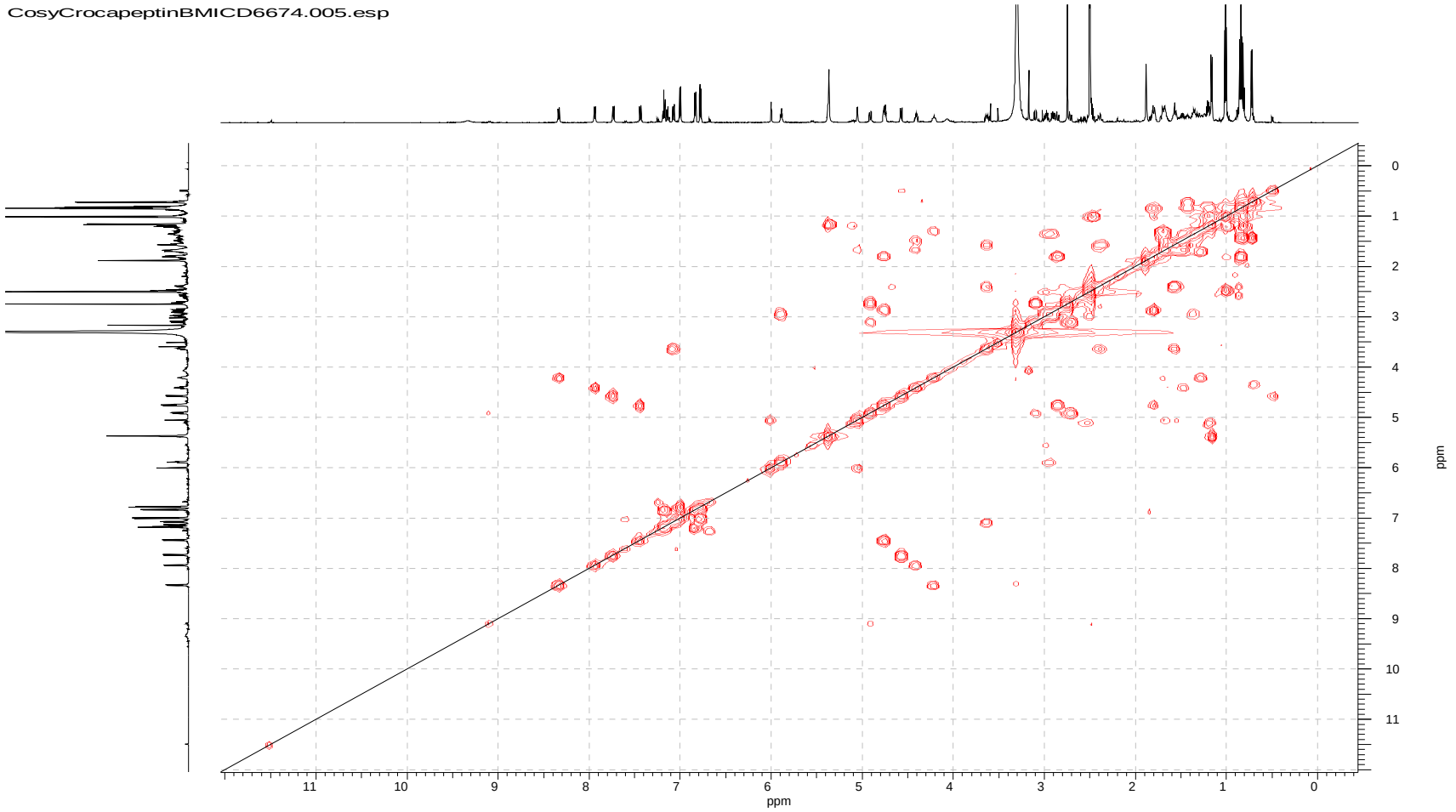
^{13}C NMR spectrum of crocapeptin A_3 (**3**) in DMSO-d_6



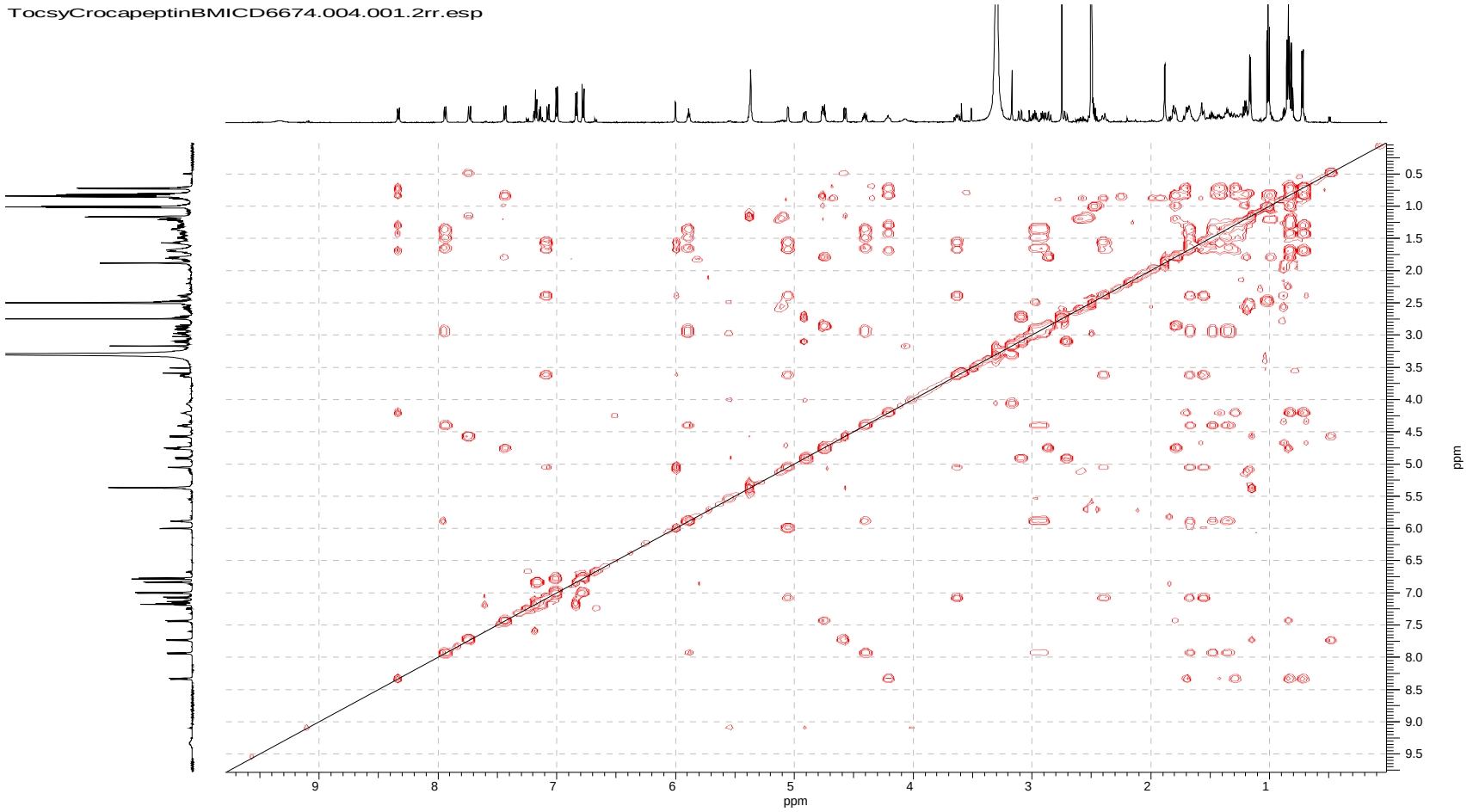
^{1}H NMR spectrum of crocapeptin B (**4**) in DMSO-d_6



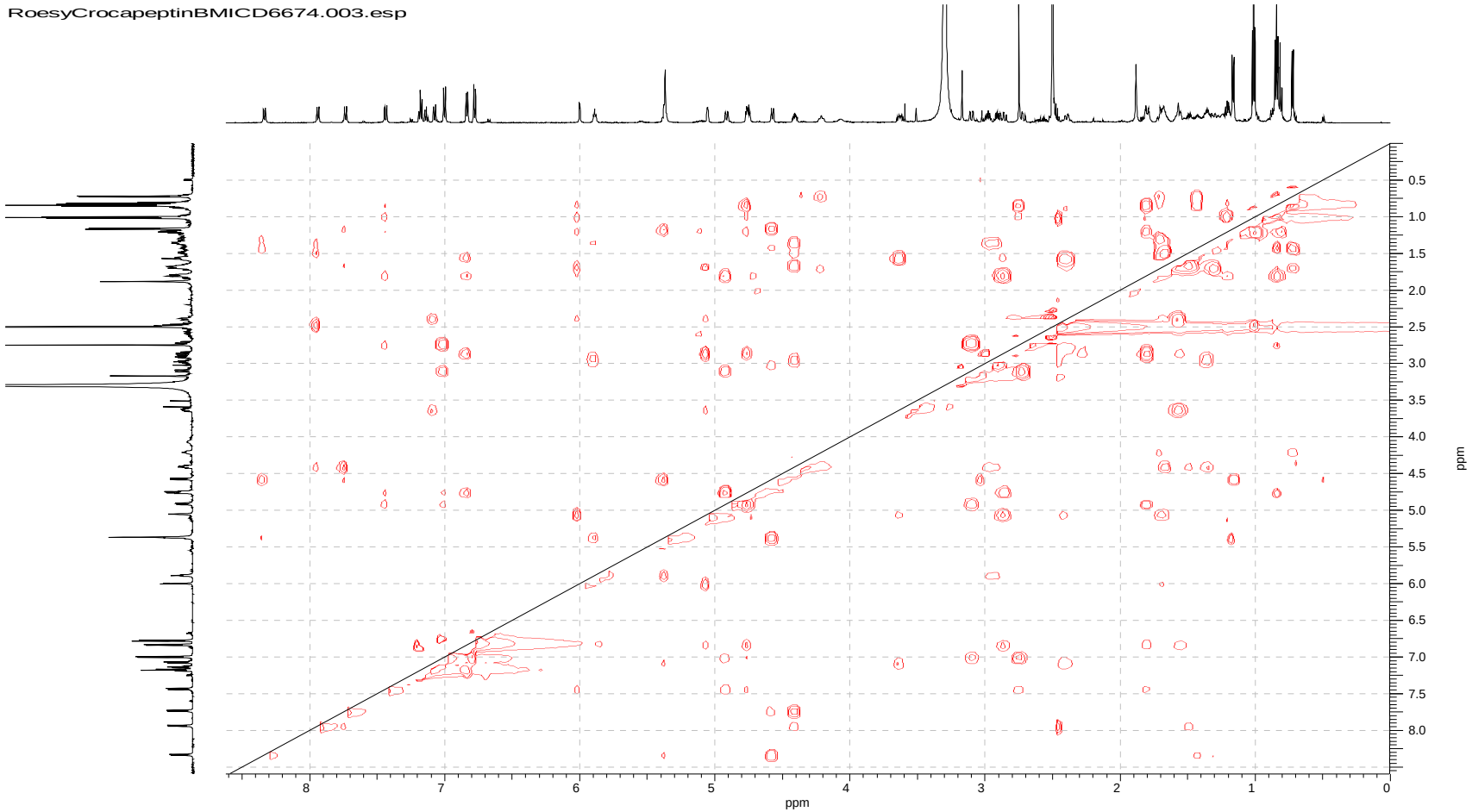
^{13}C NMR spectrum of crocapeptin B (**4**) in DMSO-d_6



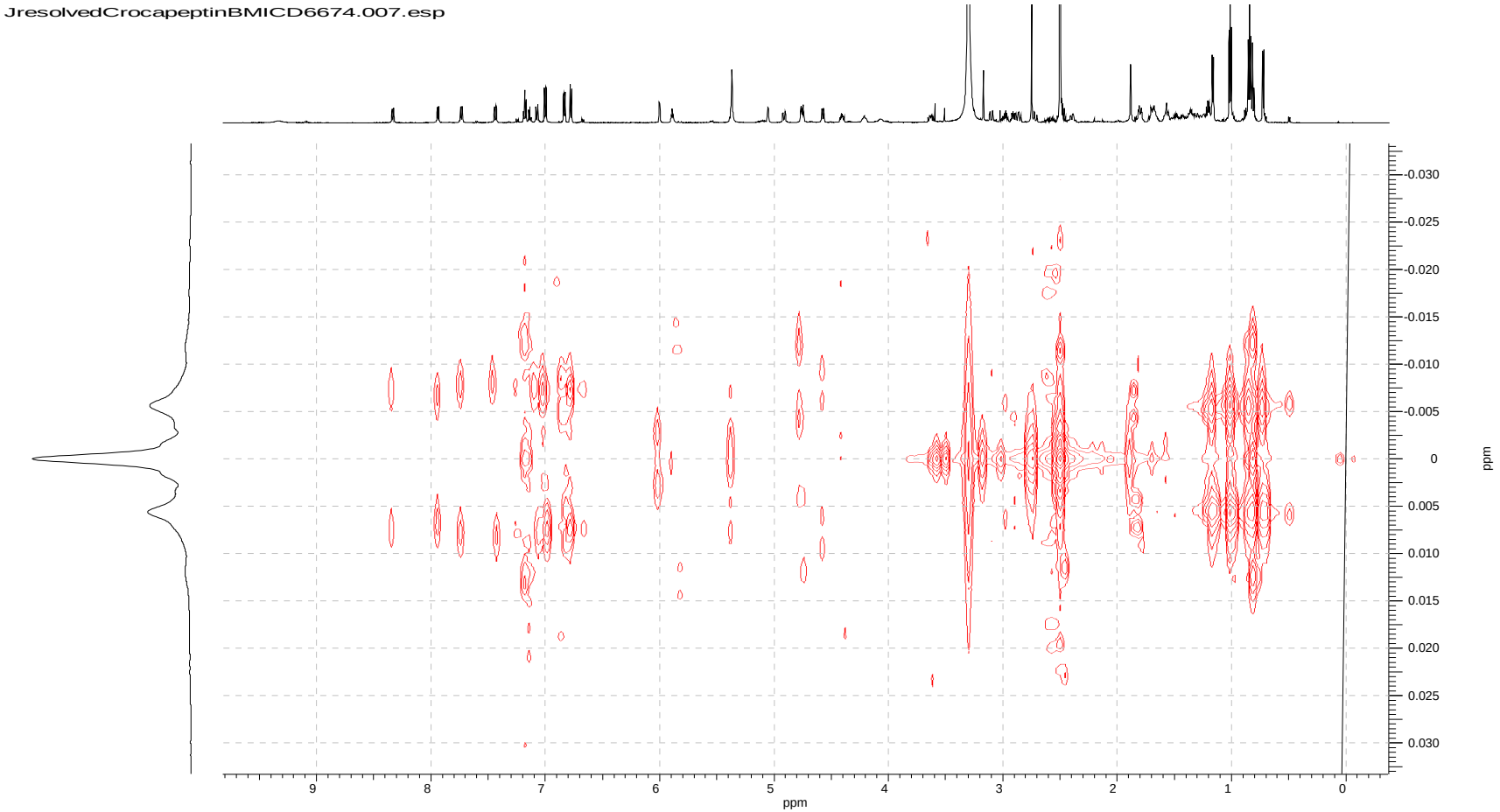
COSY spectrum of crocapeptin B (**4**) in DMSO-d_6



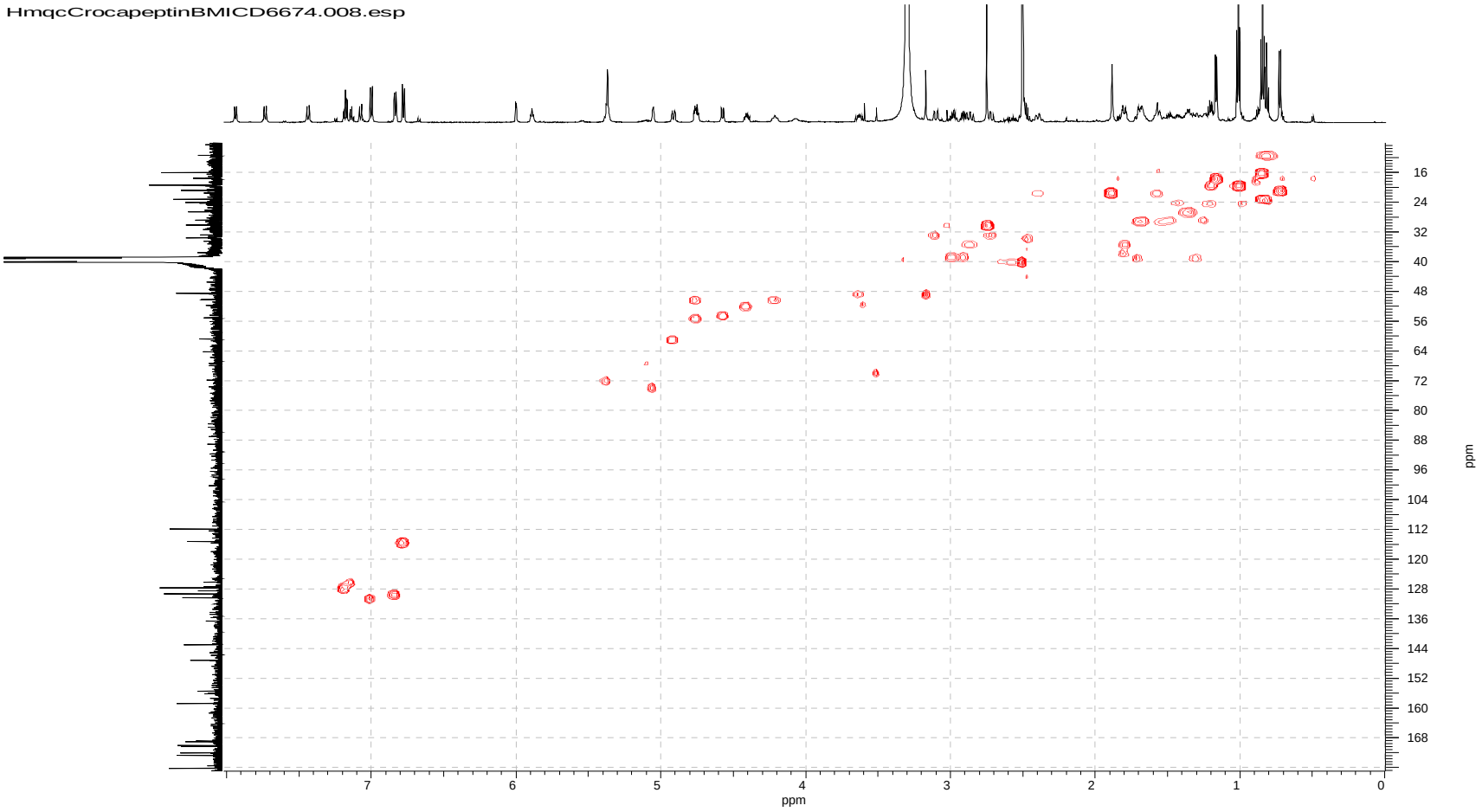
TOCSY spectrum of crocapeptin B (**4**) in DMSO-d_6



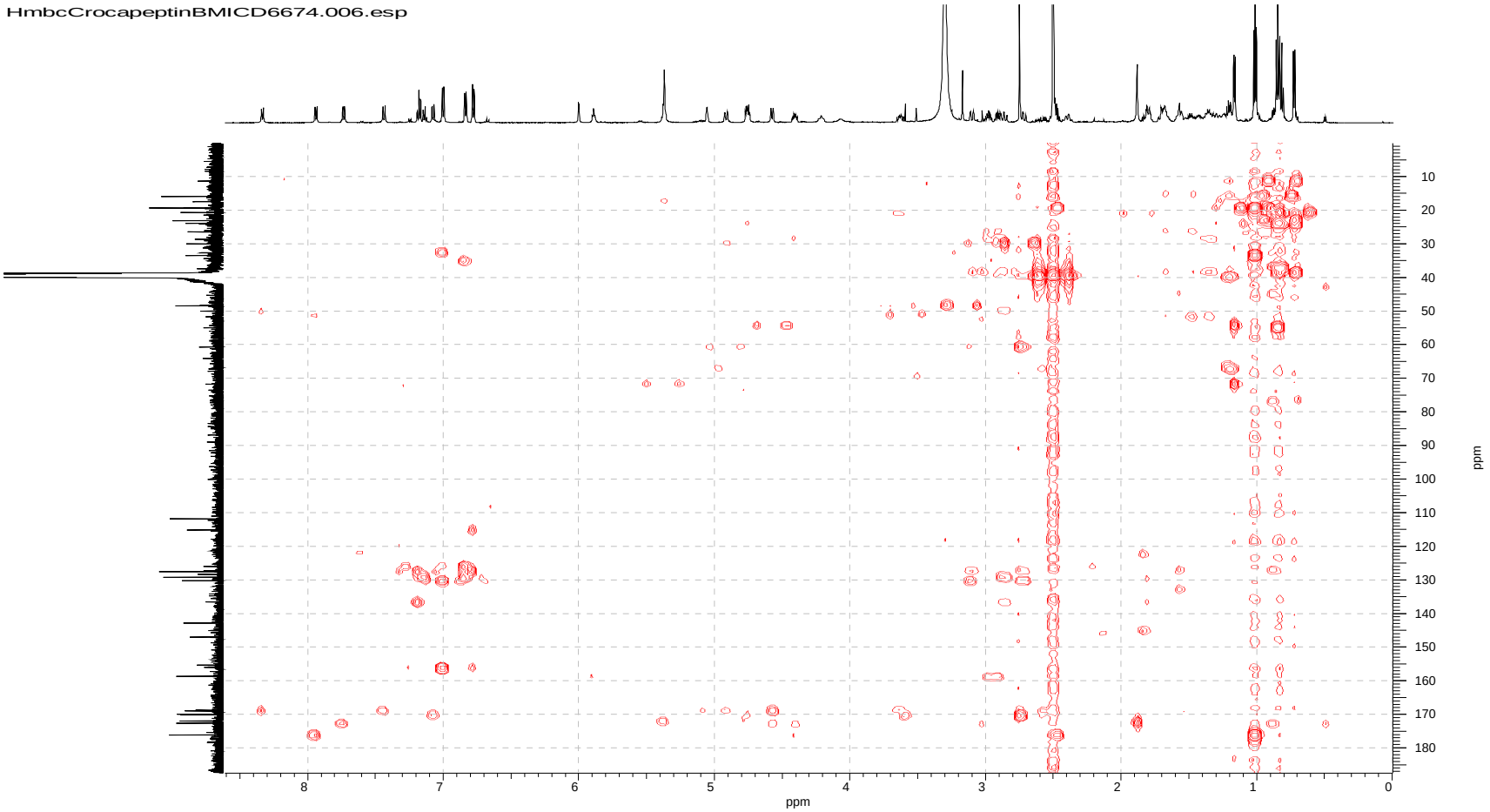
ROESY spectrum of crocapeptin B (**4**) in DMSO-d_6



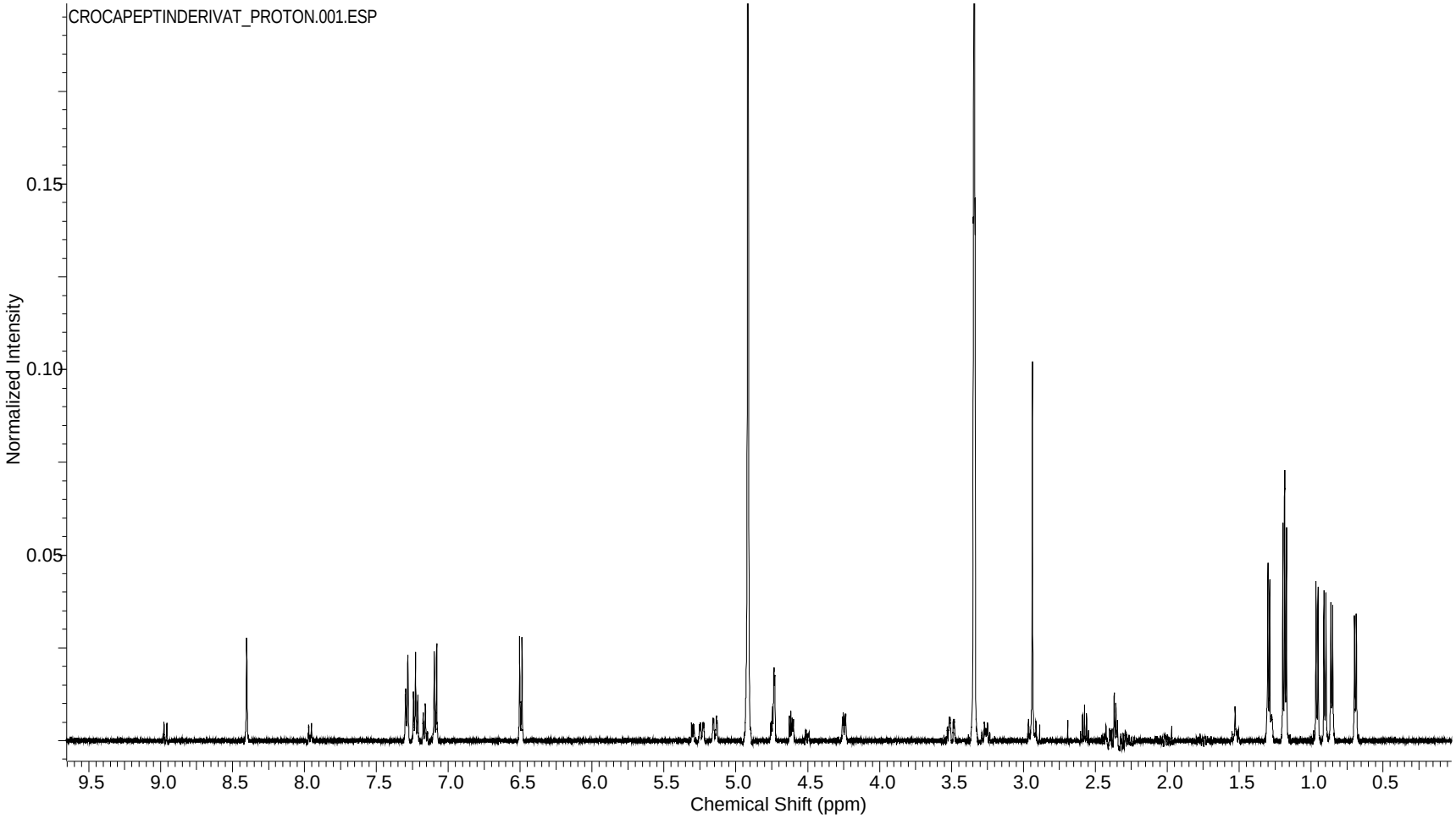
J-resolved spectrum of crocapeptin B (**4**) in DMSO-d_6



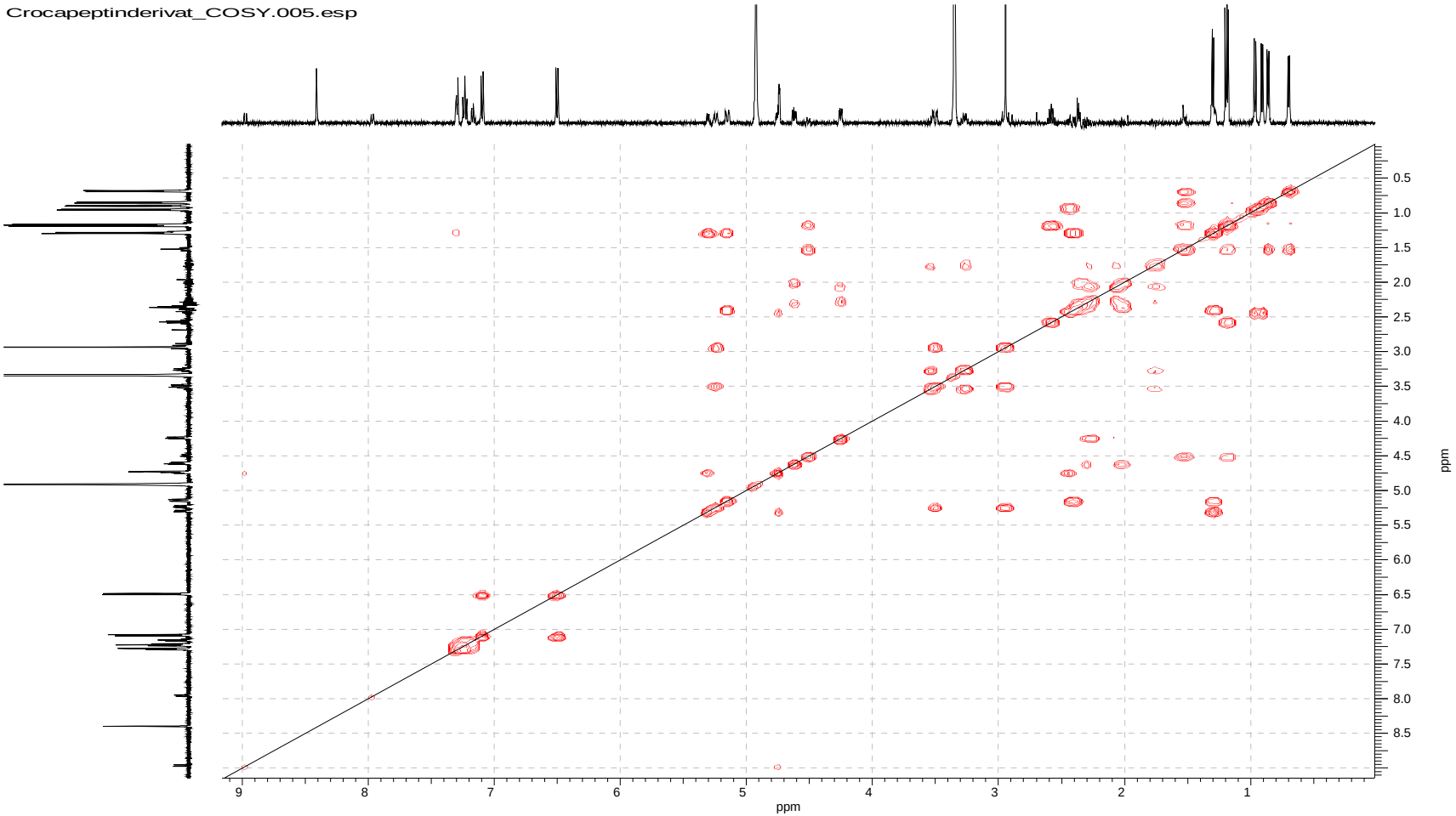
HMQC spectrum of crocapeptin B (**4**) in DMSO-d_6



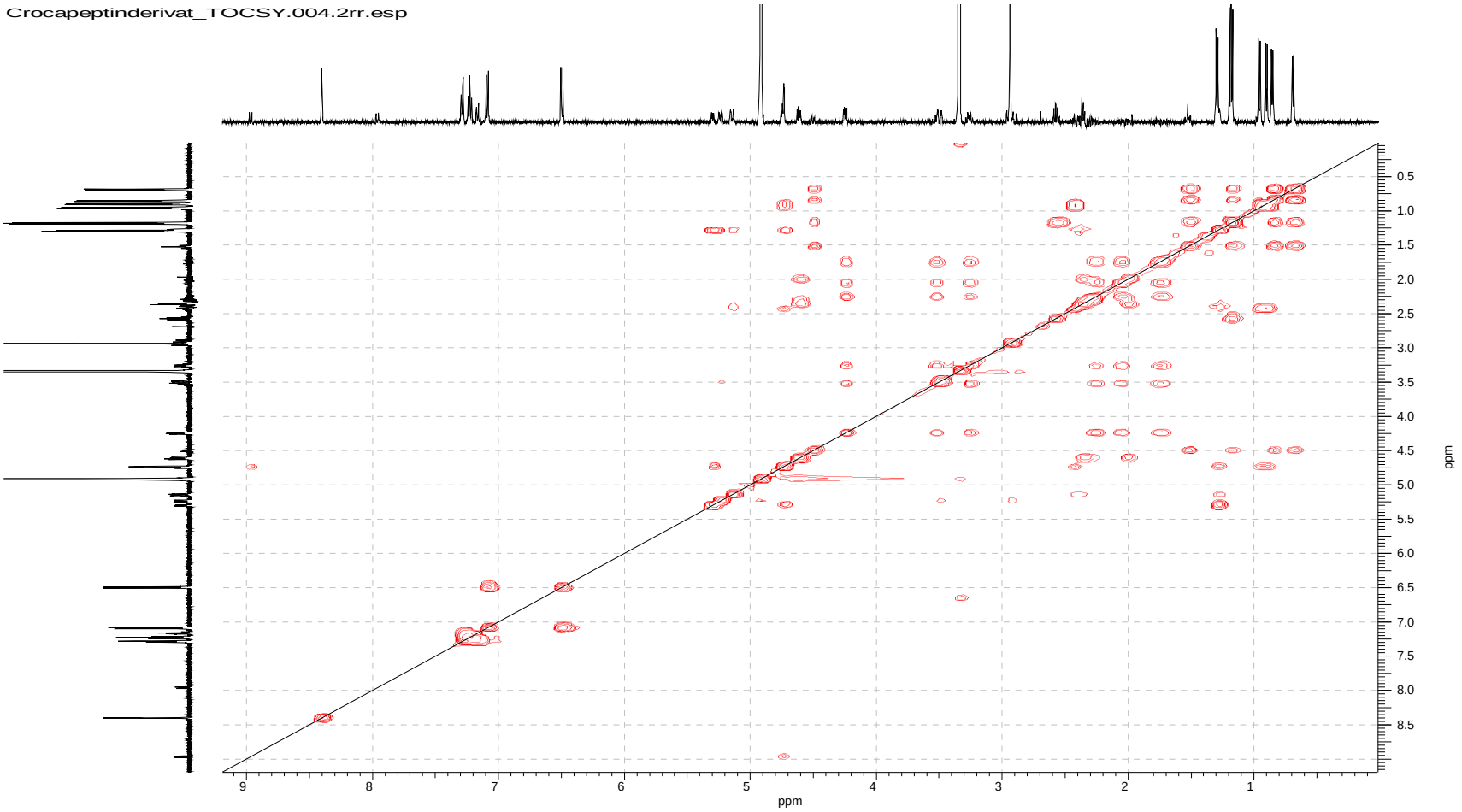
HMBC spectrum of crocapeptin B (**4**) in DMSO-d_6



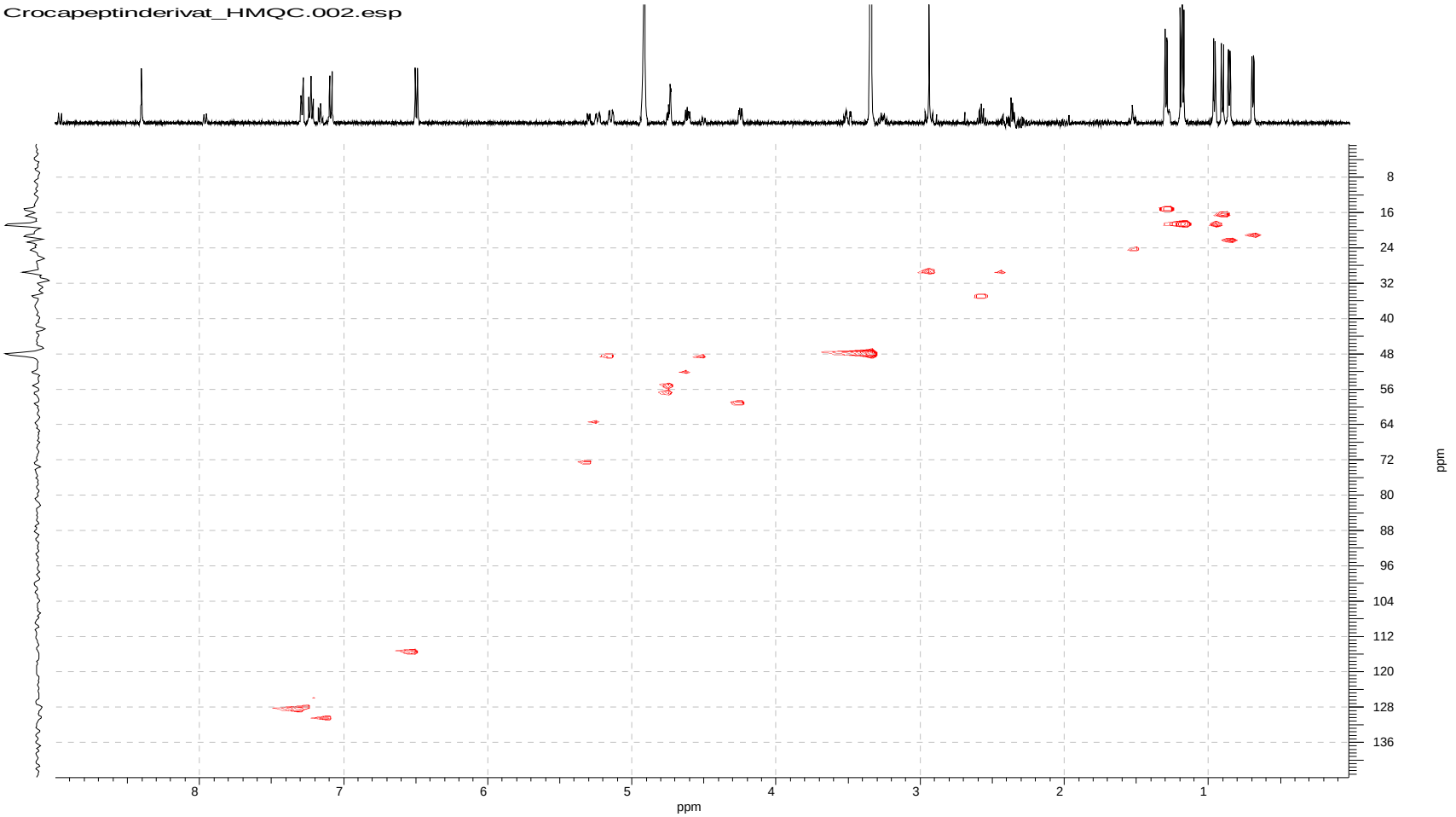
^{1}H NMR spectrum of pre-crocapeptin A_2 (**5**) in MeOH-d_4



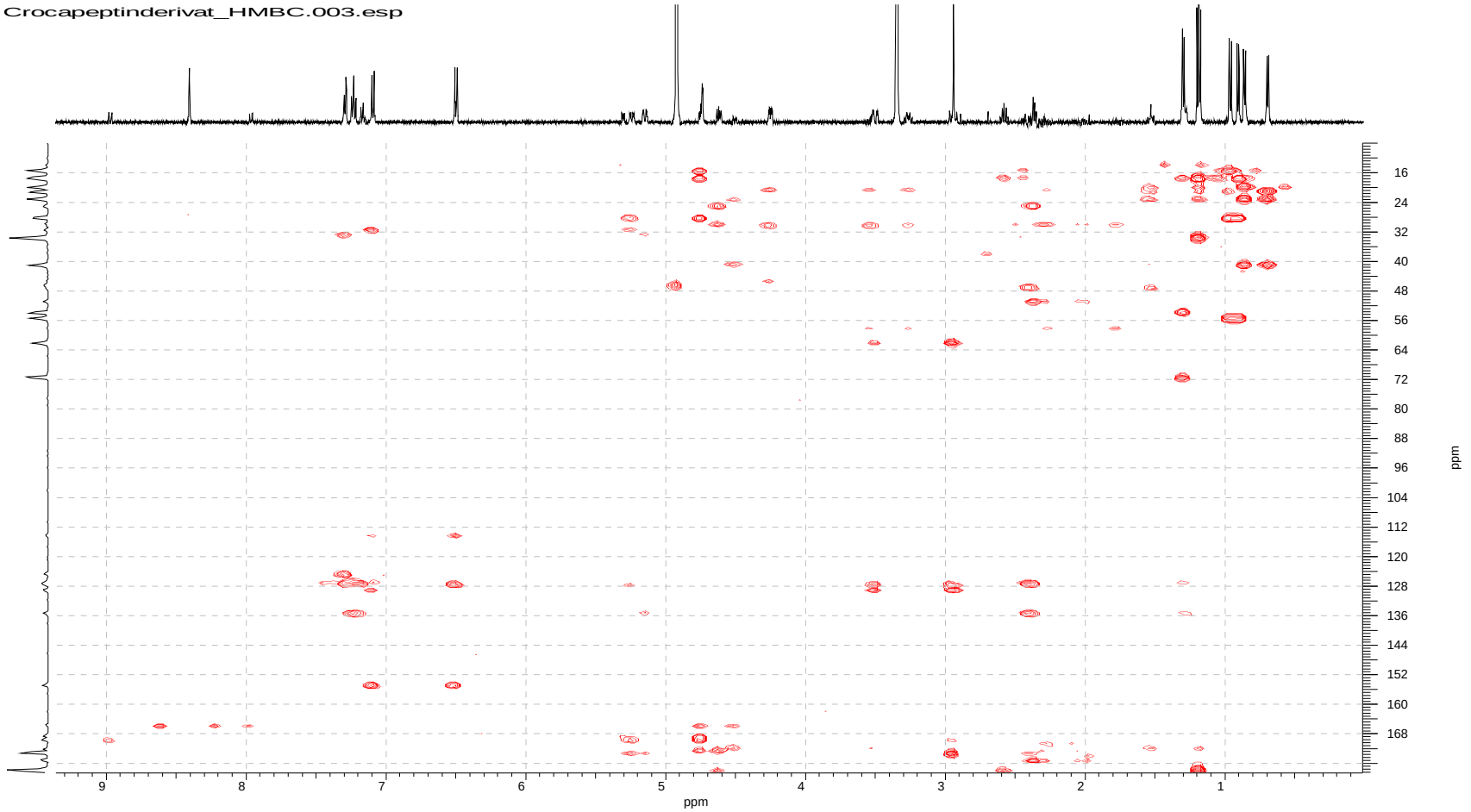
COSY spectrum of pre-crocapeptin A_2 (**5**) in MeOH-d_4



TOCSY spectrum of pre-crocapeptin A_2 (**5**) in MeOH-d_4



HMQC spectrum of pre-crocapeptin A_2 (**5**) in MeOH-d_4



HMBC spectrum of pre-crocapeptin A_2 (**5**) in MeOH-d_4

Enzymatic assay for chymotrypsin inhibition with 1 – 4.

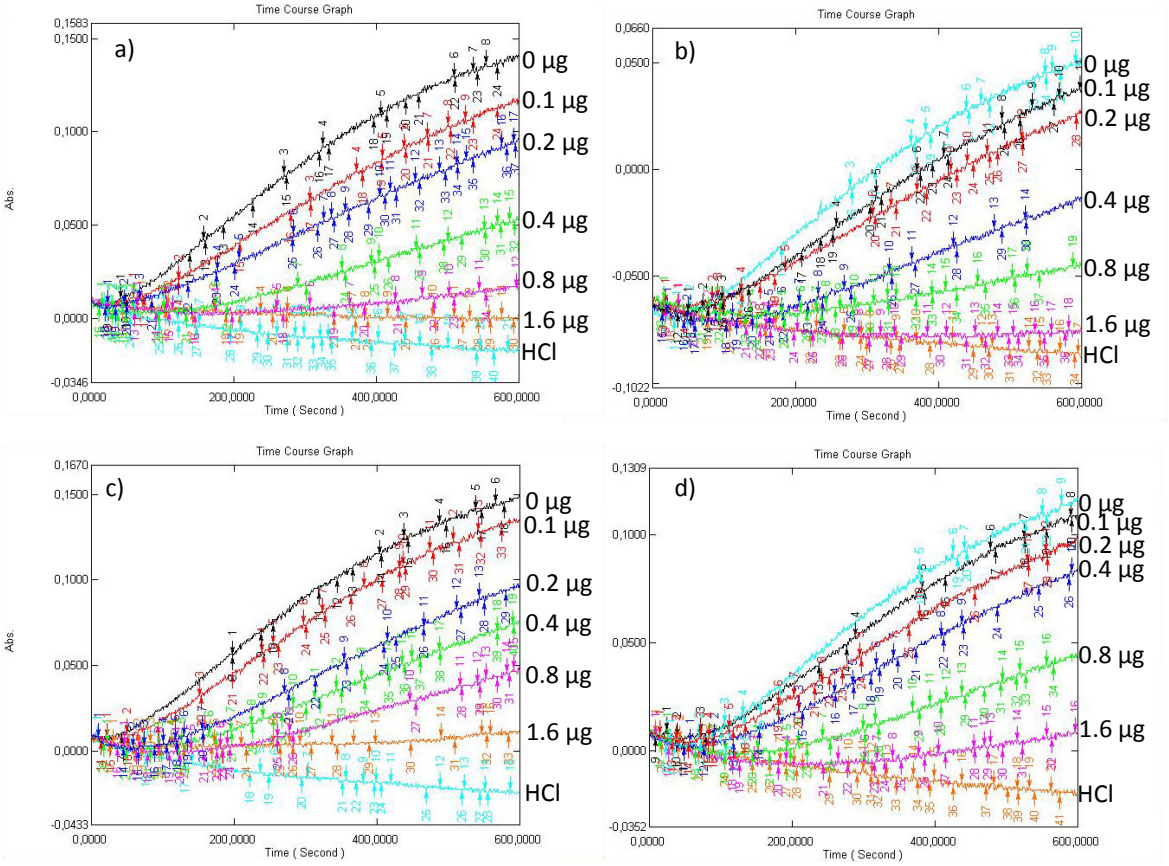


Figure S5-1: Change of the absorbance at 256 nm in the presence of 0 µg; 0.1 µg; 0.2 µg; 0.4 µg; 0.8 µg and 1.6 µg a) crocapeptin A_1 (**1**), b) crocapeptin A_2 (**2**), c) crocapeptin A_3 (**3**) and d) crocapeptin B (**4**) per 3 mL solution.

To determine the IC50 for the different compounds, the shift of the absorbance at 256 nm after 5 minutes was observed. An experiment with 100 µL HCl (1 mM) instead of enzyme solution served as reference. The detected shifts are shown in table S5. IC50 were determined to approximately 0.1 µM for crocapeptin A_1-A_3 (**1** – **3**) and approximately 0.2 µM for crocapeptin B (**4**) by assigning the concentration for half of the change of absorbance.

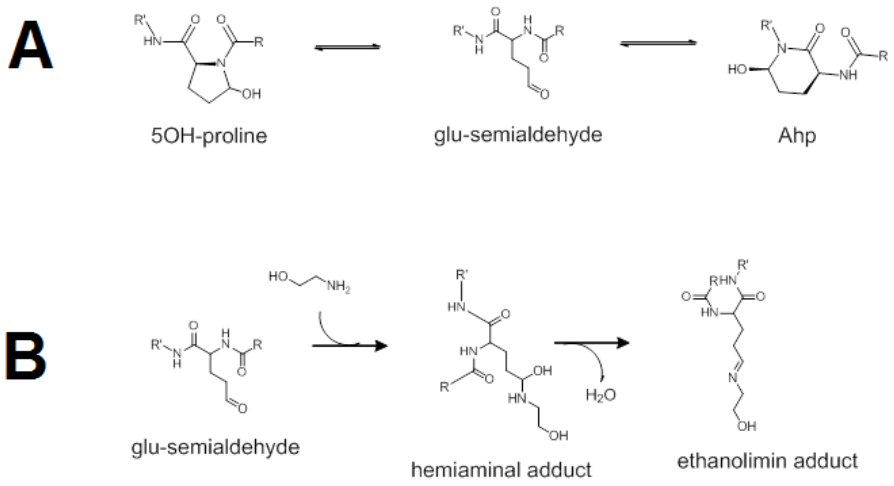
Table S5: Shifts in the absorbance at 256 nm after the addition of 0 µL, 10 µL, 20 µL, 40 µL, 80 µL and 160 µL solution of the test substances (10 µg/mL MeOH) compared to a blank run with 100 µL HCl solution (1 mM) instead of enzyme solution.

µg/3mL	Crocapeptin A1	Crocapeptin A2	Crocapeptin A3	Crocapeptin B
0	0,094	0.079	0.105	0.073
0.1	0.072	0.061	0.088	0.066
0.2	0.057	0.051	0.056	0.056
0.4	0.028	0.026	0.042	0.045
0.8	0.015	0.013	0.027	0.020
1.6	0.012	0.009	0.012	0.005

S6: In-vitro hydroxylation of pre-crocapeptin and aldehyde trapping experiments

Activity of purified recombinant proteins CpnE and CpnF were tested with isolated precursor pre-crocapeptin A1, A2 and A3 as substrates. Incubation with CpnF alone mostly leads to an instable isomer, which slowly converts to the respective crocapeptin derivative. Incubation of pre-crocapeptin with CpnE alone does not indicate any reaction. However, simultaneous incubation of pre-crocapeptins with CpnF and CpnE results mostly in crocapeptin formation. We therefore propose that the instable isomer is the hydroxyproline form of crocapeptin, since the glutamic-acid semialdehyde form would be too short-lived to be detected directly. Efforts to indirectly detect the glutamic-acid semilaldehyde intermediate were carried out by incubating the reaction mixture with an excess of ethanolamin. A fraction of the short-lived semiladehyde would then favor nucleophilic addition to the primary amino group of the ethanolamin over the adjacent peptide bond, as shown in scheme S6-1.

Scheme S6-1 A: proposed equilibrium in the crocapeptins between 5-OH-proline, glutamic-acid-semialdehyde and Ahp heterocycle states. **B:** Incubation of crocapeptins with ethanolamin leads to formation of adduct via a hemiaminal and subsequent condensation to the imine form.



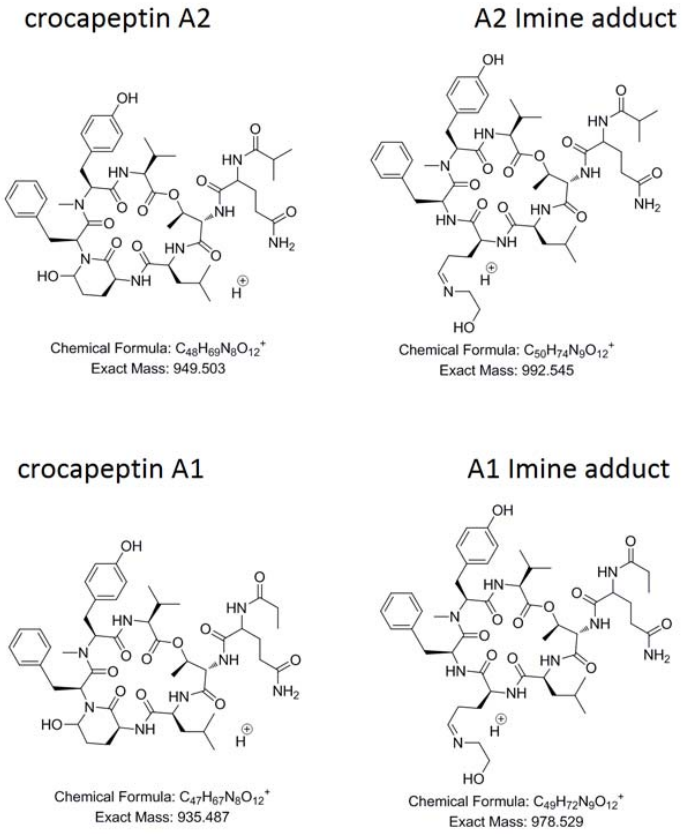


Figure S6-2: Chemical structure and masses of the crocapeptins and ethanolimine adducts formed after incubation with ethanolamine.

The in vitro hydroxylation of pre-crocapeptin was done according to the information given in the experimental section. To trap the proposed aldehyde intermediate, 10 µl of ethanolamin were added to the reaction after 30 min of incubation with the cytochrome P450 reaction mixture. Samples were then incubated for 3 h at 30°C. After incubation, compounds were extracted with 250 µL ethyl acetate, dried *in vacuo* and redissolved in MeOH. Extracts were analyzed by ESI-MS, and examined for presence of mass signals of the reaction products displayed in figure S6-2. To test addition with pure crocapeptins, 25 µg of crocapeptins A1 and A2 were incubated with 10 µl of pure ethanolamine in 50 mM Tris/HCl buffer at pH 8.0 for 3 h at 30°C. After incubation, compounds were extracted with 250 µL ethyl acetate, dried *in vacuo* and redissolved in MeOH. Extracts were analyzed by ESI-MS as described in the experimental procedures.

As shown in figure S6-3, the ethanolamine adducts were indeed formed after treatment of the hydroxylation reaction with ethanolamine, forming a peak with a mass of 987.5 Da, which is corresponds the the imine form of the adduct of crocapeptin A1. These adducts can hawever also be formed by incubation of crocapeptin with ethanolamine and subsequent extraction as described.

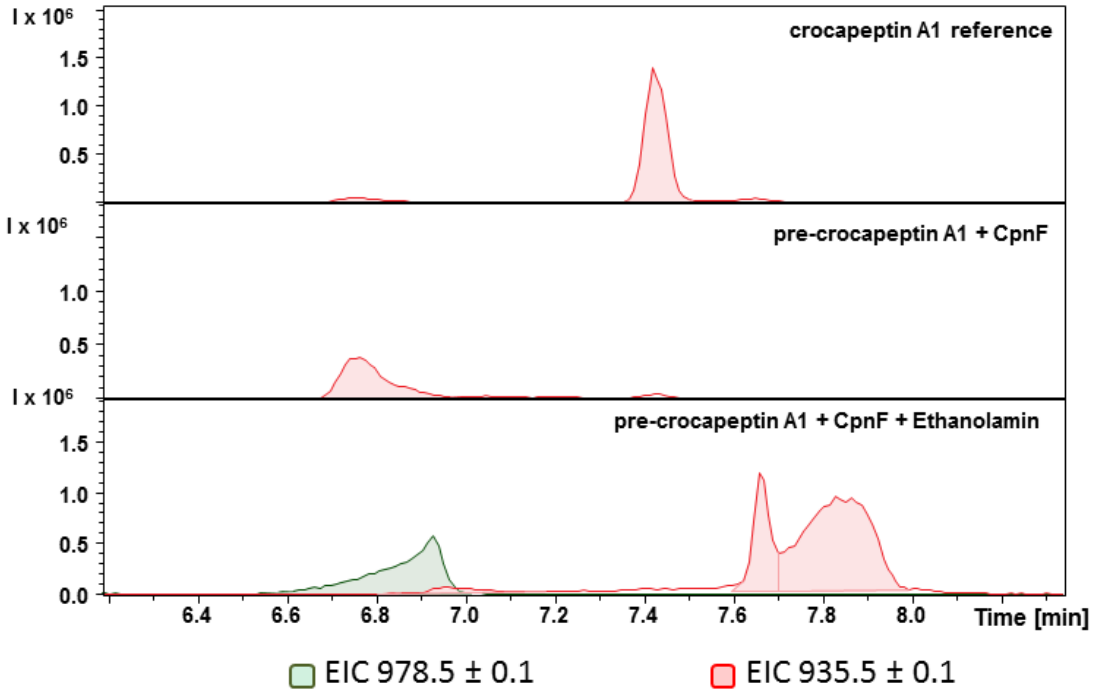


Figure S6-3: Extracted ion chromatograms of crocapeptin A1 (upper panel), pre-crocapeptin A1 after in-vitro hydroxylation with CpnF (middle panel) and pre-crocapeptin A1 after *in vitro* hydroxylation with CpnF and subsequent incubation with ethanolamine (lower panel). Masses for crocapeptin A1 (red) and the proposed ethanolimine adduct (green) are shown, the latter being detected after incubation with ethanolamine at a retention time of 6.9 min.

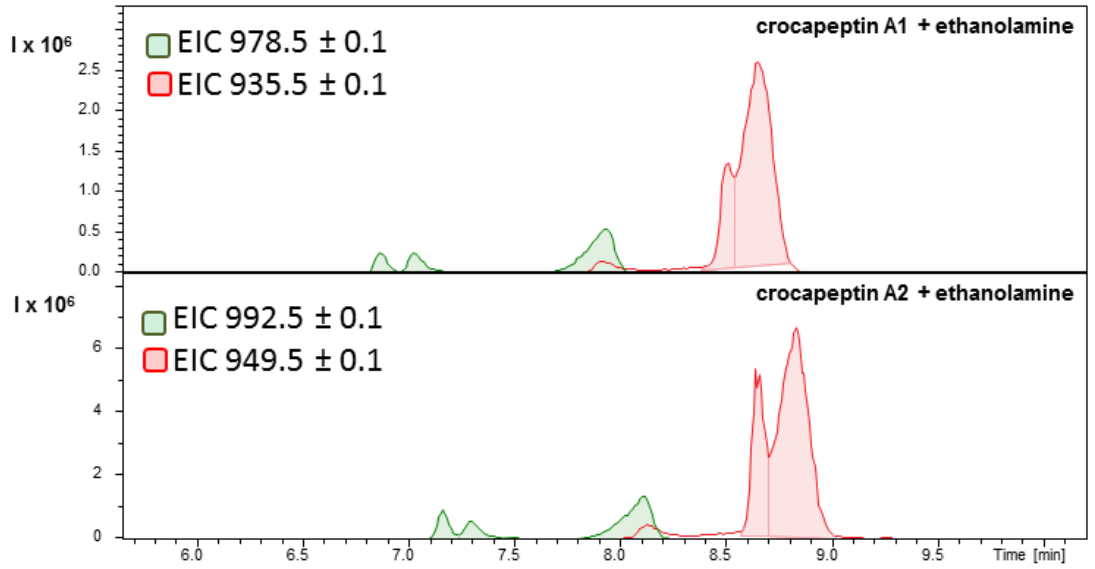


Figure S6-4: Extracted ion chromatograms of crocapeptin A1 incubated with ethanolamine (upper panel) and crocapeptin A2 incubated with ethanolamine (lower panel). Mass signals for crocapeptins are shown in red, the masses for the proposed ethanolimine adducts are shown in green.

S7: Elucidating the structure of unstable crocapeptin A_3 Intermediate

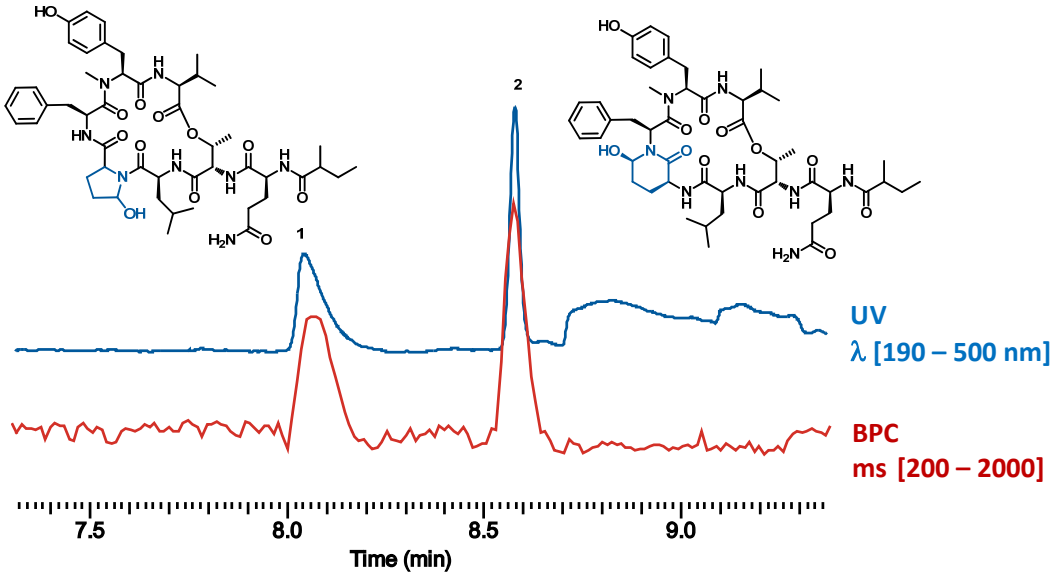


Figure S7-1: UV- and base peak chromatogram of separation method for crocapeptin A_3 (**2**) and its unstable isomeric intermediate (**1**).

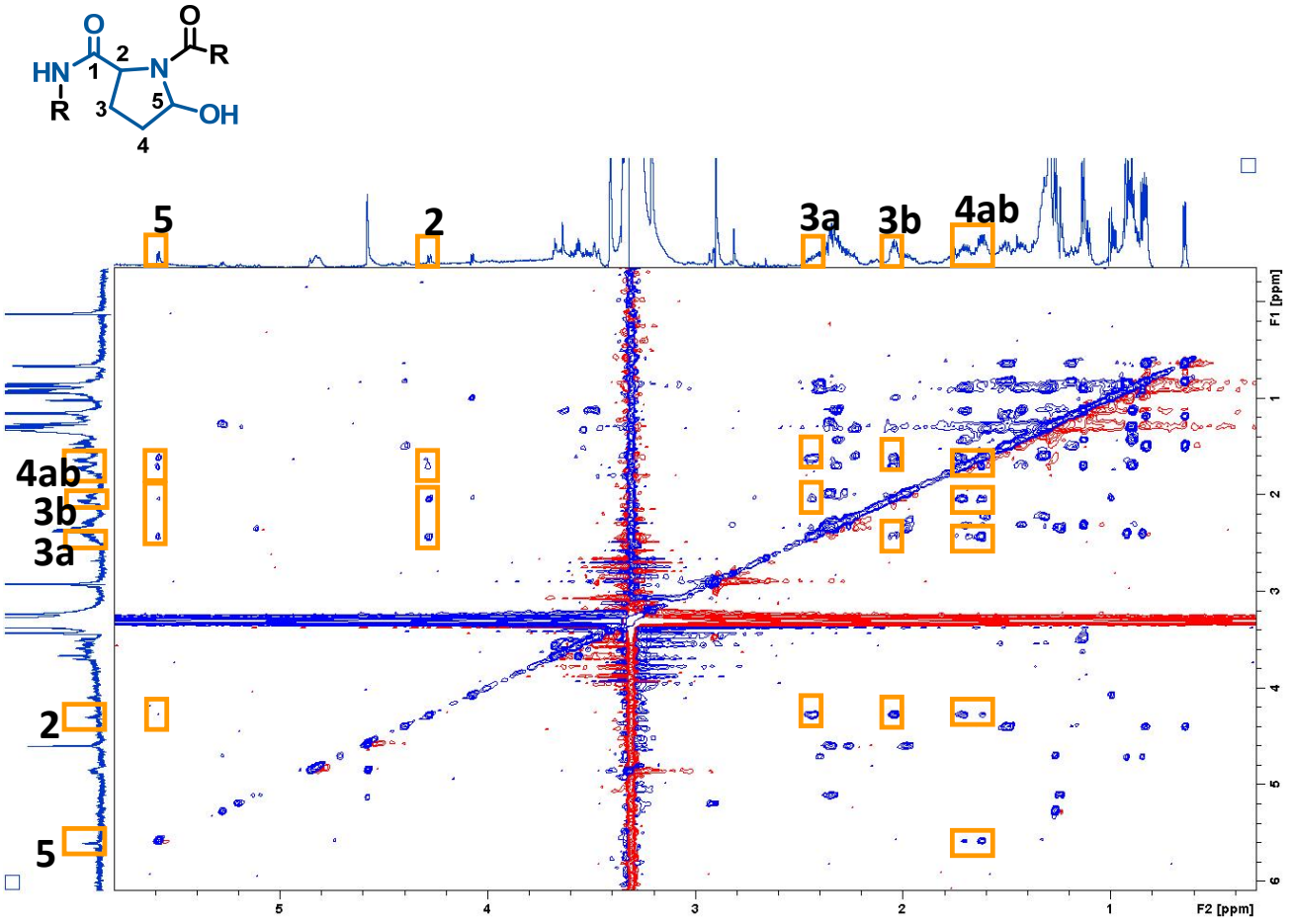


Figure S7-2: TOCSY correlations detected for the 5-hydroxyproline unit in crocapeptin A_3, MeOD, TD: 2k (F2), 512 (F1), NS24, D9 = 90 ms

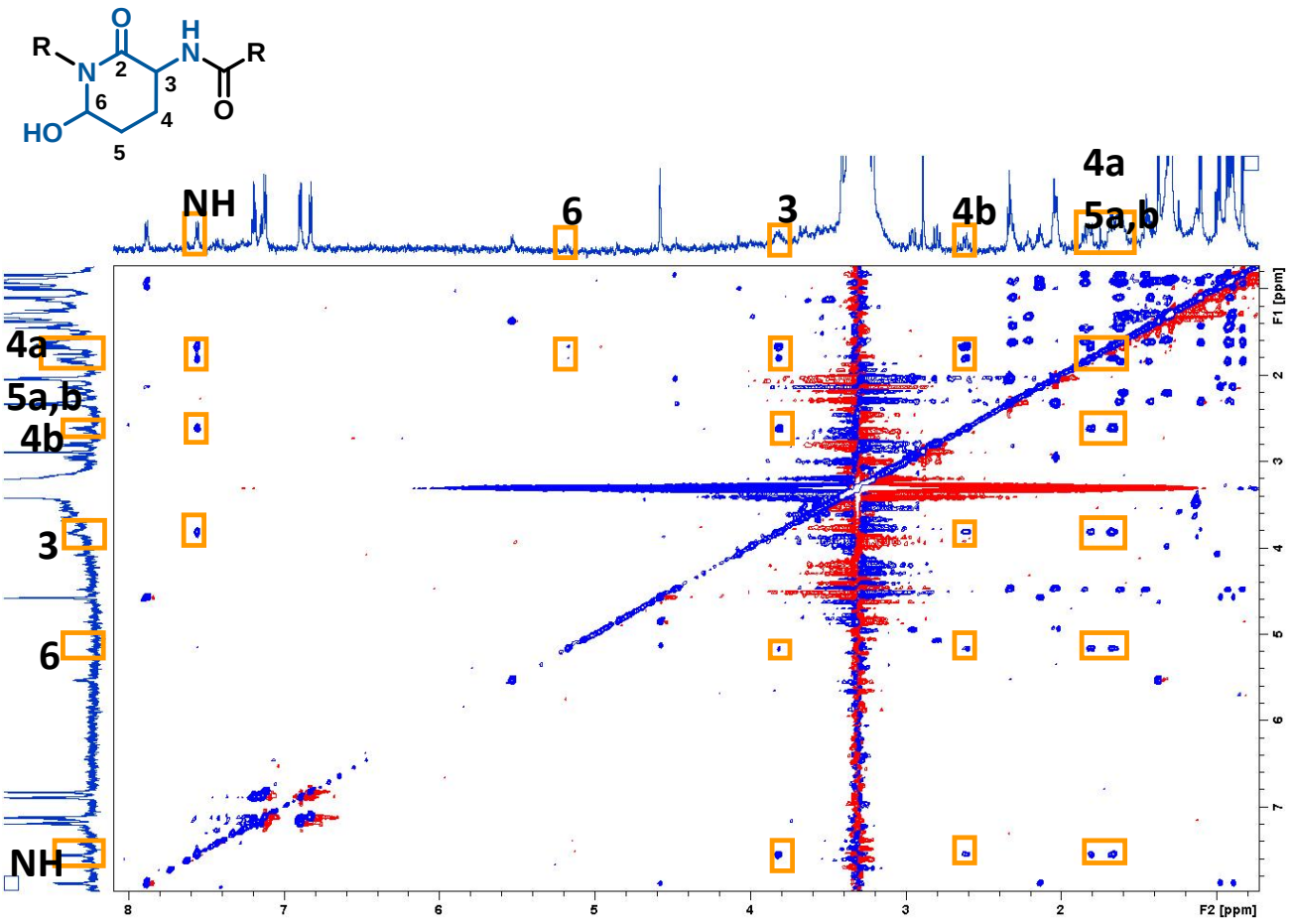


Figure S7-3: TOCSY correlations detected for the Amino-hydroxypiperidone unit in crocapeptin A_3, MeOD, TD: 2k (F2), 512 (F1), NS32, D9 = 90 ms

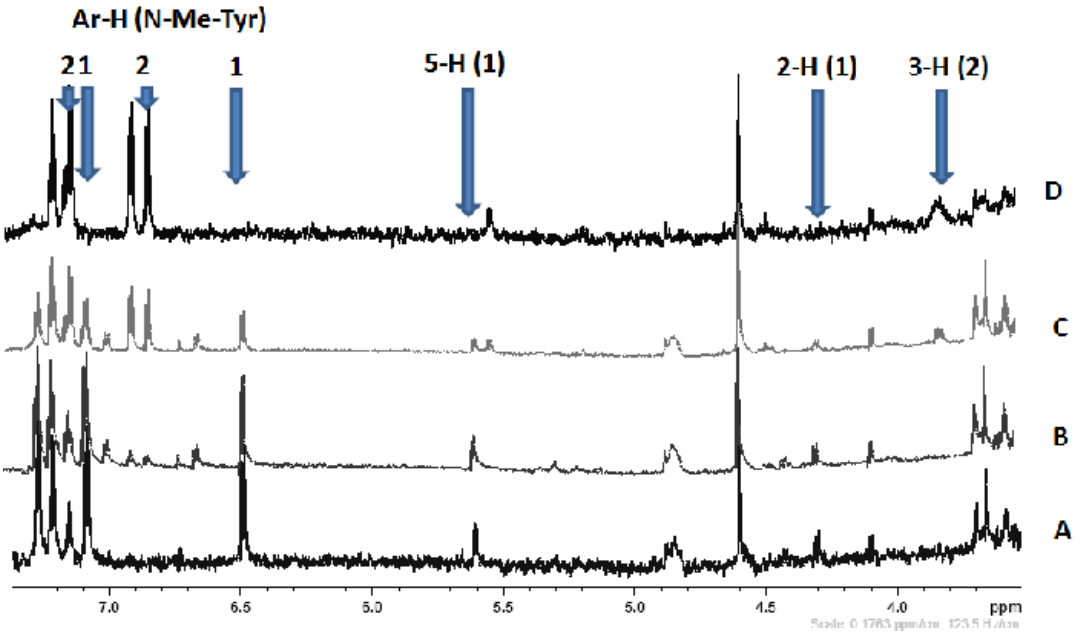


Figure S7-4: Chemical conversion from **1** to **2**: **A**: pure **1** after 10 min in MeOD, **B**: **1** with ca. 10 % of **2** after 24 h in MeOD, **C**: **1** with ca. 50 % **2** after 48 h in MeOD, **D**: pure **2**. Appearance or disappearance of significant signals is marked with arrows.

10.2 Supporting Information: Crocadepsins A and B, Depsipeptides from the myxobacterium *Chondromyces crocatus* discovered by a genome mining approach

Supporting information:

Crocadepsins A and B – Unique depsipeptides from myxobacterium *Chondromyces crocatus* explored by a genome mining approach

SI-1: Comparison of crocadepsin biosynthetic gene cluster from *Chondromyces crocatus* Cm c5 with highly similar biosynthetic gene cluster from *Chondromyces apiculatus* DSM436

***Chondromyces apiculatus* DSM436 sequence Source**
Submitter: IMTECH-CSIR; BioProjects: PRJNA192263
Assembly: GCA_000601485.1 Chondromyces apiculatus DSM 436 genome assembly scaffolds: 182 contigs: 182 N50: 120,949 L50: 31
Bacteria and source DNA available from Srikrishna Subramanian, Protein Science and Engineering, Institute of Microbial Technology, Chandigarh, India
Date: 2014/03/27
GenBank assembly accession: GCA_000601485.1 (latest)

SI-1.1: Organization of the gene clusters

A comparison of the crocadepsin biosynthetic gene cluster from *Chondromyces crocatus* Cm c5 with the homologous gene cluster found in the published genome of *Chondromyces apiculatus* DSM 436 shows high similarities. All ORFs from Cm c5 have a corresponding homologue, except for *cdpD*, encoding a putative MFS transporter gene, which is lacking in the DSM436 cluster. A homologue to EYF01035.1, a putative aminomutase, is not found in the Cm c5 genome. The ORFs up- and downstream of the clusters differ in both strains, confirming their respective boundaries. The encoded NRPS and PKS genes show identical domain organization to their respective homologue, and the predicted substrate specificity of their A-domains is almost identical. The NRPS product from DSM436 is not known, but expected to be highly similar.

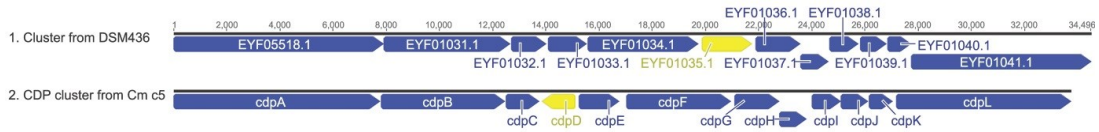


Figure SI-1.1: Graphical comparison of ORFs in the crocadepsin biosynthetic gene cluster and the homologous cluster from *Chondromyces apiculatus* DSM 436. ORFs with no corresponding homologue are shown in yellow, other ORFs are shown in blue.

Table SI-T1: Comparison of ORFs in the crocadepsin biosynthetic gene cluster and the homologous cluster from *Chondromyces apiculatus* DSM 436.

ORF in cm c5	proposed function	DMS436 homologue
cdpA	NRPS	EYF05518.1
cdpB	PKS	EYF01031.1
cdpC	cytochrome P450	EYF01032.1
cdpD	MFS transporter	-
cdpE	halogenase	EYF01033.1
cdpF	NRPS	EYF01034.1
-	aminomutase	EYF01035.1
cdpG	ABC transporter	EYF01036.1
cdpH	SyrP-like hydroxylase	EYF01037.1
cdpI	aminotransferase	EYF01038.1
cdpJ	methyltransferase	EYF01039.1
cdpK	arginase	EYF01040.1
cdpL	NRPS	EYF01041.1

SI-1.2: Substrate specificity of adenylation- and acyltransferase domains

Table SI-T2: Substrate specifity conferring Stachelhaus-codes extracted from A domains of crocadepsin NRPS genes using the NRPSpredictor2 webtool. The respective sequence homologues from DSM436 sequence were investigated, too. * The predicted substrate result from nearest-neighbor algorithm is shown, along with the sequence similarity to the closest database hit, given in percent.

A-Domain	extracted NRPS code	predicted specificity*	A-Domain	extracted NRPS code	predicted specificity*
CdpA-A1	DFWNIGMVHK	Thr 100%	**EYF05518.1-A1**	DFWNIGMVHK	Thr 100%
CdpA-A2	DASTIAAVCK	Tyr 100%	**EYF05518.1-A2**	DASTIAAVCK	Tyr 100%
CdpL-A3	DLTKVGEVGK	Asn 100%	**EYF01041.1-A3**	DLTKVGEVGK	Asn 100%
CdpL-A4	DTEHFGTPVK	beta-Lys 70%	**EYF01041.1-A4**	DTEHFGTPVK	beta-Lys 70%
CdpF-A5	DLRHLLFMAK	Ser 50%	**EYF01034.1-A5**	ELRHLAFMAK	Ser 40%

The specificity-conferring code of each A-domain was obtained from the NRPSpredictor2 webtool. [N2] Predicted substrate and observed incorporated building block in the crocadepsins at the respective positions agree for A-domains A1, A2 and A3. The modifications of these building blocks to dehydro-threonine, O-methyl-chlorotyrosin and ß-hydroxy-asparagine (OH-Asn) are carried out after loading. The substrate for A-domains A4 and A5 can not be clearly predicted from the sequences of both organisms. CdpL-A4 is expected to load either arginine-derived ß-methyl-ornithine (meOrn) or a precursor thereof, which could be arginine, ß-methyl-arginine or an α-keto-intermediate that is transiently formed prior to methylation.[IM] EYF01041.1-A4 has an identical binding pocket and is expected to load the same building block as CdpL-A4. Incorporation of α-keto-acids has been reported[CE] but is considered improbable after manual sequence analysis of CdpL-A4, because the the conserved catalytic A4 sequence motif "FDxS" is present, although slightly varied to FDTA, where the highly conserved aspartate residue is responsible for interaction with the α-amino-group of the substrate.[TW] The sequence of the α-keto-keto-acid activating A-domain CesA-A1 from the cereulid biosynthetic gene cluster contains a "HVGG" motif at this position, among other distinct features.[CS]

The substrate-prediction for A-domain A5 of CdpF or its homologue EYF01034.1 is also not clear, although aspartate is observed in the crocadepsin molecule at that position.

The substrate specificity of the AT-domains of CdpB and EYF01031.1 was investigated by multiple alignment with AT domains of known specificity. The specificity of AT domains can be predicted by identifying the residues at key positions within the domains' sequence (AT fingerprint) according to Mohanty et al.[MO] As a general rule, the QQGHS[LVIFAM]GR[FP]H[ANTGEDS][NHQ]V fingerprint sequence indicates malonate specificity, while QQGHS[QMI]GRSHT[NS]V indicates specificity for methylmalonate. The AT-domain of CdpB has the AT-fingerprint QQHSLGRAHNHV, its homologue EYF01031.1-AT has QQGHSMGRAHGHV. Both fingerprints do not clearly indicate either malonate or methylmalonate-specificity, which is in agreement with the observation that both precursors can be incorporated in crocadepsins.

SI-3: MS2-Analysis of isotope-labeled Crocadepsin A and B

Cultures of *Chondromyces crocatus* Cm c5 were cultivated in P-medium and supplemented individually with labeled amino acids D_3-methionine, $^{13}C_6$-arginine and $^{13}C_5{}^{15}N$-aspartate. After incubation, cultures were subsequently extracted and crocadepsin A and B isolated from each crude extract. The extracts were then subjected to ESI-MS/MS analysis. By comparison of the fragmentation pattern, the distribution of the methionine-derived methylations, the origin of me-Orn and incorporation of aspartic acid were investigated. The detected fragments are shown in table SI-T3.1.

Figure SI-3 shows the structure of crocadepsin A and B, the positions of labeled isotopes indicated with colors. The smallest fragment of the unlabeled crocadepsins with m/z = 392.18 contains me-Orn, since it shows incorporation of D3-Met and $^{13}C_6$-Arg (of which 5 carbons are incorporated). It does not contain the PKS-part formed from malonate or methylmalonate, because it is identical in both crocadepsin derivatives. It also does not contain the aspartate part of the molecule. Me-Orn is therefore derived from arginine and methylated by a SAM-dependent methyltransferase. The second-smallest fragment contains me-Orn, Asp, beta-hydroxy-asparagine (OH-Asn) and either acetate or propionate from the PKS. It does not contain the O-methyl-chlorotyrosine, since only one D_3-methionine label is present along with the $^{13}C_6$-arginine label. The other D_3-methionine label can therefore be assigned to O-methyl-chlorotyrosine. In extracts from the culture that was supplemented with $^{13}C_5{}^{15}N$-aspartate, this fragment was detected at equal intensity with the unlabeled fragment, which probably results from incomplete separation during workup.

Table SI-T3: Fragment analysis of isotope labeled crocadepsin A and B.

Crocadepsin A unlabeled	**Crocadepsin A D_3-Met**	**Crocadepsin A $^{13}C_6$-Arg**	**Crocadepsin A $^{13}C^{15}N$-Asp**	**fragment**
858.31	**864.30**	**863.30**	**860.30**	**[M+H]+**
840.31	846.34	845.31	842.31	[M-H2O+H]+
711.24	717.28	716.25	713.24	contains all labels
645.23	651.27	650.24	647.23	contains all labels
462.18	465.20	467.20	464.18 / 462.18	does not contain tyr
392.18	395.20	397.19	392.18	does not contain tyr, asp, ac/prop
Crocadepsin B unlabeled	**Crocadepsin B D_3-Met**	**Crocadepsin B $^{13}C_6$-Arg**	**Crocadepsin B $^{13}C^{15}N$-Asp**	**fragment**
872.31	**878.30**	**877.30**	**874.30**	**[M+H]+**
854.31	860.35	859.33	856.31	[M-H2O+H]+
725.24	731.29	730.21	727.25	contains all labels
659.23	665.28	664.26	661.24	contains all labels
476.18	479.22	481.21	478.20 / 476.20	does not contain tyr
392.18	395.20	397.19	392.18	does not contain tyr, asp, ac/prop

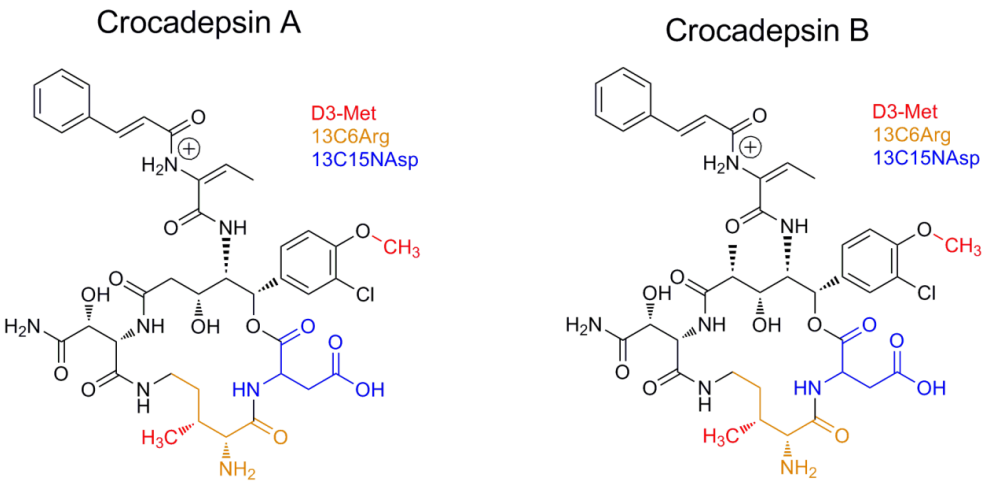


Figure SI-3.1: Chemical structure of crocadepsin A and B $[M+H]^+$ ions. The positions of labeled isotopes are indicated with colors

The MS/MS spectra of purified labeled Crocadepsins are shown in the figures below. The fragments dicussed above are highlighted in red.

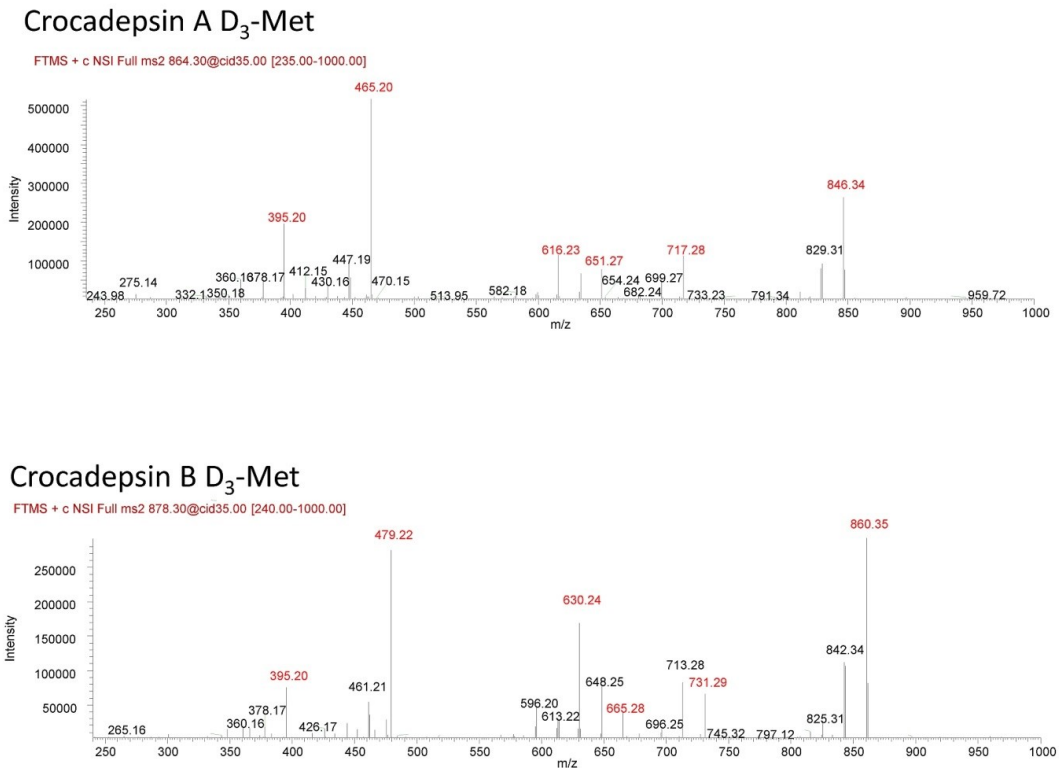


Figure SI3.2: MS/MS spectra of purified, D3-methionine labeled Crocadepsins A and B.

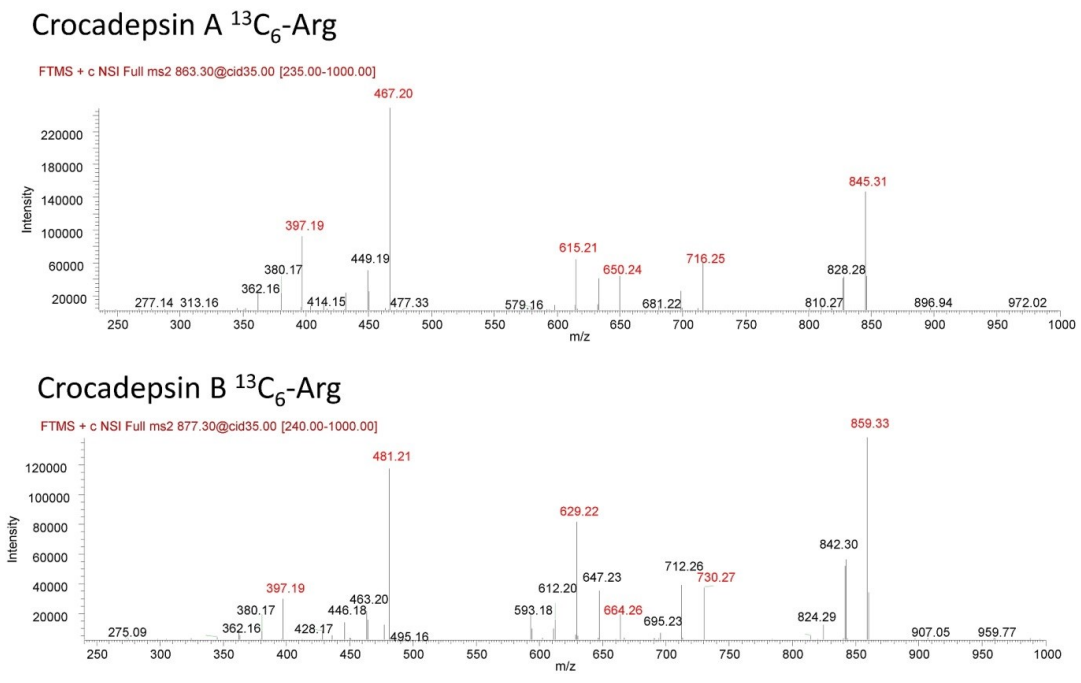


Figure SI3.3: MS/MS spectra of purified, $^{13}C_6$-arginine labeled Crocadepsins A and B.

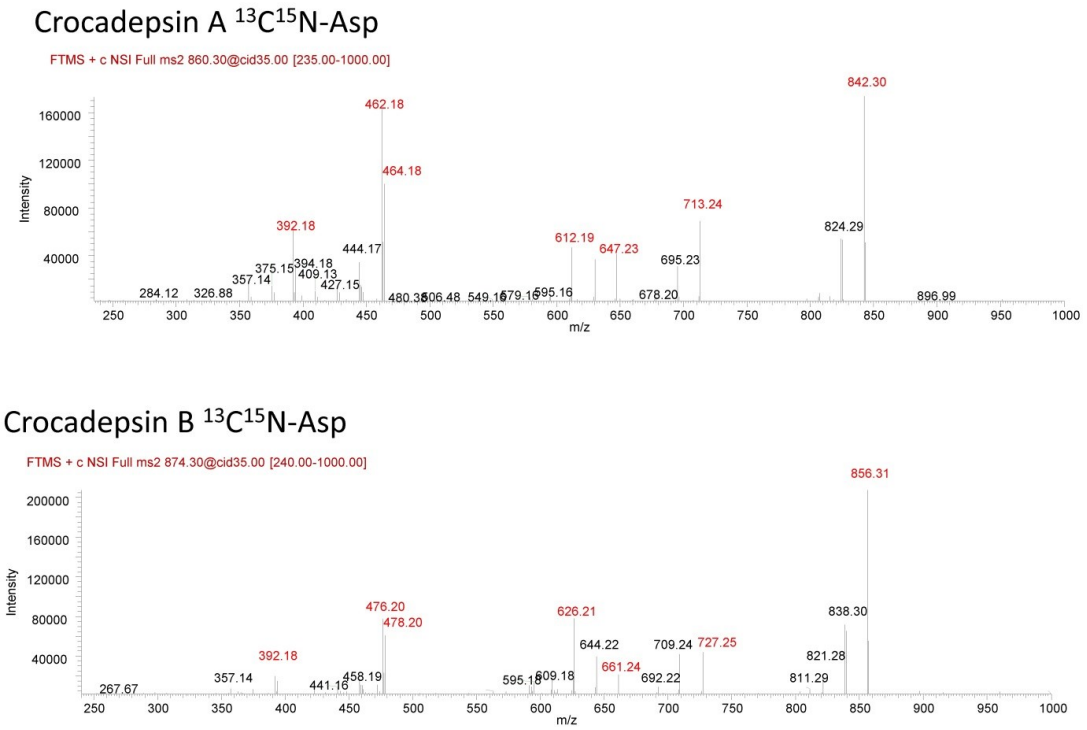


Figure SI3.4: MS/MS spectra of purified, $^{13}C^{6}N$-aspartate labeled Crocadepsins A and B.

SI-4: Mutagenesis of Cm c5, clone verification and extraction

Derivatives of the plasmid pSUP_*hyg* were used for mutagenesis of Cm c5. When a suffiently long homologous region is inserted into pSUP_*hyg*, the plasmid can integrate into the Cm c5 chromosome at the desired locus *via* a single crossing-over event. The resulting mutant is then hygromycin resistant. By cloning an internal fragment of a gene of interest into the pSUP_*hyg* cloning site, the gene is disrupted by the plasmid backbone as shown in fig. SI-4.1A, leading to a knock-out mutant. Cloning the 5' region of a gene of interest fused to a promotor leads to a mutant in which the target gene is driven by the introduced promotor, to study overexpression of a gene of interest, as shown in fig. SI-4.1B.

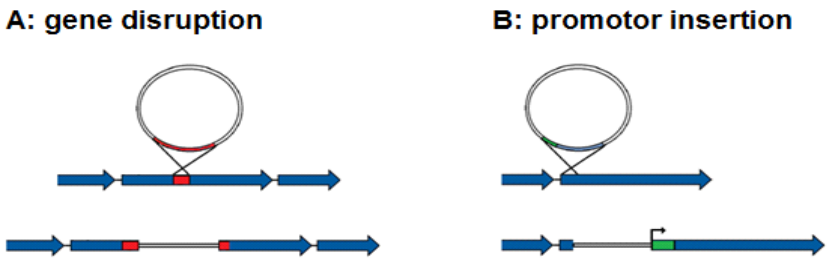


Figure S4.1A: Gene inactivation by single-crossing over mutagenesis in Cm c5 with the integrative pSUP_hyg plasmid. A: The target gene is disrupted by targeted integration of pSUP_hyg into the coding region via the homologous fragment shown in red. **Figure S4.1B:** A homologous fragment containing a promotor sequence is used to create a Cm c5 mutant with the target gene under control of the inserted promotor.

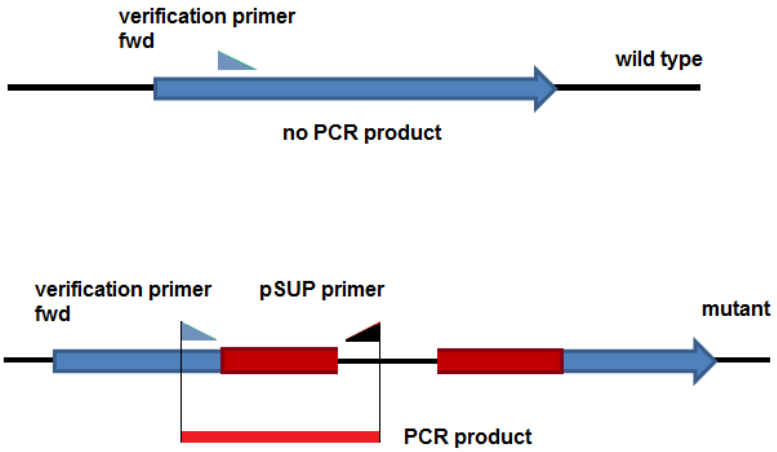


Figure SI-4.2: Genomic organization of target genes in the Cmc5 inactivation mutants. Blue bar/arrow: 5' and 3' regions of target gene inactivated by insertional mutagenesis using the respective pSUP_hyg derivatives. Light blue arrowheads: Position of primers used in a control PCR to verify disruption of the wild type allele of the target gene in the mutants.

To genetically verify the genotype of the isolated mutants, gDNA of mutant cells grown in liquid culture was extracted and tested with PCR reactions. The negative control PCR was performed with the respective verification primers that bind in the vicinity of the region of plasmid insertion. This reaction will only yield a product if no insertion has taken place, as shown in Fig. SI-4.2. In the mutant, the target gene is disrupted by the 8 kb plasmid backbone and no PCR product is obtained.

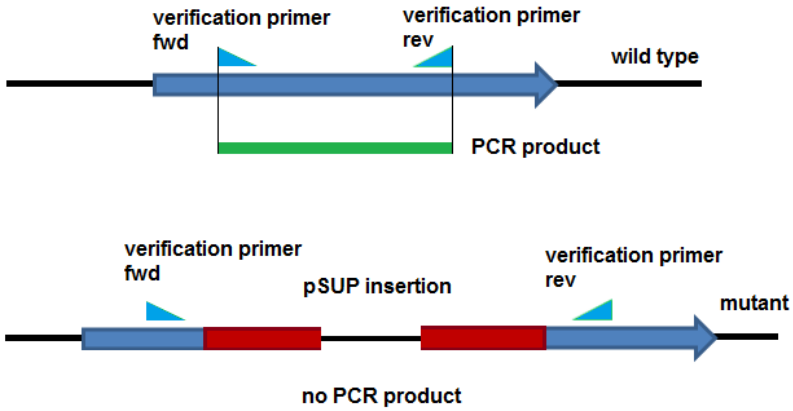


Figure SI-4.3: Genetic verification of Cmc5 mutants by PCR. Arrowheads: Position of primers used in a control amplification (blue: primers in the genomic up- and downstream regions; black: primer in the pSUP_hyg backbone) in order to verify disruption of the wild type allele of the target gene in the mutants.

The second PCR was done using one of the respective verification primers combined with a primer that binds the plasmid backbone. The correct combination of these primers leads to a product only if the pSUP_*hyg* plasmid backbone has indeed integrated at the desired position as shown in Fig. SI-4.3. Figure SI-4.3 shows the verification PCR for mutants of cdpF, cdpC, cdpG and cdpK. Figure SI-4.4 shows the negative control PCR for mutants of cdpF, cdpC, cdpG and cdpK. As negative control, gDNA template from Cm c5 wild type was used.

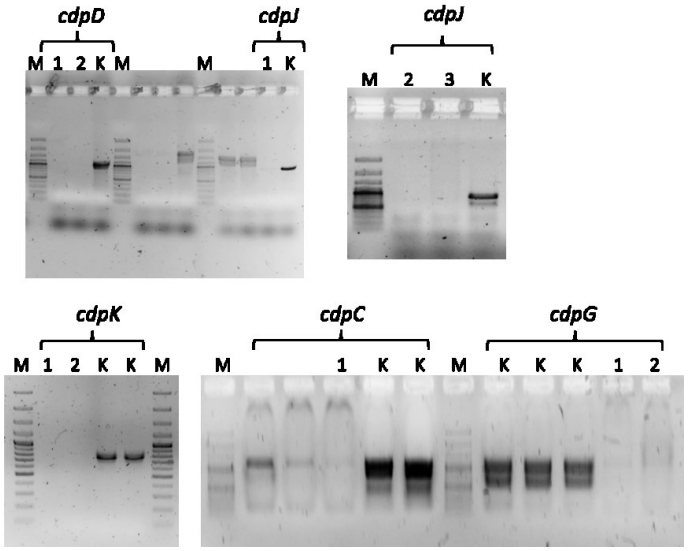


Figure SI-4.4: Genetic verification of Cmc5 mutants by PCR. DNA from Cm c5 mutants was tested with primer combinations as shown in figure Figure SI-4.3. Clones not yielding the PCR product of the positive control were considered correct. M: DNA ladder; K: positive control template; 1, 2: clone number.

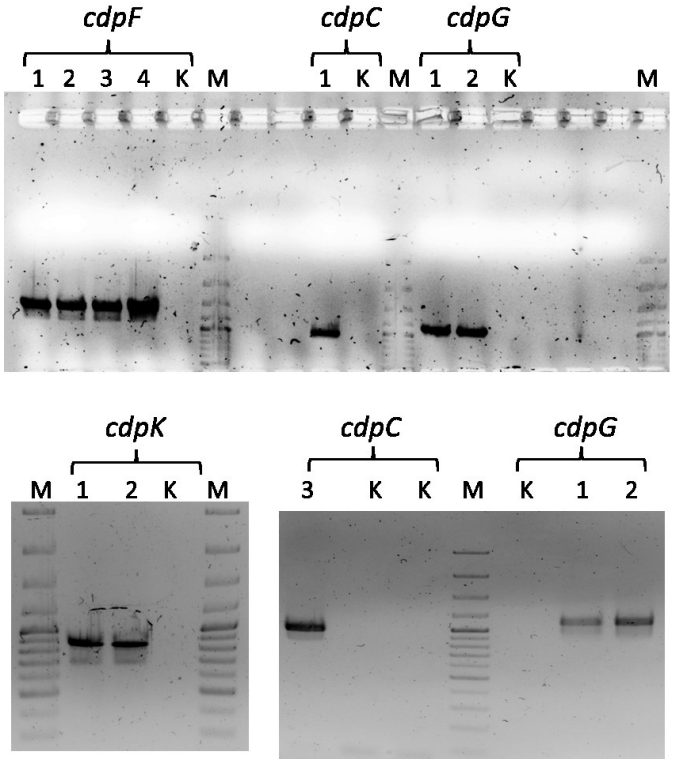


Figure SI-4.5: Genetic verification of Cmc5 mutants by PCR. DNA from Cm c5 mutants was tested with primer combinations as shown in figure Figure SI-4.2. Clones yielding the PCR product were considered correct. M: DNA ladder; K: negative control template; 1, 2, 3: clone number.

The 2 mutants of gene *cdpE* were not be verified by this method, and were considered correct after extraction and MS/MS analysis. Their chromatograms are attached below. Notably, one clone exclusively produced crocadepsin C, while the other produced crocadepsin A and C in equal amounts.

Control extracts from an independent mutant are presented for comparison. The absence of the typical satellite masses that indicate chlorinated compounds further indicates that the mass signal with $[M+H]^+$ = 838.4 indeed corresponds to non-chlorinated crocadepsin.

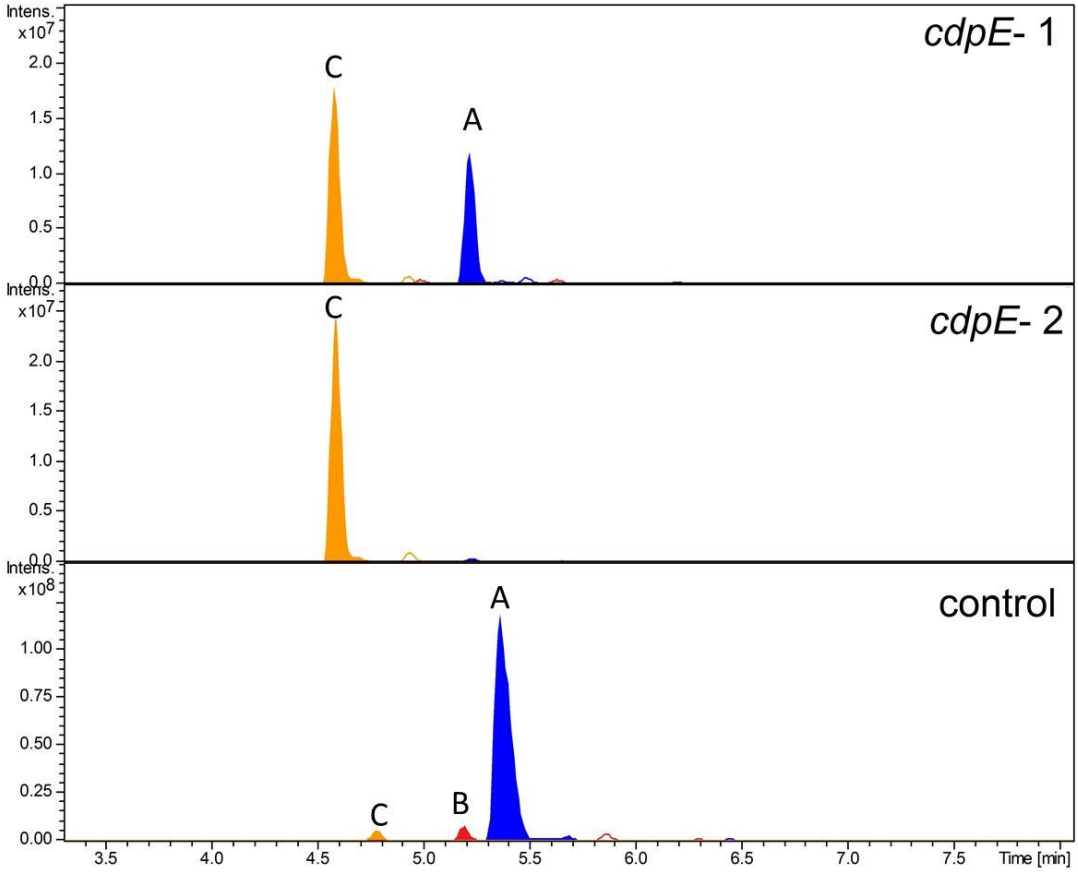


Figure SI-4.6: Extracted ion chromatograms of three individual Cm c5 mutant clones, showing mass peaks of crocadepsins A (872.3 Da), B (852.3 Da) and C (838.4 Da)

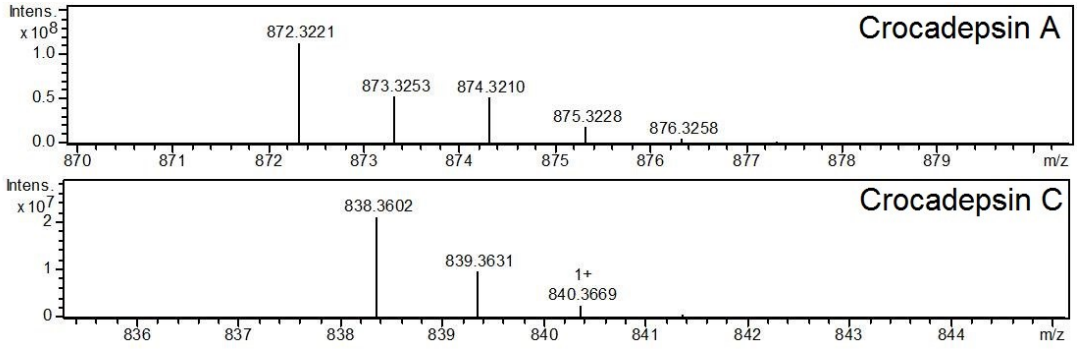


Figure SI-4.7: Mass spectra of crocadepsin A and C. Absence of the chlorine atom in crocadepsin C is indicated through the absence of the satellite isotopes, which are present in crocadepsin A.

References:

[N2] M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic acids research* **2011**, *39*, W362-367.

[IM] U. Hornemann, L. H. Hurley, M. K. Speedie, H. G. Floss, *Journal of the American Chemical Society* **1971**, *93*, 3028-3035.

[CE] N. A. Magarvey, M. Ehling-Schulz, C. T. Walsh, *Journal of the American Chemical Society* **2006**, *128*, 10698-10699.

[TW] G. L. Challis, J. Ravel, C. A. Townsend, *Chemistry & biology* **2000**, *7*, 211-224.

[CS] M. Ehling-Schulz, M. Fricker, H. Grallert, P. Rieck, M. Wagner, S. Scherer, *BMC microbiology* **2006**, *6*, 20.

[IM] G. Yadav, R. S. Gokhale, D. Mohanty, *Journal of molecular biology* **2003**, *328*, 335-363.

10.3 Supporting Information: Disciformycins A and B, Unprecedented 12-membered Macrolide-Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* active against multiresistant *Staphylococci*

Disciformycins A and B, unprecedented 12-membered Macrolide-Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* active against multiresistant *Staphylococci*

Supporting Information

Frank Surup, Konrad Viehrig, Kathrin I. Mohr, Jennifer Herrmann, Rolf Jansen, Rolf Müller*

rom@mx.uni-saarland.de

TABLE OF CONTENTS

General Experimental Procedures: Optical rotations were determined with a Perkin-Elmer 241 polarimeter, UV spectra were recorded with a Shimadzu UV-Vis spectrophotometer UV-2450. NMR spectra were recorded with a Bruker ARX 600 (^{1}H 600 MHz, ^{13}C 150 MHz) and a Bruker Ascend 700 spectrometer with a 5 mm TXI cryoprobe (^{1}H 700 MHz, ^{13}C 175 MHz). HRESIMS mass spectra were obtained with an Agilent 1200 series HPLC-UV system combined with an ESI-TOF-MS (Maxis, Bruker) [column 2.1×50 mm, 1.7 µm, C_{18} Acquity UPLC BEH (Waters), solvent A: H_2O + 0.1 % formic acid; solvent B: AcCN + 0.1 % formic acid, gradient: 5 % B for 0.5 min increasing to 100 % B in 19.5 min, maintaining 100 % B for 5 min, F_R = 0.6 mLmin^{-1}, UV detection 200-600 nm].

Cultivation A (10 L scale): *Pyxidicoccus fallax* AndGT8 was cultivated in 10 L of medium E (per liter: skimmed milk 4 g; soy meal 4 g; yeast extract 2 g; starch 10 g; $MgSO_4$ x 7 H_2O 1 g; Fe-EDTA 8 mg, glycerol 5 g; behenyl alcohol 35 mg) in a biofermenter b10 (Aktiengesellschaft für Biotechnologische Verfahrenstechnik, Schweiz) at 30 °C for 216h. The pH was regulated with potassium hydroxide (2.5 %) and sulfuric acid between 7.3 and 7.5. The stirrer speed was 100 – 400 rpm, aerated with 0.05 vvm compressed air. The dissolved oxygen content within the fermentation broth was regulated by the stirrer speed to pO_2 20 %. Fermentation was carried out with addition of 1 % adsorber resin XAD-16.

Purification A: For isolation of the active metabolites, XAD-16 was harvested by centrifugation; cells were separated from XAD-16 by flotation and discarded. The XAD was eluted with 750 mL acetone and 750 mL methanol, separately. Bioassays against *S. aureus* showed that the antibacterial activity was concentrated in the acetone extract. The organic solvent of the acetone fraction was evaporated *in vacuo* until mostly water remained. This was extracted twice with ethylacetate. The ethylacetate was removed in vacuo; the material was dissolved in 85 % aqueous methanol and extracted twice with heptane; subsequently the lower methanol phase was adjusted to 70 % methanol and extracted twice with dichloromethane. Bioassays revealing the water fraction contained the main antibacterial activity, antibiotic metabolite **1** was isolated in a bioassay guided fractionation strategy. The extract was fractionated by RP MPLC (column 480 x 30mm, ODS/AQ C18 (Kronlab), gradient 37 % to 100 % methanol in 60 min, flow 30 mL/min, UV peak detection at 210 nm). Active fractions were combined and further fractionated by preparative RP HPLC column 250 x 21 mm, VP Nucleodur C18 Gravity 5 µm, gradient 37 % to 100 % methanol in 25 min, 50 mM sodium acetate, flow 20 mL/min). A final step of preparative RP HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5 µm, gradient 28 % to 55 % methanol in 25 min, 0.5 % acetic acid, flow 20 mL/min) provided 1.0 mg of **1** as bioactive principle.

Cultivation B (70 L scale): *Pyxidicoccus fallax* AndGT8 was cultivated in 70 L of medium P (per liter: peptone [Marcor] 2 g, starch 8 g, probion single cell protein 4 g, yeast extract 2 g, $CaCl_2$ 1 g, $MgSO_4$ 1 g, Fe-EDTA 8 mg, pH 7.5) in a biofermenter P150 (Bioengineering AG, Wald CH) at 30 °C for 148h. The fermentation parameters were as described for the 10L fermentation.

Purification B: For isolation of **1**, XAD-16 and wet cell mass were harvested by centrifugation. Combined cells and XAD were washed with 30 % methanol (1.5 L), and subsequently extracted with methanol (4 L) and acetone (1 L). Methanol and acetone extracts were combined, evaporated and subjected to a solvent partition. The extract was dissolved in 70 % aqueous methanol and extracted twice with dichloromethane. The obtained material (10.0 g) was fractionated on silica gel 100 (0.063-

0.200 mm, approximately 1 kg) by a gradient from dichloromethane/methanol 98:2 to 8:2 (98:2, 95:5, 92.5:7.5 90:10, 80:20 500-750 mL each). The bioactive fraction "90:10" (219.2 mg) was subjected to preparative RP HPLC (column 250 x 21 mm, VP Nucleodur C18 Gravity 5 µm, gradient 30 % to 55 % acetonitrile in 25 min, flow 20 mL/min) and afforded 74.3 mg of material containing **1** and **2**. A final step of silica gel HPLC (column 250 x 21 mm, Nucleosil 100 7 µm, isocratic *tert*-butylmethylether/heptane 1:3 with methanol 2 %, flow 25 mL/min) provided 25.4 mg of **1** and 7.6mg of **2**.

Disciformycin A (1): colorless, amorphous powder, $[\alpha]^{20}_{D}$ –51 (*c* 0.1, MeOH); UV (MeOH) λ_{max} 229 (sh); IR (KBr) 3431, 2961, 2929, 2874, 1731, 1647, 1455, 1383, 1254, 1193, 1124, 1076, 1006, 873 cm^{-1}; ^{1}H, ^{13}C, COSY, HMBC and ROESY NMR data see table S1; HRESIMS *m/z* 547.2510 $[M+Na]^{+}$ (calcd for $C_{27}H_{40}O_{10}Na$, 547.2514).

Table S1. NMR spectroscopic data of disciformycin A (**1**) in CH_3OH-d_4 (^{1}H at 600 MHz, ^{13}C at 150 MHz).

#	^{13}C, mult.	^{1}H, mult.	COSY	HMBC	ROESY
1	169.1, C				
2	134.5, C				
3	130.0, CH	5.88, ddq (8.1, 7.3, 1.5)	4a, 4b	4, 15	15>4b
4	44.0, CH_2	4.01, dd (18.3, 8.1)	3, 4b, 15	2, 3, 5	4b, 16
		3.58, dd (18.3, 7.3)	3, 4a		4a
5	203.9, C				
6	80.8, CH	5.05, d (9.5)	7	5, 7, 8, 1'	16, 4a >4b
7	84.2, CH	4.26, d (9.5)	6	5, 6, 8, 9, 1''	9, 1'' > 16
8	134.9, C				
9	128.8, CH	5.32, brd (11.0)	10a, 10b, 16	7, 16	7, 10b, 11
10	31.8, CH_2	2.82, ddd (14.7, 11.7, 11.0)	9, 10b, 11	8, 9, 11	10b, 16, 17
		2.13, m	9, 10a, 11, 16		9, 10a, 11
11	75.4, CH	5.58, dd (11.7, 3.3)	10a, 10b	1, 9, 10, 12, 13,17	10b, 14>17
12	134.9, C				
13	123.6, CH	5.41, qq (6.6, 1.5)	14, 17	11, 14, 17	17
14	13.1, CH_3	1.74, dq (6.6, 1.5)	13, 17	12, 13	11
15	21.2, CH_3	1.96, brs	3, 4a	1, 2, 3	3
16	12.3, CH_3	1.91, dd (1.5, 1.5)	9, 10b	7, 8, 9	6, 10a
17	18.3, CH_3	1.65, qd (1.5, 1.5)	13, 14	11, 12, 13	13
1'	174.5, C				
2'	43.4, CH_2	2.30, d (7.3)	3'	1', 3', 4'/5'	4'/5'
3'	27.0, CH	2.11, m	2', 4'/5'	1', 2', 4'/5'	
4'/5'	22.7, CH_3	1.00/1.01, d (6.6)	3'	2', 3', 4'/5'	2'
1''	110.6, CH	5.13, brs	2''	7, 3'', 4''	7
2''	83.8, CH	4.06, m	1'', 3''	1'', 3'', 4''	
3''	79.1, CH	3.88, dd. (5.9, 2.9)	2'', 4''	1'', 2'', 5''	
4''	86.3, CH	3.94, td (5.7, 3.3)	3'', 5a'', 5b''	2'', 3''	
5''	63.0, CH_2	3.73, dd (12.0, 3.3)	4'', 5b''	3'', 4''	
		3.64, dd (12.0, 5.7)	4'', 5a''	3'', 4''	

Disciformycin B (2): colorless, amorphous powder, $[\alpha]_D$ +64 (*c* 0.4, MeOH); UV (MeOH) λ_{max} 221 (sh), 240 (3.67); IR (KBr) 3429, 2961, 2936, 2875, 1738, 1704, 1627, 1455, 1379, 1299, 1178, 1074, 1031, 856 cm^{-1}; ^1H, ^{13}C, COSY, HMBC and ROESY NMR data see Table S2; HRESIMS *m/z* 547.2516 $[M+Na]^+$ (calcd for $C_{27}H_{40}O_{10}Na$, 547.2514).

Table S2. NMR spectroscopic data of disciformycin B (**2**) in $CHCl_3$-*d* (^1H at 700 MHz, ^{13}C at 175 MHz).

#	^{13}C, mult.	^1H, mult.	COSY	HMBC	ROESY
1	171.5, C				
2	43.0, CH	3.34, dq (9.3, 6.6)	3, 15	3, 4, 15	4, 15
3	145.9, CH	6.59, dd (15.3, 9.3)	2, 4	2, 4, 5, 15	15>>2
4	129.9, CH	6.37, d (15.3)	3	2, 3, 5	2, 6, 16
5	192.1, C				
6	78.2, CH	5.32, d (10.3)	7	5, 7, 8, 1'	4, 16
7	81.1, CH	4.09, d (10.3)	6	5, 6, 8, 9, 16, 1''	9, 1''
8	133.3, C				
9	129.4, CH	5.42, m	10a, 10b	8, 11	7, 10b
10	32.1, CH_2	2.87, ddd (14.6,11.7,11.0)	9, 10b	8, 9, 11	10b> 16, 17
		2.04, m	9, 10a	8, 9	9, 10a
11	72.8, CH	5.33, dd (11.7, 2.8)	10a, 10b	1, 9, 10, 12, 13, 17	10b, 14
12	133.0, C				
13	123.4, CH	5.40, m	14	11, 12, 14, 17	14, 17
14	13.0, CH_3	1.71, dq (6.9, 1.3)	13	12, 13	11
15	14.1, CH_3	1.27, d (6.6)	2	1, 2, 3	2
16	12.5, CH_3	1.91, brs	9	7, 8, 9	6, 10a
17	18.0, CH_3	1.69, dq (1.5, 1.3)		11, 12 ,13	10a, 13
1'	172.5, C				
2'	42.8, CH_2	2.36, dd (14.8, 7.3)	3'	1', 3', 4'/5'	4'/5'
		2.31, dd (14.8, 7.3)			
3'	25.8, CH	2.16, tsept. (7.3, 6.8)	2', 4'/5'	1', 2', 4'/5'	4'/5
4'/5'	22.4, CH_3	1.02, d (6.8)	3'	2', 3', 4'/5'	2', 3'
1''	108.2, CH	5.19, m	2''	7, 3'', 4''	7
2''	78.2, CH	4.03, m	1'', 3''	1'', 3''	
3''	78.3, CH	4.03, m	2'', 4''	2'', 5''	
4''	88.1, CH	4.13, m	3'', 5''a, 5''b	1'', 3''	16, 5''
5''	62.1, CH_2	3.89, dd (11.6, 2.6)	4'', 5''b	3''	
		3.83, dd (11.6, 1.7)	4'', 5''a	3'', 4''	

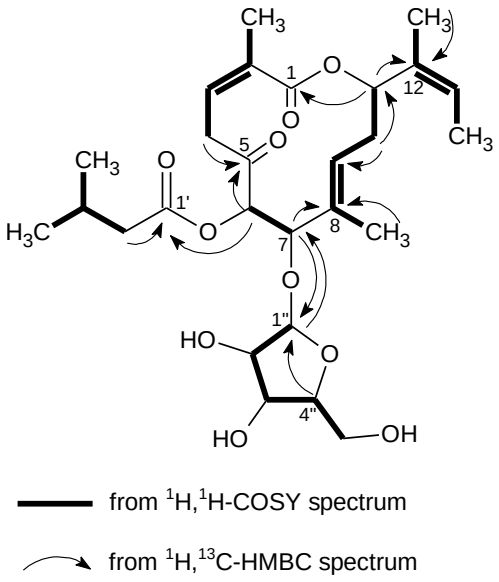


Figure S1: Selected COSY and HMBC correlations indicating the structure of disciformycin (**1**).

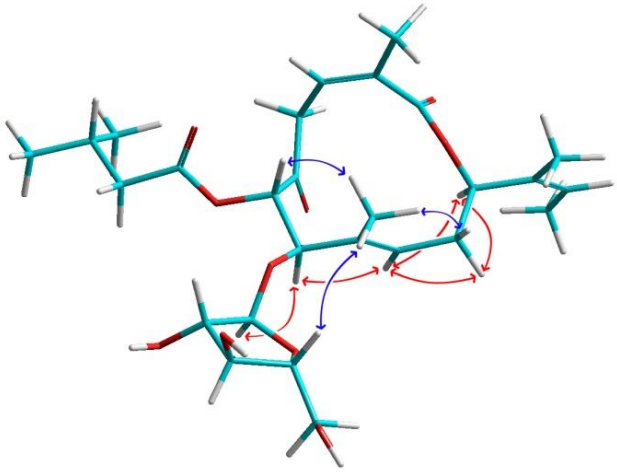


Figure S2. Model of **1** calculated by PM3 with HyperChem with selected ROESY correlations (blue above, red below main plain) indicating the conformation of **1**.

Table S3: Minimum inhibitory concentration (MIC for bacteria, yeasts and fungi) and half-inhibitory concentrations (IC_{50} for cell lines) in µg/mL of Disciformycin A (**1**) and Disciformycin B (**2**). For MICs, a total of 2 µL and 20 µL of a 1 mg/ml stock solution (6.7 and 67 µg/ml) were tested. Cell density was adjusted to about 5 x 10^6/ml. 20 µl MeOH were tested as negative control and showed no activity against the selected test organisms. For IC_{50} values, cells were seeded in 96-well microtiter plates (6 x 10^3 cells/well) and treated with **1** and **2** for 5 d. For bacteria: [a] Oxytetracyclin hydrochloride, [b] Gentamycin, [c] Vancomycin; [d] for yeasts and fungi: Nystatin; n. i. = no inhibition up to 67 µg/mL

		DSM	**1**	**2**	Ref. [a,b,c]
Cell lines	L929	**ACC-2**	> 10	> 10	/
	HCT-116	**ACC-581**	> 10	> 10	/
	CHO-K1	**ACC-110**	> 10	> 10	/
Gram +	*Bacillus subtilis*	**10**	4.2	0.83	4.2 [a]
	Paenibacillus polymyxa	36	16.6	16.6	1.7 [a]
	Staphylococcus aureus	346	16.6	3.3	0.21 [a]
	Staphylococcus aureus Newman	/	8.0	1.2	/
	Staphylococcus aureus (MRSA)	11822	4.0	0.6	0.83 [c]
	Staphylococcus aureus N315 (MRSA)	/	8.0	1.2	/
	Staphylococcus aureus Mu50 (MRSA/VRSA)	/	2.0	0.6	/
	Staphylococcus carnosus	20501	7.8	2.4	/
	Mycobacterium sp.	43270	n.i.	n.i.	0.52 [a]
	Mycobacterium diernhoferi	43524	33.3	33.3	≤ 0.052 [a]
	Micrococcus luteus	20030	67	n.i.	0.42 [a]
	Nocardioides simplex	20130	33.3	16.6	3.3 [a]
Gram -	*Pseudomonas aeruginosa*	50071	n.i.	n.i.	42.0 [b]
	Chromobacterium violaceum	30191	n.i.	n.i.	0.83 [a]
	Escherichia coli	1116	n.i.	n.i.	0.83 [a]
yeasts	*Schizosaccharomyces pombe*	70572	n.i.	n.i.	33.3 [d]
	Rhodotorula glutinis	10134	n.i.	n.i.	33.3 [d]
	Candida albicans	1665	n.i.	n.i.	67 [d]
	Pichia anomala	6766	n.i.	n.i.	33.3 [d]
fungi	*Mucor hiemalis*	2656	n.i.	n.i.	67 [d]

Comparison of observed incorporation and predicted substrate specificity of AT-domains: Assembly of the disciformycin polyketide core (see figure 3) is expected to start at carbon atom C14 with incorporation of malonate (mal) and extension by methylmalonate (mmal), mal, mmal, mal, mal, mmal. The polyketide is released as a macrolide, by formation of an ester bond between the carboxy-group at C1 and the hydroxy-group at C11. The order of the PKS modules is DifG, DifB, DifC, DifD, DifE, DifF. The specificity of the AT domains within the module sequence was predicted by sequence alignment based on occupation of critical postion (AT fingerprint) according to Mohanty et al [1]. The comparison is shown in table S1. The AT domains of modules DifB, DifF, DifE and AT1 of module DifG have the QQGHS[LVIFAM]GR [FP]H[ANTGEDS][NHQ]V motif characteristic for malonate specificity, while the AT domainss of modules DifC, DifF and AT2 from DifG have the methylmalonate-specific motif QQGHS[QMI]GRSHT[NS]V. The predicted substrate specificity agrees to the structure of the observed polyketide.

Table S4: Alignment of the AT domain sequences extracted from disciformycin PKS module sequences. Pos. A refers to the position corresponding to the prototypical *E. coli* FAS structure, while Pos. B refers to the position whithin the AT alignment. Predicted substrate specificity and observed incorporation agree for all positions.

Pos A	**11**	**63**	**90**	**91**	**92**	**93**	**94**	**117**	**200**	**201**	**231**	**250**	**255**		
Pos B	**10**	**65**	**91**	**92**	**93**	**94**	**95**	**118**	**204**	**205**	**240**	**259**	**264**	**predicted**	**observed**
DifB	**Q**	**Q**	**G**	**H**	**S**	**L**	**G**	**R**	**F**	**H**	**T**	**N**	**V**	mal	mal
DifC	**Q**	**Q**	**G**	**H**	**S**	**M**	**G**	**R**	**S**	**H**	**T**	**N**	**V**	mmal	mmal
DifD	**Q**	**Q**	**G**	**H**	**S**	**V**	**G**	**R**	**F**	**H**	**T**	**N**	**V**	mal	mal
DifE	**Q**	**Q**	**G**	**H**	**S**	**L**	**G**	**R**	**F**	**H**	**S**	**N**	**V**	mal	mal
DifF	**Q**	**Q**	**G**	**H**	**S**	**M**	**G**	**R**	**S**	**H**	**T**	**N**	**V**	mmal	mmal
DifG1	**Q**	**Q**	**G**	**H**	**S**	**V**	**G**	**R**	**Y**	**H**	**S**	**N**	**V**	mal	mal
DifG2	**Q**	**Q**	**G**	**H**	**S**	**M**	**G**	**R**	**S**	**H**	**T**	**N**	**V**	mmal	mmal

Comparison of observed and predicted stereoselectivity of KR-domains: The polyketide core of the compound contains three stereocenters and three double bonds (see Figure 1). The hydroxy-group in *S*-configuration at position C6 is expected to result from monooxygenase activity and not from keto-reduction, and is therefore excluded from the analysis. The configuration of hydroxy-groups and double bonds that result from KR activity is determined by the respective KR domain that is responsible for reduction of the keto group during metabolite assembly. Stereoselectivity can be predicted from the primary sequence of the KR-domain, since most KR domains group into two distinct types, which accept their substrate in different ways and thus produce a secondary alcohol of opposite configuration[2]. Briefly, A-type ketoreductase activity results in an *S*-configured alcohol, while B-type ketoreductases yield *D*-configured alcohols. For double bonds, which result from elimination of KR-derived hydroxy-groups, the configuration of the cryptic hydroxyacyl intermediate can be predicted by checking the ketoreductase domain sequence of the respective module. While most B-type KR-domains show a conserved “LDD”-motif in the loop region of the enzyme, A-type KRs lack this motif. Instead they often contain a conserved Trp-residue in the catalytic centre[1]. By

multiple alignments of the five KR sequences of the *dif*-cluster, loop region and catalytic region where investigated for presence of the key residues, as shown in table S2. The KR-domain of DifB is clearly A-type, because the LDD motif is absent and the critical Trp in the catalytic centre is present. KR-domainss of DifC, DifD and DifF are all B-type, with the conserved LDD-motif being replaced by "LED" in DifD and "LQD" in DifF. The KR sequence of DifG shows neither the conserved LDD-motif nor the conserved Trp in the catalytic region. A prediction of its stereoselectivity is therefore not possible by this method.

Observed and predicted stereochemistry agree for the KRs of DifC and DifD. For the DifG KR, no clear predicition of the cryptic hydroxyacyl intermediate is possible. Reduction by DifB, which is clearly A-type, should lead to an S-configured OH-group at C11, but an R-configured OH-group is observed here. The double bond indroduced by module DifF is either trans-α,β in Disciformycin A or cis-β,γ in Disciformycin B and the DifF KR-domain is predicted to be B-type. Therefore, the trans-α,β double bond of 1 and the cis-β,γ double bond of 2 should result from an *R*-configured alcohol.

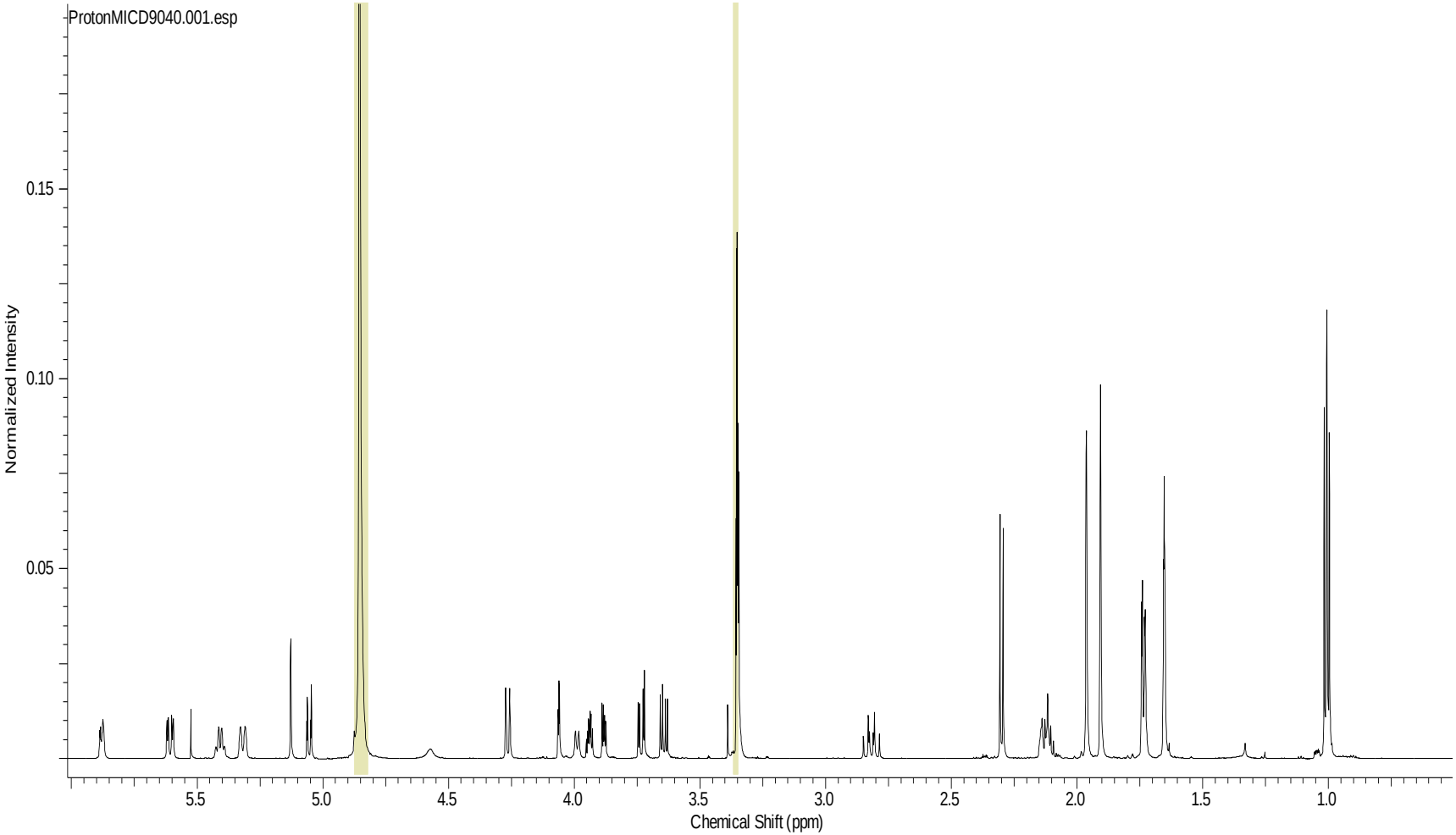
Table S5: Alignment of the KR domain sequences from the Disciformycin PKS and comparison of predicted and observed product stereochemistry. Indicator residues in the sequence are shown in blod letters. KR domains of DifC, DifD and DiF show the "L[DEQ]D" motif in the loop region, which is tyipcal for B-type KRs. DifB contains the conserved Trp residue in the catalytic region, indicating an A-type KR. DifG lacks indicating residues.

	Loop region	Catalytic region	KR type	predicted	observed
DifB	HAAGISREVPL	SSGASV**W**GSAGLAP**Y**AAANE	A	*S*-OH	*R*-OH
DifC	HLAGV**LDD**ATL	SAGATLLGSPGQAN**Y**AAVNA	B		*E*
DifD	HAAGV**LED**GVL	SSATATLGAAGQTN**Y**AAANA	B	*R*-OH	*R*-OH
DifF	HCAGV**LQD**GAL	SSVSSMLGSPGQGN**Y**AAANA	B		*Z*/ *E* shifted
DifG	HLAGVIRPELL	SSASAVLNSPLLGS**Y**AAANA	-	-	*Z*

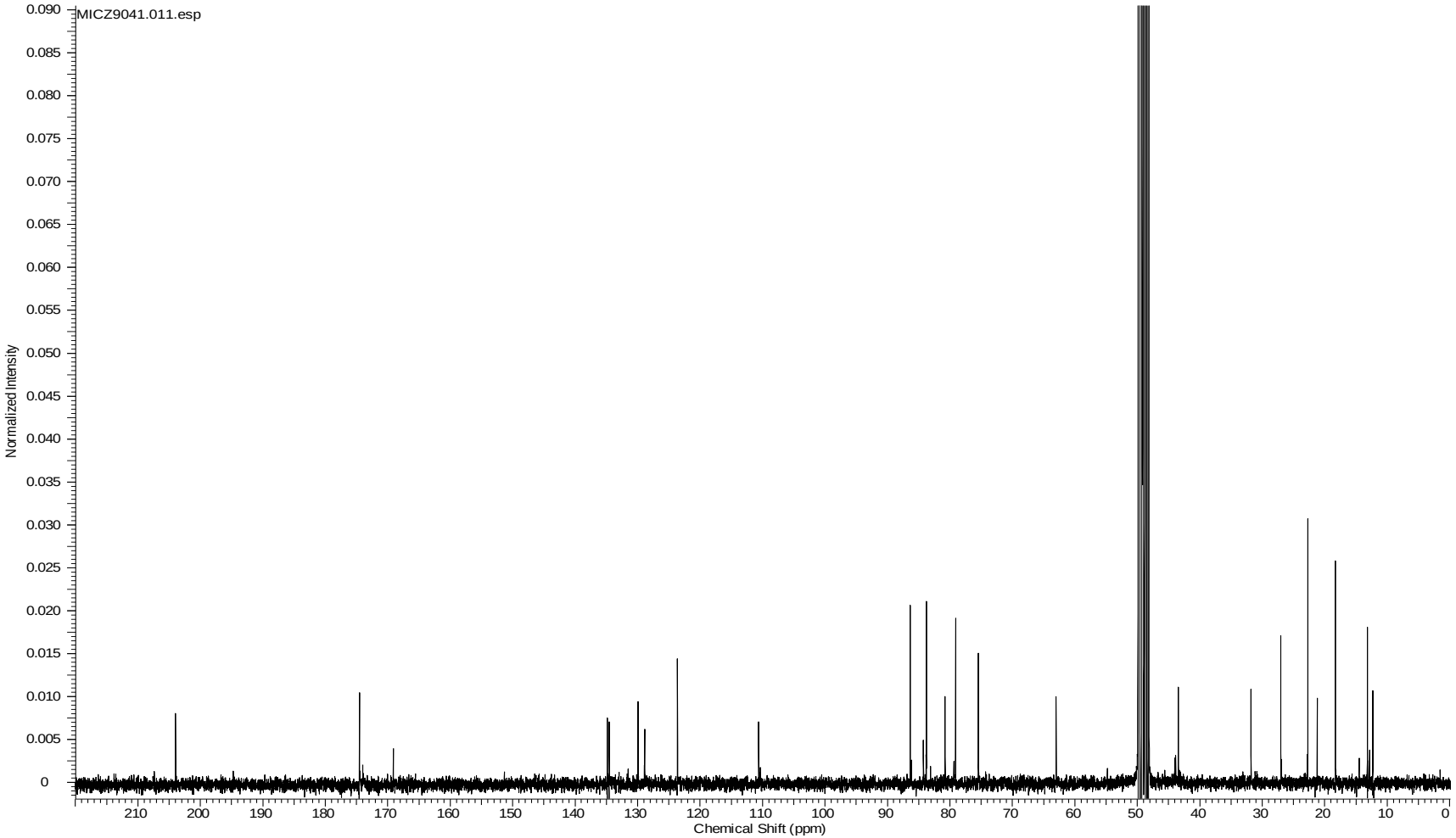
Domain of unknown function in DifF: The last module of the Disciformycin assembly line, DifF contains domain of unknown function in its sequence, between the DH and KR domain. This position is usually occupied by an ER-domain in fully reductive modules. Indeed, the Rossman fold for NADP(H)-binding is present in the sequence, as obtained from BLAST analysis. However, other typical features of ER domains are absent. We therefore propose that it is an inactive ER domain. It can however not be excluded, that this domain is responsible for the abovementioned isomerization or double bond migration of the double bond between C3 and C2 or C4.

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

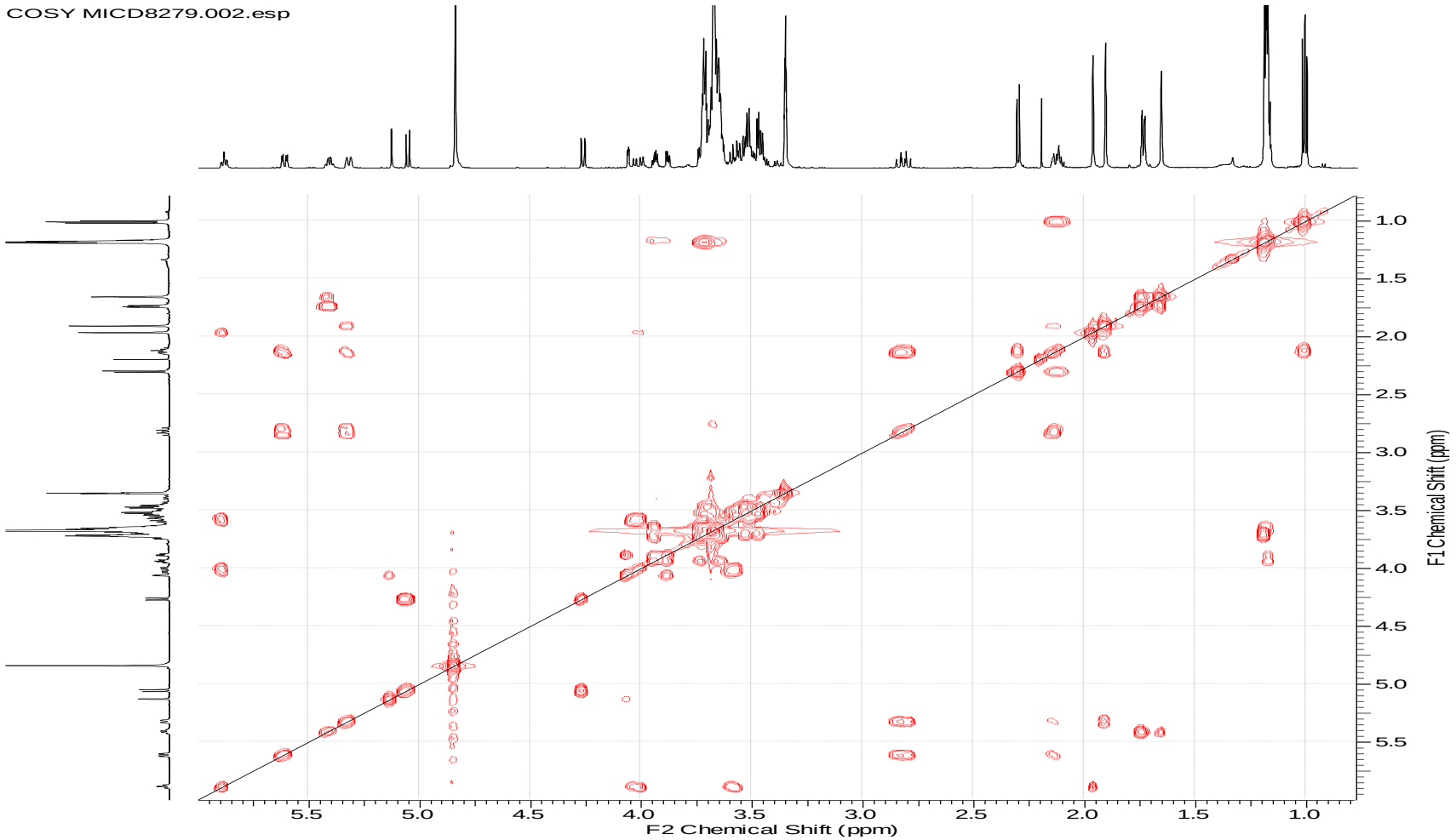
[2] S. Essig, S. Bretzke, R. Muller, D. Menche, *Journal of the American Chemical Society* **2012**, *134*, 19362-19365.



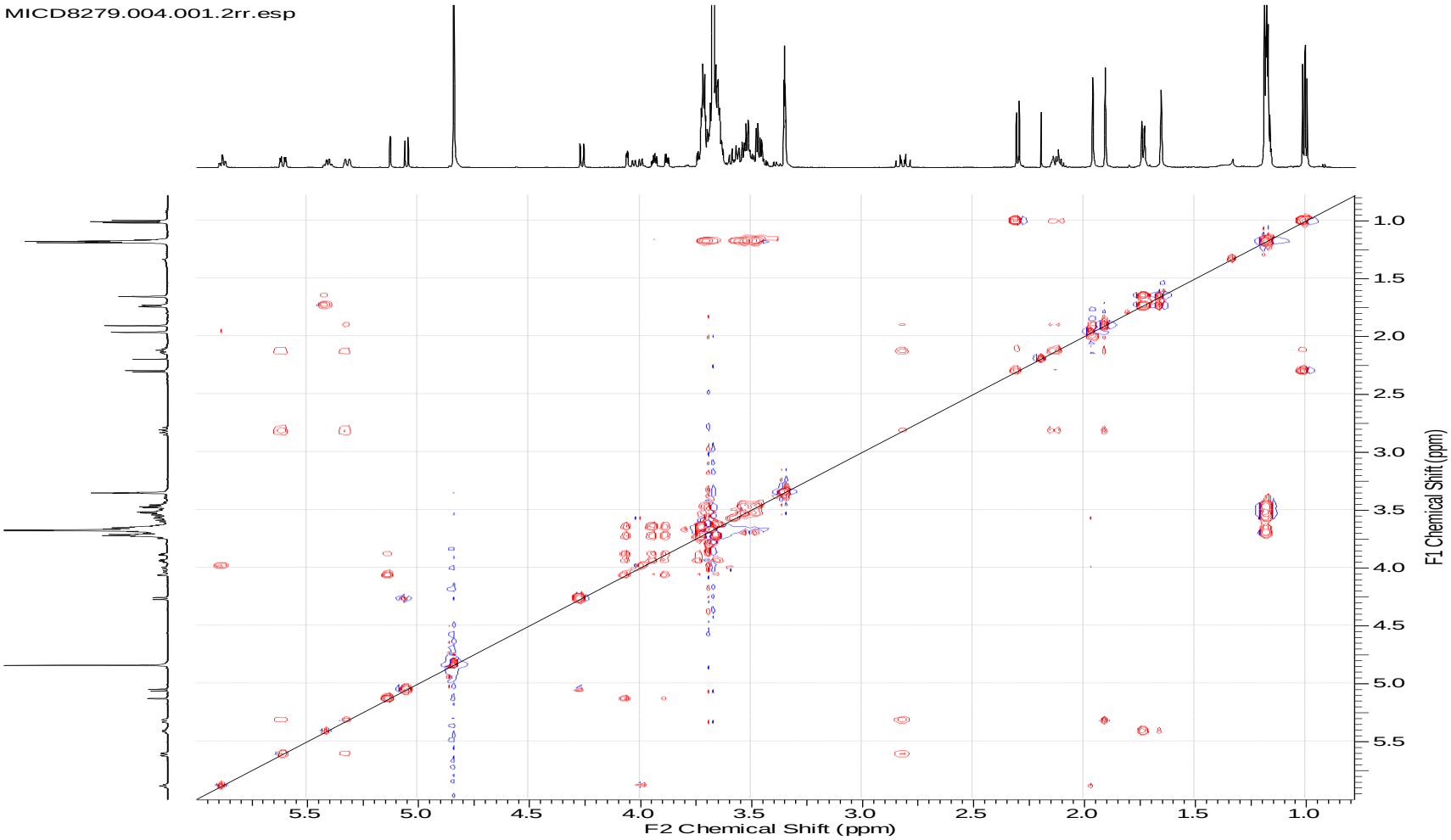
^{1}H NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



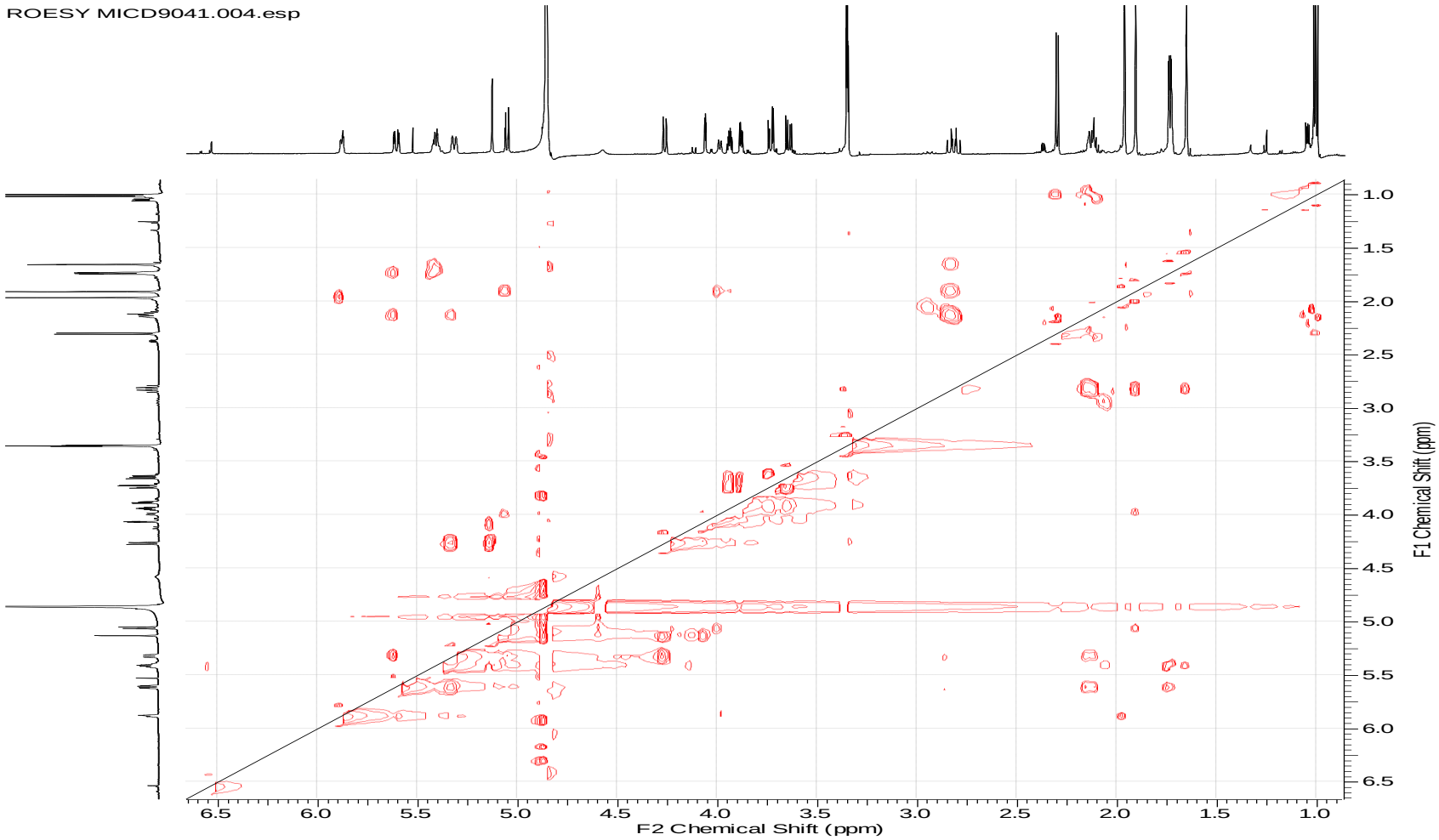
^{13}C NMR spectrum (150 MHz, methanol-d_4) of disciformycin A (**1**).



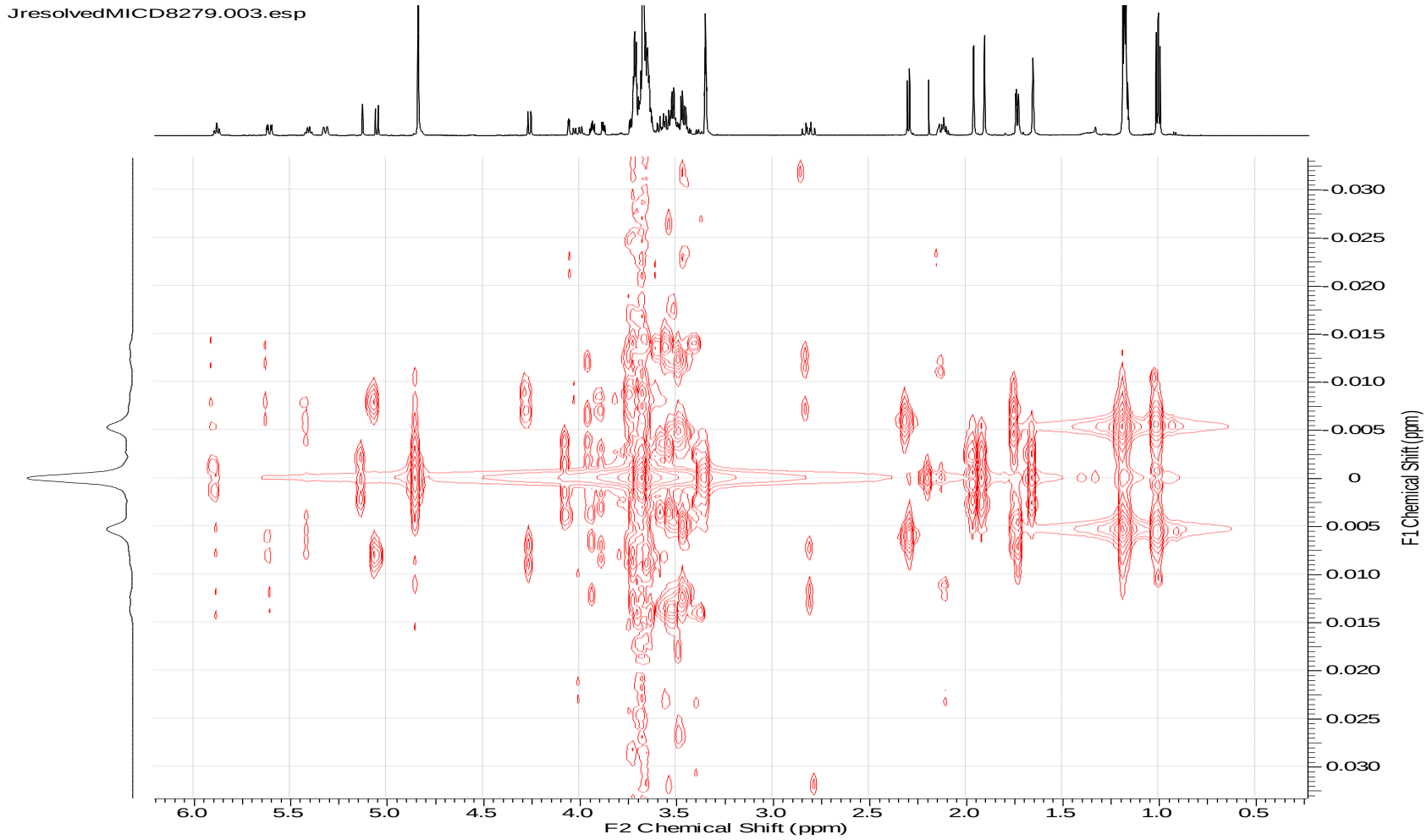
COSY NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



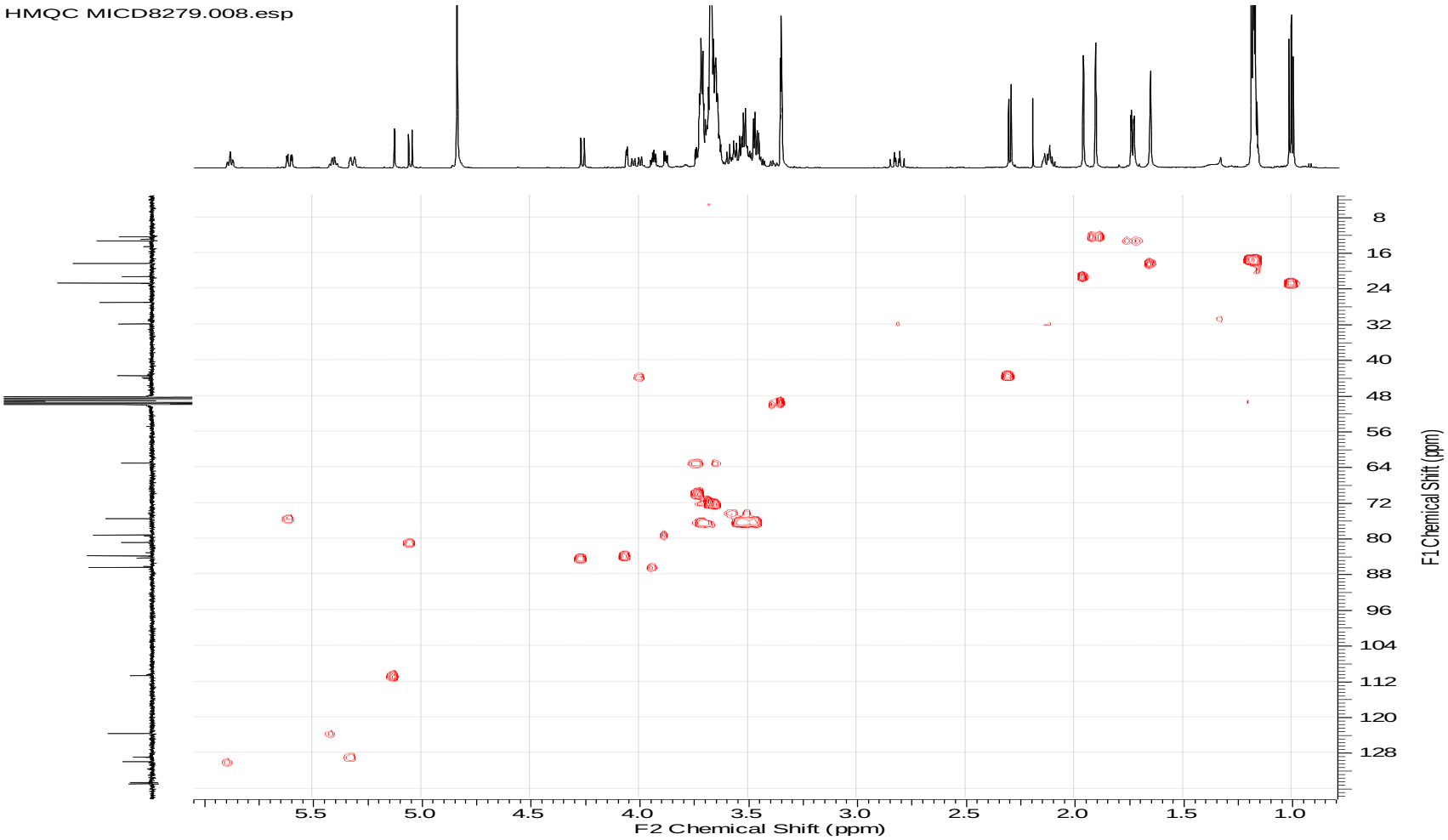
TOCSY NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



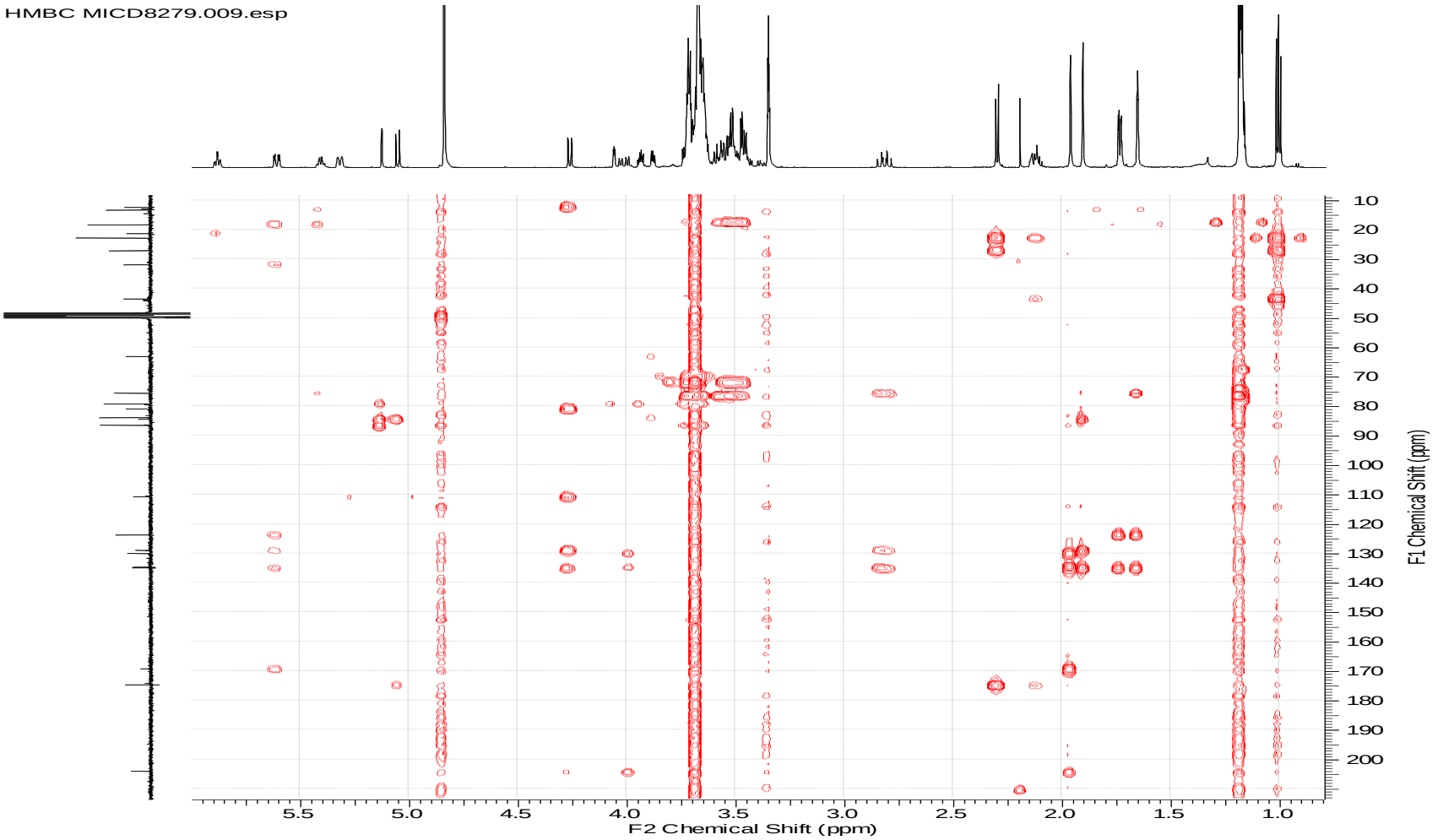
ROESY NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



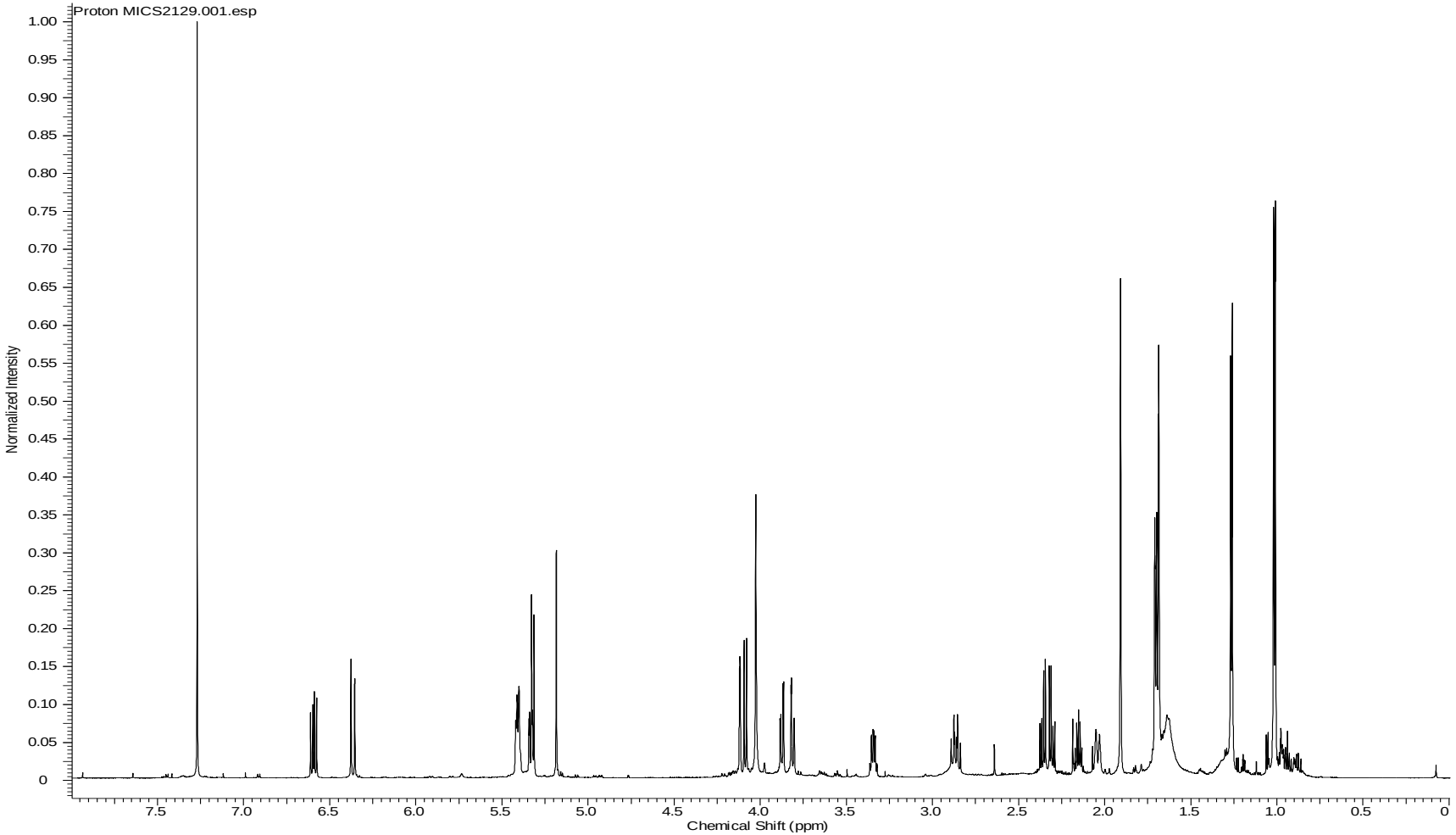
J-resolved NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



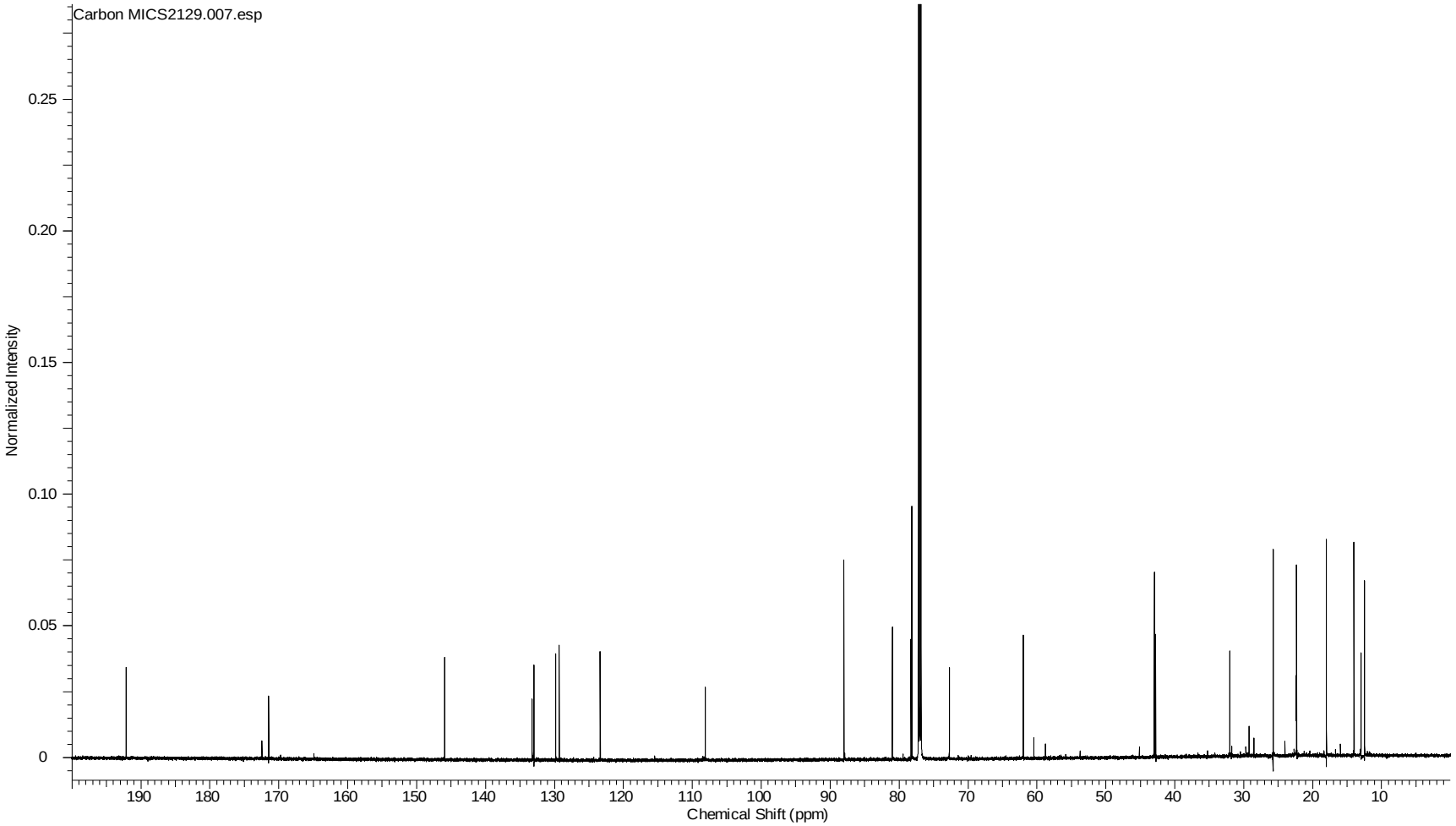
HMQC NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



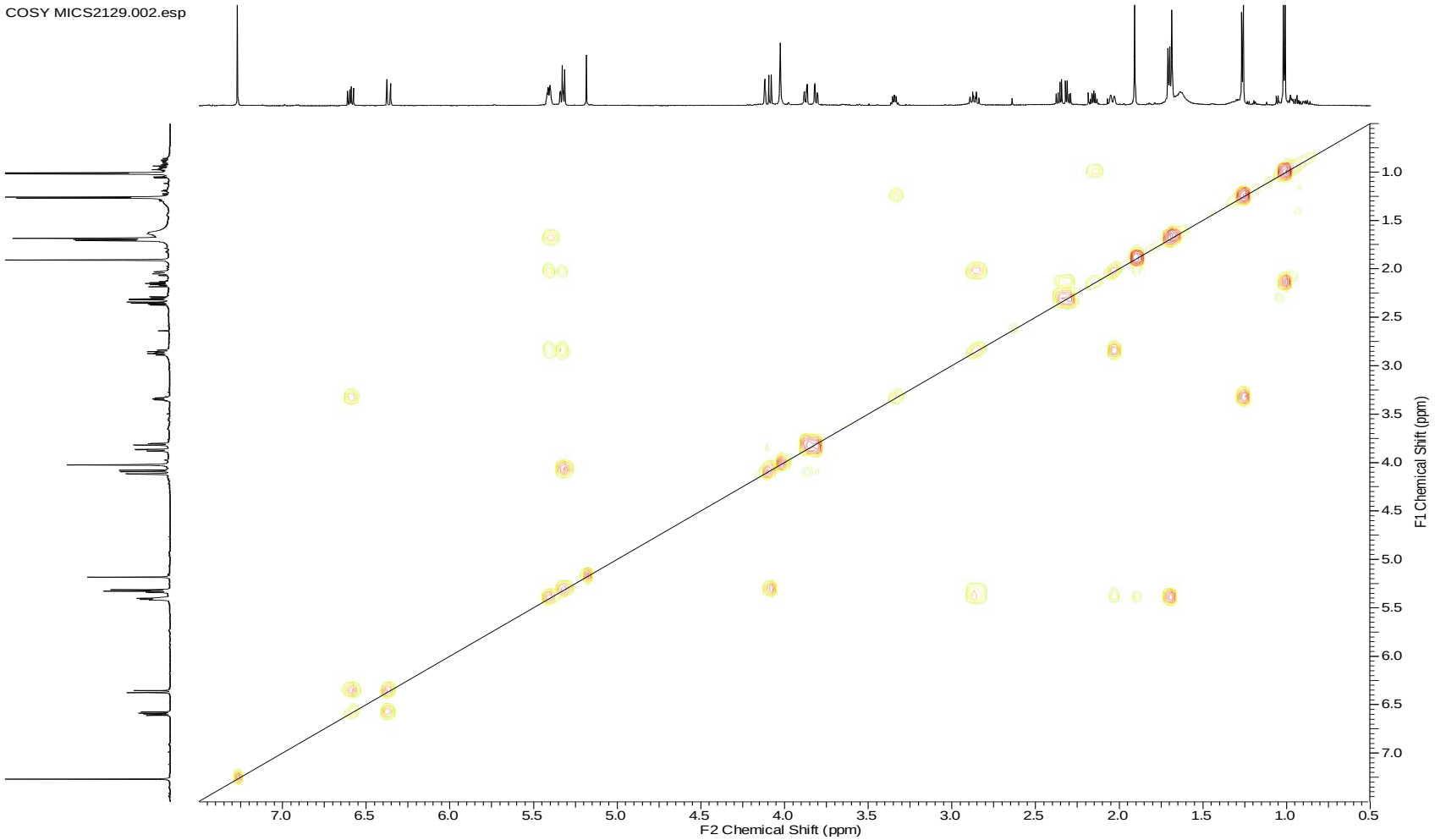
HMBC NMR spectrum (600 MHz, methanol-d_4) of disciformycin A (**1**).



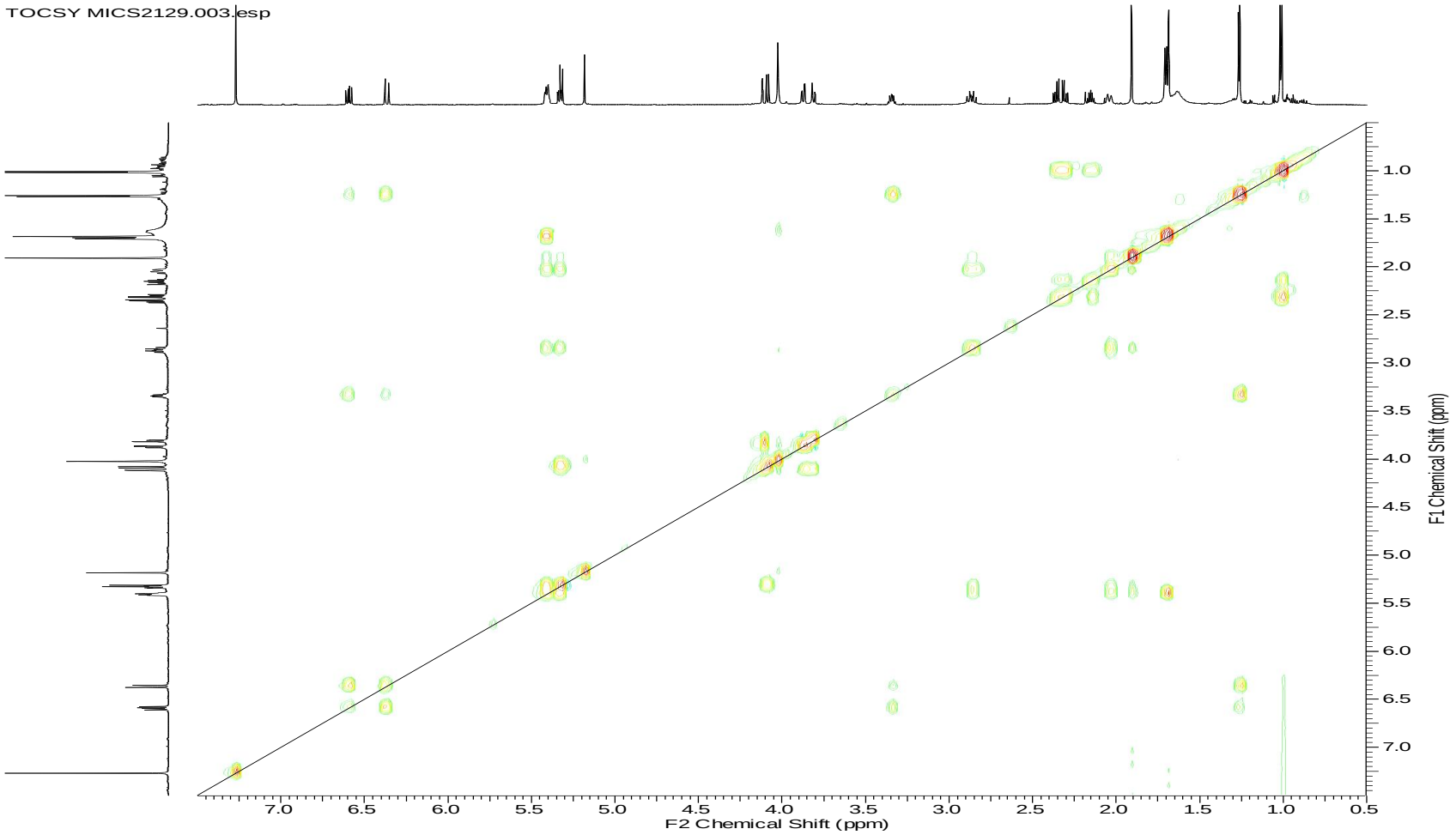
^{1}H NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).



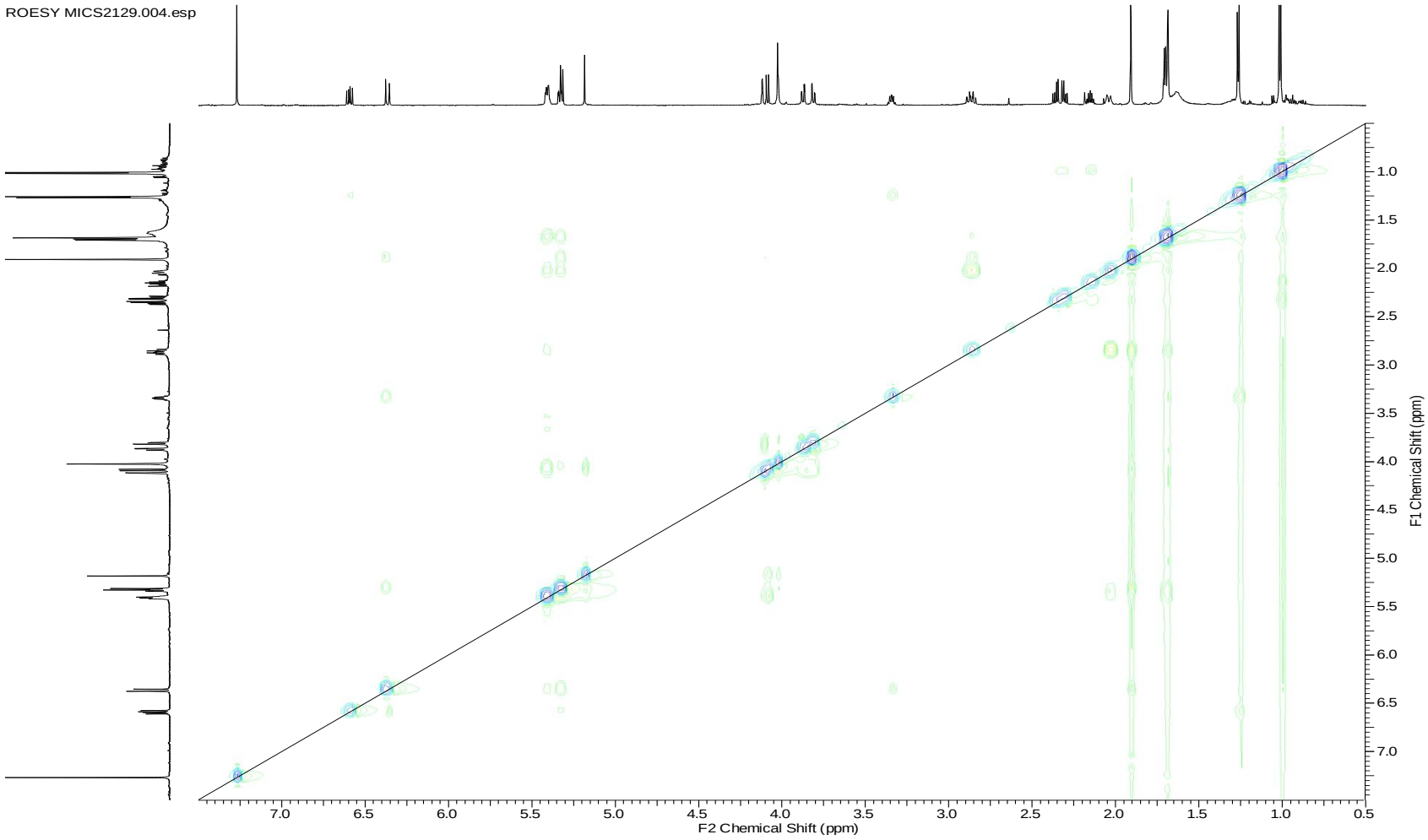
^{13}C NMR spectrum (175 MHz, chloroform-*d*) of disciformycin B (**2**).



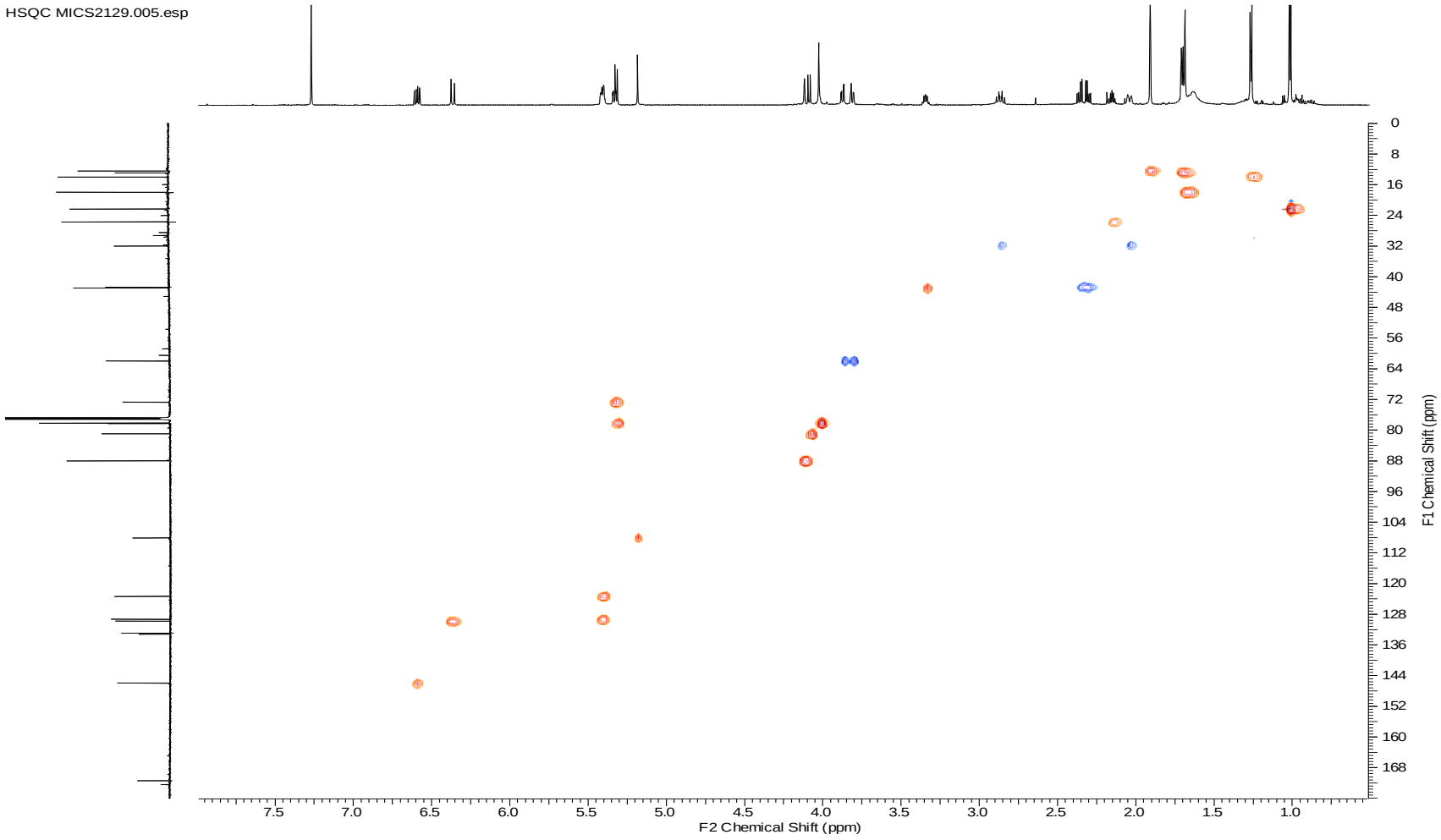
COSY NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).



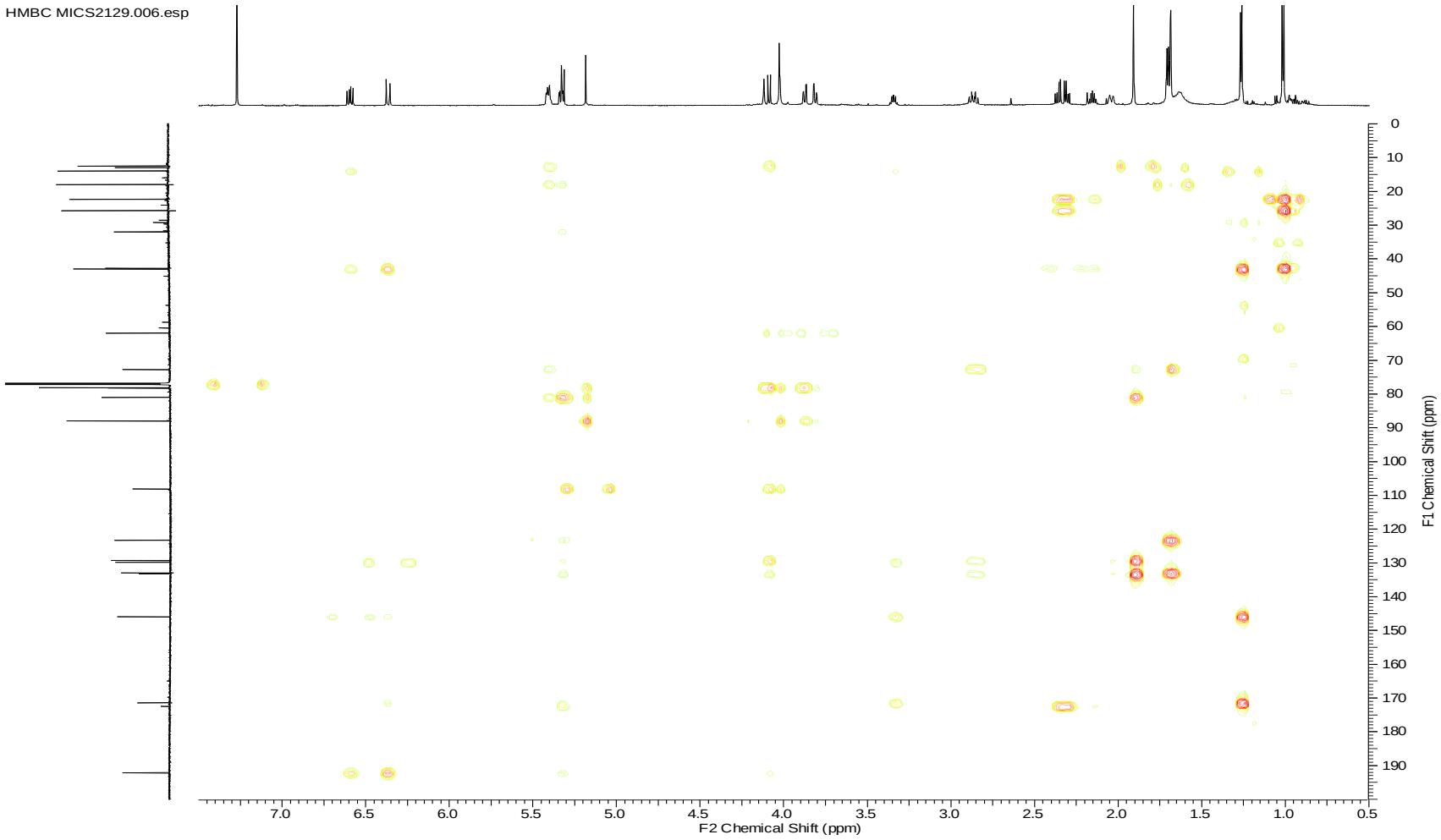
TOCSY NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).



ROESY NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).



HSQC NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).



HMBC NMR spectrum (700 MHz, chloroform-*d*) of disciformycin B (**2**).

10.4 Supporting Information: Heterologous Expression of Disciformycin Precursors in the Mxyobacterium *Myxococcus xanthus*

Heterologous Expression of Disciformycin Precursors in the Mxyobacterium *Myxococcus xanthus*

Supporting information

NMR-Data of heterologously produced disciformycin precursor P2 and the by-product P1:

Table 1: NMR spectroscopic data of **P1** in DMSO-d_6 (^{1}H at 700 MHz).

#	^{13}C, mult.	^{1}H, mult.	COSY	HMBC
1	176.5, C			
2	37.0, CH	2.44, m[a)]	3,14	1,3,4,5,14
3	26.6, CH2	3_a: 1.48, m[b)] 3_b: 1.56, m[c)]	2,4	1,2,4,5,14
4	38, CH2	2.43, m[a)]	3,2	1,2,3,5,14
5	206.7, C			
6	49.1, CH2	6_a: 2.70, dd (5.8,12.9) 6_b: 2.37, dd (10.9,12.8)[a)]	6_a, 6_b, 7	5,7,8
7	73.8, CH	4.30, dd (5.8, 10.9)	6_a, 6_b	6,8,15
8	137.8, C			
9	124.2, CH	5.13 (d, 11.4)	10_a, 10_b, 15	7,10,15
10	32.3	10_a: 2.25, m 10_b: 2.11, m	9, 11	8,9,11,12
11	73	4.88, m	10_a, 10_b, 12_a, 12_b	1
12	26.6, CH2	12_a: 1.50, m[b)] 12_b: 1.57, m[c)]	13	10,11,13
13	9.5, CH3	0.81, t (7.5)	12_a, 12_b	11,12
14	18.2, CH3	1.02, d (6.94)	2	1
15	9.3, CH3	1.47, m[c)]	9	7,8,9

[a)], [b)], [c)], [d)] : Overlapping ^{1}H and ^{1}H, ^{1}H-COSY signals.

Table 2: NMR spectroscopic data of **P2** in DMSO-d_6 (^{1}H at 700 MHz).

#	^{13}C, mult.	^{1}H, mult.	COSY	HMBC
1	175.2, C			
2	37.3, CH	2.42, m [a)]	15, 3_b [b)]	15,3,4
3	26.6, CH2	3_a: 1.57, m 3_b: 1.49, m	2,4 [b)]	1,2,4,5,15
4	38.2, CH2	2.45, m [a)]	3_b,3_a [b)]	2,3,5
5	207, C			
6	49.2, CH2	6_a: 2.69, dd (5.9, 10.9) 6_b: 2.43, m [a)]	7	5,7,8
7	73.8, CH	4.32, dd ((5.9, 11.1)	6_a, 6_b	6,8,9,16
8	138.5, C			
9	124.0, C	5.23, m	10_a,10_b,16	7,10,11,16
10	30.4, CH2	10_a: 1.89, m 10_b: 2.54, m [c)]	10_a: 9,11,16 10_b: 9,11	
11	70.2, CH	5.69, dd (11.9,2.9)	10_a,10_b,	1,9,10,12,13
12	133.7, C			
13	121.9, CH	5.3, m	14,17	11,12,14,17
14	12.5, CH3	1.63, m [d)]	13 [d)]	10,11,12,13,17
15	18;CH3	0.98, d (6.9)	2	1,2,3
16	9.5, CH3	1.52, m	9,10_a,16	7,8,9,10,11
17	17.9, CH3	1.64, m [d)]	14 [d)]	10,11,12,13,14

NMR-spectra of P2:

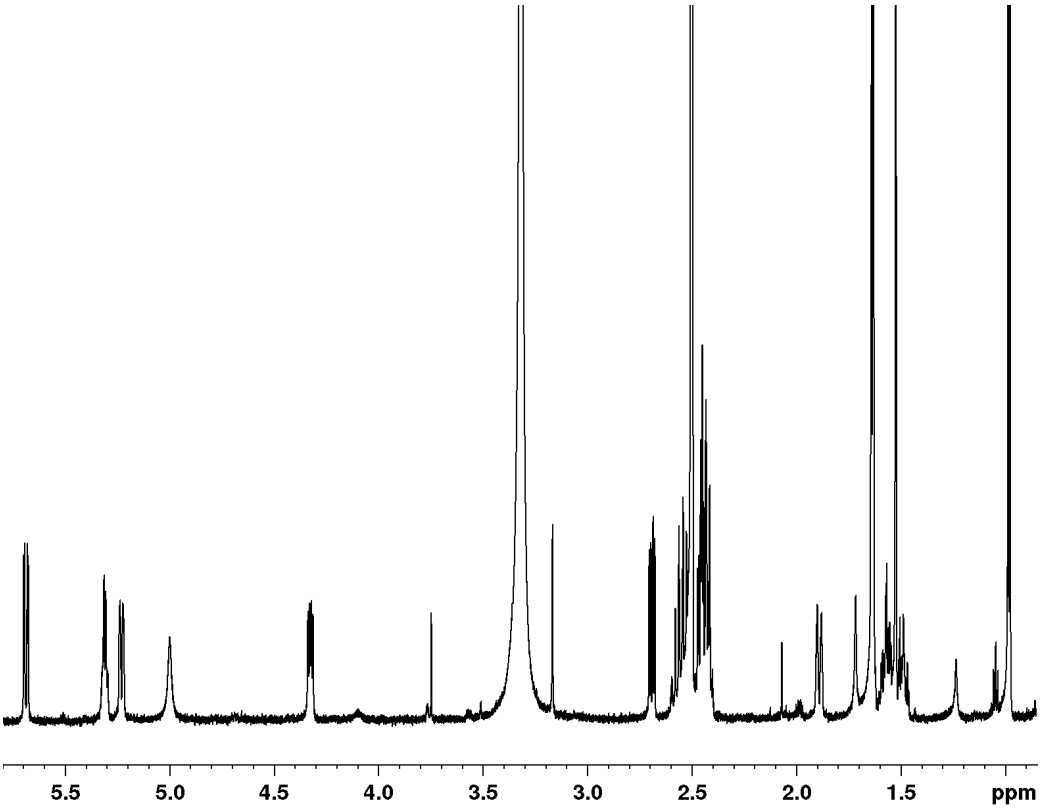


Figure 1: ^{1}H NMR spectrum (700 MHz, DMSO-d_6) of **P2**

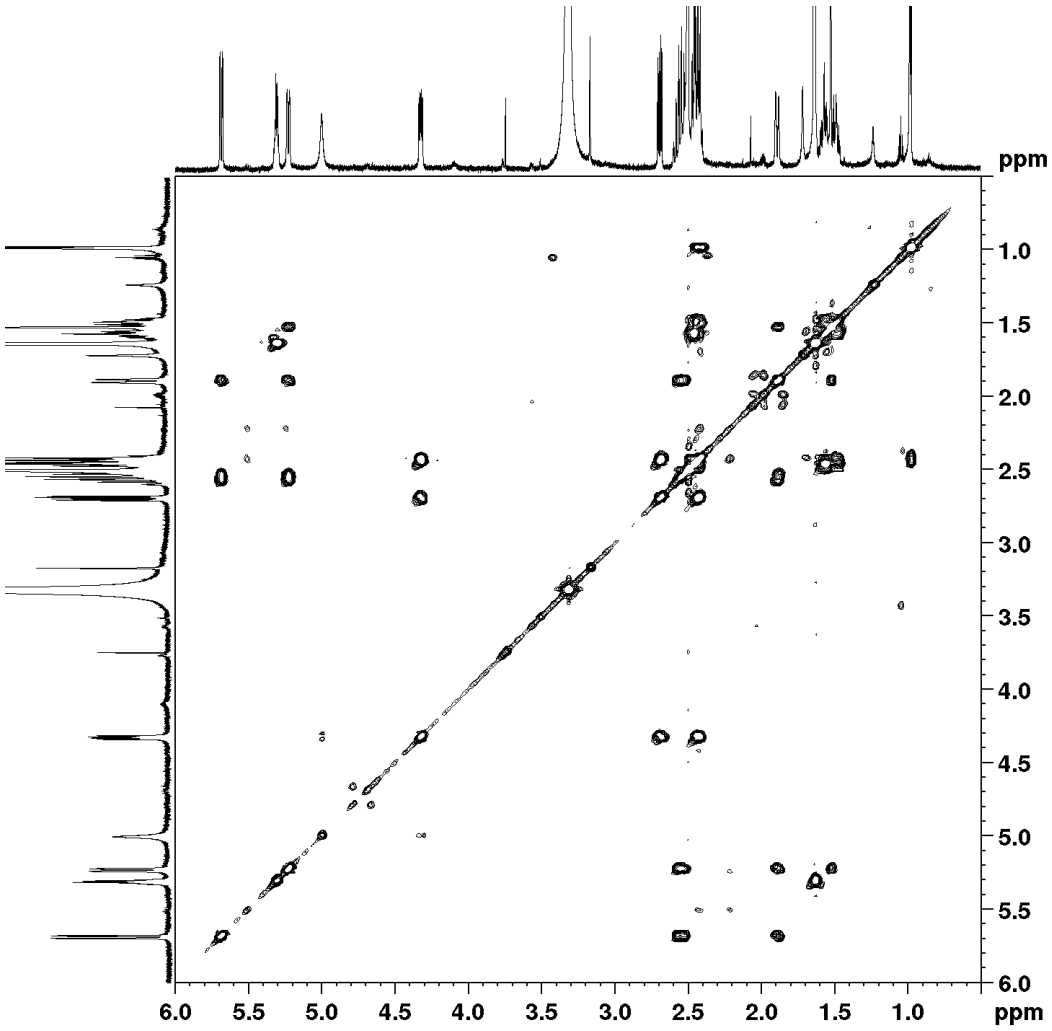


Figure 2: COSY NMR spectrum (700 MHz, DMSO-d_6) of **P2**.

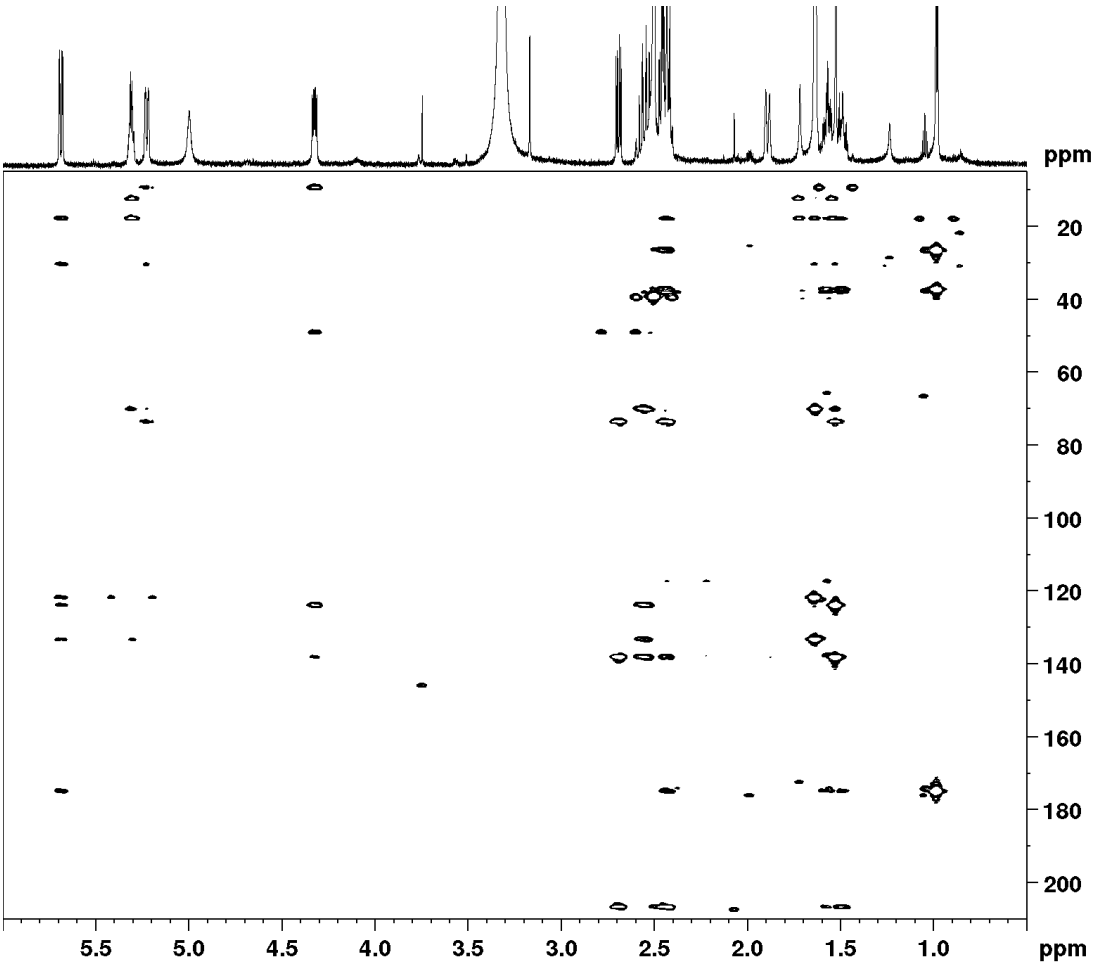


Figure 3: HMBC NMR spectrum (700 MHz, DMSO-d_6) of **P2**.

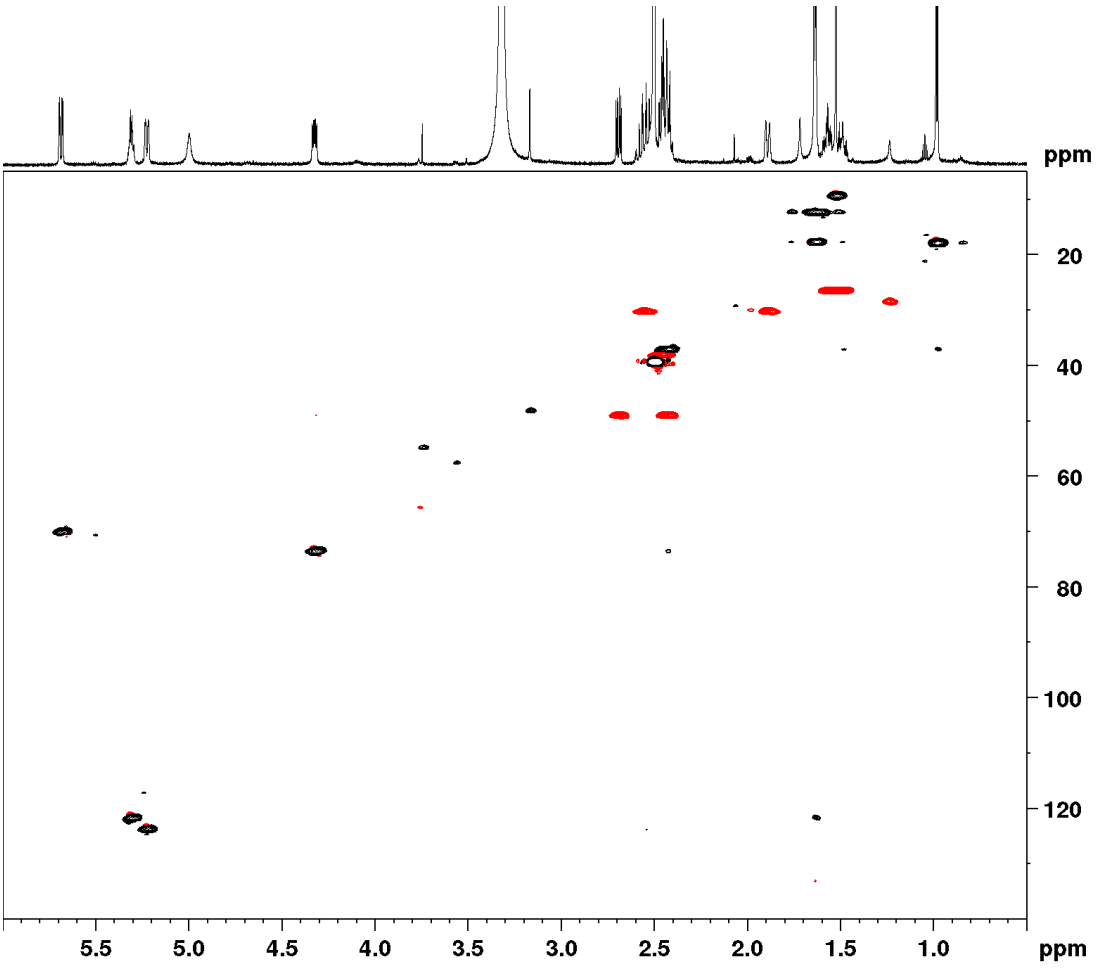


Figure 4: HMQC NMR spectrum (700 MHz, DMSO-d_6) of **P2**.

NMR spectra of P1 are not presented here, since the compound is regarded as an unwanted side-product and not a disciformycin precursor.

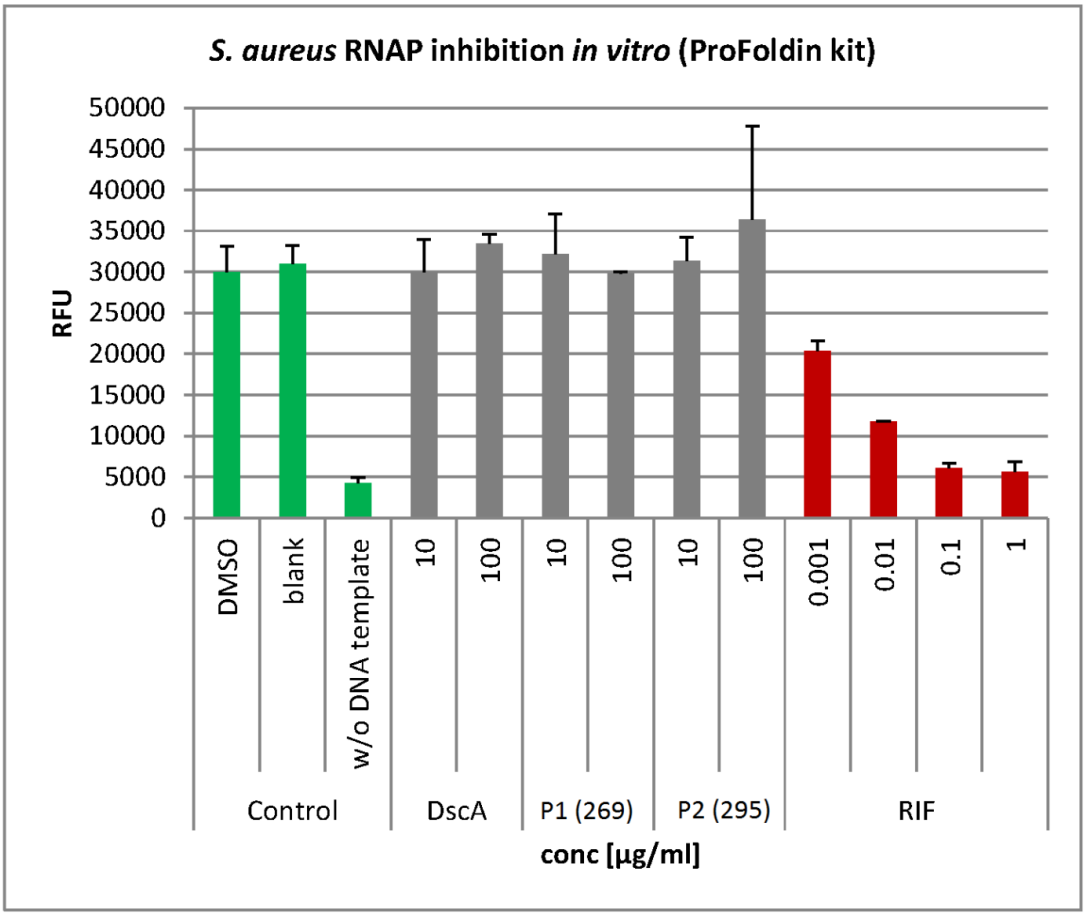


Figure 5: RNAP inhibitory activity of disciformycin A (DscA) and precursors P1 and P2 tested *in vitro*. Rifampicin (RIF) was used as positive control. Neither disciformycin nor P1 and P2 exhibited any *in vitro* activity.

10.5 Supporting Information: Jahnellamides, α-Keto-β-Methionine-Containing Peptides from the Terrestrial Myxobacterium *Jahnella* sp.: Structure and Biosynthesis

Jahnellamides, α-Keto-β-Methionine-Containing Peptides from the Terrestrial Myxobacterium *Jahnella* sp.: Structure and Biosynthesis

*Alberto Plaza, Konrad Viehrig, Ronald Garcia, Rolf Müller**

Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Pharmaceutical Biotechnology, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany

rom@mx.uni-saarland.de

TABLE OF CONTENTS

Experimental Details:

General Experimental Procedures.

Optical rotations were measured with a Jasco polarimeter, IR spectra were recorded on a Perkin-Elmer FT-IR Spectrum One spectrometer, and UV spectra were recorded on an Agilent 8453 spectrophotometer. NMR spectra were recorded in CD3OD and DMSO-d_6 on a Bruker Ascend 700 spectrometer with a 5 mm TXI cryoprobe (^{1}H at 700 MHz, ^{13}C at 175 MHz). DQFCOSY, 2D-HOHAHA, HSQC, HMBC, and ROESY experiments were recorded using standard pulse programs. HSQC experiments were optimized for $^1J_{C\text{-}H}$ = 145 Hz, and HMBC spectra were optimized for $^{2,3}J_{C\text{-}H}$ = 8 and 6 Hz. LC-HRMS data was performed on a Dionex Ultimate 3000 RSLC system using a Waters BEHC18, 100 x 2.1 mm, 1.7 µm dp column. Separation of 2 µl sample was achieved by a linear gradient with (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µl/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in18 min to end up with a 2 min step at 95 % B before reequilibration with initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the maXishr-ToF mass spectrometer (BrukerDaltonics, Bremen, Germany) using the standard ESI source. Mass spectra were acquired in centroid mode ranging from 150 – 2000 m/z at 2 Hz scan speed.

Cultivation of Strain SBSr007

Jahnella sp. SBSr007 was isolated from a dried soil sample containing decayed plant materials using mineral salt agar. The strain was recognized after weeks of cultivation at 30°C. Isolation and purification was performed by repeated transfer of the swarm material into new agar medium. Strain SBSr007 was identified based on morphological characteristics of growth stages and by 16S rRNA gene amplification. Its identity was further confirmed by phylogenetic analysis showing its clustering to *Jahnella* clade in *Polyangiaceae* family. SBSr007 was cultivated and maintained in buffered VY/2 agar medium containing 5mM HEPES. The bacterial culture was subcultivated in the same agar medium every two weeks. Fermentation was performed in 5 x 0.2 L flasks containing 100 mL buffered VY/2 broth. The flasks cultures were grown for 30d with replacement of the medium every 24h. Cell pellets were combined in all the flasks and harvested by centrifugation.

Isolation: The cell pellet was harvested by centrifugation and lyophilized overnight, followed by extraction with 3 x 300 mL methanol. The combined fractions yielded an orange-brown crude extract which was further partitioned between hexane and MeOH/H_2O 7:3 (v/v) to yield 180.0 mg of crude extract out of the aqueous phase. Subsequently, the extract was purified by semi-preparative HPLC (Jupiter Proteo C12, 250 x 10 mm, 4 µm column, DAD at 220 and 320 nm) eluting with a linear gradient of 20-50% ACN/H_2O in 0.1% FA in 50 min to afford compounds **1** (0.6 mg, t_R = 19.5 min), **2** (0.2 mg, t_R = 19.7 min), microsclerodermin D (1.0 mg, t_R = 25.6 min), and pedein A (1.2 mg, t_R = 30.8 min).

Jahnellamide A (1): Colorless amorphous powder; $[\alpha]^{20}_D$ + 18.7 (*c* 0.4, MeOH); IR (film) ν_{max} 3286, 2933, 1646, 1604, 1542, 1355, 1236, 1058, 752, 715 cm^{-1}; UV (ACN) λmax 224, 344 nm; ^{1}H and ^{13}C NMR data, see Table S1 and S2; HR-ESI-MS *m/z* 947.3702 $[M + H]^+$ corresponding to a molecular formula of $C_{44}H_{54}N_{10}O_{12}S$ (calcd for $C_{44}H_{55}N_{10}O_{12}S$, 947.3722).

Jahnellamide B (2): Colorless amorphous powder; $[\alpha]^{20}_D$ + 26.6 (*c* 0.2, MeOH); UV (ACN) λmax 222, 344 nm; ^{1}H and ^{13}C NMR data, see Table S3; HR-ESI-MS *m/z* 931.3762 $[M + H]^+$ corresponding to a molecular formula of $C_{44}H_{54}N_{10}O_{11}S$ (calcd for $C_{44}H_{55}N_{10}O_{11}S$, 931.3772).

LC/MS Analysis of L/D-FDLA Derivatives. Approximately 0.25 mg of compound **1** was hydrolyzed with 6 N HCl (0.5 mL) for 16 h at 100 °C, dried, and dissolved in H2O (100 μL). To a 50 μL aliquot of each were added 1 N $NaHCO_3$ (20 μL) and 1% 1-fluoro-2,4-dinitrophenyl-5-L-leucinamide (L-FDLA or D-FDLA solution in acetone, 100 μL), and the mixtures were heated to 40 °C for 40 min, allowed to cool to rt, neutralized with 2 N HCl (20 μL), and evaporated to dryness. Residues were dissolved in CH_3CN and analyzed by LC-MS. All measurements were performed on

a Dionex Ultimate 3000 RSLC system using a Waters BEH C18, 100 x 2.1 mm, 1.7 µm column by injection of 1 µl sample. Separation was achieved by a gradient using (A) H_2O + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 550 µl/min and 45 °C. The gradient was as follows: starting at 5 % B to increase to 10 % B in 1 min, from 1 to 15 min increase to 35 % B, from 15 to 22 min increase to 50% B, from 22 to 25 min increase to 80 % B. After a 1 min hold at 80 % B the system was reequilibrated with initial conditions for 5 minutes. UV data was acquired at 340 nm and MS detection was performed simultaneously. Coupling the HPLC to the MS was supported by an Advion Triversa Nanomate nano-ESI system attached to a Thermo Fisher Orbitrap. LC flow is split to 500 nL/min before entering the ion source. Mass spectra were acquired in centroid mode ranging from 150 – 1000 *m/z* at a resolution of R = 30000.
Retention times (t_R, min) of the FDLA-derivatized amino acids for compound **1**: L-Gln 13.4, D-Gln 14.3 *m/z* 442 $[M+H]^+$ (analyzed as Glu); L-Dap 8.74, 9.54, D-Dap 8.14, 9.66 *m/z* 399 $[M+H]^+$; L-Pro 14.5, D-Pro 16.2 *m/z* 410 $[M+H]^+$; 4*S*-OH-L-Glu 15.6, 4*R*-OH-L-Glu 17.0, 4*S*-OH-D-Glu 16.5, 4*R*-OH-D-Glu 16.0 *m/z* 458 $[M+H]^+$.

For the determination of the absolute configuration of α-keto-β-methionine the following procedure was applied: To a stirred solution of **1** (0.25 mg) in 5% NaOH (0.5 ml), 35% H_2O_2 (0.1 ml) was added dropwise. The reaction was stirred at 65 C for 40 min and subsequently evaporated to dryness. Hydrolysis, derivatization and LC-MS analysis were performed as previously described, giving peaks for L-Met 18.1, D-Met 18.5 *m/z* 476 $[M+H]^+$ (analyzed as Met(O_2)).

Feeding Experiments (MS detection):
Cultivation of SBSr007 was performed in 25 mL buffered VY/2 broth. The flasks cultures were grown for 7d. Labeled L-[U-^{13}C, ^{15}N]methionine, L-[ring-$^{13}C_6$]phenylalanine, [$^2H_5$]benzoic acid, and [$^2H_7$]cinnamic acid were dissolved in DMSO, and 50 µL of each were added to individual cultures in three portions (on first, third and fifth day) to a final concentration of 0.2 mM. The individual cell pellets were harvested by centrifugation and lyophilized overnight, followed by extraction with 3 x 25 mL methanol. The organic solvent was removed and the residuals dissolved in 500 µL methanol and analyzed by HR-ESI-MS.

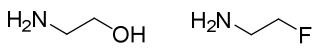
Feeding Experiments (NMR detection):
[1-^{13}C] sodium acetate, [2-^{13}C] sodium acetate:
Cultivation of SBSr007 was performed in 400 mL (8x50 mL) buffered VY/2 broth. Every flask was fed with either [1-^{13}C] sodium acetate or [2-^{13}C] sodium acetate. Labeled precursors were dissolved in DMSO, and 50 µL of each were added to individual cultures in three portions (on first, third and fifth day) to a final concentration of 10 mM. The individual cell pellets were harvested by centrifugation and lyophilized overnight, followed by extraction with methanol. This crude extract was further partitioned between hexane and MeOH/H_2O 7:3 (v/v) and subsequently purified by semi-preparative HPLC (Jupiter Proteo C12, 250 x 10 mm, 4 µm column). See isolation conditions.
L-[2,3,3-$^2H_3$]serine:
Cultivation of SBSr007 was performed in 8x50 mL buffered VY/2 broth 400 mL cultures (8x50 mL) of SBSr007. Every flask was fed with L-[2,3,3-$^2H_3$]serine. Serine was dissolved in DMSO, and 50 µL of each were added to individual cultures in three portions (on first, third and fifth day) to a final concentration of 0.2 mM. The individual cell pellets were harvested by centrifugation and lyophilized overnight, followed by extraction with methanol. This crude extract was further partitioned between hexane and MeOH/H_2O 7:3 (v/v) and subsequently purified by semi-preparative HPLC (Jupiter Proteo C12, 250 x 10 mm, 4 µm column). See isolation conditions.

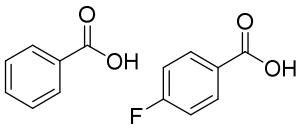
Precursor-directed biosynthesis:

Cultivation of SBSr007 was performed in 25 mL buffered VY/2 broth. The flasks cultures were grown for 7d. Precursors were dissolved in DMSO, and 50 µL of each were added to individual

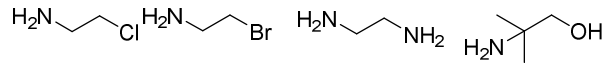
cultures in three portions (on first, third and fifth day) to a final concentration of 0.2 mM. The individual cell pellets were harvested by centrifugation and lyophilized overnight, followed by extraction with 3 x 25 mL methanol. The organic solvent was removed and the residuals dissolved in 500 µL methanol and analyzed by HR-ESI-MS.
The complete list of substrates used is the following

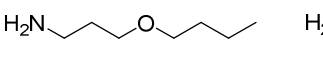
Incorporated: Yield for the fluorinated analogues was ~125 µg/L

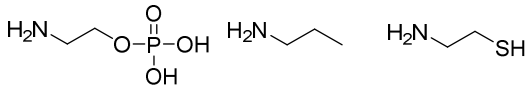


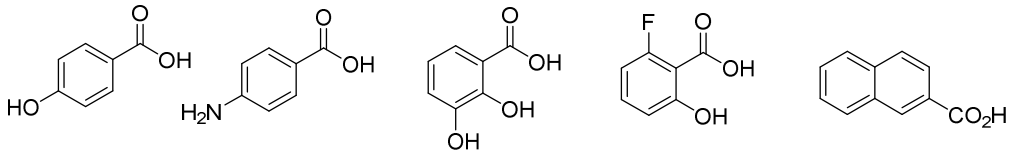


Non Incorporated:

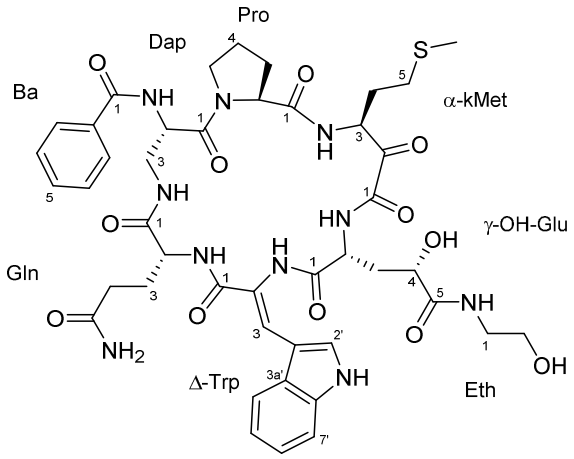








NMR Data



1

Table S1. NMR Spectroscopic Data for Jahnellamide A (**1**) (CD_3OD)

	δ_C[a]	δ_H[b] (*J* in Hz)	HMBC[c]
		Dap	
1	172.2		
2	52.4	4.58 dd (11.5, 4.4)	1, 3, 1_{Ba}
3a	42.0	4.34 dd (9.2, 4.4)	1, 2, 1_{Gln}
3b		3.25 dd (11.5, 12.6)	1, 2, 1_{Gln}
		Pro	
1	176.7		
2	61.3	4.33 m	1, 3, 1_{Dap}
3a	29.7	1.78 m	1, 4, 5
3b		1.29 m	1, 2, 4
4a	26.1	1.45 m	3, 5
4b		1.10 m	3
5a	49.1	4.13 t (8.6)	2, 3, 4, 1_{Dap}
5b		3.45 ddd (11.5, 9.8, 6.0)	4
		α-kMet	
1	174.2		
2	99.4		
3	53.8	4.83 dd (8.4, 2.3)	2, 4, 5, 1_{Pro}
4a	28.0	2.28 m	3, 5
4b		1.81 m	2, 3, 5
5a	30.7	2.70 m	3, 4, 6
5b		2.63 m	3, 4, 6
6	15.0	2.07 s	5
		γ-OH-Glu	
1	171.8		
2	53.7	4.86 m	1, 3, 4, $1_{α\text{-kMet}}$
3a	37.3	2.67 m	1, 2
3b		2.30 m	1, 2, 4, 5
4	69.5	4.16 dd (11.4, 2.3)	2, 3, 5
5	177.8		
		Δ-Trp	
1	167.7		
2	120.4		
3	127.2	8.13 s	1, 2', 9', $1_{γ\text{-OH-Glu}}$
1'			
2'	130.1	7.71 s	2, 3, 3', 4', 7', 8', 9'
3'	110.7		
3'a	129.5	9'	129.5
4'	119.1	7.81 dd (7.1, 1.4)	3', 6', 7', 8'
5'	122.2	7.19 dt (7.1, 1.4)	7', 9'
6'	123.8	7.20 dt (7.1, 1.4)	4', 8'
7'	113.1	7.43 dd (7.1, 1.4)	2', 5'
7'a	137.5		
		Gln	
1	175.0		
2	57.6	4.31 dd (7.7, 3.8)	1, 3, 4, $1_{Δ\text{-Trp}}$
3a	26.8	2.24 m	1, 2, 4, 5
3b		2.07 m	1, 2, 4, 5
4a	31.9	2.55 dd (9.5, 1.8)	2, 3, 5

4b		2.47 dd (9.0, 1.8)	2, 3, 5
5	179.2		
		Ba	
1	170.3		
2	134.5		
3/7	128.7	7.81[d]	1, 4, 5
4/6	129.7	7.43[d]	2, 3, 5
5	133.2	7.52 tt (7.4, 1.3)	2, 3, 4
		Eth	
1	42.7	3.36 dd (11.4, 5.4)	2, $5_{\gamma\text{-OH-Glu}}$
2	61.6	3.62 t (5.4)	1

[a]Recorded at 175 MHz; referenced to residual CD_3OD at δ 49.15 ppm.
[b]Recorded at 700 MHz; referenced to residual CD_3OD at δ 3.31 ppm.
[c]Proton showing HMBC correlation to indicated carbon.
[d]Proton showing ROESY correlation to indicated proton. [d] Overlapped.

Table S2. NMR Spectroscopic Data for Jahnellamide A (**1**) (DMSO-d_6)

	δ_C[a]	δ_H[b] (*J* in Hz)	ROESY[c]
Dap			
1	171.5		
2	51.1	4.24 m	NH, NH-3, $5a_{Pro}$
3a	39.9	4.10 [d]	NH
3b		3.05 t (11.6)	NH, NH-3
NH		8.74 d (6.4)	2, 3a, 3b, 3_{Ba}
NH-3		6.78 dd (10.1, 2.0)	3b, 2_{Gln}, NH_{Gln}
Pro			
1	173.3		
2	59.0	4.14 t (7.6)	$NH_{\alpha\text{-kMet}}$
3a	27.6	1.63 m	$NH_{\alpha\text{-kMet}}$
3b		1.27 m	
4a	24.5	1.45 m	
4b		1.19 m	
5a	46.7	3.89 m	2_{Dap}
5b		3.27 m	
α-kMet			
1	161.2		
2	192.0		
3	53.0	4.50 dd (11.5, 5.2)	NH, $2_{\gamma\text{-OH-Glu}}$
4a	29.0	2.02 m	NH
4b		1.92 m	
5	28.2	2.60 m	6, NH
6	14.1	2.05 s	5
NH		9.33 d (4.7)	3, 4a, 4b, 5a, 2_{Pro}
γ-OH-Glu			
1	167.5		
2	49.4	4.69 m	4, NH, $3_{\alpha\text{-kMet}}$, $2'_{\Delta\text{-Trp}}$, $NH_{\Delta\text{-Trp}}$
3a	34.7	2.29 m	NH
3b		2.19 m	
4	67.6	3.91 m	2, NH, OH-4, NH_{Eth}
5	173.8		
NH		8.93 d (9.9)	2, 3a, 4, $NH_{\Delta\text{-Trp}}$
OH-4		5.81 d (6.0)	4, NH_{Eth}
Δ-Trp			
1	164.2		
2	123.9		
3	124.0	7.85 s	
NH-1'		11.8 d (2.2)	2'
2'	128.0	7.48 d (2.2)	NH, $2_{\gamma\text{-OH-Glu}}$
3'	108.7		
3'a	126.7		
4'	117.8	7.74 d (7.5)	6'
5'	121.9	7.18 t (7.5)	7'
6'	120.2	7.15 t (7.5)	4'
7'	111.6	7.43 d (7.5)	5'
7'a	134.8		
NH		8.41 s	NH_{Gln}, $2_{\gamma\text{-OH-Glu}}$, $NH_{\gamma\text{-OH-Glu}}$
Gln			
1	171.1		
2	54.7	4.23 m	NH, NH-3_{Dap}
3a	25.8	1.99 m	
3b		1.86 m	
4	30.8	2.28 m	NH, NH_{2a}, NH_{2b}

5	175.0		
NH		8.54 br s	2, NH-3_{Dap}, $NH_{\Delta\text{-Trp}}$
NH_{2a}		7.79 br s	4
NH_{2b}		7.05 br s	4
	Ba		
1	165.8		
2	132.8		
3/7	127.3	7.84 t (7.2)	4, NH_{Dap},
4/6	128.0	7.45 t (7.2)	3
5	131.3	7.53 t (7.2)	
	Eth		
1	40.7	3.19 m	
2	59.4	3.43 dd (8.9, 5.4)	OH
NH		7.82 t (6.1)	OH-$4_{\gamma\text{-OH-Glu}}$, $4_{\gamma\text{-OH-Glu}}$
OH		4.69 t (5.4)	2

[a]Recorded at 175 MHz; referenced to residual DMSO-d_6 at δ 39.51 ppm.
[b]Recorded at 700 MHz; referenced to residual DMSO-d_6 at δ 2.50 ppm.
[c]Proton showing ROESY correlation to indicated proton. [d]Overlapped.

Table S3. NMR Spectroscopic Data for Jahnellamide B (**2**) (CD_3OD)

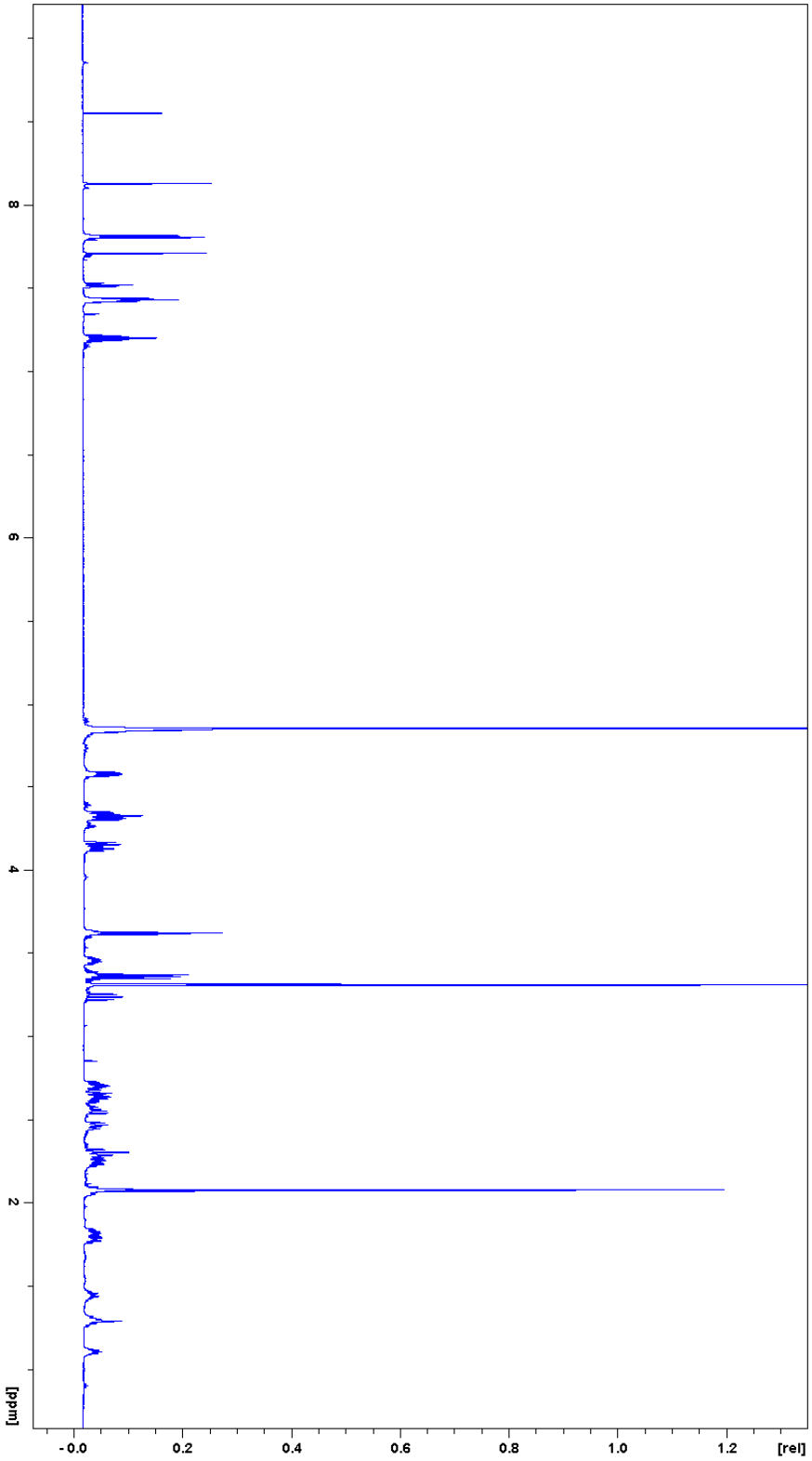
	δ_C[a]	δ_H[b] (*J* in Hz)
		Dap
1	172.2	
2	52.4	4.58 dd (8.8, 4.3)
3a	42.0	4.30 dd (9.3, 4.3)
3b		3.26 dd (12.4, 8.8)
		Pro
1	176.7	
2	61.3	4.33 m
3a	29.7	1.78 m
3b		1.29 m
4a	26.1	1.45 m
4b		1.10 m
5a	49.1	4.12 t (8.7)
5b		3.45 m
		α-kMet
1	174.2	
2	99.4	
3	53.8	4.78 dd (11.3, 2.4)
4a	28.0	2.28 m
4b		1.81 m
5a	30.7	2.70 m
5b		2.63 m
6	15.0	2.07 s
		Glu
1	171.8	
2	56.3	4.61 m
3	28.3	2.42 m
4	33.7	2.45 m
5	177.8	
		Δ-Trp
1	167.7	
2	120.4	
3	127.2	8.13 s
1'		
2'	129.8	7.66 s
3'	110.7	
4'	119.1	7.81 dd (7.3, 1.1)
5'	122.2	7.19 dt (7.3, 1.1)
6'	123.8	7.20 dt (7.3, 1.1)
7'	113.1	7.43 dd (7.3, 1.1)
8'	137.5	
9'	129.5	
		Gln
1	175.0	
2	57.6	4.31[c]
3a	26.8	2.24 m
3b		2.07 m
4a	31.9	2.55 m
4b		2.47 m
5	179.2	

	Ba	
1	170.3	
2	134.5	
3/7	128.7	7.81[c]
4/6	129.7	7.43[c]
5	133.2	7.52 tt (7.4, 1.3)
	Eth	
1	43.0	3.32[c]
2	61.4	3.60 dd (11.3, 5.6)

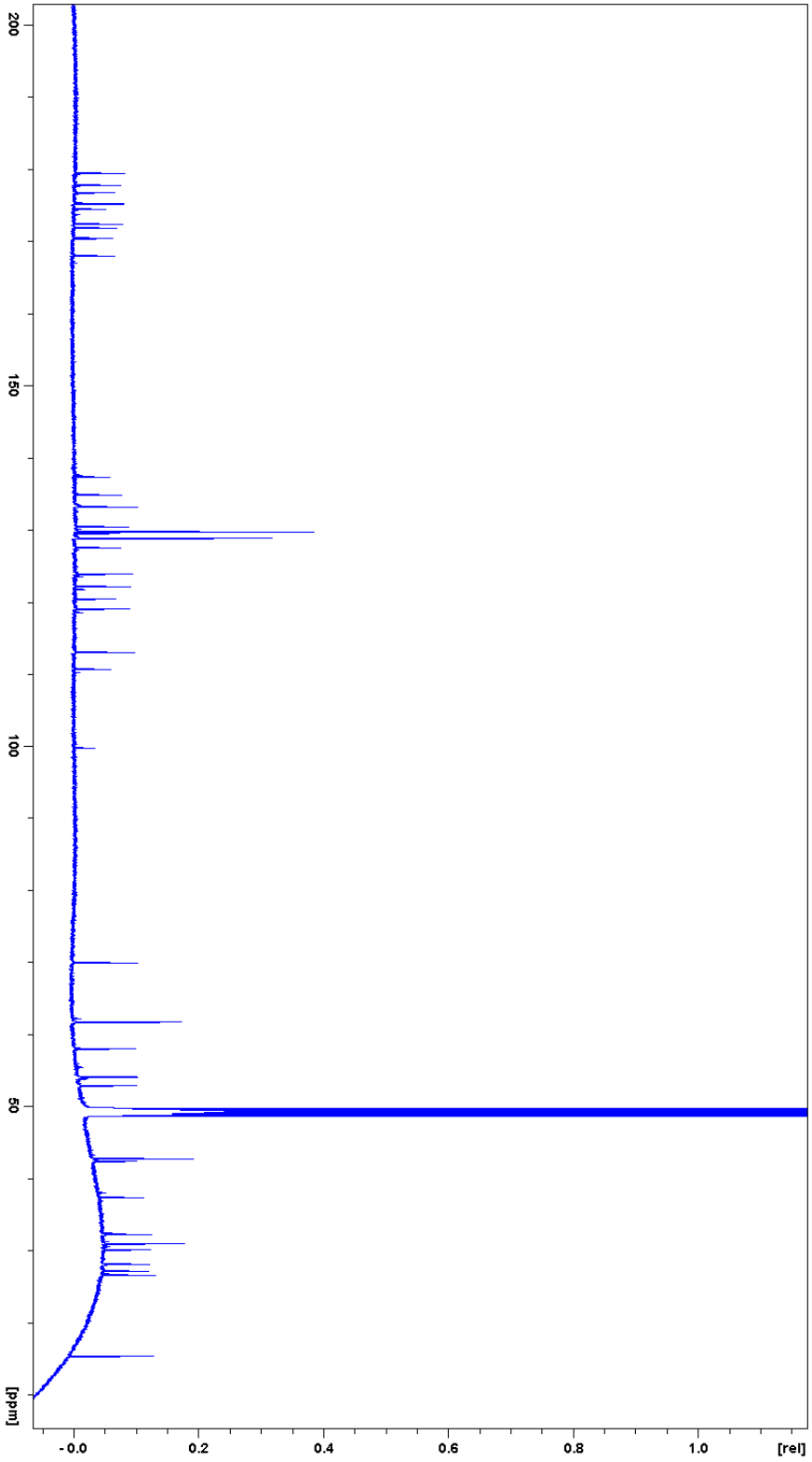
[a]Recorded at 175 MHz; referenced to residual CD_3OD at δ 49.15 ppm. [b]Recorded at 700 MHz; referenced to residual CD_3OD at δ 3.31 ppm. [c] Overlapped.

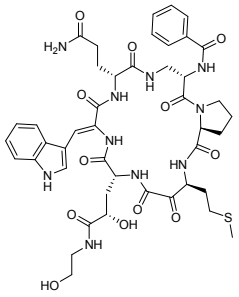
Table S4. Deduced Functions of ORFs in the Jahnellamide Biosynthetic Gene Cluster

ORF	predicted function in jahnellamide biosynthesis	closest protein similarity	source organism	identity [%]	reference
1	membrane protein	Hypothetical protein blr6493	[*Bradyrhizobium japonicum* USDA 110]	55	NP_773133.1
2	SbnB-like diaminopropionate synthetase	ornithine cyclodeaminase	[*Beggiatoa* sp. PS]	44	WP_008474879.1
3	SbnA-like diaminopropionate synthetase	cysteine synthase	[*Paenibacillus mucilaginosus*]	52	YP_006190628.2
4	MFS permease	MFS superfamily protein	[uncultured marine group II euryarchaeote]	31	EHR76320.1
5	transmembrane receptor	TonB-dependent receptor	[*Sorangium cellulosum* So0157-2]	60	YP_008151552.1
6	ferritin-like protein/isomerase	Hypothetical protein	[*Azospirillum* sp. B510]	33	AGS49410.1
jahA	NRPS	NRPS	[*Anabaena variabilis* ATCC 29413]	40	ADL59762.1
jahB	PKS/NRPS hybrid	PKS/NRPS hybrid	[*Corallococcus coralloides* DSM 2259]	42	YP_005370175.1
7	aminomethyltransferase	hypothetical protein	[*Rudaea cellulosilytica*]	50	WP_018971787.1
8	desaturase / amine oxidase	FAD-dependent oxidoreductase	[*Candidatus Koribacter versatilis* Ellin345]	60	YP_590346.1
9	desaturase / amine oxidase	FAD-dependent oxidoreductase	[*Candidatus Koribacter versatilis* Ellin345]	34	YP_590346.1
10	PLP-dependent decarboxylase	putative threonine synthase	[*Anaerolinea thermophila* UNI-1]	24	YP_004173811.1
11	SAM-dependent methyltransferase	SAM-dependent methyltransferase	[*Roseovarius sp.* TM1035]	29	WP_008280394.1

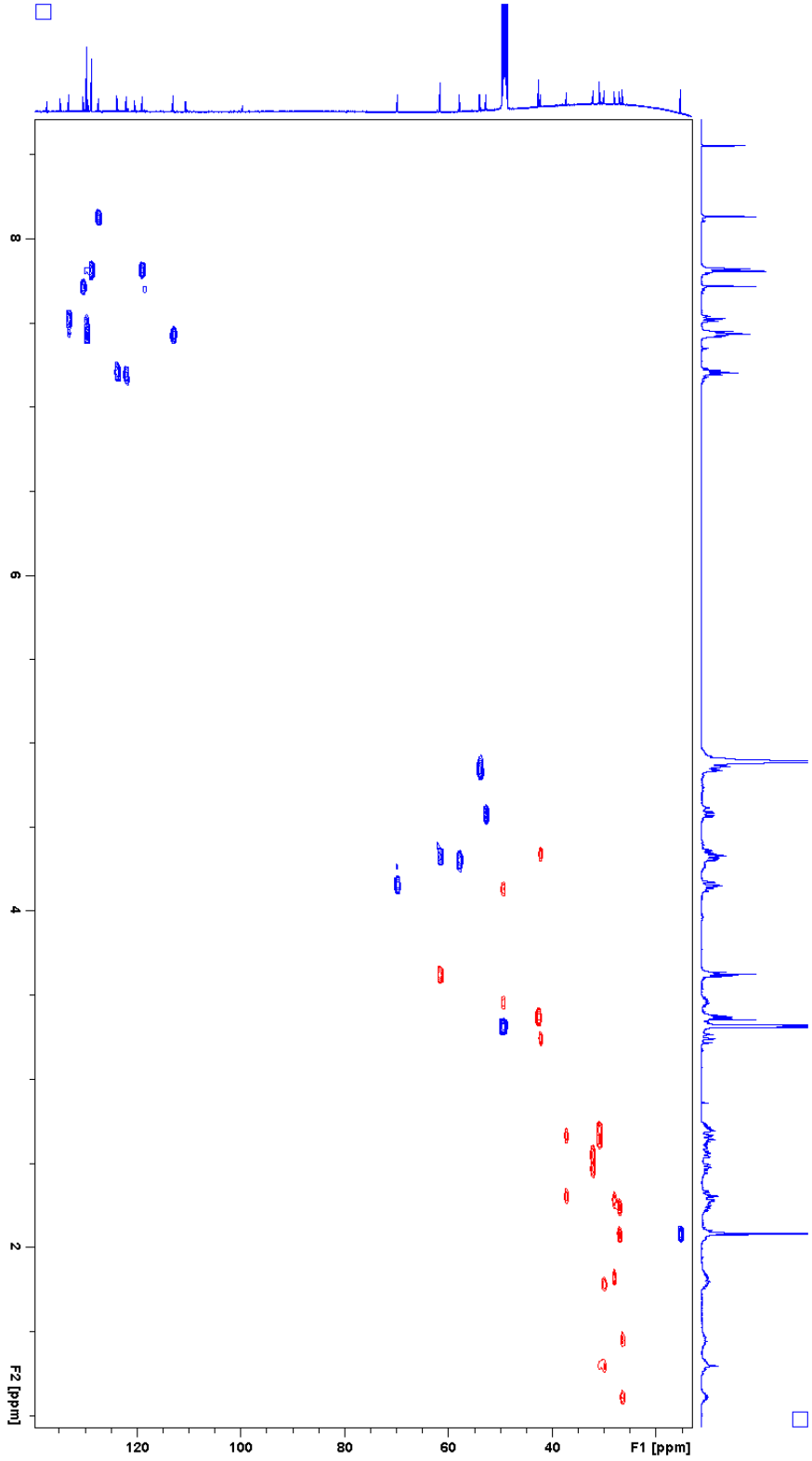


^{1}H NMR spectrum of jahnellamide A (**1**) in CD_3OD



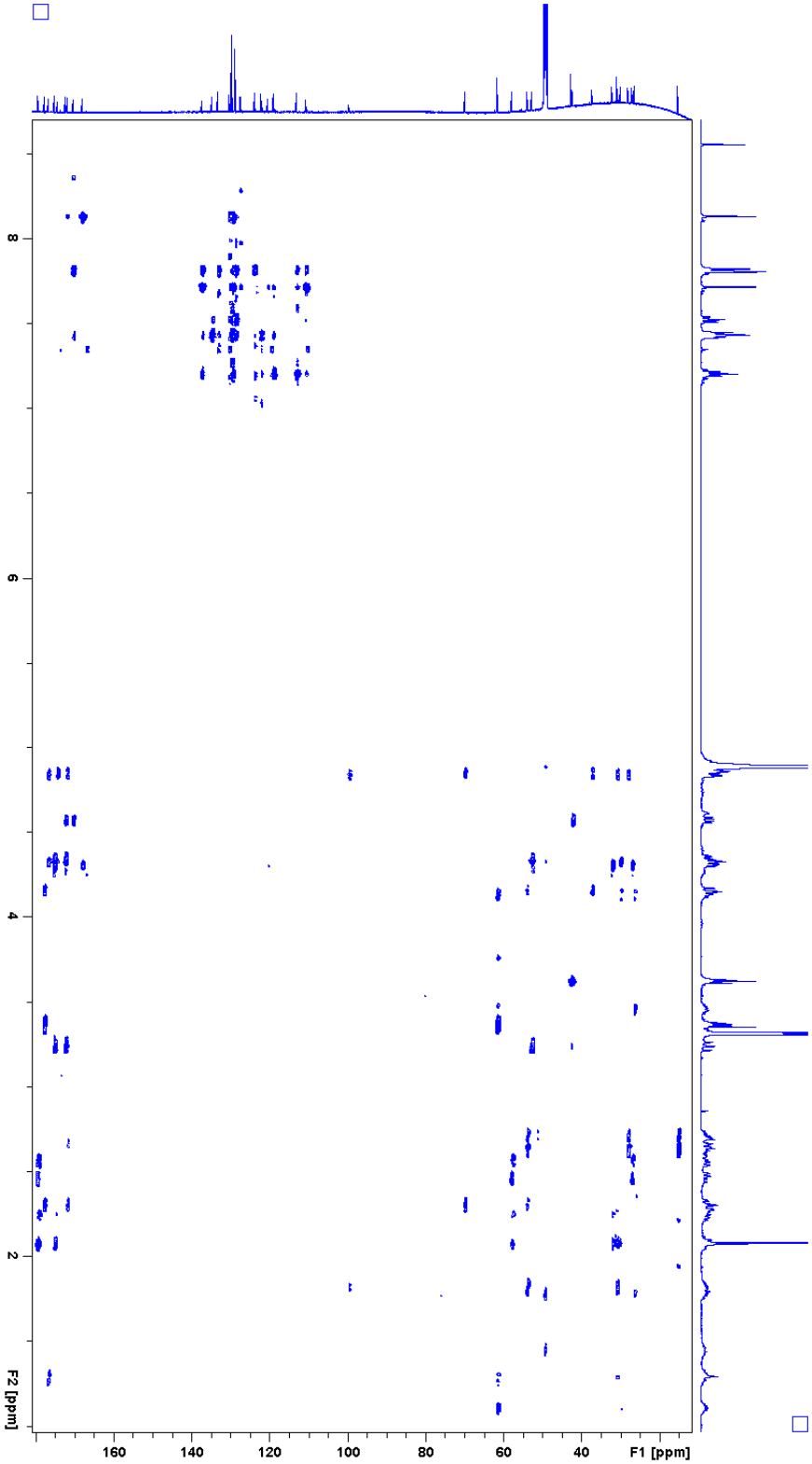


^{13}C NMR spectrum of jahnellamide A (**1**) in CD_3OD

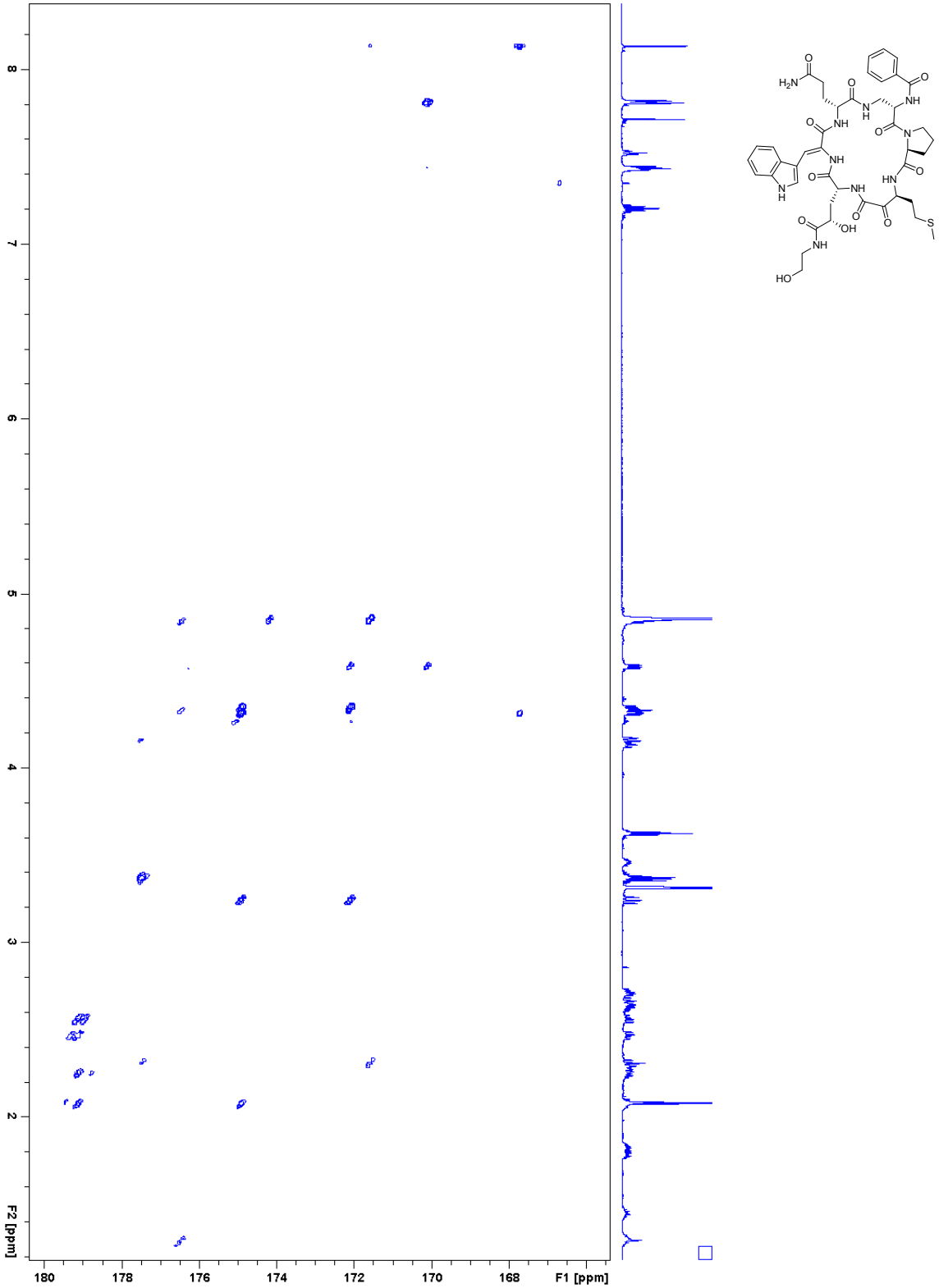


HSQC spectrum of jahnellamide A (**1**) in CD_3OD

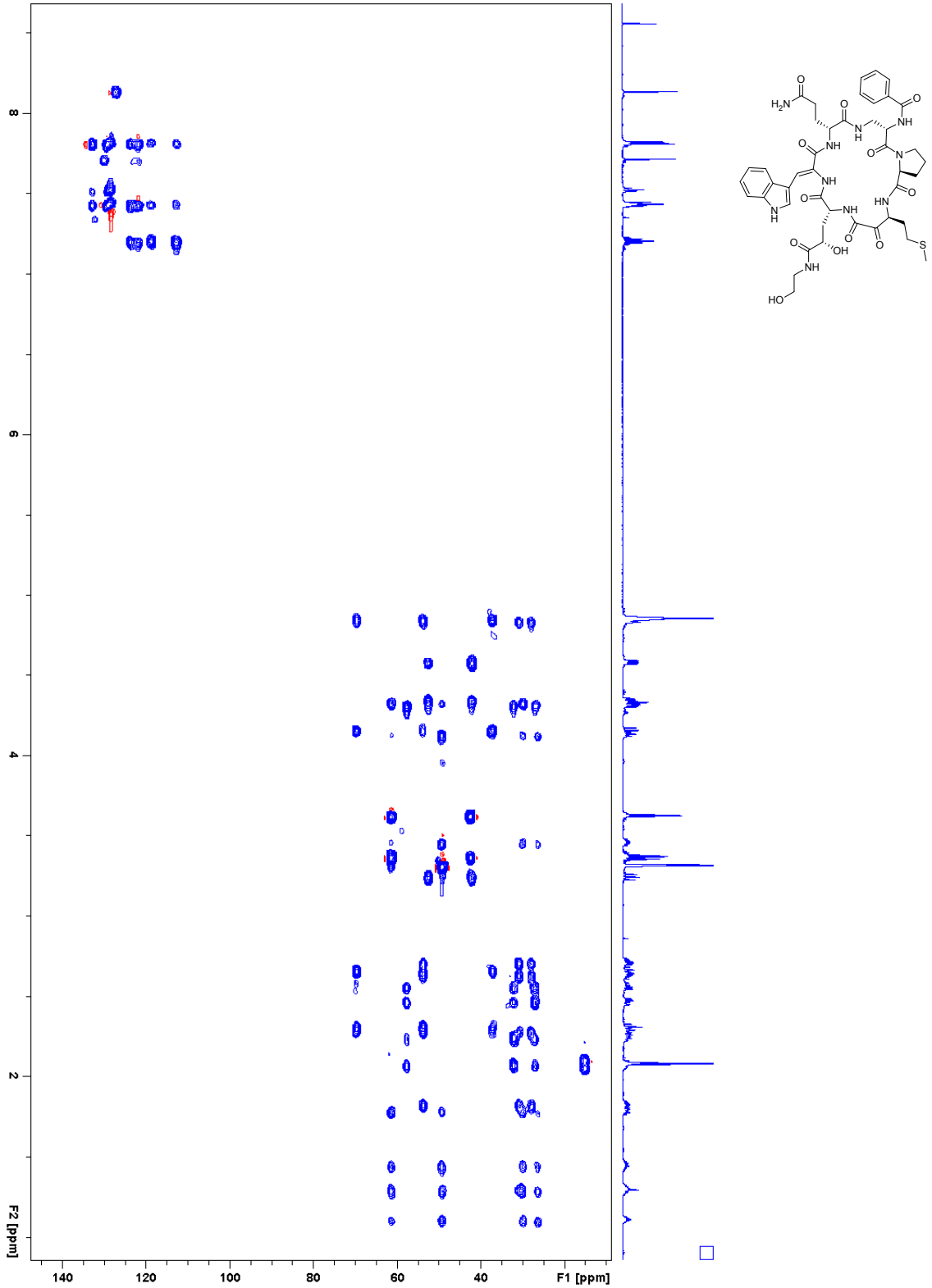
S15



HMBC spectrum of jahnellamide A (**1**) in CD_3OD

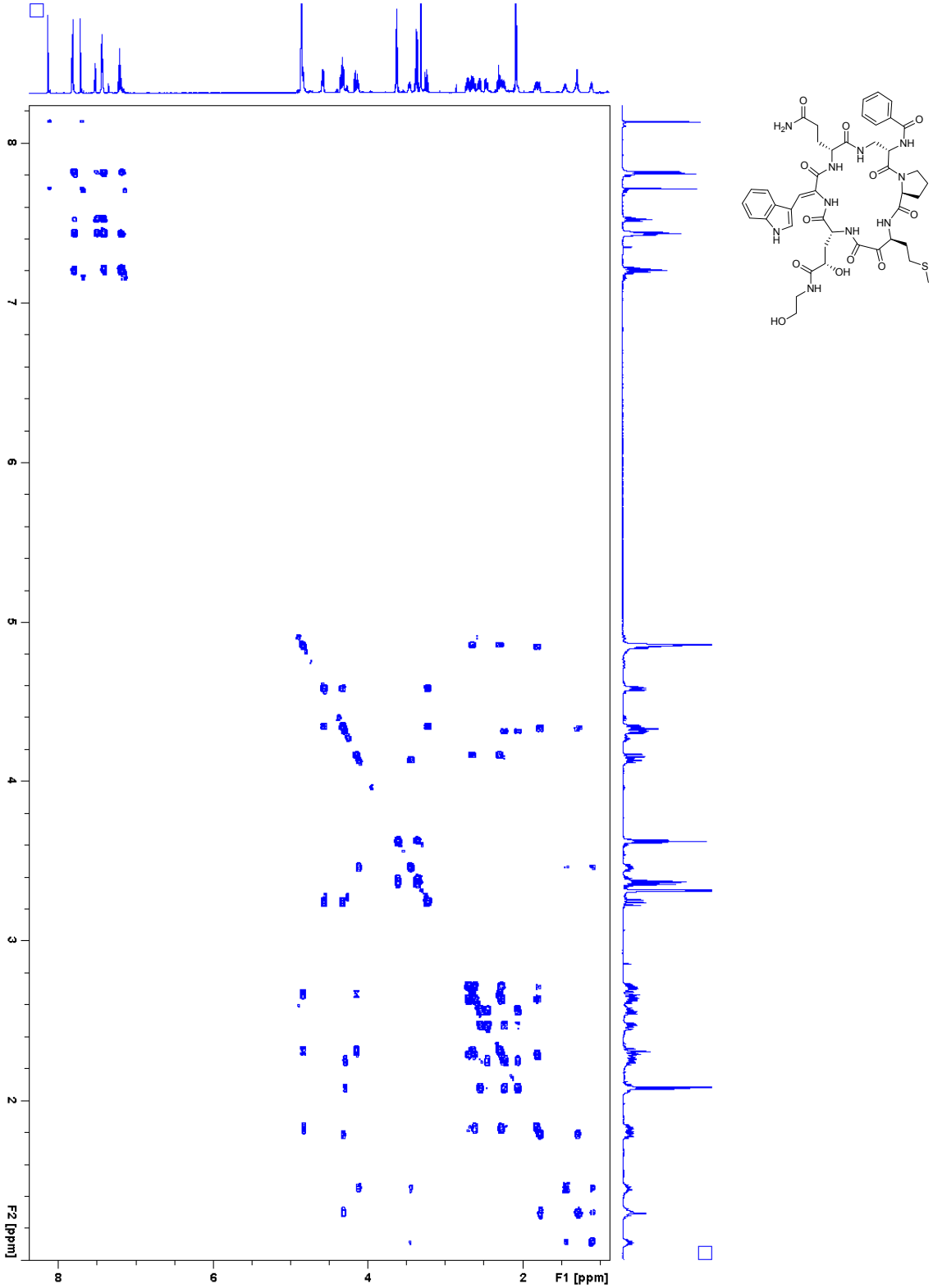


Semi-selective HMBC spectrum of jahnellamide A (**1**) in CD_3OD



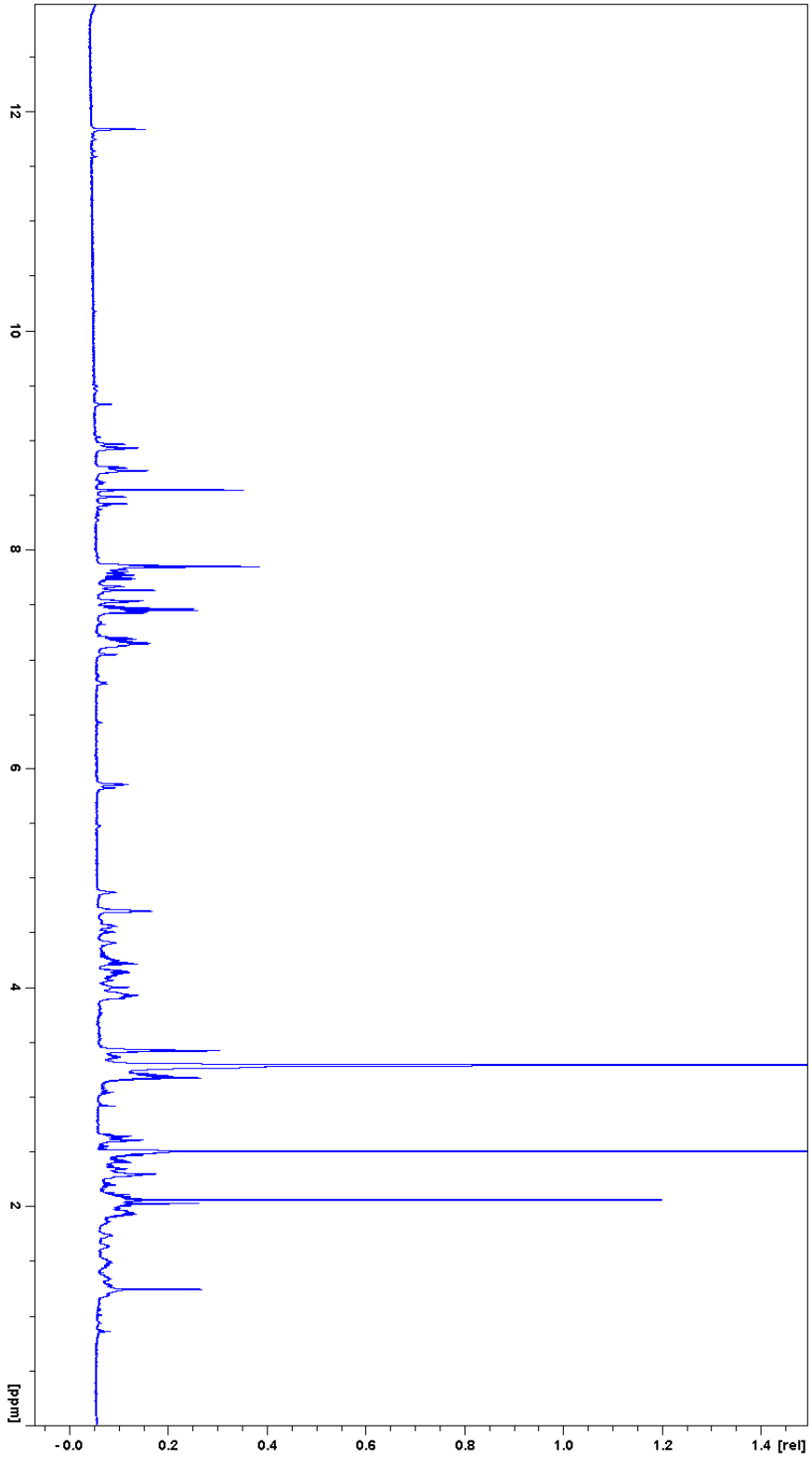
HSQC-TOCSY spectrum of jahnellamide A (**1**) in CD_3OD

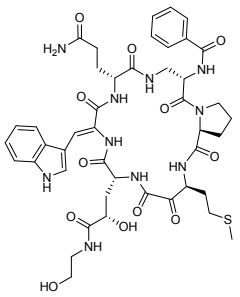
S18



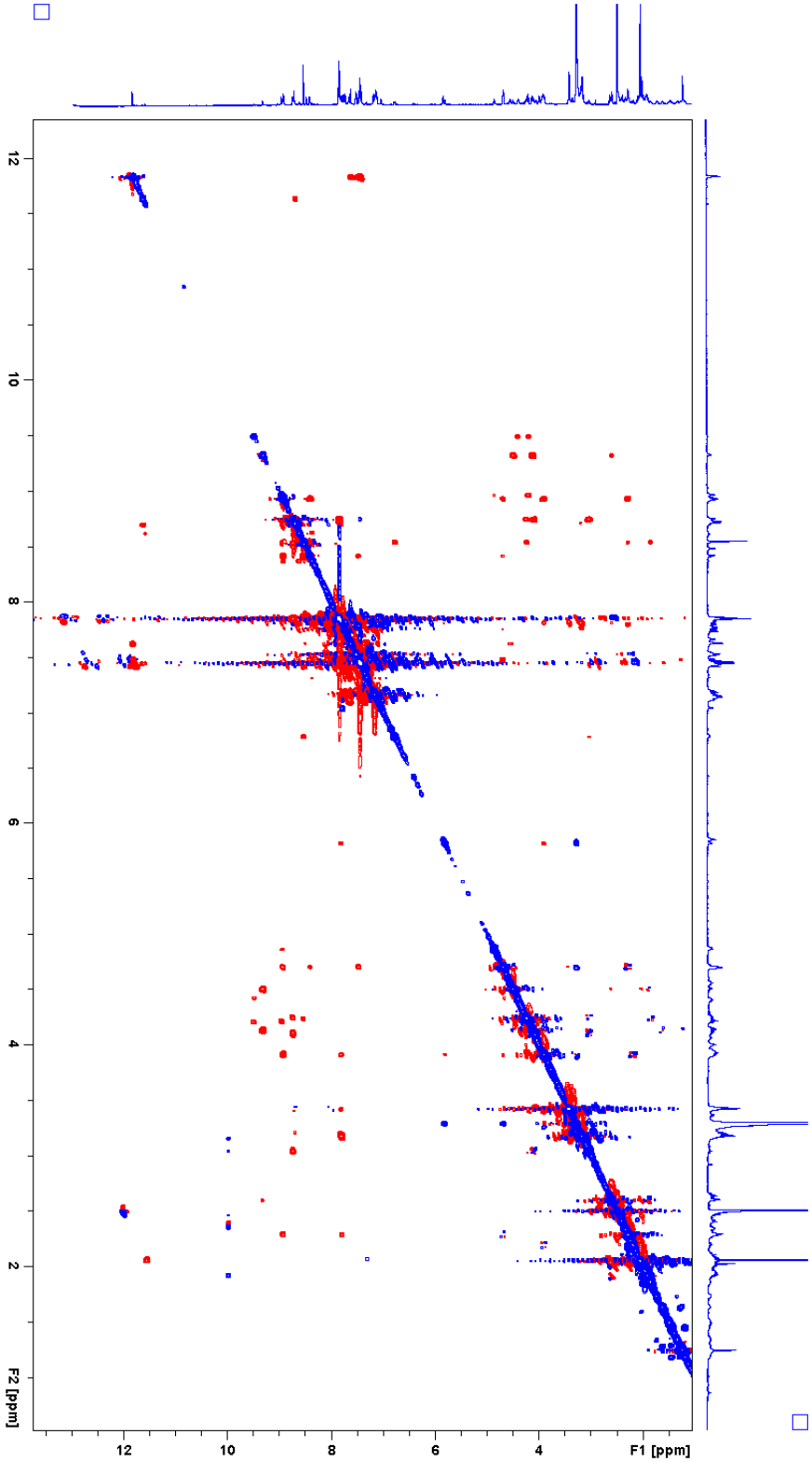
COSY spectrum of jahnellamide A (**1**) in CD_3OD

S19

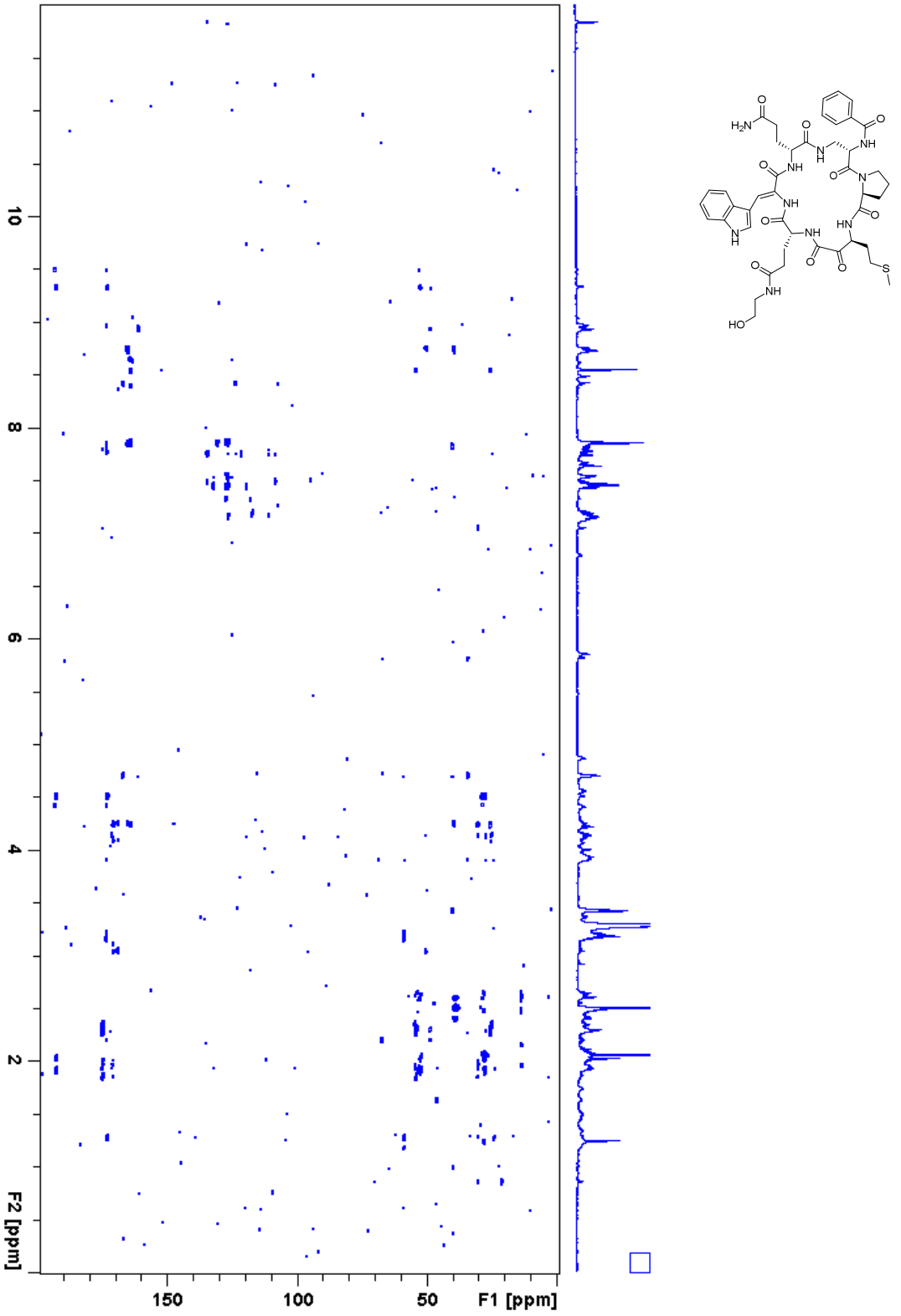




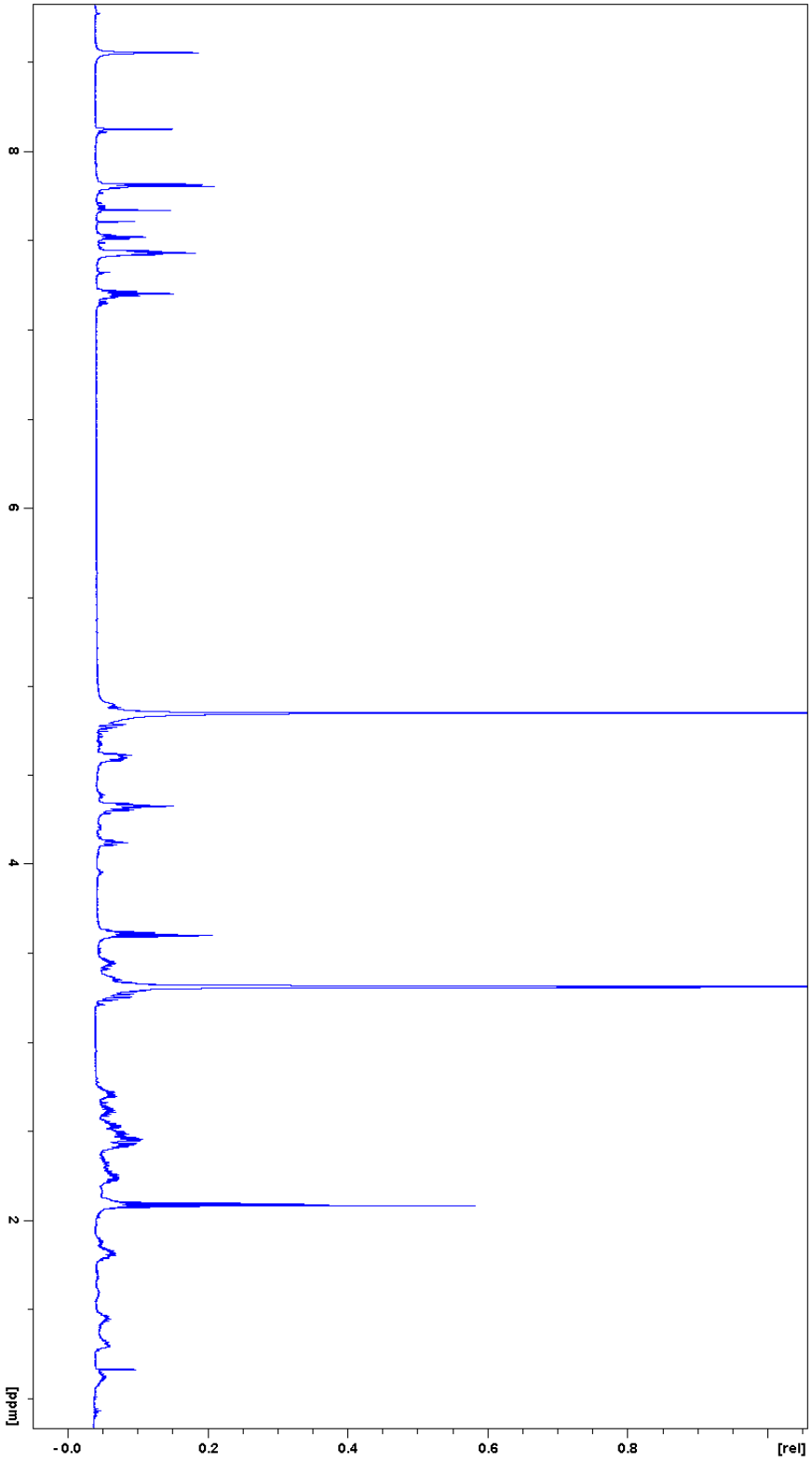
^{1}H NMR spectrum of jahnellamide A (**1**) in DMSO-d_6 (T = 305 K)

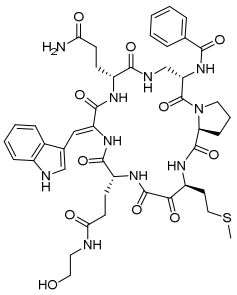


ROESY spectrum of jahnellamide A (**1**) in DMSO-d_6 (T = 305 K)

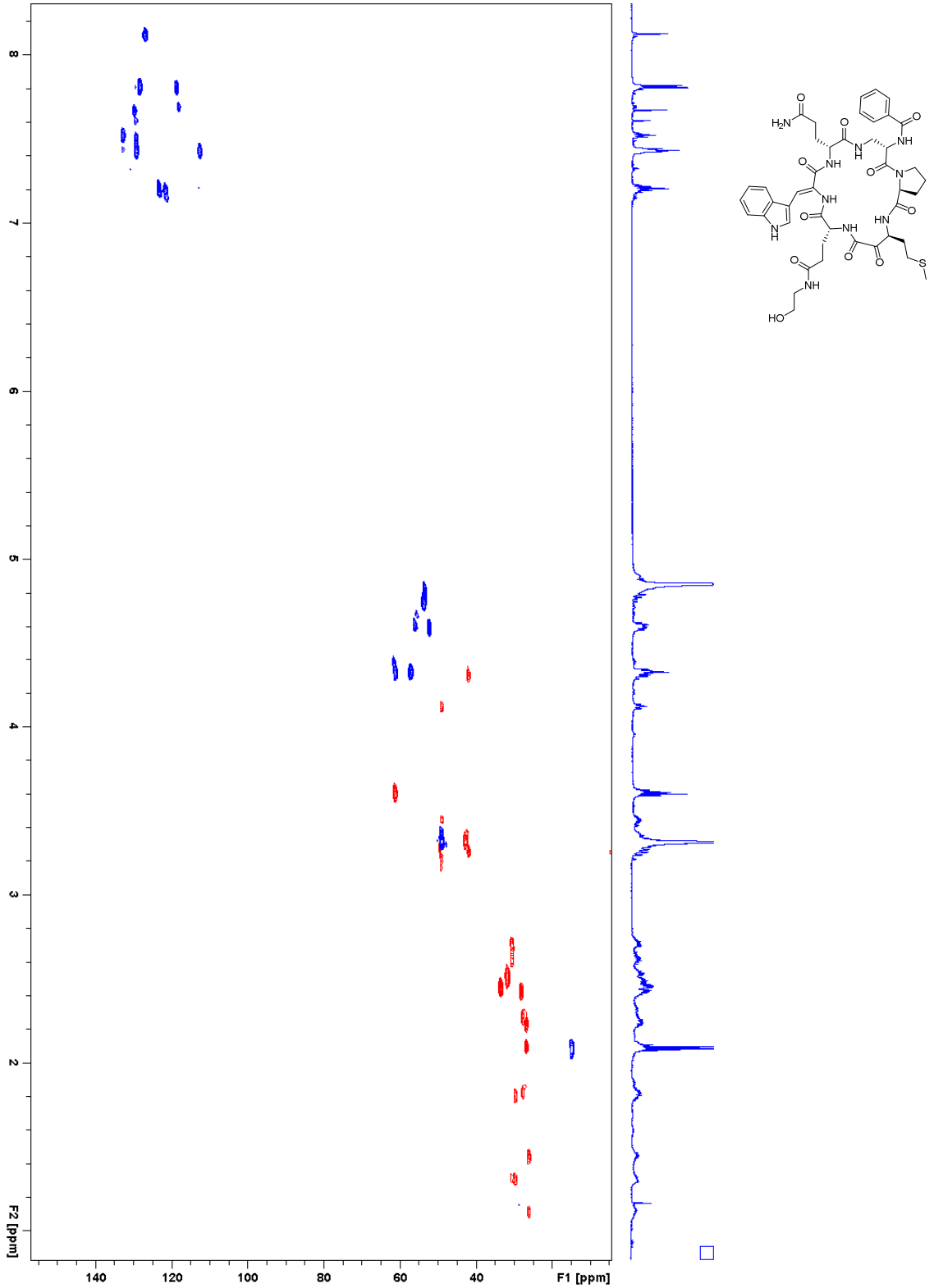


HMBC spectrum of jahnellamide A (**1**) in DMSO-d_6 (T = 305 K)

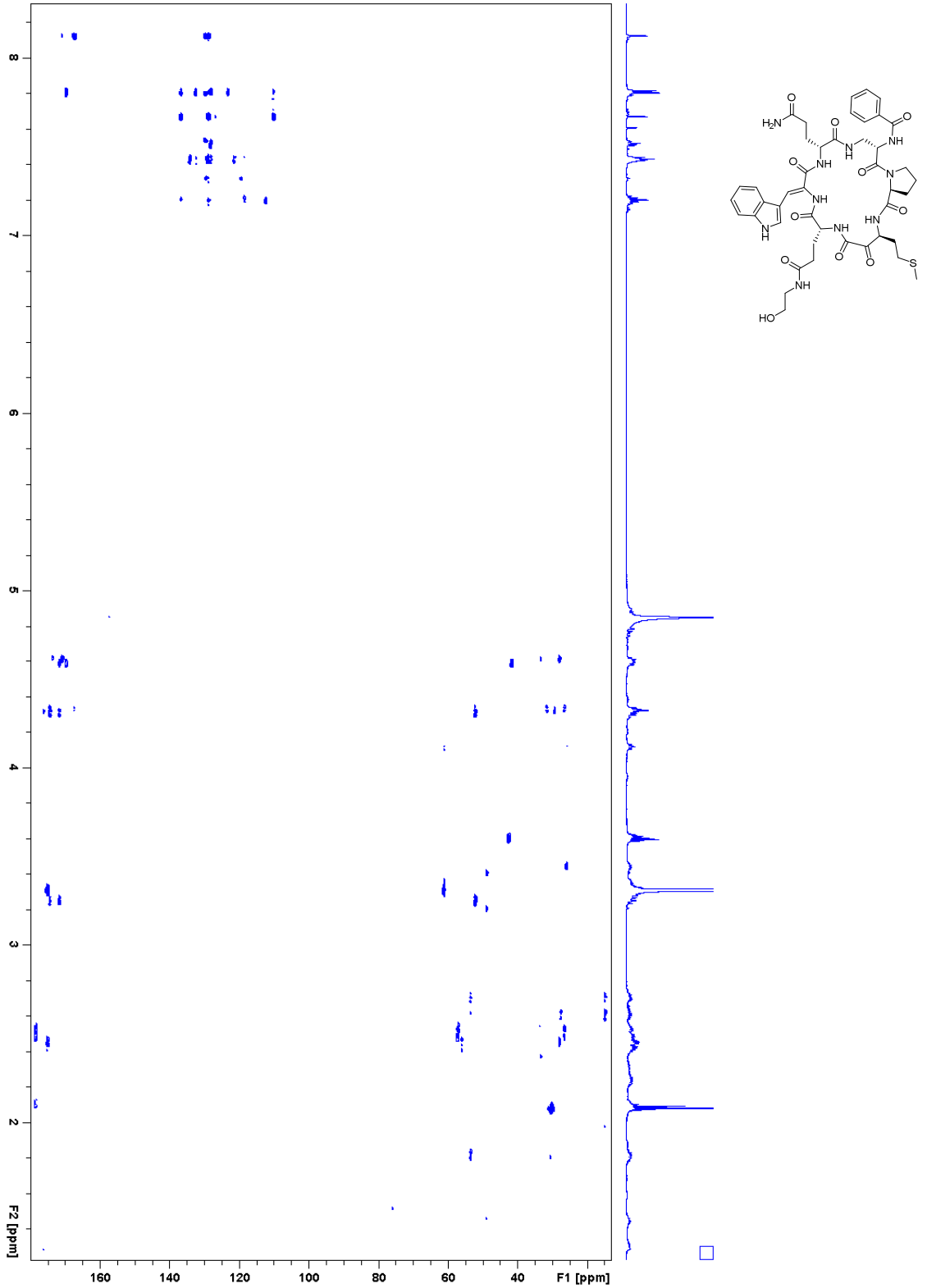




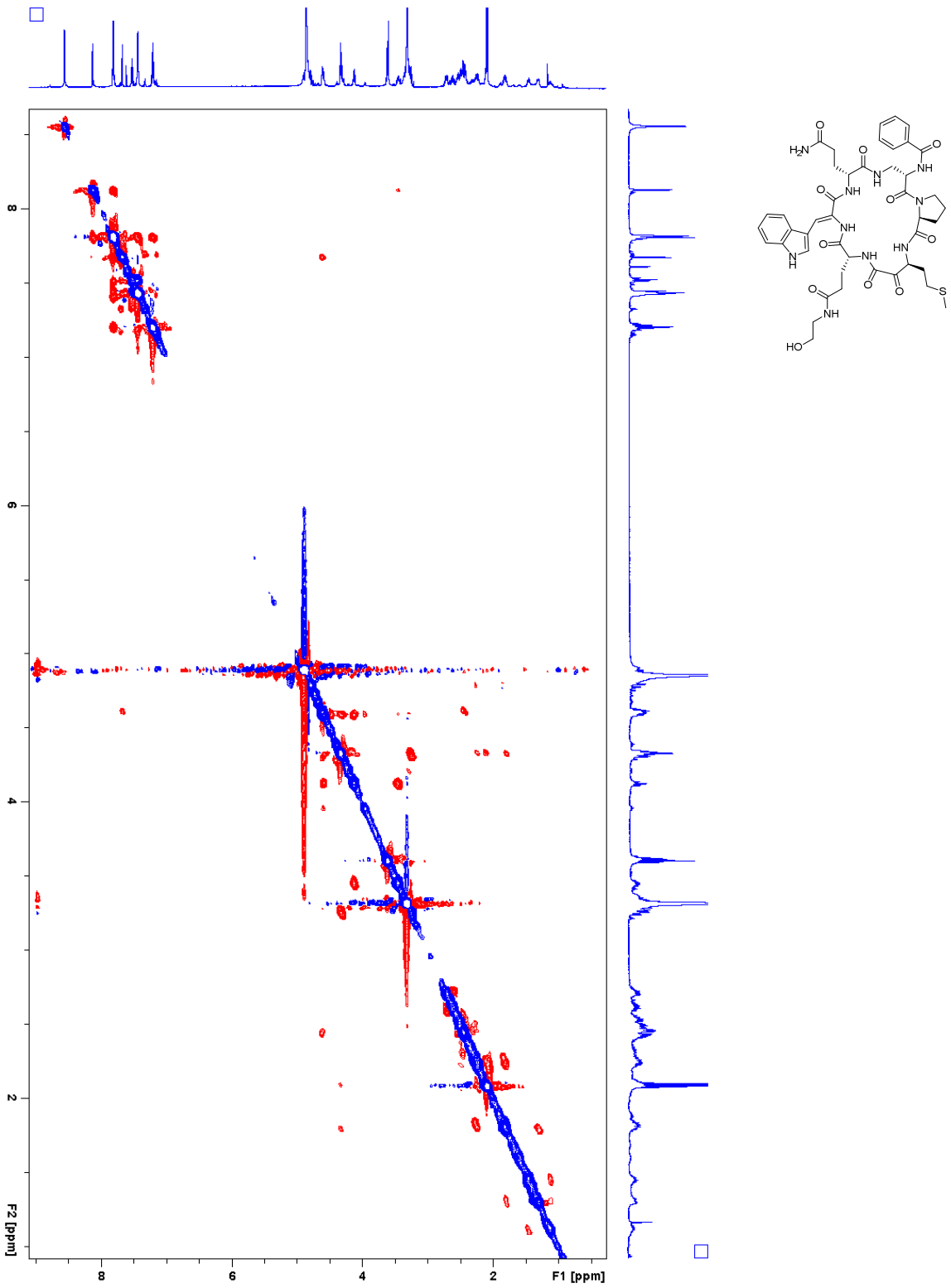
^{1}H NMR spectrum of jahnellamide B (**2**) in CD_3OD



HSQC spectrum of jahnellamide B (**2**) in CD_3OD



HMBC spectrum of jahnellamide B (**2**) in CD_3OD



ROESY spectrum of jahnellamide B (**2**) in CD_3OD

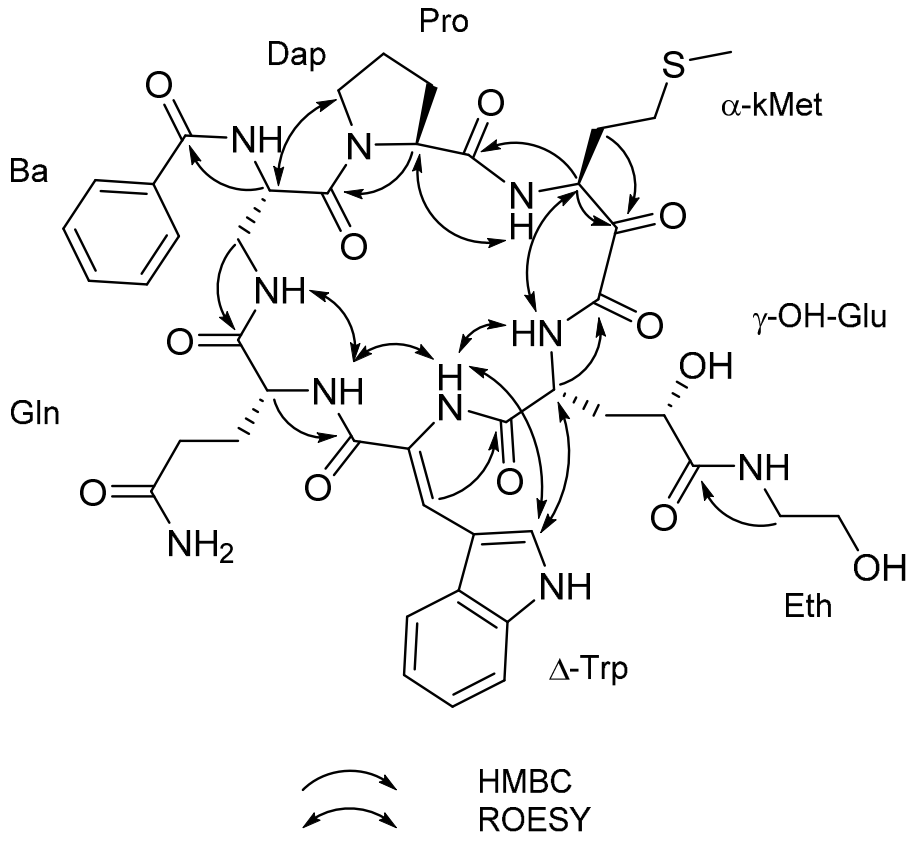


Figure S1. Key HMBC and ROESY correlations for **1**

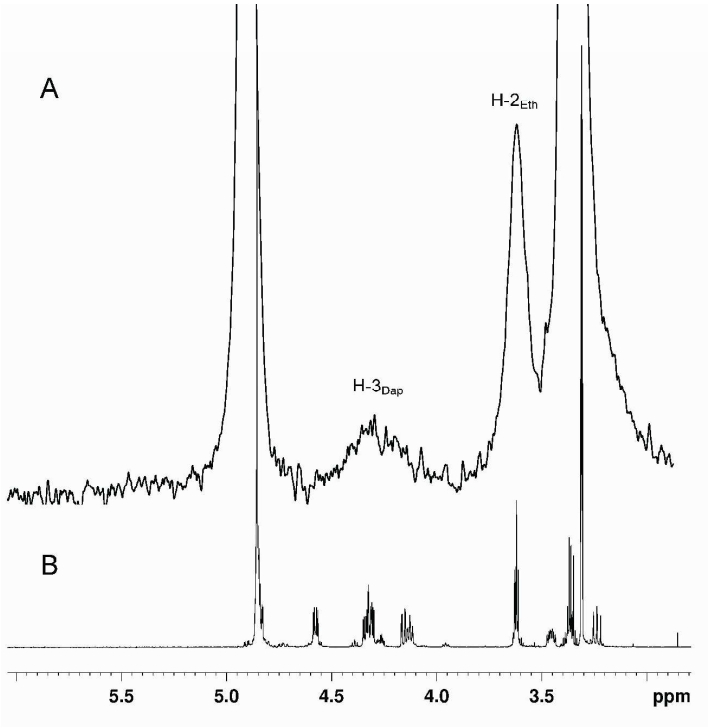


Figure S2. Partial NMR spectra of **1** obtained from feeding with L-[2,3,3-$^2H_3$]serine. A) ^1H NMR (CD_3OD, 700 MHz) spectrum and B) ^2H NMR (MeOH, 107.7 MHz) spectrum showing enhancements for protons at δ 4.34 (H-3_{Dap}) and δ 3.62 (H-2_{Eth}).

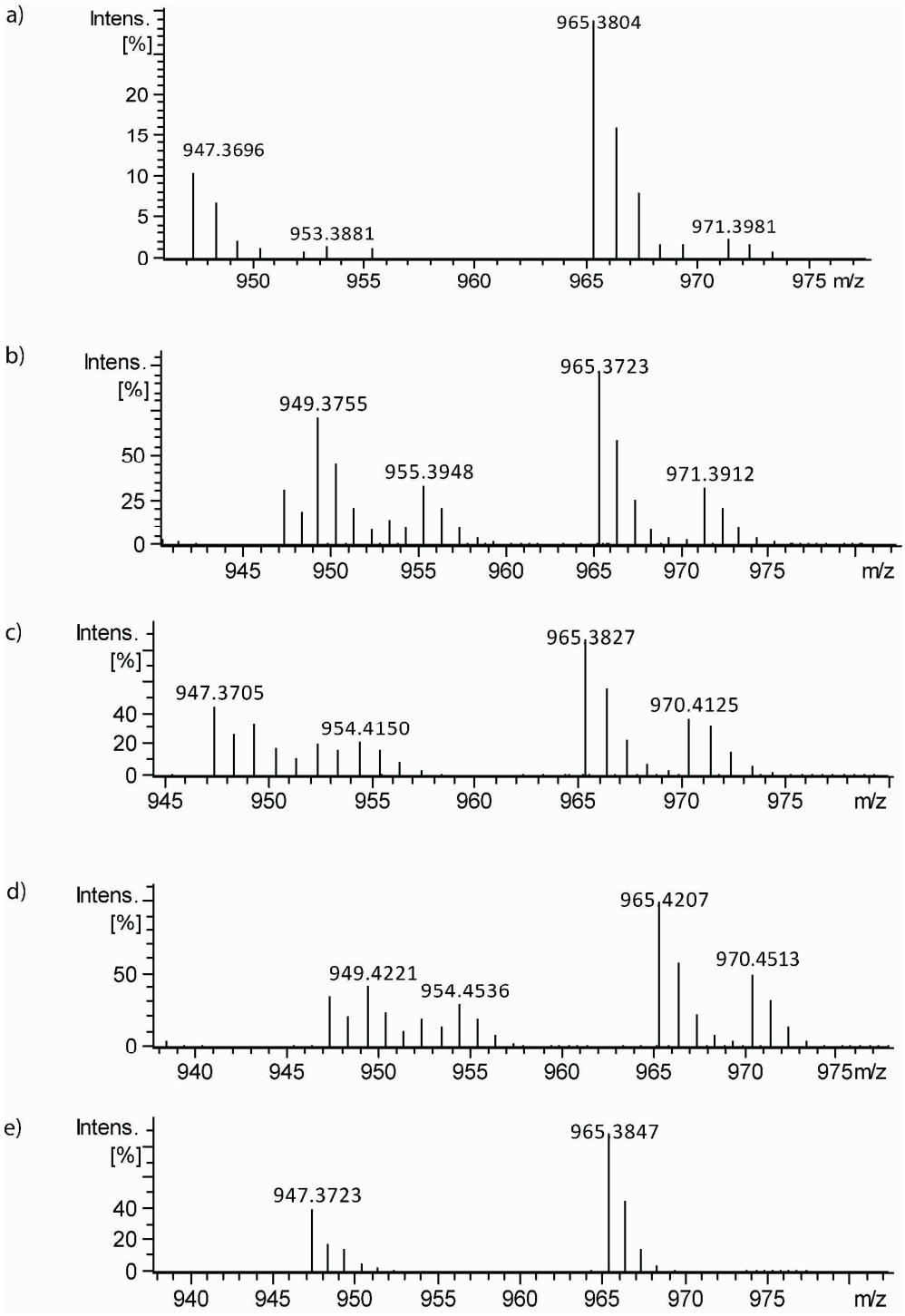


Figure S3. ESI-MS spectra of jahnellamide A obtained from feeding experiments with a) L-[U-^{13}C, ^{15}N]methionine, b) L-[ring-$^{13}C_6$]phenylalanine, c) [$^{2}H_5$]benzoic acid, d) [$^{2}H_7$]cinnamic acid, and e) non labeled jahnellamide A.

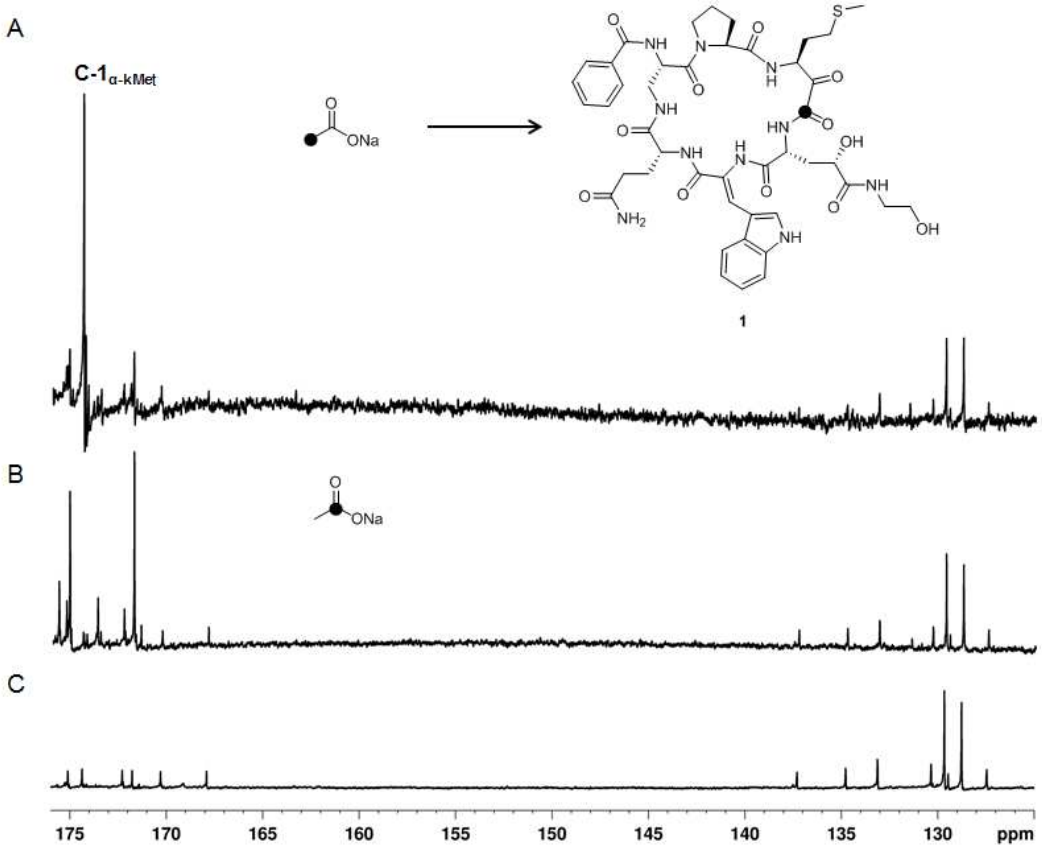


Figure S4. Downfield region of ^{13}C NMR spectra of jahnellamide A enriched with A) [2-^{13}C] acetate, B) [1-^{13}C] acetate, and C) non labeled jahnellamide A. Black dots indicated labeled carbons.

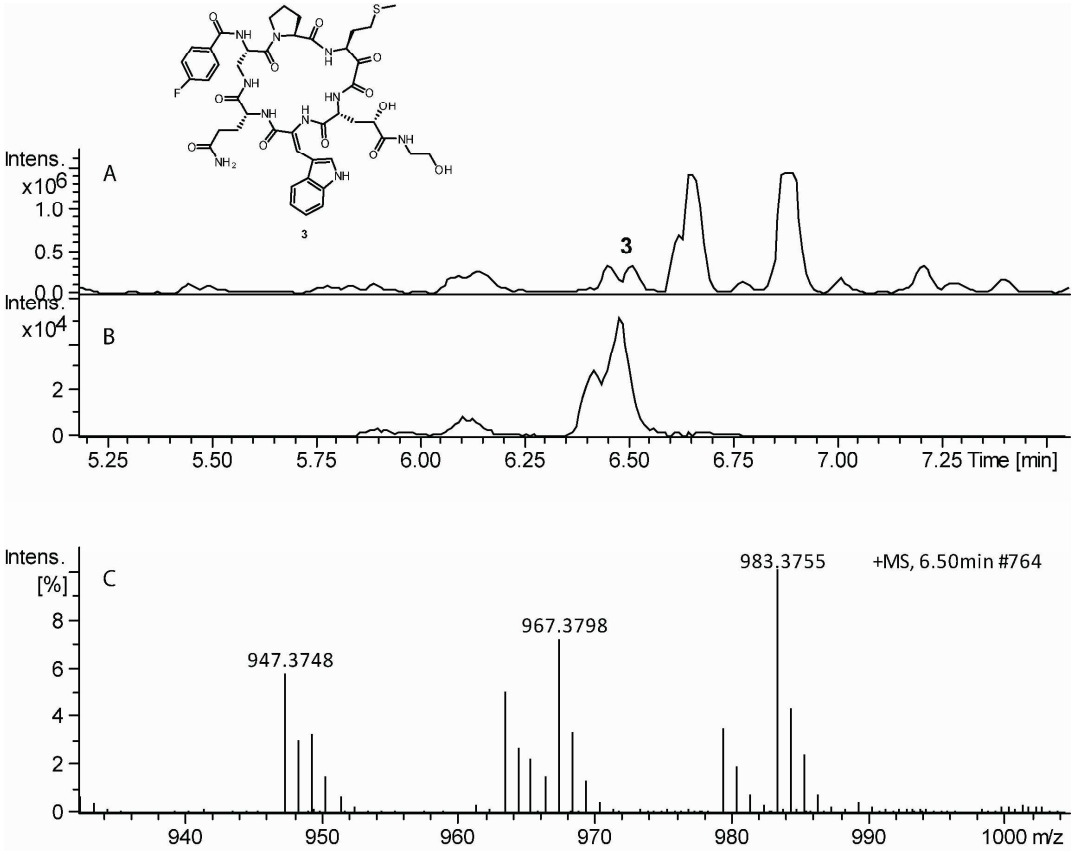


Figure S5. LC-MS analysis of a methanol extract of SBSr007 showing the presence of compound **3**. A) UV Chromatogram. B) Extracted ion chromatogram for *m*/*z* 983.3 $[M+H]^+$. C) MS spectrum of **3**.

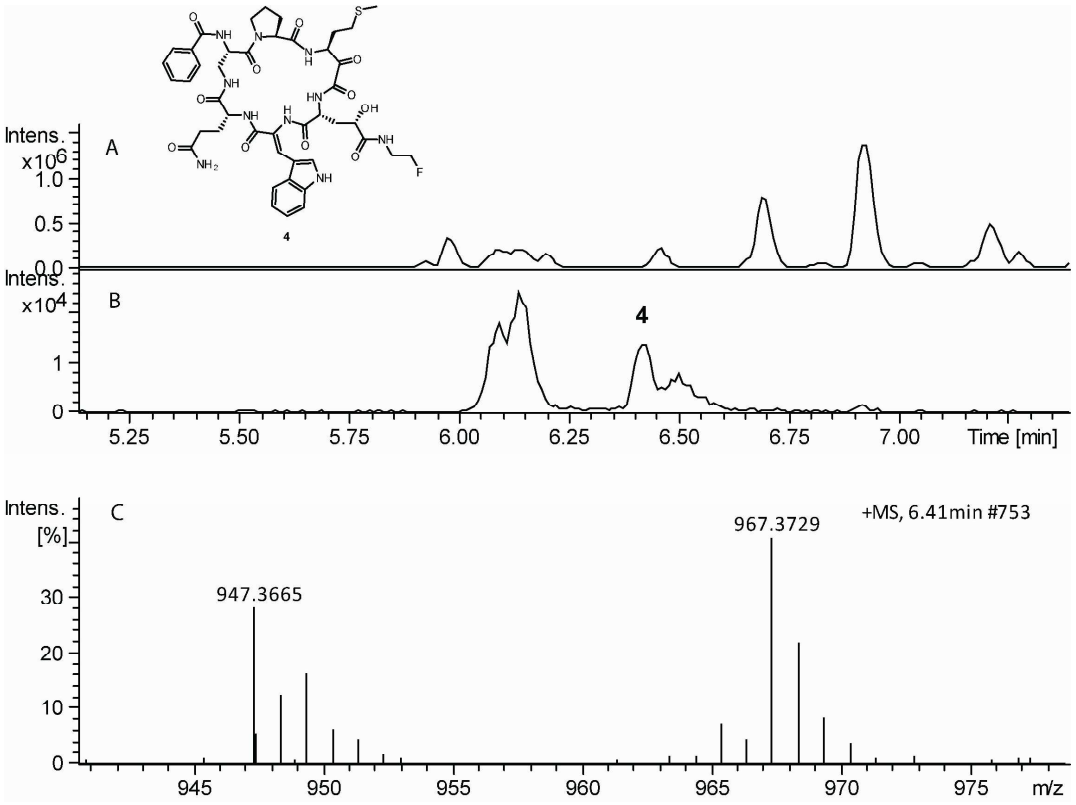


Figure S6. LC-MS analysis of a methanol extract of SBSr007 showing the presence of compound **4**. A) UV Chromatogram. B) Extracted ion chromatogram for *m/z* 967.3 $[M+H]^+$. C) MS spectrum of **4**.