Heterologous Expression for the Discovery and Biosynthesis Studies of Myxobacterial Natural Products

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

Von
Yunsheng Gao
Saarbrücken
2023

Tag des Kolloquiums: 21. Juli 2023

Dekan: Prof. Dr. Ludger Santen

Berichterstatter: Prof. Dr. Rolf Müller

Prof. Dr. Andriy Luzhetskyy

Akad. Mitglied: Dr. Stefan Boettcher

Vorsitz: Prof. Dr. Uli Kazmaier

Acknowledgements

I have had the pleasure of working in HIPS-MINS group for the last four years. First of all, I would like to express my deepest appreciation to my supervisor, Prof. Rolf Müller, for providing me with this opportunity and also for his invaluable inspiration and guidance throughout my PhD studies. I am also extremely grateful to Prof. Andriy Luzhetskyy and Dr. Chengzhang Fu for their attendance at my thesis committee meetings and their helpful suggestions.

I would like to extend my gratitude to Dr. Ronald Garcia for providing me with myxobacterial strains and assistance with strain cultivation, Dr. Fabian Panter for guiding me with metabolome analysis and compound purification, Dr. Asfandyar Sikandar for his expertise in protein expression, insightful discussions and invaluable proofreading assistance of my thesis, Joy Birkelbach for her contribution to the disorazole project and translation of the abstract, Dr. Chantal D. Bader and Dr. Christine Walt for their collaboration in the sorangibactin project, and all group members for their assistance, advice and discussions.

I am indebted to Prof. Fu Yan for his invaluable guidance when I first arrived in Germany, as well as to Prof. Youming Zhang, Prof. Hailong Wang, and Dr. Zaichao Ma from Shandong University, who assisted me in applying for a PhD scholarship.

I would like to thank all of my colleagues and friends who have provided me with consistent support over the past years, particularly Christina Decker and Ellen Merckel, all technical assistants, and the analytic team. Special thanks go to Dr. Jun Feng and Xi Zheng for the sincere and unforgettable friendship.

I would also like to acknowledge the China Scholarship Council (CSC) and the Helmholtz Centre for Infection Research (HZI) for providing me with my PhD stipend.

Last but not least, I am deeply grateful to my family for their endless love and support, especially my wife for accompanying me to Germany and taking care of our newborn baby.

Life and research have not been always easy, especially during the COVID-19 pandemic, which highlights the supports mentioned above are particularly precious to me.

Yunsheng Gao
2023, Saarbrücken

Vorveröffentlichungen der Dissertation

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

Yunsheng Gao, Joy Birkelbach, Chengzhang Fu, Jennifer Herrmann, Herbert Irschik, Bernd Morgenstern, Kerstin Hirschfelder, Ruijuan Li, Youming Zhang, Rolf Jansen, and Rolf Müller, The disorazole Z family of highly potent anticancer natural products from *Sorangium cellulosum*: structure, bioactivity, biosynthesis and heterologous expression, *Microbiology Spectrum*, 2023, accepted

Yunsheng Gao, Joy Birkelbach, Asfandyar Sikandar, Youming Zhang, and Rolf Müller, Unusual oxidations mediated by a Coq4 domain-containing hypothetical protein in disorazole biosynthesis, manuscript in preparation

Yunsheng Gao, Christine Walt, Chantal D. Bader, and Rolf Müller, Genome-guided discovery of the myxobacterial thiolactone-containing sorangibactins, *ACS Chemical Biology*, 2023, 18(4), 924-932, DOI: 10.1021/acschembio.3c00063

Abstract

Myxobacteria have immense underexplored biosynthetic potential for the discovery of novel natural products as revealed by large-scale genome sequencing. Due to their slow growth and intractability for *in situ* genetic engineering of most myxobacterial strains, heterologous expression could represent an effective approach for genome mining of the huge untapped gene cluster resources and investigation of natural product biosynthesis. In this thesis, the disorazole biosynthetic pathways from *S. cellulosum* strains were characterized by extensive gene cluster engineering and subsequent heterologous expression in *M. xanthus* DK1622, resulting in the generation of several novel derivatives. In addition, the tailoring *O*-methylation and unusual oxidations mediated by Coq4 domain-containing hypothetical proteins were uncovered. Furthermore, heterologous expression of a coelibactin-like NRPS gene cluster from the *Sorangiineae* strain MSr11367 in *M. xanthus* DK1622 led to the discovery of sorangibactins featuring an unprecedented thiolactone moiety. Molecular engineering of the sorangibactin biosynthetic gene cluster revealed unusual biosynthesis features, such as the P450 enzyme-mediated oxazoline dehydrogenation and the thioesterase domain-mediated C-terminal γ-thiolactone formation. The novel (bio)chemistries presented in this study can serve as a promising starting point for further investigations.

Zusammenfassung

Myxobakterien verfügen über ein immenses, noch nicht ausreichend erforschtes Biosynthesepotential für die Entdeckung neuer Naturstoffe, was durch angelegte Genomsequenzierung gezeigt werden konnte. Aufgrund ihres langsamen Wachstums und der Schwierigkeit, die meisten Myxabakterien *in situ* genetisch zu verändern, könnte die heterologe Expression einen effektiven Ansatz darstellen, um die riesigen, noch nicht erschlossenen Gencluster-Ressourcen und die Biosynthese von Naturstoffen zu erforschen. In dieser Arbeit wurden die Disorazol-Biosynthesewege von *S. cellulosum*-Stämmen durch umfangreiche Modifizierung des Genclusters und anschließende heterologe Expression in *M. xanthus* DK1622 charakterisiert, was zur Bildung mehrerer neuer Derivate führte. Darüber hinaus wurden die *O*-Methylierung und ungewöhnliche Oxidation aufgedeckt, die durch hypothetische Proteine mit Coq4-Domäne vermittelt werden. Darüber hinaus führte die heterologe Expression eines Coelibactin-ähnlichen NRPS-Genclusters aus dem *Sorangiineae*-Stamm MSr11367 in *M. xanthus* DK1622 zur Entdeckung von Sorangibactinen, die eine beispiellose Thiolactoneinheit aufweisen. Durch molekularbiologische Eingriffe in das Sorangibactin-Biosynthese Gencluster wurden ungewöhnliche Biosynthesemerkmale aufgedeckt, wie die durch das P450-Enzym vermittelte Oxazolindehydrierung und die durch die Thioesterasedomäne vermittelte C-terminale γ-Thiolactonbildung. Die in dieser Studie vorgestellten neuen (Bio)chemien können als vielversprechender Ausgangspunkt für weitere Untersuchungen dienen.

Table of Contents

Chapter 1. Introduction

1.1 Natural products as sources of new drugs

Natural products, in general, are biosynthetic organic-compounds with low molecular weight that are produced mainly during secondary metabolism, in most cases, by plants, fungi and bacteria.[1] These biomolecules serve various functions, such as defense and competition (i.e. “chemical warfare”), communication and regulation (e.g. quorum sensing), and nutrient acquisition (e.g. siderophores). During the long period of evolutionary selection, natural products have been optimized to possess privileged structures that enable them to bind to specific biological targets and exert a variety of functions, such as antibacterial, antifungal, antiviral, cytotoxic, insecticidal, and herbicidal activities.[2,3] Consequently, natural products and their derivatives are a treasure trove for the discovery and development of new drugs and agrochemicals.[4,5] In fact, among all 1881 approved drugs between January 1981 and September 2019, approximately half of them are natural products related (unaltered natural products, semisynthetic derivatives, or synthetic drugs based on natural products) (Figure 1), and it is noteworthy that this proportion is even higher for antibacterial and small anticancer drugs.[6] As the accelerated antibiotic resistance is becoming a big threat worldwide, new drugs are urgently needed.[7] As a consequence, the continuous discovery of novel natural products is of great importance for pharmaceutical research.[8]

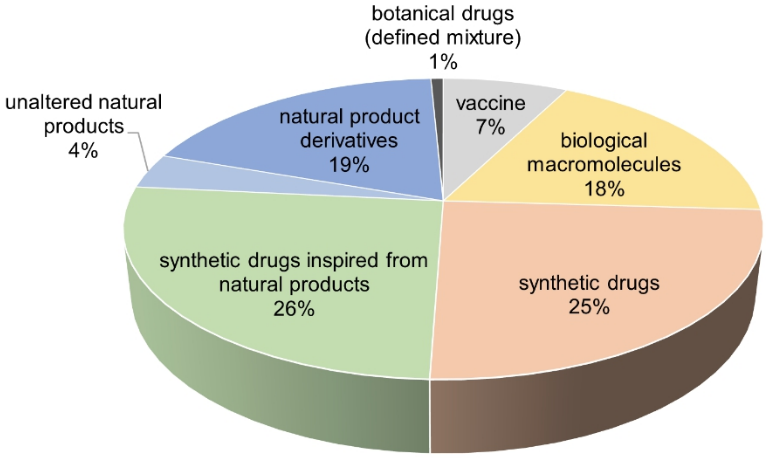
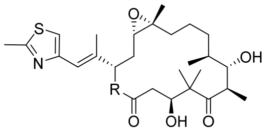


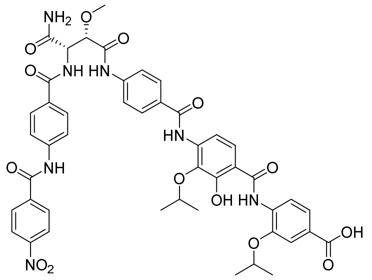
Figure 1 All 1881 approved drugs from 01-01-1981 to 30-09-2019. This figure is adapted from Reference [6].

1.2 Myxobacteria as underexploited sources of natural products

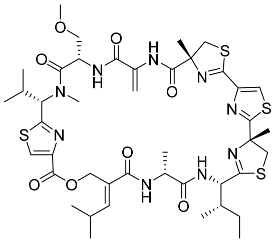
Myxobacteria are classified in the order *Myxococcales*, which belongs to Gram-negative δ-proteobacteria and can be further divided into three suborders: *Cystobacteraceae*, *Nannocystineae* and *Sorangiineae*. These microorganisms are mostly soil-dwelling and possess several distinguished features, such as gliding motility on agar surface, “social behavior” mediated by chemical communication, morphological differentiation during complex life cycle, and the large size of genome (approximately 10 Mb on average). Under starvation conditions, the rod-shaped vegetative cells aggregate together, leading to the formation of fruiting bodies and myxospores.[9] These unique characteristics make myxobacteria of great interest for basic microbiology research. In addition, myxobacteria are known as proficient producers of secondary metabolites with diverse structures and various biological activities (Figure 2), which are of great interest for natural product research.[10,11]



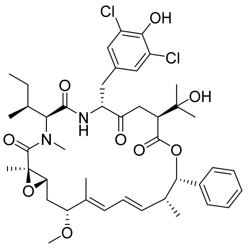
Epothilone B (R=O), from *Sorangium cellulosum*
Ixabepilone/Ixempra® (R=NH), US FDA 2007
anti-cancer durg



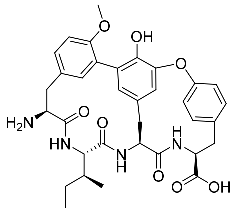
Cystobactamid 919-1, from *Cystobacter* sp. Cbv34
topoisomerase inhibitor, antibacterial activity



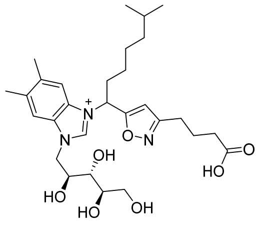
Thiamyxin A, from *Myxococcaceae* strain MCy9487
RNA virus inhibitor, antiviral activity



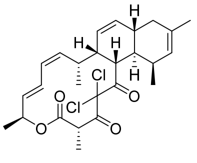
Nannocystin A from *Nannocystis* sp. MB1016
elongation factor 1 inhibitor, antifungal and cytotoxic activities



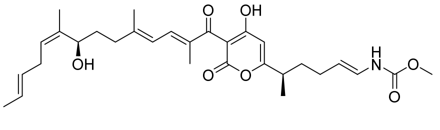
Cittilin A, unusual RiPP from *Myxococcus xanthus*
Prokaryotic carbon storage regulator A inhibitor



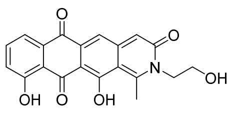
Myxadazole A, from *Myxococcus* sp. SDU36
Cardiovascular activity



Chlorotonil A, from *Sorangium cellulosum* So ce1525
antimalarial and antibacterial activities



Corallopyronin A, from *Corallococcus coralloides* Cc c127
RNA polymerase inhibitor, antibacterial activity



Pyxidicycline A, from *Pyxidicoccus fallax* An d48
topoisomerase inhibitor, cytotoxic activity

Figure 2 Representative myxobacterial natural products.

Myxobacteria-derived natural products occupy a large proportion of chemical space. A wide variety of structurally diverse compounds have been isolated in the last decades including but not limited to polyketides, peptides, hybrid polyketide-peptide products, terpenoids, and nucleosides, many of which have shown promising bioactivities and unusual modes of action (Figure 2).[12-15]

1.3 The biosynthetic logic of natural products

Natural products are biosynthesized by enzymatic machineries. Genes encoding these biosynthetic enzymes are usually clustered in close vicinity in a specific region of the bacterial genome, which is therefore termed a biosynthetic gene cluster (BGC). This property makes the identification, cloning and mobilization for heterologous expression of complete biosynthetic pathways straight forward. A typical BGC comprises core biosynthetic genes and tailoring genes. The former are responsible for backbone formation using basic building blocks such as amino acids and coenzyme A (CoA)-activated carboxylic acids, while the latter may endow the final natural products with structural novelty and diversity through auxiliary modifications, such as oxidoreduction, acylation, glycosylation, and halogenation. In addition, transport-related genes, regulatory genes, and resistance genes are also frequently present within a BGC.

The majority of myxobacterial natural products discovered to date are polyketides and peptides derived, which are mainly biosynthesized by modular polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) gene cluster in an assembly-line fashion, following the colinearity rule. Some peptide natural products can also be ribosomally synthesized and further modified after translation, which are termed as ribosomally synthesized and post-translationally modified peptides (RiPPs). The following sections will introduce the biosynthetic logics of these three main categories of (myxo)bacterial natural products.

1.3.1 Modular polyketide synthase (PKS)

Modular PKS, also known as type I PKS, is a multifunctional and multienzyme complex composed of numerous catalytic domains that are further bundled together to form functional modules.[16] A minimal set of a PKS elongation module consists of a ketosynthase (KS) domain, an acyltransferase (AT) domain, and an acyl carrier protein (ACP) domain (Figure 3). The

ACP domain, also known as a thiolation (T) domain, is responsible for substrate tethering after being activated by post-translational modification at the serine residue by a phosphopantetheinyl transferase (PPTase) (Figure 3). The AT domain selects a specific acyl-CoA (usually malonyl-CoA or methylmalonyl-CoA) and loads the acyl units to the phosphopantetheinyl arm (Ppant arm) of the activated ACP domain. In some cases, each PKS module lacks an integrated AT domain and instead shares a standalone *trans*-acting AT. This type of PKS gene cluster is therefore termed *trans*-AT PKS.[17,18] The KS domain resembles a fatty acid synthase and typically catalyzes intermodular chain translocation and decarboxylative Claisen condensation, resulting in the two-carbon unit elongation of the growing polyketide. In the starter module, a decarboxylating KS domain (KS_Q, $_Q$ refers to the replacement of the active-site cysteine with a glutamine) is often observed.[19] In addition to these core domains, other optional catalytic domains are frequently used in the PKS assembly line to further modify the ACP-bound intermediate, for example, the β-ketoacyl reductase (KR) domain, the β-hydroxyacyl dehydratase (DH) domain, the enoyl reductase (ER) domain, and the *C*-methyltransferase (cMT) domain (Figure 3). The terminal module usually contains a thioesterase (TE) domain (type I TE), which can offload the mature polyketide chain from ACP, typically by hydrolysis or macrocyclization, resulting in linear or cyclic products, respectively.[20] In some cases, chain release can also be catalyzed by a standalone thioesterase (type II TE) or by other enzymes such as reductase, oxygenase, and FabH-like enzyme.[21] The released polyketide can undergo further modifications by post-tailoring enzymes. Noteworthily, unprecedented structures and unusual biosynthetic features are frequently found from myxobacterial natural products.[22]

1.3.2 Non-ribosomal peptide synthetase (NRPS)

Like modular PKS, NRPS is often a giant megasynthetase.[16] A minimal set of an NRPS elongation module comprises an adenylation (A) domain, a condensation (C) domain and a peptidyl carrier protein (PCP) domain (Figure 3). The A domain is responsible for selecting and activating carboxylate building blocks, including but not limited to 20 proteinogenic amino acids, through an adenylation reaction that uses adenosine triphosphate (ATP) as a co-substrate. In some cases, the function of A domains is dependent on accessory MbtH-like proteins.[23] The C domain catalyzes peptide bond formation between the upstream PCP-bound donor and the downstream PCP-bound acceptor. According to the stereoselectivity of the upstream PCP-bound donor, C domains can be categorized into two phylogenetic groups: $^{L}C_{L}$ (linking an L-amino acid to a peptide ending with an L-amino acid) and $^{D}C_{L}$ (linking an L-amino acid to a peptide ending with a D-amino acid).[24] In addition to catalyzing canonical peptide bond formation, C domains can also perform versatile and unusual functions in natural product biosynthesis, such as epimerization, dehydration, β-lactam formation, and chain release.[25] In lipopeptides biosynthesis, the starter condensation (Cs) domain of the first NRPS module can direct lipoinitiation by conjugating an acyl moiety to the PCP-bound aminoacyl thioester.[26] In the biosynthesis of thiazoline- and oxazoline-containing peptides, heterocyclization (HC or Cy) domains are used, which can catalyze condensation, cyclization, and dehydration of a PCP-tethered cysteine/serine/threonine in a stepwise manner. In the presence of a flavin mononucleotide (FMN)-binding oxidase (Ox) domain, the thiazoline and oxazoline moieties can be further converted into thiazole and oxazole, respectively. Other optional catalytic domains, such as the epimerization (E) domain and the *N*-methyltransferase (nMT) domain, are also frequently used in the NRPS assembly line (Figure 3). In some cases, the nascent PCP-bound peptides can undergo on-line modifications by *trans*-acting enzymes. In addition, the released peptide may be further modified by post-tailoring enzymes.

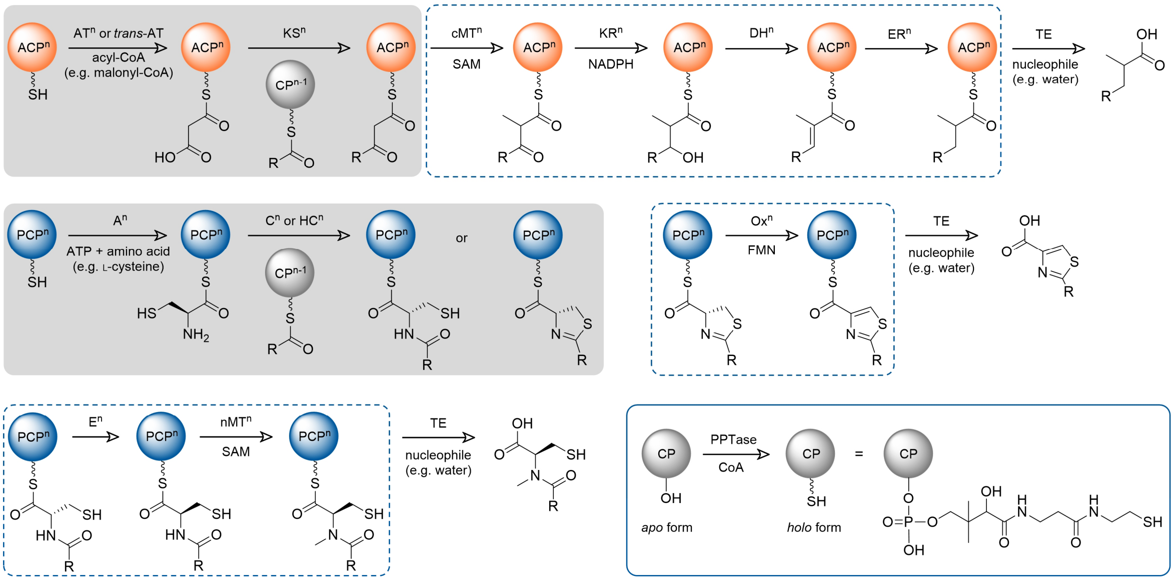


Figure 3 The biosynthetic logic of polyketides by modular PKS and non-ribosomal peptides by NRPS. Shaded areas illustrate the essential elements for a PKS/NRPS module, while optional catalytic steps are depicted within dashed boxes. The activation of the carrier protein is indicated within the solid box. The symbol "R" represents the nascent polyketide or peptide chain, while the superscript "n" denotes the module number.

1.3.3 PKS-NRPS hybrids

As mentioned above, the PKS and NRPS assembly lines share analogous modular architecture and biosynthetic logic, which has enabled the combination of these two systems during evolution. Alongside the use of tailoring enzymes for diversifying polyketides or peptides, hybridization of PKS and NRPS is another strategy employed by nature to generate structural diversity. The majority of myxobacteria-derived natural products contain both peptide and polyketide elements are biosynthesized by hybrid PKS-NRPS assembly lines (Table 1). Two primary considerations must be taken into account for such kind of hybrid systems. On the one hand, the intermodular interaction and biosynthesis transition between the hybrid interface (PCP/KS and ACP/C) are crucial for functionality. Studies using the epothilone biosynthetic machinery as a model system have shown that while the terminal recognition sequence is critical, it is not the only element required for protein-protein interactions at the interface.[27,28] On the other hand, PKS and NRPS typically use dedicated PPTase for carrier protein (CP) activation, as a consequence of which, the PPTase with substrate promiscuity is required for the hybrid system to convert both *apo* ACP and PCP to their corresponding *holo* forms (Figure 3). The MtaA enzyme from the myxothiazol biosynthetic pathway is an example of such a PPTase, as it has been shown to have a substrate promiscuity and was used to activate heterologously expressed PKS and NRPS.[29,30]

Table 1 Myxobacteria-derived type I PKS, NRPS and hybrid PKS-NRPS BGCs, and the corresponding natural product synthesized by the organism.

Name	Producing organism	Type of BGC	Year of publication
Myxothiazol	*Stigmatella aurantiaca* DW4/3-1	PKS-NRPS	1999
Epothilone	*Sorangium cellulosum* SMP44	PKS-NRPS	2000
Myxochelin	*Stigmatella aurantiaca* Sg a15	NRPS	2000
Myxalamid	*Stigmatella aurantiaca* Sg a15	PKS-NRPS	2001
Soraphen	*Sorangium cellulosum* So ce26	type I PKS	2002
Stigmatellin	*Stigmatella aurantiaca* Sg a15	type I PKS	2002
Melithiazol	*Melittangium lichenicola* Me l46	PKS-NRPS	2003
Tubulysin	*Angiococcus disciformis* An d48	PKS-NRPS	2004
Cystothiazole	*Cystobacter fuscus* AJ-13278	PKS-NRPS	2005

Name	**Producing organism**	**Type of BGC**	**Year of publication**
Disorazole	*Sorangium cellulosum* So ce12	*trans*-AT PKS-NRPS	2005
Myxochromide	*Stigmatella aurantiaca* DW4/3-1	PKS-NRPS	2005
Myxovirescin	*Myxococcus xanthus* DK1622	*trans*-AT PKS-NRPS	2006
Chondramide	*Chondromyces crocatus* Cm c5	PKS-NRPS	2006
Chivosazole	*Sorangium cellulosum* So ce56	*trans*-AT PKS-NRPS	2006
Spirangien	*Sorangium cellulosum* So ce90	type I PKS	2006
Ambruticin	*Sorangium cellulosum* So ce10	type I PKS	2006
Jerangolid	*Sorangium cellulosum* So ce307	type I PKS	2006
Aurafuron	*Stigmatella aurantiaca* DW4/3-1	type I PKS	2007
Ajudazol	*Chondromyces crocatus* Cm c5	PKS-NRPS	2008
DKxanthene	*Myxococcus xanthus* DK1622	PKS-NRPS	2008
Etnangien	*Sorangium cellulosum* So ce56	*trans*-AT PKS	2008
Chondrochloren	*Chondromyces crocatus* Cm c5	PKS-NRPS	2008
Corallopyronin	*Corallococcus coralloides* B035	*trans*-AT PKS-NRPS	2010
Thuggacin	*Sorangium cellulosum* So ce895	PKS-NRPS	2010
Sorangicin	*Sorangium cellulosum* So ce12	*trans*-AT PKS	2010
Leupyrrin	*Sorangium cellulosum* So ce690	PKS-NRPS	2011
Aetheramide	*Aetherobacter rufus* SBSr003	PKS-NRPS	2012
Rhizopodin	*Stigmatella aurantiaca* Sg a15	*trans*-AT PKS-NRPS	2012
Myxoprincomide	*Myxococcus xanthus* DK1622	PKS-NRPS	2012
Microsclerodermin	*Jahnella* sp. MSr9139	PKS-NRPS	2013
Microsclerodermin	*Sorangium cellulosum* So ce38	PKS-NRPS	2013
Gephyronic acid	*Cystobacter violaceus* Cb vi76	type I PKS	2013
Myxopyronin	Myxococcus fulvus Mx f50	*trans*-AT PKS-NRPS	2013
Crocacin	*Chondromyces crocatus* Cm c5	PKS-NRPS	2014
Disciformycin	*Pyxidicoccus fallax* AndGT8	type I PKS	2014
Gulmirecin	*Pyxidicoccus fallax* HKI 727	type I PKS	2014
Cystobactamid	*Cystobacter* sp. Cbv34	NRPS	2014
Phenylnannolone	*Nannocystis pusilla* B150	type I PKS	2014
Cystomanamide	*Cystobacter fuscus* MCy9118	PKS-NRPS	2014
Phenylnannolone	*Nannocystis pusilla* B150	type I PKS	2014

Name	Producing organism	Type of BGC	Year of publication
Nannocystin	*Nannocystis* sp. MB1016	PKS-NRPS	2015
Bengamide	*Myxococcus virescens* ST200611	PKS-NRPS	2015
Chlorotonil	*Sorangium cellulosum* So ce1525	*trans*-AT PKS	2015
Macyranone	*Cystobacter fuscus* MCy9118	PKS-NRPS	2015
Haliamide	*Haliangium ochraceum* SMP-2	PKS-NRPS	2016
Haliangicin	*Haliangium ochraceum* SMP-2	type I PKS	2016
Haprolid	*Byssovorax cruenta* Har1	PKS-NRPS	2016
Antalid	*Polyangium spumosum* MSr6761	PKS-NRPS	2016
Ripostatin	*Sorangium cellulosum* So ce377	PKS	2016
Phenalamide	*Myxococcus stipitatus* DSM 14675	PKS-NRPS	2016
Myxochromide	*Myxococcus xanthus* DK1622	PKS-NRPS	2017
Pyrronazol	*Nannocystis pusilla* Ari7	PKS-NRPS	2017
Pyxipyrrolone	*Pyxidicoccus fallax* MCy9557	*trans*-AT PKS-NRPS	2017
Crocadepsin	*Chondromyces crocatus* Cm c5	PKS-NRPS	2018
Vioprolide	*Cystobacter violaceus* Cb vi35	NRPS	2018
Chloromyxamide	*Myxococcus* sp. MCy10608	NRPS	2018
Fulvuthiacene	*Myxococcus fulvus* MCy9280	PKS-NRPS	2019
Argyrin	*Cystobacter* sp. SBCb004	NRPS	2019
Thaxteramide	*Jahnella thaxteri* MSr9139	PKS-NRPS	2019
Archangiumide	*Archangium violaceum* SDU8	*trans*-AT PKS	2021
Sorangipyranone	*Myxobacterial* strain MSr12020	type I PKS	2021
Sandarazol	*Sandaracinus* sp. MSr10575	*trans*-AT PKS-NRPS	2021
Myxadazole	*Myxococcus* sp. SDU36	PKS-NRPS	2021
Myxofacycline	*Corallococcus* sp. MCy9072	PKS-NRPS	2021
Thiamyxin	*Myxococcaceae* strain MCy9487	PKS-NRPS	2022
Corramycin	*Corallococcus coralloides* MCy10984	PKS-NRPS	2022
Archangiumin	*Archangium gephyra* SDU49	PKS-NRPS	2023
Pyxipyrrolone	*Corallococcus exercitus* SDU142	*trans*-AT PKS-NRPS	2023
Myxovalargin	*Corallococcus coralloides* 1071	NRPS	2023
Icumazole	*Sorangium cellulosum* So ce836	PKS-NRPS	2023

1.3.4 Ribosomally synthesized and post-translationally modified peptides (RiPPs)

Ribosomally synthesized and post-translationally modified peptides (RiPPs) (e.g. lanthipeptides, lasso peptides, and thioamitides) are another class of pharmaceutically important natural products with diverse bioactivities.[31-33] These molecules are rarely discovered from myxobacteria. To date, only three different structural classes of myxobacteria-derived RiPPs (crocagin, cittilin, and myxarylin) were reported.[34-36] This limited number of discoveries is not an indication that myxobacteria are poor producers of RiPPs, but rather suggests that their biosynthetic potential remains largely untapped. The subsequent genome mining section will explore this potential in greater detail.

Unlike modular PKS and NRPS systems forming multi-domain complexes, the RiPPs biosynthesis mainly relies on separate tailoring enzymes for post-translational modifications (PTMs) of the ribosomally synthesized precursor peptides (typically 20−110 amino acids in length), consisting of the core peptide and leader/follower peptide (Figure 4).[37] This unique property makes the engineering and *in vitro* reconstitution of RiPPs biosynthetic pathways much easier. Besides, the catalytic promiscuity of tailoring enzymes enables the generation of a compound library by simply changing the core sequence.[38-40] However, despite the concise biosynthetic logic, the accumulating evidence suggests that complex protein-protein interactions are critical for efficient catalysis during post-translational modifications in RiPPs biosynthesis.[41] Although the core peptides are only derived from proteinogenic amino acids, the matured RiPPs natural products can exhibit high structure diversity due to various post-translational modifications, for example, different peptide crosslinks catalyzed either by cytochrome P450 enzymes (CYPs) (e.g. cittilin and myxarylin) or by radical *S*-adenosyl methionine (RaS) enzymes (e.g. streptide and darobactin) and the pyrroloindoline moiety formation guided by unusual peptide-binding proteins (e.g. crocagin).[42-46] In addition to the peptide crosslinks, which are usually found in NRPS biosynthesis, many RiPPs products emulate other NRPS structure features such as the D-amino acid, acetylation, heterocyclization, and glycosylation.[47-49] The huge reserve of structure diversification provides an excellent starting point for the discovery of novel enzymology and bioactivity from this rapidly growing class of natural products.[31]

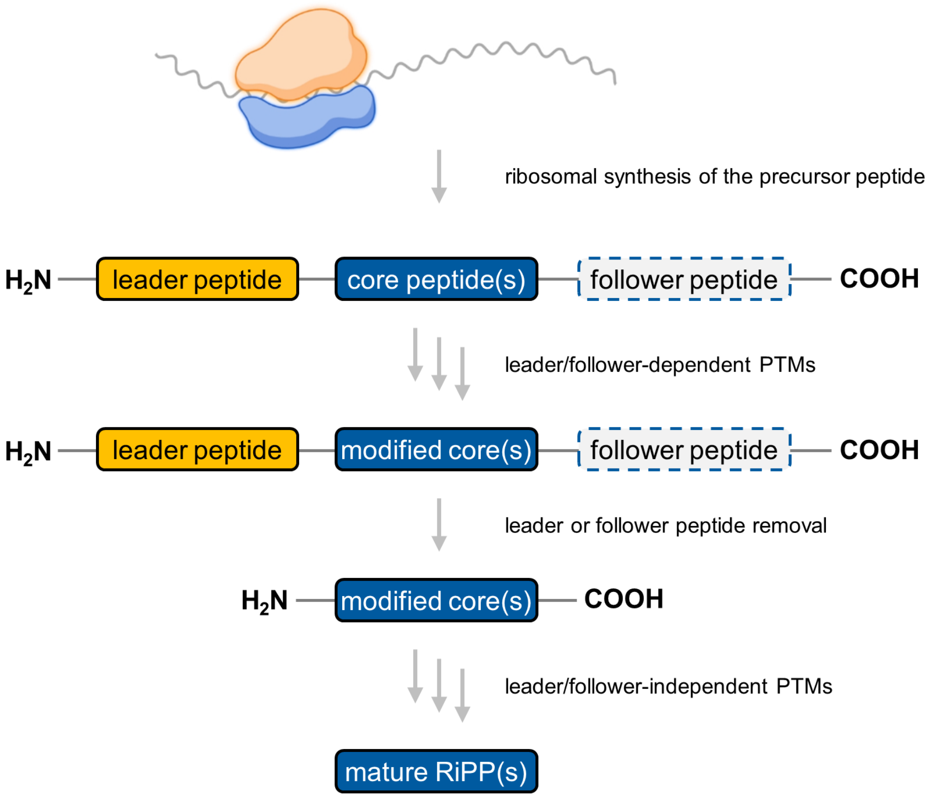


Figure 4 The biosynthetic logic of RiPPs. In some cases, the precursor peptide may comprise multiple core peptides, leading to the generation of a group of mature RiPP products.[31]

1.4 Genome mining for natural products discovery

The continuous discovery of novel compounds is the cornerstone of natural product research. Currently, the primary strategies employed for identifying natural products can be classified into four major categories: bioactivity-guided fractionation, cultivation-based screening, metabolome mining, and genome mining (Figure 5).[50]

Tracking bioactivities is a commonly used and effective method to discover potential antibiotics.[51] However, this bioassay-based approach is heavily reliant on the available test panels and frequently leads to the rediscovery of known compounds.[52] The cultivation-based approach could help to elicit secondary metabolism by cultivating the producer strain in different media (e.g. composition, pH, and additives), under different growth conditions (e.g. temperature, aeration, and illumination), or together with other strains (i.e. co-cultivation).[53] Metabolome mining represents a prioritization approach for natural products isolation. Following dereplication against databases, unknown compounds exhibiting abundant production, conspicuous UV adsorption, or distinct structural features can be prioritized for isolation, for example, by statistical analysis of LC-MS datasets, GNPS molecular networks, and NMR-based metabolic profiling.[54-56] It is undeniable that these conventional "compound-first" approaches have been major pipelines for natural products discovery.[57]

Over the last few years, the number of available genome sequences has grown exponentially. An extraordinary number of bioinformatically predicted BGCs are uncovered and greatly outnumber the experimentally confirmed natural products, indicating that the biosynthetic potential of novel natural products is largely untapped, which has also been observed for myxobacteria.[58] Combined with the accumulated knowledge about natural product biosynthesis, the "gene-to-compound" is therefore rapidly emerging as an alternative strategy to the traditional "compound-first" approaches.[59,60] Many methods have been developed, including *in situ* genetic manipulation (inactivation, activation, or overexpression), heterologous expression of cryptic BGCs, isotopic labeling, and total chemical synthesis of the bioinformatically predicted BGC products (syn-BNP) (Figure 5).[61-64]

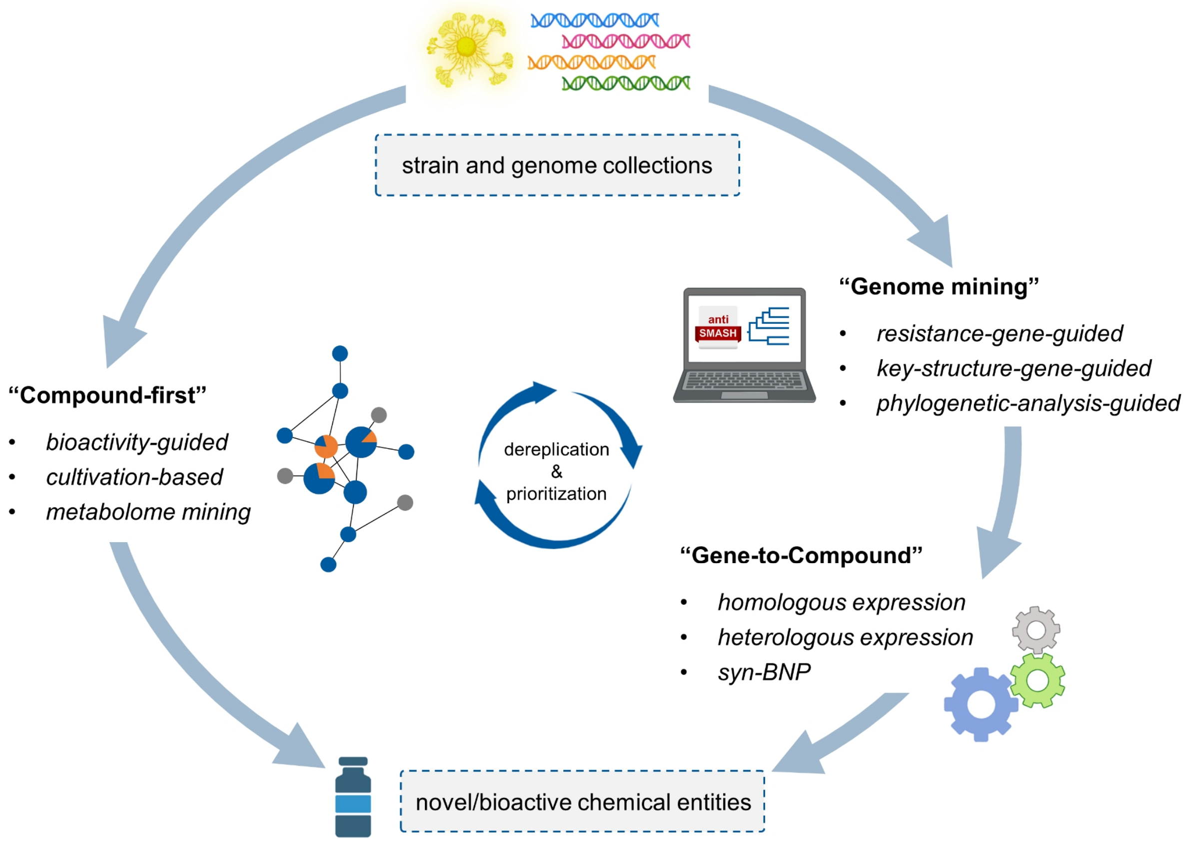


Figure 5 Strategies for natural products discovery with a focus on genome mining.

1.4.1 Gene cluster annotation and prioritization

Identification and annotation of secondary metabolite biosynthetic loci from a large-scale DNA sequence collection is the starting point for genome mining. To facilitate this, numerous bioinformatics platforms have been developed for automatic *in silico* analysis.[65] Among these, antiSMASH (antibiotics & secondary metabolite analysis shell) is currently the most widely used tool capable of comprehensively detecting 71 types of BGCs covering the most classes of known natural products.[66,67] A variety of other tools have also been designed for specialized secondary metabolites, such as RiPPER, RRE-Finder, and RiPPMiner for the mining of RiPPs.[68-70] Noteworthily, manual double-checking, such as using BLSAT (basic local alignment search tool), is often necessary for more precise annotations of the detected biosynthetic pathways, especially for hypothetical proteins.[71] Currently, more than 140,000 antiSMASH-annotated BGCs are deposited in antiSMASH database, which is expanding rapidly, contrasting with the approximately 2,500 experimentally validated BGCs deposited in the MIBiG (minimum information about a biosynthetic gene cluster) database.[72,73] Therefore, the dereplication and careful selection of appropriate gene cluster candidates are prerequisite for the downstream characterization of natural products. The prioritization of gene clusters is often guided by resistance genes, key structure genes, phylogenetic analysis, or a combination of these approaches.[74,75]

Microorganisms employ several BGC-associated strategies for self-resistance, such as safeguarding the target and inactivating/transporting the product.[76] Inspired by the former strategy, “target-directed genome mining” is developed,[77] which serves as a valuable approach to prioritize cryptic gene clusters to discover novel compounds that are likely to interact with a specific target.[78] Moreover, this approach can be used to uncover the biosynthetic gene clusters and the targets of known compounds.[76,79] The myxobacterial cystobactamid biosynthetic pathway contains a pentapeptide repeat protein (PRP) for self-defense by protecting the host gyrase.[80] By probing the PRP, new topoisomerase inhibitors, pyxidicyclines and uncommon alkylpyrones, were discovered from the myxobacteria *Pyxidicoccus fallax* An d48 and *Myxococcus xanthus* DK1622, respectively.[78,81] Different from the PRP, the target variant encoded by a duplicated house-keeping gene, which is colocalized with a BGC, is so called “biosynthesis-associated resistance determinant”.[79] In the mode of action and biosynthesis studies of griselimycins, which target DnaN (the DNA polymerase sliding clamp) for anti-tuberculosis bioactivity, it was discovered that the biosynthetic pathway of griselimycins includes an additional copy of DnaN that confers self-resistance.[82] This kind of self-resistance mechanism also exist in the myxobacterial bengamide biosynthetic pathway, which contains an additional copy of methionine aminopeptidase.[83] Xiaoyu et al. developed the rational target-directed genome mining strategy and discovered a group of thiotetronic acid antibiotics that target the fatty acid synthase.[77] This strategy has led to the identification of many previously uncharacterized biosynthetic pathways related to known bioactive natural products. For example, the chuangxinmycin gene cluster was discovered by genome scanning of DNA regions containing tryptophanyl-tRNA synthetase genes,[84] the restricticin/lanomycin biosynthetic pathway was discovered by genome mining of uncharacterized BGCs containing a lanosterol 14α-demethylase (CYP51) gene,[85] and the proteasome inhibitor fellutamide B biosynthetic pathway was discovered by characterizing a BGC containing a gene encoding a putative proteasome subunit.[86] Ellis el at. summarized the biosynthesis-associated resistance determinants that could be used as queries for genome mining.[79] It is important to consider that mining these determinants may lead to the identification of false-positive gene clusters. The target-directed genome mining strategy can also be employed to determine the mode of action of known compounds. For example, Yan et al. rediscovered aspterric acid as a new herbicide with a novel mode of action, targeting DHAD (dihydroxy acid dehydratase).[87] Furthermore, by employing this genome mining strategy, Elizabeth et al. discovered clipibicyclene as the genuine metabolite of a pyrrolizidine alkaloid-related BGC, targeting ClpP (caseinolytic protease).[88] Researchers can use the ARTS (antibiotic resistant target seeker) tool for comparative genome mining by targeting self-resistance genes.[89,90] In a review by Yan et al., the recent development in self-resistance-gene-directed natural products discovery was summarized.[91]

The discovery of enediyne natural products by Shen Lab is a prime example of structure-gene-guided genome mining.[92] Enediyne natural products feature a conserved core structure, consisting of two acetylenic groups conjugated to a double bond within the nine- or ten-membered carbocycle that is biosynthesized by a conserved warhead cassette consisting of five genes.[93] Consequently, this conserved warhead cassette can be used as probes for genome mining of enediyne biosynthetic gene clusters.[94-97] In addition to enediyne, Bode's group recently characterized the benzoxazolinate biosynthesis and performed the genome mining using the core-structure biosynthetic gene to explore benzoxazolinate-containing natural products.[98] Katherine et al. summarized case studies targeting chemical features for genome mining of specialized natural products.[99]

The construction of sequence similarity network through phylogenetic analysis represents another approach for gene cluster prioritization from a large genome data collection, also known as evolutionary genome mining.[100] This approach can be used to classify the detected BGCs into gene cluster families (GCFs), therefore providing a bird's-eye view of the evolutionary relationships of these BGCs and the chemical diversity of their corresponding products. For example, Peter et al. performed a global phylogenomic analysis of prokaryotic BGCs and characterized the widely distributed gene clusters of unknown function,[101] Jorge et al. developed the BiG-SCAPE/CORASON platform to explore the large-scale biosynthetic diversity and showcased the identification of new detoxins,[102] Elizabeth et al. prioritized phylogenetically divergent gene clusters and discovered the glycopeptide antibiotics complestatin and corbomycin with new biological mode of action,[103] just to mention a few. Studies have shown that the chances to discover chemical diversity of natural products correlates with the taxonomic distance between the producing strains.[104] Consequently, distinct families of natural products are more likely to be discovered from new or rare genera.

1.4.2 Functional characterization of cryptic gene clusters

Currently, the majority of detected biosynthetic gene clusters (BGCs) remain cryptic, with their corresponding products unknown. These cryptic gene clusters can be categorized into three classes: 1) orphan BGCs, whose products have not yet been characterized; 2) silent BGCs, which are not (sufficiently) expressed under certain conditions; and 3) unfunctional BGCs. Therefore, the two major tasks for functional characterization of cryptic BGCs are linking orphan gene clusters to secondary metabolites and activating silent biosynthetic gene clusters.[105] Genomics-driven strategies, such as *in situ* inactivation, genomisotopic approaches, and heterologous expression, are typically used for the former task. For the latter, promoter exchange, regulator manipulation, reporter-guided screening, and heterologous expression are commonly employed.[106,107]

The exchange of native promoters with strong constitutive or inducible promoters using homologous recombination is a widely used method for *in situ* eliciting gene cluster expression by circumventing the tight transcriptional regulation of the host. For instance, sandarazols, the myxobacteria-derived toxins, were discovered by activating a plasmid-encoded gene cluster using a vanillate inducible promoter.[108] Noteworthily, genome engineering using strain-specific recombinases is more efficient than the commonly used single-crossover method for bioprospecting, as only about 80 bp homology arms are sufficient for homologous recombination. This kind of recombineering systems have been established for *Photorhabdus*, *Xenorhabdus*, *Burkholderiales*, and *Pseudomonas* strains, and have been successfully used for activating cryptic gene clusters by promoter insertion.[109-114]

Manipulating regulators by inducing the expression of activator genes and deletion/inactivation of repressor genes is another promising method for activating silent BGCs. For instance, the silent totopotensamides biosynthetic gene cluster was activated by knocking out transcriptional repressors and overexpressing activators genes.[115] In addition to directly manipulating regulator genes, Huimin Zhao's group developed a transcription factor decoy strategy to activate silent biosynthetic gene clusters by overexpressing the regulator-binding DNA to sequester repressors, thereby obviating the need to identify the regulator genes.[116] It is worth noting that some global regulators, such as the ScmR of *Burkholderia thailandensis* E264 and the ROK-like protein of *Myxococcus xanthus* DK1622, have pleiotropic effects on secondary metabolism, and thus are promising targets for high-throughput gene cluster activations.[117,118]

Using the reporter gene translationally fused with the silent gene cluster of interest and the luminescence as readout, Mohammad developed a high-throughput screening method to identify activators and natural products of silent gene clusters.[119] However, if genetic manipulation of the strain is not feasible or the cryptic gene clusters are from environmental DNA (eDNA), heterologous expression would be an alternative strategy.[120] Several studies have reported successful activation of cryptic gene clusters through heterologous expression. For example, the new antibiotic taromycin was discovered by heterologous expression of a silent lipopeptide BGC,[121] a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster was activated after reconstruction and heterologous expression,[122] and the potent indolotryptoline antiproliferative agents lazarimides were discovered after promoter refactoring and heterologous expression of a “dead” eDNA-derived gene cluster,[123] just to mention a few.

1.4.3 Chemical synthesis of the bioinformatically predicted products

Although promising gene clusters can be prioritized *in silico* without difficulties, *in vivo* activation often fails to yield the corresponding natural products, most probably due to the complex biosynthesis and regulation network of the organism. Additionally, secondary metabolites are generally produced in low quantities, hindering their further application. Chemical synthesis of the bioinformatically predicted products is emerging as an alternative approach, allowing for the efficient translation of the prioritized gene clusters into putative antibiotics *in vitro*, bypassing the need for strain cultivation and manipulation.

Chemical synthesis of pretubulysin, a postulated biosynthetic intermediate that cannot be obtained in the yield achievable through fermentation, facilitated its structural confirmation and biological characterization.[124] Sean Brady's group discovered many synthetic-bioinformatic compounds with promising bioactivities and diverse modes of action, further highlighting the power of this approach in natural products discovery.[125] Examples include the improved albicidin/cystobactamid analogues, the humimycins against MRSA (methicillin-resistant *S. aureus*), the lapcin inhibiting topoisomerase I/II, as well as macolacin, cilagicin and menaquinone-binding antibiotics against multidrug-resistant pathogens.[126-131]

This approach is particularly advantageous for the discovery of antimicrobial peptides from metagenome data. For example, bioactive syn-BNP cyclic peptides inspired by NRPS BGCs from the human microbiome were discovered.[132] In addition, combined with deep machine learning, antimicrobial peptides were identified from the human gut microbiome and chemically synthesized, most of which (>83%) exhibited antibacterial activity at a concentration of 60 μM in liquid media.[133]

This chemical synthesis approach greatly accelerates the genome mining process by translating the (meta-)genomic DNA data directly into chemical entities, bypassing the *in vivo* gene expression process. However, it is currently mainly applied to peptide metabolites. On the one hand, *in silico* analysis of biosynthetic pathways enables relatively accurate prediction of peptide structures, which is more challenging for polyketides. On the other hand, solid-phase synthesis allows for the efficient production of large quantities of peptides, enabling the structure-activity relationship studies, as well as modes of action and pharmacokinetics studies of these syn-BNP compounds.

1.4.4 Combining genome mining with bioactivity screening and metabolome mining

Advancements in genomics, bioinformatics and instrumental analytics have propelled natural product research into the era of big data. The integration of gene cluster information, analytical datasets, and bioactivity profiles has facilitated the annotation, dereplication, prioritization, and discovery of novel and/or bioactive compounds, alleviating blind screening.[134-136] For instance, Roland et al. developed a MS-guided genome mining approach for high-throughput characterization of ribosomal and non-ribosomal peptide natural products and their respective biosynthetic genes.[137] Integration of NMR-based metabolic profiling and genome mining enabled the discovery of archangiumide, which features a rare allenic group, from the myxobacterium *Archangium violaceum* SDU8.[138] The discovery of myxoprincomide with a very low abundance by combining targeted mutagenesis and comparative metabolome analysis highlights the importance of complementing metabolome mining to genome mining in the discovery of natural products with low titer.[139] Min Xu et al. established a high-throughput functional genome mining approach, which combines genome mining and bioactivity-guided screening to discover bioactive natural products and their corresponding biosynthetic gene clusters.[140] This approach led to the discovery and characterization of a new class of lantibiotic (class V) termed lexapeptide.[141] Synergizing available approaches and making full use of databases are expected to accelerate the process of bridging the gap between isolated compounds and genome-encoded potential (Figure 5).[142,143]

1.5 Gene cluster engineering and heterologous expression

Myxobacteria, particularly the *Sorangiineae* strains have proven to be challenging for genetic manipulation, making natural products discovery through genome mining and investigation of biosynthetic pathways in native producer strains difficult. Additionally, many promising compounds are produced by slow-growing myxobacteria, often with unstable and poor yields, which hinders the fermentative production process. Therefore, gene cluster cloning and engineering (e.g. using the Red/ET recombineering, Figure 6) and subsequent heterologous expression can facilitate genome-guided discovery, biosynthesis studies, and highly efficient production of desired components or generation of non-natural products.[144]

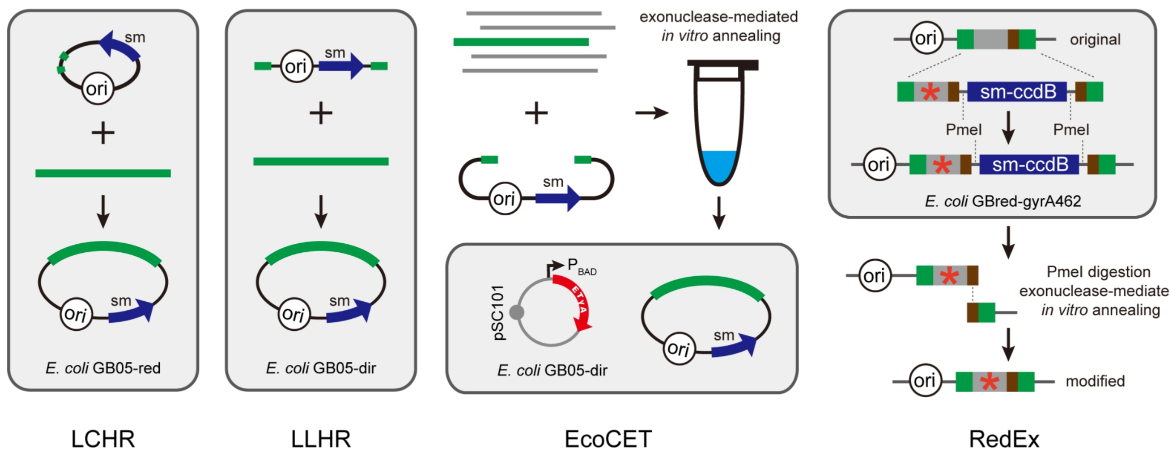


Figure 6 Gene cluster cloning and engineering with the versatile Red/ET recombineering technology. LCHR: Redαβ-mediated linear-circular homologous recombination;[145] LLHR: RecET-mediated linear-linear homologous recombination;[145] ExoCET: Exonuclease *in vitro* assembly combined with RecET-mediated LLHR;[146] RedEx: Redαβ-mediated LCHR combined with *ccdB* counterselection and exonuclease-mediated *in vitro* annealing;[147] sm: selection marker; ori: the origin of replication of a plasmid.

1.5.1 Gene cluster cloning

The acquisition of a target gene cluster is a prerequisite for subsequent engineering and heterologous expression. *De novo* chemical synthesis of artificial gene clusters is becoming increasingly cost-effective and particularly advantageous for customized biosynthetic pathways (e.g. codon optimization, operon re-organization, and multiple biosynthetic engineering for combinatorial biosynthesis).[148-150] However, subsequent assembly of the synthesized DNA fragments is usually needed. Gene cluster cloning techniques (e.g. PCR assembly, cosmid/BAC library construction, and direct cloning approaches) are more commonly used, especially when the genomic DNA is readily available and the products of biosynthetic pathways are unknown. In addition to the conventional restriction/ligation-dependent cloning method, many other methods for assembling DNA fragments into a vector have been developed, such as T4 DNA polymerase or T5 exonuclease-dependent assembly, Gibson Assembly, Golden Gate Cloning, and *in vivo* homologous recombination using *E. coli* or *S. cerevisiae*.[151-156]

Cosmid/BAC library construction enables the high-throughput cloning of large genomic DNA fragments (>30 kb) into plasmids, but the generation of the library and the screening of target constructs can be time-consuming and laborious. Direct cloning represents another commonly used approach for efficiently capturing the target gene cluster into an expression vector, for example, TAR cloning, ExoCET (Figure 6), CATCH, CAPTURE, and CAT-FISHING.[157-160] These PCR-independent methods have the advantage of high fidelity, with rare occurrences of mutations during cloning, thereby avoiding the need for resequencing of the generated expression constructs.

1.5.2 Biosynthetic engineering

Once the biosynthetic gene cluster is cloned into an expression vector, molecular engineering of the biosynthetic pathway in the most commonly used *E. coli* host becomes much more feasible compared with genome engineering of the native producer strain. Methodologies for engineering of large biosynthetic pathways are mainly based on the versatile Red/ET recombineering technology (Figure 6) and the CRISPR-Cas system.[147,161-164] Applications of biosynthetic engineering mainly include promoter substitution, gene/domain/module deletion, point mutation, and combinatorial biosynthesis.

Promoter engineering used for activating of silent gene clusters has been discussed above. Further, multiple promoter substitutions could greatly improve the expression level, especially using the heterologous host-derived strong constitutive promoters. Using this approach, an increased heterologous production for spinosad was achieved.[155,165] Similar logic was also used for the enhanced heterologous production of salinomycin and disorazole.[166,167] Gene knockout/inactivation is commonly used for functional characterizations and can also be used for the production of desired intermediates and precursor-directed mutabiosynthesis.

PKS/NRPS engineering for the generation of rationally designed non-natural products is becoming of great interest,[168-170] for example, the alteration of A domain substrate specificity by point mutation or subdomain swapping,[171-173] the alteration of PKS substrate specificity by AT engineering,[174-176] and the alteration of polyketide stereochemistry by module exchange or KR domain mutations,[177,178] just to mention a few. However, successful cases mainly owed to the broad substrate promiscuity, intrinsically tolerating various non-natural intermediates, e.g., editing the rapamycin biosynthetic gene cluster resulted in the production of almost all the desired derivatives by AT-exchange, module deletion/insertion, and swapping reductive loop domains.[179] Nevertheless, the non-natural products are usually produced in low amounts.

Therefore, it is still challenging to engineer PKS/NRPS BGCs without broadly applicable rules and guidelines,[180] especially for biosynthetic pathways with poor substrate tolerance. Regarding this, at least three points should be considered: 1) the domain/module fusion-points that are crucial for protein-protein interaction; 2) the tolerance of gatekeeper domains to substrate variation as the non-natural substrate may result in a premature release or stall the biosynthetic machinery; 3) the efficient release of the desired products ensuring an efficient turnover of the biosynthetic pathway.

From the biosynthesis perspective, a catalytic module begins with a KS/C domain and ends with a CP domain. This classical domains-unit corresponds well with one building block incorporation. Whereas, characterization of the PKS gene cluster evolution revealed that AT_n-(DH_n-KR_n)-ACP_n migrates together with KS_{n+1} instead of the KS_n.[181,182] Characterization of the NRPS interdomain linker regions also revealed that the conserved C-A interface is structurally flexible and an ideal fusion point for domain exchange.[180] As a consequence, an alternative module definition rather suggests AT_n-ACP_n-KS_{n+1} for PKS and A_n-PCP_n-C_{n+1} for NRPS (Figure 7). These re-defined boundaries are quite useful for PKS and NRPS engineering, combinational biosynthesis, and even *de novo* design of artificial biosynthetic pathways, for example, the generation of truncated polyketides with variable length by contracting PKS assembly lines, morphing aureothin-type modular polyketide synthases by modules excision/insertion, and the production of FK228 derivatives with superb yields by A-PCP-C domain unit replacements.[183-185] Bode's group further developed an alternative XUC concept for NRPS engineering (Figure 7), in which the exchange units C_{Asub}-A-PCP-C_{Dsub} are fused in the linker region between two sub-domains in the C domain, overcoming the problem of the specificity of downstream C domains.[186]

In addition to respecting the fusion-sites and the downstream domain specificity, docking domains also need to be considered and are becoming another area of interest for PKS/NRPS engineering. Insights into the structure of docking domains could help to understand the functional interaction between catalytic units of megasynthetase.[187,188] Besides, reprogramming the docking domains or using synthetic zippers may help to generate peptide diversification.[189,190] The TE selectivity could determine the substrate's fate, whereas the type II TE could free CP domains from stalled substrates and boost the activity of engineered megasynthases.[191]

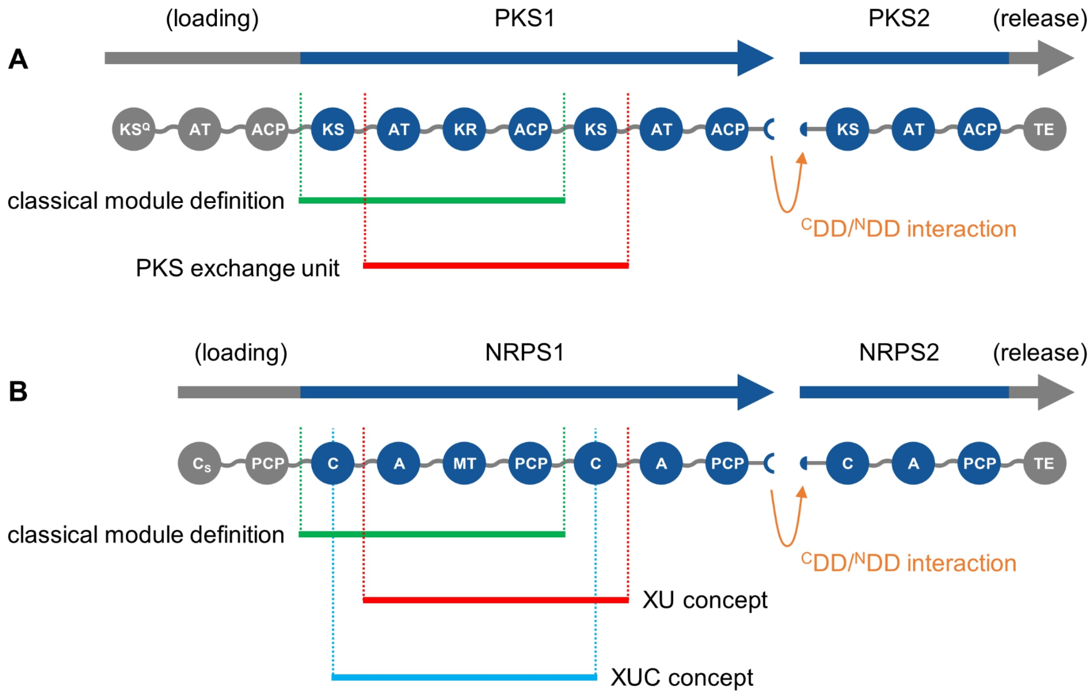


Figure 7 The concept of exchange units for PKS and NRPS engineering. XU: the eXchange Unit;[180] XUC: the eXchange Unit Condensation domain;[186] ^{N}DD: N-terminal docking domain; ^{C}DD: C-terminal docking domain.

Hybriding biosynthetic gene clusters with close evolutionary origins are more likely to generate the functional chimera retaining a near-wild-type activity. Besides, it is believed that the diversity of PKS/NRPS is generated through recombination events during the evolutionary process.[192-194] Therefore, evolution-guided engineering of PKS/NRPS megasynthetases using synthetic biology-inspired strategies would be of great interest for combinatorial biosynthesis.[195-197]

1.5.3 Heterologous expression

Heterologous expression is of great interest in functional characterization of cryptic biosynthetic pathways (Figure 8), optimization of known natural products, and generation of non-natural derivatives.[198-201] Specifically, heterologous expression has the following advantages: 1) clearly links the characterized gene cluster with compounds; 2) heterologous engineering of the cloned gene cluster is easier than genetic manipulation of the original producer strains, some of which cannot be genetically manipulated; 3) improved heterologous production of desired components can be achieved by chassis optimization and metabolic engineering; 4) efficient generation of non-natural products can be achieved by combinatorial biosynthesis and synthetic biology; 5) metagenome mining using the eDNA can be feasible bypassing the cultivation of the original producer strain as many of which are not yet culturable under laboratory conditions.

Figure 8 shows the general workflow of heterologous expression. After cloning and engineering of the prioritized gene cluster, the expression construct can be transferred to amenable heterologous hosts usually by electroporation or conjugation. Then, the biosynthetic gene cluster could be integrated into the chromosome of the host by homologous recombination, transposition, or site-specific recombination. Harnessing natural homologous recombination is a traditionally used method for integration of exogenous genes into the chromosome; however, low efficiency is usually observed.[202] Transposition represents an efficient method to transfer the large gene cluster randomly into the chromosome of various heterologous hosts such as *M. xanthus* DK1622.[202] Whereas, different integration positions may affect the gene expression level, resulting in varied production yields among different mutant clones.[203,204] Site-specific integration, which is mediated by a phage integrase, has been commonly used for *Streptomyces* with high efficiency and specificity, and has also been established for *M. xanthus*.[205,206] In some cases, heterologous expression can be based on multi-copy and autonomously replicating plasmids;[144,162] however, this approach is not yet available for myxobacteria.

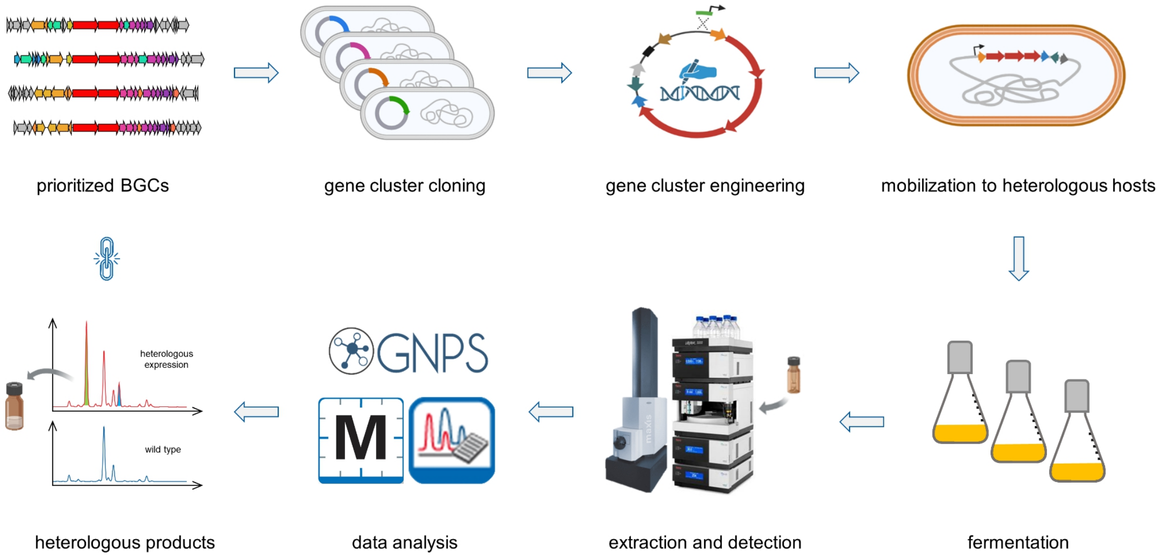


Figure 8 The general workflow of heterologous expression approach.

Most of heterologous expression experiments were done in *Streptomyces* hosts.[207,208] Far fewer (no more than twenty until now, Table 2) myxobacteria-derived biosynthetic pathways have been examined through heterologous expression compared to that of the streptomycetes. In addition, most of these experiments used *Myxococcus xanthus* as the heterologous host, with the aim to either improve the yields or investigate the corresponding biosynthetic pathways. Whereas, functional characterization of myxobacteria-derived cryptic BGCs through heterologous expression has not been reported prior to the sorangibactin study.[209]

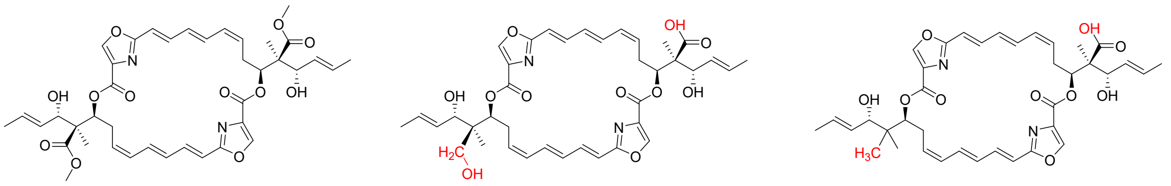
Table 2 Summary of heterologous expression of myxobacteria-derived natural products.

Compounds	Heterologous hosts	Yields	References
Epothilones	*Streptomyces coelicolor*	50−100 μg/L	Li Tang et al., 2000 [210]
Epothilones	*Myxococcus xanthus*	100−400 μg/L	Bryan Julien et al., 2002 [211]
Epothilones	*Myxococcus xanthus*	23 mg/L [a]	Janice Lau et al., 2002 [212]
Epothilone C and D	*Escherichia coli*	< 1 μg/L [b]	Sarah C. Mutka et al., 2006 [213]
Epothilone B and D	*Streptomyces venezuelae*	< 1 μg/L	Sung Ryeol Park et al., 2008 [214]
Epothilones	*Myxococcus xanthus*	about 100 μg/L [b]	Corina Oßwald et al., 2014 [148]
Epothilones	*Schlegelella brevitalea*	307 μg/L	Xiaoying Bian et al., 2017 [215]
Epothilones	*Schlegelella brevitalea*	82 mg/L [c]	Yucong Yu et al., 2020 [216]
Soraphen A	*Streptomyces lividans*	< 0.3 mg/L	Ross Zirkle et al., 2004 [217]
Myxothiazol	*Pseudomonads putida*	0.6 mg/L	Frank Gross et al., 2006 [218]
Myxothiazol	*Myxococcus xanthus*	about 20 mg/L	Olena Perlova et al., 2006 [219]
Myxochromide S	*Pseudomonads putida*	40 mg/L	Silke C. Wenzel et al., 2005 [220]
Myxochromide S	*Myxococcus xanthus*	about 500 mg/L	Jun Fu et al., 2008 [140]
Myxochromide S	*Corallococcus macrosporus*	600 mg/L	Olena Perlova et al., 2009 [221]
Pretubulysin A	*Myxococcus xanthus*	0.19 mg/L	Yi Chai et al., 2012 [222]
Bengamides	*Myxococcus xanthus*	5−10 mg/L	Silke C. Wenzel et al., 2015 [83]
Haliangicin	*Myxococcus xanthus*	about 11 mg/L	Yuwei Sun et al., 2016 [223]
Disorazole A2	Myxococcus xanthus	about 1 mg/L	Qiang Tu et al., 2016 [224]
Disorazole F2	*Burkholderia thailandensis*	38.3 mg/L [c]	Zong-Jie Wang et al., 2022 [167]
Myxopyronin A	*Myxococcus xanthus*	about 156 mg/L	Hilda Sucipto et al., 2017 [225]
Corallopyronin A	*Myxococcus xanthus*	about 37 mg/L	Hilda Sucipto et al., 2017 [225]
Corallopyronin A	*Myxococcus xanthus*	about 100 mg/L [c]	Domen Pogorevc et al., 2019 [226]
Vioprolide B and D	*Myxococcus xanthus*	about 80 mg/L	Fu Yan et al., 2018 [227]
Argyrin A and B	*Myxococcus xanthus*	160 mg/L	Domen Pogorevc et al., 2019 [228]
Cystobactamid 919-1	*Myxococcus xanthus*	8.1 mg/L	Sebastian Groß et al., 2021[149]

[a] fed-batch fermentation; [b] codon optimization; [c] promoter engineering.

1.6 Outline of this dissertation

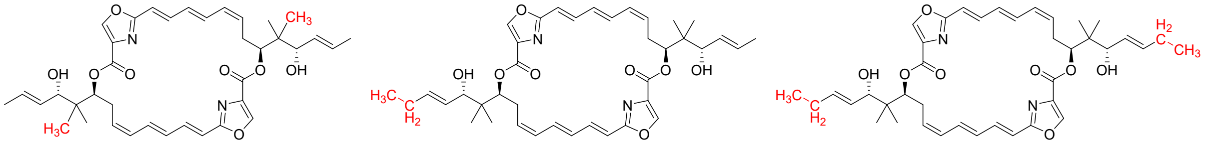
The aim of the work described in this thesis was to discover myxobacterial natural products through heterologous expression and investigate their biosynthesis using the established heterologous expression system, covering *in silico* bioinformatics, *in vivo* characterization, and *in vitro* reconstitution. Two families of myxobacterial natural products (the disorazole cytotoxic polyketides and the sorangibactin iron-chelating peptides; Figure 9) and their biosynthetic pathways were intensively characterized.



Disorazole Z1
Chemical Formula: $C_{40}H_{46}N_2O_{12}$
Exact Mass: 746.3051

Disorazole Z9
Chemical Formula: $C_{38}H_{44}N_2O_{11}$
Exact Mass: 704.2945

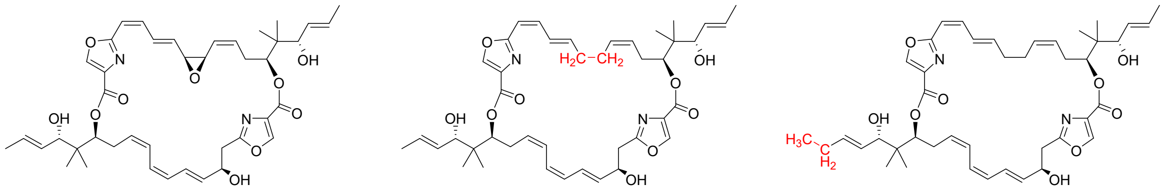
Disorazole Z10
Chemical Formula: $C_{38}H_{44}N_2O_{10}$
Exact Mass: 688.2996



Disorazole Z11
Chemical Formula: $C_{38}H_{46}N_2O_8$
Exact Mass: 658.3254

Disorazole Z12
Chemical Formula: $C_{39}H_{48}N_2O_8$
Exact Mass: 672.3411

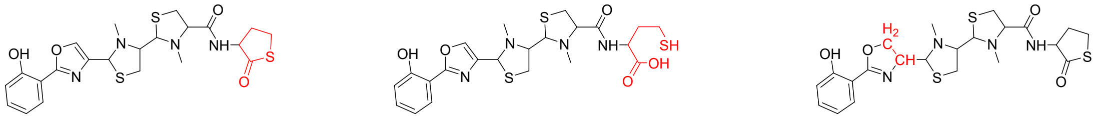
Disorazole Z13
Chemical Formula: $C_{40}H_{50}N_2O_8$
Exact Mass: 686.3567



Disorazole A2
Chemical Formula: $C_{42}H_{52}N_2O_{10}$
Exact Mass: 744.3622

Disorazole F2
Chemical Formula: $C_{42}H_{54}N_2O_9$
Exact Mass: 730.3829

Disorazole F7
Chemical Formula: $C_{43}H_{56}N_2O_9$
Exact Mass: 744.3986



Sorangibactin A
Chemical Formula: $C_{22}H_{26}N_4O_4S_3$
Exact Mass: 506.1116

Sorangibactin B
Chemical Formula: $C_{22}H_{28}N_4O_5S_3$
Exact Mass: 524.1222

Sorangibactin C
Chemical Formula: $C_{22}H_{28}N_4O_4S_3$
Exact Mass: 508.1273

Figure 9 Chemical structures of disorazole and sorangibactin compounds produced through heterologous expression. The major structural features of newly characterized products are highlighted in red.

The disorazole family of natural products are specifically produced by the myxobacteria *Sorangium cellulosum* strains, which are difficult to genetically manipulate. The discovery and heterologous expression of the disorazole Z biosynthetic gene cluster enabled biosynthesis studies, pathway engineering, and production of intermediates/derivatives. These findings can be a stepping-stone for the pharmaceutical development of this promising family of anti-cancer natural products for (pre-)clinical studies. These results are presented in **Chapter 2.**

During our attempt to hybridize the disorazole A and disorazole Z biosynthetic pathways for combinational biosynthesis of non-natural disorazole products using the heterologous host *M. xanthus* DK1622, we happened to discover the unusual function of the hypothetical protein DisE containing a Coq4 domain involved in the enigmatic oxidations (the epoxidation in disorazole A biosynthesis and the site-specific oxidation of the *gem*-dimethyl groups in disorazole Z biosynthesis). Gene cluster engineering, characterization of intermediates, and functional investigation of *disE* are discussed in **Chapter 3.**

The sorangibactins were discovered by heterologous expression of a coelibactin-like NRPS gene cluster, which showed a widespread cross-order occurrence of cryptic homologous BGCs among *Sorangiineae*, *Streptomyces*, and *Pseudomonas* strains. The sorangibactin A features an unprecedented γ-thiolactone moiety and represents a novel family of siderophores. Heterologous expression of the engineered gene clusters also facilitated discoveries of an unprecedented function of a cytochrome P450 dependent-enzyme involved in oxazoline dehydrogenation and an unusual thioesterase (TE) domain that contains a rare active-site cysteine involved in C-terminal thiolactone formation. The novel (bio)chemistry can serve as a starting point for biosynthetic investigations. These results are presented in **Chapter 4.**

In addition, a dozen cryptic gene clusters from myxobacteria were prioritized followed by cloning, engineering, and heterologous expression using the host *M. xanthus* DK1622; however, with limited success regarding compound production. Opportunities/challenges of the genome miming approach for natural products discovery and thoughts about successful heterologous expression (influencing factors and potential solutions) are discussed in the section of **5.1**. Furthermore, insights of biosynthesis and rational engineering of PKS/NRPS are discussed in other sections of **Chapter 5.**

1.7 References

[1] J. J. Hug, D. Krug, R. Müller, Bacteria as genetically programmable producers of bioactive natural products, *Nat Rev Chem*, **2020**, 4(4), 172-193.

[2] E. K. Davison and M. A. Brimble, Natural product derived privileged scaffolds in drug discovery, *Curr Opin Chem Biol*, **2019**, 52, 1-8.

[3] F. Hemmerling and J. Piel, Strategies to access biosynthetic novelty in bacterial genomes for drug discovery, *Nat Rev Drug Discov*, **2022**, 21(5), 359-378.

[4] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch et al., Natural products in drug discovery: advances and opportunities, *Nat Rev Drug Discov*, **2021**, 20(3), 200-216.

[5] O. Loiseleur, Natural Products in the Discovery of Agrochemicals, *Chimia (Aarau)*, **2017**, 71(12), 810-822.

[6] D. J. Newman and G. M. Cragg, Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019, *J Nat Prod*, **2020**, 83(3), 770-803.

[7] D. M. P. de Oliveira, B. M. Forde, T. J. Kidd et al., Antimicrobial Resistance in ESKAPE Pathogens, *Clin Microbiol Rev*, **2020**, 33(3), e00181-19.

[8] M. Miethke, M. Pieroni, T. Weber et al., Towards the sustainable discovery and development of new antibiotics, *Nat Rev Chem*, **2021**, 5(10), 1-24.

[9] P. Cao, A. Dey, C. N. Vassallo et al., How myxobacteria cooperate, *J Mol Biol*, **2015**, 427(23), 3709-3721.

[10] K. Gerth, S. Pradella, O. Perlova et al., Myxobacteria: proficient producers of novel natural products with various biological activities—past and future biotechnological aspects with the focus on the genus *Sorangium*, *J Biotechnol*, **2003**, 106(2-3), 233-253.

[11] T. F. Schäberle, F. Lohr, A. Schmitz et al., Antibiotics from myxobacteria, *Nat Prod Rep*, **2014**, 31(7), 953-972.

[12] S. C. Wenzel and R. Müller, Myxobacteria—'microbial factories' for the production of bioactive secondary metabolites, *Mol BioSyst*, **2009**, 5(6), 567-574.

[13] K. J. Weissman and R. Müller, Myxobacterial secondary metabolites: bioactivities and modes-of-action, *Nat Prod Rep*, **2010**, 27(9), 1276-1295.

[14] J. Herrmann, A. A. Fayad, R. Müller, Natural products from myxobacteria: novel metabolites and bioactivities, *Nat Prod Rep*, **2017**, 34(2), 135-160.

[15] C. D. Bader, F. Panter, R. Garcia et al., Sandacrabins–Structurally unique antiviral RNA polymerase inhibitors from a rare myxobacterium, *Chem Eur J*, **2022**, 28(10), e202104484.

[16] M. A. Fischbach and C. T. Walsh, Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms, *Chem Rev*, **2006**, 106(8), 3468-3496.

[17] J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2010**, 27(7), 996-1047.

[18] E. J. N. Helfrich and J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2016**, 33(2), 231-316.

[19] T. Chisuga, A. Nagai, A. Miyanaga et al., Structural Insight into the Reaction Mechanism of Ketosynthase-Like Decarboxylase in a Loading Module of Modular Polyketide Synthases, *ACS Chem Biol*, **2022**, 17(1), 198-206.

[20] M. E. Horsman, T. P. A. Hari, C. N. Boddy, Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate?, *Nat Prod Rep*, **2016**, 33(2), 183-202.

[21] R. F. Little and C. Hertweck, Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis, *Nat Prod Rep*, **2022**, 39(1), 163-205.

[22] S. C. Wenzel and R. Müller, Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry, *Nat Prod Rep*, **2007**, 24(6), 1211-1224.

[23] K. J. Esquilín-Lebrón, T. O. Boynton, L. J. Shimkets et al., An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622, *J Bacteriol*, **2018**, 200(21), e00346-18.

[24] N. Roongsawang, S. P. Lim, K. Washio et al., Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases, *FEMS Microbiol Lett*, **2005**, 252(1), 143-151.

[25] S. Dekimpe and J. Masschelein, Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis, *Nat Prod Rep*, **2021**, 38(10), 1910-1937.

[26] L. Zhong, X. Diao, N. Zhang et al., Engineering and elucidation of the lipoinitiation process in nonribosomal peptide biosynthesis, *Nat Commun*, **2021**, 12(1), 296.

[27] S. E. O'Connor, H. Chen, C. T. Walsh, Enzymatic assembly of epothilones: the EpoC subunit and reconstitution of the EpoA-ACP/B/C polyketide and nonribosomal peptide interfaces, *Biochemistry*, **2002**, 41(17), 5685-5694.

[28] S. E. O'Connor, C. T. Walsh, F. Liu, Biosynthesis of epothilone intermediates with alternate starter units: engineering polyketide-nonribosomal interfaces, *Angew Chem Int Ed*, **2003**, 42(33), 3917-3921.

[29] N. Gaitatzis, A. Hans, R. Müller et al., The *mtaA* gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases, *J Biochem*, **2001**, 129(1), 119-124.

[30] X. Bian, F. Huang, F. A. Stewart et al., Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering, *ChemBioChem*, **2012**, 13(13), 1946-1952.

[31] M. Montalbán-López, T. A. Scott, S. Ramesh et al., New developments in RiPP discovery, enzymology and engineering, *Nat Prod Rep*, **2021**, 38(1), 130-239.

[32] G. Zhong, Z.-J. Wang, F. Yan et al., Recent Advances in Discovery, Bioengineering, and Bioactivity-Evaluation of Ribosomally Synthesized and Post-translationally Modified Peptides, *ACS Bio & Med Chem Au*, **2023**, 3(1), 1-31.

[33] C. Ongpipattanakul, E. K. Desormeaux, A. DiCaprio et al., Mechanism of Action of Ribosomally Synthesized and Post-Translationally Modified Peptides, *Chem Rev*, **2022**, 122(18), 14722-14814.

[34] K. Viehrig, F. Surup, C. Volz et al., Structure and biosynthesis of crocagins: polycyclic posttranslationally modified ribosomal peptides from *Chondromyces crocatus*, *Angew Chem Int Ed*, **2017**, 56(26), 7407-7410.

[35] J. J. Hug, J. Dastbaz, S. Adam et al., Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from *Myxococcus xanthus*, *ACS Chem Biol*, **2020**, 15(8), 2221-2231.

[36] J. J. Hug, N. A. Frank, C. Walt et al., Genome-guided discovery of the first myxobacterial Biarylitide Myxarylin reveals C-N biaryl crosslinking in RiPP biosynthesis, *Molecules*, **2021**, 26(24), 7483.

[37] P. G. Arnison, M. J. Bibb, G. Bierbaum et al., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat Prod Rep*, **2013**, 30(1), 108-160.

[38] S. Groß, F. Panter, D. Pogorevc et al., Improved broad-spectrum antibiotics against Gram-negative pathogens via darobactin biosynthetic pathway engineering, *Chem Sci*, **2021**, 12(35), 11882-11893.

[39] C. E. Seyfert, C. Porten, B. Yuan et al., Darobactins Exhibiting Superior Antibiotic Activity by Cryo-EM Structure Guided Biosynthetic Engineering, *Angew Chem Int Ed*, **2023**, 62(2), e202214094.

[40] M. Lopatniuk, F. Riedel, J. Wildfeuer et al., Development of a *Streptomyces*-based system for facile thioholgamide library generation and analysis, *Metab Eng*, **2023**, in press, DOI: 10.1016/j.ymben.2023.04.015.

[41] A. Sikandar and J. Koehnke, The role of protein-protein interactions in the biosynthesis of ribosomally synthesized and post-translationally modified peptides, *Nat Prod Rep*, **2019**, 36(11), 1576-1588.

[42] H. Lee and W. A. van der Donk, Macrocyclization and Backbone Modification in RiPP Biosynthesis, *Annu Rev Biochem*, **2022**, 91, 269-294.

[43] Y. Zhao, E. Marschall, M. Treisman et al., Cytochrome P450$_{\mathrm{Blt}}$ Enables Versatile Peptide Cyclisation to Generate Histidine- and Tyrosine-Containing Crosslinked Tripeptide Building Blocks, *Angew Chem Int Ed*, **2022**, 61(37), e202204957.

[44] A. Caruso, R. J. Martinie, L. B. Bushin et al., Macrocyclization via an Arginine-Tyrosine Crosslink Broadens the Reaction Scope of Radical *S*-Adenosylmethionine Enzymes, *J Am Chem Soc*, **2019**, 141(42), 16610-16614.

[45] S. Adam, D. Zheng, A. Klein et al., Unusual peptide-binding proteins guide pyrroloindoline alkaloid formation in crocagin biosynthesis, *Nat Chem*, **2023**, 15(4), 560-568.

[46] S. Guo, S. Wang, S. Ma et al., Radical SAM-dependent ether crosslink in daropeptide biosynthesis, *Nat Commun*, **2022**, 13(1), 2361.

[47] S. Mordhorst, F. Ruijne, A. L. Vagstad et al., Emulating nonribosomal peptides with ribosomal biosynthetic strategies, *RSC Chem Biol*, **2023**, 4(1), 7-36.

[48] F. Hubrich, N. M. Bösch, C. Chepkirui et al., Ribosomally derived lipopeptides containing distinct fatty acyl moieties, *Proc Natl Acad Sci USA*, **2022**, 119(3), e2113120119.

[49] L. Franz, U. Kazmaier, A. W. Truman et al., Bottromycins - biosynthesis, synthesis and activity, *Nat Prod Rep*, **2021**, 38(9), 1659-1683.

[50] C. D. Bader, F. Panter, R. Müller, In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites, *Biotechnol Adv*, **2020**, 39, 107480.

[51] Y. Imai, K. J. Meyer, A. Iinishi et al., A new antibiotic selectively kills Gram-negative pathogens, *Nature*, **2019**, 576(7787), 459-464.

[52] M. Tulp and L. Bohlin, Rediscovery of known natural compounds: nuisance or goldmine?, *Bioorg Med Chem*, **2005**, 13(17), 5274-5282.

[53] R. Pan, X. Bai, J. Chen et al., Exploring structural diversity of microbe secondary metabolites using OSMAC strategy: a literature review, *Front Microbiol*, **2019**, 10, 294.

[54] F. Panter, D. Krug, R. Müller, Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the *Myxococcus fulvus* Secondary Metabolome, *ACS Chem Biol*, **2019**, 14(1), 88-98.

[55] L. Etzbach, A. Plaza, R. Garcia et al., Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria, *Org Lett*, **2014**, 16(9), 2414-2417.

[56] M. Wang, J. J. Carver, V. V. Phelan et al., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nat Biotechnol*, **2016**, 34(8), 828-837.

[57] L. Katz and R. H. Baltz, Natural product discovery: past, present, and future, *J Ind Microbiol Biotechnol*, **2016**, 43(2-3), 155-176.

[58] K. Gregory, L. A. Salvador, S. Akbar et al., Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential, *Microorganisms*, **2019**, 7(6).

[59] N. Ziemert, M. Alanjary, T. Weber, The evolution of genome mining in microbes - a review, *Nat Prod Rep*, **2016**, 33(8), 988-1005.

[60] Y. Luo, R. E. Cobb, H. Zhao, Recent advances in natural product discovery, *Curr Opin Biotechnol*, **2014**, 30, 230-237.

[61] B. O. Bachmann, S. G. van Lanen, R. H. Baltz, Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?, *J Ind Microbiol Biotechnol*, **2014**, 41(2), 175-184.

[62] H. Gross, V. O. Stockwell, M. D. Henkels et al., The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters, *Chem Biol*, **2007**, 14(1), 53-63.

[63] C. S. McCaughey, J. A. van Santen, J. J. J. van der Hooft et al., An isotopic labeling approach linking natural products with biosynthetic gene clusters, *Nat Chem Biol*, **2022**, 18(3), 295-304.

[64] X. Vila-Farres, J. Chu, D. Inoyama et al., Antimicrobials Inspired by Nonribosomal Peptide Synthetase Gene Clusters, *J Am Chem Soc*, **2017**, 139(4), 1404-1407.

[65] T. Weber and H. U. Kim, The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production, *Synth Syst Biotechnol*, **2016**, 1(2), 69-79.

[66] M. H. Medema, K. Blin, P. Cimermancic et al., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, *Nucleic Acids Res*, **2011**, 39(suppl_2), W339-W346.

[67] K. Blin, S. Shaw, A. M. Kloosterman et al., antiSMASH 6.0: improving cluster detection and comparison capabilities, *Nucleic Acids Res*, **2021**, 49(W1), W29-W35.

[68] J. Santos-Aberturas, G. Chandra, L. Frattaruolo et al., Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool, *Nucleic Acids Res*, **2019**, 47(9), 4624-4637.

[69] A. M. Kloosterman, K. E. Shelton, G. P. van Wezel et al., RRE-Finder: a Genome-Mining Tool for Class-Independent RiPP Discovery, *mSystems*, **2020**, 5(5), e00267-20.

[70] P. Agrawal, S. Khater, M. Gupta et al., RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links, *Nucleic Acids Res*, **2017**, 45(W1), W80-W88.

[71] S. F. Altschul, W. Gish, W. Miller et al., Basic local alignment search tool, *J Mol Biol*, **1990**, 215(3), 403-410.

[72] K. Blin, M. H. Medema, R. Kottmann et al., The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters, *Nucleic Acids Res*, **2017**, 45(D1), D555-D559.

[73] B. R. Terlouw, K. Blin, J. C. Navarro-Muñoz et al., MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters, *Nucleic Acids Res*, **2023**, 51(D1), D603-D610.

[74] J. J. L. Malit, H. Y. C. Leung, P.-Y. Qian, Targeted Large-Scale Genome Mining and Candidate Prioritization for Natural Product Discovery, *Mar Drugs*, **2022**, 20(6), 398.

[75] L. Y. Li, Y. L. Hu, J. L. Sun et al., Resistance and phylogeny guided discovery reveals structural novelty of tetracycline antibiotics, *Chem Sci*, **2022**, 13(43), 12892-12898.

[76] K. H. Almabruk, L. K. Dinh, B. Philmus, Self-Resistance of Natural Product Producers: Past, Present, and Future Focusing on Self-Resistant Protein Variants, *ACS Chem Biol*, **2018**, 13(6), 1426-1437.

[77] X. Tang, J. Li, N. Millán-Aguiñaga et al., Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining, *ACS Chem Biol*, **2015**, 10(12), 2841-2849.

[78] F. Panter, D. Krug, S. Baumann et al., Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria, *Chem Sci*, **2018**, 9(21), 4898-4908.

[79] E. C. O'Neill, M. Schorn, C. B. Larson et al., Targeted antibiotic discovery through biosynthesis-associated resistance determinants: target directed genome mining, *Crit Rev Microbiol*, **2019**, 45(3), 255-277.

[80] S. Baumann, J. Herrmann, R. Raju et al., Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity, *Angew Chem Int Ed*, **2014**, 53(52), 14605-14609.

[81] J. J. Hug, F. Panter, D. Krug et al., Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria, *J Ind Microbiol Biotechnol*, **2019**, 46(3-4), 319-334.

[82] A. Kling, P. Lukat, D. V. Almeida et al., Targeting DnaN for tuberculosis therapy using novel griselimycins, *Science*, **2015**, 348(6239), 1106-1112.

[83] S. C. Wenzel, H. Hoffmann, J. Zhang et al., Production of the Bengamide Class of Marine Natural Products in Myxobacteria: Biosynthesis and Structure-Activity Relationships, *Angew Chem Int Ed*, **2015**, 54(51), 15560-15564.

[84] X. Xu, H. Zhou, Y. Liu et al., Heterologous Expression Guides Identification of the Biosynthetic Gene Cluster of Chuangxinmycin, an Indole Alkaloid Antibiotic, *J Nat Prod*, **2018**, 81(4), 1060-1064.

[85] N. Liu, E. D. Abramyan, W. Cheng et al., Targeted Genome Mining Reveals the Biosynthetic Gene Clusters of Natural Product CYP51 Inhibitors, *J Am Chem Soc*, **2021**, 143(16), 6043-6047.

[86] H.-H. Yeh, M. Ahuja, Y.-M. Chiang et al., Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in *Aspergillus nidulans* Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor, *ACS Chem Biol*, **2016**, 11(8), 2275-2284.

[87] Y. Yan, Q. Liu, X. Zang et al., Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action, *Nature*, **2018**, 559(7714), 415-418.

[88] E. J. Culp, D. Sychantha, C. Hobson et al., ClpP inhibitors are produced by a widespread family of bacterial gene clusters, *Nat Microbiol*, **2022**, 7(3), 451-462.

[89] M. Alanjary, B. Kronmiller, M. Adamek et al., The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery, *Nucleic Acids Res*, **2017**, 45(W1), W42-W48.

[90] M. D. Mungan, M. Alanjary, K. Blin et al., ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining, *Nucleic Acids Res*, **2020**, 48(W1), W546-W552.

[91] Y. Yan, N. Liu, Y. Tang, Recent developments in self-resistance gene directed natural product discovery, *Nat Prod Rep*, **2020**, 37(7), 879-892.

[92] E. Zazopoulos, K. Huang, A. Staffa et al., A genomics-guided approach for discovering and expressing cryptic metabolic pathways, *Nat Biotechnol*, **2003**, 21(2), 187-190.

[93] Z.-X. Liang, Complexity and simplicity in the biosynthesis of enediyne natural products, *Nat Prod Rep*, **2010**, 27(4), 499-528.

[94] B. Shen, Hindra, X. Yan et al., Enediynes: Exploration of microbial genomics to discover new anticancer drug leads, *Bioorg Med Chem Lett*, **2015**, 25(1), 9-15.

[95] J. D. Rudolf, X. Yan, B. Shen, Genome neighborhood network reveals insights into enediyne biosynthesis and facilitates prediction and prioritization for discovery, *J Ind Microbiol Biotechnol*, **2016**, 43(2-3), 261-276.

[96] X. Yan, H. Ge, T. Huang et al., Strain Prioritization and Genome Mining for Enediyne Natural Products, *mBio*, **2016**, 7(6), e02104-16.

[97] X. Yan, J.-J. Chen, A. Adhikari et al., Genome Mining of *Micromonospora yangpuensis* DSM 45577 as a Producer of an Anthraquinone-Fused Enediyne, *Org Lett*, **2017**, 19(22), 6192-6195.

[98] Y.-M. Shi, J. J. Crames, L. Czech et al., Genome Mining Enabled by Biosynthetic Characterization Uncovers a Class of Benzoxazolinate-Containing Natural Products in Diverse Bacteria, *Angew Chem Int Ed*, **2022**, 61(51), e202206106.

[99] K. D. Bauman, K. S. Butler, B. S. Moore et al., Genome mining methods to discover bioactive natural products, *Nat Prod Rep*, **2021**, 38(11), 2100-2129.

[100] M. G. Chevrette, A. Gavrilidou, S. Mantri et al., The confluence of big data and evolutionary genome mining for the discovery of natural products, *Nat Prod Rep*, **2021**, 38(11), 2024-2040.

[101] P. Cimermancic, M. H. Medema, J. Claesen et al., Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters, *Cell*, **2014**, 158(2), 412-421.

[102] J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney et al., A computational framework to explore large-scale biosynthetic diversity, *Nat Chem Biol*, **2020**, 16(1), 60-68.

[103] E. J. Culp, N. Waglechner, W. Wang et al., Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling, *Nature*, **2020**, 578(7796), 582-587.

[104] T. Hoffmann, D. Krug, N. Bozkurt et al., Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria, *Nat Commun*, **2018**, 9(1), 803.

[105] Y.-M. Chiang, S.-L. Chang, B. R. Oakley et al., Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms, *Curr Opin Chem Biol*, **2011**, 15(1), 137-143.

[106] P. J. Rutledge and G. L. Challis, Discovery of microbial natural products by activation of silent biosynthetic gene clusters, *Nat Rev Microbiol*, **2015**, 13(8), 509-523.

[107] D. Mao, B. K. Okada, Y. Wu et al., Recent advances in activating silent biosynthetic gene clusters in bacteria, *Curr Opin Microbiol*, **2018**, 45, 156-163.

[108] F. Panter, C. D. Bader, R. Müller, The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium, *Angew Chem Int Ed*, **2021**, 60(15), 8081-8088.

[109] J. Yin, H. Zhu, L. Xia et al., A new recombineering system for *Photorhabdus* and *Xenorhabdus*, *Nucleic Acids Res*, **2015**, 43(6), e36.

[110] X. Wang, H. Zhou, H. Chen et al., Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species, *Proc Natl Acad Sci USA*, **2018**, 115(18), E4255-E4263.

[111] R. Li, H. Shi, X. Zhao et al., Development and application of an efficient recombineering system for *Burkholderia glumae* and *Burkholderia plantarii*, *Microb Biotechnol*, **2021**, 14(4), 1809-1826.

[112] H. Chen, T. Sun, X. Bai et al., Genomics-Driven Activation of Silent Biosynthetic Gene Clusters in *Burkholderia gladioli* by Screening Recombineering System, *Molecules*, **2021**, 26(3), 700.

[113] W. Zheng, X. Wang, H. Zhou et al., Establishment of recombineering genome editing system in *Paraburkholderia megapolitana* empowers activation of silent biosynthetic gene clusters, *Microb Biotechnol*, **2020**, 13(2), 397-405.

[114] W. Zheng, X. Wang, Y. Chen et al., Recombineering facilitates the discovery of natural product biosynthetic pathways in *Pseudomonas parafulva*, *Biotechnol J*, **2021**, 16(8), e2000575.

[115] R. Chen, Q. Zhang, B. Tan et al., Genome Mining and Activation of a Silent PKS/NRPS Gene Cluster Direct the Production of Totopotensamides, *Org Lett*, **2017**, 19(20), 5697-5700.

[116] B. Wang, F. Guo, S.-H. Dong et al., Activation of silent biosynthetic gene clusters using transcription factor decoys, *Nat Chem Biol*, **2019**, 15(2), 111-114.

[117] D. Mao, L. B. Bushin, K. Moon et al., Discovery of *scmR* as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264, *Proc Natl Acad Sci USA*, **2017**, 114(14), E2920-E2928.

[118] S. Izzat, S. Rachid, A. Ajdidi et al., The ROK like protein of *Myxococcus xanthus* DK1622 acts as a pleiotropic transcriptional regulator for secondary metabolism, *J Biotechnol*, **2020**, 311, 25-34.

[119] M. R. Seyedsayamdost, High-throughput platform for the discovery of elicitors of silent bacterial gene clusters, *Proc Natl Acad Sci USA*, **2014**, 111(20), 7266-7271.

[120] C. N. Teijaro, A. Adhikari, B. Shen, Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts, *J Ind Microbiol Biotechnol*, **2019**, 46(3-4), 433-444.

[121] K. Yamanaka, K. A. Reynolds, R. D. Kersten et al., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A, *Proc Natl Acad Sci USA*, **2014**, 111(5), 1957-1962.

[122] Y. Luo, H. Huang, J. Liang et al., Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster, *Nat Commun*, **2013**, 4(1), 2894.

[123] D. Montiel, H.-S. Kang, F.-Y. Chang et al., Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters, *Proc Natl Acad Sci USA*, **2015**, 112(29), 8953-8958.

[124] A. Ullrich, Y. Chai, D. Pistorius et al., Pretubulysin, a potent and chemically accessible tubulysin precursor from *Angiococcus disciformis*, *Angew Chem Int Ed Engl*, **2009**, 48(24), 4422-4425.

[125] J. Chu, B. Koirala, N. Forelli et al., Synthetic-Bioinformatic Natural Product Antibiotics with Diverse Modes of Action, *J Am Chem Soc*, **2020**, 142(33), 14158-14168.

[126] Z. Wang, A. Kasper, R. Mehmood et al., Metagenome-Guided Analogue Synthesis Yields Improved Gram-Negative-Active Albicidin- and Cystobactamid-Type Antibiotics, *Angew Chem Int Ed*, **2021**, 60(41), 22172-22177.

[127] J. Chu, X. Vila-Farres, D. Inoyama et al., Discovery of MRSA active antibiotics using primary sequence from the human microbiome, *Nat Chem Biol*, **2016**, 12(12), 1004-1006.

[128] Z. Wang, N. Forelli, Y. Hernandez et al., Lapcin, a potent dual topoisomerase I/II inhibitor discovered by soil metagenome guided total chemical synthesis, *Nat Commun*, **2022**, 13(1), 842.

[129] Z. Wang, B. Koirala, Y. Hernandez et al., A naturally inspired antibiotic to target multidrug-resistant pathogens, *Nature*, **2022**, 601(7894), 606-611.

[130] Z. Wang, B. Koirala, Y. Hernandez et al., Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance, *Science*, **2022**, 376(6596), 991-996.

[131] L. Li, B. Koirala, Y. Hernandez et al., Identification of structurally diverse menaquinone-binding antibiotics with in vivo activity against multidrug-resistant pathogens, *Nat Microbiol*, **2022**, 7(1), 120-131.

[132] J. Chu, X. Vila-Farres, S. F. Brady, Bioactive Synthetic-Bioinformatic Natural Product Cyclic Peptides Inspired by Nonribosomal Peptide Synthetase Gene Clusters from the Human Microbiome, *J Am Chem Soc*, **2019**, 141(40), 15737-15741.

[133] Y. Ma, Z. Guo, B. Xia et al., Identification of antimicrobial peptides from the human gut microbiome using deep learning, *Nat Biotechnol*, **2022**, 40(6), 921-931.

[134] J. R. Doroghazi, J. C. Albright, A. W. Goering et al., A roadmap for natural product discovery based on large-scale genomics and metabolomics, *Nat Chem Biol*, **2014**, 10(11), 963-968.

[135] F. Panter, C. D. Bader, R. Müller, Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics, *Chem Sci*, **2021**, 12(17), 5994-6010.

[136] S. K. Hight, T. N. Clark, K. L. Kurita et al., High-throughput functional annotation of natural products by integrated activity profiling, *Proc Natl Acad Sci USA*, **2022**, 119(49), e2208458119.

[137] R. D. Kersten, Y.-L. Yang, Y. Xu et al., A mass spectrometry-guided genome mining approach for natural product peptidogenomics, *Nat Chem Biol*, **2011**, 7(11), 794-802.

[138] J.-Q. Hu, J.-J. Wang, Y.-L. Li et al., Combining NMR-Based Metabolic Profiling and Genome Mining for the Accelerated Discovery of Archangiumide, an Allenic Macrolide from the Myxobacterium *Archangium violaceum* SDU8, *Org Lett*, **2021**, 23(6), 2114-2119.

[139] N. S. Cortina, D. Krug, A. Plaza et al., Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome, *Angew Chem Int Ed*, **2012**, 51(3), 811-816.

[140] M. Xu, Y. Wang, Z. Zhao et al., Functional Genome Mining for Metabolites Encoded by Large Gene Clusters through Heterologous Expression of a Whole-Genome Bacterial Artificial Chromosome Library in *Streptomyces* spp, *Appl Environ Microbiol*, **2016**, 82(19), 5795-5805.

[141] M. Xu, F. Zhang, Z. Cheng et al., Functional Genome Mining Reveals a Class V Lanthipeptide Containing a D-Amino Acid Introduced by an $F_{420}H_2$-Dependent Reductase, *Angew Chem Int Ed*, **2020**, 59(41), 18029-18035.

[142] M. A. Schorn, S. Verhoeven, L. Ridder et al., A community resource for paired genomic and metabolomic data mining, *Nat Chem Biol*, **2021**, 17(4), 363-368.

[143] Q. Wu, B. A. Bell, J.-X. Yan et al., Metabolomics and Genomics Enable the Discovery of a New Class of Nonribosomal Peptidic Metallophores from a Marine *Micromonospora*, *J Am Chem Soc*, **2023**, 145(1), 58-69.

[144] J. J. Zhang, X. Tang, B. S. Moore, Genetic platforms for heterologous expression of microbial natural products, *Nat Prod Rep*, **2019**, 36(9), 1313-1332.

[145] J. Fu, X. Bian, S. Hu et al., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting, *Nat Biotechnol*, **2012**, 30(5), 440-446.

[146] H. Wang, Z. Li, R. Jia et al., ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes, *Nucleic Acids Res*, **2018**, 46(5), e28.

[147] C. Song, J. Luan, R. Li et al., RedEx: a method for seamless DNA insertion and deletion in large multimodular polyketide synthase gene clusters, *Nucleic Acids Res*, **2020**, 48(22), e130.

[148] C. Osswald, G. Zipf, G. Schmidt et al., Modular construction of a functional artificial epothilone polyketide pathway, *ACS Synth Biol*, **2014**, 3(10), 759-772.

[149] S. Groß, B. Schnell, P. A. Haack et al., In vivo and in vitro reconstitution of unique key steps in cystobactamid antibiotic biosynthesis, *Nat Commun*, **2021**, 12(1), 1696.

[150] F. Yan, C. Burgard, A. Popoff et al., Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production, *Chem Sci*, **2018**, 9(38), 7510-7519.

[151] M. Z. Li and S. J. Elledge, Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC, *Nat Methods*, **2007**, 4(3), 251-256.

[152] Y. Xia, K. Li, J. Li et al., T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis, *Nucleic Acids Res*, **2019**, 47(3), e15.

[153] D. G. Gibson, L. Young, R.-Y. Chuang et al., Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat Methods*, **2009**, 6(5), 343-345.

[154] J. E. Bird, J. Marles-Wright, A. Giachino, A User's Guide to Golden Gate Cloning Methods and Standards, *ACS Synth Biol*, **2022**, 11(11), 3551-3563.

[155] C. Song, J. Luan, Q. Cui et al., Enhanced Heterologous Spinosad Production from a 79-kb Synthetic Multioperon Assembly, *ACS Synth Biol*, **2019**, 8(1), 137-147.

[156] Z. Shao, H. Zhao, H. Zhao, DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways, *Nucleic Acids Res*, **2009**, 37(2), e16.

[157] J. J. Zhang, K. Yamanaka, X. Tang et al., Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination, *Methods Enzymol*, **2019**, 621, 87-110.

[158] W. Jiang, X. Zhao, T. Gabrieli et al., Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters, *Nat Commun*, **2015**, 6, 8101.

[159] B. Enghiad, C. Huang, F. Guo et al., Cas12a-assisted precise targeted cloning using in vivo Cre-*lox* recombination, *Nat Commun*, **2021**, 12(1), 1171.

[160] M. Liang, L. Liu, F. Xu et al., Activating cryptic biosynthetic gene cluster through a CRISPR-Cas12a-mediated direct cloning approach, *Nucleic Acids Res*, **2022**, 50(6), 3581-3592.

[161] H. Wang, X. Bian, L. Xia et al., Improved seamless mutagenesis by recombineering using *ccdB* for counterselection, *Nucleic Acids Res*, **2014**, 42(5), e37.

[162] H. Wang, Z. Li, R. Jia et al., RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression, *Nat Protoc*, **2016**, 11(7), 1175-1190.

[163] M. N. Abbasi, J. Fu, X. Bian et al., Recombineering for Genetic Engineering of Natural Product Biosynthetic Pathways, *Trends Biotechnol*, **2020**, 38(7), 715-728.

[164] R. E. Cobb, Y. Wang, H. Zhao, High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system, *ACS Synth Biol*, **2015**, 4(6), 723-728.

[165] G.-Y. Tan, K. Deng, X. Liu et al., Heterologous Biosynthesis of Spinosad: An Omics-Guided Large Polyketide Synthase Gene Cluster Reconstitution in *Streptomyces*, *ACS Synth Biol*, **2017**, 6(6), 995-1005.

[166] C. Jiang, H. Zhou, H. Sun et al., Establishing an efficient salinomycin biosynthetic pathway in three heterologous *Streptomyces* hosts by constructing a 106-kb multioperon artificial gene cluster, *Biotechnol Bioeng*, **2021**, 118(12), 4668-4677.

[167] Z.-J. Wang, X. Liu, H. Zhou et al., Engineering of *Burkholderia thailandensis* strain E264 serves as a chassis for expression of complex specialized metabolites, *Front Microbiol*, **2022**, 13, 1073243.

[168] M. Winn, J. K. Fyans, Y. Zhuo et al., Recent advances in engineering nonribosomal peptide assembly lines, *Nat Prod Rep*, **2016**, 33(2), 317-347.

[169] M. Klaus and M. Grininger, Engineering strategies for rational polyketide synthase design, *Nat Prod Rep*, **2018**, 35(10), 1070-1081.

[170] C. Beck, J. F. G. Garzón, T. Weber, Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines, *Biotechnol Bioprocess Eng*, **2020**, 25(6), 886-894.

[171] X. Bian, A. Plaza, F. Yan et al., Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives in vivo, *Biotechnol Bioeng*, **2015**, 112(7), 1343-1353.

[172] H. Kries, D. L. Niquille, D. Hilvert, A subdomain swap strategy for reengineering nonribosomal peptides, *Chem Biol*, **2015**, 22(5), 640-648.

[173] W. L. Thong, Y. Zhang, Y. Zhuo et al., Gene editing enables rapid engineering of complex antibiotic assembly lines, *Nat Commun*, **2021**, 12(1), 6872.

[174] C. D. Reeves, S. Murli, G. W. Ashley et al., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations, *Biochemistry*, **2001**, 40(51), 15464-15470.

[175] H. Wang, J. Liang, Q. Yue et al., Engineering the acyltransferase domain of epothilone polyketide synthase to alter the substrate specificity, *Microb Cell Fact*, **2021**, 20(1), 86.

[176] E. Kalkreuter, J. M. CroweTipton, A. N. Lowell et al., Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units, *J Am Chem Soc*, **2019**, 141(5), 1961-1969.

[177] J.-M. Massicard, C. Soligot, K. J. Weissman et al., Manipulating polyketide stereochemistry by exchange of polyketide synthase modules, *Chem Commun (Cambridge, UK)*, **2020**, 56(84), 12749-12752.

[178] E. E. Drufva, N. R. Spengler, E. G. Hix et al., Site-Directed Mutagenesis of Modular Polyketide Synthase Ketoreductase Domains for Altered Stereochemical Control, *ChemBioChem*, **2021**, 22(7), 1122-1150.

[179] K. Kudo, T. Hashimoto, J. Hashimoto et al., In vitro Cas9-assisted editing of modular polyketide synthase genes to produce desired natural product derivatives, *Nat Commun*, **2020**, 11(1), 4022.

[180] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck et al., *De novo* design and engineering of non-ribosomal peptide synthetases, *Nat Chem*, **2018**, 10(3), 275-281.

[181] L. Zhang, T. Hashimoto, B. Qin et al., Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides, *Angew Chem Int Ed*, **2017**, 56(7), 1740-1745.

[182] A. T. Keatinge-Clay, Polyketide Synthase Modules Redefined, *Angew Chem Int Ed*, **2017**, 56(17), 4658-4660.

[183] H. Peng, K. Ishida, Y. Sugimoto et al., Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases, *Nat Commun*, **2019**, 10(1), 3918.

[184] L. Su, L. Hôtel, C. Paris et al., Engineering the stambomycin modular polyketide synthase yields 37-membered mini-stambomycins, *Nat Commun*, **2022**, 13(1), 515.

[185] K. Gong, M. Wang, Q. Duan et al., High-yield production of FK228 and new derivatives in a *Burkholderia* chassis, *Metab Eng*, **2023**, 75, 131-142.

[186] K. A. J. Bozhüyük, A. Linck, A. Tietze et al., Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains, *Nat Chem*, **2019**, 11(7), 653-661.

[187] R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft et al., The structure of docking domains in modular polyketide synthases, *Chem Biol*, **2003**, 10(8), 723-731.

[188] T. J. Buchholz, T. W. Geders, F. E. Bartley et al., Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains, *ACS Chem Biol*, **2009**, 4(1), 41-52.

[189] C. Hacker, X. Cai, C. Kegler et al., Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS, *Nat Commun*, **2018**, 9(1), 4366.

[190] K. A. J. Bozhueyuek, J. Watzel, N. Abbood et al., Synthetic Zippers as an Enabling Tool for Engineering of Non-Ribosomal Peptide Synthetases, *Angew Chem Int Ed*, **2021**, 60(32), 17531-17538.

[191] F. Pourmasoumi, S. De, H. Peng et al., Proof-Reading Thioesterase Boosts Activity of Engineered Nonribosomal Peptide Synthetase, *ACS Chem Biol*, **2022**, 17(9), 2382-2388.

[192] M. Baunach, S. Chowdhury, P. Stallforth et al., The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity, *Mol Biol Evol*, **2021**, 38(5), 2116-2130.

[193] A. S. Brown, M. J. Calcott, J. G. Owen et al., Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines, *Nat Prod Rep*, **2018**, 35(11), 1210-1228.

[194] A. Wlodek, S. G. Kendrew, N. J. Coates et al., Diversity oriented biosynthesis via accelerated evolution of modular gene clusters, *Nat Commun*, **2017**, 8(1), 1206.

[195] K. J. Weissman, Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology, *Nat Prod Rep*, **2016**, 33(2), 203-230.

[196] R. H. Baltz, Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways, *ACS Synth Biol*, **2014**, 3(10), 748-758.

[197] K. Alam, J. Hao, Y. Zhang et al., Synthetic biology-inspired strategies and tools for engineering of microbial natural product biosynthetic pathways, *Biotechnol Adv*, **2021**, 49, 107759.

[198] S. E. Ongley, X. Bian, B. A. Neilan et al., Recent advances in the heterologous expression of microbial natural product biosynthetic pathways, *Nat Prod Rep*, **2013**, 30(8), 1121-1138.

[199] L. Huo, J. J. Hug, C. Fu et al., Heterologous expression of bacterial natural product biosynthetic pathways, *Nat Prod Rep*, **2019**, 36(10), 1412-1436.

[200] Y. Luo, B.-Z. Li, D. Liu et al., Engineered biosynthesis of natural products in heterologous hosts, *Chem Soc Rev*, **2015**, 44(15), 5265-5290.

[201] L. Li, L. W. MacIntyre, S. F. Brady, Refactoring biosynthetic gene clusters for heterologous production of microbial natural products, *Curr Opin Biotechnol*, **2021**, 69, 145-152.

[202] J. Fu, S. C. Wenzel, O. Perlova et al., Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition, *Nucleic Acids Res*, **2008**, 36(17), e113.

[203] L.-P. Zhu, X.-J. Yue, K. Han et al., Allopatric integrations selectively change host transcriptomes, leading to varied expression efficiencies of exotic genes in *Myxococcus xanthus*, *Microb Cell Fact*, **2015**, 14, 105.

[204] B. Bilyk, L. Horbal, A. Luzhetskyy, Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074, *Microb Cell Fact*, **2017**, 16(1), 5.

[205] V. Magrini, C. Creighton, P. Youderian, Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration, *J Bacteriol*, **1999**, 181(13), 4050-4061.

[206] B. Julien, Characterization of the integrase gene and attachment site for the *Myxococcus xanthus* bacteriophage Mx9, *J Bacteriol*, **2003**, 185(21), 6325-6330.

[207] H.-S. Kang and E.-S. Kim, Recent advances in heterologous expression of natural product biosynthetic gene clusters in *Streptomyces* hosts, *Curr Opin Biotechnol*, **2021**, 69, 118-127.

[208] M. Myronovskyi and A. Luzhetskyy, Heterologous production of small molecules in the optimized *Streptomyces* hosts, *Nat Prod Rep*, **2019**, 36(9), 1281-1294.

[209] Y. Gao, C. Walt, C. D. Bader et al., Genome-Guided Discovery of the Myxobacterial Thiolactone-Containing Sorangibactins, *ACS Chem Biol*, **2023**, 18(4), 924-932.

[210] L. Tang, S. Shah, L. Chung et al., Cloning and heterologous expression of the epothilone gene cluster, *Science*, **2000**, 287(5453), 640-642.

[211] B. Julien and S. Shah, Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*, *Antimicrob Agents Chemother*, **2002**, 46(9), 2772-2778.

[212] J. Lau, S. Frykman, R. Regentin et al., Optimizing the heterologous production of epothilone D in *Myxococcus xanthus*, *Biotechnol Bioeng*, **2002**, 78(3), 280-288.

[213] S. C. Mutka, J. R. Carney, Y. Liu et al., Heterologous production of epothilone C and D in *Escherichia coli*, *Biochemistry*, **2006**, 45(4), 1321-1330.

[214] S. R. Park, J. W. Park, W. S. Jung et al., Heterologous production of epothilones B and D in *Streptomyces venezuelae*, *Appl Microbiol Biotechnol*, **2008**, 81(1), 109-117.

[215] X. Bian, B. Tang, Y. Yu et al., Heterologous Production and Yield Improvement of Epothilones in *Burkholderiales* Strain DSM 7029, *ACS Chem Biol*, **2017**, 12(7), 1805-1812.

[216] Y. Yu, H. Wang, B. Tang et al., Reassembly of the Biosynthetic Gene Cluster Enables High Epothilone Yield in Engineered *Schlegelella brevitalea*, *ACS Synth Biol*, **2020**, 9(8), 2009-2022.

[217] R. Zirkle, J. M. Ligon, I. Molnár, Heterologous production of the antifungal polyketide antibiotic soraphen A of *Sorangium cellulosum* So ce26 in *Streptomyces lividans*, *Microbiology*, **2004**, 150(Pt 8), 2761-2774.

[218] F. Gross, M. W. Ring, O. Perlova et al., Metabolic engineering of *Pseudomonas putida* for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation, *Chem Biol*, **2006**, 13(12), 1253-1264.

[219] O. Perlova, J. Fu, S. Kuhlmann et al., Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*, *Appl Environ Microbiol*, **2006**, 72(12), 7485-7494.

[220] S. C. Wenzel, F. Gross, Y. Zhang et al., Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering, *Chem Biol*, **2005**, 12(3), 349-356.

[221] O. Perlova, K. Gerth, S. Kuhlmann et al., Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate *Corallococcus macrosporus* GT-2, *Microb Cell Fact*, **2009**, 8, 1.

[222] Y. Chai, S. Shan, K. J. Weissman et al., Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis, *Chem Biol*, **2012**, 19(3), 361-371.

[223] Y. Sun, Z. Feng, T. Tomura et al., Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway, *Sci Rep*, **2016**, 6, 22091.

[224] Q. Tu, J. Herrmann, S. Hu et al., Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering, *Sci Rep*, **2016**, 6, 21066.

[225] H. Sucipto, D. Pogorevc, E. Luxenburger et al., Heterologous production of myxobacterial α-pyrone antibiotics in *Myxococcus xanthus*, *Metab Eng*, **2017**, 44, 160-170.

[226] D. Pogorevc, F. Panter, C. Schillinger et al., Production optimization and biosynthesis revision of corallopyronin A, a potent anti-filarial antibiotic, *Metab Eng*, **2019**, 55, 201-211.

[227] F. Yan, D. Auerbach, Y. Chai et al., Biosynthesis and heterologous production of vioprolides: rational biosynthetic engineering and unprecedented 4-methylazetidinecarboxylic acid formation, *Angew Chem Int Ed*, **2018**, 57(28), 8754-8759.

[228] D. Pogorevc, Y. Tang, M. Hoffmann et al., Biosynthesis and Heterologous Production of Argyrins, *ACS Synth Biol*, **2019**, 8(5), 1121-1133.

Chapter 2. The disorazole Z family of highly potent anticancer natural products from *Sorangium cellulosum*: structure, bioactivity, biosynthesis and heterologous expression

Microbiology Spectrum, accepted as a Research Article, 2023

Yunsheng Gao+,[a,c,d] Joy Birkelbach+,[a] Chengzhang Fu,[a,d] Jennifer Herrmann,[a] Herbert Irschik,[b] Bernd Morgenstern,[e] Kerstin Hirschfelder,[a] Ruijuan Li,[c] Youming Zhang,[c] Rolf Jansen*,[b] and Rolf Müller*[a,d]

[a] Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

[b] Department of Microbial Drugs, Helmholtz Centre for Infection Research (HZI), Inhoffenstraße 7, 38124 Braunschweig, Germany

[c] Helmholtz International Lab for Anti-Infectives, Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, 266237 Qingdao, China

[d] Helmholtz International Lab for Anti-Infectives, Helmholtz Center for Infection Research (HZI), 38124 Braunschweig, Germany

[e] Department of Inorganic Chemistry, Saarland University, Campus C4.1, 66123 Saarbrücken, Germany

[+] These authors contributed equally to this work

[*] To whom correspondence should be addressed:

rolf.mueller@helmholtz-hips.de and rolf.jansen@helmholtz-hzi.de

Keywords

myxobacteria, disorazole Z, anticancer natural products, direct cloning, heterologous expression

Author's efforts:

The author performed *in silico* analysis, direct cloning, promoter engineering, gene knockout, and heterologous expression of the disorazole Z biosynthetic gene cluster, as well as the subsequent metabolome and quantification analysis. The author constructed protein expression plasmids, purified the methyltransferase and performed the *in vitro* reaction. The author did large-scale fermentation using the heterologous expression system and purified the corresponding disorazole Z compounds. The author prepared the manuscript with inputs from other authors.

Other's efforts:

Joy Birkelbach made significant contributions by analyzing the NMR data and preparing the manuscript. Jennifer Herrmann and Kerstin Hirschfelder contributed to the biological evaluation section. Rolf Jansen and Herbert Irschik contributed to isolation and characterization of disorazole from *Sorangium cellulosum*, as well as draft preparation. Chengzhang Fu contributed to biosynthesis analysis and draft preparation. Bernd Morgenstern contributed to crystallographic analysis. Youming Zhang and Ruijuan Li contributed to supervision of the research. Rolf Müller supervised the research and revised the manuscript.

Signatures:

Yunsheng Gao: Yunsheng Gao
Chengzhang Fu:
Herbert Irschik:
Kerstin Hirschfelder:
Youming Zhang:
Rolf Müller:

Joy Birkelbach:
Jennifer Herrmann:
Bernd Morgenstern:
Ruijuan Li:
Rolf Jansen: Rolf Jansen

2.1 Abstract

Myxobacteria serve as a treasure trove for secondary metabolites. During the course of our ongoing search for bioactive natural products, a novel subclass of disorazoles termed disorazole Z was discovered. Ten disorazole Z family members were purified from a large-scale fermentation of the myxobacterium *Sorangium cellulosum* So ce1875 and characterized by ESI-HRMS, X-ray, NMR, and Mosher ester analysis. Disorazole Z compounds are characterized by the lack of one polyketide extension cycle resulting in a shortened monomer in comparison to disorazole A, which finally forms a dimer in the bis-lactone core structure. In addition, an unprecedented modification of a geminal dimethyl group takes place to form a carboxylic acid methyl ester. The main component disorazole Z1 shows comparable activity in effectively killing cancer cells to disorazole A1 via binding to tubulin, which we show to induce microtubule depolymerization, endoplasmic reticulum delocalization and eventually apoptosis. The disorazole Z biosynthetic gene cluster (BGC) was identified and characterized from the alternative producer *S. cellulosum* So ce427 and compared to the known disorazole A BGC, followed by heterologous expression in the host *Myxococcus xanthus* DK1622. Pathway engineering by promoter substitution and gene deletion paves the way for detailed biosynthesis studies and efficient heterologous production of disorazole Z congeners.

2.2 Introduction

Myxobacteria are soil-dwelling Gram-negative delta proteobacteria and known as prolific source of natural products often exhibiting novel chemical scaffolds and unique biological modes-of-action.[1-4] Disorazoles are a family of macrocyclic dilactones with dimeric or pseudodimeric structures, and represent one group of myxobacterial secondary metabolites showing promising cytotoxic activity.[5,6] Furthermore, this family of compounds was shown to block the invasion of group A *Streptococci* into human epithelial cells via targeting the host factor ezrin.[7] Therefore, disorazoles have potential to be developed as anti-cancer or anti-infective drugs.

Previously, 29 disorazole A family compounds (disorazole A1 and 28 variants) were isolated from the fermentation culture of the myxobacterium *Sorangium cellulosum* So ce12.[8] The main component disorazole A1 (**1**) showed extremely high cytotoxicity with IC_{50} values in the range of 2−42 pmol/L against a panel of cancer cell lines including a multidrug-resistant KB line by inhibiting tubulin polymerization and inducing apoptosis.[9] In 2005, the disorazole A biosynthetic gene cluster from *S. cellulosum* So ce12 (*dis12* gene cluster) was identified by transposon mutagenesis.[10,11] The core biosynthetic gene cluster encodes a multifunctional megasynthetase consisting of polyketide synthases (PKS), a nonribosomal peptide synthetase (NRPS) and a dedicated acyltransferase (AT), and thus belongs to the family of *trans*-AT PKS-NRPS hybrid gene clusters.[12,13] In 2016, the *dis12* gene cluster was subcloned and heterologously expressed in *Myxococcus xanthus* DK1622 leading to the production of disorazole A2 (**2**) as the major compound.[14]

In the course of our ongoing screening for biologically active natural products from myxobacteria, a novel subclass of disorazole, i.e., disorazole Z was discovered to be produced by several *S. cellulosum* strains such as So ce1875 and So ce427.[15] Structure elucidation and characterization of disorazole Z has not been published to date. Nevertheless and due to its outstanding anticancer activity, disorazole Z was subjected to preclinical evaluation as a cytotoxic component in a drug-targeting approach for the treatment of luteinizing hormone-releasing hormone (LHRH) receptor overexpressing cancers by conjugation with the receptor-targeting moiety D-Lys6-LHRH.[16-19] Compared to doxorubicin-D-Lys6-LHRH, the disorazole Z conjugate demonstrated an increased cytotoxicity *in vitro* in HCC 1806 and MDA-MB-231 triple-negative breast cancer cells.[20] In the meantime, disorazole Z was also identified as maytansine-site ligand, which was confirmed by solving the crystal structure of the tubulin-disorazole Z complex.[21]

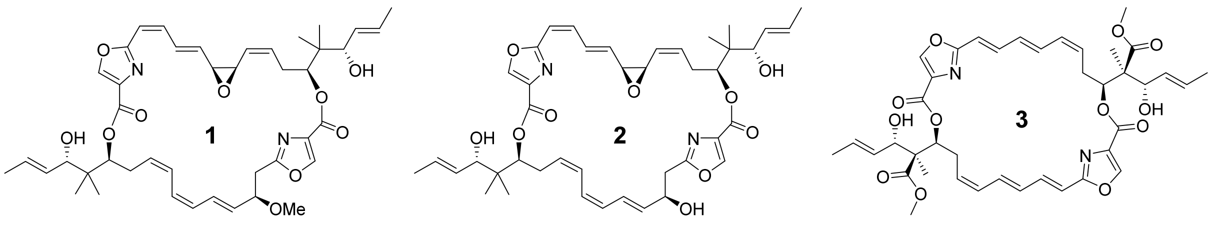
Due to the novelty in structures, their intriguing biological activities, and promising potential for clinical application, massive efforts were made to develop chemical synthesis routes. Total synthesis has been achieved for some disorazole congeners and non-natural analogs, for example, disorazole A1 (**1**) and also a simplified disorazole Z.[22,23] Nevertheless, biotechnological methods still present significant advantages towards obtaining such type of complex natural products.[24] Furthermore, taking advantage of synthetic biotechnology, rational engineering of the biosynthetic pathway in an advantageous heterologous host could also facilitate high-efficient production of certain target components or generate non-natural compounds.[25,26]

In this work, large-scale fermentation afforded the isolation of ten disorazole Z family members featuring two-carbon shorter monomers and unprecedented modification of a geminal dimethyl group compared to disorazole A. We report the first full structure elucidation of the disorazole Z congeners using NMR, Mosher ester and X-ray analysis. Furthermore, we characterized the bioactivity of disorazole Z1 (**3**). In addition, the disorazole Z biosynthetic pathway was identified by comparative analysis and heterologous expression. Promoter engineering and gene deletion experiments were carried out using the heterologous expression system which allowed the function of a methyltransferase involved in disorazole Z biosynthesis to be assigned.

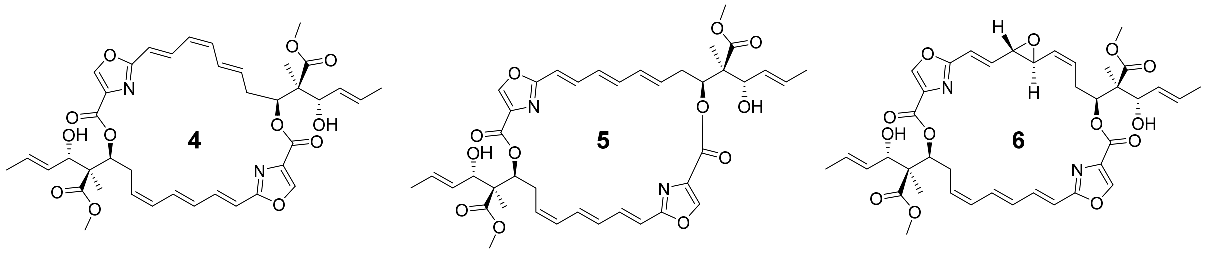
2.3 Results and discussion

2.3.1 Isolation and full structure elucidation of the disorazole Z family of compounds

After large scale fermentation of *S. cellulosum* So ce1875, we were able to isolate ten disorazole Z congeners (**3**−**12**), achieving a production titer of 60−80 mg/L disorazole Z1 (**3**). An overview of the chemical structures of these compounds is given in Figure 1. Details of fermentation and purification are provided in the Methods.

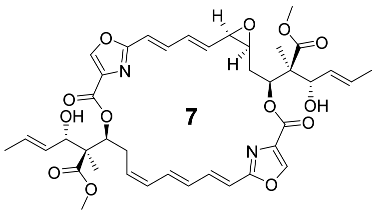


Disorazole A1 $C_{43}H_{54}N_2O_{10}$ Disorazole A2 $C_{42}H_{52}N_2O_{10}$ Disorazole Z1 $C_{40}H_{46}N_2O_{12}$

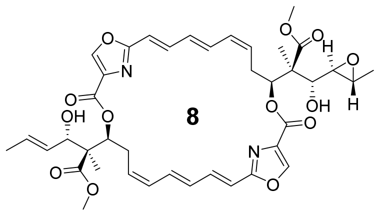


Disorazole Z2 $C_{40}H_{46}N_2O_{12}$ Disorazole Z3 $C_{40}H_{46}N_2O_{12}$ Disorazole Z4 $C_{40}H_{46}N_2O_{13}$

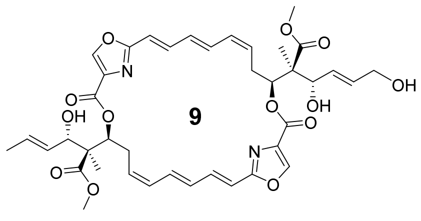
Figure 1 The chemical structures of disorazole Z family compounds and disorazole A for comparison.



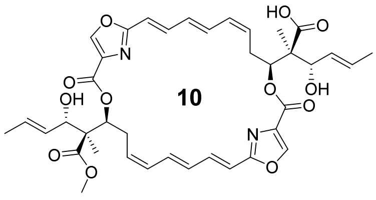
Disorazole Z5
$C_{40}H_{46}N_2O_{13}$



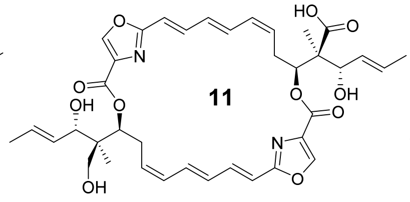
Disorazole Z6
$C_{40}H_{46}N_2O_{13}$



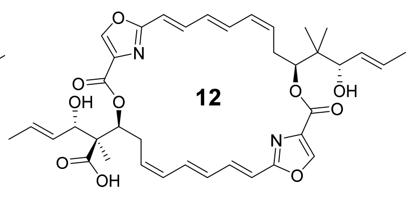
Disorazole Z7
$C_{40}H_{46}N_2O_{13}$



Disorazole Z8
$C_{39}H_{44}N_2O_{12}$



Disorazole Z9
$C_{38}H_{44}N_2O_{11}$



Disorazole Z10
$C_{38}H_{44}N_2O_{10}$

Figure 1 (continued) The chemical structures of disorazole Z family compounds and disorazole A for comparison.

Molecular ion clusters $[M+H]^+$ (*m/z* 747.3132), $[M+Na]^+$ and $[M-H_2O+H]^+$ in ESI-HRMS revealed the elemental formula $C_{40}H_{46}N_2O_{12}$ of disorazole Z1 (**3**) from which 19 double bond equivalents were calculated. Surprisingly, the ^{13}C NMR spectrum showed only half the number of carbon signals solely explicable for a completely symmetrical dimer. The ^1H NMR spectrum in acetone-d_6 presented 23 protons, which were correlated to their corresponding carbons in a $^1H,^{13}C$ HSQC NMR spectrum with the exception of one doublet of a secondary alcohol proton (OH-26) at 4.16 ppm. The $^1H,^1H$ COSY NMR spectrum allowed to assign two main structural parts A and B (Figure 2). H-5 of part A and a ^1H singlet (3-H) at 8.53 ppm were connected by a long-range coupling. The corresponding carbon C-3 had no correlation in the HMBC NMR spectrum. However, the direct $^1J_{C,H}$ coupling of 213 Hz indicated its role in a hetero aromatic ring.[27] The further ring carbons C-2 and C-4 were identified by their HMBC correlations with H-3 while the exact connection of the oxazole ring to structural part A was shown by the HMBC correlation between C-4 and H-5. However, H-6 did not show any HMBC correlation appearing as a flat and broad proton signal. The continuation of structural part A at the opposite oxymethine C-12 (δ_C 76.28) was indicated by an apparent low-field acylation shift (δ_H 5.45) of the H-12 proton signal that provided the only HMBC correlation of the carboxyl carbon C-13 (δ_C 159.65). By HMBC correlations with H-12, H-26, and the hydroxyl group OH-26, the quaternary sp^3 carbon C-25 (δ_C 56.26) was revealed as the interconnection of parts A and B.

Further, the HMBC correlations showed the direct attachment of the methyl group C-30 to C-25. The last open position was filled by a carboxymethyl ester at C-25 as indicated by three HMBC correlations of the carboxyl carbon C-31 (δ_C 173.81) with the methyl group C-30, the alcohol methine H-26 and the acylated oxymethine H-12. The stereochemistry of the *cis* and *trans* double bonds was derived from vicinal coupling constants of about 11 Hz and 15 Hz, respectively. With all NMR structural elements assigned, 18 double bond equivalents were consumed for two identical structural parts. Using the last equivalent both halves were connected to give the point symmetrical dilactone ring of **3**, in which all identical structural elements have the same vicinity.

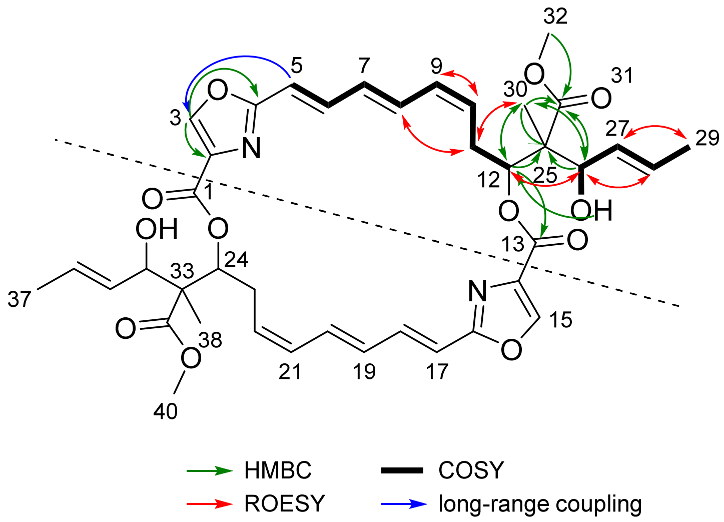


Figure 2 NMR correlations of structural parts A and B of disorazole Z1. ^{1}H,^{1}H-COSY correlations (fat bonds) and their interconnection by selected ^{1}H,^{13}C-HMBC correlations (green arrows), ^{1}H,^{1}H-NOESY correlations (red arrows) and ^{1}H,^{1}H-long-range coupling (blue arrow).

X-ray analysis revealed the relative configuration of disorazole Z1 (**3**). In order to elucidate the absolute configuration Mosher's method based upon ^{1}H and ^{13}C NMR chemical shift differences was applied.[28-31] Consequently, all asymmetric centers (12, 25, 26) of **3** are in the (*S*)-configuration as shown in Figure 1. A full description of the stereo-chemical analyses is given in the supporting information (SI). The structural assignment is in full agreement with the structure of the tubulin-disorazole Z complex published recently.[21]

Besides disorazole Z1 (**3**), nine more variants have also been isolated and characterized. Two of them turned out to be isomers of **3** with different double bond geometries, i.e., disorazole Z2 (**4**, $\Delta^{7,8}$-*cis*-disorazole Z) and disorazole Z3 (**5,** $\Delta^{9,10}$-*trans*-disorazole Z). Four more polar disorazole Z variants were identified as isomers with the ESI-HRMS-derived elemental formula $C_{40}H_{46}N_2O_{13}$. They contained one additional oxygen, either incorporated as an epoxide group replacing different double bonds, i.e., disorazole Z4 (**6**, 7,8-epoxy-disorazole Z), disorazole Z5 (**7**, 9,10-epoxy-disorazole Z) and disorazole Z6 (**8**, 27,28-epoxy-disorazole Z), or incorporated as primary unsaturated alcohol at the end of a side chain, i.e., disorazole Z7 (**9**, 29-hydoxy-disorazole Z). Three disorazole Z variants with smaller molecular weight were proposed to be biosynthetic intermediates featuring different degrees of modification on geminal methyl groups, i.e., disorazole Z8 (**10**, 31-*O*-desmethyl-disorazole Z), disorazole Z9 (**11**, 31-*O*-desmethyl-39-hydroxy-disorazole Z) and disorazole Z10 (**12**, 39-*O*-desmethyl-25,25-dimethyl-disorazole Z). Details of the structure elucidation of these analogues are given in the supporting information (SI).

2.3.2 Discovery and comparative analysis of disorazole Z biosynthetic gene cluster

The significant structural differences between the disorazole Z and A family compounds motivated us to investigate the disorazole Z biosynthetic pathway. The draft genome sequence of the alternative producer strain *S. cellulosum* So ce427 was obtained by Illumina sequencing and submitted to antiSMASH for annotation of biosynthetic gene clusters.[32] A *trans*-AT PKS-NRPS gene cluster (*dis427* gene cluster, GenBank accession number OQ408282) which exhibited significant overall similarity to the known *dis12* gene cluster (GenBank accession number DQ013294 or AJ874112), was speculated to be responsible for the biosynthesis of disorazole Z (Figure 3). Intriguingly, this gene cluster has two main distinguishing features: on the one hand, one PKS module is lacking in the second polyketide synthase gene *disB*$_{427}$, which corresponds nicely with the two-carbon shortened monomeric unit of disorazole Z compared to one half side of the disorazole A congener; on the other hand, a methyltransferase gene *disF*$_{427}$ was found downstream of *disD*$_{427}$, which is not present in the *dis12* gene cluster and could be assigned to the *O*-methylation function needed for the methylation of the carboxyl group found in disorazole Z. The carboxyl group most likely arises from the oxidation of one of the methyl groups at C-25 and C-33 positions. However, no gene surrounding the corresponding BGC encoding an expected oxidative function could be assigned for this hypothetic step. The organization of the remaining part of the disorazole Z pathway is quite similar to that of the disorazole A pathway.

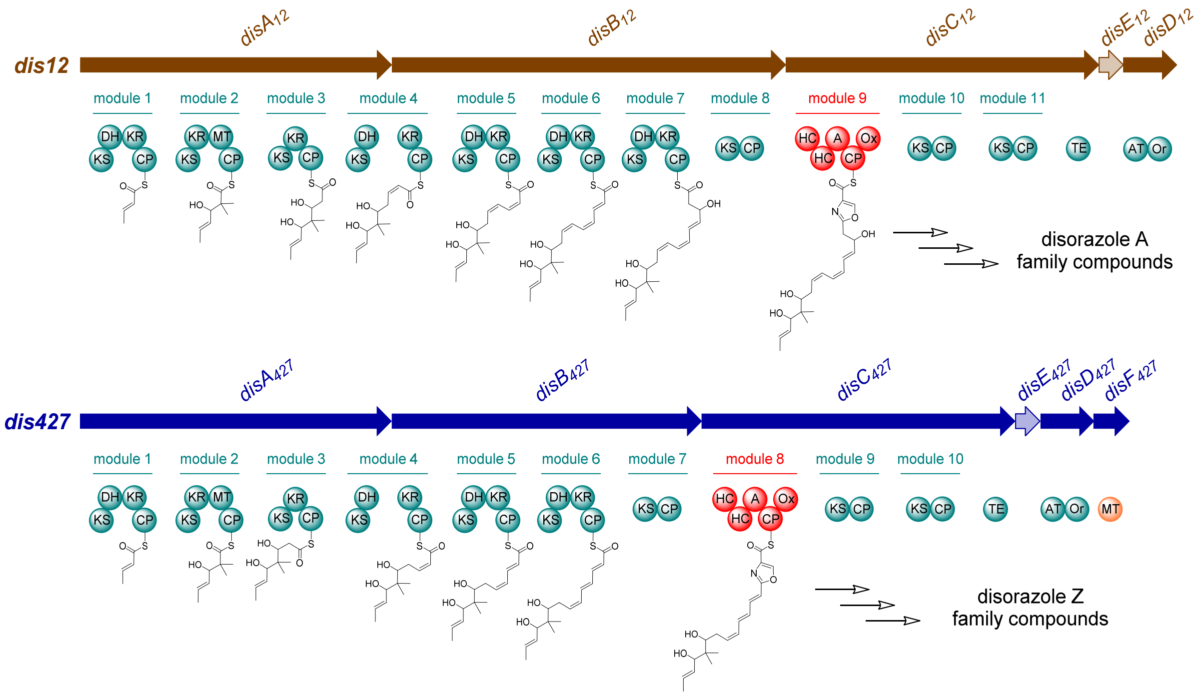


Figure 3 Comparison of proposed biosynthetic pathways between disorazole A and Z. *dis12*: the disorazole A biosynthetic gene cluster from *S. cellulosum* So ce12, *dis427*: the disorazole Z biosynthetic gene cluster from *S. cellulosum* So ce427, KS: ketosynthase, DH: dehydratase, KR: ketoreductase, CP: carrier protein, MT: methyltransferase, HC: heterocyclization domain, A: adenylation domain, Ox: oxidation domain, TE: thioesterase, AT: acyl transferase, Or: oxidoreductase. Oxidation of the methyl group and formation of the methyl ester may occur prior to or post-release of the bis-lactone.

The biosynthesis of the half-side of the disorazole Z dilactone begins with condensing one malonyl-CoA with the starter acetate. After additional five rounds of extension with malonyl-CoA, the ketosynthase (KS) domain in module 7 is not expected to extend the polyketide chain because the catalytic histidine of KS7 is substituted with alanine (Figure S7). The mutation of the conserved motif C-H-H to C-A-H most likely causes malfunction of this KS domain, as the histidine residue is known to play an essential role in decarboxylation and condensation reactions.[33] However, this type of KS was proven to be a gatekeeper and still capable of transferring the polyketide chain between different domains in FR901464 biosynthesis.[34] The adenylation domain of DisC$_{427}$ activates serine, also found in the disorazole A pathway. Similar tandem heterocyclization (HC) domains were proven to be essential for vibriobactin and anguibactin biosynthesis.[35,36] It is assumed likely that the HC domains of DisC$_{427}$ work in the same fashion. However, it is also possible that one of the HC domains condenses the serine with the polyketide chain, and the other HC domain might perform the cyclization of the serine moiety to form the oxazoline ring which is finally oxidized to its oxazole form by the oxidation domain. The extension of the nascent intermediate bound to the carrier protein of module 9 is supposed to stop here because the two following PKS modules are most likely non-extending ones, which is suggested by annotation as non-condensing KS9 and KS10 domains similar to KS7 (Figure S7). Finally, the termination of the assembly line and the cyclization of the two monomeric subunits are likely to be similar to the mechanism described for enterobactin, elaiophylin and conglobatin biosynthesis,[37-39] eventually forming the bis-lactone structure. The carboxylic acid methyl ester in disorazole Z might be introduced before or after release from the assembly line by stereospecific oxidation of a methyl group giving rise to the free carboxylic acid and subsequent methyl ester formation. Nevertheless, the oxidation process involved remains enigmatic and needs further studies. Intriguingly, after expression of the *dis427* gene cluster disorazole Z1 is also formed being the major heterologous product in the host *M. xanthus* DK1622 (see below), which indicates the missing functionalities to be either encoded within the BGC or the heterologous host to harbor similar genomic functions capable of oxidizing the precursor. The final step of methyl ester formation could finally be connected to DisF encoded in the corresponding BGC.

2.3.3 Heterologous production and biosynthesis of disorazole Z

Since *Sorangium cellulosum* strains are slow growing and unsuitable for *in situ* genetic manipulation, we cloned the disorazole Z biosynthetic gene cluster (from $disA_{427}$ to $disF_{427}$) from the genomic DNA of So ce427 for heterologous expression and functional verification. One-step capture of large gene clusters is usually challenging especially for myxobacteria due to their complex genomes. Therefore, the gene cluster was divided into three smaller parts for cloning and then assembled into a p15A-cm vector using RecET-mediated linear-linear homologous recombination (LLHR) (Figure S9).[40] After that, the chloramphenicol resistance gene was replaced by a *km-int* cassette coding for phage Mx8 integrase, which could be employed for site-specific integration of the gene cluster into the genome of the heterologous host *M. xanthus* DK1622.[41] The *dis427* gene cluster under the control of its native promoters was successfully expressed in *M. xanthus* DK1622, leading to the production of disorazole Z1 (**3**) as the primary compound, which has been confirmed by HPLC-MS and NMR analysis (Figure 4A, Table S3, Figure S17, Figure S18). However, the production yield is only about 0.2 mg/L under standard fermentation procedure. Replacement of the native promoter before *disA*$_{427}$ with a tetracycline-inducible Ptet promoter resulted in at least a 4-fold increase in heterologous production of **3** (Figure S10). Furthermore, inserting a strong constitutive Papr promoter before *disD*$_{427}$ led to a further improved yield but not by much. When the gene cluster was under the control of a vanillate-inducible Pvan promoter, a nearly 9-fold increase was achieved (Figure S10). Nevertheless, the production yield of **3** (about 1.8 mg/L) is still much lower than using the original producer strain under optimized fermentation conditions which may be due to the low level of phylogenetic similarity between *Sorangium* and *Myxococcus*. Further systematic genetic engineering to improve the yield using heterologous expression is therefore still required to achieve competitiveness. Nevertheless, this system, for the first time, allows for genetic manipulation of the disorazole Z biosynthesis as the native host was found genetically intractable despite significant efforts.

As mentioned above, the methyltransferase gene was found exclusively in type Z disorazole gene clusters and was thus supposed to be involved in methyl ester formation. In order to verify its function in disorazole Z biosynthesis, we replaced *disF*$_{427}$ with a gentamycin resistance gene using Redαβ-mediated linear-circular homologous recombination (LCHR; see SI).[42] Deletion of *disF*$_{427}$ completely abolished disorazole Z1 (**3**) production in *M. xanthus* DK1622, whereas the two compounds disorazole Z9 (**11**) and disorazole Z10 (**12**) accumulated in the culture broth, which were confirmed by HPLC-MS and NMR analysis (Figure 4A, Table S12, Table S14, Figure S55, Figure S61). The methyltransferase DisF$_{427}$ was then expressed and purified as N-terminal His-taged recombinant protein using *E. coli* BL21 (DE3). Incubating purified DisF$_{427}$ with **11** or **12** and *S*-adenosyl methionine resulted in almost complete conversion to the corresponding methylated compounds *in vitro* at 30 °C in 1 h (Figure 4B, Figure 4C). These results clearly demonstrated the methyltransferase DisF to be responsible for methyl ester formation in disorazole Z biosynthesis. The accumulation of **12** as the major component in the absence of *disF*$_{427}$ also indicated that stereospecific oxidation and subsequent methylation might occur initially on one side of the symmetric substrate. However, it remains unclear how the geminal methyl group was oxidized to the hydroxy and furthermore to the carboxyl group which implies novel tailoring biochemical steps and motivates further investigation which is ongoing in our laboratory.

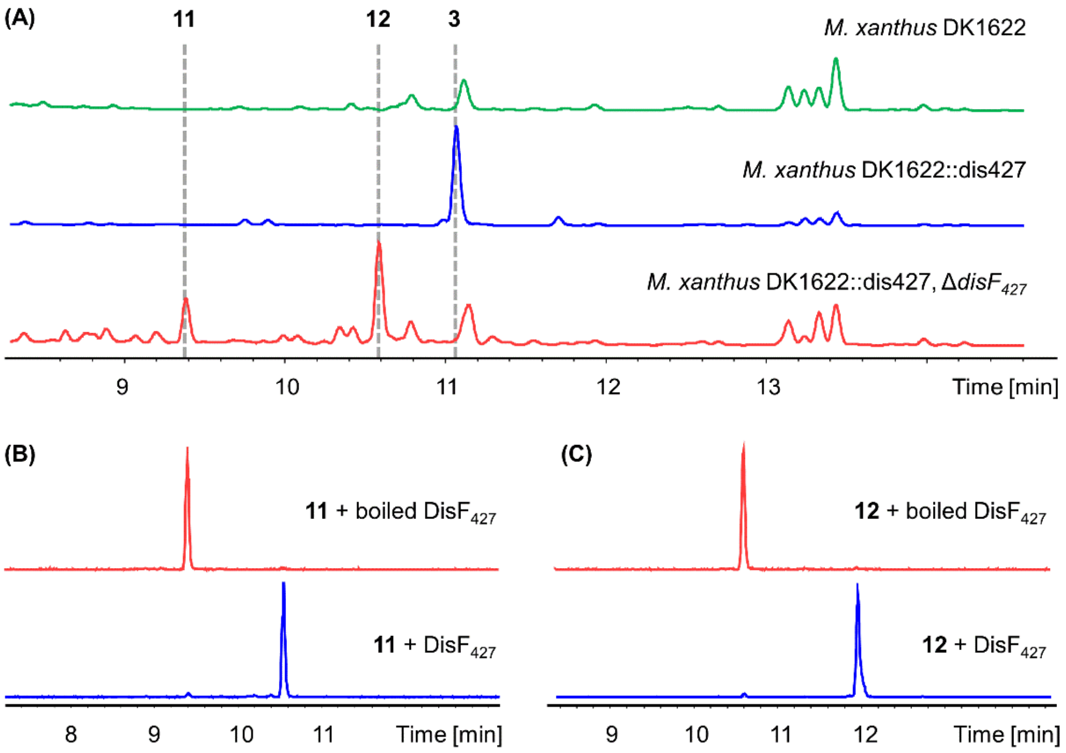


Figure 4 HPLC-MS analysis of disorazole Z formation in *M. xanthus* and *in vitro*. (A) HPLC chromatogram analyzed by UV at 320 nm. Blue trace: heterologous production of disorazole Z1 (**3**) in *M. xanthus* DK1622 mutant harboring the *dis427* gene cluster. Red trace: heterologous production of disorazole Z9 (**11**) and Z10 (**12**) after deletion of *disF*$_{427}$. Green trace: *M. xanthus* DK1622 wild type. (B) *in vitro* reaction of **11** with DisF$_{427}$ (blue trace) and boiled DisF$_{427}$ (red trace) shown as extracted ion chromatograms (EIC) of *m/z* 689.31 ± 0.05 and *m/z* 703.31 ± 0.05. (C) *in vitro* reaction of **12** with DisF$_{427}$ (blue trace) and boiled DisF$_{427}$ (red trace) shown as EIC of *m/z* 705.30 ± 0.05 and *m/z* 719.30 ± 0.05.

2.3.4 Biological activity of disorazole Z

Disorazole Z1 (**3**) was tested on a small panel of human cancer cell lines and displayed very pronounced cytotoxic activity with IC_{50}s in the range of 0.07−0.43 nM (Table S16). In comparison to the previously described disorazole A1 (**1**), it showed similar activity although being by tendency less potent (by a factor of 4 to 5) when tested on hepatocellular carcinoma (HepG2) and osteosarcoma (U-2 OS) cells.

In order to explore the effects of disorazole Z on microtubule dynamics, U-2 OS cells were treated with disorazoles followed by immunostaining of α-tubulin and fluorescence microscopy. After 5 h treatment with **1** or **3**, a slightly higher density of interphase microtubules around the nuclear periphery was observed, which resembles a local destabilization. The same effect was already described for e.g. the microtubule destabilizing agent vinblastine.[43] Interestingly, the acetylated microtubule population, which plays an important role in dynamic cellular processes, was much more affected. This might be caused by the ability of disorazoles to preferentially suppress dynamic mechanisms at the binding sites at the end of microtubules. After prolonged treatment, microtubules were completely depolymerized and the low-abundance acetylated tubulin population was not detectable anymore by immunostaining (Figure 5A). In particular endoplasmic reticulum (ER) dynamics are directly associated with acetylated microtubules – an effect termed ER sliding.[44] Thus, we studied whether specific responses of the ER can be observed after treatment of cells with disorazoles. Both compounds, **1** and **3**, induced a delocalization of the ER structure, which co-aligns with (depolymerized) microtubules. However, this event does not trigger ER stress as studied through applying GRP78/BiP immunostaining to disorazole-treated cells (Figure 5B). The 78-kDa glucose-regulated protein (GRP78) functions as a chaperone and is a master regulator of the unfolded protein response (UPR)[45] and it is found upregulated after treatment of cells with the ER calcium ATPase inhibitor thapsigargin. However, disorazoles did not directly induce ER stress although dynamics of the ER are probably greatly impaired due to the delocalization of the complex.

To further assess **3** and its ability to induce apoptotic processes, the mitochondrial membrane potential (MMP) and caspase activation were determined. Following microtubule depolymerization and a concurrent cell cycle arrest at the G2/M checkpoint, many tubulin binding agents are described to induce a loss of MMP followed by cytochrome-*c* release and activation of the caspase cascade.[46] Here, we could demonstrate that treatment of cells with **3** at nanomolar concentrations results in mitochondrial swelling (5 h), followed by a complete loss of MMP (24 h). In line with these findings, we found caspase-3/7 activation upon 24 h treatment of cells with **3** at concentrations as low as 0.3 nM (Figure 5C).

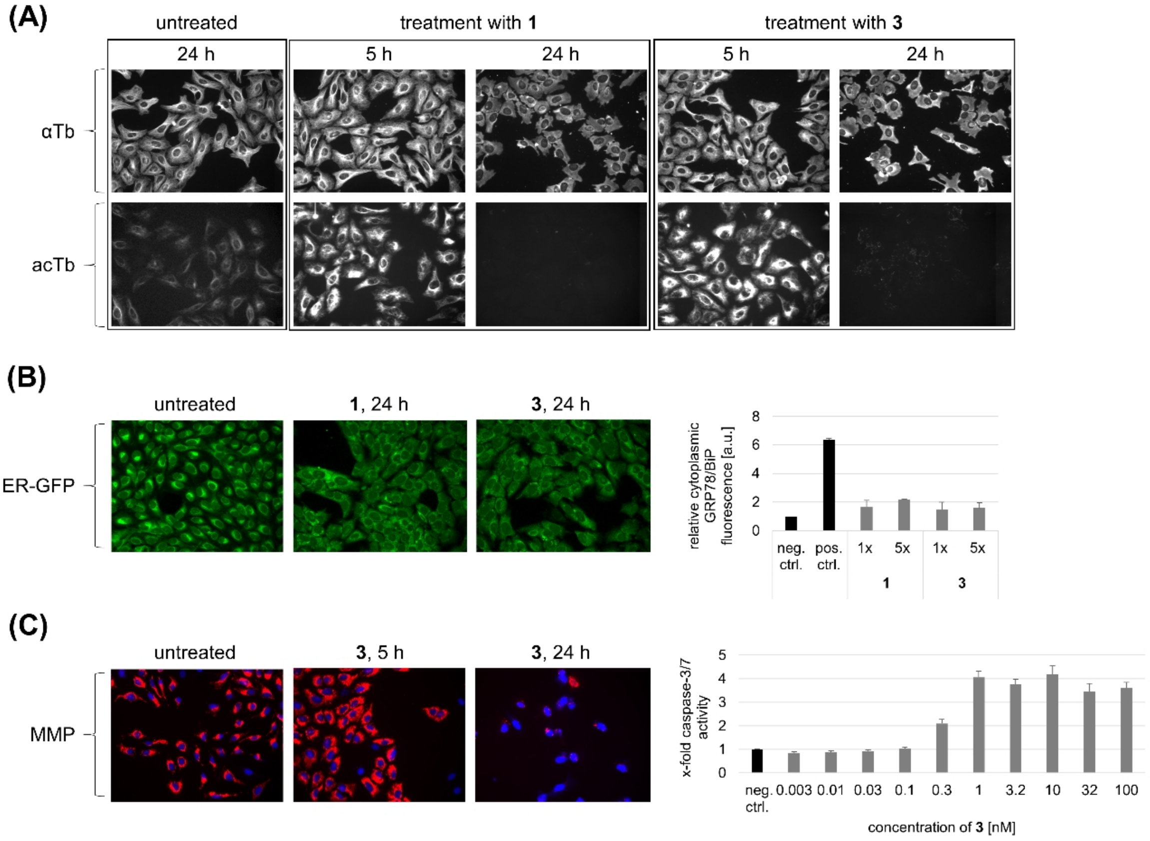


Figure 5 Biological characterization of disorazole Z. (A) Influence on microtubules. U-2 OS cells were treated with disorazole A1 (**1**) and disorazole Z1 (**3**) at their respective IC_{50}. α-tubulin (αTb) and acetylated tubulin (acTb) were visualized by immunofluorescence. (B) Influence on the endoplasmic reticulum (ER). Left panel: U-2 OS cells were transfected with CellLight® ER-GFP for live cell imaging of the ER after treatment with **1** and **3** at their respective IC_{50}. Right panel: Relative immunofluorescence quantification of high-content (HC) images of the ER stress marker protein GRP78/BiP after 24 h treatment of U-2 OS cells with **1** and **3** at their respective 1-fold and 5-fold IC_{50}s; positive control: 1 µM thapsigargin, negative control: untreated (was set to 1); data represent mean ± SD. (C) Induction of the intrinsic apoptosis pathway. Left panel: U-2 OS ells were stained with tetramethyl rhodamine methyl ester (TMRM) as fluorescence marker of the mitochondrial membrane potential (MMP) after treatment of U-2 OS with **3** at its 5-fold IC_{50}. Right panel: Quantification of caspase-3/7 activity in HepG2 cells after treatment with **3** in serial dilution; negative control: untreated (was set to 1); data represent mean ± SD.

2.4 Conclusion and outlook

In this study, we describe the full structural elucidation of ten novel disorazole congeners exhibiting a significantly modified basic structure compared to disorazole A and thus group as a new subclass of disorazole anticancer drugs termed disorazole Z. This family of compounds possesses a shortened polyketide chain in each half-side of the bis-lactone ring and a carboxymethyl ester at the position where a geminal methyl group is installed in disorazole A. The discovery of the disorazole Z biosynthetic gene cluster and comparison to the disorazole A biosynthetic pathway allowed to understand the structural differences between these two types of disorazoles. The successful heterologous expression of the disorazole Z gene cluster in an amenable host organism paved the way for detailed biosynthesis studies, e.g., elucidating the intriguing biosynthetic steps involving the oxidation of one methyl center in the geminal dimethyl group, and rational biosynthetic engineering to further improve the yield of disorazole Z. The system will also allow to generate non-natural disorazole family compounds through combinatorial biosynthesis. Activity assays of disorazole Z1 and disorazole A1 revealed similar biological activities in cancer cell lines and thus great potential for this family compounds to be employed as anti-tumor drugs, an approach which is being followed as a peptide-drug conjugate.

2.5 Methods

2.5.1 General experimental procedures

Melting points were measured on a Büchi-510 melting point apparatus. UV data were recorded on a Shimadzu UV/Vis-2450 spectrophotometer in methanol (UVASOL, Merck). IR data were recorded on a Bruker Tensor 27 IR spectrophotometer. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker Avance III 700, DMX 600, UltraShield 500 or DPX 300 NMR spectrometers, locked to the deuterium signal of the solvent. Data acquisition, processing and spectral analysis were performed with standard Bruker software and ACD/NMRSpectrus. Chemical shifts are given in parts per million (ppm) and coupling constants in Hertz (Hz). Analytical RP HPLC was carried out with an Agilent 1260 HPLC system equipped with a diode-array UV detector (DAD) and a Corona Ultra detector (Dionex) or a maXis ESI TOF mass spectrometer (ESI-HRMS, Bruker Daltonics). HPLC conditions: Waters Acquity C18 column 100×2 mm, 1.7 µm; solvent A: H_2O, 0.1% formic acid; solvent B: acetonitrile, 0.1% formic acid; gradient: 5%

B for 1 min, increasing to 95% B in 18 min; flow rate 0.6 mL/min; column temperature 45 °C. All elemental formulae were assigned using the high-resolution data of molecular ion clusters measured with a Bruker maXis ESI-TOF mass spectrometer and calculated with the SmartFormula tool of the Compass DataAnalysis program (Bruker). The myxobacterial strain *Sorangium cellulosum* So ce1875 was isolated in 2001 from soil with plant residues collected near Holbrook, Arizona, in 1996, and can be obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures) under the depository DSM 53600.

2.5.2 Fermentation of *S. cellulosum* So ce1875

A fermentation (300 L) was inoculated with 10 L pre-cultured *S. cellulosum* So ce1875 in 2-L-Erlenmeyer flaks. The fermentation medium contained 0.8% starch (Cerestar), 0.3% soya meal, 0.05% Casitone, 0.02% soya peptone, 0.1% $MgSO_4 \times 7H_2O$, 0.075% $CaCl_2 \times 2H_2O$, 8 mg/L Na-Fe-EDTA, 1% Amberlite XAD-16 resin. pH was 7.3 before autoclaving. 0.25% glucose × H_2O was added after autoclaving. The strain was cultivated at 30 °C with a pO_2 level at 20% for 14 days. At the end of fermentation the XAD resin was collected from the culture by sieving. A production 76 mg/L disorazole Z1 (**3**) was analyzed by HPLC.

2.5.3 Extraction and isolation of disorazole Z from *S. cellulosum* So ce1875

The XAD adsorber resin (3.03 kg) from a large-scale (300 L) fermentation of *S. cellulosum* strain So ce1875 was separated from adhering cells by floatation with water before it was eluted in a glass column with 2 bed volumes of 30% aqueous methanol followed by 3 bed volumes of 100% methanol. The methanol eluate was evaporated to an aqueous mixture, diluted with water and extracted twice with equal portions of ethyl acetate. After evaporation the aqueous oil was subjected to a 90% methanol-heptane partitioning removing lipid products with three equal portions of heptane. The aqueous oil was diluted with water and extracted with dichloromethane (DCM) to give 72 g crude extract after evaporation of the solvent. Crystallization from ethanol provided 36.7 g of raw crystalline disorazole Z1 (**3**), including several minor structural variants and about 1.7 equivalents of the solvent.

A portion of 10.7 g raw crystals was dissolved in DCM and toluene and evaporated to dryness twice before the material was subjected to Si-flash chromatography [Reveleris Silica Cartridge, 330 g, 61×223 mm (Grace), equilibrated with DCM; flow rate 90 mL/min; solvent A = DCM, solvent B = acetone; gradient: 0% B for 15 min, to 10% B in 15 min, 50 min at 10% B, to 20% B in 5 min, 10 min at 20% B, to 100% B in 35 min, 10 min at 100% B]. 8.2 g of disorazole **3** was obtained in the first peak (UV at 280 nm, ELS) between 40 and 85 min and crystalized from ethanol to give 6.8 g of disorazole **3** as crystals containing 1.7 equivalents of ethanol after drying in vacuum. Further fractions were collected with disorazole variants.

Fraction 3 (91−96 min), 760 mg, was separated by RP-MPLC [column 30×480 mm, ODS-AQ, 25 µ (YMC); solvent A = 50% aqueous methanol; solvent B = 100 % methanol; gradient: 20% to 34 % B in 120 min, to 45% B in 20 min, to 100% B in 10 min, flow 30 mL/min, UV detection 310 nm] to give fractions according to UV peaks. The fractions were evaporated and the remaining water was extracted with ethyl acetate yielding disorazole Z5 (**7**) (387 mg), disorazole Z6 (**8**), and disorazole Z1 (**3**) (35 mg) after evaporation to dryness.

Fraction 4 (97−109 min) (428 mg) was separated similarly with a gradient 20% B to 32% B in 105 min, to give disorazole Z4 (**6**) enriched (113 mg, amorphous) and 234 mg, which was crystalized from ethanol and furnished **6** (195 mg, cryst. containing 1.2 equivalents of ethanol).

Fraction 6 (628 mg) was separated similarly with a gradient 10% B to 25% B in 130 min to give a main peak of disorazole Z7 (**9**) (387 mg), that was crystalized from ethanol yielding 310 mg of white crystals.

Si-flash chromatography of mother liquor (38 g) of a disorazole Z crystallisation provided further fractions, which were separated by RP-MPLC.

Fraction 12 (2.7 g) was separated to by RP-MPLC [column 60×500 mm, YMC ODS-AQ 120 Å, 21 µ; solvent A = 50% methanol; solvent B = methanol; gradient: 15% B to 26% in 225 min, to 50% B in 120 min; flow 65 mL/min; UV detection 313 nm] to give disorazole Z8 (**10**) (593 mg) and *O*-desmethyl-dimethyl-disorazole Z10 (**12**) (503 mg) as main peaks, recovered after evaporation of the methanol and extraction with ethyl acetate.

Fraction 15 (500 mg) was separated be RP-MPLC [column 30×480 mm, ODS-AQ, 120 Å, 16 µ (YMC), solvent A = 50% aqueous methanol; solvent B = 100% methanol, gradient: 15% for 3 min, 15% B to 30% B in 173 min; flow 30 mL/min; UV detection 310 nm] to give a main peak at 125 min with 80 mg of disorazole Z9 (**11**).

A sample of 6 g disorazole Z from Si-Flash chromatography was further separated by RP-MPLC in two runs [column 60×500 mm, YMC ODS-AQ 120 Å, 21 µ; solvent A = 50% methanol; solvent B = methanol; gradient: 20% B for 160 min, to 30% B in 240 min; flow 60 mL/min; UV detection 313 nm] to give a mixture of disorazols (680 mg) eluting in front of **3**. This mixture was again separated by RP-MPLC [column 30×480 mm, ODS-AQ, 15 µ (YMC), solvent A = 50% aqueous methanol; solvent B = 100% methanol; gradient: 30% B for 47 min, 30% B to 40 % B in 40 min, to 100% B in 30 min; flow 30 mL/min, UV detection 310 nm] to give dirorazole Z3 (**5**) (284 mg). Stepwise crystallization from methanol/water gave about 200 mg of **5**.

Enriched disorazole Z1 (**3**) (2.33 g) from Si-Flash chromatography was separated by RP-MPLC [column 60×500 mm, YMC ODS-AQ 120 Å, 21 µ; solvent A = 50% methanol; solvent B = methanol; gradient: 20% B for 160 min, to 30% B in 240 min; flow 60 mL/min; UV detection 313 nm] to give **3** and a mixture of disorazols (48 mg) eluting after **3**. This fraction was separated by preparative thin layer chromatography (four plates, 20×20 cm, silica gel, 0.5 mm, solvent DCM/MeOH 19:1, UV detection 254 nm) to give after elution with the same solvent mixture **3** and at higher R_f value **5** (19 mg).

Disorazole Z1 (3): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 11.09 min; cryst. from ethanol with 1.7 equivalent ethanol $C_{40}H_{46}N_2O_{12} \times 1.7\ C_2H_6O_1$ M = 825.12; m.p. 128−130 °C; $[\alpha]^{22}_D$ +169.8 (*c* = 1.01, CH_3OH); UV (MeOH) (amorph): λ_{max} (log ε) 314 (4.948), 340 (sh) nm; UV (chloroform) (cryst.): λ_{max} (log ε) 315 (4.969), 322, 340 (sh) nm; IR (KBr) (amorph): ν_{max} 3486 (m), 2950 (m), 1730 (s), 1614 (m), 1312 (m), 1234 (m), 1153 (m), 1115 (s), 996 (s), 973 (m) cm^{-1}; NMR data, see Table S1; ESI-HRMS *m/z* 747.3132 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{12}$, 747.3123), *m/z* 769.2949 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3022 $[M-H_2O+H]^+$ (calcd. for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z2 ($\Delta^{7,8}$-*cis*-disorazole Z) (4): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 11.17 min; cryst. from methanol; m.p. 133−135 °C; $[\alpha]^{22}_D$ +305 (*c* = 0.41, $CHCl_3$); UV (chloroform) (cryst.): λ_{max} (log ε) 313 (5.014), 299, 342 (sh) nm; NMR data, see Table S4; ESI-HRMS *m/z* 747.3128 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{12}$, 747.3124), *m/z* 769.2944 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3020 $[M-H_2O+H]^+$ (calcd. for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z3 ($\Delta^{9,10}$-*trans*-disorazole Z) (5): $C_{40}H_{46}N_2O_{12}$ M = 746.80; t_R = 10.85 min; $[\alpha]^{22}_D$ –59.6 (*c* = 0.12, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 314 (4.948), 340 (sh) nm; NMR data, see Table S5; ESI-HRMS *m/z* 747.3123 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{12}$, 747.3124), *m/z* 769.2935 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{12}Na$, 769.2943), *m/z* 729.3012 $[M-H_2O+H]^+$ (calcd. for $C_{40}H_{45}N_2O_{11}$, 729.3018).

Disorazole Z4 (7,8-epoxy-disorazole Z) (6): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 10.09 min; m.p. 132−138 °C; $[\alpha]^{22}_D$ +170.9 (*c* = 0.23, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 266 (4.601), 328 (4.652), 344 (4.558), 276, 314 (sh) nm; NMR data, see Table S6; ESI-HRMS *m/z* 763.3074 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2896 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 745.2975 $[M-H_2O+H]^+$ (calcd. for $C_{40}H_{45}N_2O_{12}$, 745.2967), *m/z* 727.2866 $[M-2H_2O+H]^+$ (calcd. for $C_{40}H_{43}N_2O_{11}$, 727.2861), *m/z* 807.2991 $[M+HCOOH-H]^-$ (calcd. for $C_{41}H_{47}N_2O_{15}$, 807.2982).

Disorazole Z5 (9,10-epoxy-disorazole Z) (7): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 9.87 min; m.p. 137−148 °C; $[\alpha]^{22}_D$ +135.1 (*c* = 0.28, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 301 (4.823), 345 (4.416), 291, 323 (sh) nm; NMR data, see Table S7; ESI-HRMS *m/z* 763.3085 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 780.3351 $[M+NH_4]^+$ (calcd. for $C_{40}H_{50}N_3O_{13}$, 780.3338), *m/z* 785.2903 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 1542.6339 $[2M+NH_4]^+$ (calcd. for $C_{80}H_{96}N_5O_{26}$ 1542.6338), *m/z* 745.2977 $[M+H-H_2O]^+$ (calcd. for $C_{40}H_{45}N_2O_{12}$ 745.2967).

Disorazole Z6 (27,28-epoxy-disorazole Z) (8): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 10.04 min; m.p. 121−125 °C; $[\alpha]^{22}_D$ +20.3 (*c* = 0.272, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 315 (4.924), 300, 322, 342 (sh) nm; NMR data, see Table S8; ESI-HRMS *m/z* 763.3079 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2898 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{13}Na$, 785.2892), *m/z* 745.2971 $[M-H_2O+H]^+$ (calcd. for $C_{40}H_{45}N_2O_{12}$, 745.2967).

Disorazole Z7 (29-hydroxy-disorazole Z) (9): $C_{40}H_{46}N_2O_{13}$ M = 762.80; t_R = 8.96 min; $[\alpha]^{22}_D$ -11.6 (*c* = 0.23, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 316 (4.893), 300, 342 (sh) nm; NMR data, see Table S9; ESI-HRMS *m/z* 763.3075 $[M+H]^+$ (calcd. for $C_{40}H_{47}N_2O_{13}$ 763.3073), *m/z* 785.2883 $[M+Na]^+$ (calcd. for $C_{40}H_{46}N_2O_{13}Na$, 785.2892).

Disorazole Z8 (31-*O*-desmethyl-disorazole Z) (10): $C_{39}H_{44}N_2O_{12}$ M = 732.77; t_R = 9.38 min; $[\alpha]^{22}_D$ +305.2 (*c* = 0.41, $CHCl_3$); UV (MeOH) (amorph): λ_{max} (log ε) 316 (4.954), 300, 343 (sh) nm; NMR data, see Table S10; ESI-HRMS *m/z* 733.2958 $[M+H]^+$ (calcd. for $C_{39}H_{45}N_2O_{12}$, 733.2967), *m/z* 755.2779 $[M+Na]^+$ (calcd. for $C_{39}H_{44}N_2O_{12}Na$, 755.2786), *m/z* 715.2866 $[M-H_2O+H]^+$ (calcd. for $C_{39}H_{43}N_2O_{11}$, 715.2861).

Disorazole Z9 (31-*O*-desmethyl-39-hydroxy-disorazole Z) (11): $C_{38}H_{44}N_2O_{11}$ M = 704.76; t_R = 9.35 min; $[\alpha]^{22}_D$ +55.8 (*c* = 0.26, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 317 (4.942), 300, 344 (sh) nm; NMR data, see Table S11; ESI-HRMS *m/z* 705.3023 $[M+H]^+$ (calcd. for $C_{38}H_{45}N_2O_{11}$ 705.3018), *m/z* 687.2920 $[M-H_2O+H]^+$ (calcd. for $C_{38}H_{43}N_2O_{10}$, 687.2912), *m/z* 1409.5975 $[2M+H]^+$ (calcd. for $C_{76}H_{89}N_4O_{22}$, 1409.5963).

Disorazole Z10 (*O*-desmethyl-dimethyl-disorazole Z) (12): $C_{38}H_{44}N_2O_{10}$ M = 688.76; t_R = 10.59 min; $[\alpha]^{22}_D$ +25.4 (*c* = 0.24, $CHCl_3$); UV ($CHCl_3$): λ_{max} (log ε) 316 (4.825), 300, 342 (sh) nm; NMR data, see Table S13; ESI-HRMS *m/z* 689.3057 $[M+H]^+$ (calcd. for $C_{38}H_{44}N_2O_{10}$ 689.3069), *m/z* 671.2952 $[M-H_2O+H]^+$ (calcd. for $C_{38}H_{43}N_2O_9$, 671.2963).

2.5.4 Preparation of disorazole Z Mosher esters

To prepare disorazole Z (*S*)-Mosher ester (**13**), 20 mg of crystalline disorazole Z1 (**3**) were twice dissolved in pyridine and toluene and evaporated to dryness. The residue was dissolved in 0.2 mL of dry pyridine. 25 µL of (*R*)-(-)-α-Methoxy-α-trifluoromethylphenylacetyl chloride were added in three portions to the stirred solution for 24 h. The mixture was dissolved with pyridine and sodium hydrogen carbonate solution (1%), extracted with DCM, washed with water twice and dried by evaporation with toluene. The residue was purified by Si-flash chromatography [12 g Silica gel 40 µm, Reveleris (Grace), solvent A: petroleum ether, solvent B: ethyl acetate, gradient: 0% B for 1 min, in 1 min to 9% B, 9% B for 1 min, in 4.7 min to 36% B, for 2.5 min 36% B, flow 36 mL/min]. The man peak was collected and evaporated to dryness yielding 39 mg of **13**. $C_{60}H_{60}F_6N_2O_{16}$ M = 1179.1; NMR data, see Table S2.

The disorazole Z (*R*)-Mosher ester (**14**) was prepared analogously using (*S*)-(+)-α-methoxy-α-trifluoromethylphenylacetyl chloride yielding 19 mg of **14**. $C_{60}H_{60}F_6N_2O_{16}$ M = 1179.1; NMR data, see Table S2.

2.5.5 Crystallography and X-ray analysis

Crystallographic data of 3, 6 and 7: Crystals suitable for single-crystal X-ray analysis were obtained from ethanol. The data set for all structures was collected using a Bruker X8 Apex diffractometer. Graphite-monochromated $Mo_{K\alpha}$ radiation ($\lambda = 0.71073$ Å) was used throughout. Data were collected at 133(2) K (disorazole Z1 (**3**) = sh3137_a_sq, disorazole Z5 (**7**) = sh3279), or 152(2) K (disorazole Z4 (**6**) = sh3191), respectively, and corrected for absorption effects using the multi-scan method. The structures were solved by direct methods using SHELXS-97[47] and were refined by full matrix least squares calculations on F^2 (SHELXL2018[48]) in the graphical user interface Shelxle[49].

Refinement: All non H-atoms were located in the electron density maps and refined anisotropically. C-bound H atoms were placed in positions of optimized geometry and treated as riding atoms. Their isotropic displacement parameters were coupled to the corresponding carrier atoms by a factor of 1.2 (CH, CH2) or 1.5 (CH3, OH) for **3** = sh3137_a_sq and **6** = sh3191. Restraints of 0.84 (0.01) Å were used for O-H bond lengths. For **7** = sh3279 the O bonded H atoms were located in the electron density maps. Their positional parameters were refined using isotropic displacement parameters which were set at 1.5 times the Ueq value of the parent atom. *Disorder:* For **7** = sh3279, each oxirane atom is not fully occupied (O4a: 0.76, O4b: 0.24), furthermore, the propylene group of the main compound as well as one solvent ethanol molecule are split over two positions. Their occupancy factors refined to 0.87 and 0.76, respectively for the major components. For **6** = sh3191, two of the ethanol solvent molecules and the propylene residue of the structure were split over two positions. Their occupancy factors refined to 0.55, 0.86 and 0.81 for the major components, respectively. *SQUEEZE:* For **3** = sh3137_a_sq, the unit cell contains approximately 2 EtOH solvent molecules (occupancy factor less than 1.0 for each EtOH), which have been treated as a diffuse contribution to the overall scattering without specific atom positions by SQUEEZE/PLATON.

Data deposit: Crystallographic data for the structure have been deposited with the Cambridge Crystallographic Data Centre, CCDC, 12 Union Road, Cambridge CB21EZ, UK. Copies of the data can be obtained free of charge on quoting the depository number: 2236753 for disorazole Z1 (**3**), 1834868 for disorazole Z4 (**6**) and 1834869 for disorazole Z5 (**7**).

2.5.6 Cloning and engineering of the *dis427* gene cluster

The myxobacterium *Sorangium cellulosum* So ce427 was cultivated using CYH medium (1.5 g/L casitone, 1.5 g/L yeast extract, 4 g/L starch, 1 g/L soya meal, 1 g/L glucose, 1 g/L calcium chloride dihydrate, 0.5 g/L magnesium sulphate heptahydrate, 5.96 g/L HEPES, 4 mg/L Na-

Fe-EDTA, pH=7.3) at 30 °C. Clumpy cells were collected by centrifugation and then homogenized to become suspension for isolation of genomic DNA according to the published protocol.[42] The genomic DNA was treated with RNase A and appropriate DNA restriction endonucleases (MluCI or BstXI) to eliminate RNA contamination and to release the gene cluster. The digested DNA was recovered by phenolic chloroform extraction and ethanol precipitation and finally dissolved in autoclaved Milli-Q water. The linear cloning vectors (p15A-cm or pBR322-amp) containing homology arms with homology to the end of released genomic DNA fragment were achieved by PCR. The digested DNA and corresponding cloning vector were then used for electroporation of *E. coli* GB05-dir expressing RecET recombinase for linear-linear DNA homologous recombination (LLHR).[42] Recombinant plasmids were isolated from antibiotic-resistant colonies and verified by restriction digestion and Sanger/Illumina sequencing. Correct recombinant plasmids p15A-cm-MluCI-dis427 and pBR322-amp-BstXI-dis427 were then digested with MluCI and BstXI respectively to release the cloned fragments. These two fragments were then assembled with a PCR-amplified fragment and a p15A-cm vector by LLHR leading to the generation of p15A-cm-dis427 plasmid. To construct p15A-km-int-dis427 and p15A-km-int-Ptet-dis427, the *km-int* or *km-int-Ptet* cassette (Table S15) was amplified by PCR and electroporated into *E. coli* GB05-red harboring p15A-cm-dis427 and expressing Redαβ recombinase for linear-circular homologous recombination (LCHR)[42] to replace the *cm* cassette. Similarly, the expression plasmid p15A-km-int-Pvan-dis427 was constructed based on p15A-km-int-dis427 by LCHR using *apr-Pvan-disA* cassette. The p15A-km-int-Ptet-Papr-dis427 and p15A-km-int-Ptet-dis427-gent-delF were constructed based on p15A-km-int-Ptet-dis427 by LCHR using *Papr-disD* and *gent-delF* cassettes, respectively.

2.5.7 Electroporation of *M. xanthus* DK1622

M. xanthus DK1622 was inoculated into CTT medium (10 g/L casitone, 10 mM Tris-HCl, 8 mM magnesium sulphate, 1 mM potassium phosphate, pH=7.6) and incubated at 30 °C with shaking until exponential phase. For preparation of electro-competent cells, 1.8 mL culture was transferred into a 2-mL microcentrifuge tube with a hole punctured in the cap. Cells were centrifuged, washed twice, and finally resuspended in 50 µL autoclaved Milli-Q water. After addition of 1−3 µg plasmid DNA, the mixture was then transferred into a 1-mm cuvette. Electroporation was performed at 650 V, 400 Ω and 25 µF using a Bio-Rad Gene Pulser Xcell electroporation system. The pulsed cells were mixed with 1.6 mL fresh CTT medium and transferred back into the 2-mL microcentrifuge tube. After recovery at 30 °C for 6 h in an Eppendorf thermomixer, the cells were mixed with 10 mL CTT soft agar (0.6% agar, supplemented with 50 µg/mL kanamycin) and poured onto a CTT agar plate (1.2% agar, supplemented with 50 µg/mL kanamycin). The plate was incubated at 30 °C and colonies appeared after 4−7 days. Kanamycin-resistant colonies were picked out and inoculated into a new CTT agar plate supplemented with 50 µg/mL kanamycin. In order to verify intact integration of the biosynthetic gene cluster, three pairs of primers located at different positions of the gene cluster were used for colony PCR. Cells were scraped from the agar plate, suspended in 50 µL autoclaved Milli-Q water, and incubated at 100 °C for 20 min. After centrifugation, 1−2 µL of the supernatant was used as the PCR template.

2.5.8 Heterologous production of disorazole Z in *M. xanthus* DK1622

M. xanthus DK1622 mutants that integrated with disorazole biosynthetic gene cluster were inoculated in 1.6 mL CTT medium supplemented with 50 µg/mL kanamycin in a 2-mL microcentrifuge tube with a hole punctured in the cap and incubated at 30 °C in an Eppendorf thermomixer for 1 day. After that, 1 mL of the culture was inoculated into 50 mL CTT medium supplemented with 50 µg/mL kanamycin in a 300-mL baffled flask and incubated at 30 °C, 180 rpm for 2 days. After addition of inducer (0.5 µg/mL anhydrotetracycline when the Ptet promoter was used, 1 mM vanillate when the Pvan promoter was used) and 1 mL XAD-16 resin, incubation was continued for two more days. Cells and XAD-16 resin were collected by centrifugation and resuspended in methanol for extraction. After filtration, the extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-MS analysis using the method described above.

2.5.9 Isolation of disorazole Z from *M. xanthus* DK1622 mutants

For heterologous production of disorazole Z1 (**3**), the fully-grown culture of *M. xanthus* DK1622::km-int-Ptet-Papr-dis427 or *M. xanthus* DK1622::km-int-Pvan-dis427 was inoculated 1 to 100 into 2 L CTT medium supplemented with 50 µg/mL kanamycin in 5-L unbaffled flasks and cultivated at 30 °C, 180 rpm for 2 days. After addition of inducer (0.5 µg/mL anhydrotetracycline when the Ptet promoter was used, 1 mM vanillate when the Pvan promoter was used) and 2% XAD-16 resin, incubation was continued for three more days.

Cells and XAD-16 resin were collected by centrifugation, lyophilized to dryness, and extracted stepwise with methanol. The methanol extract was concentrated and partitioned with n-hexane to remove nonpolar impurity. After evaporation *in vacuo*, the methanol extract was then dissolved in Milli-Q water and extracted twice with equal portions of ethyl acetate. The ethyl acetate extract was evaporated to dryness, redissolved in methanol, and fractionated using Si-Flash chromatography on a Biotage system. Gradient was set as follow: 0−20 CV, hexane to ethyl acetate; 20−40 CV, ethyl acetate to MeOH; 40−45 CV, MeOH. Fractions containing targeted compounds were combined and further separated by semi-preparative HPLC on a Waters XSelect Peptide BEH C18 column (250×10 mm, 5 µm) using A (Milli-Q water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid) as mobile phase at a flow rate of 5 mL/min and column temperature of 45 °C. Gradient conditions were set as follows: 0−4 min, 5% B; 4−5 min, to 53% B; 5−20 min, 53% B; 20−21 min, to 95% B; 21−25 min, 95% B, 25−26 min, to 5% B; 26−30 min, 5% B. Fractions were detected by UV at 320 nm and were collected using an AFC-3000 fraction collector based on the retention time of 17.7 min. Heterologous production and purification of disorazole Z9 (**11**) and Z10 (**12**) were in a similar manner using the strain *M. xanthus* DK1622::km-int-Ptet-dis427-delF. For purification of **11** using semi-preparative HPLC, separation was achieved using a Waters XBridge Peptide BEH C18 column (250×10 mm, 5 µm) using the following gradient: 0−5 min, 5% B; 5−15 min, to 40% B; 15−25 min, 40% B; 25−26 min, to 95% B; 26−31 min, 95% B; 31−32 min, to 5% B; 32−35 min, 5% B. Fractions were detected by UV at 320 nm and were collected using an AFC-3000 fraction collector based on the retention time of 21.8 min. Similarly, **12** was purified using a modified gradient: 0−5 min, 5% B; 5−15 min, to 33% B; 15−50 min, 33% B; 50−51 min, to 95% B; 51−56 min, 95% B; 56−57 min, to 5% B; 57−60 min, 5% B. Fractions were detected by UV at 320 nm and were collected using an AFC-3000 fraction collector based on the retention time of 46.0 min.

2.5.10 Purification and *in vitro* reaction of the recombinant protein DisF$_{427}$

DNA fragment containing *disF$_{427}$* was amplified by PCR using So ce427 genomic DNA as template and oligos HisTEV-disF-F and HisTEV-disF-R. The plasmid pHis-TEV[50] was linearized by NcoI/XhoI and assembled with the PCR fragment by Gibson Assembly generating the protein expression construct pHisTEV-disF427. After Sanger sequencing, this construct was electroporated into *E. coli* BL21 (DE3). The resulting strain was cultivated in a 300-mL flask at 37 °C overnight in 50 mL LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH=7.0) supplemented with 50 µg/mL kanamycin. 20 mL overnight culture was used for inoculation of 2-L fresh LB medium supplemented with 50 µg/mL kanamycin in a 5-L flask. After cultivation at 37 °C until OD_{600} was about 0.6, the culture was cooled to 16 °C, induced with IPTG at the final concentration of 0.1 mM, and further cultivated at 16 °C overnight. Cells were harvested by centrifugation at 4 °C, resuspended in ice-cold lysis buffer (25 mM Tris, 200 mM NaCl, 10% glycerol, 20 mM imidazole, pH=8.0), and lysed using a Continuous Flow Cell Disrupter (Constant Systems) at 25000 psi and 4 °C. After centrifugation at 23500 rpm and 4 °C for 30 min, the cell debris was removed and the supernatant was loaded onto a 5-mL HisTrap HP column (GE healthcare) for nickel affinity chromatography using the ÄKTA protein purification system. Fractions containing recombinant protein in elution buffer (25 mM Tris, 200 mM NaCl, 10% glycerol, 250 mM imidazole, pH=8.0) were collected and loaded onto a HiPrep 26/10 desalting column (GE healthcare) to remove imidazole using desalting buffer (25 mM Tris, 200 mM NaCl, 10% glycerol, pH=8.0). The eluents were collected, concentrated, and stored at -80 °C.

The *in vitro* reaction was carried out at 30 °C for 1 h in a 50-µL mixture containing 25 mM Tris, 200 mM NaCl, 2 mM $MgCl_2$, 2 mM *S*-adenosyl methionine, 5 µL (about 1 mg/mL) recombinant protein, 0.5 µL disorazole Z9 (**11**) or Z10 (**12**) solution (1 mg/mL in MeOH). Boiled (100 °C for 10 min) recombinant protein was used for negative control. After addition of 50 µL MeOH, the mixture was vortexed and centrifuged at 15000 rpm for 15 min. 2 µL of the supernatant was used for HPLC-MS analysis.

2.5.11 Biological characterization

Half-inhibitory concentration (IC_{50}): Cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) or the American Type Culture Collection (ATCC) and were handled according to standard procedures as recommended by the depositor. Cells were seeded in 96-well plates and treated with disorazole A1 (**1**) and disorazole Z1 (**3**) in serial dilution for 48 h. Viability was determined by adding resazurin sodium salt for 3 h. Fluorescence measurements were performed using a SpectraMax T5 plate reader (Molecular Devices). Readouts were referenced and IC_{50} values were determined by sigmoidal curve fitting using OriginPro software. Data for **3** were determined as duplicates in three independent experiments.

Immunostaining and HC imaging: U-2 OS cells were seeded in 96-well imaging plates. After overnight equilibration, the cells were treated with **1** and **3** as assigned and incubated for up to 24 h. Cells were fixed with cold (-20 °C) acetone/MeOH (1:1) for 10 min. After washing with PBS, the cells were permeabilized with 0.01% Triton X-100 in PBS. The following primary antibodies (Sigma) were used: α-tubulin mAb, acetylated tubulin mAb, and GRP78/BiP mAb. For labeling, cells were incubated with primary antibody for 45 min at 37 °C, followed by incubation with the secondary antibody (Alexa Fluor® 488 goat anti-mouse or anti-rabbit; Molecular Probes) under the same conditions. After washing with PBS, the nuclear stain Hoechst33342 (5 µg/mL) was applied for 10 min. Samples were imaged on an automated microscope (BD Pathway 855) suitable for high-content screening. In case of GRP78 immunostaining, fluorescence intensity was determined in the cytoplasmic segment as defined by a ring around nuclei. The relative intensity of GRP78 fluorescence in this region of interest was used a measure for ER stress/unfolded protein response (UPR).

Live-cell imaging ER: U-2 OS cells were seeded in 96-well imaging plates and transfected using CellLight ER-GFP (Thermo Fisher Scientific) according to the manufacturer's protocol. After treatment with **1** and **3** in triplicates as indicated live cells were imaged on an automated microscope (BD Pathway 855).

Mitochondrial membrane potential (MMP): U-2 OS cells were seeded at 8×10^3 cells/well in 96-well imaging plates and were treated after overnight equilibration with **3** as assigned. Following treatment, the cells were washed twice with PBS and 100 µL of a staining solution (50 nM TMRM and 5 µg/mL Hoechst33342 in assay buffer) was added. The cells were stained for 30 min at 37 °C and after washing with assay buffer the samples were examined on an automated microscope (BD Pathway 855).

Determination of caspase-3/7 activity: HepG2 cells were treated with **3** as indicated and lysed in buffer containing 25 mM HEPES, 5 mM $MgCl_2$, 1 mM EGTA and 0.1% (v/v) Triton X-100. Protein lysates were mixed with substrate buffer containing 50 mM HEPES, 0.1% (w/v) CHAPS, 1% (w/v) sucrose, 0.2% (w/v) DTT and 0.05 mM Ac-DEVD-7-amino-4-trifluoro-methylcoumarin (AFC). Generation of free AFC was determined after 2 h at 37 °C by measurement of fluorescence using a SpectraMax T5 plate reader (Molecular Devices). For normalization, protein concentrations were determined by Bradford assay (absorption: 595 nm) using an Ultra Microplate Reader EL808 (BioTek Instruments). Data were determined in three independent experiments.

2.6 Supporting information

2.6.1 Structure elucidation of disorazole Z congeners

Structure elucidation of disorazole Z1: NMR based structure elucidation of disorazole Z1 is explained in the section of **2.3.1**. The crystal structure determination shows the constitution of the cyclic molecule with the relative configuration of the stereo centers C-12, C-25 and C-26 (Figure S1). In view of the relative configuration in the crystal, the stereo chemical NMR data, i.e. proton coupling constants and NOE correlations, were accessed critically. Indeed, with a dihedral angle of 179° the X-ray structure was not compatible with the strong NOE between methyl group C-30 and H-26 observed in solution and also not with the absence of a strong NOE between the methyl group C-30 and methine H-27 (Table S1).

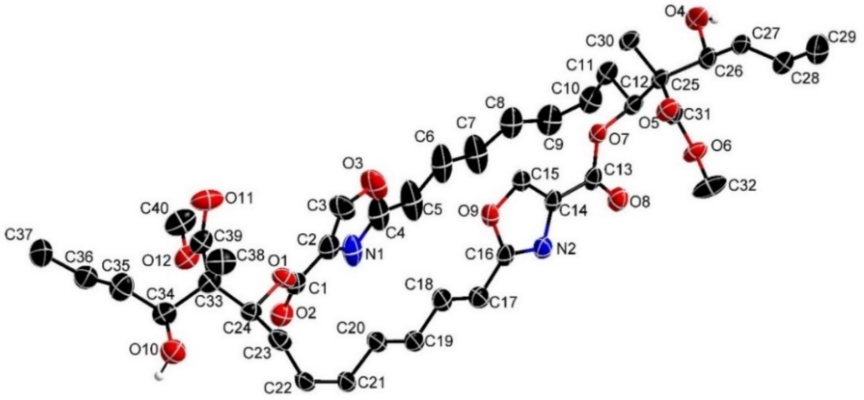
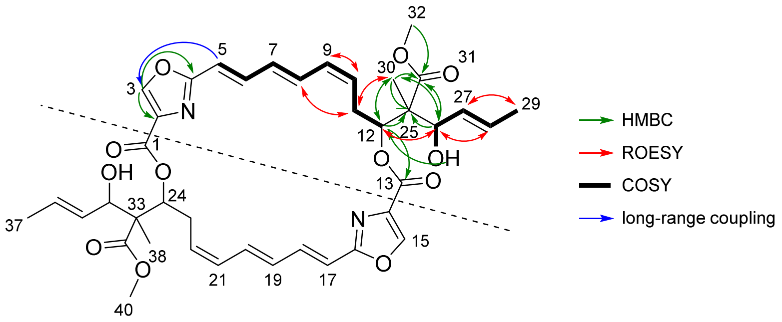


Figure S1 Crystal structure of **3**-$(EtOH)_2$. **3** = $C_{40}H_{46}N_2O_{12}$ in the crystal (displacement ellipsoids are drawn at 50% probability level). The C bonded hydrogen atoms are omitted for clarity. Both EtOH solvent molecules are squeezed by the program PLATON.

In the ^{13}C NMR spectra in acetone-d_6 a signal doubling of C-25 to C-27 revealed the presence of at least two rotamers in a ratio near 1:1 while the ^{1}H NMR signal of H-26 still appeared as a broad doublet after H/D exchange. Thus, the side chain conformation was explored using the "Conformational Search" module in HyperChem and finally calculated by PM3. The optimized conformation (Figure S2, side chain on the right side) with a dihedral angel of 66° between methyl group C-30 and H-26 now accounted for the observed NOEs (Table S1) by distances of 2.4 or 2.6 Å between methyl C-30 and H-26 or H-11, respectively. Contrary to the conformation in the crystal (Figure S1), the hydroxyl group is freely exposed in the predominant conformer in solution (Table S1, the NOE 30/26 was also the most intensive in DMSO-d_6 and CD_3OD).

Table S1 NMR data of disorazole Z1 (**3**) in acetone-d_6.[a]



Pos.	δ_H, m (J [Hz])	COSY[b]	NOESY[b]	δ_C, type	HMBC[b,c]
1/13	-	-	-	159.76, C	12
2/14	-	-	-	135.84, C	3
3/15	8.53, s	5	-	145.16, CH	(${}^1J_{H,C}$ 213 Hz)
4/16	-	-	-	162.45, C	3, 5
5/17	6.19, d (15.5)	3, 6	overlap with 7	116.61, CH	7
6/18	6.81, m br.	5, 7	overlap with 8	137.77, CH	7, 8
7/19	6.17, dd (14.9, 11.1)	6, 8	-	131.92, CH	5, 9
8/20	6.77, dd (14.5, 11.9)	7, 9	11	135.10, CH	9, 10
9/21	6.11, t (11.1)	8, 10	10	132.89, CH	7, 11
10/22	5.69, dt br. (9.5, 9.5)	11a, 11b, 9	9, 11, 12,	131.19, CH	11, 12
11a/23a	2.71, m	10, 11b, 12	8, 30	30.31, CH_2	9, 10, 12
11b/23b	2.68, m	10, 11a, 12	8, 10, 12, 26, 30	30.31, CH_2	9, 10, 12
12/24	5.45, dd (9.4, 1.5)	11a, 11b	10, 26, 30	76.28, CH	9, 11, 30
25/33	-	-	-	56.26, C[d] 56.24	OH, 12, 26, 30
26/34	4.45, dd br. (4.9, 7.6[e])	OH, 27	OH, 11, 12, 28, 30	75.65, CH[d] 75.52	OH, 27, 28, 29, 30
27/35	5.65, ddquin (15.4, 7.2, 1.3)	26, 28, 29	29	131.69, CH[d] 131.66	OH, 26, 29
28/36	5.74, dqd (15.4, 6.5, 0.9)	27, 29	26, 29	128.84, CH	26, 27, 29
29/37	1.70, dd (6.5, 0.9)	27, 28	27, 28	18.06, CH_3	27, 28
30/38	1.36, s	-	OH, 11, 26	13.66, CH_3	12, 26
31/39	-	-	-	173.92, C	12, 26, 30, 32
32/40	3.6, s	-	30	51.97, CH_3	-
26/34-OH	4.16, d (4.9)	26	26, 30	-	-

[a] ${}^1H/{}^{13}C$ at 700/150.9 MHz; [b] For overview, correlations are given for northern part only, although they are valid for the southern part, too; [c] Carbon showing HMBC correlations to indicated protons; [d] signal doubling ratio 1:1; [e] 7.2 Hz after H/D exchange.

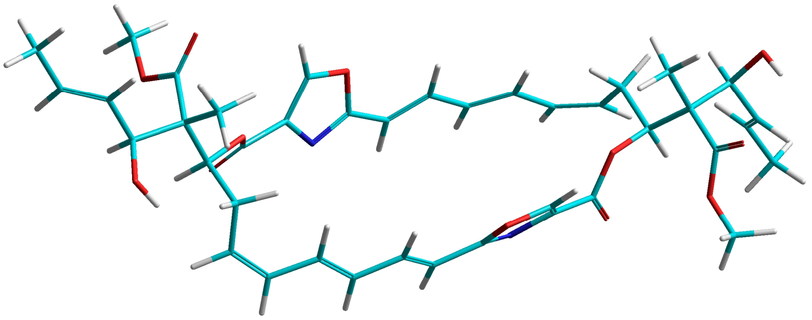


Figure S2 Side chain of disorazol Z1 (**3**) after conformational search and PM3 optimization (right side).

Since the X-ray analysis provided only the relative configuration of **3**, the absolute configuration was determined by chemical derivatization and NMR comparison of the α-methoxy-α-trifluoromethylphenyl acetic (MTPA) ester diastereomers, i.e. the application of the Mosher method based upon the ^{1}H and ^{13}C NMR chemical shift differences.[28-31] The (*S*)- and (*R*)-bis-methoxy-trifluoromethylphenyl acetic acid (MTPA) esters of **3** were easily prepared using an excess of the (*R*)- and (*S*)-MTPA chlorides in pyridine even without any catalyst.

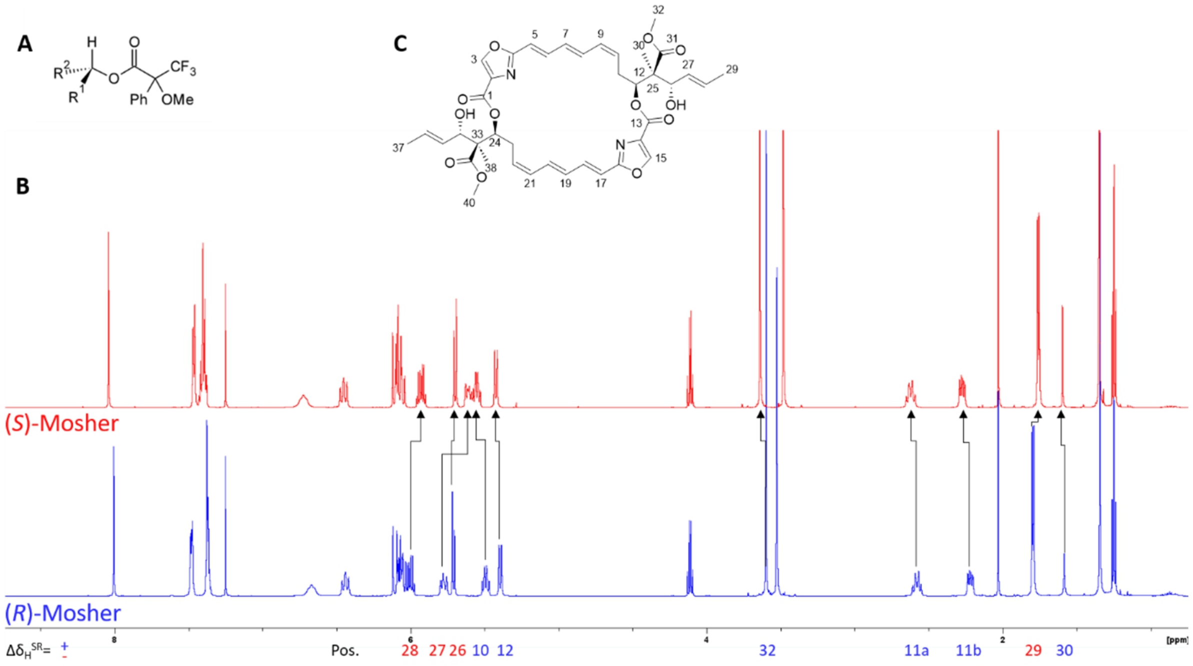


Figure S3 Mosher ester analysis of disorazole Z1 (**3**). A: Configuration of Mosher esters. B: ^{1}H NMR spectra and shift differences of (*S*)- and (*R*)-Mosher esters of **3**. C: Absolute configuration of **3**.

Defined by convention the signs of the $\Delta\delta^{SR} = \delta_S - \delta_R$ values for protons and carbons residing in R^1 will be positive and those in R^2 will be negative in Figure S3A. Both, the ^{1}H- and ^{13}C-NMR shift differences of the disorazole Z (*S*)- and (*R*)-Mosher esters (Table S2) unambiguously indicated that the negative $\Delta\delta^{SR}$ values are located at the end of side-chain (C-27 to C-29), which therefore was assigned as R^2. Positive values were observed at positions 10, 11, 12, 30, and 31, all residing in the molecular part R^1 (Figure S3B). Consequently, all asymmetric centers (12, 25, 26) of 3 are in the (*S*)-configuration as shown in (Figure S3C).

Table S2 ^{1}H- and ^{13}C-NMR data (^{1}H 600 MHz; ^{13}C 150 MHz) of the (*S*)- and (*R*)-Mosher esters of **3** in $CDCl_3$.

Pos.	δ_S	- δ_R	= $\Delta\delta^{SR}$	δ_{CS}	- δ_{CR}	= $\Delta\delta^{SR}$
10	5.57	5.51	+0.06	129.19	129.15	+0.04
11a	2.64	2.60	+0.04	29.47	29.25	+0.22
11b	2.29	2.24	+0.05			
12	5.44	5.41	+0.03	74.19	74.12	+0.07
25	-			53.42	53.54	-0.12
26	5.72	5.73	-0.01	80.02	80.05	-0.03
27	5.63	5.80	-0.17	123.86	123.92	-0.06
28	5.95	6.02	-0.07	135.32	135.98	-0.66
29	1.78	1.81	-0.03	17.95	18.03	-0.08
30	1.37	1.36	+0.01	16.29	15.85	+0.44
31	-			171.93	171.71	+0.22
32	3.66	3.62	+0.04	52.24	52.20	+0.04

Disorazole Z1 (**3**) isolated from the heterologous producer *M. xanthus* DK1622::km-int-Ptet-dis427 was characterized by ESI-HRMS and NMR, accordingly. Comparison of ^{1}H and ^{13}C NMR data of **3** isolated from both producers revealed shift differences <0.1 ppm ($\Delta\delta_H^b$-δ_H^c) and <0.3 ppm ($\Delta\delta_C^b$-δ_C^c). Thus, both compounds are identical (Table S3, Figure S17, and Figure S18).

Table S3 Comparison of ^{1}H and ^{13}C NMR data of **3** in acetone-d_6 isolated from *S. cellulosum* and *M. xanthus*.[a]

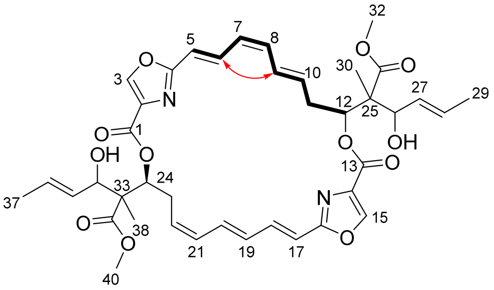
Pos.	δ_H[b]	δ_C[b]	δ_H[c]	δ_C[c]	$\Delta\delta_H$[b]-δ_H[c]	$\Delta\delta_C$[b]-δ_C[c]
1/13	-	159.76	-	159.60	-	0.16
2/14	-	135.84	-	135.69	-	0.15
3/15	8.53	145.16	8.56	145.16	0.03	0.00
4/16	-	162.45	-	162.30	-	0.15
5/17	6.19	116.61	6.17	116.49	0.02	0.12
6/18	6.81	137.77	6.80	137.60	0.01	0.17
7/19	6.17	131.92	6.15	131.78	0.02	0.14
8/20	6.77	135.10	6.78	135.02	0.01	0.08
9/21	6.11	132.89	6.10	132.74	0.01	0.15
10/22	5.69	131.19	5.68	131.17	0.01	0.02
11a/23a	2.71	30.31	2.68	30.17	0.03	0.14
11b/23b	2.68	30.31	-	-	-	-
12/24	5.45	76.28	5.45	76.14	0.00	0.14
25/33	-	56.26/ 56.24[d]	-	56.13	-	0.13/ 0.1[d]
26/34	4.45	75.65/ 75.52[d]	4.43	75.60	0.02	0.05/ 0.08[d]
27/35	5.65	131.69/ 131.66[d]	5.64	131.58	0.01	0.11/ 0.08[d]
28/36	5.74	128.84	5.74	128.77	0.00	0.07
29/37	1.70	18.06	1.70	18.00	0.00	0.06
30/38	1.36	13.66	1.36	13.51	0.00	0.15
31/39	-	173.92	-	173.81	-	0.11
32/40	3.61	51.97	3.61	51.88	0.00	0.09
26/34-OH	4.16	-	4.22	-	0.08	-

[a] ^{1}H/^{13}C at 700/175 MHz; [b] isolated from *S. cellulosum* So ce1875; [c] isolated from *M. xanthus* DK1622::km-int-Ptet-dis427; [d] signal doubling.

Structure elucidation of disorazole *cis*/*trans*-isomers Z2 and Z3:

In the analytical RP-HPLC two disorazole Z variants were observed eluting shortly in front and behind disorazole Z1 (**3**). Since ESI-HRMS provided the elemental formulae $C_{40}H_{46}N_2O_{12}$ similar to the mayor compound **3** for both, they were recognized as isomers. Their ^{1}H NMR spectra showed two well separated pairs of oxazole signals at about 8.5 ppm, which revealed the asymmetry of their structures. In the NMR data of disorazole Z2 (**4**) (Table S4) the signals of the side chains were nearly unchanged compared to **3** as well as the signals of the south part of the lactone ring (Table S1, Figure S11). In the ^{1}H,^{1}H COSY spectrum the full correlation sequence of H-5 to H-12 of the north part could be assigned unambiguously which characterized **4** as a $\Delta^{5,6}$-*trans*, $\Delta^{7,8}$-*cis*, $\Delta^{9,10}$-*trans* isomer. This stereo chemistry was indicated by the vicinal coupling constants of about 15 Hz for the *trans* double bonds and 11 Hz for the *cis* double bond. The $\Delta^{7,8}$-*cis* geometry was additionally shown by the strong NOE correlation between H-6 and H-9 in the ^{1}H,^{1}H ROESY spectrum.

Table S4 NMR data of disorazole Z2 (**4**) in acetone-d_6[a].

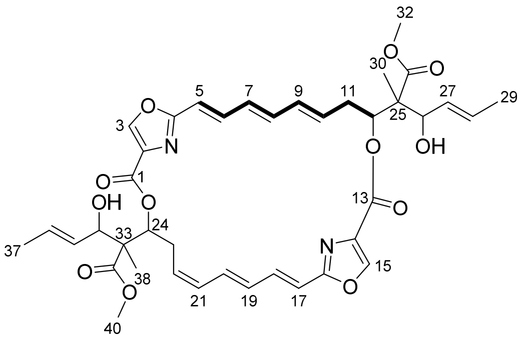


Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C, type	HMBC[b]
1	-	-	-	159.56, C	24
2[c]	-	-	-	135.59, C	3
3	8.58, s	5	18, 20	145.19, CH	-
4	-	-	-	162.13, C	3, 5, 6
5	6.24, d (15.4)	3, 6	7	117.98, CH	6, 7
6[c]	6.81, dd (15.4, 11.7)	5, 7	9	131.64, CH	5, 8
7	5.93, t (11.4)	6, 8	5, 8	127.42, CH	5, 6, 9
8	6.08, t (11.0)	7, 9	7, 10	135.13, CH	6, 9, 10
9	6.49, dd (14.9, 11.6)	8, 10, 11b	6, 11b	130.44, CH	7, 11ab
10	5.83, ddd (15.2, 9.5, 6.1)	9, 11ab	8, 11a, 12	134.98, CH	8, 11ab, 12
11a	2.79, m	10, 11b	10, 11b, 12, 26	35.06, CH$_2$	9, 10, 12
11b	2.31, dt (12.8, 10.6)	9, 10, 11a	9, 11a		
12	5.42, dd (11.2, 1.3)	11ab	10, 11a, 26, 30	76.25, CH	11ab, 26, 30
13	-	-	-	159.91, C	12
14[c]	-	-	-	135.86, C	15
15	8.64, s	17	6, 9	145.21, CH	- (213.7 Hz)
16	-	-	-	162.73, C	15, 17, 18
17	6.11, d (15.8)	15, 18	-	115.81, CH	18, 19
18	6.96, dd (15.8, 11.0)	17, 19	20	138.83, CH	17, 19, 20
19	6.13, dd (14.9, 11.4[g])	18	-	130.83, CH	17, 21
20	6.87, dd (14.9, 11.6)	21	18, 23	135.94, CH	18, 22
21	6.13, t (11.4[g])	20, 22	22	132.59, CH	19, 20, 23
22[d]	5.70, m	21, 23	21, 23, 24	131.52, CH	20, 21, 23, 24
23	2.71, m	22, 24	20, 24, 34, 38	30.11, CH$_2$	21, 22, 24
24	5.46, dd (9.9, 1.5)	23	22, 23, 34, 38	76.31, CH	23, 34, 38
25	-	-	-	55.88, C	11b, 12, 26, 26-OH, 30
26	4.45, m (6.6, 4.4)	27	11a, 12, 28, 30	75.22, CH	12, 27, 28, 30
26-OH	4.13, d (4.8)	-	-	-	-
27[d]	5.61, ddq (15.4, 7.0, 1.5)	26, 28, 36	12, 29, 30	131.63, CH	24, 28
28[e]	5.74, m	27, 29	26, 29	128.77, CH	26, 27, 29
29	2.06, m	28	27	18.04, CH$_3$	27, 28
30	1.35, m	-	11ab, 12, 26, 27	13.46, CH$_3$	12, 26
31	-	-	-	173.91, C	12, 26, 30, 32
32[f]	3.64, s	-	12, 27, 28	51.98, CH$_3$	12, 26
33	-	-	-	56.33, C	24, 34, 34-OH, 38
34	4.43, t (6.2)	34-OH	23, 24, 36, 38	75.64, CH	24, 35, 36, 38
34-OH	4.15, d (4.2)	34	-	-	-
35[d]	5.63, ddq (15.8, 7.0, 1.5)	-	37, 38	131.59, CH	36, 37
36[e]	5.74, m	37	34, 37	128.85, CH	35, 34, 37
37	1.70, d (5.5)	36	35, 36	18.04, CH$_3$	35, 36
38	1.69, d (5.5)	-	23, 24, 34, 35	13.57, CH$_3$	24, 34
39	-	-		173.91, C	24, 34, 38, 40
40[f]	3.60, s	-	24, 35	51.96, CH$_3$	-

[a] 1H/13C at 600/150 MHz; [b] Carbon showing HMBC correlations to indicated protons; [c-f] interchangeable ^{13}C assignments; [g] after H/D exchange.

In the ^{1}H NMR spectrum of $\Delta^{9,10}$-*trans*-disorazole Z (**5**) the clearly separated signals of 17-, 18-, and 20-H with coupling constants of about 15 Hz as well as the shifts of 23-Ha and -Hb suggested a carbon skeleton similar to disorazole Z1 (**3**) for the southern part (Table S1, Figure S11). Although the signals of 21- and 22-H of the double bond were hidden in multiplets, the NOE between 20-H and the methylene group protons at C-23 unambiguously indicated the expected $\Delta^{21,22}$-*cis* double bond geometry. In the northern part a ^{1}H,^{1}H COSY sequence of 5-H to 10-H could be established. While 10-H was completely overlapping with H-12 and H-22, the large coupling constants of about 15 Hz in the signals of H-5, -6, -8 and -9 suggested the all-*trans* geometry for the triene. The assignment of the $\Delta^{9,10}$-*trans* bond was supported by a strong NOE between 9-H and the 11-Hb proton of the methylene group, which was shifted high-field to 2.29 ppm.

Table S5 NMR data of disorazole Z3 (**5**) in methanol-d_4.[a]



Pos	δ_C	type	δ_H	m (*J* [Hz])	COSY	ROESY	H in HMBC
1	161.26	C					24
2	135.72	C					3
3	146.08	CH	8.32	s	5		-
4	163.00	C					3, 5, 6
5	117.49	CH	6.30	d (15.4)	3, 6	7	7
6	137.86	CH	6.81	dd (15.4, 11.1)	5, 7	8	5, 8
7	133.01	CH	6.03	br dd (14.4, 11.0)	6, 8	5[b]	5, 9[d]
8	137.35	CH	6.16	dd (14.8, 10.8)	7, 9	6, 10	6, 9
9	139.01	CH	6.06	br dd (15.3, 10.3)	8, 10	11b[b]	7, 8, 11ab
10	131.64	CH	5.56	br m	9, 11b	8, 11a	8, 11ab
11a	36.48	CH_2	2.71	m	11b, 12	10, 11b, 12, 26, 30	9, 10
11b			2.29	td (11.9, 11.2)	11a, 12	9, 11a, 30	
12	76.87	CH	5.57	dd (11.6, 3.2)	11ab	11a, 26, 30	11b, 26, 30
13	161.80	C					12
14	134.99	C					15
15	145.86	CH	8.43	s	17		-

Pos	δ_C	type	δ_H	m (*J* [Hz])	COSY	ROESY	H in HMBC
16	163.35	C					15, 17, 18
17	116.88	CH	6.14	d (15.3)	15, 18		19
18	139.10	CH	6.96	dd (15.4, 11.3)	17, 19	20	17, 19, 20
19	133.01	CH	6.05	dd (14.3, 11.3)	18, 20	[c]	17, 18[d], 21
20	135.06	CH	6.71	br dd (14.8, 11.4)	19, 21	18, 23b	18, 21, 22
21	133.57	CH	6.03	br t (10.3)	20, 22	22[c]	19, 20, 23b
22	131.78	CH	5.56	t (10.5)	21, 23a	21, 23a	20, 23ab, 24
23a	30.93	CH_2	2.72	br d (3.4)	22, 23b, 24	20, 22, 23b, 38	21, 22, 24
23b			2.59	br dd (13.7, 6.6)	22, 23a, 24	23a, 24, 34, 38	
24	78.01	CH	5.52	dd (10.9, 1.2)	23ab	23a, 34, 38	23ab, 34, 38
25	56.58	C					11b, 12, 26, 30
26	75.91	CH	4.32	d (7.5)	27	11a, 12, 27, 28, 30	27, 28, 29, 30
27	131.27	CH	5.60	ddq (15.3, 7.6, 1.5)	26, 28	26, 29, 30	26, 28, 29
28	130.41	CH	5.72	br m	27, 29	26, 29	27, 29
29	18.17	CH_3	1.71	dd (6.5, 1.3)	28	27, 28	27, 28
30	13.55	CH_3	1.33	s		11ab, 12, 26, 27, 32	12, 26
31	175.21	C					12, 26, 30, 32
32	52.47	OCH_3	3.67	s		30	-
33	56.72	C					24, 34, 38
34	77.32	CH	4.30	d (7.7)	35	23b, 24, 35, 36 38	24, 35, 36, 37, 38
35	131.45	CH	5.68	ddq (15.1, 7.7, 1.2)	34, 36	38, 37, 34	34, 36, 37
36	130.30	CH	5.74	m	37, 35	37, 34	34, 36, 37
37	18.14	CH_3	1.74	dd (6.2, 1.3)	36	35, 36	35, 36
38	13.41	CH_3	1.45	s		23ab, 34, 24, 35, 40	24, 34
39	175.11	C					24, 34, 38, 40
40	52.35	OCH_3	3.59	m		38	-

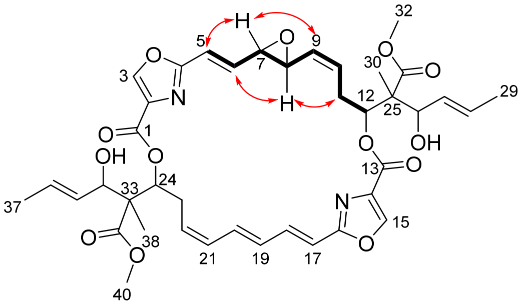
[a] ^{1}H 700 MHz, ^{13}C 175 MHz; [b] 7-H close to 9-H; [c] H-19 and H-21 overlap; [d] C-7 and C-19 overlap.

Structure elucidation of disorazole epoxides Z4, Z5 and Z6:

A group of three more polar disorazole Z variants with retention times of 15-16 min were also isomers which according to their ESI-HRMS-derived elemental formulas $C_{40}H_{46}N_2O_{13}$ contained one additional oxygen. While the NMR data of the mayor structural part of the main representative disorazole Z4 (**6**) (Table S6) were similar to disorazole Z1 (**3**) (Table S1), the COSY NMR spectrum presented a new sequence of correlations for the northern halve which connected completely visible ^{1}H signals between 5-H and 12-H. Thus, instead of a triene the isomer **6** contained a 7,8-epoxide (δ_C 60.4 and 56.5 ppm) flanked by a $\Delta^{5,6}$-*trans* and a $\Delta^{9,10}$-*cis* double bond. The proton signals of H-7 and H-8 (δ_H 3.25 and 3.76) had a small vicinal coupling constant of 1.7 Hz suggesting a *trans* configuration of the epoxide. This was supported by the strong NO effects between 7-H with 5-H and 9-H on one side and between

8-H and 6-H and 11-H_2 on the other side. Assuming the epoxide might be generated by an oxidation of **3** in a late stage of the biosynthesis, the epoxide ring should be directed to the outside of the bislactone ring and the absolute configuration of both epoxide carbons in **6** should be *R*.

Table S6 NMR data of disorazole Z4 (**6**) in methanol-d_4.^{a}



Pos	δ_C	type	δ_H	m (*J* [Hz])	COSY	ROESY	H in HMBC
1	160.61	C	-	-	-	-	24
2	135.59	C	-	-	-	-	3
3	145.92	CH	8.50	s	5	18, 20	
4	161.84	C		-	-	-	3, 5, 6
5	121.37	CH	6.46	d (15.8)	3, 6	7	6, 7
6	137.71	CH	6.06	dd (15.9, 9.4)	5, 7	7, 8, 15	5, 7, 8
7	60.42	CH	3.21	dd (9.4, 1.8)	6, 8	5, 6, 8, 9, 10	5, 9
8	56.49	CH	3.72	dd (9.7, 1.8)	7, 9	6, 7, 11, 15	6, 7, 10
9	131.08	CH	5.06	dd (10.7, 9.9)	8, 10	7, 8	7, 8, 11
10	134.04	CH	5.85	ddd (11.0, 10.0, 7.7)	9, 11	7, 11, 12	8, 11, 12
11	30.53	CH_2	2.66	m	10, 12	8, 10, 12, 26, 30	9, 10, 12
12	77.16	CH	5.47	dd (8.4, 4.0)		10, 11, 26, 30	9, 11, 26, 30
13	160.65	C	-	-	-	-	12
14	135.00	C	-	-	-	-	15
15	145.75	CH	8.48	s	-	6, 8, 26, 40	-
16	163.60	C	-	-	-	-	15, 17, 18
17	115.93	CH	6.18	br d (15.4)^{c}	18	^{c}	18, 19
18	139.36	CH	6.58	dd (15.6, 11.2)	17, 19	3, 20	19, 20
19	131.82^{b}	CH	6.19	br dd (14.7, 11.4)^{c}	18, 20	^{c}	17, 20, 22
20	136.56	CH	6.67	dd (15.0, 11.4)	19	3, 18, 23	18, 19, 22
21	133.29	CH	6.17	br t (11.0)^{c}	22	22^{c}	19, 23
22	131.82^{b}	CH	5.76	br t (9.5)	21, 23	21, 23, 24	23, 24
23	30.27	CH_2	2.63	m	22, 24	20, 22, 24, 34, 38	21, 22
24	77.77	CH	5.42	dd (8.4, 2.2)	23	22, 23, 34, 38	23, 34, 38
25	56.88	C	-	-	-	-	12, 26, 30
33	57.02	C	-	-	-	-	24, 34, 38
26, 34	76.56	CH	4.33	d (7.3)	27/35	11, 12, 23, 24, 27/35, 28/36, 30/38	12, 24, 28/36

Pos	δ_C	type	δ_H	m (*J* [Hz])	COSY	ROESY	H in HMBC
27, 35	131.39	CH	5.61	br ddq (15.4, 7.7, 1.60)	26/34, 28/36 29/37	26/34, 29/37, 32/40, 38/35	26/34, 28/36, 29/37
28, 36	130.34	CH	5.72	dq (15.3, 6.6)	27/35, 29/37	26/34, 29/37, 32/40, 38/35	26/34, 27/35, 29/37
29, 37	18.12	CH_3	1.72	dt (6.2, 1.8)	27/35, 28/36	27/35, 28/36	27/35, 28/36
30, 38	13.68	CH_3	1.37	s	-	11/23, 12, 24, 26/34, 27/35, 28/36	12, 24, 26/34,
30, 38	13.61	CH_3	1.35	s	-	-	-
31, 39	175.16	C	-	-	-	-	12, 24, 26/34, 30/38, 32/40
32, 40	52.46	CH_3	3.61	s	-	26/34, 27/35, 28/36	-
32, 40	52.45	CH_3	3.61	s	-	-	-

[a] $^1H/^{13}C$ = 600/150 MHZ; [b] double intensity; [c] H-17, -19, -21 overlapping multipletts.

This configuration was proven by X-ray analysis of the 7,8-epoxy-disorazole Z (**6**) after crystallization from ethanol. The crystals contained about two solvent molecules per one molecule of **6**, which are not presented in Figure S4. The X-ray analysis provided a relative configuration shown in Figure S4.

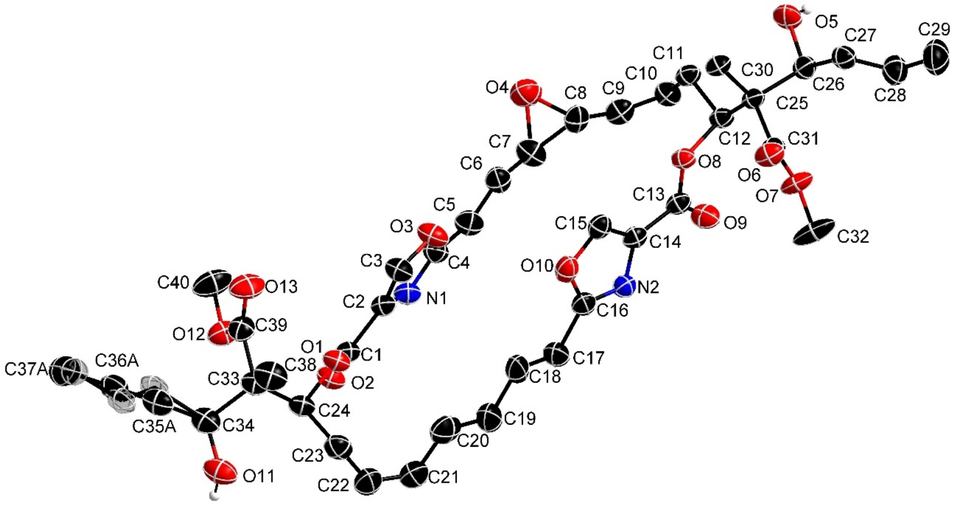
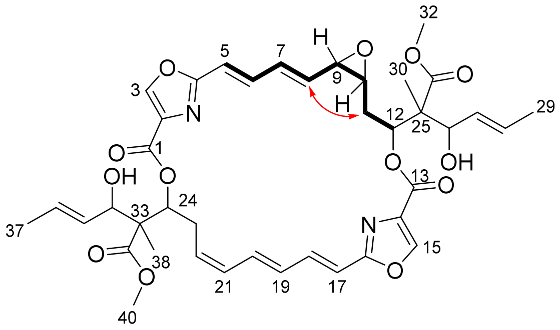


Figure S4 Crystal structure of **6**-(EtOH)$_2$. **6** = $C_{40}H_{46}N_2O_{13}$ in the crystal (displacement ellipsoids are drawn at 50% probability level). The C bonded hydrogen atoms and both solvent ethanol molecules are omitted for clarity. The disordered atoms are highlighted with transparency.

The second isomer with the elemental formula $C_{40}H_{46}N_2O_{13}$ was easily recognized as 9,10-epoxide disorazole Z (**7**), because the methylene protons H-11a at 2.76 ppm and the conspicuously high-field shifted H-11b at 1.50 ppm showed COSY correlation with the epoxide proton H-10 (3.27 ppm) (Table S7). The vicinal coupling $J_{9,10}$ of 4 Hz was large for epoxides and indicated a *cis* geometry. As proof, a strong NOE between H-8 and 11-Hb was observed, very similar to the NOE between H-20 and the methylene group protons H-23ab, which characterized the *cis* configuration of the $\Delta^{21,22}$- double bond in the southern part of **7**. On the other side of the *cis*-epoxide three clearly visible dd-signals of H-8, H-7 and H-6, which included a vicinal coupling of about 15 Hz, showed the *trans* configuration of the diene in the northern part of **7**. The relative configuration of disorazole Z5 (**7**) was unambiguously shown by the X-ray structure of **7** after crystallization from ethanol (Figure S5).

Table S7 NMR data of disorazole Z5 (**7**) in methanol-d_4.[a]



Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C	H in HMBC
1	-	-	-	160.66, C	24
2	-	-	-	135.58, C	-
3	8.42, br s	-	18, 20, 26/34, 30/38, 32/40	145.52, CH	-
2	-	-	-	163.00, C	3, 5, 10
5	6.14, d (15.8)	10	7	117.87, CH	7, 10
6	6.72, dd (15.4, 11.0)	5, 7	8	136.82, CH	7, 8
7	6.41, dd (15.0, 11.0)	8, 10	5, 9	136.16, CH	9[b], 17/5, 18, 22
8	5.68, dd (15.0, 9.9)	7, 9	10, 11b	135.46, CH	9, 10
9	3.34, dd (9.9, 4.2)	8, 10	7	58.91, CH	7, 8, 9, 11ab, 12
10	3.27, dt (10.5, 3.8)	9, 11a/11b	11a, 12	58.91, CH	7, 8, 9, 11ab, 12
11a	2.76, dd (14.7, 3.7)	10, 11b	10, 11b, 12, 26	32.11, CH_2	10, 12
11b	1.50, dt (14.8, 10.6)	10, 11a, 12	8, 11a		
12	5.57, d (10.6)	11b	10, 11a, 26/34	74.69, CH	12, 26, 30
13	-	-	-	160.98, C	12
14	-		-	135.85, C	-
15	8.37, s		8, 10, 26/34, 30/38, 32/40	145.63, CH	-
16	-	-	-	163.56, C	15, 17, 18
17[c]	6.15, d (15.4)	18	19[c]	116.28, CH	18, 19[c]
18	6.64, dd (15.6, 10.8)	17, 19c	-	139.06, CH	19[c], 20, 21
19[c]	6.16, dd (11.0, 15.0)	20, 22	17[c], 21	132.10, CH	18/20, 21[c]
20	6.61, dd (11.2, 6.4)	18, 19, 21[c]	23ab	136.22, CH	9[b], 17/5, 18, 22

Pos.	δ_H, m (J [Hz])	COSY	ROESY	δ_C	H in HMBC
21[c]	6.15, t (11.0)	18, 20	19[c], 22	133.36, CH	19[c], 23ab
22	5.72, m	19, 23ab	21[c], 23ab	131.69, CH	20, 23ab, 24
23	2.62, m	22, 24	20, 22, 24, 26/34	30.37, CH_2	21[c], 22, 24
24	5.41, dd (6.4, 4.6)	23ab	22, 23ab, 26/34, 27/35,	77.52, CH	23ab, 34, 38
25/33	-	-	-	56.90, C	12/24, 26/34, 30/38
25/33	-	-	-	56.86, C	12/24, 26/34, 30/38
26/34	4.32, d (7.3)	27/35	23ab, 24, 28/36, 30/38	76.70, CH	27/35, 28/36, 30/38
26/34	4.33, d (7.3)	27/35	11a, 12, 28/36, 30/38	76.51, CH	27/35, 28/36, 30/38
27/35	5.61, ddt (15.2, 7.9, 1.7)	26/34, 28/36, 29/37	24, 29/37	131.42, CH	28/36, 29/37
27/35	5.61, ddt (15.2, 7.9, 1.7)			131.37, CH	28/36, 29/37
28/36	5.74, m	27/35, 29/37	26/34, 29/37	130.42, CH	10, 26/34, 27/35, 29/37
28/36	5.74, m			130.31, CH	10, 26/34, 27/35
29/37	1.73, br s	27/35, 28/36	27/35	18.12, CH_3	27/35, 28/36
30/38	1.37, s	-	23ab, 24, 26/34, 27/35, 28/36	13.67, CH_3	12/24, 26/34
30/38	1.35, s	-	11ab, 12, 26/34, 27/35, 28/36	13.50, CH_3	12/24, 26/34
31/39	-	-	-	175.18, C	12/24, 26/34, 30/38, 32/40
31/39	-	-	-	175.04, C	12/24, 26/34, 30/38, 32/40
32/40	3.62, s	-	27/35, 28/36	52.49, CH	-
32/40	3.59, s	-	27/35, 28/36	52.44, CH	-

[a] ^{1}H 600 MHz, ^{13}C 150 MHz; Because the ^{1}H signals of the side chains closely overlap, they are assigned pairwise, and direct CH assignments may be interchanged; [b] C-20/C-18 overlap; [c] H-17, -19, -21 overlap.

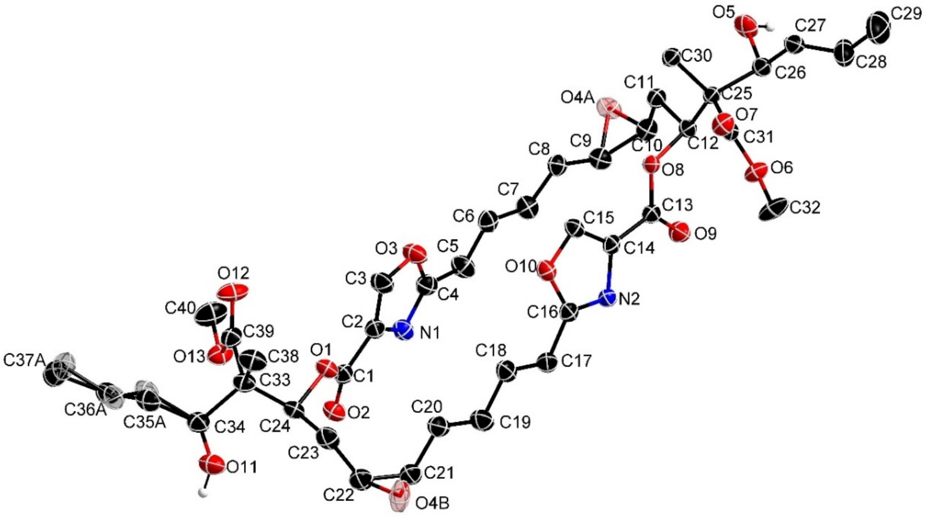


Figure S5 Crystal structure of **7**-(EtOH)$_2$. **7** = $C_{40}H_{46}N_2O_{13}$ in the crystal (displacement ellipsoids are drawn at 50% probability level). The C bonded hydrogen atoms and both solvent ethanol molecules are omitted for clarity. The disordered atoms are highlighted with transparency.

The NMR data of the third isomer disorazole Z6 (**8**) with the elemental formula $C_{40}H_{46}N_2O_{13}$ were nearly identical to disorazole Z1 (**3**) except those of one side chain. There, the former $\Delta^{27,28}$ double bond was replaced by an epoxide with characteristic ^{1}H (δ_H 2.87 and 2.92 ppm) and ^{13}C NMR shifts (δ_C 61.34 and 54.54 ppm) (Table S8). The vicinal ^{1}H coupling constant $J_{27,28}$ = 2.1 Hz was small and thus suggested the *trans* configuration of the epoxide. This finding was supported by the ROESY correlations observed between methyl group (C-29) and H-27 and between H-28 and H-26. Considering, that the oxidation of the former double bond would occur from the least hindered side, i.e. opposite to the hydroxyl group, the absolute configuration of the epoxide **8** was suggested as 27*R*,28*R* (Figure S6).

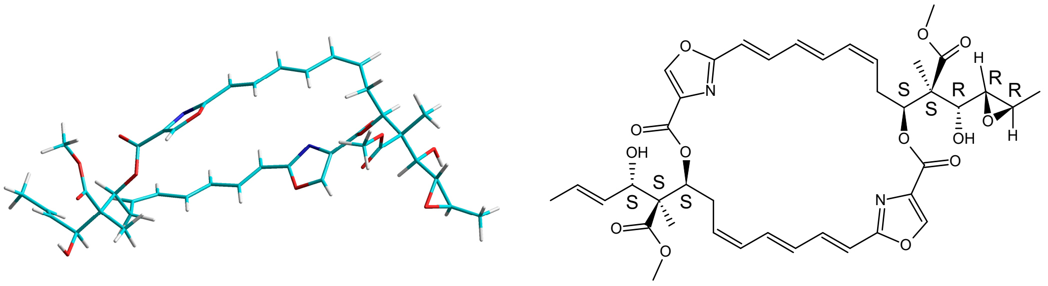
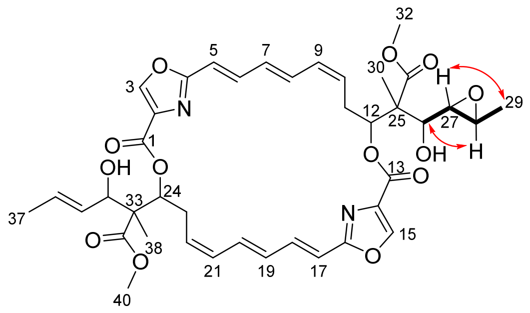


Figure S6 Model and structure of disorazole Z6 (**8**) with 25*S*,26*R*,27*R*,28*R*-configuration.

Table S8 NMR data of disorazole Z6 (**8**) in methanol-d_4.[a]



Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C, type	H in HMBC
1/13	-	-	-	161.03, C	12/24
2/14	-	-	-	136.00, C	3/15, 5/17
2/14	-	-	-	135.92, C	3/15, 5/17
3	8.48, s	-	34, 40, 38	145.94, CH	-[c]
15	8.49, s	-	27, 30, 32	146.05, CH	-[c]
4/16	-	-	-	163.88, C	n.a.
4/16	-	-	-	163.85, C	n.a.
5/17	6.13, m	6/18	n.a.	116.59, CH	n.a.
5/17	-	-	n.a.	116.52, CH	n.a.
6/18	6.79, m	5/17	n.a.	139.41, CH	n.a.
6/18	-	-	n.a.	139.47, CH	n.a.

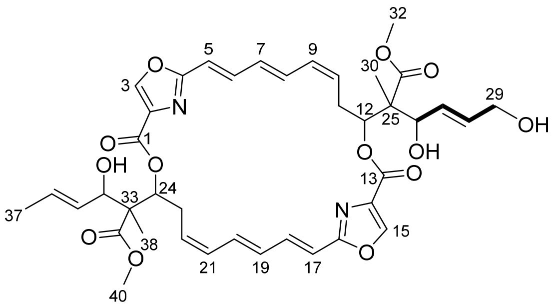
Pos.	δ_H, m (*J* [Hz])	COSY	ROESY	δ_C, type	H in HMBC
7/19	6.13, m		n.a.	132.58, CH	n.a.
7/19	-	-	n.a.	132.49, CH	n.a.
8/20	6.68, m	7/19, 9/21	n.a.	136.29, CH	n.a.
8/20	-	-	n.a.	136.34, CH	n.a.
9/21	6.13, m		n.a.	133.79, CH	n.a.
9/21	-	-	n.a.	133.63, CH	n.a.
10/22	5.71, m		n.a.	131.99, CH	12/24
10/22	-	-	-	131.72, CH	12/24
11a	2.70, m		30	31.13, CH_2	7/19, 10/22, 12
11b	2.64, m	10/22, 12	26, 30	-	-
12	5.46, d (10.1)	11	10, 11b, 26[b], 27, 30	77.36, CH	26
25	-	-	-	56.39, C	12, 26, 30
26	3.70, d (5.9)	27	11b[b], 12, 27, 28, 30	75.24, CH	12, 27, 30
27	2.87, dd (5.9, 2.1)	26, 28	15, 26, 29, 30	61.34, CH	26, 29
28	2.92, qd (5.9, 2.1)	27, 29	26[b], 29	54.54, CH	29
29	1.29, d (5.0)	28	26[b], 27, 28	17.86, CH_3	28
30	1.45, s	-	11ab, 12, 26, 27	14.41, CH_3	12, 26
31	-	-	-	175.21, C	12, 26, 30, 32/26
32	3.69, m	-	11b[b], 12, 27, 28, 30	52.99, CH_3	-
23	2.61, m	10/22, 24	24, 34, 38	30.79, CH_2	9/21, 24
24	5.44, d (10.2)	23	22, 23, 34, 38	77.70, CH	23, 34, 38
33	-	-	-	57.18, C	24, 34, 38
34	4.33, d (7.6)	35	3, 23, 24, 35, 36, 38	76.92, CH	24, 30/38
35	5.63, m		34, 37, 38	131.81, CH	36, 37
36	5.74, m	37	34, 37, 38, 40	130.59, CH	34, 35, 37
37	1.73, dd (6.3, 1.3)	10/22, 35, 36	35, 36	18.43, CH_3	35, 36
38	1.39, s	-	3, 23, 24, 34, 35	13.98, CH_3	24, 34
39	-		-	175.46, C	24, 34, 38, 40
40	3.69, s	-	3, 34, 35, 36	52.71, CH_3	-

[a] ^1H 600 MHz, ^{13}C 150 MHz; [b] methoxy-32 and H-26 ^1H signals overlap, however methoxy ROESY correlation signals are expected as very small, like those of methoxy group C-40; [c] carbon signals may be interchanged; n.a. not analyzed due to ^1H signal overlap.

Structure elucidation of disorazole carbon skeleton variants Z7-Z10:

Considerably more polar (t_R = 8.96 min), a fourth isomer disorazole Z7 (**9**) of the elemental formula $C_{40}H_{46}N_2O_{13}$ was identified by HPLC-UV-ESI-HRMS. The NMR data of the isolated compound again only differed from disorazole Z1 (**3**) (Table S1, Figure S11) in one side chain, which instead of a methyl group ended in a primary unsaturated alcohol (δ_H 4.11, δ_C 63.0).

Table S9 NMR data disorazole Z7 (**9**) in methanol-d_4.[a]

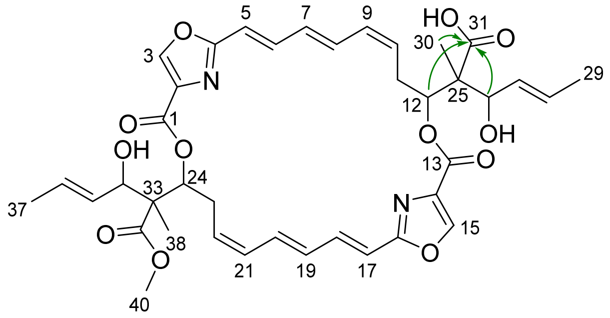


Pos.	δ_H, m (*J* [Hz])	COSY	δ_C, type	H in HMBC
1/13	-	-	160.82, C	12[b]
1/13	-	-	160.75, C	12[b]
2/14	-	-	135.70, C	-
2/14	-	-	135.68, C	-
3/15	8.52, s	-	145.75, CH	-
3/15	8.48, s	-	145.65, CH	-
4/16	-	-	163.57, C	3, 5[b]
5/17	6.12, d (15.0)	6/18	116.25, CH	7[b]
6/18	6.79, t br[d] (12.5)	5/17	139.15, CH	5, 7, 8[b]
7/19	6.13, dd (15.0, 9.5)	8/20	132.20, CH	9[b]
8/20	6.68, t br[e] (12.7)	7/19	136.06, CH	9, 10[b]
9/21	6.13, t (10.6)	10/22	133.33, CH	7, 11[b]
10/22	5.70, m	9/21	131.75, CH	8, 10, 11[b]
10/22	-	-	131.70, CH	8, 10, 11[b]
11/23	2.65, m	10/22, 12/24	30.65, CH$_2$	9, 10, 12[b]
11/23	-	-	30.51, CH$_2$	9, 10, 12[b]
12/24	5.45, d br (9.1)	11/23	77.43, CH	9[b], 11, 26, 30, 34
12/24	5.43, d br (9.1)	11/23	77.34, CH	9[b], 11, 26, 30, 34
25/33	-	-	56.89, C	12[b], 25[b], 26, 34
25/33	-	-	56.75, C	12[b], 25[b], 26, 34
26	4.45, d (5.5)	27, 28	75.87, CH	12, 27, 28
27	5.86, m (15.7, 5.4)[c]	26, 28	130.49, CH	26, 28, 29
28	5.86, m (15.3, 3.8)[c]	26, 29	134.18, CH	26, 27, 29
29	4.11, d (3.7)	26, 27	63.00, CH$_2$	27, 28
30	1.39, s	-	14.01, CH$_3$	12[b], 26
34	4.33, d (7.7)	35	76.66, CH	24, 35, 36, 38
35	5.63, ddq (15.4, 7.7, 1.5)	34, 36, 37	131.43, CH	34, 36, 37
36	5.74, dq (15.4, 6.2)	35, 37	130.34, CH	34, 35, 37
37	1.73, dd (6.2, 1.1)	35, 36	18.14, CH$_3$	35, 36
38	1.38, s	-	13.69, CH$_3$	12[b], 34
31/39	-	-	175.17, C	12[b], 26, 30[b], 32[b], 34
32/40	3.66, s	-	52.55, CH$_3$	-
32/40	3.62, s	-	52.44, CH$_3$	-

[a] ^{1}H 600 MHz, ^{13}C 150 MHz; [b] correlations are given for the northern part only, although they are valid for the southern part too; [c] from *J*-resolved NMR spectrum; [d] not observed in *J*-resolved NMR spectrum; [e] from *J*-resolved NMR spectrum 11.0, 15.6 Hz.

With a RP-HPLC retention time of 9.38 min, disorazole Z8 (**10**) was the second most polar variant. The elemental formula $C_{39}H_{44}N_2O_{12}$ showed a loss of a methylene compared to disorazole Z1 (**3**). The NMR data of **10** (Table S10) were most similar to **3** for the bislactone ring and one side chain (C-33 - C-40) (Table S1, Figure S11). However, the ^{1}H and ^{13}C spectra contained only one methoxy signal which correlated in the HMBC spectrum with the ester carbon C-39 (δ_C 175.17). The respective carbon C-31 (δ_C 176.66) of the second side chain was assigned as carboxylic acid, which only had HMBC correlations with the oxymethines C-12 and C-26 and methyl group C-30.

Table S10 NMR data of disorazole Z8 (**10**) in methanol-*d*4.[a]



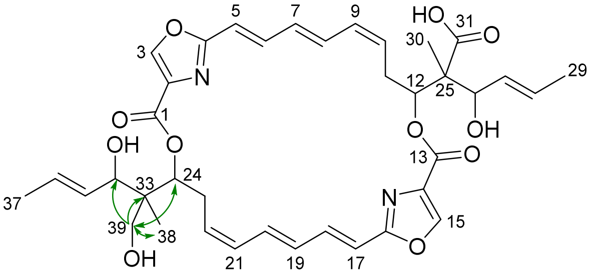
H	δ_H	m	*J*[Hz]	COSY	δ_C, type	H in HMBC
1/13	-			-	160.85, C	12/24
1/13	-			-	160.72, C	12/24
2/14	-			-	135.90, C	3/15
2/14	-			-	135.68, C	3/15
3/15	8.52	s	-	-	145.62, CH	-
3/15	8.51	s	-	-	145.56, CH	-
4/16	-			-	163.57, C	3/15, 5/17
4/16	-			-	163.49, C	3/15, 5/17
5/17	6.16	d	15.0	6/18	116.28, CH	7/19
5/17	-			6/18	116.20, CH	7/19
6/18	6.81	t	11.7 br.	5/17, 7/19	139.07, CH	7/19, 8/20
6/18	-			5/17, 7/19	139.14, CH	7/19, 8/20
7/19	6.17	dd	15.0, 11.0	6/18, 8/20	133.34, CH	5/17, 8/20, 9/21
7/19	-			6/18, 8/20	133.29, CH	5/17, 8/20, 9/21
8/20	6.72	dt	14.7, 11.7	7/19, 9/21	136.13, CH	10/22
8/20	-			7/19, 9/21	136.00, CH	10/22
9/21	6.16	t	11.0	8/20, 10/22	132.11, CH	5/17, 7/19, 11b/23b
9/21	-			8/20, 10/22	132.19, CH	5/17, 7/19, 11b/23b
10/22	5.74	m[b]	-	9/21, 11/23	131.95, CH	8/20, 11b/23b, 12/24
10/22	-		-	-	131.67, CH	8/20, 11b/23b, 12/24
11/23	2.73	m	-	10/22, 12/24	30.71, CH_2	7/19, 10/22

H	δ_H	m	J[Hz]	COSY	δ_C, type	H in HMBC
11/23	2.66	m	-	10/22, 12/24	30.50, CH_2	7/19, 10/22
12/24	5.47	d	10.3	11/23	77.41, CH	11ab/23ab, 30, 38
12/24	-			-	77.38, CH	11ab/23ab, 30, 38
25	-			-	56.31, C	26, 30
33	-			-	56.89, C	34, 38
26	4.41	d	7.3	27	76.34, CH	28, 30
34	4.37	d	7.7	35	76.64, CH	36, 38
35	5.67	ddq	15.4, 7.7, 1.6	34, 36, 37	131.59, CH	26/34, 28/36, 29/37
27	5.72	m[b]		26, 28	131.43, CH	26/34, 28/36, 29/37
28	5.79	m[b]		27, 29	129.99, CH	26, 29/37
36	5.79	m[b]		35, 37	130.33, CH	29/37, 34
29/37	1.77	d	6.6	28, 35, 36	18.16, CH_3	27/35, 28/36
30	1.38	s	-	-	13.91, CH_3	12/24, 26
31	-			-	176.66, C	12/24, 26, 30
38	1.42	s	-	-	13.69, CH_3	12/24, 34
39	-			-	175.17, C	12/24, 34, 38, 40
40	3.66	s	-	-	52.45, CH_3	-

[a] ^1H 600 MHz, ^{13}C 150 MHz; [b] with 27, 35, 36, 10/22-H. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z8 (**10**) misses CH_3-32.

The variant disorazole Z9 (31-*O*-desmethyl-39-hydroxy-disorazole Z) (**11**) with a retention time of 9.35 min had the elemental formula $C_{38}H_{44}N_2O_{11}$, which showed the formal loss of C_2H_2O. The absence of any methoxy group in the NMR data of **11** (Table S11) instantly indicated that both side chains were affected. The data of the lactone core were comparable to disorazole Z1 (**3**), although the symmetry of the molecule and the complete overlap of the respective NMR signals were lost (Table S11, Table S1, Figure S11). In the side chain of the northern halve a free carboxylic acid at C-31 (176.78 ppm) comparable to the mono ester of **10** was assigned from the NMR data. A second carboxyl carbon was also absent in the ^{13}C NMR spectrum of **11**. Instead, a new primary alcohol C-39 ($\delta_{H/C}$ 3.61/66.34 ppm) was assigned to its position from the HMBC correlations observed with lactone C-24, quaternary carbon C-33, secondary alcohol C-34, and methyl group C-38. The exchange of a carboxyl substituent against a primary alcohol group resulted in a high-field shift of 10 ppm for the quaternary carbon C-33 (δ_C 46.91 ppm).

Table S11 NMR data of disorazole Z9 (11) in methanol-d_4.[a]



H	δ_H	m	J [Hz]	ROESY	δ_C, type	H in HMBC
1	-		-	-	161.70, C	24
2/14	-		-	-	136.10, C	3/15
2/14	-		-	-	135.95, C	3/15
3	8.53	s	-	34, 39	145.58, CH	-
4/16	-		-	-	163.61, C	3/15, 5/17
4/16	-		-	-	163.52, C	3/15, 5/17
5/17	6.13	m	-	-	116.20, CH	7/19
5/17	6.13	m	-	-	115.99, CH	7/19
6/18	6.84	m	-	-	139.31, CH	7/19, 8/20
6/18	6.84	m	-	-	139.17, CH	7/19, 8/20
7/19	6.13	m	-	-	133.27, CH	[b]
7/19	6.13	m	-	-	132.95, CH	[b]
8/20	6.78	dd	14.5, 11.9	-	136.45, CH	[b]
8/20	6.71	dd	14.3, 12.1	-	136.07, CH	[b]
9/21	6.13	m	-	-	132.78[b], CH	[b]
10/22	5.7	m	-	-	132.12[b], CH	[b]
27/35	5.7	m	-	-	132.05[b], CH 131.95[b], CH 131.88[b], CH 131.63[b], CH	[b] [b] [b] [b]
28/36	5.7	m	-	-	129.62, CH	26/34
28/36	5.7	m	-	-	129.96, CH	26/34
29/37	1.75	br d	6	-	18.18, CH_3	[b]
29/37	1.74	br d	6.6	-	18.15, CH_3	[b]
11a	2.72	br dt	13.7, 10.4	11b, 30	30.76, CH_2	12
11b	2.62	dd	13.9, 7.3	11a, 12, 26, 30		12
12	5.45	d	10.3	11b, 26, 30	77.41, CH	11ab, 26, 30
	-			-	56.29, C	12, 26, 30
26	4.37	d	7.0	11, 12, 30	76.42, CH	12, 30
30	1.35	s	-	11ab, 12, 26	14.06, CH_3	12, 26
31	-		-	-	176.78, C	12, 26, 30
13	-		-	-	161.03, C	12
15	8.47	s	-	26	145.58, CH	-
23a	2.88	br dt	13.4, 10.5	23b	30.40, CH_2	24
23b	2.53	dd	13.9, 7.3	23a, 24, 34, 38, 39		24
24	5.28	d	10.3	23b, 34, 38, 39	77.52, CH	23ab, 34, 38, 39
33	-			-	46.91, C	24, 34, 38, 39
34	4.13	d	6.6	23b, 24, 38, 39	76.90, CH	24, 38, 39
38	1.02	s	-	23b, 24, 34, 39	15.53, CH_3	24, 34, 39
39	3.61	m	-	23b, 24, 34, 38	66.34, CH_2	24, 34, 38

[a] ^{1}H/^{13}C at 600/150 MHz; [b] overlapping signals. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z9 (**11**) misses CH_3-32 and CH_3-40.

Comparison of ^{1}H and ^{13}C NMR data of disorazol Z9 (**11**) isolated from *S. cellulosum* So ce1875 and *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF revealed a shift differences <0.1 ppm ($\Delta\delta_H{}^b$-$\delta_H{}^c$). Thus, both compounds are identical (Table S12, Figure S55). Greater differences of carbon shifts can be explained due to their extraction from HMBC experiments.

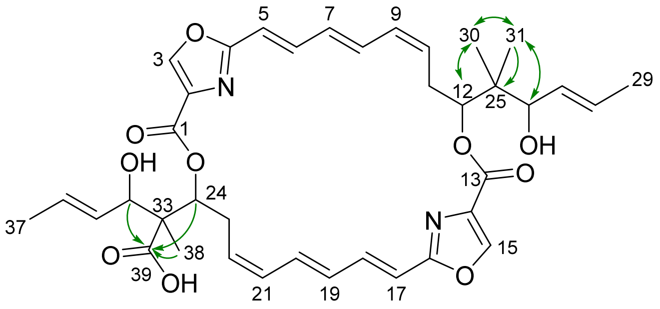
Table S12 Comparison of ^{1}H and ^{13}C NMR data of **11** in methanol-d_4 isolated from *S. cellulosum* and *M. xanthus*.[a]

Pos.	$\delta_H{}^b$	$\delta_C{}^b$	$\delta_H{}^c$	$\delta_C{}^c$	$\Delta\delta_H{}^b$-$\delta_H{}^c$	$\Delta\delta_C{}^b$-$\delta_C{}^c$
2/14	-	136.1	-	136.02	-	0.08
2/14	-	135.95	-	136.02	-	0.07
4/16	-	163.61	-	163.53	-	0.08
4/16	-	163.52	-	163.53	-	0.01
5/17	6.13	116.2	6.12	116.02	0.01	0.18
5/17	6.13	115.99	6.12	116.02	0.01	0.03
6/18	6.84	139.31	6.82	139.21	0.02	0.10
6/18	6.84	139.17	6.82	139.21	0.02	0.04
7/19	6.13	133.27	6.12	133.17	0.01	0.10
7/19	6.13	132.95	6.12	133.17	0.01	0.22
8/20	6.78	136.45	6.78	136.48	0.00	0.03
8/20	6.71	136.07	6.73	136.36	0.02	0.29
9/21	6.13	132.78[d]	6.12	132.05	0.01	[d]
10/22	5.7	132.12[d]	5.72	131.97	0.02	[d]
27/35	5.7	132.05[d] 131.95[d] 131.88[d] 131.63[d]	5.72	131.97	0.02	[d] [d] [d] [d]
28/36	5.7	129.62	5.72	129.43	0.02	0.19
28/36	5.7	129.96	5.72	129.43	0.02	0.53
29/37	1.75	18.18	1.74	18.24	0.01	0.06
29/37	1.74	18.15	1.74	18.24	0.00	0.09
11a	2.72	30.76	2.75	31.04	0.03	0.28
11b	2.62	-	2.60	-	0.02	-
12	5.45 -	77.41 56.29	5.40 -	77.53 56.09	0.05 -	0.12 0.20
26	4.37	76.42	4.38	76.69	0.01	0.27
30	1.35	14.06	1.29	15.21	0.06	1.15
31	- -	176.78 161.03	- -	178.19 161.71	- -	1.41 0.68
15	8.47	145.58	8.57	145.74	0.10	0.16
23a	2.88	30.4	2.87	30.37	0.01	0.03
23b	2.53	-	2.53	-	0.00	-
24	5.28 -	77.52 46.91	5.28	77.43 46.89	0.00	0.09 0.02
34	4.13	76.9	4.13	76.77	0.00	0.13
38	1.02	15.53	1.01	15.47	0.01	0.06
39	3.61 -	66.34 161.7	3.61	66.21 161.71	0.00	0.14 0.01
3	8.53	145.58	8.57	145.74	0.04	0.16

[a] ^{1}H/^{13}C at 700/175 MHz; [b] isolated from *S. cellulosum* So ce1875; [c] isolated from *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF; [d] overlapping signals.

The elemental formula $C_{38}H_{44}N_2O_{10}$ of disorazole Z10 (39-*O*-desmethyl-25,25-dimethyl-disorazole Z) (**12**) with a retention time of 10.59 min in analytic HPLC showed the formal loss of $C_2H_2O_2$. In the NMR spectra both methoxy signals were absent. Similar to variants **10** and **11** only one carboxyl group C-39 at 176.75 ppm remained as free carboxylic acid on one side chain. Its position was retrieved from HMBC correlations with C-38, -34, and -24. The second side chain featured two geminal methyl groups C-30 and C-31 similar to disorazole A1 (**1**).[8] These were identified from their mutual HMBC correlations and positioned from their HMBC correlations with C-12, C-25 and C-26 together with their NOE interactions.

Table S13 NMR data of disorazole Z10 (**12**) in methanol-d_4.[a]



Pos.	δ_H	m (J [Hz])	COSY	ROESY	δ_C	type	H to C HMBC
1	-	-	-	-	160.99	C	24
2	-	-	-	-	135.95	C	3
3	8.47	s	-	-	145.60	CH	
4/16	-	-	-	-	163.59 163.54	C C	3/15, 5/17 3/15, 5/17
5/17	6.13c	m	6/18	-	116.05 116.22	CH CH	7/19 7/19
6/18	6.83	m	5/17	-	139.27 139.15	CH	7/19, 8/20
7/19	6.13[c]	m	-	-	133.28[b] 132.88[b]	CH CH	
8/20	6.72	dd (15.5, 12.3)	9/21	11a/23a	136.26 136.07	CH CH	7/19, 9/21, 10/22 7/19, 9/21, 10/22
9/21	6.13[c]	m	8/20, 10/22	-	132.63[b] 132.13[b]	CH CH	
10	5.69[d]	m	9/21, 23ab	-	131.62[b]	CH	12, 26
11a	2.71	m	10, 12	8/20, 11b, 30, 31	29.81	CH_2	9, 10, 12
11b	2.43	dd (13.9, 7.3)	10, 12	11a, 26, 30, 31			
12	5.13	d (10.3)	11ab	11b, 26, 27/28, 30, 31	79.15	CH	11a/b, 26, 30, 31
13	-	-	-	-	161.75	C	12
14	-	-	-	-	136.07	C	15

Pos.	δ_H	m (J [Hz])	COSY	ROESY	δ_C	type	H to C HMBC
15	8.56	s	-	-	145.53	CH	
22			9/21, 23ab	-	131.93[b]	CH	24, 34
23a	2.72	m	22, 24	8/20, 23b, 38	30.76	CH_2	21, 22, 24
23b	2.62	dd (13.6, 7.3)	22, 24	22, 23a, 24, 34, 38			
24	5.45	d (9.9)	23ab	23b, 34, 38	77.41	CH	34, 23a/b, 38
25	-	-	-	-	42.94	C	12, 26, 30, 31
26	3.89	d (7.7)	29	15, 11b, 12, 27/28, 30, 31	77.99	CH	12, 27, 28, 30, 31
27	5.59	ddq (15.4, 8.1, 1.5)	28, 29	11b, 12, 26, 29/37, 30, 31	131.93[b]	CH	
28	5.69[d]	m	27, 29	-	129.78	CH	
29/37	1.72	dd (6.6, 1.5)	26/34, 27/35, 28/36	27/35, 28/36	18.14	CH_3	27/35, 28/36
					18.18	CH_3	27/35, 28/36
30	0.98	s	31	11a, 12, 26, 36, 31	19.66	CH_3	12, 26, 31
31	1.05	s	30	11a, 12, 26, 36, 30	19.44	CH_3	12, 26, 30
33	-	-	-	-	56.30	C	23b, 24, 34, 38
34	4.37	d (7.3)	37	23b, 24, 3, 38	76.39	CH	24, 35, 36, 38
35	5.76	m	36, 37	-	129.96	CH	
36	5.69[d]	m	37	-	131.62[b]	CH	
38	1.35	s	-	23ab, 24, 34, 27/35	14.02	CH_3	24, 34
39	-	-	-	-	176.75	C	24, 34, 38

[a] $^1H/^{13}C$ 600/150 MHz; [b, c, d] interchangeable. For comparability reasons side chain numbering starts with position 25 and 33, respectively. Disorazole Z10 (**12**) misses CH_3-32 and CH_3-40.

Comparison of ^1H and ^{13}C NMR data of disorazol Z10 (**12**) isolated from *S. cellulosum* So ce1875 and *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF revealed a shift differences <0.2 ppm ($\Delta\delta_H^b$-δ_H^c). Thus, both compounds are identical (Table S14 and Figure S61). Greater differences of carbon shifts can be explained due to their extraction from HMBC experiments.

Table S14 Comparison of ^1H and ^{13}C NMR data of **12** in methanol-d_4 isolated from *S. cellulosum* and *M. xanthus*.[a]

Pos.	δ_H^b	δ_C^b	δ_H^c	$\delta_C^{c,d}$	$\Delta\delta_H^b$-δ_H^c	$\Delta\delta_C^b$-$\delta_C^{c,d}$
1	-	160.99	-	160.30	-	0.69
2	-	135.95	-	135.26	-	0.69
3	8.47	145.6	8.30	144.68	0.17	0.92
4/16	-	163.59	-	162.19	-	1.40
4/16	-	163.54	-	162.19	-	1.35
5/17	6.13	116.05	6.11	114.31	0.02	1.74
5/17	-	116.22	6.11	114.31	-	1.91
6/18	6.83	139.27	6.81	137.42	0.02	1.85
6/18	-	139.15	6.81	137.42	-	1.73
7/19	6.13	133.28	6.12	131.21	0.01	2.07
7/19	-	132.88	6.12	131.21	-	1.67

Pos.	$\delta_{H}{}^{b}$	$\delta_{C}{}^{b}$	$\delta_{H}{}^{c}$	$\delta_{C}{}^{c,d}$	$\Delta\delta_{H}{}^{b}$-$\delta_{H}{}^{c}$	$\Delta\delta_{C}{}^{b}$-$\delta_{C}{}^{c,d}$
8/20	6.72	136.26	6.74	134.66	0.02	1.60
8/20	-	136.07	6.74	134.66	-	1.41
9/21	6.13	132.63	6.12	130.29	0.01	2.34
9/21	-	132.13	6.12	130.29	-	1.84
10	5.69	131.62	5.72	130.91	0.03	0.71
11a	2.71	29.81	2.71	28.48	0.00	1.34
11b	2.43	29.81	2.42	28.48	0.01	1.34
12	5.13	79.15	5.13	77.34	0.00	1.81
13	-	161.75	-	160.30	-	1.45
14	-	136.07	-	135.26	-	0.81
15	8.56	145.53	8.62	144.36	0.06	1.17
22	-	131.93	5.72	130.91	-	1.02
23a	2.72	30.76	2.79	29.92	0.07	0.85
23b	2.62	30.76	2.58	29.92	0.04	0.85
24	5.45	77.41	5.36	76.08	0.09	1.33
25	-	42.94	-	41.54	-	1.40
26	3.89	77.99	3.89	76.42	0.00	1.57
27	5.59	131.93	5.59	130.29	0.00	1.64
28	5.69	129.78	5.73	127.14	0.04	2.64
29/37	-	18.14	1.73	16.89	-	1.25
29/37	1.72	18.18	1.73	16.89	0.01	1.29
30	0.98	19.66	1.05	18.14	0.06	1.52
31	1.05	19.44	0.98	18.30	0.08	1.14
33	-	56.3	-	54.46	-	1.84
34	4.37	76.39	4.39	75.46	0.02	0.93
35	5.76	129.96	5.69	128.08	0.07	1.88
36	5.69	131.62	5.59	130.29	0.10	1.33
38	1.35	14.02	1.24	14.99	0.11	0.97
39	-	176.75	-	178.23	-	1.48

[a] $^{1}H/^{13}C$ at 700/175 MHz; [b] isolated from *S. cellulosum* So ce1875; [c] isolated from *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF; [d] δ_C extracted from HMBC.

2.6.2 Supplementary tables

Table S15 DNA oligos used in this study.

Name	**Sequence (5’ to 3’)**	**Application**
p15A-cm-MluCI-dis427-F	CCTGGCCGAGGATCGCGCCCACCGCCTGATCAGCGGCCCCCGCATGCTCCAGCAGCTCGCCAGGAGATCCCGCCAGCGGCAATTAGATCCGAAAACCCCAAGTTACG	for amplification of the p15A-cm-MluCI vector
p15A-cm-MluCI-dis427-R	AGTCAGGCTATGCCCTCGTTTTATAGACCACAAAATCGCGGTGTCAACGCGGTGCGCGAGAATACAAAGAGTCCGCGCAAAATTAGATCCTTTCTCCTCTTTAGATC	
pBR322-amp-BstXI-dis427-F	GCTGGGGAGCTCGGCCTTCCGGGAGGACTACGGCCTGCGCTACGCCTACGTCGCCGGATCCATGGTCGAGGGCATCTCGTCCAGCGAGCTGGAGATCCGAAAACCCCAAGTTACG	for amplification of the pBR322-amp-BstXI vector
pBR322-amp-BstXI-dis427-R	GGCGAGGTGAATGGCGGTGAGGGACGACGCGCACGCGGTGTCCACCGCCATGCAGGGCCCGTGCAGGTTGAGGAAATACGCCACATGATTGGGATCTAAAGAGGAGAAAGGATCT	

Name	Sequence (5’ to 3’)	Application
dis427-F3-F	CTGGGGAGCTCGGCCTTCCGGGAGGACTAC	for amplification of a BGC fragment
dis427-F3-R	GCTGTCCGCCGAGAGGTGGTTCCGTGTGGG	
p15A-cm-dis427-F	GAGCGGCTCGTTCGCCCTCGCGGAGCTCCGCGGCCGCGCCCCCGCGACGGACCGGTGGGTCCCACACGGAACCACCTCTCAGATCCGAAAACCCCAAGTTACG	for amplification of the p15A-cm vector
km-int-F	ACTAGTGCTTGGATTCTCACCAA	for amplification of the *km-int* cassette
km-int-R	CAACGCGGTGCGCGAGAATACAAAGAGTCCGCGCAAAATTAAGACGGACGAGGTGCTCACTCA	
Ptet-disA-R	ATGCCGATGATGGCAATGCCGTCCTGCTCTACGTGATTCCCGCTGCTCATAGATCCTTTCTCCTCTTTAGATC	for amplification of the *km-int-Ptet* cassette
Papr-disD-F	ATCACGCAGGTCGCAGGCACGAGCAGGGCTCTCTCATCTAGGAGGCGCTTACGCTCAGTGGAACGAGGTT	for amplification of the *Papr-disD* cassette
Papr-disD-R	CCTCCCATCCCTGGCGACTGCGACCCCTGCCCCGGAAACATGACGGCCTTCATAATCTGTACCTCCTTAAGTCAGCCAATCGACTGGCGAGC	
gent-delF-F	AGGCCTCGGCCGATCTCCTGGACCACCGCCTGCGCGCGCTGTCGCGGTAAGAAGGCACGAACCCAGTTGAC	for amplification of the *gent-delF* cassette
gent-delF-R	CTGAGCCTTTCGTTTTATTTGATGCCTGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTTTAGGTGGCGGTACTTGGGTC	
apr-Pvan-disA-F	AATTTTGCGCGGACTCTTTGTATTCTCGCGCACCGCGTTGACACCGCGATTCAGCCAATCGACTGGCGAGCGG	for amplification of the *apr-Pvan-disA* cassette
apr-Pvan-disA-R	GCCATGCCGATGATGGCAATGCCGTCCTGCTCTACGTGATTCCCGCTGCTCATATGCGTTTCCTCGCATCGTG	
dis427-chk01-F	GGATCGTGAGTACCTGGAGAAG	for colony PCR
dis427-chk01-R	GAGCGTCCGGGAGGTCGTGGGC	
dis427-chk02-F	GCAGAAGTACGTGGGCCTCAGC	
dis427-chk02-R	CGACGAGCAGGGTGGCGTATCC	
dis427-chk03-F	CGACCTCTACCTGAAGCACGATG	
dis427-chk03-R	CTGCACCTGATAATCGACCTGGC	
HisTEV-disF-F	TTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCCTTGCAAAACTCATGTTGTCTCA	for cloning disF_{427} into pHis-TEV
HisTEV-disF-R	CTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTACGCGCTCCGCTGGGCGATGG	

Table S16 Half-inhibitory concentrations (IC_{50}, mean ± SD) of disorazole Z1 (**3**) in comparison of disorazole A1 (**1**) on selected human cancer cell lines.

Cell line	Origin	**1** IC_{50} [nM]	**3** IC_{50} [nM]
A549	lung carcinoma	0.35 ± 0.15	0.21 ± 0.11
HCT-116	colon carcinoma	0.24 ± 0.02	0.25
HepG2	hepatocellular carcinoma	0.28 ± 0.07	0.06
HL-60	promyelocytic leukemia	0.07 ± 0.01	nd
KB-3.1	cervix carcinoma	0.15 ± 0.02	0.07 ± 0.02
KB-V1	cervix carcinoma, MDR	0.17 ± 0.01	0.11 ± 0.02
U-2 OS	osteosarcoma	0.43 ± 0.16	0.09

nd: not determined; MDR: multi-drug resistant.

2.6.3 Supplementary figures

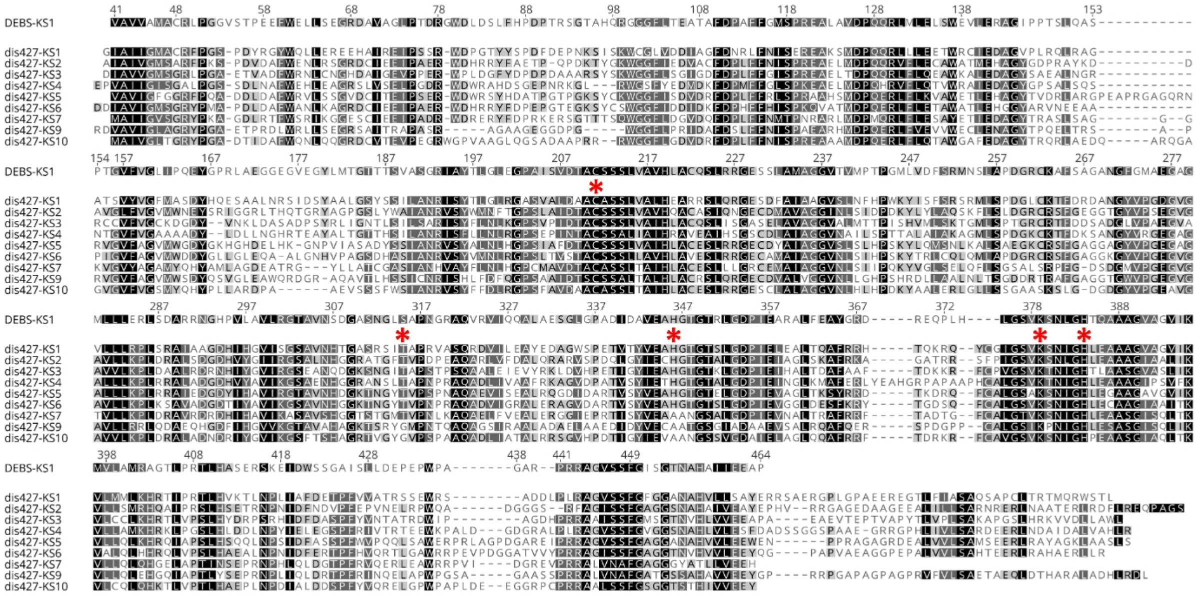


Figure S7 The aligned KS domains. DEBS-KS1 is the first KS domain of 6-deoxyerythronolide B synthase.[33] The dis427-KS with corresponding module numbers are KS domains from the *S. cellulosum* So ce427 disorazole Z biosynthetic pathway. The conserved active sites are marked by red asterisks.

Figure S8 The aligned CP domains. DEBS-CP2 is the CP domain from the second module of 6-deoxyerythronolide B synthase.[51] The dis427-CP with corresponding module numbers are CP domains from the *S. cellulosum* So ce427 disorazole Z biosynthetic pathway. The serine residue for phosphopantetheinyl arm attachment is marked by a red asterisk.

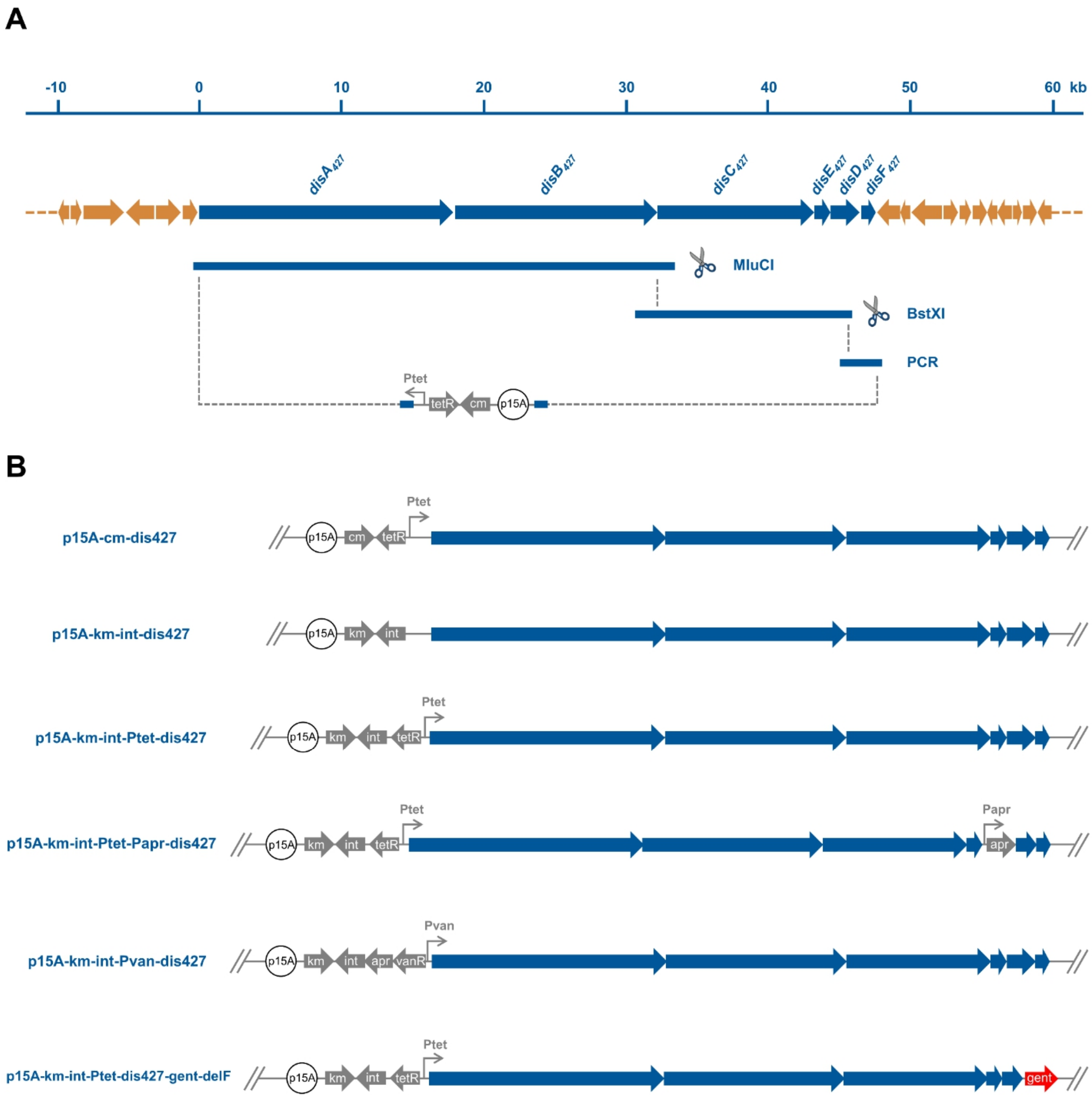
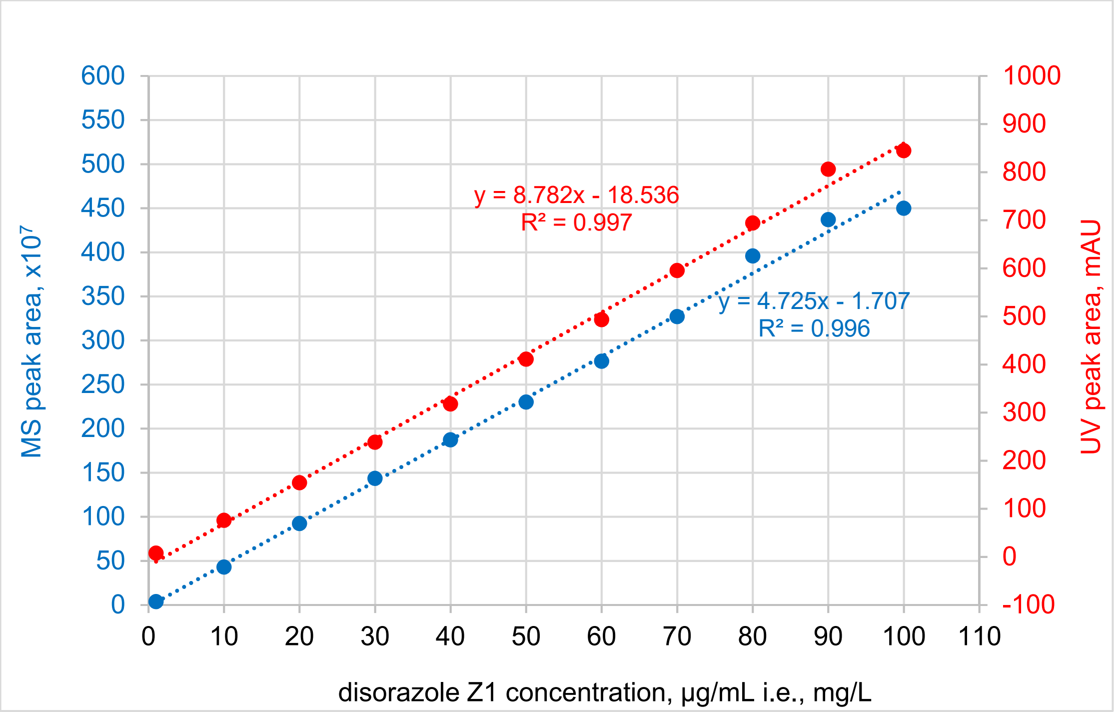


Figure S9 Cloning and engineering of the *dis427* gene cluster for heterologous expression. A: the disorazole Z biosynthesis gene cluster was cloned from the genomic DNA of *S. cellulosum* So ce427 using LLHR as described in methods section. B: scheme of plasmids containing *dis427* gene cluster generated in this study, the expression constructs are created using LCHR as described in methods section.

A



B

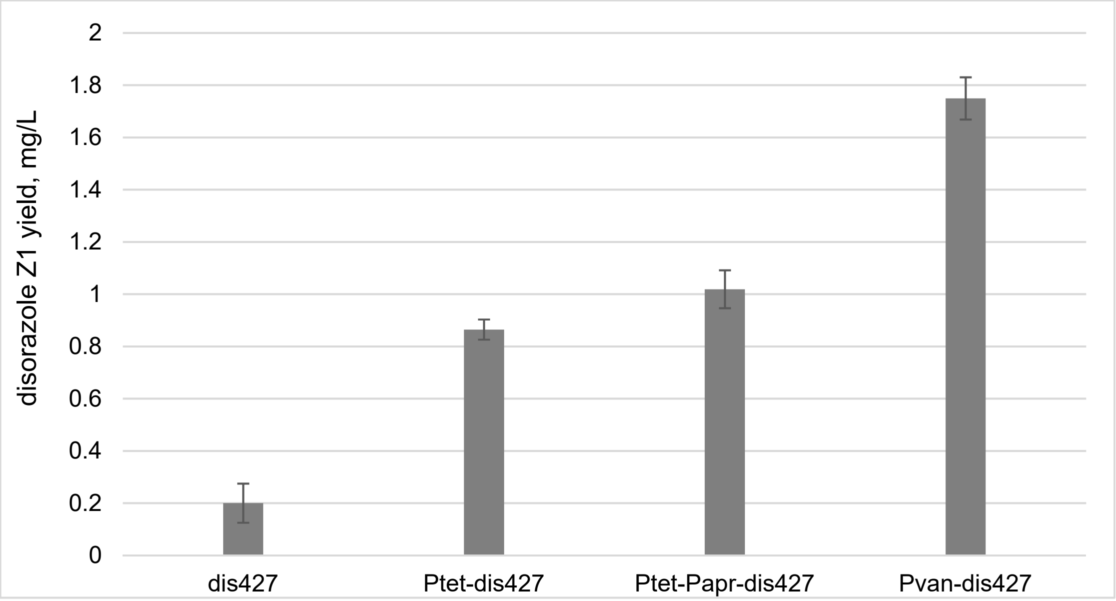


Figure S10 Quantitative analysis of disorazole Z1 (**3**) using HPLC-UV-MS. A: the standard curve of **3**. B: heterologous production yields of **3** in *M. xanthus* DK1622 before and after promoter engineering.

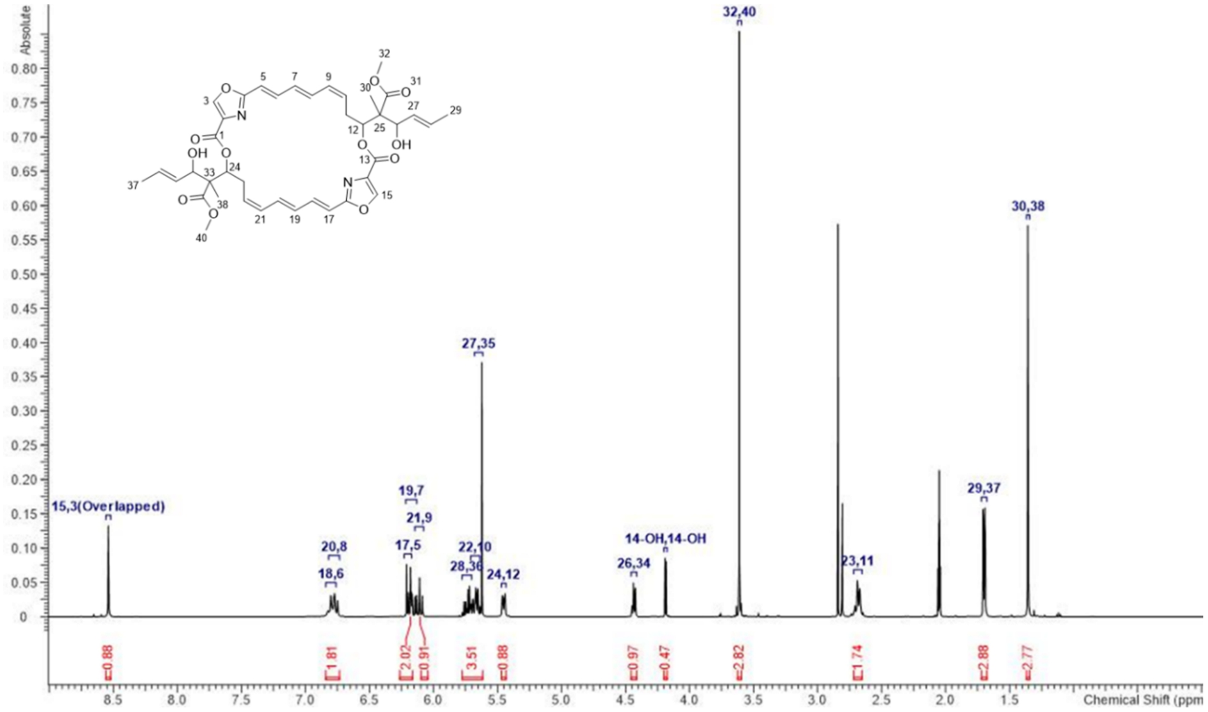


Figure S11 ^{1}H NMR spectrum of disorazole Z1 (**3**) in acetone-*d*$_6$ (500 MHz).

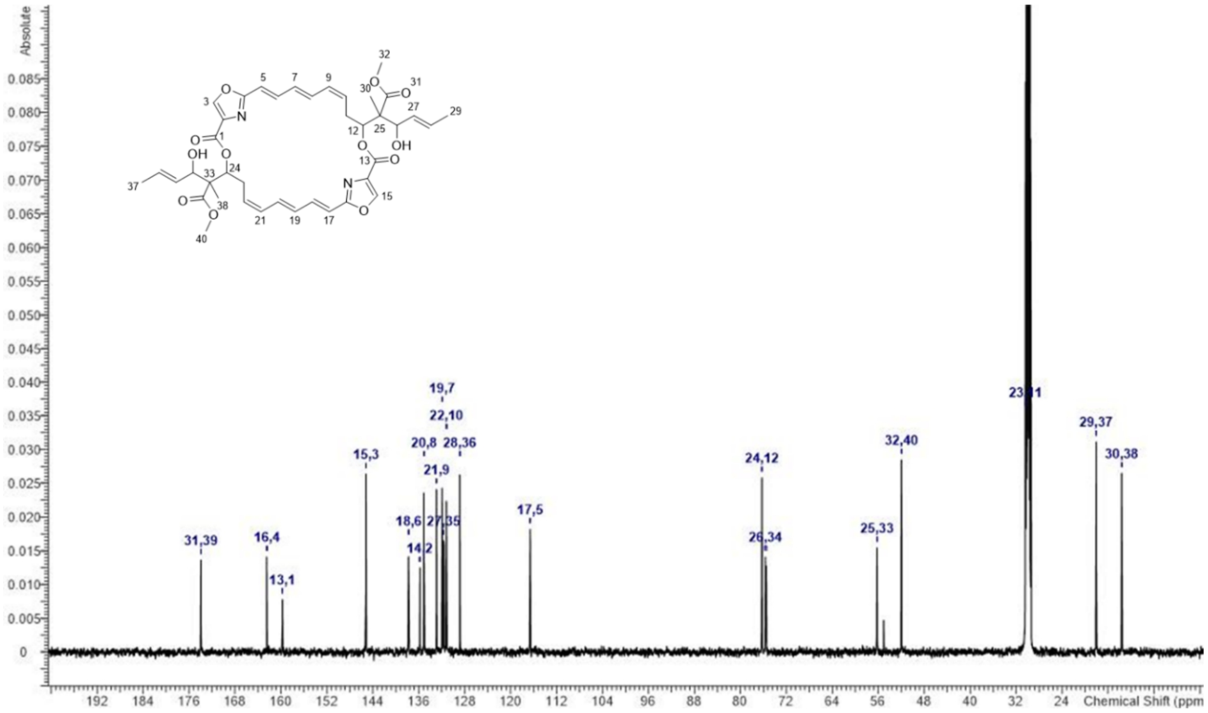


Figure S12 ^{13}C NMR spectrum of disorazole Z1 (**3**) in acetone-*d*$_6$ (100 MHz).

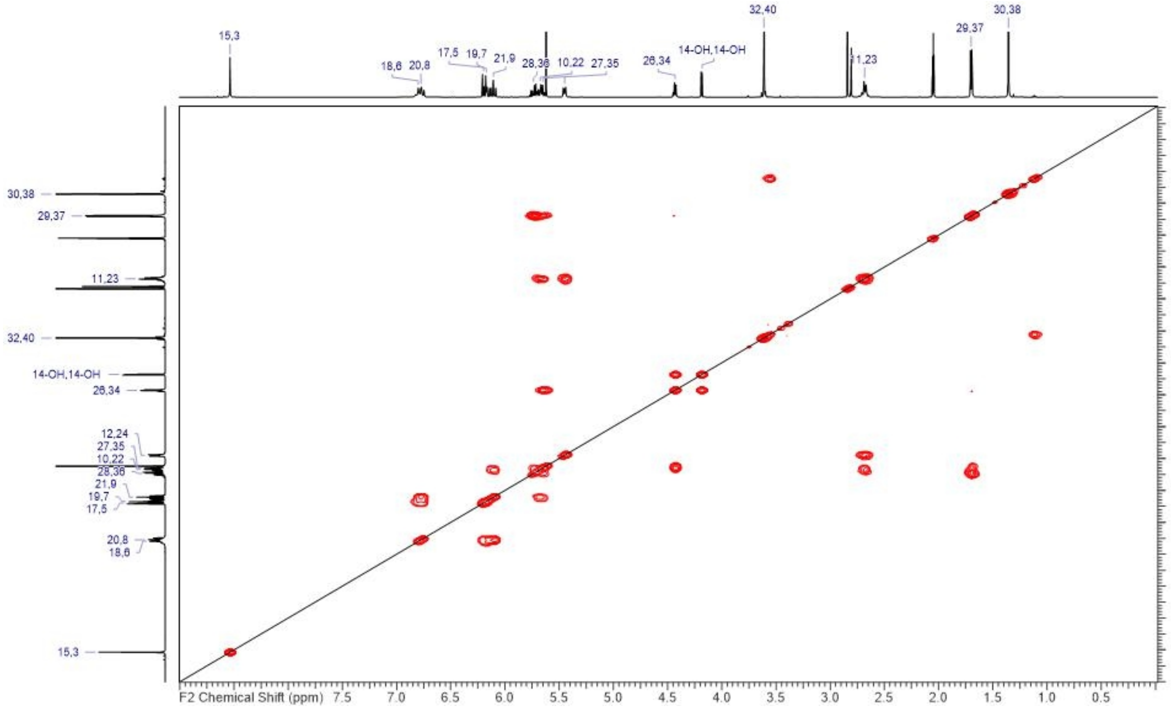


Figure S13 ^{1}H,^{1}H-COSY NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.

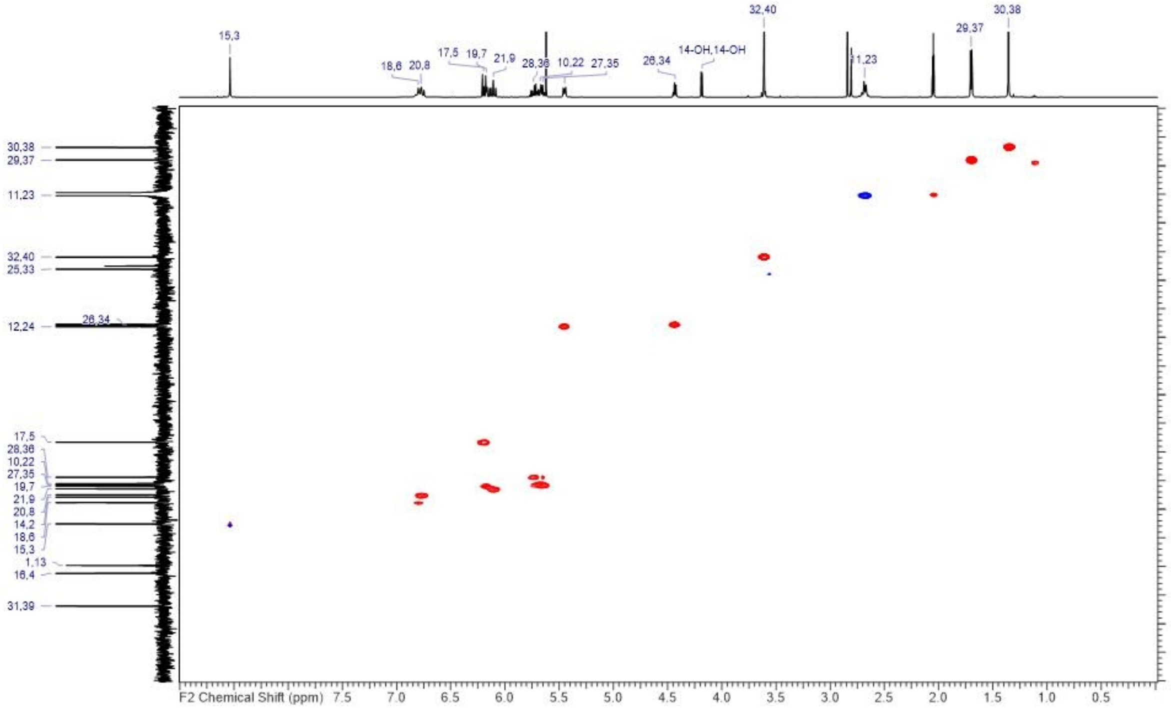


Figure S14 HSQC NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.

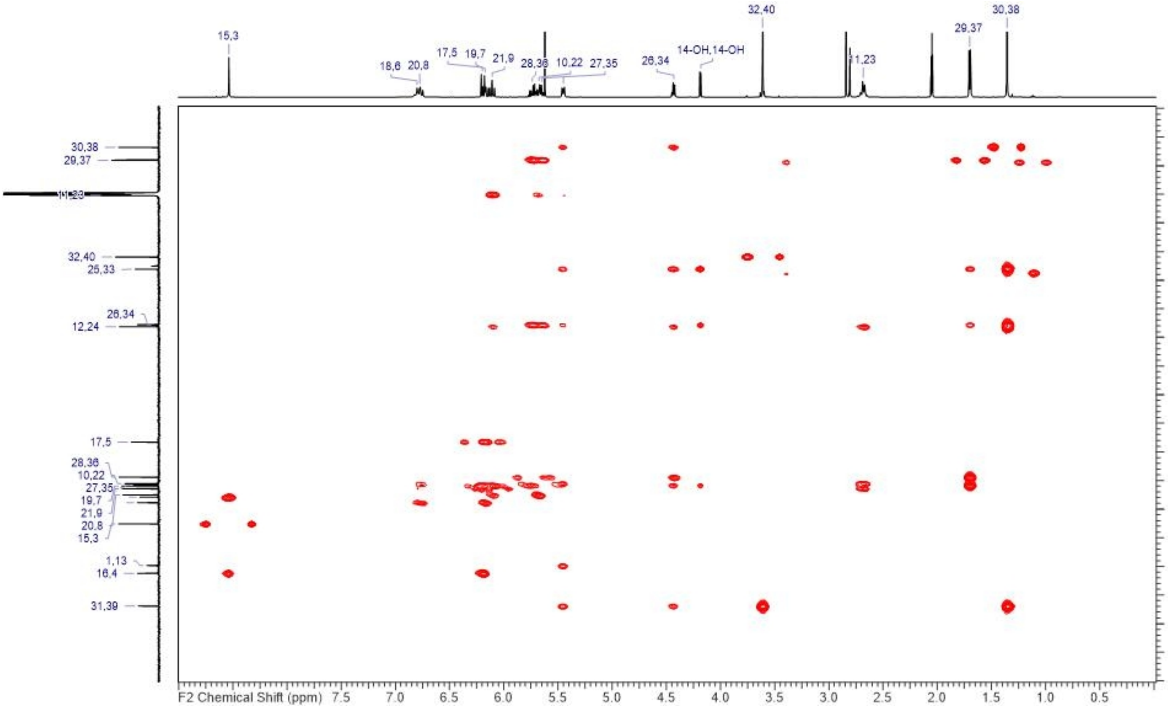


Figure S15 HMBC NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.

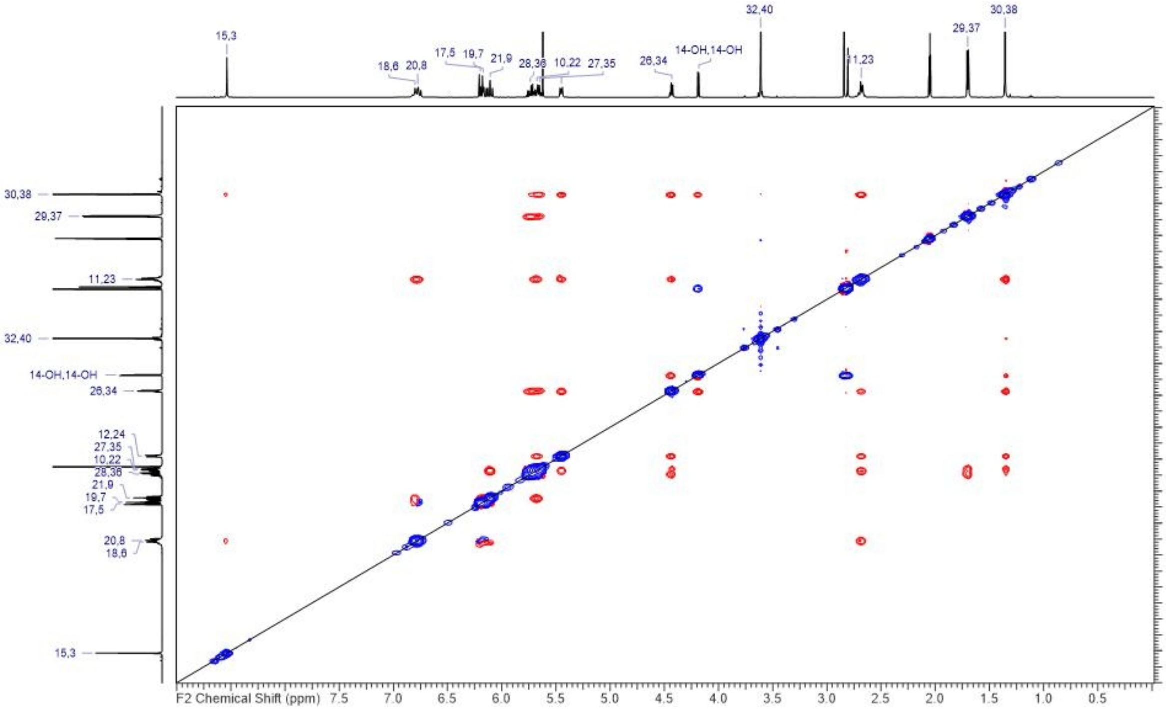


Figure S16 ^{1}H,^{1}H-ROESY NMR spectrum of disorazole Z1 (**3**) in acetone-d_6.

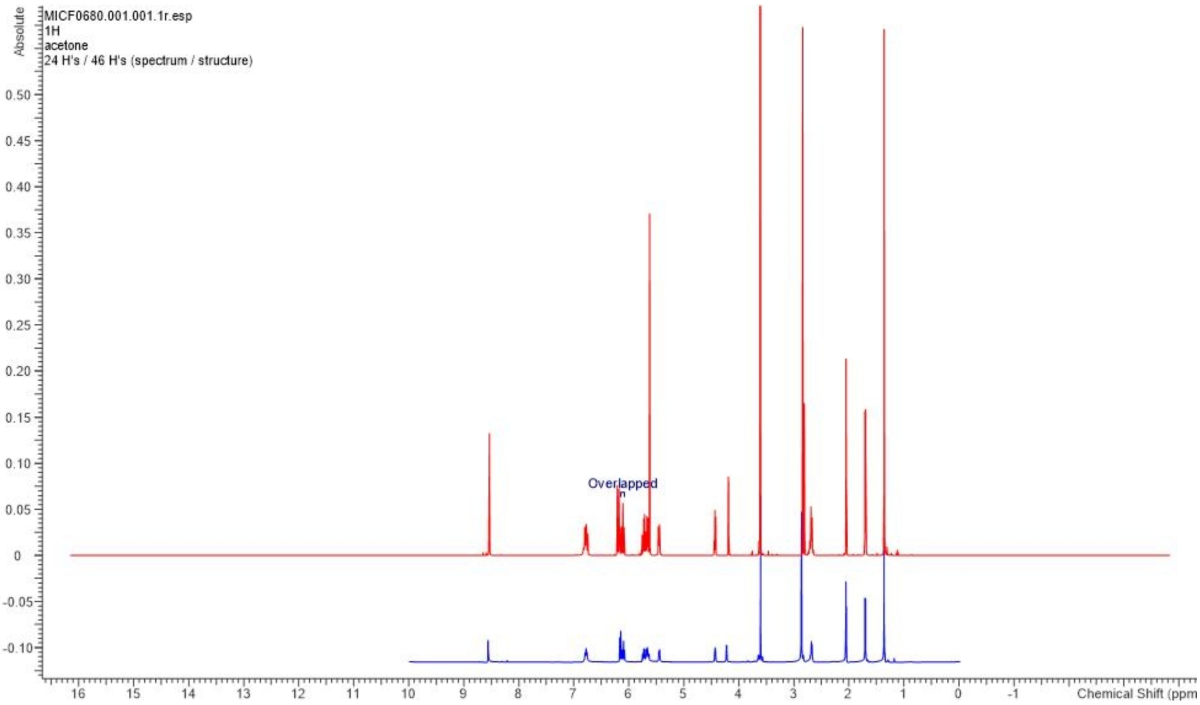


Figure S17 Comparison of ^{1}H NMR spectra of **3** in acetone-d_6 isolated from *S. cellulosum* and *M. xanthus*. Top/red: isolated from *S. cellulosum* So ce1875; Bottom/blue: isolated from *M. xanthus* DK1622::km-int-Ptet-dis427.

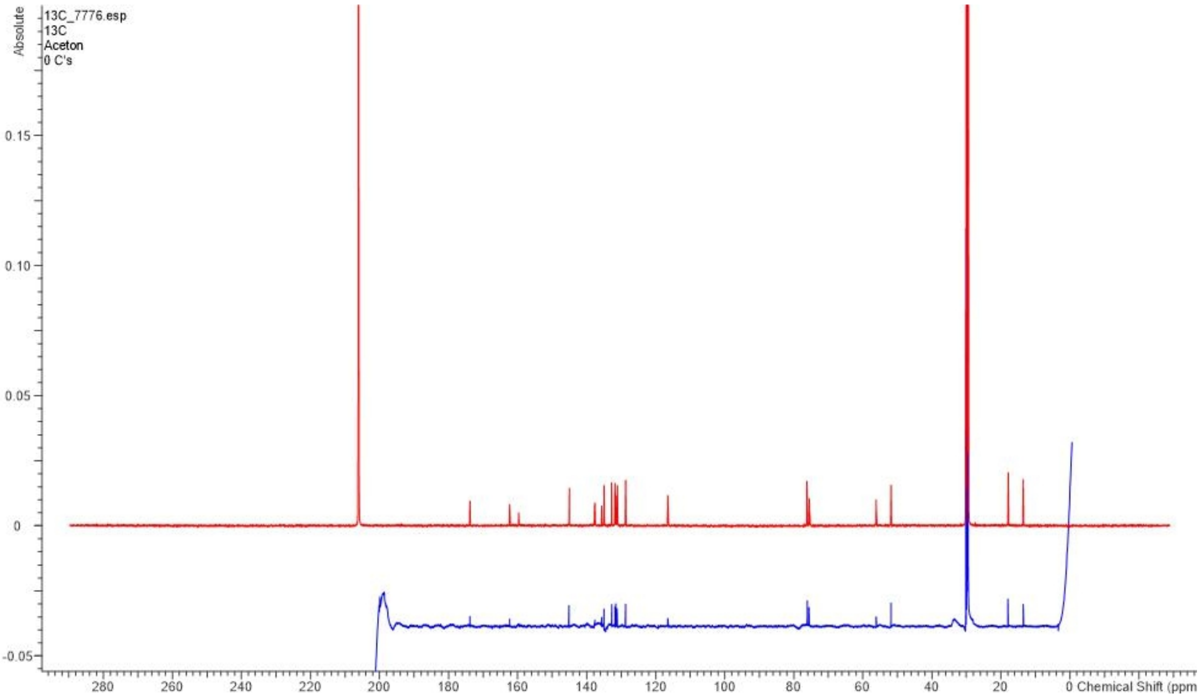


Figure S18 Comparison of ^{13}C NMR spectra of **3** in acetone-d_6 isolated from *S. cellulosum* and *M. xanthus*. Top/red: isolated from *S. cellulosum* So ce1875; Bottom/blue: isolated from *M. xanthus* DK1622::km-int-Ptet-dis427.

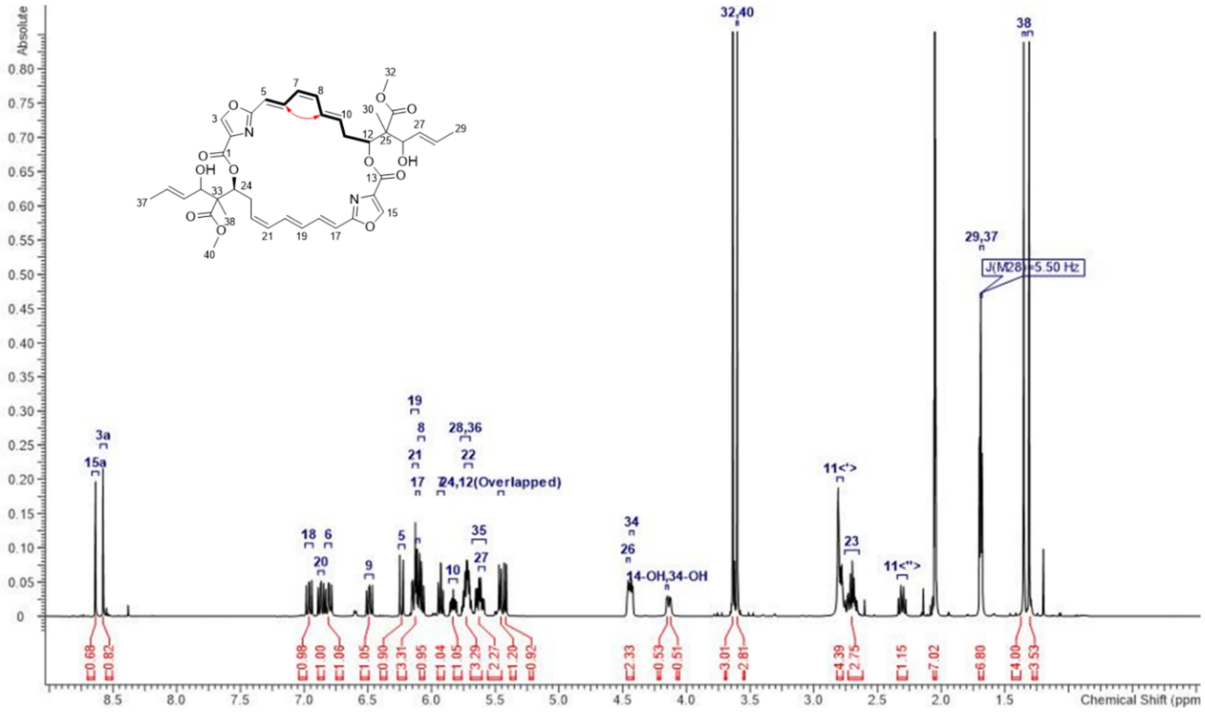


Figure S19 ^{1}H NMR spectrum of $\Delta^{7,8}$-*cis*-disorazole Z (**4**) in acetone-d_6 (500 MHz).

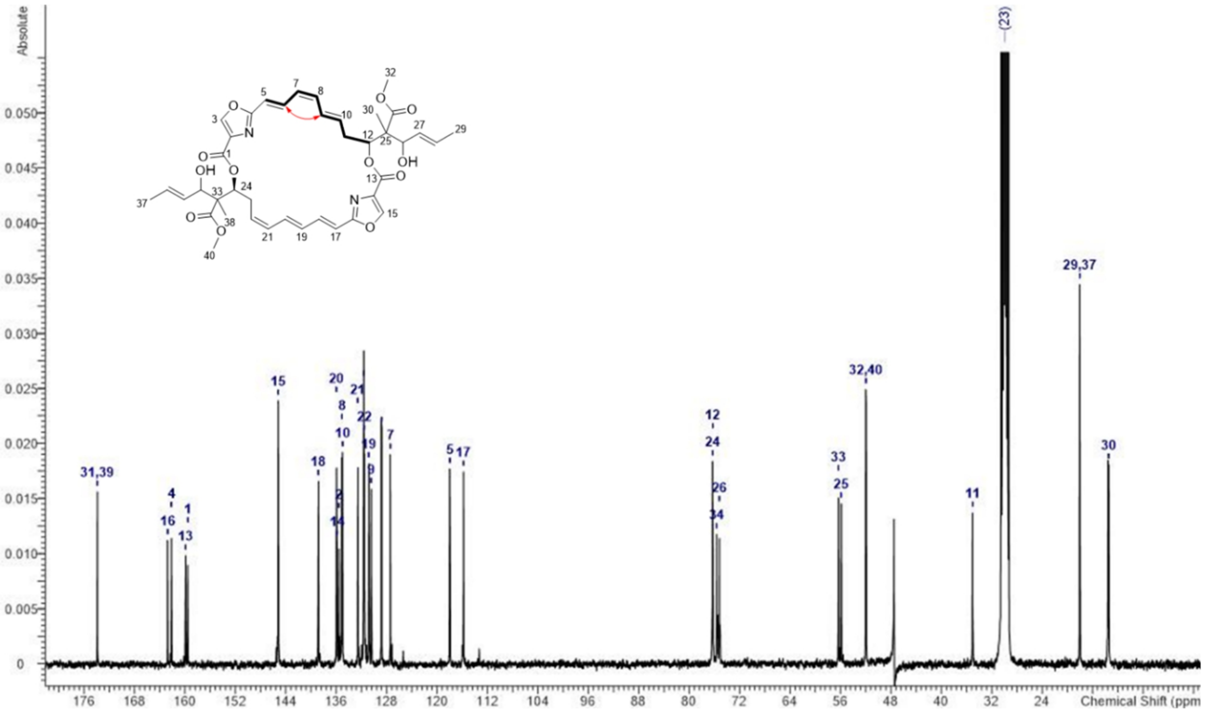


Figure S20 ^{13}C NMR spectrum of $\Delta^{7,8}$-*cis*-disorazole Z (**4**) in acetone-d_6 (100 MHz).

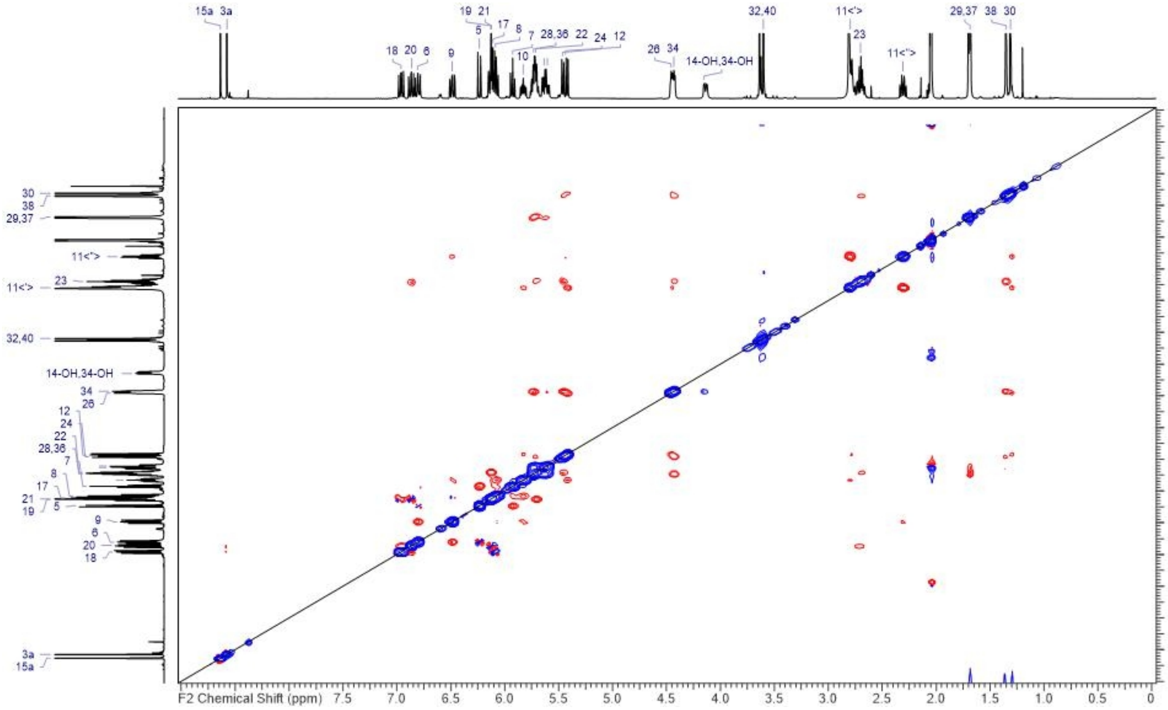


Figure S21 ^{1}H,^{1}H-ROESY NMR spectrum of Δ^{7,8}-*cis*-disorazole Z (**4**) in acetone-d_6.

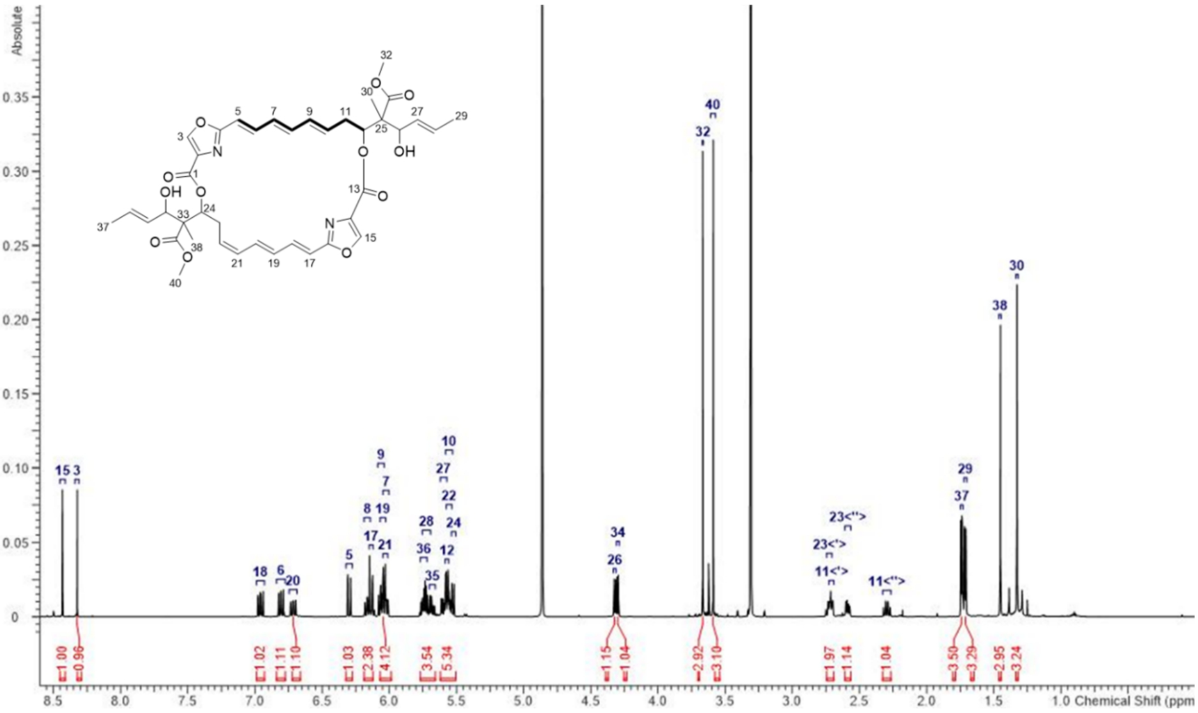


Figure S22 ^{1}H NMR spectrum of Δ^{9,10}-*trans*-disorazole Z (**5**) in methanol-d_4 (700 MHz).

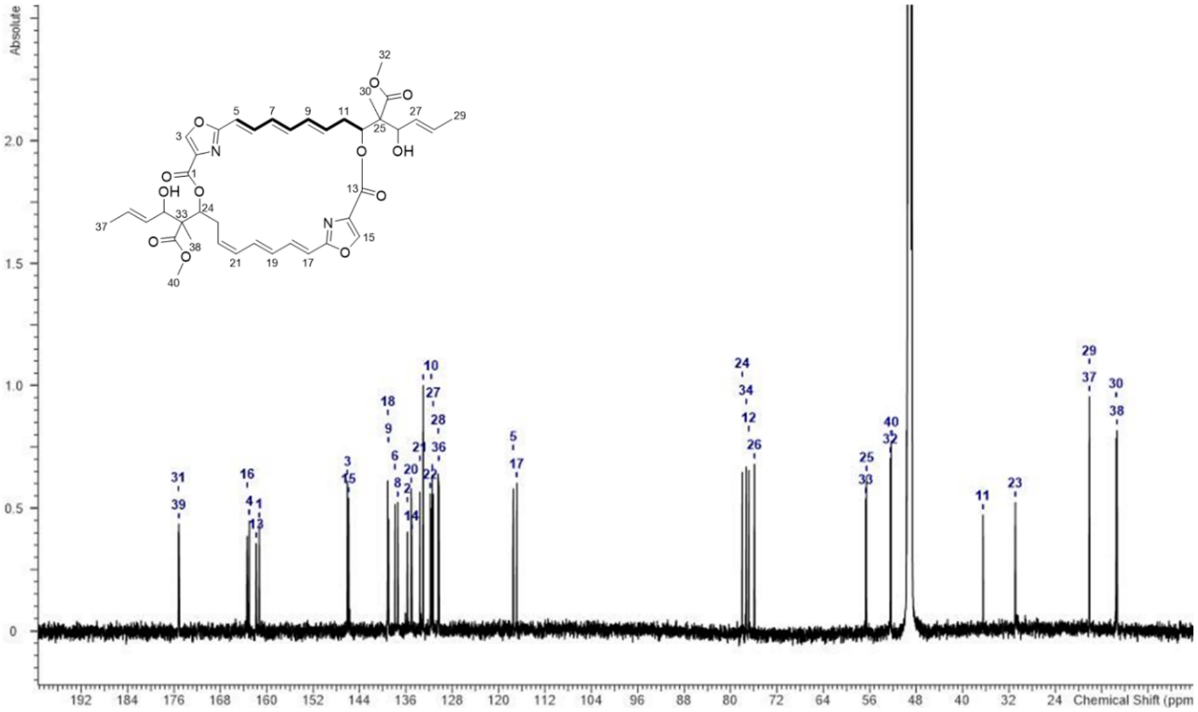


Figure S23 ^{13}C NMR spectrum of Δ^{9,10}-*trans*-disorazole Z (**5**) in methanol-d_4 (175 MHz).

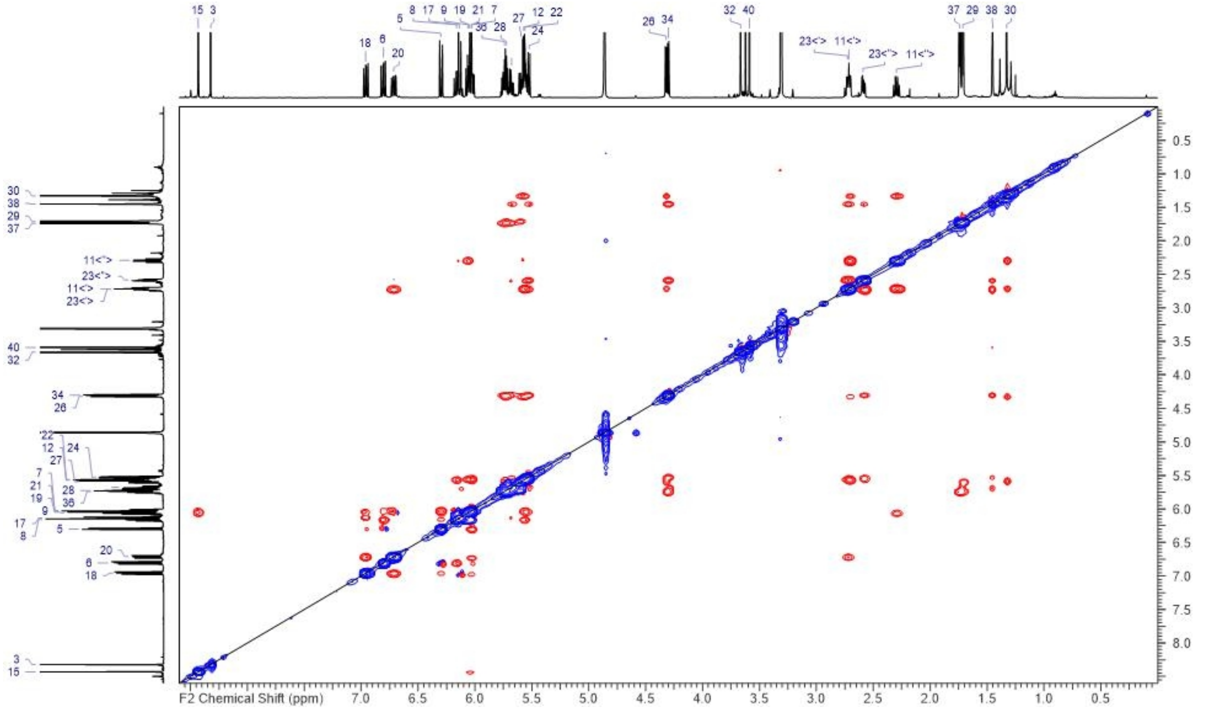


Figure S24 ^{1}H,^{1}H-ROESY NMR spectrum of Δ^{9,10}-*trans*-disorazole Z (**5**) in methanol-d_4.

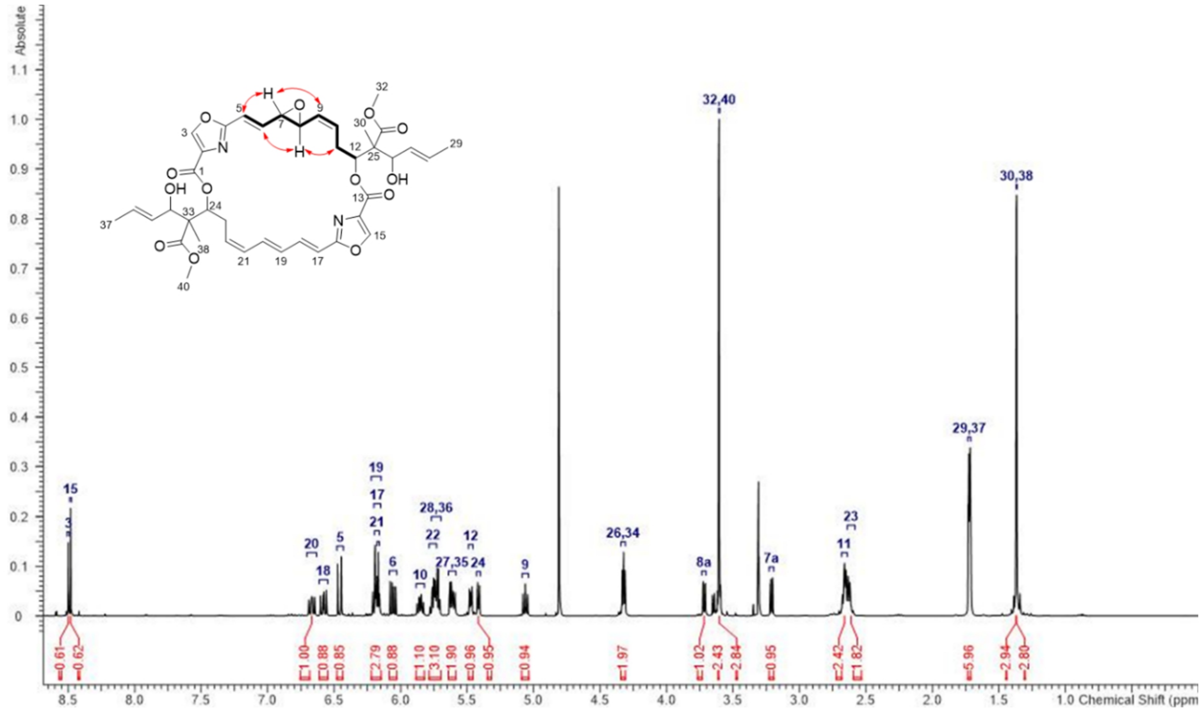


Figure S25 ^{1}H NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-*d*$_4$ (600 MHz).

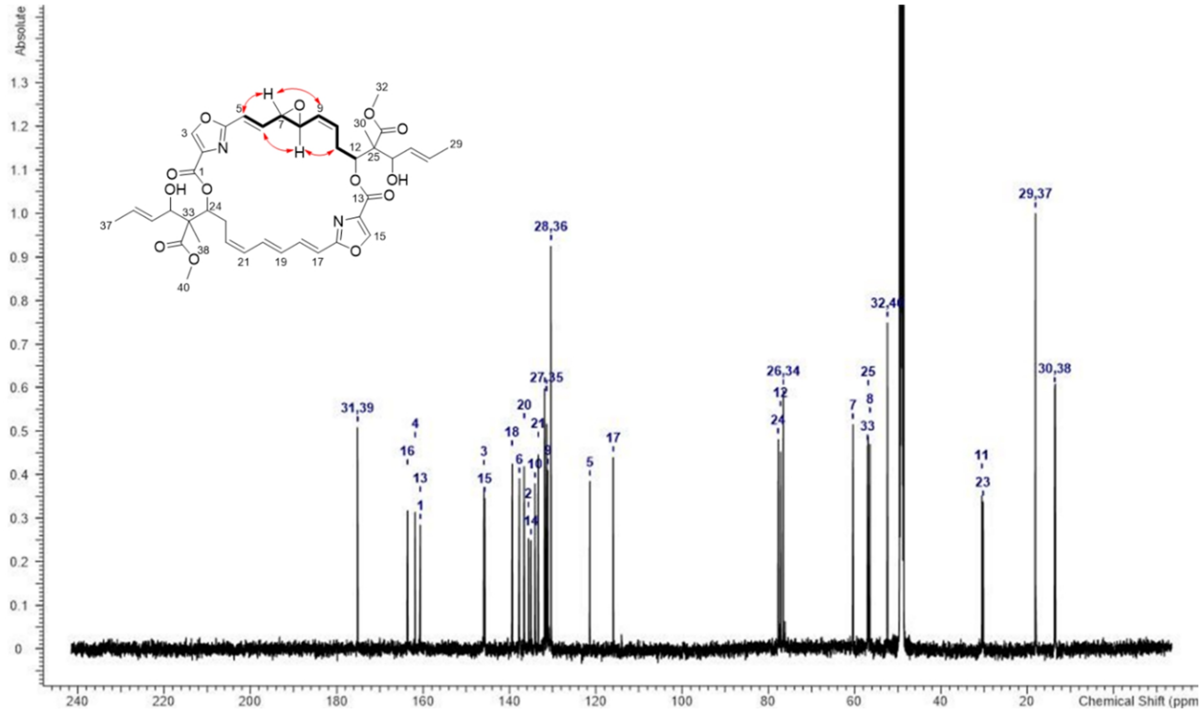


Figure S26 ^{13}C NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-*d*$_4$ (150 MHz).

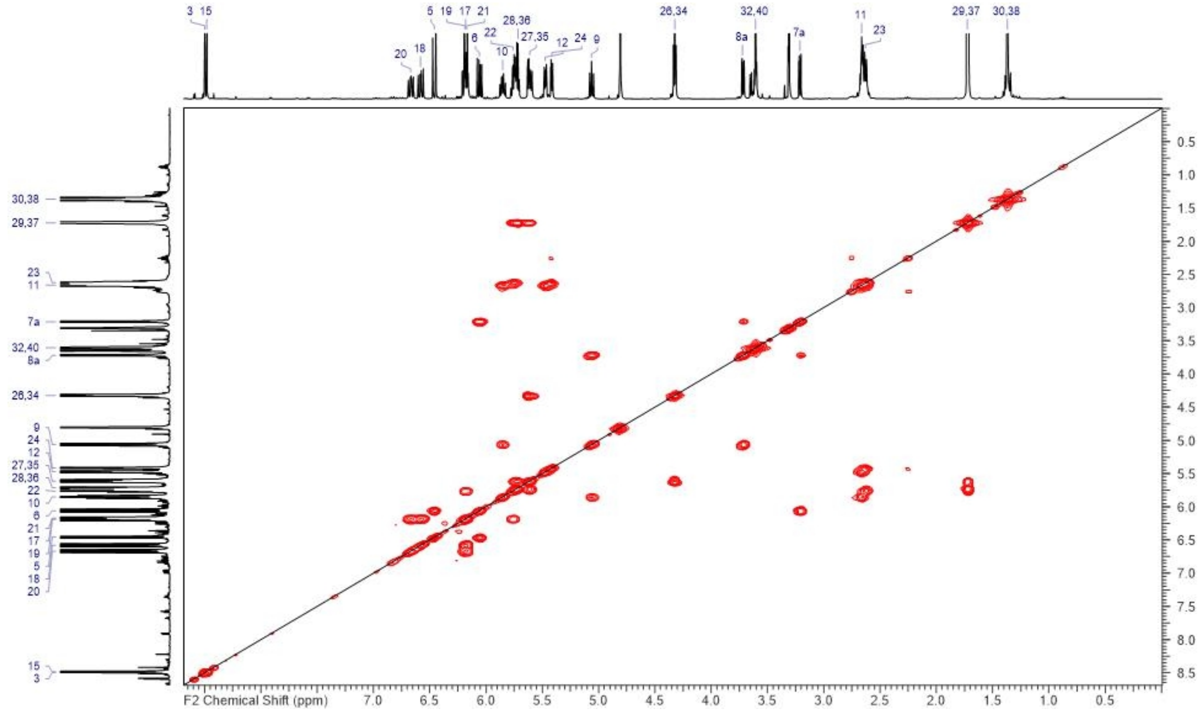


Figure S27 ^{1}H,^{1}H-COSY NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-d_4.

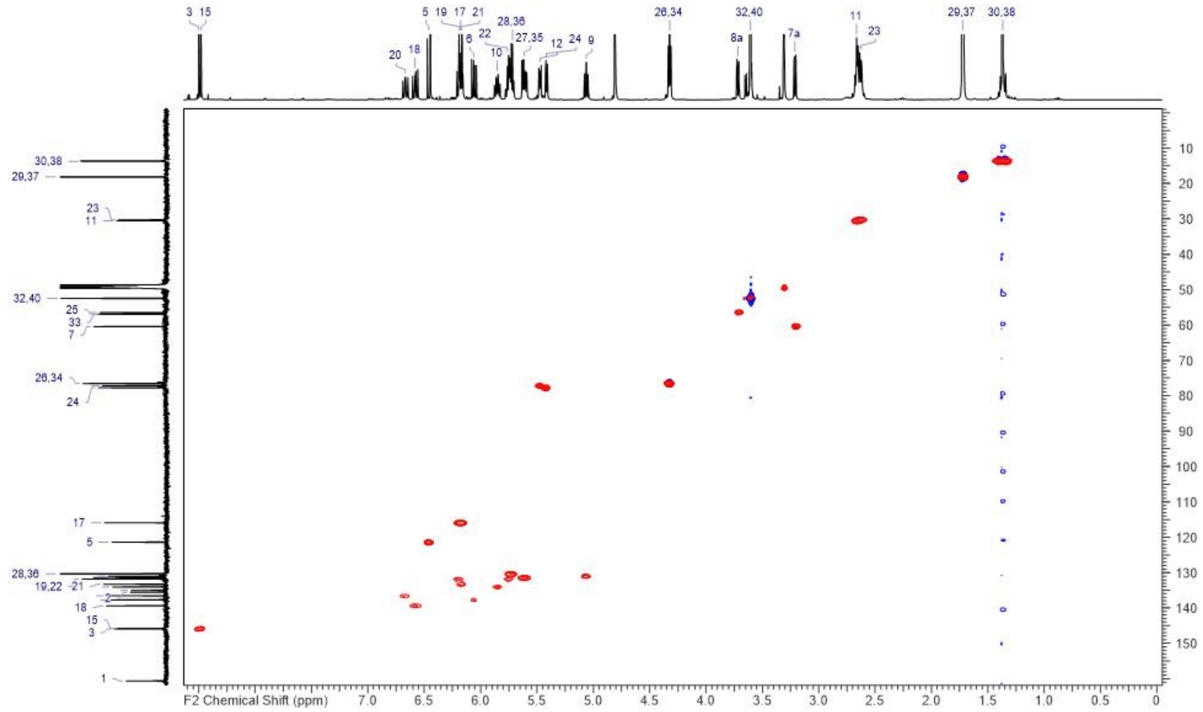


Figure S28 HMQC-NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-d_4.

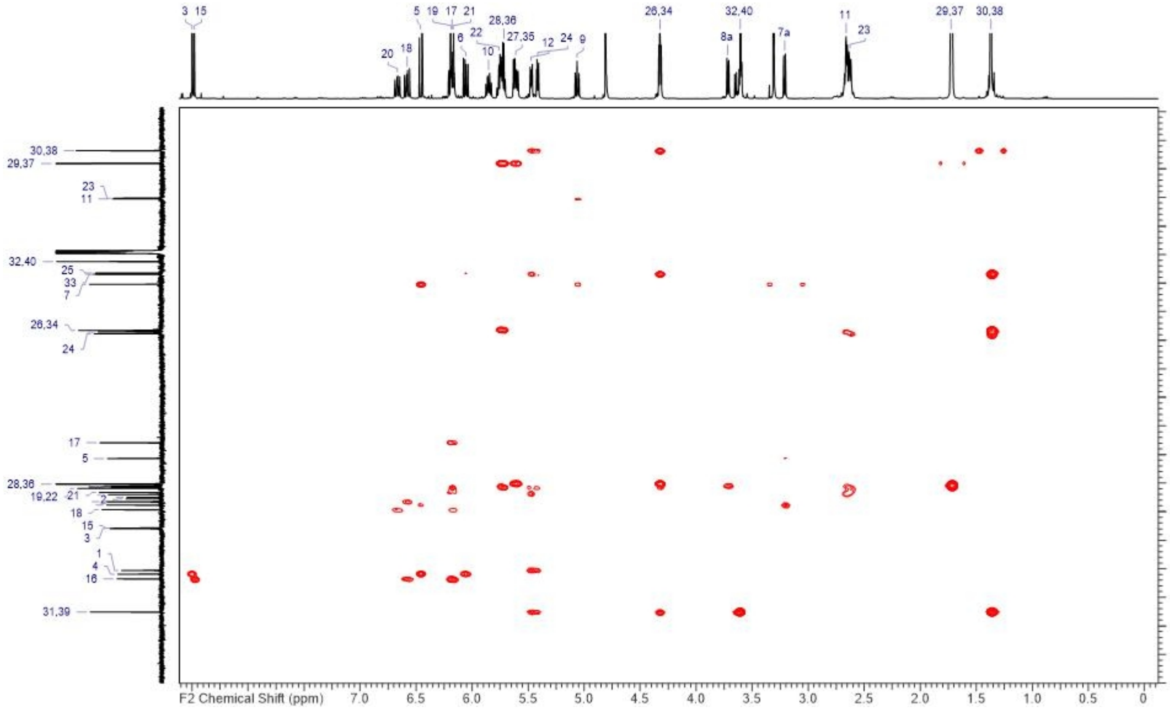


Figure S29 HMBC NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-d_4.

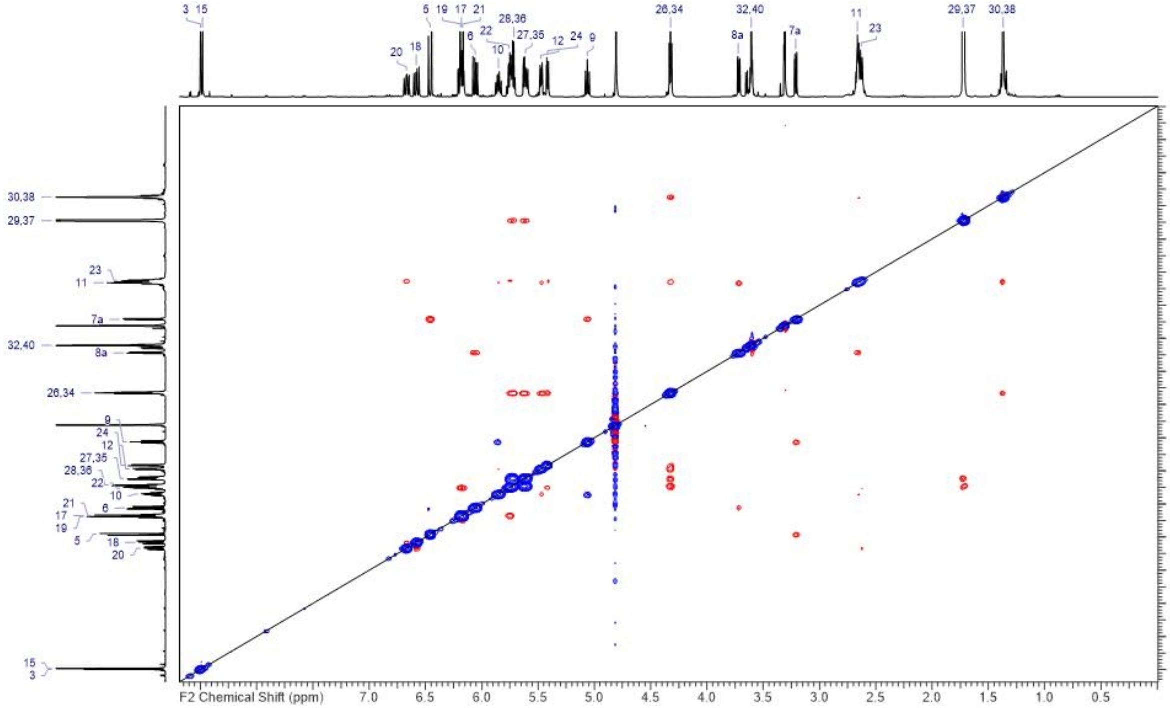


Figure S30 ^{1}H,^{1}H-ROESY NMR spectrum of 7,8-epoxy-disorazole Z (**6**) in methanol-d_4.

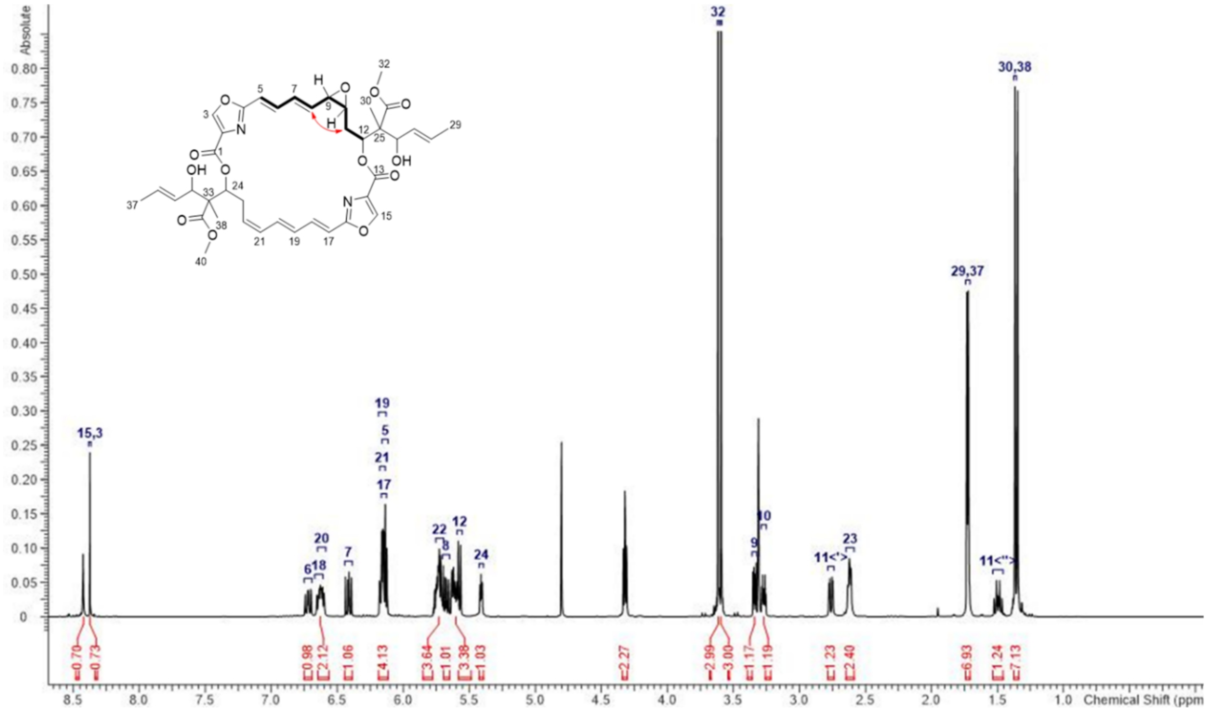


Figure S31 ^{1}H NMR spectrum of 9,10-epoxy-disorazole Z (**7**) in methanol-d_4 (600 MHz).

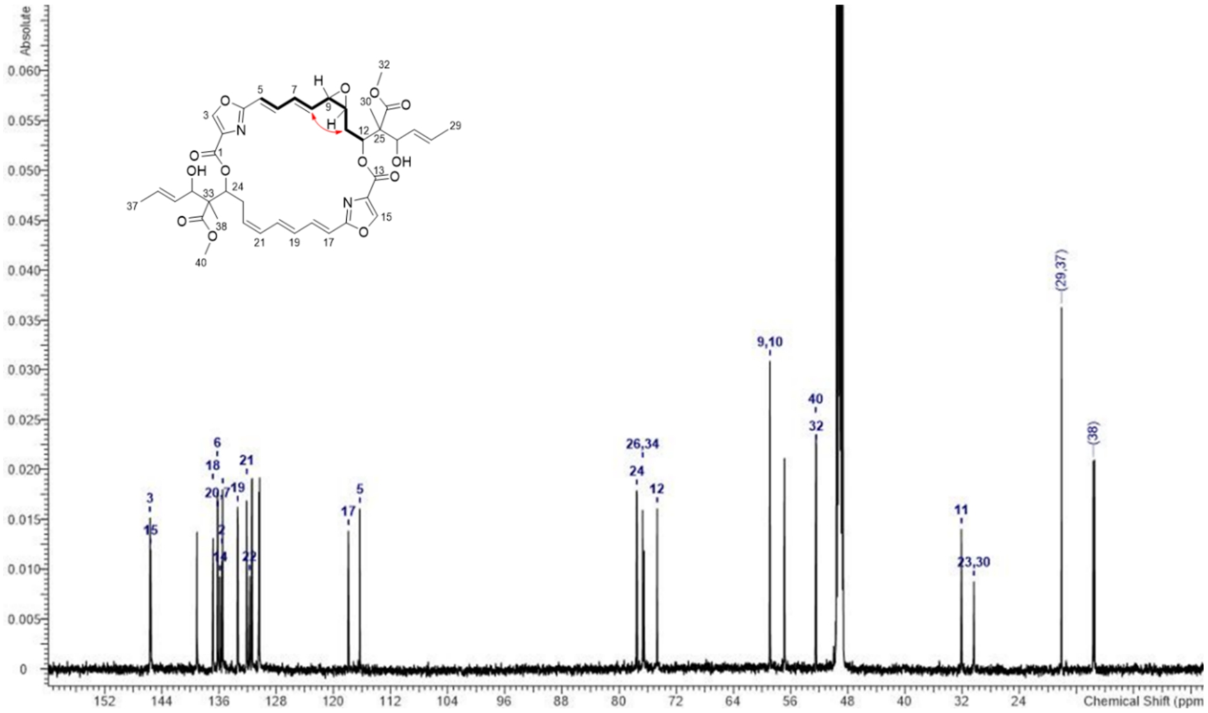


Figure S32 ^{13}C NMR spectrum of 9,10-epoxy-disorazole Z (**7**) in methanol-d_4 (150 MHz).

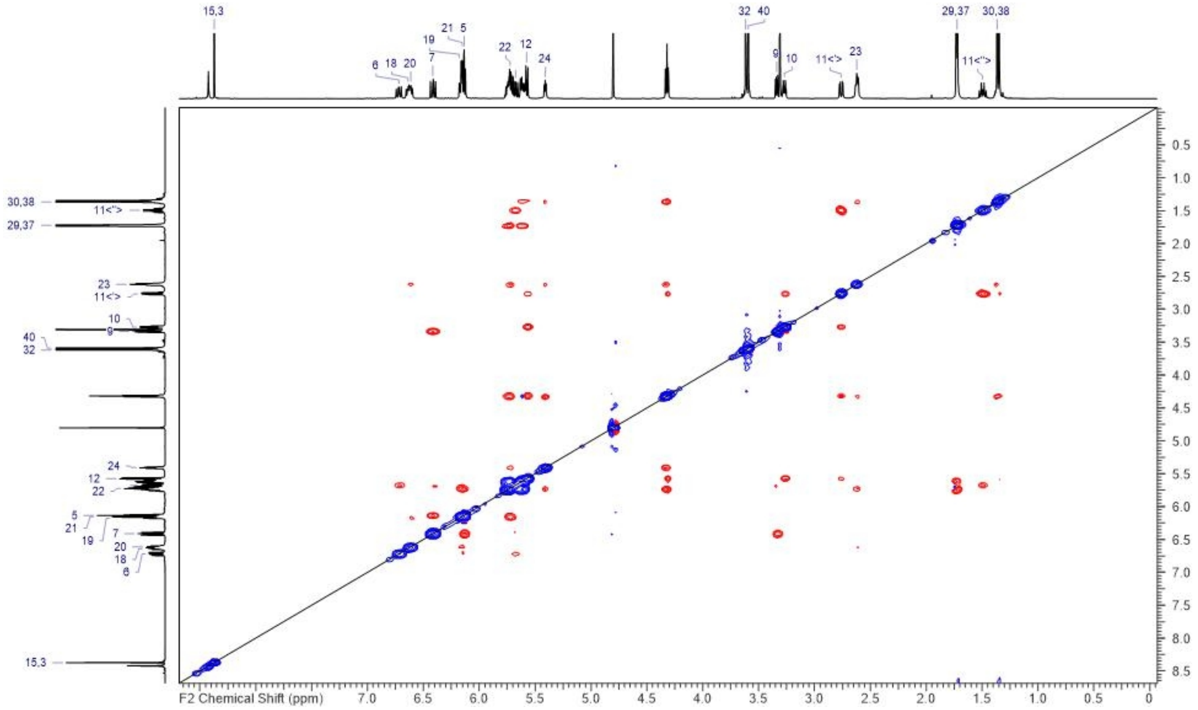


Figure S33 ^{1}H,^{1}H-ROESY NMR spectrum of 9,10-epoxy-disorazole Z (**7**) in methanol-d_4.

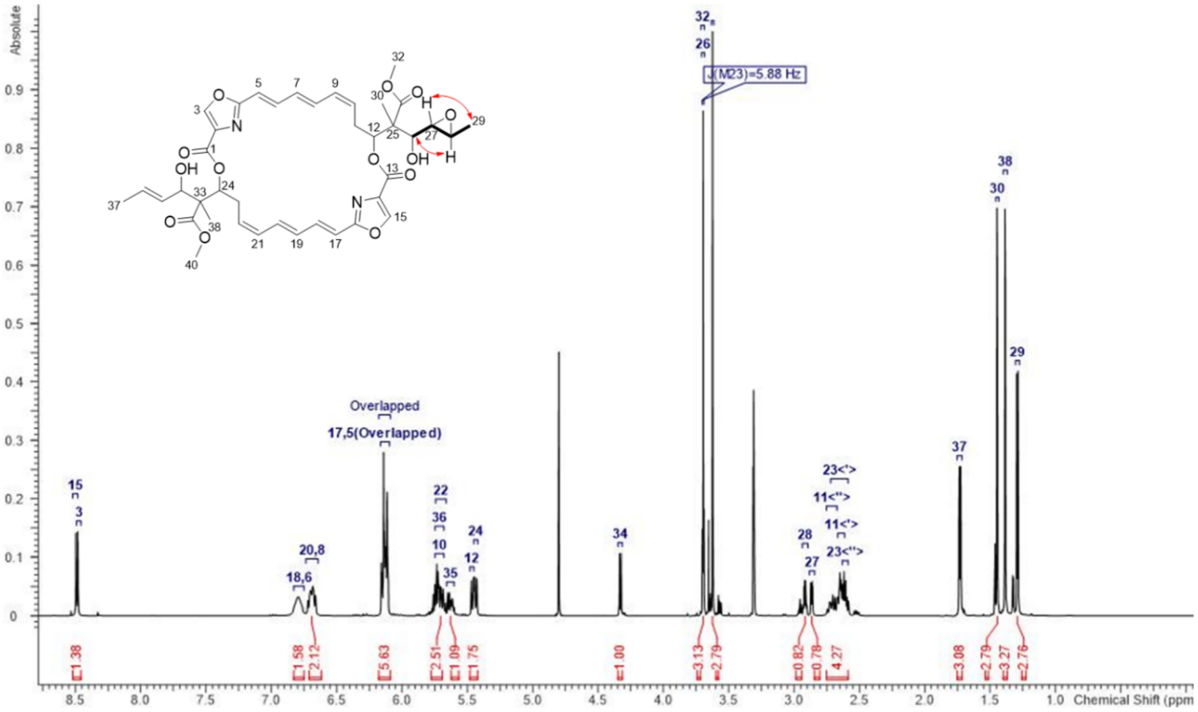


Figure S34 ^{1}H NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4 (600 MHz).

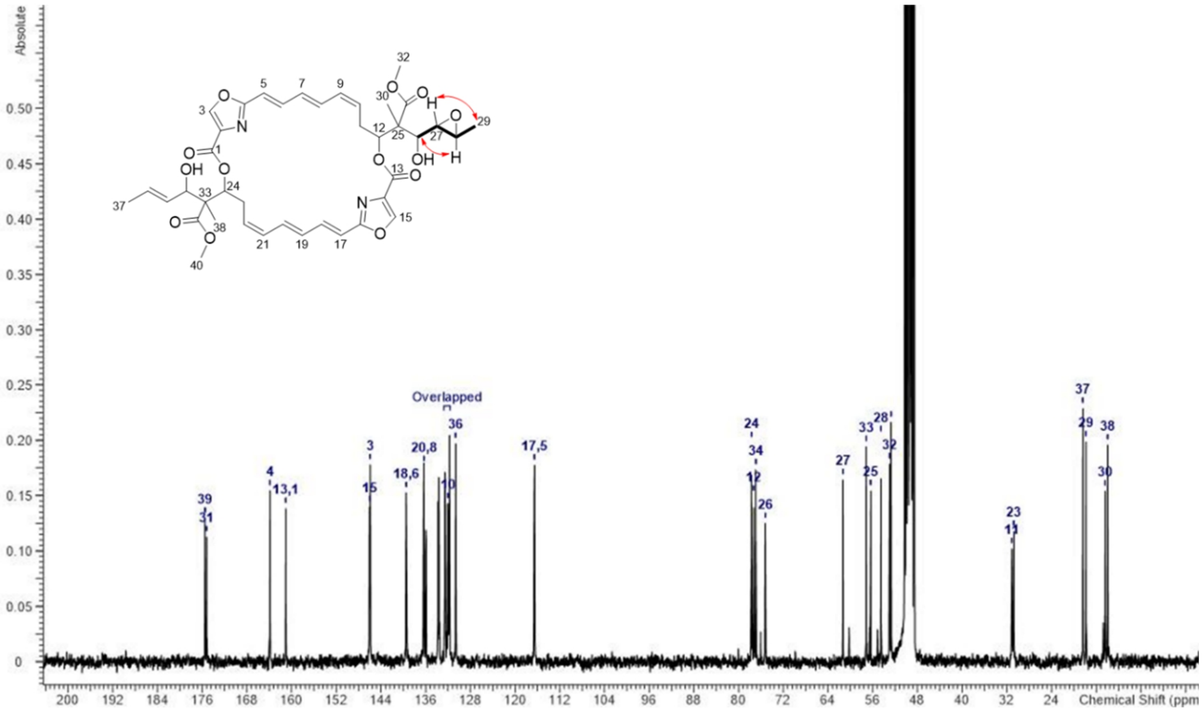


Figure S35 ^{13}C NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4 (150 MHz).

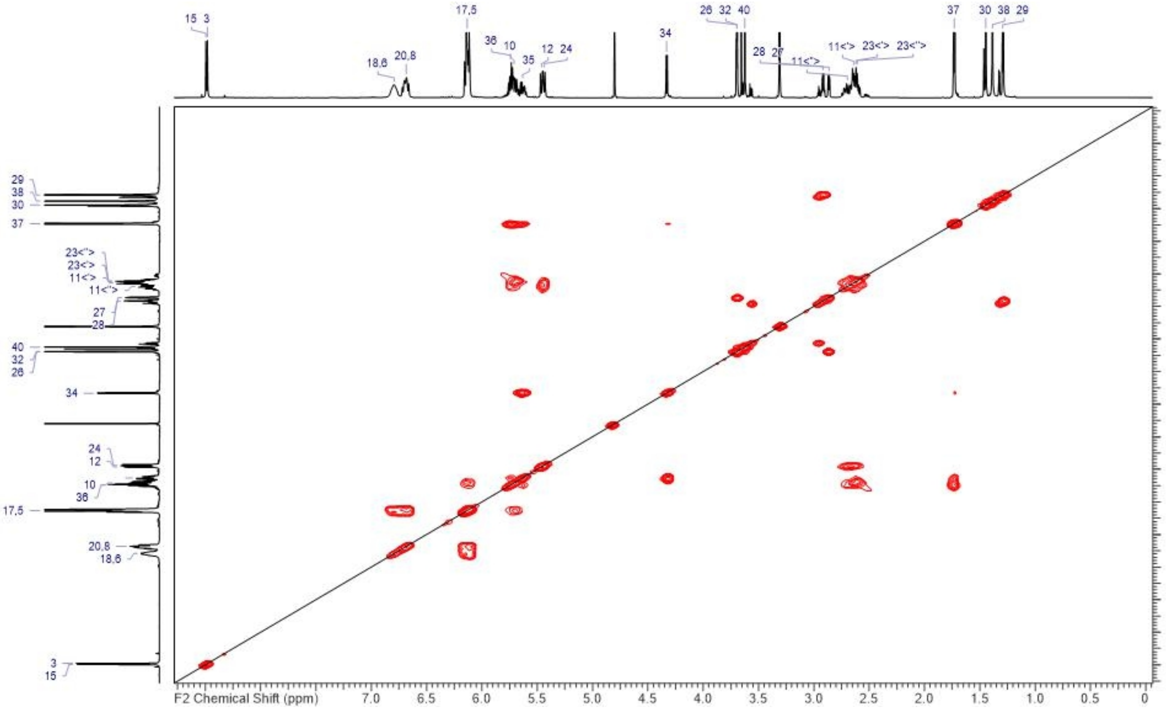


Figure S36 ^{1}H,^{1}H-COSY NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4.

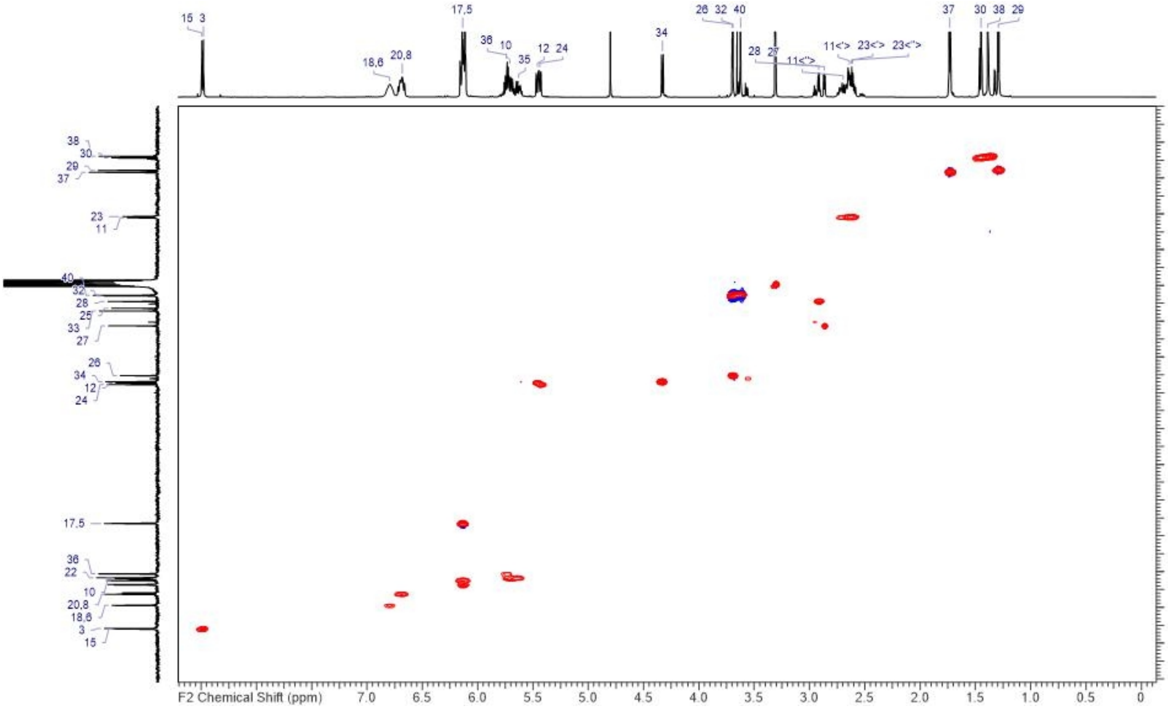


Figure S37 HMQC-NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4.

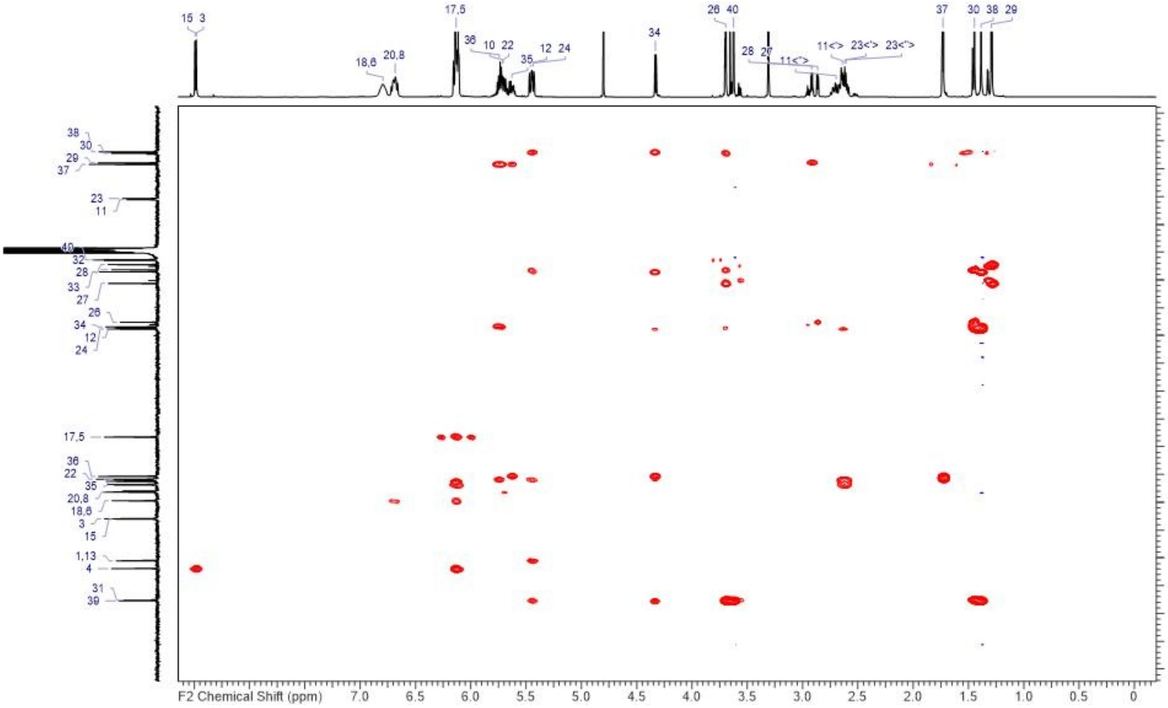


Figure S38 HMBC NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4.

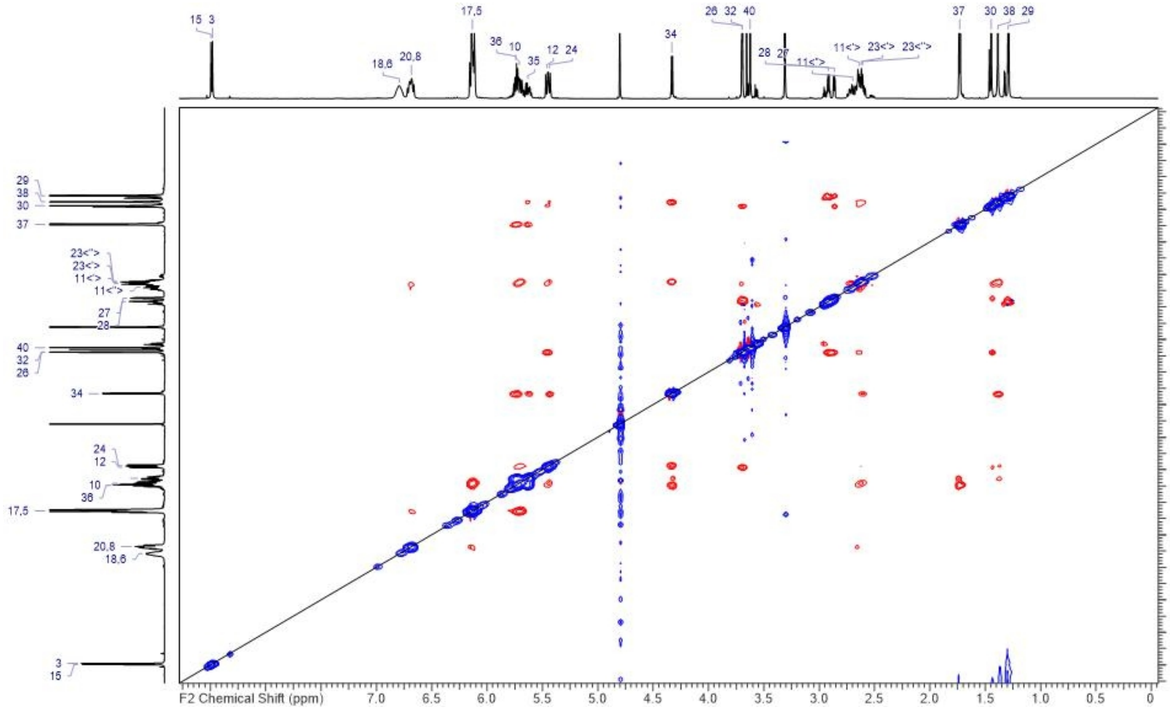


Figure S39 ^{1}H,^{1}H-ROESY NMR spectrum of 27,28-epoxy-disorazole Z (**8**) in methanol-d_4.

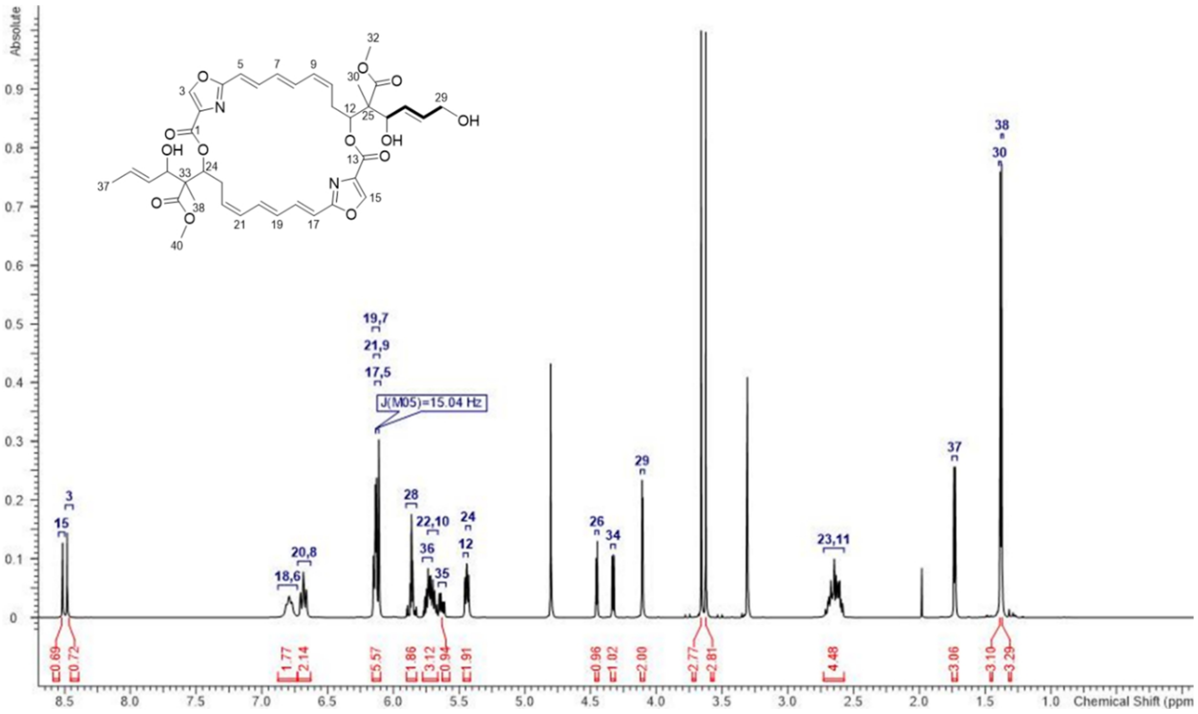


Figure S40 ^{1}H NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-d_4 (600 MHz).

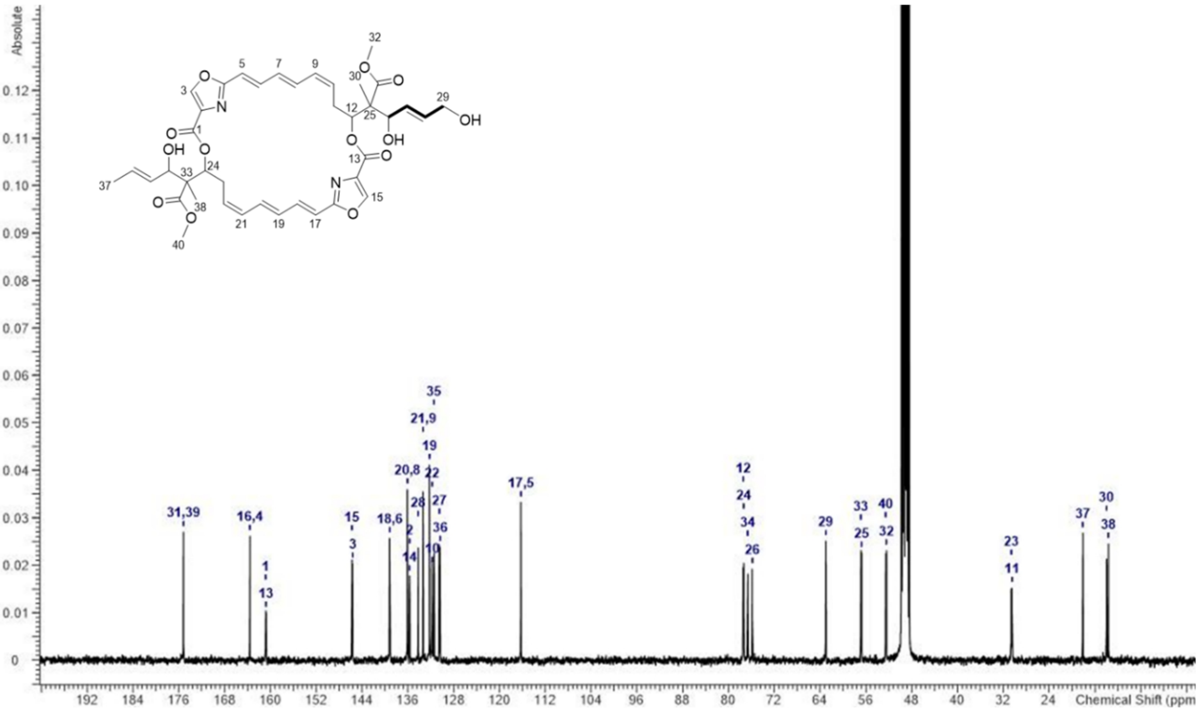


Figure S41 ^{13}C NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-*d*4 (150 MHz).

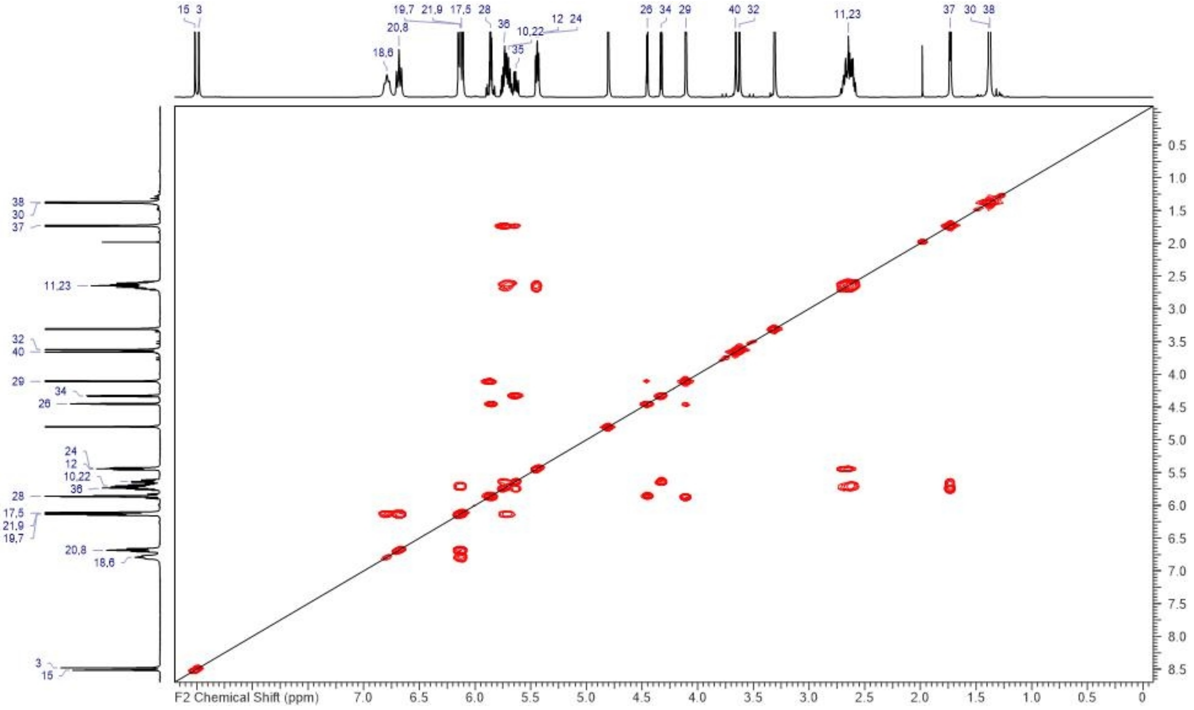


Figure S42 ^{1}H,^{1}H-COSY NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-*d*4.

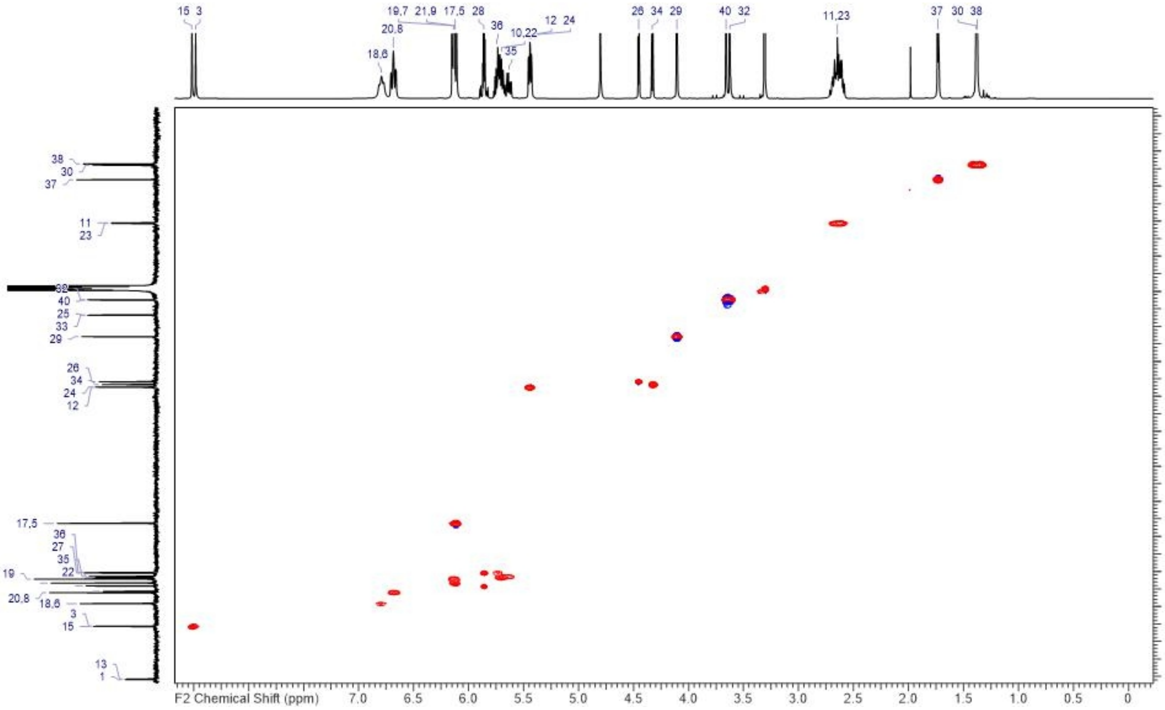


Figure S43 HMQC-NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-*d*4.

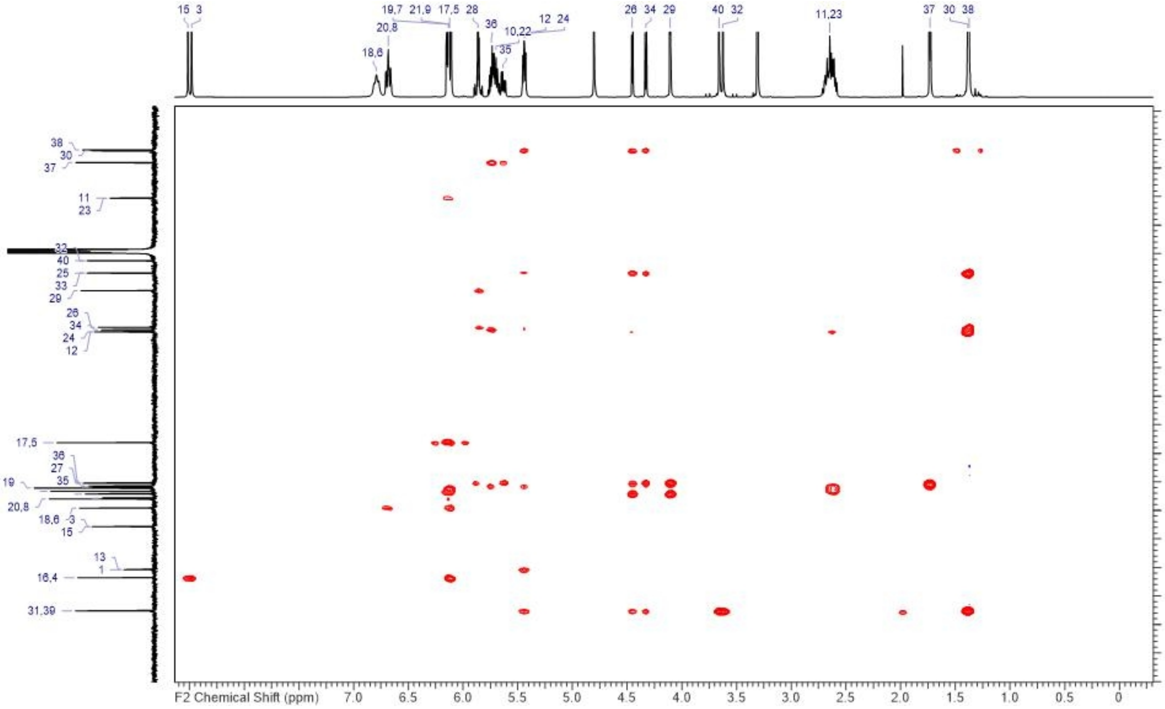


Figure S44 HMBC NMR spectrum of 29-hydroxy-disorazole Z (**9**) in methanol-*d*4.

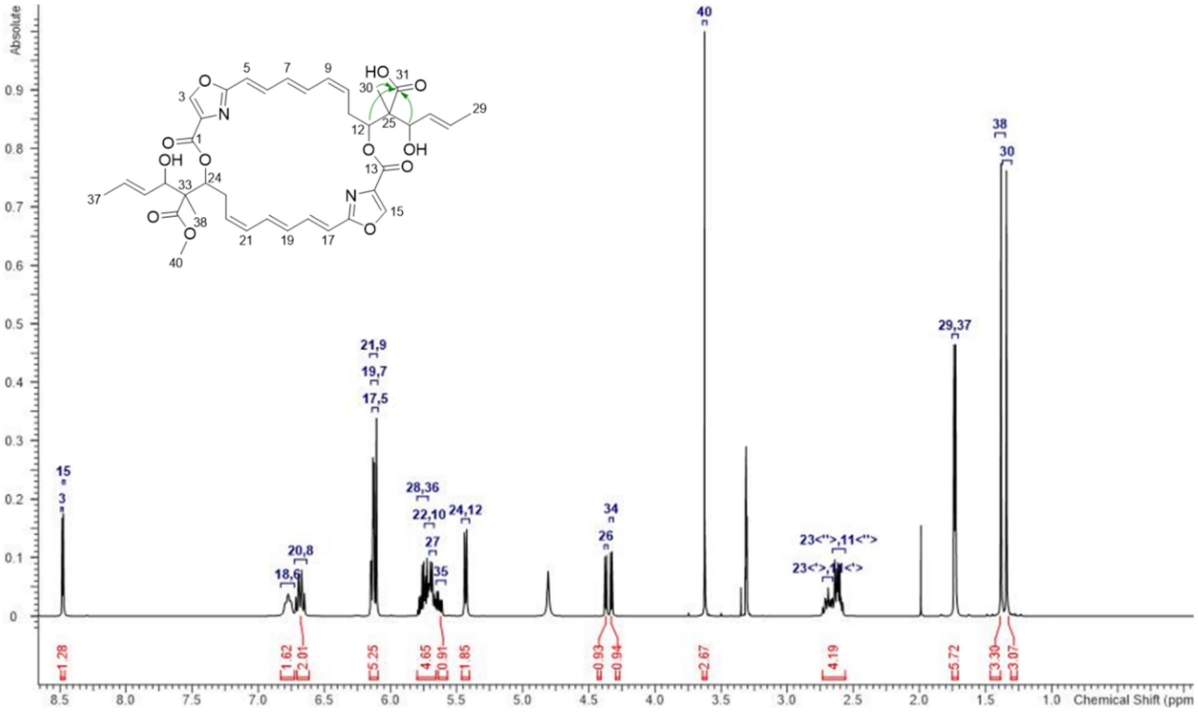


Figure S45 ^{1}H NMR spectrum of 31-*O*-desmethyl-disorazole Z (**10**) in methanol-d_4 (600 MHz).

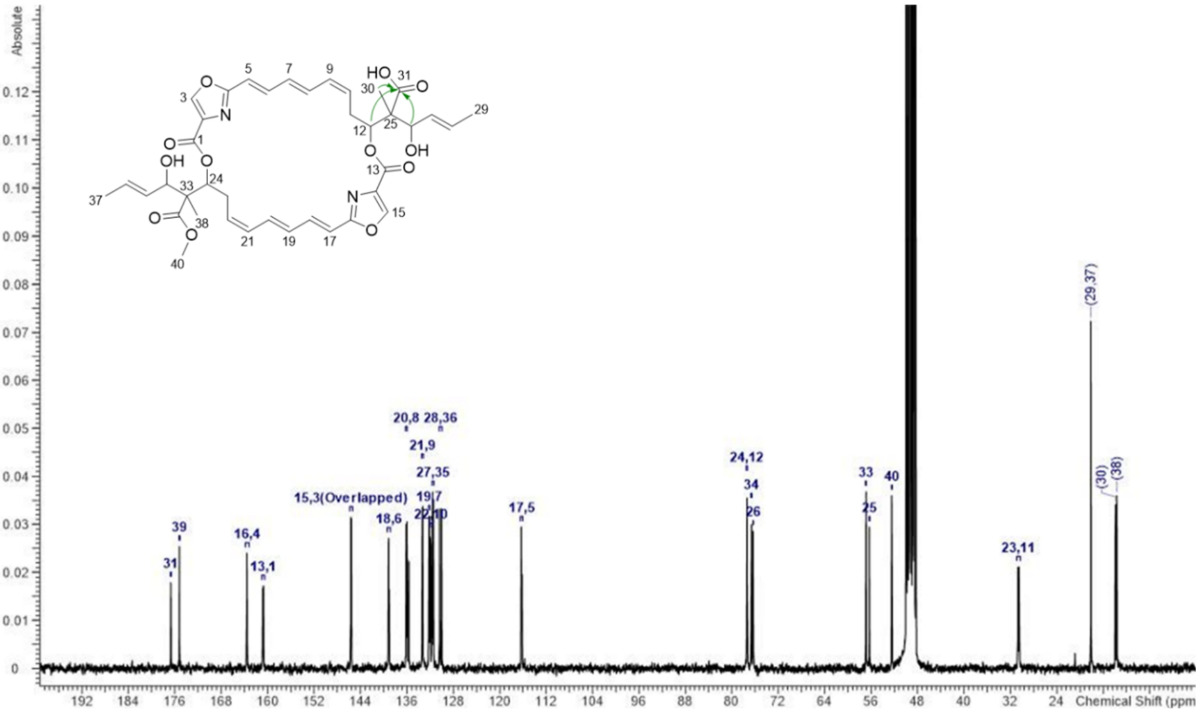


Figure S46 ^{13}C NMR spectrum of 31-*O*-desmethyl-disorazole Z (**10**) in methanol-d_4 (150 MHz).

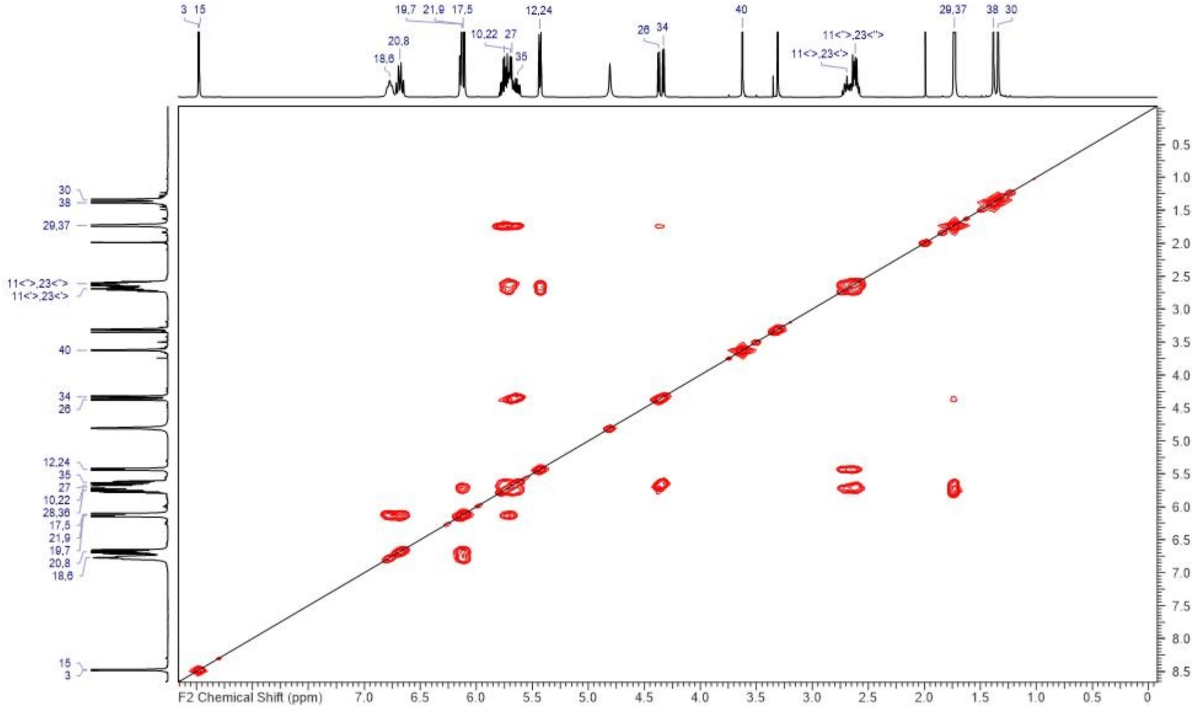


Figure S47 ^{1}H,^{1}H-COSY NMR spectrum of 31-*O*-desmethyl-disorazole Z (**10**) in methanol-d_4.

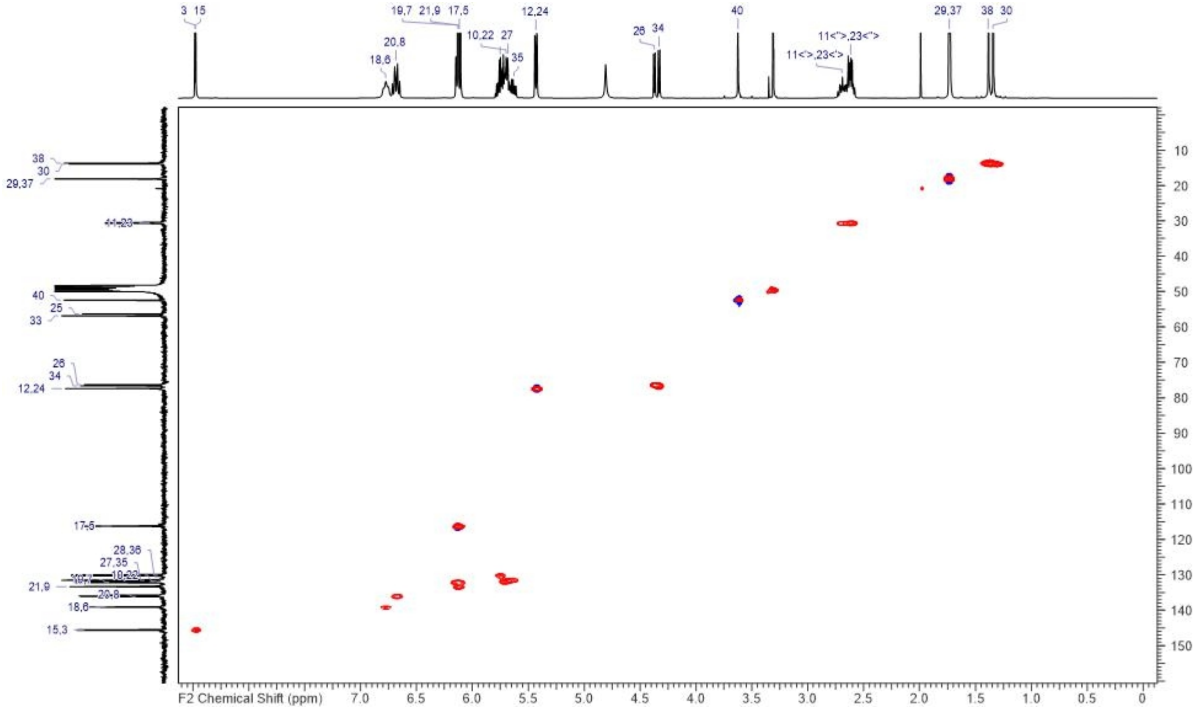


Figure S48 HMQC-NMR spectrum of 31-*O*-desmethyl-disorazole Z (**10**) in methanol-d_4.

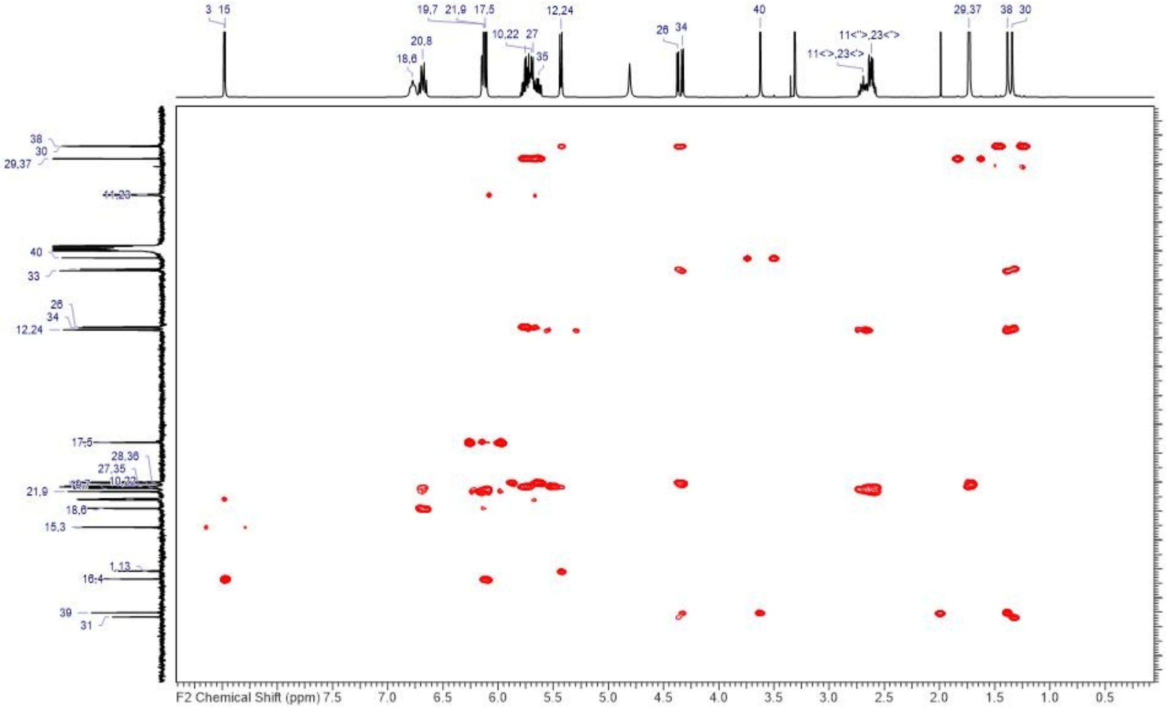


Figure S49 HMBC NMR spectrum of 31-*O*-desmethyl-disorazole Z (**10**) in methanol-d_4.

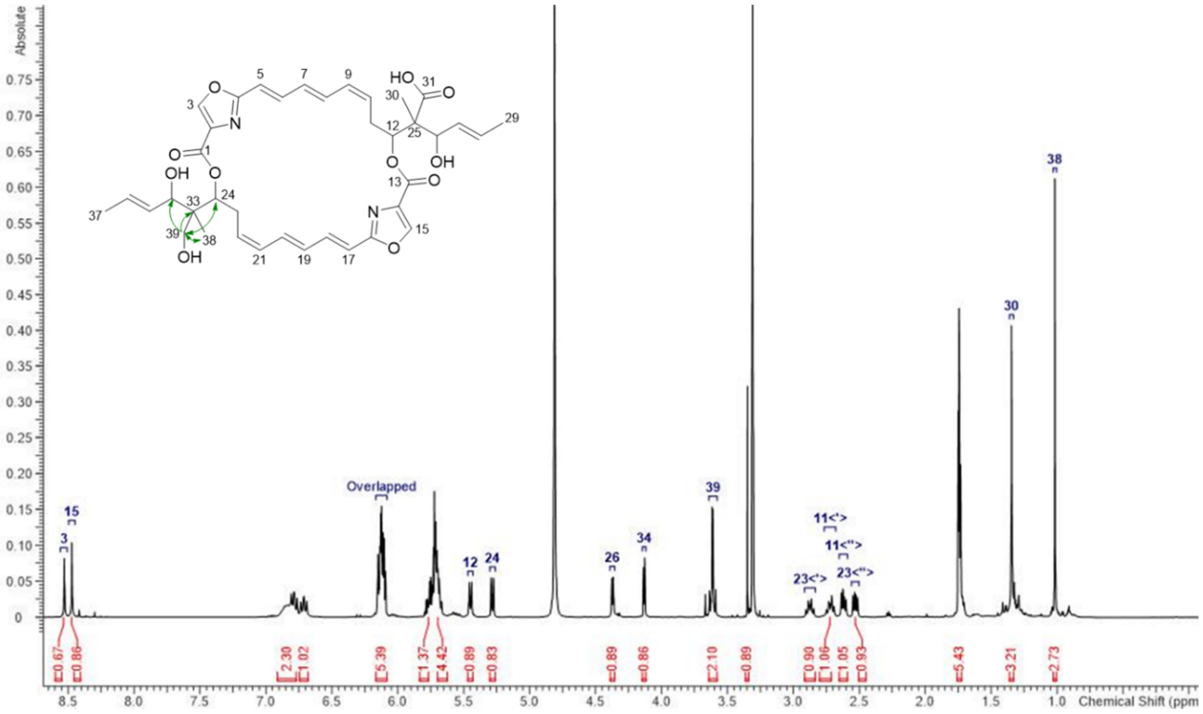


Figure S50 ^{1}H NMR spectrum of 31-*O*-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-d_4 (600 MHz).

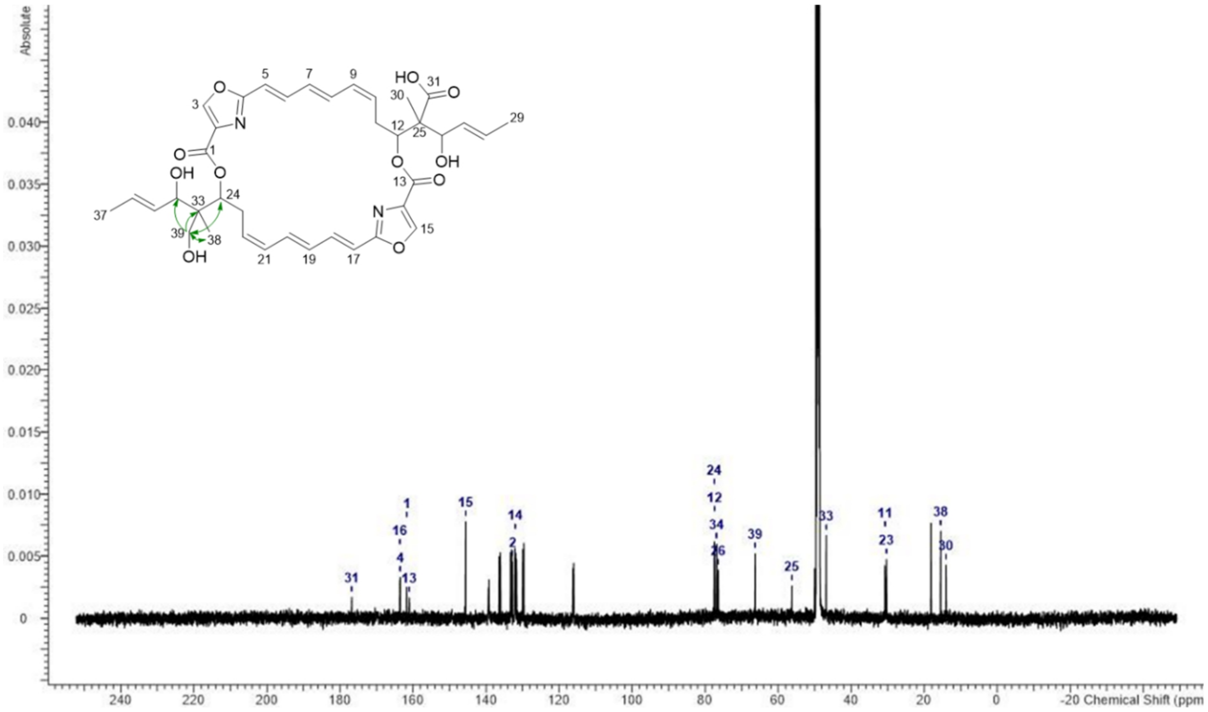


Figure S51 ^{13}C NMR spectrum of 31-*O*-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-*d*4 (150 MHz).

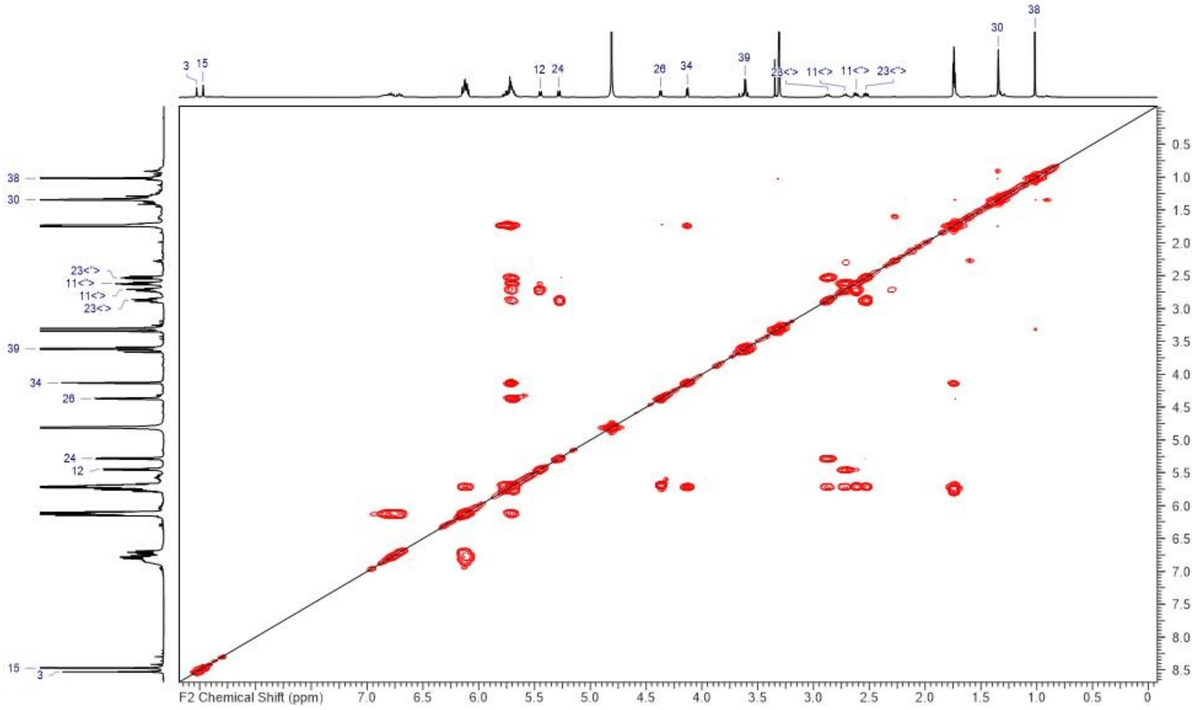


Figure S52 ^{1}H,^{1}H-COSY NMR spectrum of 31-*O*-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-*d*4.

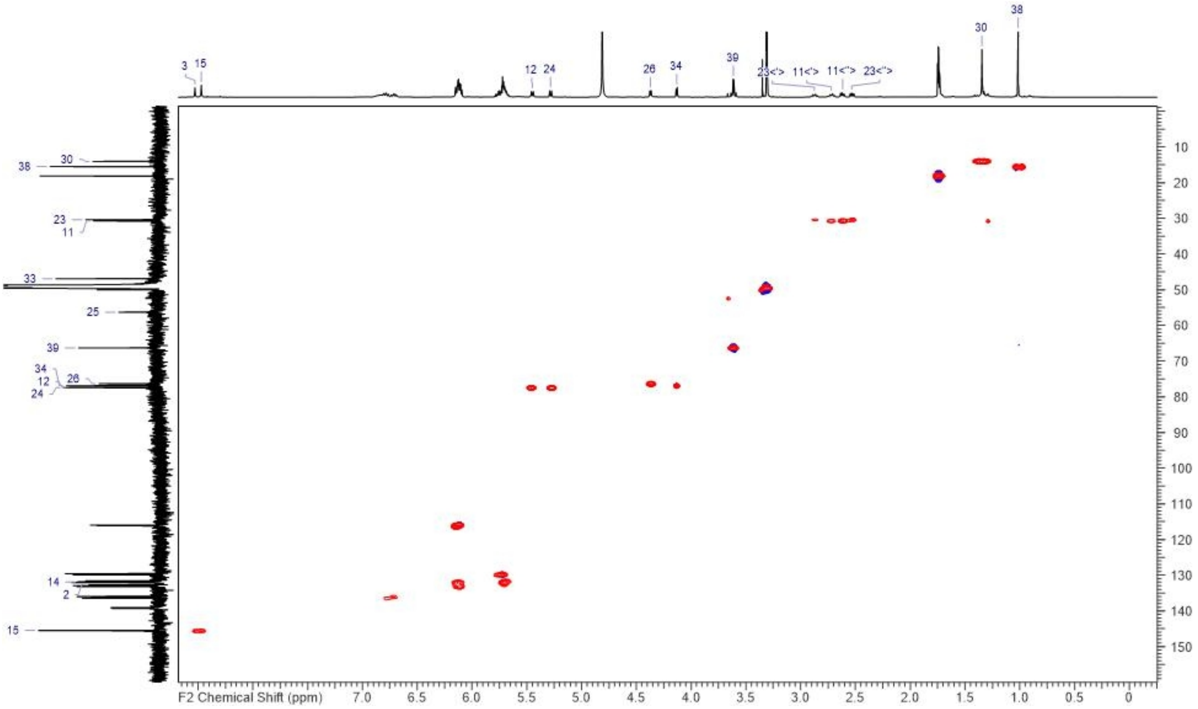


Figure S53 HMQC-NMR spectrum of 31-*O*-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-d_4.

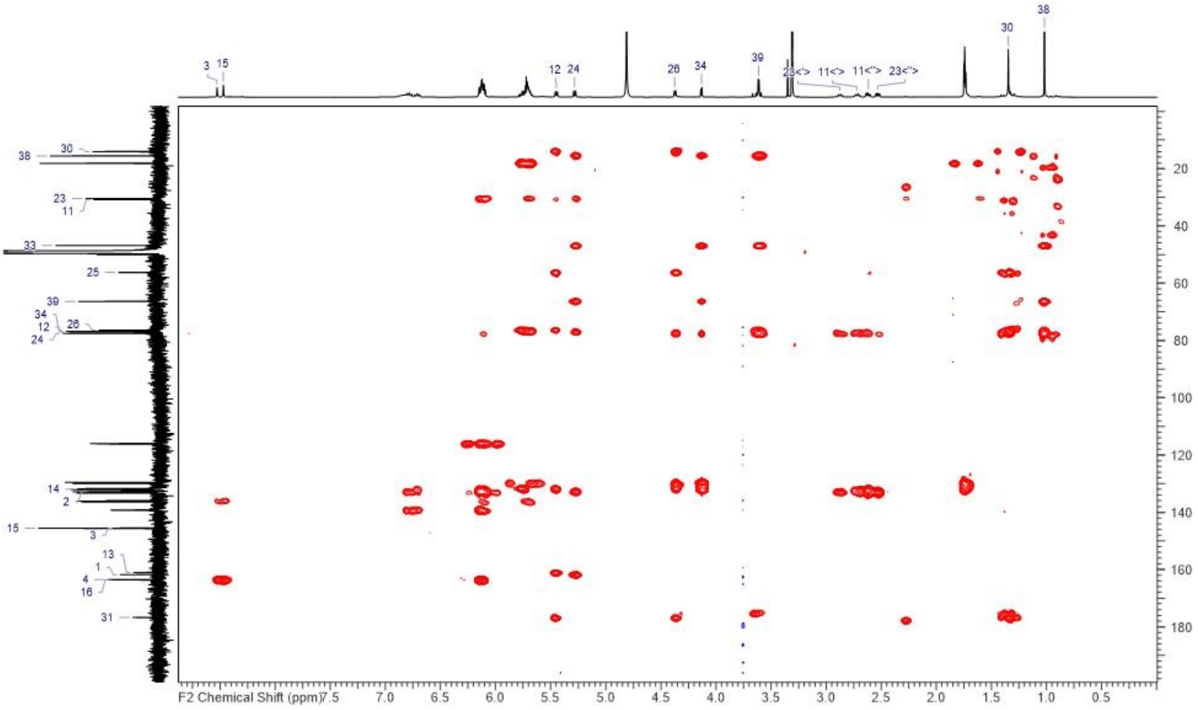


Figure S54 HMBC NMR spectrum of 31-*O*-desmethyl-39-hydroxy-disorazole Z (**11**) in methanol-d_4.

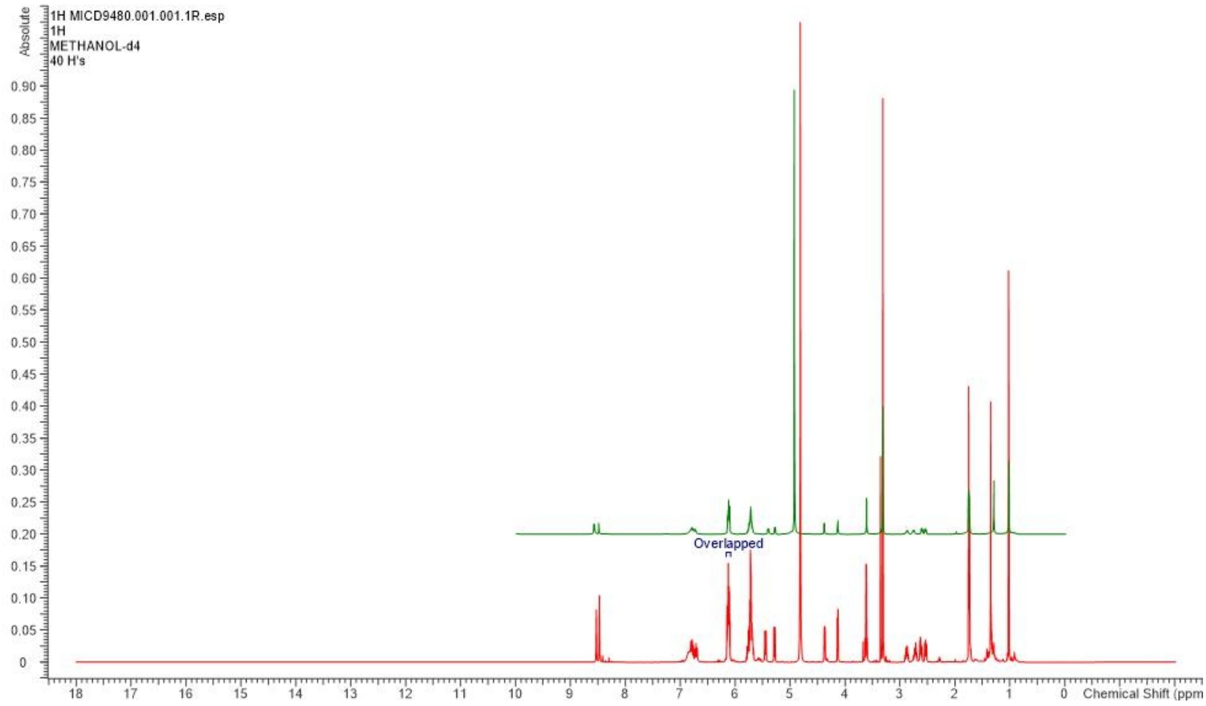


Figure S55 Comparison of ^{1}H NMR spectra of **11** in methanol-d_4 isolated from *S. cellulosum* and *M. xanthus*. Bottom/red: isolated from *S. cellulosum* So ce1875; Top/green: isolated from *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF.

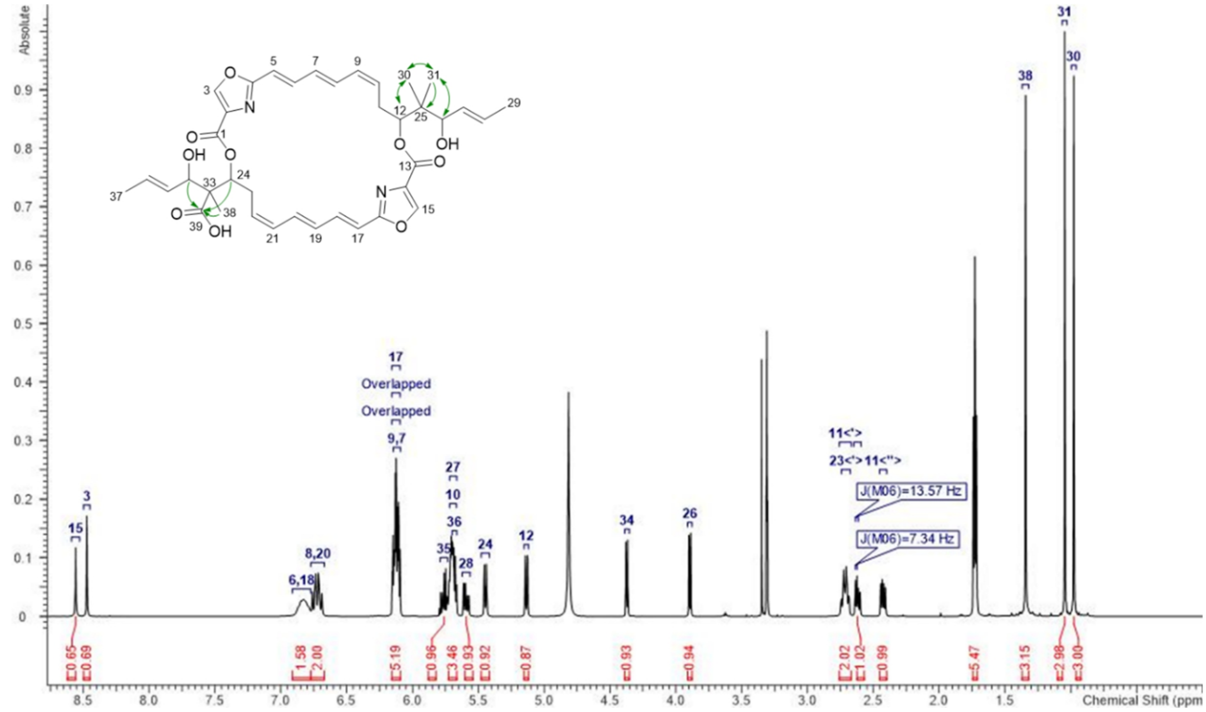


Figure S56 ^{1}H NMR spectrum of *O*-desmethyl-dimethyl-disorazole Z (**12**) in methanol-d_4 (600 MHz).

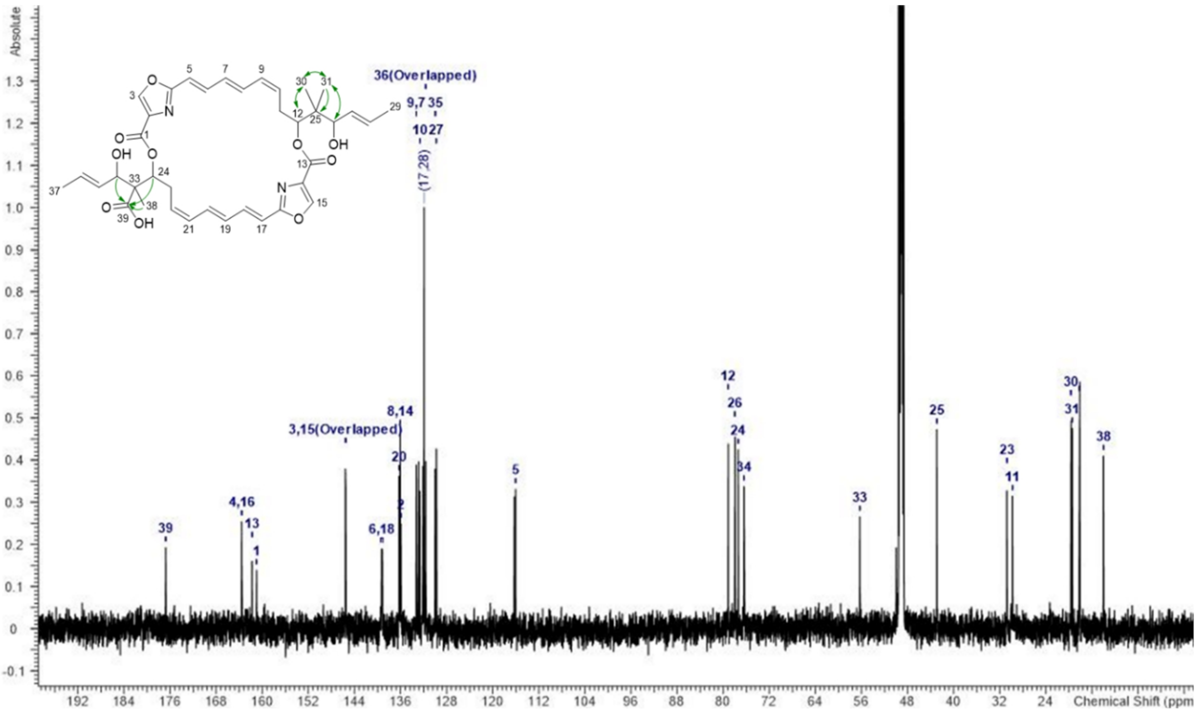


Figure S57 ^{13}C NMR spectrum of *O*-desmethyl-dimethyl-disorazole Z (**12**) in methanol-*d*$_4$ (150 MHz).

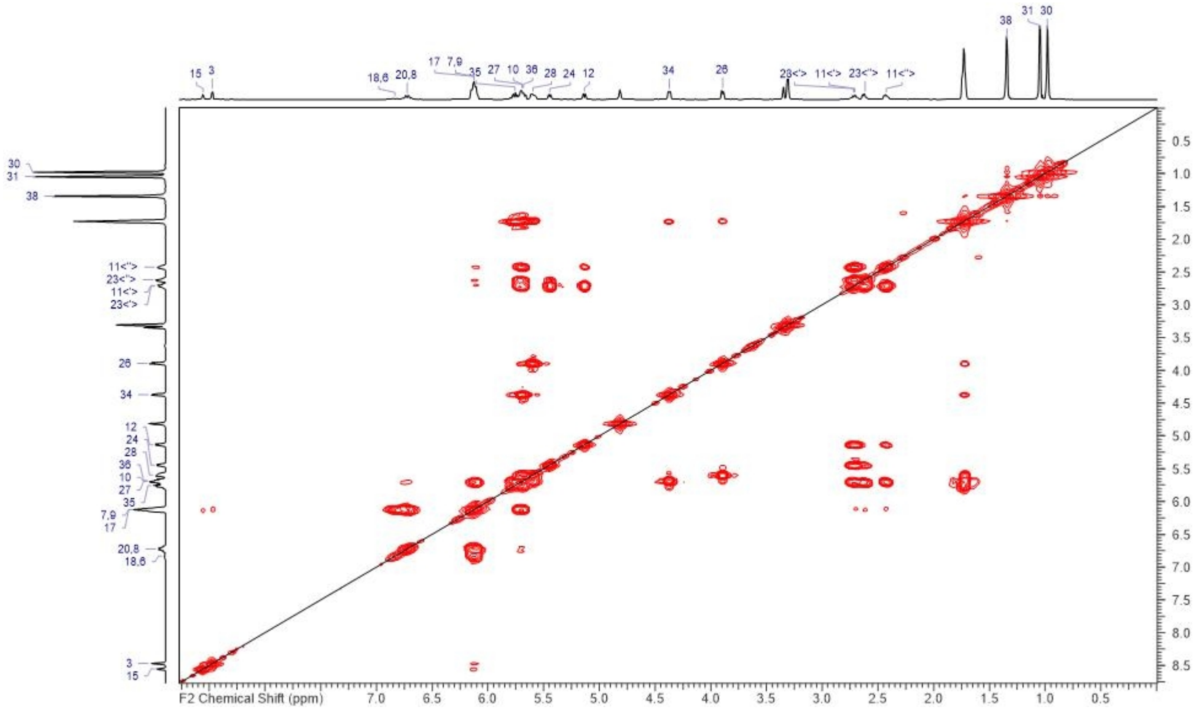


Figure S58 ^{1}H,^{1}H-COSY NMR spectrum of *O*-desmethyl-dimethyl-disorazole Z (**12**) in methanol-*d*$_4$.

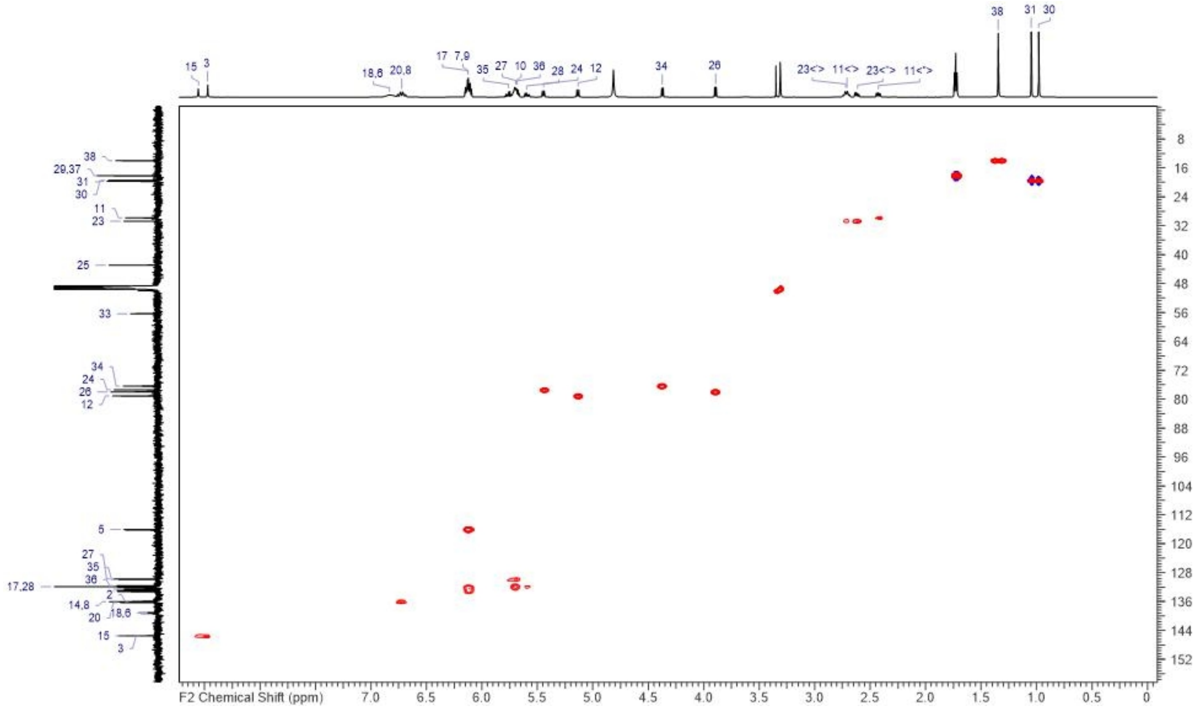


Figure S59 HMQC-NMR spectrum of *O*-desmethyl-dimethyl-disorazole Z (**12**) in methanol-d_4.

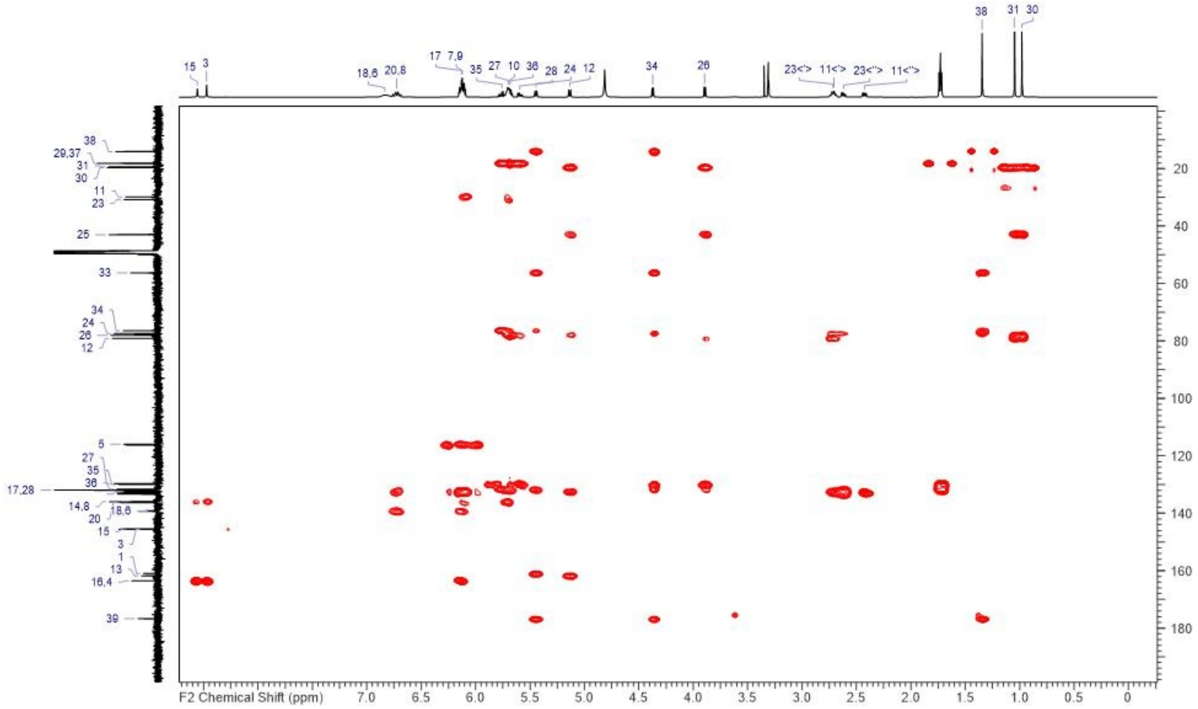


Figure S60 HMBC NMR spectrum of *O*-desmethyl-dimethyl-disorazole Z (**12**) in methanol-d_4.

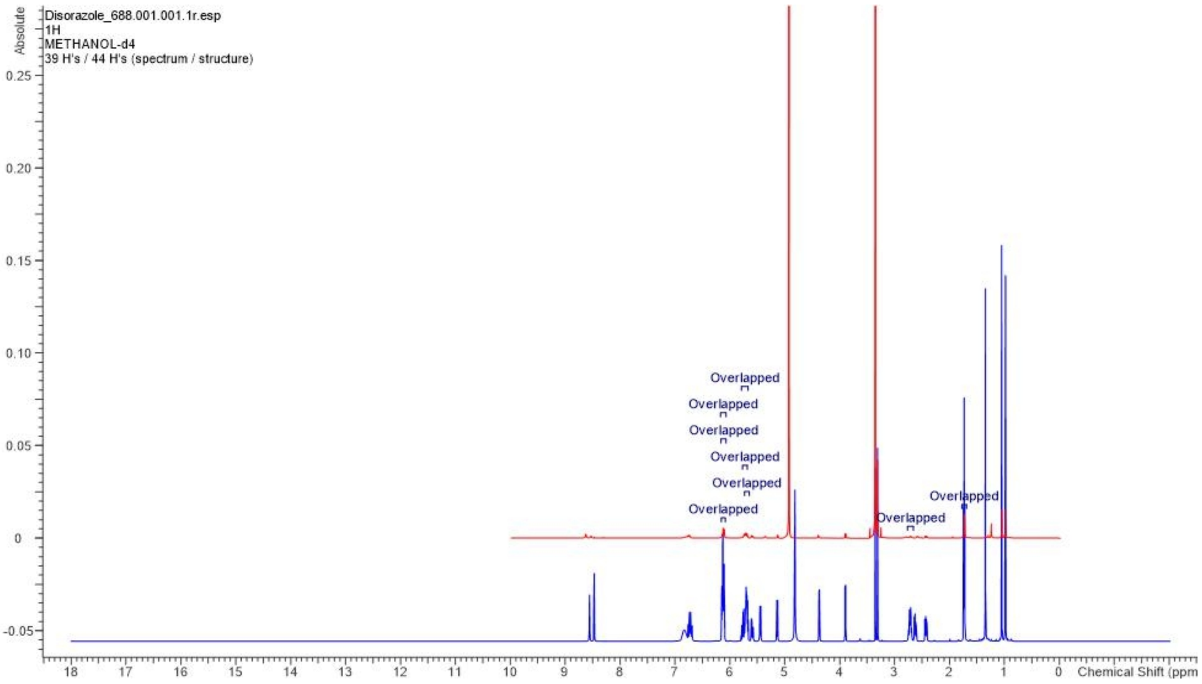


Figure S61 Comparison of ^{1}H NMR spectra of **12** in methanol-d_4 isolated from *S. cellulosum* and *M. xanthus*. Bottom/blue: isolated from *S. cellulosum* So ce1875; Top/red: isolated from *M. xanthus* DK1622::km-int-Ptet-dis427-gent-delF.

2.7 Acknowledgements

The authors greatly acknowledge Prof. Marc Stadler, Reinhard Sterlinski, Axel Schulz, Wolfgang Kessler and Kerstin Schober for their contributions to large-scale fermentation and downstream processing, especially Steffen Bernecker for establishing and monitoring the fermentation process. The authors thank Volker Huch for crystallographic analysis, instrumentation and technical assistance.

2.8 References

[1] K. Gerth, S. Pradella, O. Perlova et al., Myxobacteria: proficient producers of novel natural products with various biological activities—past and future biotechnological aspects with the focus on the genus *Sorangium*, *J Biotechnol*, **2003**, 106(2-3), 233-253.

[2] S. C. Wenzel and R. Müller, Myxobacteria—'microbial factories' for the production of bioactive secondary metabolites, *Mol BioSyst*, **2009**, 5(6), 567-574.

[3] K. J. Weissman and R. Müller, Myxobacterial secondary metabolites: bioactivities and modes-of-action, *Nat Prod Rep*, **2010**, 27(9), 1276-1295.

[4] J. Herrmann, A. A. Fayad, R. Müller, Natural products from myxobacteria: novel metabolites and bioactivities, *Nat Prod Rep*, **2017**, 34(2), 135-160.

[5] H. Irschik, R. Jansen, K. Gerth et al., Disorazol A, an efficient inhibitor of eukaryotic organisms isolated from myxobacteria, *J Antibiot (Tokyo)*, **1995**, 48(1), 31-35.

[6] C. D. Hopkins and P. Wipf, Isolation, biology and chemistry of the disorazoles: new anti-cancer macrodiolides, *Nat Prod Rep*, **2009**, 26(5), 585-601.

[7] K. Rox, M. Rohde, G. S. Chhatwal et al., Disorazoles Block Group A Streptococcal Invasion into Epithelial Cells Via Interference with the Host Factor Ezrin, *Cell Chem Biol*, **2017**, 24(2), 159-170.

[8] R. Jansen, H. Irschik, H. Reichenbach et al., Disorazoles, Highly Cytotoxic Metabolites from the Sorangicin-Producing Bacterium *Sorangium cellulosum*, Strain So ce12, *Liebigs Ann Chem*, **1994**, 1994(8), 759-773.

[9] Y. A. Elnakady, F. Sasse, H. Lünsdorf et al., Disorazol A1, a highly effective antimitotic agent acting on tubulin polymerization and inducing apoptosis in mammalian cells, *Biochem Pharmacol*, **2004**, 67(5), 927-935.

[10] R. Carvalho, R. Reid, N. Viswanathan et al., The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation, *Gene*, **2005**, 359, 91-98.

[11] M. Kopp, H. Irschik, S. Pradella et al., Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes, *ChemBioChem*, **2005**, 6(7), 1277-1286.

[12] J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2010**, 27(7), 996-1047.

[13] E. J. N. Helfrich and J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2016**, 33(2), 231-316.

[14] Q. Tu, J. Herrmann, S. Hu et al., Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering, *Sci Rep*, **2016**, 6, 21066.

[15] H. Irschik, R. Jansen, F. Sasse, Patent EP1904505B1, **2013**.

[16] E. Guenther, O. Schaefer, M. Teifel et al., Patent EP2066679B1, **2016**.

[17] B. Aicher, K. Hirschfelder, R. Jansen et al., Disorazol Z: A highly cytotoxic natural compound with antitumor properties, *Mol Cancer Ther*, **2011**, 10(11_Supplement), C214.

[18] E. Guenther, K. Paulini, B. Aicher et al., Targeting Disorazol Z to LHRH-receptor positive tumors by the cytotoxic conjugate AEZS-125, *Cancer Res*, **2008**, 68(9_Supplement), 782.

[19] B. Aicher, T. Schuster, L. Blumenstein et al., Highly Potent Cytotoxic Conjugates of Disorazol Z Linked to a LHRH Receptor Targeting Peptide Interfere with Cell Cycle Progression in Human Cancer Cell Lines and Suppress Tumor Growth in a LHRH Receptor Positive Ovarian Cancer Xenograft Model, *Eur J Cancer*, **2012**, 48(48), 101.

[20] S. Seitz, S. Buchholz, A. V. Schally et al., Triple negative breast cancers express receptors for LHRH and are potential therapeutic targets for cytotoxic LHRH-analogs, AEZS 108 and AEZS 125, *BMC Cancer*, **2014**, 14, 847.

[21] G. Menchon, A. E. Prota, D. Lucena-Agell et al., A fluorescence anisotropy assay to discover and characterize ligands targeting the maytansine site of tubulin, *Nat Commun*, **2018**, 9(1), 2106.

[22] R. Schäckel, B. Hinkelmann, F. Sasse et al., The synthesis of novel disorazoles, *Angew Chem Int Ed*, **2010**, 49(9), 1619-1622.

[23] K. C. Nicolaou, G. Bellavance, M. Buchman et al., Total Syntheses of Disorazoles A1 and B1 and Full Structural Elucidation of Disorazole B1, *J Am Chem Soc*, **2017**, 139(44), 15636-15639.

[24] M. Miethke, M. Pieroni, T. Weber et al., Towards the sustainable discovery and development of new antibiotics, *Nat Rev Chem*, **2021**, 5(10), 1-24.

[25] S. E. Ongley, X. Bian, B. A. Neilan et al., Recent advances in the heterologous expression of microbial natural product biosynthetic pathways, *Nat Prod Rep*, **2013**, 30(8), 1121-1138.

[26] L. Huo, J. J. Hug, C. Fu et al., Heterologous expression of bacterial natural product biosynthetic pathways, *Nat Prod Rep*, **2019**, 36(10), 1412-1436.

[27] H. Hiemstra, H. A. Houwing, O. Possel et al., Carbon-13 nuclear magnetic resonance spectra of oxazoles, *Can J Chem*, **1979**, 57(23), 3168-3170.

[28] J. A. Dale, D. L. Dull, H. S. Mosher, α-Methoxy-α-trifluoromethylphenylacetic acid, a versa-tile reagent for the determination of enantiomeric composition of alcohols and amines, *J Org Chem*, **1969**, 34(9), 2543-2549.

[29] G. L. Lemière, J. J. Willaert, R. A. Dommisse et al., Determination of the absolute configuration and enantiomeric purity of alcohols from the ^{13}C-NMR spectra of the corresponding MTPA esters, *Chirality*, **1990**, 2(3), 175-184.

[30] T. Pehk, E. Lippmaa, M. Lopp et al., Determination of the absolute configuration of chiral secondary alcohols; new advances using ^{13}C- and 2D-NMR spectroscopy, *Tetrahedron: Asymmetry*, **1993**, 4(7), 1527-1532.

[31] T. R. Hoye, C. S. Jeffrey, F. Shao, Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons, *Nat Protoc*, **2007**, 2(10), 2451-2458.

[32] M. H. Medema, K. Blin, P. Cimermancic et al., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, *Nucleic Acids Res*, **2011**, 39(suppl_2), W339-W346.

[33] T. Robbins, J. Kapilivsky, D. E. Cane et al., Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase, *Biochemistry*, **2016**, 55(32), 4476-4484.

[34] H.-Y. He, M.-C. Tang, F. Zhang et al., *Cis*-Double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis, *J Am Chem Soc*, **2014**, 136(12), 4488-4491.

[35] C. G. Marshall, M. D. Burkart, T. A. Keating et al., Heterocycle formation in vibriobactin biosynthesis: alternative substrate utilization and identification of a condensed intermediate, *Biochemistry*, **2001**, 40(35), 10655-10663.

[36] M. Di Lorenzo, M. Stork, H. Naka et al., Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in *Vibrio anguillarum*, *Biometals*, **2008**, 21(6), 635-648.

[37] C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey et al., Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization, *Chem Biol*, **1999**, 6(6), 385-400.

[38] Y. Zhou, P. Prediger, L. C. Dias et al., Macrodiolide formation by the thioesterase of a modular polyketide synthase, *Angew Chem Int Ed*, **2015**, 54(17), 5232-5235.

[39] Y. Zhou, A. C. Murphy, M. Samborskyy et al., Iterative Mechanism of Macrodiolide Formation in the Anticancer Compound Conglobatin, *Chem Biol*, **2015**, 22(6), 745-754.

[40] J. Fu, X. Bian, S. Hu et al., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting, *Nat Biotechnol*, **2012**, 30(5), 440-446.

[41] V. Magrini, C. Creighton, P. Youderian, Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration, *J Bacteriol*, **1999**, 181(13), 4050-4061.

[42] H. Wang, Z. Li, R. Jia et al., RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression, *Nat Protoc*, **2016**, 11(7), 1175-1190.

[43] M. A. Jordan, Mechanism of action of antitumor drugs that interact with microtubules and tubulin, *Curr Med Chem Anticancer Agents*, **2002**, 2(1), 1-17.

[44] B. Short, Acetylated microtubules let the ER slide, *J Cell Biol*, **2010**, 190(3), 280.

[45] A. S. Lee, The ER chaperone and signaling regulator GRP78/BiP as a monitor of endoplasmic reticulum stress, *Methods*, **2005**, 35(4), 373-381.

[46] N. André, D. Braguer, G. Brasseur et al., Paclitaxel induces release of cytochrome *c* from mitochondria isolated from human neuroblastoma cells, *Cancer Res*, **2000**, 60(19), 5349-5353.

[47] G. M. Sheldrick, A short history of SHELX, *Acta Crystallogr A*, **2008**, A64(Pt 1), 112-122.

[48] G. M. Sheldrick, Crystal structure refinement with SHELXL, *Acta Crystallogr C*, **2015**, C71(Pt 1), 3-8.

[49] C. B. Hübschle, G. M. Sheldrick, B. Dittrich, ShelXle: a Qt graphical user interface for SHELXL, *J Appl Crystallogr*, **2011**, 44(Pt 6), 1281-1284.

[50] H. Liu and J. H. Naismith, A simple and efficient expression and purification system using two newly constructed vectors, *Protein Expr Purif*, **2009**, 63(2), 102-111.

[51] V. Y. Alekseyev, C. W. Liu, D. E. Cane et al., Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase, *Protein Sci*, **2007**, 16(10), 2093-2107.

Chapter 3. Unusual oxidations mediated by a Coq4 domain-containing hypothetical protein in disorazole biosynthesis

Manuscript (Research Article) in preparation

Yunsheng Gao,[a,b] Joy Birkelbach,[a] Asfandyar Sikandar,[a] Youming Zhang,[a,b] and
Rolf Müller*[a,b]

[a] Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

[b] Helmholtz International Lab for Anti-Infectives, Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, 266237 Qingdao, China

[*] To whom correspondence should be addressed: rolf.mueller@helmholtz-hips.de

Note: the author list is tentative according to current contributions.

Keywords

disorazole, biosynthesis, Coq4, oxidation, heterologous expression, P450

Author's efforts:

The author designed the research, performed *in silico* analysis and all molecular biology experiments, including promoter engineering, gene knockout, gene swapping, and heterologous expression. The author performed fermentation, metabolome analysis, compounds purification, and substrate feeding experiments. The author wrote the manuscript with inputs from other authors.

Other's efforts:

Joy Birkelbach made significant efforts to NMR measurements and structure elucidation. Youming Zhang contributed to supervision of the research. Rolf Müller supervised the research and revised the manuscript.

Signatures:

Yunsheng Gao: Yunsheng Gao

Joy Birkelbach:

Asfandyar Sikandar:

Youming Zhang:

Rolf Müller:

3.1 Abstract

Disorazoles are a family of macrodilactone natural products with promising anti-cancer bioactivity which are specifically produced by myxobacteria, especially by strains of *Sorangium cellulosum*. Two types of disorazoles, i.e., disorazole A and disorazole Z, together with their corresponding biosynthetic gene clusters, were discovered and characterized. However, biosynthesis of the epoxy group in disorazole A and oxidations of the geminal dimethyl group in disorazole Z to a carboxylic acid remained enigmatic. Here, we report that the hypothetical protein DisE that contains a Coq4 domain is essential for these tailoring modifications. In the absence of DisE, five disorazole intermediates/derivatives accumulated using the heterologous expression system in *Myxococcus xanthus* DK1622 and further characterized by NMR. Using this highly conserved DisE as a probe, 14 additional disorazole biosynthetic gene clusters were discovered from our in-house myxobacterial genome database. Furthermore, we identified cytochrome P450-dependent enzymes from the disorazole producers as well as the heterologous host *M. xanthus* DK1622 that likely collaborate with DisE for the site-specific oxidations. The results presented in this study provide a method for targeted production of disorazole intermediates/derivatives and, in addition, pave the way for in-depth studies on the functional mechanisms of the Coq4 domain-containing hypothetical proteins that are involved in unusual oxidation steps.

3.2 Introduction

Disorazoles are a family of myxobacterial macrodilactones possessing symmetric or pseudo-symmetric structures and potent cytotoxicity through blocking the assembly of microtubules.[1] Previously, 29 disorazole A compounds and 10 disorazole Z congeners were isolated from the *Sorangium cellulosum* strain So ce12 and So ce1875, respectively.[2,3] These compounds are produced by *trans*-AT PKS/NRPS biosynthetic pathways.[4,5] The disorazole A biosynthetic gene cluster from So ce12 (*dis12* BGC) and the disorazole Z BGC from So ce427 (*dis427* BGC) were characterized by heterologous expression in *Myxococcus xanthus* DK1622, leading to the production of disorazole A2 (**1**) and disorazole Z1 (**2**) as the primary compound, respectively.[3,6] Besides, the *dis12* BGC was also heterologously expressed in the *Burkholderia thailandensis* strain E264; however, the intermediate disorazole F2 (**3**) that lacks the epoxy group was found to be the product.[7,8]

Both disorazole A and disorazole Z biosynthetic pathways comprise three multidomain-containing megasynthetases (DisA, DisB, and DisC), a discrete DisD that contains an acyltransferase domain, and a hypothetical protein DisE. Besides, the disorazole Z biosynthetic pathway contains an additional methyltransferase DisF, which was assigned to *O*-methylation during the formation of the methyl ester group.[3] However, no predicted epoxidase for disorazole A biosynthesis or oxygenase(s) for site-specific oxidation of the geminal dimethyl group for disorazole Z biosynthesis were found in the identified biosynthetic pathways. Even though, disorazole A2 and disorazole Z1 were produced when the gene clusters were expressed in the heterologous host *M. xanthus* DK1622. It is likely that genes encoding for oxygenases, such as the cytochrome P450-dependent enzymes, are not clustered with the core PKS/NRPS region, and homologues from the phylogenetically closely-related host *M. xanthus* DK1622 rather than from *B. thailandensis* E264 could complement the missing functions for auxiliary modifications; however, experimental proof is still missing.

Significantly, a hypothetical protein DisE that contains a Coq4 domain is conserved in both types of disorazole biosynthetic pathways, the function of which has not yet been assigned to disorazole biosynthesis. The Coq4 is known as a coenzyme Q (ubiquinone) biosynthesis protein from yeast and human mitochondria, which has no other known catalytic domain but is supposed to be an indispensable accessory factor for the assembly of a multi-enzyme complex in ubiquinone biosynthesis.[9-11] Besides, Coq4 homologs were found to be present in prokaryotic species that do not produce ubiquinone, which may suggest a broader role of this protein family.[11] However, the basic function of the Coq4 family of proteins is still poorly understood.

In this study, we discovered the Coq4 domain-containing hypothetical protein DisE to be involved in disorazole biosynthesis, playing an indispensable role in tailoring oxidations. Based on the heterologous expression systems using the host *M. xanthus* DK1622, we found that deletion of $disE_{427}$ in the disorazole Z biosynthetic pathway resulted in complete abolition of disorazole Z1 (**2**) production, and instead led to the accumulation of the intermediate disorazole Z11 (**4**), which is a simplified form of **2** lacking the methyl ester group but retaining the geminal dimethyl group. Similarly, deletion of $disE_{12}$ in the disorazole A biosynthetic pathway completely abolished disorazole A2 (**1**) production and led to the accumulation of the intermediate disorazole F2 (**3**), which lacks the epoxy group modification. Additionally, inactivation of cytochrome P450 genes in *M. xanthus* DK1622 showed that P450-dependent enzymes collaborate with DisE for these enigmatic oxidations.

3.3 Results and discussion

3.3.1 The hypothetical gene *disE$_{427}$* is involved in disorazole Z biosynthesis

Having access to the heterologous expression systems for both disorazole A and disorazole Z,[3,6] we sought to hybridize these two biosynthetic pathways to produce non-natural disorazoles in the host *M. xanthus* DK1622. Upon replacing *disE$_{427}$-disD$_{427}$-disF$_{427}$* with *disD$_{12}$*, the disorazole Z1 (**2**) production was abolished completely (Figure S1). However, we observed the unexpected accumulation of a proposed intermediate, termed disorazole Z11 (**4**), which has a mass that was 88 Da lower than **2** (Figure S1). This compound was purified and characterized by NMR (Table S1), which showed no further modification on the geminal dimethyl groups compared with **2** (Figure 1). Total synthesis was reported for this simplified disorazole in the year of 2010.[12]

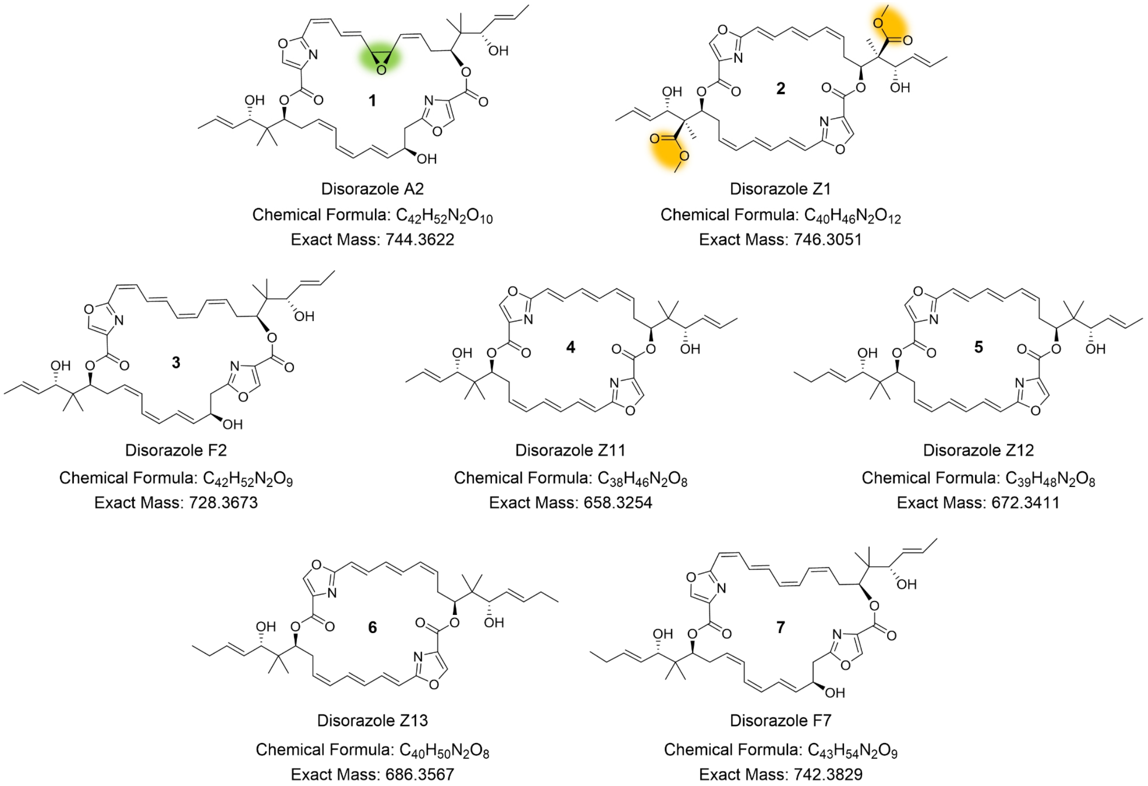


Figure 1 Chemical structures of disorazole compounds isolated from *M. xanthus* mutants that contain the *disE*-knockout expression constructs. Disorazole A2 and disorazole Z1 are shown for comparison, with the epoxy group and methyl ester groups highlighted in green and orange, respectively.

The accumulation of this unmodified intermediate indicates that at least a portion of the *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$ gene cassette is involved in oxidations that convert one methyl group to a hydroxyl group and subsequently to a carboxyl group. Our previous study revealed that the methyltransferase DisF$_{427}$ was not involved in oxidations, as oxidized intermediates were still produced in the absence of *disF*$_{427}$.[3] Coincidentally, DisD contains an uncharacterized oxidoreductase (Or) domain behind the acetyltransferase (AT) domain.[13,14] Therefore, we knocked out the Or domain sequence of the *disD*$_{427}$ (Figure S2), in the meantime, we also replaced the *disD*$_{427}$ with the full-length *disD*$_{12}$ or truncated *disD*$_{12}$ (only AT domain or AT domain together with the linker region) (Figure S3). Surprisingly, disorazole Z1 was still produced in all cases in the heterologous host *M. xanthus* DK1622, even when the full-length *disD*$_{427}$ was knocked out (Figure S4). This suggests the presence of a *trans*-AT homologue in *M. xanthus* DK1622 that is capable of complementing the function of DisD$_{427}$. One possible candidate for this *trans*-AT homologue is the myxovirescin *trans*-AT TaV (MXAN_3942) (Figure S5).[15] These results indicate that the *disE*$_{427}$ is essential for oxidation steps in disorazole Z biosynthesis.

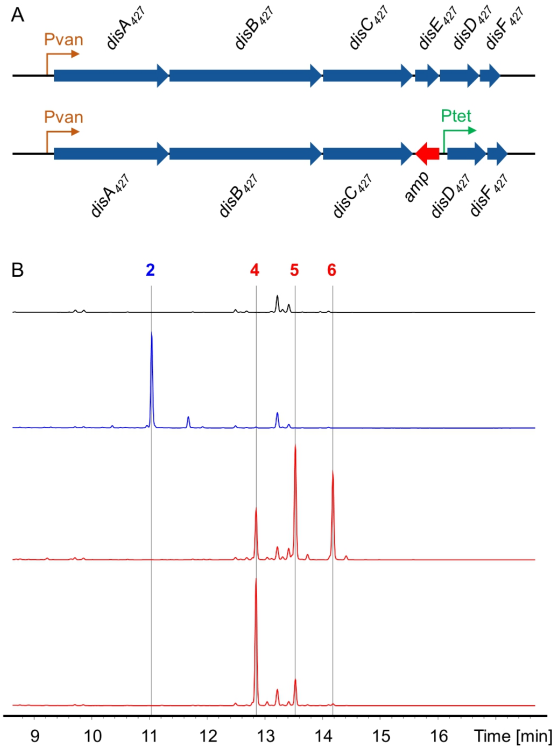


Figure 2 Production of disorazole Z intermediates/derivatives after deletion of *disE*$_{427}$. (A) Scheme of *dis427* gene cluster and its modifications. (B) HPLC-MS analysis of crude extracts from *M. xanthus* DK1622 (black trace) and DK1622 mutants containing the Pvan-*dis427* gene cluster (blue trace) or Pvan-*dis427*-Δ*disE* gene cluster (red trace). The chromatograms are shown as UV absorption at 320 nm.

As expected, deletion of $disE_{427}$ or $disE_{427}$-$disD_{427}$-$disF_{427}$ resulted in the abolishment of disorazole Z1 (**2**) production and the accumulation of disorazole Z11 (**4**) (Figure 2, Figure S6, and Figure S7), consistent with previous observations. Notably, we were able to purify and characterize two new disorazole derivatives, termed disorazole Z12 (**5**) and disorazole Z13 (**6**), which feature the prolonged side chain(s) by an additional methyl group (Figure 1, Figure 2, Table S3, and Table S4). Interestingly, the yield of **5** or **6** was found to be higher than that of **4** in some mutants (Figure 2). The reasons for the accumulation of compounds **5** and **6**, as well as the observed differences of metabolic profiles among different mutants, are currently unclear and require further investigation.

3.3.2 The hypothetical gene $disE_{12}$ is involved in disorazole A biosynthesis

Since $disE_{427}$ was found to be responsible for oxidations of the geminal dimethyl groups and $disF_{427}$ for subsequent methylation, we sought to generate novel disorazoles by replacing the $disE_{12}$ with $disE_{427}$-$disF_{427}$ and expressing the hybrid gene cluster in the host *M. xanthus* DK1622 (Figure S8). However, the expected disorazole A derivatives with methyl ester modifications were not detected in the crude extracts. Unexpectedly, production of disorazole A2 (**1**) was abolished, while the intermediate disorazole F2 (**3**) and a new derivative, termed disorazole F7 (**7**), were accumulated (Figure 1, Figure S8). These compounds were purified and analyzed by NMR (Table S5, Table S7). Unlike compound **1**, compounds **3** and **7** lack the epoxy group. The accumulation of **3** in the absence of DisE_{12} indicates that this hypothetical protein is involved in epoxy group formation, contrary to the previous results.[6] To confirm the involvement of $disE_{427}$ in disorazole A biosynthesis, we conducted further gene knockout experiments (Figure 3). To avoid potential polar effects, we inserted a Papr promoter upstream of $disD_{12}$ during the knockout of $disE_{12}$. In addition, we also performed gene knockout experiments targeting $disD_{12}$ and $disE_{12}$-$disD_{12}$. Deletion of $disE_{12}$ resulted in the accumulation of disorazole F2 (**3**) and the abrogation of disorazole A2 (**1**). Upon deletion of $disD_{12}$, the production of both **1** and **3** was significantly reduced by at least tenfold. These findings strongly support the essential role of $disE_{12}$ in heterologous production of **1** in the host *M. xanthus* DK1622, and suggest that the host's *trans*-AT homologue may potentially complement DisD_{12} in disorazole A biosynthesis.

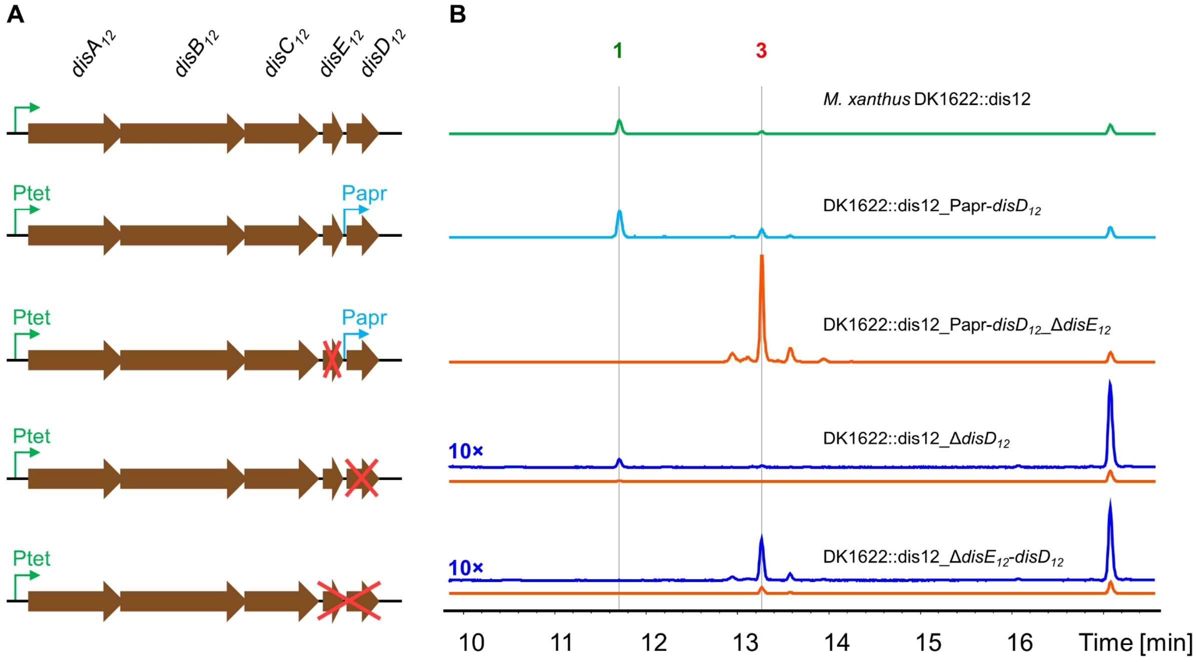


Figure 3 Engineering and heterologous expression of the *dis12* gene cluster in *M. xanthus* DK1622. (A) Scheme of *dis12* gene cluster and its modifications. (B) HPLC-MS analysis of crude extracts from *M. xanthus* mutants that contain the corresponding gene clusters. The chromatograms are shown as EIC of *m/z* (745.37; 729.38) ± 0.2.

3.3.3 *In silico* investigation of the hypothetical protein DisE

To predict the function of the hypothetical protein DisE, we searched the Pfam database,[16] and found that DisE contains a Coq4 domain (Pfam entry: PF05019) with a highly conserved HDXXH-$(X)_{11}$-E motif;[11] however, the detailed function of this motif is still unclear. The AlphaFold-predicted structure of $DisE_{427}$ showed an overall similarity (RMSD: 2.732) to the Alr8543 protein (PDB: 6E12) from Cyanobacteria *Nostoc* sp. PCC 7120 with an unknown function (Figure S9).[17] Besides, we found no convincing homology between $DisE_{427}$ and any proteins of known function.

A search in our in-house myxobacterial genome database using $DisE_{427}$ revealed that the top 14 BLAST hits are all located in disorazole biosynthetic gene clusters, mostly from *Sorangium cellulosum* except for one BGC that is from *Aetherobacter rufus*. Together with the *Sorangium cellulosum* So ce12, So ce427, and So ce1875, we have discovered a total of 16 myxobacterial strains that harbor the disorazole biosynthetic gene cluster. Through protein alignment and phylogenetic tree analysis, we found that these 16 hypothetical proteins can be divided into two groups with an exception of the DisE from *A. rufus* strain MSr9331 (Figure 4,

Figure S10, and Figure S11). This grouping is consistent with their corresponding types of gene clusters and natural products. However, the disorazole biosynthetic gene cluster from *A. rufus* strain MSr9331 seems to be silent as we could not detect any disorazole family compounds from its crude extract, or even seems to be unfunctional as we cannot detect any heterologous product after expression of the promoter-engineered gene cluster in *M. xanthus*.

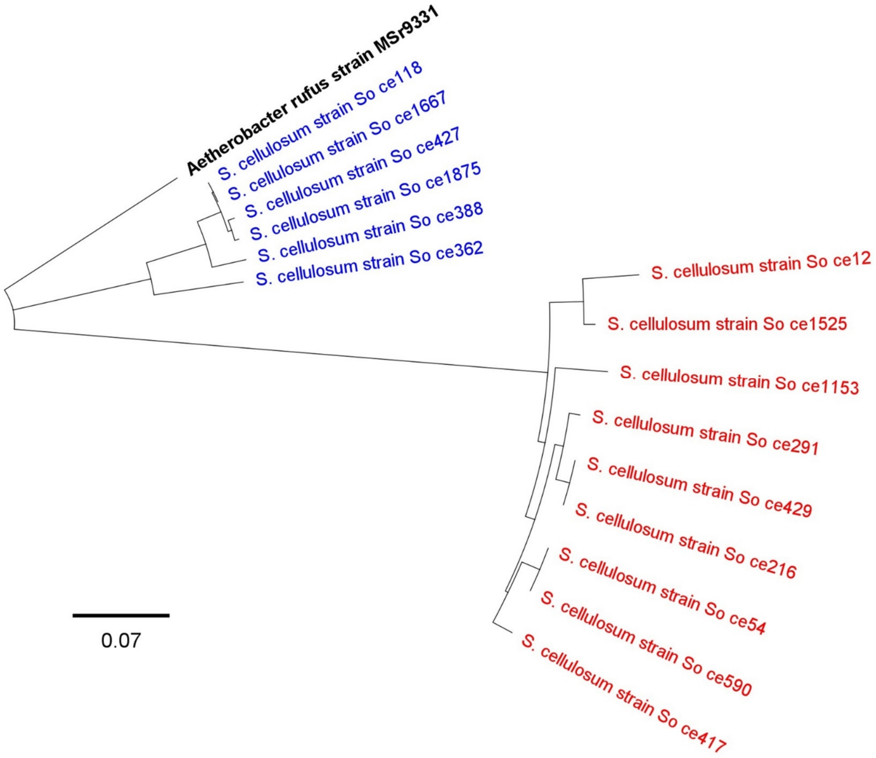


Figure 4 Phylogenetic analysis of DisE proteins from 16 myxobacterial disorazole biosynthetic pathways. The *Sorangium cellulosum* strains that contain the disorazole Z or disorazole A biosynthetic gene cluster are marked in blue or red, respectively. Additionally, the myxobacterium *Aetherobacter rufus* strain MSr9331 also contains a disorazole biosynthetic gene cluster but is not known to produce disorazole.

Following the top 14 hits, we found two BLAST hits from *Cystobacterineae* strains MCy13299 and MCy12312, which were not located in disorazole biosynthetic gene clusters but still showed relatively high sequence similarities (55.1% and 54.6%, respectively) to DisE$_{427}$. Both hits are located in the vicinity of genes encoding P450-dependent enzymes. Interestingly, we also found a Coq4 domain-containing protein (MXAN_1114) in *M. xanthus* DK1622 that shows 37.1% similarity and 23.7% identity with DisE$_{427}$. Adjacent to MXAN_1114 is a gene that encodes a carboxyltransferase. Although this type of gene cassette, which encodes a Coq4 domain-containing protein and a carboxyltransferase, is widespread in other myxobacterial strains, its function remains unknown.

While Coq4 domain-containing proteins are widespread in myxobacteria, it is still unclear whether these proteins are involved in ubiquinone biosynthesis or other (secondary) metabolite biosynthetic pathways. The Coq4 domain-containing protein located within a known biosynthetic pathway is rarely reported. To our knowledge, only the FtyA from 3-formyl-L-tyrosine (fty) biosynthetic pathway has been found to contain a Coq4 domain; however, the function of FtyA is unclear, as the deletion of *ftyA* had no effect on heterologous production of fty using the *E. coli* host.[18]

3.3.4 *In vivo* conversion by substrate complementation

To assess whether the hypothetical protein alone is sufficient for conversion in the heterologous host, we cloned *disE*$_{12}$ or *disE*$_{427}$ under the control of a vanillate-inducible Pvan promoter and transferred them separately to *M. xanthus* DK1622 for substrate feeding experiments (see the Methods section). As shown in Figure 5A, feeding of disorazole F2 (**3**) to *M. xanthus* DK1622::Pvan-*disE*$_{12}$ resulted in the conversion of **3** to disorazole A2 (**1**). Notably, the addition of vanillate improved the conversion rate, while no conversion was observed when **3** was fed to the wild-type *M. xanthus* DK1622. Similarly, feeding of disorazole Z11 (**4**) to *M. xanthus* DK1622::Pvan-disE$_{427}$ resulted in the conversion of **4** to disorazole Z10 (Figure 5B). These results suggest that the epoxy group formation in disorazole A and oxidations of the geminal dimethyl group in disorazole Z can occur after TE-mediated polyketide chain release. To purify the hypothetical proteins for *in vitro* reactions, we first attempted to express DisE in the *E. coli* host BL21(DE3). Although soluble expression of the recombinant N-terminal 6×His-tagged hypothetical proteins DisE$_{427}$ and DisE$_{12}$ was achieved with decent production yields, size-exclusion chromatography using a HiLoad 16/60 Superdex 200 column or a Superose 6 Increase column revealed that the soluble recombinant proteins were eluted at the void volume, indicating that the proteins were improperly folded. Despite several attempts to optimize the expression conditions, including using an N-terminal Sumo tag, a C-terminal His tag, and co-expression with TaKaRa chaperone plasmids, we were unable to obtain properly folded proteins using the *E. coli* system. Therefore, we attempted to purify DisE using the myxobacterial host. We constructed a *vanR-Pvan-His-TEV* cassette and inserted it upstream of *disE*$_{427}$ in the *dis427* gene cluster to enable a fusion expression in *M. xanthus* DK1622. The DisE$_{427}$ was expressed as an N-terminal 6×His-tagged protein under the control of the vanillate-inducible promoter. In the absence of vanillate, the compound **4** was accumulated, while supplementation with 1 mM vanillate induced the expression of *disE*$_{427}$ and resulted in the production of disorazole Z1 (**2**) as the primary product (Figure S12). This indicated that the N-terminal His-tagged DisE$_{427}$ was still functional in *M. xanthus* DK1622 and verified once again that *disE*$_{427}$ is essential for methyl ester formation in disorazole Z biosynthesis. We attempted to purify the hypothetical protein using this expression system; however, the production was lower than the detection limit checked by SDS-PAGE, most likely due to the degradation by endogenous proteases. Despite inducing the protein expression under lower temperature (18 °C instead of 30 °C) and adding protease inhibitors during purification, we were still unable to purify the protein using the *M. xanthus* host.

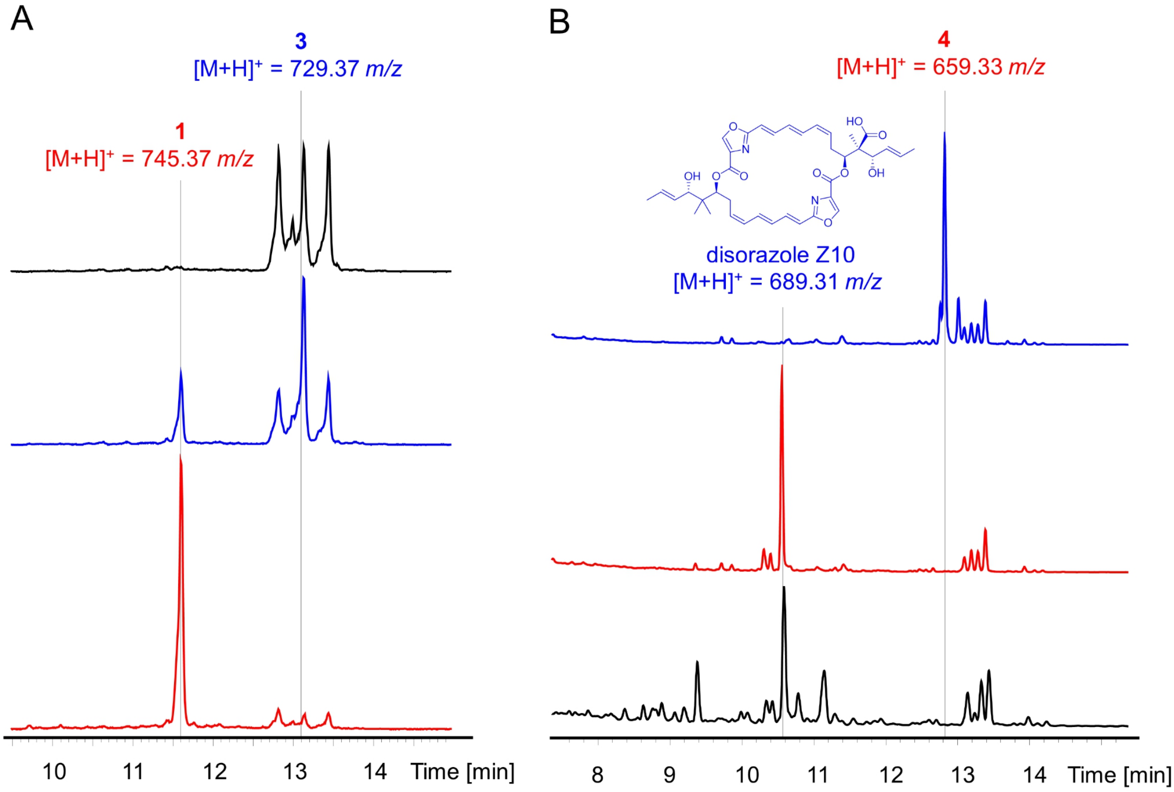


Figure 5 HPLC-MS analysis of crude extracts from *M. xanthus* DK1622::Pvan-*disE* after feeding of the corresponding substrates. (A) Feeding of disorazole F2 (**3**) to the culture of wild-type *M. xanthus* DK1622 (black trace), *M. xanthus* DK1622::Pvan-*disE*$_{12}$ (blue trace), and *M. xanthus* DK1622::Pvan-*disE*$_{12}$ supplemented with 1 mM vanillate (red trace); the chromatograms are shown as EIC of (745.37; 729.37) ± 0.05 *m/z*. (B) Feeding of disorazole Z11 (**4**) to the culture of wild-type *M. xanthus* DK1622 (blue trace) and *M. xanthus* DK1622::Pvan-*disE*$_{427}$ supplemented with 1 mM vanillate (red trace), where the compound **4** was converted into disorazole Z10, which is the primary product of *M. xanthus* DK1622::*dis427*_ Δ*disF*$_{427}$ (black trace);[3] the chromatograms are shown as UV absorption at 320 nm.

3.3.5 Proposed oxygenase partners of DisE

Although the hypothetical protein DisE in *M. xanthus* DK1622 is sufficient for *in vivo* conversions, it is unlikely that DisE functions as an oxygenase and is capable of catalyzing an *in vitro* reaction without other oxygenase partners. It is probable that the oxygenase partners of DisE are cytochrome P450-dependent enzymes (CYPs) that are capable of catalyzing diverse reactions in secondary metabolism, the majority of which involve oxygenations, such as the carbon hydroxylation and the epoxide formation.[19,20] The genome of *M. xanthus* DK1622 contains seven CYP-encoding genes.[21] We found that inactivating either MXAN_0683 or MXAN_1743 through single-crossover homologous recombination (Figure S13) significantly reduced the production of disorazole Z1 (**2**) and led to the accumulation of disorazole Z11 (**4**) (Figure 6), with no observable effect on the strain growth. These two proteins belong to the list of seven CYP candidates.

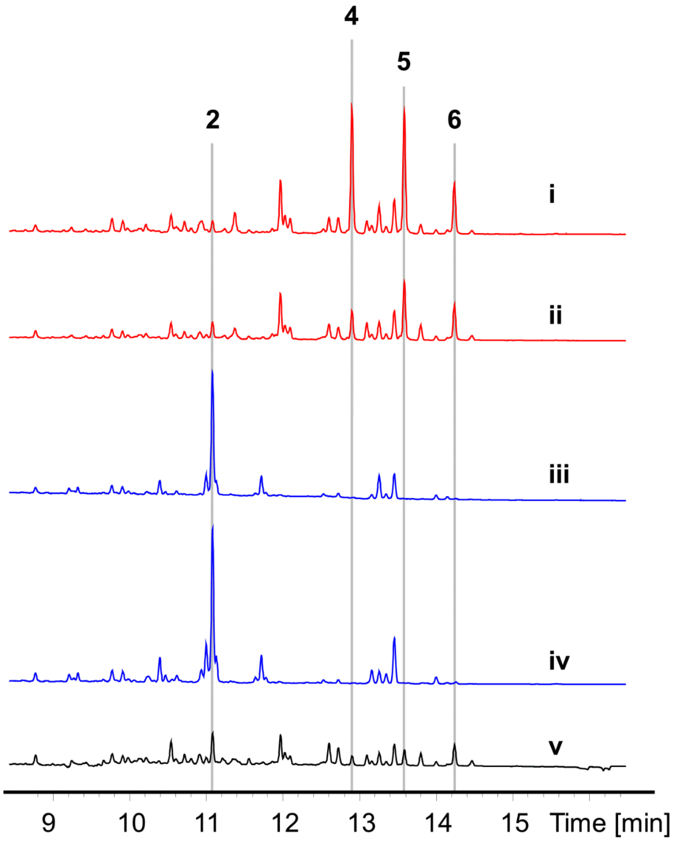


Figure 6 HPLC-MS analysis of crude extracts from the CYP-inactivation mutants of *M. xanthus* DK1622::*dis427*. The chromatograms are shown as UV absorption at 320 nm for inactivation of MXAN_0683 (i), MXAN_1743 (ii), MXAN_2304 (iii), MXAN_4919 (iv), and MXAN_7298 (v). The inactivations that significantly affected disorazole Z production are highlighted in red, while those with no significant impact or limited influence are indicated in blue and black, respectively. Inactivation of other two CYP genes negatively affected the strain growth.

These findings suggest that MXAN_0683 and MXAN_1743 are potential oxygenase partners of DisE during the heterologous biosynthesis of disorazole Z which may be able to complement each other in this functionality. In contrast, inactivation of MXAN_2304 or MXAN_4919 did not significantly affect the disorazole Z production, while inactivation of MXAN_7298 had a moderate effect (Figure 6). However, inactivation of MXAN_3943 or MXAN_4127 negatively affected the strain growth.

It is possible that more than one CYPs could serve as the oxygenase partners of DisE_{427}, since single inactivation of a CYP was not sufficient to completely abolish the production of disorazole Z1. However, multiple inactivation of CYPs in *M. xanthus* DK1622 is impractical. In addition, at least 18 CYP-encoding genes were found in the genome of *S. cellulosum* So ce427, and unfortunately, this disorazole Z producer strain proved to be intractable for genetic manipulation. Comparison of sequence identities between seven CYPs from *M. xanthus* DK1622 and 18 CYPs from *S. cellulosum* So ce427 revealed MXAN_1743 as the most promising candidate (Figure S14), which requires further *in vitro* validation. The coordinated action of a CYP with a hypothetical protein has been reported in the myxobacterial crocapeptin biosynthetic pathway, where the helper protein CpnE could accelerate the hydroxylation reaction catalyzed by the CYP CpnF.[22] However, no convincing homology was found between DisE and CpnE.

It is worth noting that, in addition to TaV, which resembles the *trans*-AT domain of DisD, *M. xanthus* DK1622 has a standalone oxygenase TaN (MXAN_3934) that shows a high sequence similarity to the Or domain of DisD (Figure S15).[15] The oxygenase TaN is proposed to perform hydroxylation in myxovirescin biosynthesis; however, this has not yet been confirmed experimentally.[15] Therefore, in addition to CYPs, the hypothetical protein DisE could also cooperate with the Or domain of DisD or with the oxygenase TaN (Figure S16).

3.4 Conclusion

Molecular engineering and subsequent heterologous expression of the disorazole biosynthetic pathways led to the discovery of the essential role played by the hypothetical protein DisE, which contains a Coq4 domain, in disorazole biosynthesis. We found that the deletion of $disE_{12}$ and $disE_{427}$ resulted in the production of disorazole F2 lacking the epoxide group and disorazole Z11 without the formation of a methyl ester group in *M. xanthus* DK1622, respectively. We also found that DisE is highly conserved in disorazole biosynthetic pathways and may act as an accessory factor for recruiting oxygenase(s) (such as P450 enzymes) and/or presenting substrates to its oxygenase partner. Its homologues are widely distributed across myxobacteria and are encoded by function-unknown chromosome regions that have not yet been studied.

Our study has uncovered the unique oxidation steps mediated by the Coq4 domain-containing hypothetical protein DisE, which is involved in the post-modifications of disorazole bis-lactones. Further studies focusing on the functional expression of DisE-CYP pair(s) and subsequent *in vitro* biochemical investigations may help to elucidate the mystery functions of the Coq4 domain-containing proteins.

3.5 Methods

3.5.1 General experimental procedures

E. coli strain GB2005, GB05-red, GB05-dir, GBred-gyrA462, and GBdir-pir116 were used for molecular cloning and engineering,[23] and cultivated in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH=7.0) at 37 °C. HPLC-purified DNA oligos were synthesized by Sigma-Aldrich and are listed in Table S8. The model myxobacterial strain *Myxococcus xanthus* DK1622 was used as the host for heterologous expression, and cultivated in CTT medium (10 g/L casitone, 10 mM Tris-HCl, 8 mM magnesium sulfate, 1 mM potassium phosphate, pH=7.6) at 30 °C. Transformation of *E. coli* and *M. xanthus* were performed by electroporation according to the previously established protocols.[24] *M. xanthus* DK1622 mutants containing exogenous genes generated in this study are listed in Table S9. The following antibiotics were supplemented to the medium when the cultivated strain carries the corresponding resistance genes: 10 μg/mL for oxytetracycline, 15 μg/mL for chloramphenicol, 50 μg/mL for kanamycin and apramycin, and 100 μg/mL for ampicillin.

3.5.2 HPLC-MS measurements

The HPLC measurements were performed on a Thermo Scientific Dionex UltiMate 3000 RSLC system using a Waters Acquity BEH C18 column (100 mm × 2.1 mm, 1.7 μm) in combination with a Waters VanGuard BEH C18 1.7 μm guard. Separation was achieved using a linear 5−95% gradient of acetonitrile (supplemented with 0.1% formic acid) in ddH$_2$O (supplemented with 0.1% formic acid) for 18 min at a flow rate of 0.6 mL/min and the column temperature of 45 °C. UV spectra were recorded by a diode array detector at 200−600 nm. The HPLC system was coupled with a maXis 4G HR-TOF or an amaZon speed ion trap MS system (Bruker Daltonics) equipped with an ESI source. MS spectra were recorded by positive ion mode ranging from 150 to 2500 *m/z*.

3.5.3 Construction of *M. xanthus* mutants containing engineered gene clusters

The disorazole A biosynthetic gene cluster (*dis12* BGC comprising *disA*$_{12}$, *disB*$_{12}$, *disC*$_{12}$, *disE*$_{12}$, and *disD*$_{12}$) and the disorazole Z biosynthetic gene cluster (*dis427* BGC comprising *disA*$_{427}$, *disB*$_{427}$, *disC*$_{427}$, *disE*$_{427}$, *disD*$_{427}$, and *disF*$_{427}$) were integrated into the chromosome of *Myxococcus xanthus* DK1622, respectively, resulting in the generation of *M. xanthus* DK1622::*dis12* and *M. xanthus* DK1622::*dis427* (Mutants No. 1 and 2 in Table S9). To replace *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$ by *disD*$_{12}$, the *Papr-apr-disD*$_{12}$ cassette was constructed by overlap-extension PCR using the oligos No. 1−4 (Table S8), and electroporated into the *E. coli* GB05-red containing p15A-km-tps-Ptet-dis427 for linear-plus-circular homologous recombination (LCHR).[23] The resulting plasmid was used for electroporation of *M. xanthus* DK1622, resulting in the generation of DK1622::*dis427_ΔdisE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$*_disD*$_{12}$ (Mutants No. 3 in Table S9). The mutants No. 4, 5 and 6 (Table S9) were constructed similarly except that oligo pairs No. 5&6, 5&7 and 5&8 (Table S8) containing different homology arms were used, respectively. To construct mutant No. 7, 8, 9, 10 and 11 (Table S9), the *Papr-apr* cassettes comprising the apramycin resistance gene together with its promoter region and flanking different homology arms were amplified using oligo pairs No. 9&10, 11&10, 5&10, 1&12 and 1&13 (Table S8), respectively, and electroporated into the *E. coli* GB-red containing p15A-km-tps-Ptet-dis427 for LCHR, followed by electroporation of *M. xanthus* DK1622 using the resulted expression constructs. To replace *disE*$_{12}$ by *disE*$_{427}$-*disF*$_{427}$, the *Papr-apr-disE*$_{427}$-*disF*$_{427}$ cassette was constructed by overlap extension PCR using the oligos No. 14−19 (Table S8), and electroporated into the *E. coli* GB05-red containing p15A-km-tps-Ptet-dis12 for LCHR. To construct mutant No. 13, 14, 15, and 16 (Table S9), the *Papr-apr* cassettes comprising the apramycin resistance gene together with its promoter region and flanking different homology arms were amplified using oligo pairs No. 20&21, 22&21, 20&23 and 22&23 (Table S8), respectively, and electroporated into the *E. coli* GB-red containing p15A-km-tps-Ptet-dis12 for

LCHR, followed by electroporation of *M. xanthus* DK1622 using the resulted expression constructs. For fusion expression of a N-terminal 6xHis tag and a TEV protease cleavage site with the *disE₄₂₇* under the control of a Pvan promoter, an *amp-Pvan-His-TEV* cassette was constructed by two rounds of PCR using oligos No. 24&25 and 24&26 (Table S8), and electroporated into the *E. coli* GB05-red containing p15A-km-tps-Ptet-dis427 for LCHR. To construct DK1622::Pvan-*disE*$_{427}$, the plasmid p15A-km-tps-Pvan-disE$_{427}$ was generated by Gibson Assembly between a *p15A-km-tps-Pvan* cassette and the *disE*$_{427}$ amplified by PCR using oligos No. 27&28, and used for electroporation of *M. xanthus* DK1622. The mutant DK1622::Pvan-*disE*$_{12}$ was generated similarly using the plasmid p15A-km-tps-Pvan-disE$_{12}$, which was constructed using oligos No. 29&30 (Table S8).

The *disE*$_{427}$-knockout experiment was also repeated based on p15A-km-int-Pvan-dis427 plasmid using the oligos No. 31&32 for LCHR, resulting in the generation of DK1622::*dis427_ΔdisE*$_{427}$ strain with much higher productions of disorazole Z11, Z12 and Z13, which was used for fermentation and compounds purification.

3.5.4 NMR measurements

NMR data were recorded on an UltraShield 500 MHz (^{1}H at 500 MHz, ^{13}C at 125 MHz) or an AVANCE III 700 MHz (^{1}H at 700 MHz, ^{13}C at 175 MHz) NMR system equipped with a 5-mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) were calculated in ppm, and coupling constants (J) were calculated in Hz. Spectra were recorded in methanol-d_4 and adjusted to the solvent signals (δ_H 3.31, δ_C 49.2). Measurements were conducted in 5-mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA). For the two-dimensional experiments, HMBC, HSQC, and gCOSY standard pulse programs were used. HMBC and HSQC experiments were optimized for $^{2,3}J_{C\text{-}H}$ = 6 Hz and $^{1}J_{C\text{-}H}$ = 145 Hz, respectively.

3.5.5 Inactivation of P450 enzyme-encoding genes in *M. xanthus* DK1622

About a 1-kb fragment of each P450 enzyme-encoding gene was amplified by PCR using the genomic DNA of *M. xanthus* DK1622 as the template, and then assembled with the *pR6K-tet* cassette by linear-plus-linear homologous recombination (LLHR) using the *E. coli* strain GBdir-pir116, resulting in the generation of the plasmid pR6K-tet-MXAN_0683, pR6K-tet-MXAN_1743, pR6K-tet-MXAN_2304, pR6K-tet-MXAN_3943, pR6K-tet-MXAN_4127, pR6K-tet-MXAN_4919, and pR6K-tet-MXAN_7298, respectively. These plasmids were used for electroporation of *M. xanthus* DK1622::*dis427*. Mutants resulting from single crossover inactivation (No. 20-24, Table S9) were selected using kanamycin and oxytracycline, and further checked by colony PCR (Figure S13). The correct mutants were then used for fermentation and subsequent metabolome analysis.

3.5.6 Fermentation and compounds purification

Fermentation of *M. xanthus* strains were carried out according to the previously established protocols.[3] The methanol extract of cells and XAD was partitioned with n-hexane, redissolved in Milli-Q water, and extracted with chloroform. The chloroform phase was fractionated using Si-Flash chromatography on a Biotage system. Gradient was set as follow: 0−20 CV, hexane to ethyl acetate; 20−40 CV, ethyl acetate to MeOH; 40−45 CV, MeOH. Fractions containing targeted compounds were combined and further separated by semi-preparative HPLC using A (Milli-Q water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid) as mobile phase at a flow rate of 5 mL/min and the column temperature of 45 °C. For purification of disorazole Z11, Z12 and Z13, a Waters XSelect Peptide BEH C18 column (250×10 mm, 5 µm) and an isocratic elution method (65% B) were used. Fractions were detected by UV absorption at 320 nm and were collected using an AFC-3000 fraction collector based on retention time of 10.4 min for disorazole Z11, 14.0 min for disorazole Z12, and 19.0 min for disorazole Z13. For disorazole F2 and F7, separation was achieved using a Waters XBridge Peptide BEH C18 column (250×10 mm, 5 µm) and the following gradient: 0−5 min, 5% B; 5−10 min, to 60% B; 10−30 min, to 75% B; 30−31 min, to 95% B; 31−36 min, 95% B; 36−37 min, to 5% B; 37−40 min, 5% B. Fractions were detected by UV absorption at 320 nm and were collected using an AFC-3000 fraction collector based on retention time of 20.0 min for disorazole F2 and 22.8 min for disorazole F7. Impure fractions were further purified using a Waters XSelect CSH Phenyl-Hexyl column (250×10 mm, 5 µm) and isocratic elution methods (60% B for disorazole F2, t_R = 9.25 min; 62% B for disorazole F7, t_R = 10.05 min).

3.5.7 Substrate complementation for *in vivo* conversion

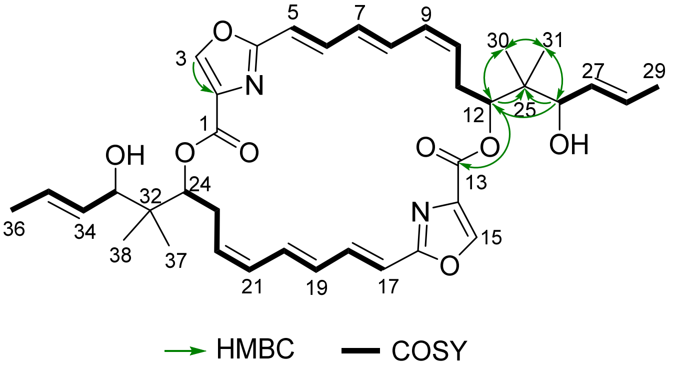
The *M. xanthus* DK1622::Pvan-*disE*$_{12}$ was cultivated in CTT medium at 30 °C with shaking until the exponential growth. Then, 1.6 mL of culture was added into a 2-mL microcentrifuge tube with a hole punctured in the cap, followed by addition of 1 mM vanillate and 2 µL disorazole F2 solution (1 mg/mL in methanol). The culture was further incubated at 30 °C for 6 h in an Eppendorf thermomixer, lyophilized to dryness, and extracted with 100 µL methanol. After centrifugation, 2 µL of the methanol extract was used for HPLC-MS measurement. Feeding of disorazole Z11 to *M. xanthus* DK1622::Pvan-*disE₄₂₇* was carried out following the same procedure.

3.6 Supporting information

3.6.1 Structure elucidation of disorazole intermediates/derivatives

Disorazole Z11 (**4**) shows a $[M+H]^+$ signal at *m/z* 659.3327 with corresponds to the molecular formula $C_{38}H_{46}N_2O_8$. Interestingly, ^1H and ^{13}C NMR spectra showed only half of the expected signals, in the same fashion as disorazole Z1 (**2**),[3] indicating a completely symmetrical molecule. High similarity of proton and carbon shifts of **4** with the northern side chain of disorazole Z10 (Table S2), as well as COSY and HMBC correlations of **4**, indicated that both side chains feature two geminal methyl groups.[3]

Table S1 NMR data of disorazole Z11 (**4**) in methanol-d_4 (^1H 700 MHz; ^{13}C 175 MHz).

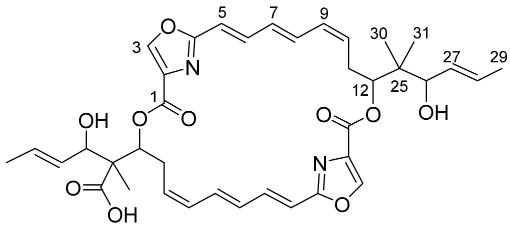


Pos.	δ_H, m (*J* [Hz])	δ_C, type	COSY	HMBC[a]
1/13	-	161.8, C	-	-
2/14	-	136.0, C	-	-
3/15	8.58, br s	145.6, CH	-	2/14 (wk), 4/16 (wk)
4/16	-	163.6, C	-	-
5/17	6.15, m	116.0, CH	6/18	4/16, 6/18, 7/19
6/18	6.91, br s	139.4, CH	5/17, 7/19	-
7/19	6.13, m	131.9, CH	6/18, 8/20	4/16, 5/17, 6/18
8/20	6.77, br t (12.6)	136.3, CH	7/19, 9/21	6/18 (wk)
9/21	6.12, m	132.9, CH	8/20, 10/22	8/20, 10/22, 11/23, 12/24
10/22	5.70, m	132.7, CH	9/21, 11ab/23ab	7/19, 8/20
11a/23a	2.73, m	29.8, CH_2	10/22, 11b/23b, 12/24	10/22, 12/24

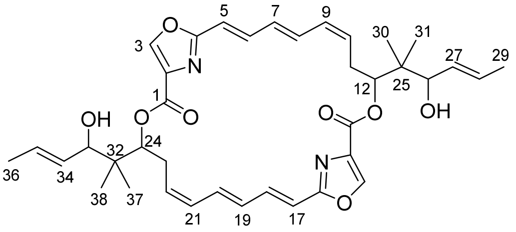
Pos.	δ_H, m (J [Hz])	δ_C, type	COSY	HMBC[a]
11b/23b	2.43, br dd (13.9, 7.2)	-	10/22, 11a/23a, 12/25	9/21, 20/22, 12/24, 25/32
12/24	5.16, br d (10.3)	79.0, CH	11ab/23ab	1/13, 9/21, 10/22, 11/23, 25/32, 26/33, 31/38
25/32	-	42.9, C	-	-
26/33	3.90, d (8.0)	78.0, CH	27/34, 29/36	12/24, 25/32, 27/34, 28/35, 30/37
27/34	5.59, ddq (15.3, 7.9, 1.4)	131.9, CH	26/33, 28/35, 29/36	26/33, 28/35, 29/36
28/35	5. 68, m	129.8, CH	27/35, 29/36	26/33, 29/36
29/36	1.73, dd (6.4, 1.4)	18.2, CH_3	26/33, 27/34, 28/35	25/32, 26/33, 27/34, 28/36
30/37	1.05, s	19.4, CH_3	31/38	12/24, 25/32, 26/33, 31/38
31/38	0.98, m	19.6, CH_3	30/37	12/24, 25/32, 26/33, 30/37

[a] Proton showing HMBC correlations to indicated carbons.

Table S2 Comparison of ^{1}H and ^{13}C NMR data of **4** with the northern part of disorazole Z10 in methanol-d_4.[a]



Disorazole Z10 [b]



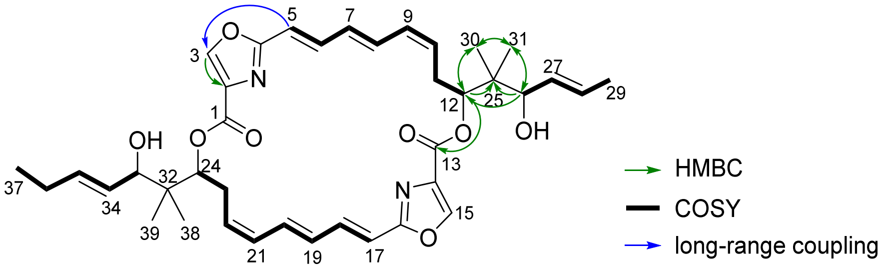
Disorazole Z11 (**4**) [d]

Pos.	δ_H^b	$\delta_C^{b,c}$	δ_H^d	δ_C^d	$\Delta\delta_H^b$-δ_H^d	$\Delta\delta_C^{b,c}$-δ_C^d
1	-	160.3	-	161.8	-	1.5
2	-	135.3	-	136.0	-	0.7
3	8.30	144.7	8.58	145.6	0.28	0.9
4/16	-	162.2	-	163.6	-	1.4
5/17	6.11	114.3	6.15	116.0	0.04	1.7
6/18	6.81	137.4	6.91	139.4	0.10	2.0
7/19	6.12	131.2	6.13	131.9	0.01	0.7
8/20	6.74	134.7	6.77	136.3	0.03	1.6
9/21	6.12	130.3	6.12	132.9	0.00	2.6
10	5.72	130.9	5.70	132.7	0.02	1.8
11a	2.71	28.5	2.73	29.8	0.02	1.3
11b	2.42	-	2.43	-	0.01	-
12	5.13	77.3	5.16	79.0	0.03	1.7
25	-	41.5	-	42.9	-	1.4
26	3.89	76.4	3.90	78.0	0.01	1.6
27	5.59	130.3	5.59	131.9	0.00	1.6
28	5.73	127.1	5.68	129.8	0.05	2.7
29/37	1.73	16.9	1.73	18.2	0.00	1.3
30	1.05	18.1	1.05	19.4	0.00	1.3
31	0.98	18.3	0.98	19.6	0.00	1.3

[a] ^{1}H/^{13}C at 700/150.9 MHz; [b] isolated from *S. cellulosum* So ce1875;[3] [c] δ_C extracted from HMBC; [d] isolated from *M. xanthus* DK1622::*dis427*_Δ*disE*$_{427}$.

The congener disorazole Z12 (**5**) exhibits the molecular formula $C_{39}H_{48}N_2O_8$, and showed similar 1D and 2D NMR data to that of disorazole Z11 (**4**). However, a further methylene group ($\delta_{\text{H-36}}$ 2.09, $\delta_{\text{C-36}}$ 26.7) could be observed, which showed COSY and HMBC correlations to the southern side chain positions 33, 34, 35 and 37 ($\delta_{\text{H-33}}$ 3.91, $\delta_{\text{C-33}}$ 78.0; $\delta_{\text{H-34}}$ 5.57, $\delta_{\text{C-34}}$ 129.6; $\delta_{\text{H-35}}$ 5.72, $\delta_{\text{C-35}}$ 136.9; $\delta_{\text{H-37}}$ 1.03, $\delta_{\text{C-37}}$ 14.2). Thus, **5** possesses a southern side chain that is prolonged by one methylene unit.

Table S3 NMR data of disorazole Z12 (**5**) in methanol-d_4 (^{1}H 500 MHz; ^{13}C 125 MHz).

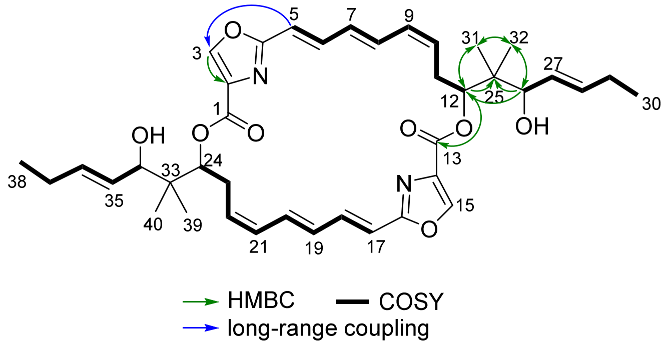


Pos.	δ_H, m (J [Hz])	δ_C, type	COSY	HMBC[a]
1/13	-	161.8, C	-	-
2/14	-	136.0, C	-	-
3/15	8.57, br s	145.6, CH	5/17	2/14, 4/16
4/16	-	163.6, C	-	-
5/17	6.15, d (15.5)	116.0, CH	3/15, 6/18	4/16, 6/18, 7/19
6/18	6.91, br s	139.4, CH	5/17, 7/19	-
7/19	6.13, m	131.9, CH	6/18, 8/20	4/16, 5/17, 6/18, 8/20, 9/21
8/20	6.77, br t (13.5)	136.3, CH	7/19, 9/21, 10/22	6/18, 9/21
9/21	6.11, m	132.9, CH	10/22, 11ab/23ab	7/19, 8/20, 11/23
10/22	5.70, m	132.6, CH	8/20, 9/21, 11ab/23ab	7/19, 11/23, 12/24
11a/23a	2.43, dd (13.8, 7.4)	29.8, CH_2	9/21, 10/22, 11b/23b, 12/24	9/21, 12/24, 25/32
11b/23b	2.73, m	-	9/21, 10/22, 11a/23a, 12/24	9/21, 12/24, 25/32
12/24	5.16, br dd (10.2, 2.5)	79.1, CH	11ab/23ab	1/13, 9/21, 11/23, 25/32, 26/33, 30/38, 31/39
25	-	42.9, C	-	-
26	3.90, d (7.6)	78.0, CH	27, 28, 29	12/24, 25/32, 27, 30/38, 31/39
27	5.61, m	131.9, CH	26, 28, 29	26, 28, 29
28	5.68, m	129.8, CH	26, 27, 29	26, 27, 29
29	1.72, dd (6.3, 1.2)	18.2, CH_3	26, 27, 28	27, 28
30/38	0.98, s	19.6, CH_3	-	12/24, 25/32, 26/33, 31/39
31/39	1.05/1.06, s	19.4, CH_3	-	12/24, 25/32, 26/33. 30/38
32	-	42.9, C	-	-
33	3.91, d (7.8)	78.0, CH	34, 36	12/24, 25/32, 27/34, 28/35, 30/38, 31/39
34	5.57, m	129.6, CH	33, 34	33, 35, 36
35	5.72, m	136.9, CH	34, 36, 37	33, 34, 36, 37
36	2.09, m	26.7, CH_2	33, 34, 35, 37	34, 35, 37
37	1.03, t (7.5)	14.2, CH_3	36	35, 36

[a] Proton showing HMBC correlations to indicated carbons.

Disorazole Z13 (**6**) exhibits the molecular formula $C_{40}H_{50}N_2O_8$. Similar to disorazole Z11 (**4**), it also shows only half of the expected signals in 1D and 2D NMR spectra, which points towards a completely symmetrical dimer. However, similar to disorazole Z12 (**5**), a further methylene group ($\delta_{\text{H-29/37}}$ 2.09, $\delta_{\text{C-29/37}}$ 26.7) can be observed in all measured spectra, which shows COSY and HMBC correlations to the side chain positions 27/35, 28/36 and 30/38 ($\delta_{\text{H-27/35}}$ 5.57, $\delta_{\text{C-27/35}}$ 129.6; $\delta_{\text{H-28/36}}$ 5.73, $\delta_{\text{C-28/36}}$ 136.9; $\delta_{\text{H-30/38}}$ 1.03, $\delta_{\text{C-30/38}}$ 14.2). Therefore, an additional methylene group prolongs both side chains of **6**.

Table S4 NMR data of disorazole Z13 (**6**) in methanol-d_4 (^{1}H 500 MHz; ^{13}C 125 MHz).

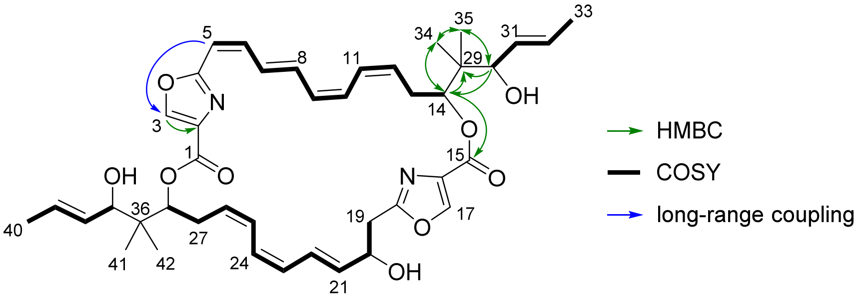


Pos.	δ_H, m (J [Hz])	δ_C, type	COSY	HMBC[a]
1/13	-	161.8, C	-	-
2/14	-	136.0, C	-	-
3/15	8.57, br s	145.6, CH	5/17 (wk)	2/14, 4/16
4/16	-	163.6, C	-	-
5/17	6.15, d (15.2)	116.0, CH	3/15, 6/18	4/16, 6/18, 7/19
6/18	6.91, br s	139.4, CH	5/17	-
7/19	6.12, m	132.0, CH	8/20, 9/21, 10/22	5/17, 6/18, 8/20, 9/21
8/20	6.77, br t (13.1)	136.3, CH	7/19, 9/21, 10/22	6/18, 10/22
9/21	6.10, m	132.9, CH	8/20, 10/22, 11ab/23ab	7/19, 11/23
10/22	5.70, m	132.6, CH	7/19, 8/20, 9/21, 11ab/23ab	8/20, 11/23, 12/24
11a/23a	2.73, m	29.8, CH_2	9/21, 10/22, 11b/23b, 12/24	9/21, 10/22, 12/24, 25/33
11b/23b	2.43, br dd (13.7, 7.3)	-	9/21, 10/22, 11a/23a, 12/24	9/21, 10/22, 12/24, 25/33
12/24	5.16, br d (10.2)	79.1, CH	11ab/23ab	1/13, 10/22, 11/23, 25/33, 26/34, 31/39, 32/40
25/33	-	42.9, C	-	-
26/34	3.91, d (8.0)	78.0, CH	27/35, 28/36, 29/37	12/24, 25/33, 27/35, 28/36, 31/39, 32/40
27/35	5.57, m	129.6, CH	26/34, 28/36, 29/37	25/33, 26/34, 28/36, 29/37, 30/38
28/36	5.73, m	136.9, CH	26/34, 27/35, 29/37, 30/38	25/33, 26/34, 27/35, 29/37, 30/38
29/37	2.09, m	26.7, CH_2	26/34, 27/35, 28/36, 30/38	27/35, 28/36, 30/38
30/38	1.03, t (7.5)	14.2, CH_3	28/36, 29/37	28/36, 29/37
31/39	0.98, s	19.6, CH_3	-	12/24, 25/33, 26/34, 32/40
32/40	1.06, s	19.4, CH_3	-	12/24, 25/33, 26/34, 31/39

[a] Proton showing HMBC correlations to indicated carbons.

HPLC-HRMS analysis of **3** revealed the elemental formula $C_{42}H_{52}N_2O_9$ and a retention time of 13.2 min, which matched disorazole F2 in our in-house database MxBase.[25] Further, the identity of **3** as disorazole F2 was proven by comparison of its ^{1}H NMR data with that of disorazole F2 (Table S6), and reinforced by COSY and HMBC correlations (Table S5).[2]

Table S5 NMR data of disorazole F2 (**3**) in methanol-d_4 (^{1}H 500 MHz; ^{13}C 125 MHz).



Pos.	δ_H, m (J [Hz])	δ_C, type	COSY	HMBC^a
1	-	162.3, C	-	-
2	-	134.4, C	-	-
3	8.30, s	146.3, CH	5, 6	2, 4, 5 (wk)
4	-	163.7, C	-	-
5	6.07, d (11.8)	112.8, CH	3, 6, 7	3 (wk), 4, 6, 7, 8 (wk)
6	6.56, m	138.7, CH	3, 5, 7, 8	4, 8, 9
7	7.14, dd (14.2, 12.2)	130.2, CH	5, 6, 8	5, 6
8	6.79, dd (14.2, 12.2)	136.7, CH	6, 7, 9, 10	5, 6, 7, 9, 10
9	6.03,br t (10.4)	130.2, CH	8, 10, 12	7, 8, 10, 11
10	6.47, m	130.1, CH	8, 9, 11	8, 12
11	6.52, m	128.1, CH	10, 12, 13ab	9, 10, 12, 13, 14
12	5.60, m	132.4, CH	9, 11, 13ab	10, 13, 14
13a	2.35, br dd (12.9, 6.0)	29.5, CH_2	11, 12, 13b, 14	11, 12, 14
13b	2.81, m	-	11, 12, 13a, 14	11, 12, 14, 29
14	5.26, dd (11.2, 1.7)	78.6, CH	13ab	12, 13, 15, 29, 30, 35
15	-	162.2, C	-	-
16	-	133.8, C	-	-
17	8.33, s	146.9, CH	-	16, 18
18	-	164.3, C	-	-
19a	2.83, m	38.3, CH_2	19b, 20	17 (wk), 18, 20, 21
19b	2.91, dd (14.4, 7.4)	-	19a, 20	17 (wk), 18, 20, 21
20	4.30, br dd (11,6, 5.6)	70.5, CH	19ab, 21, 22	18, 19, 21, 22, 23
21	5.83, dd (15.0, 5.6)	136.9, CH	20, 22, 23, 24	19, 20, 23, 24
22	6.49, m	125.8, CH	20, 21, 23, 24	20, 21, 23, 24
23	6.00, br t (10.4)	130.4, CH	21, 22, 24, 25, 26	21, 22, 25
24	6.34, t (11.1)	126.3, CH	22, 23, 25, 26	22, 23, 26
25	6.43, m	128.3, CH	23, 24, 26, 27ab	23, 26, 27, 28
26	5.47, dt (5.5, 10.8)	130.5, CH	23, 24 (wk), 25, 27ab	24, 27, 28
27a	2.30, br dd (12.2, 5.5)	29.4, CH_2	25, 26, 27b, 28	25, 26, 28

Pos.	δ_H, m (J [Hz])	δ_C, type	COSY	HMBC[a]
27b	2.85, m	-	25, 26, 27a, 28	25, 26, 28, 36
28	5.30, dd (11.5, 2.5)	78.7, CH	27ab	1, 26, 27, 36, 37, 41/42
29	-	42.8, C	-	-
30/37	3.86, dd (7.6, 1.5)	77.9, CH	32/39	14/28, 29/36, 31/38, 32/39, 34/41
31/38	5.61, m	131.8, CH	30/37, 32/39, 33/40	30/37, 33/40
32/39	5.66, m	129.8, CH	31/38, 33/41	29/36, 30/37, 31/38, 33/40
33/40	1.71/1.72, br s	18.2, CH$_3$	31/38, 32/39	29/36, 30/37, 31/38, 32/39
34/41	0.98/0.99, s	19.6, CH$_3$	-	29, 30/37, 14, 35/42
35/42	1.03, s	19.5, CH$_3$	-	28, 30/37, 32/39, 34/41, 36
36	-	42.8, C	-	-

[a] Proton showing HMBC correlations to indicated carbons.

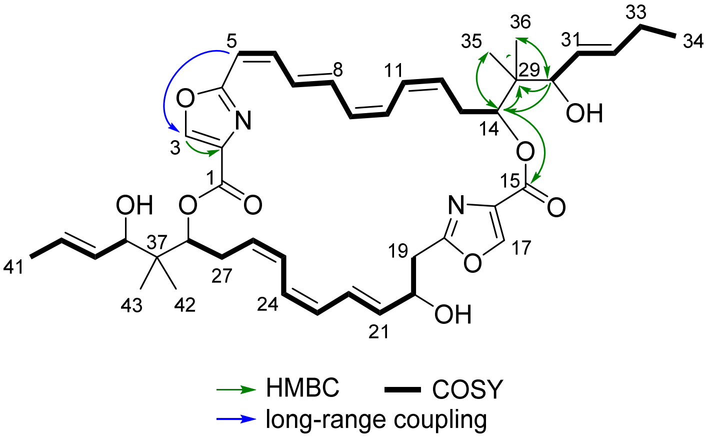
Table S6 Comparison of ^{1}H data of disorazole F2 (**3**) in methanol-d_4 isolated from *S. cellulosum* and *M. xanthus*.[a]

Pos.	$\delta_H{}^b$	$\delta_H{}^c$	$\Delta\delta_H{}^b$-$\delta_H{}^c$
1	-	-	-
2	-	-	-
3	8.32	8.30	0.02
4	-	-	-
5	6.11	6.07	0.04
6	6.59	6.56	0.03
7	7.17	7.14	0.03
8	6.82	6.79	0.03
9	6.07	6.03	0.04
10	6.51	6.47	0.04
11	6.58	6.52	0.06
12	5.51	5.60	0.09
13a	2.88	2.81	0.07
13b	2.39	2.35	0.04
14	5.29	5.26	0.03
15	-	-	-
16	-	-	-
17	8.36	8.33	0.03
18	-	-	-
19a	2.98	2.91	0.07
19b	2.88	2.83	0.05
20	4.34	4.30	0.04
21	5.85	5.83	0.02
22	6.53	6.49	0.04
23	6.03	6.00	0.03
24	6.37	6.34	0.03
25	6.47	6.43	0.04
26	5.64	5.47	0.17
27a	2.85	2.85	0.00
27b	2.34	2.30	0.04
28	5.33	5.30	0.03
29	-	-	-
30	3.90	3.86	0.04
31	5.64	5.66	0.02
32	5.71	5.61	0.10
33	1.76	1.72	0.04
34	1.08	0.98	0.10
35	1.02	1.03	0.01
36	-	-	-
37	3.89	3.86	0.03
38	5.64	5.66	0.02
39	5.71	5.61	0.10
40	1.75	1.72	0.03
41	1.07	0.98/0.99	0.09/0.08
42	1.01	1.03	0.02

[a] ^{1}H at 700 MHz; [b] isolated from *S.* cellulosum So ce12;[2] [c] isolated from *M. xanthus* DK1622::*dis12*_Papr-*disD$_{12}$*_Δ*disE$_{12}$*.

Disorazole F7 (**7**) exhibited a $[M+H]^+$ signal at *m/z* 743.3902, which can be assigned to a molecular formula of $C_{43}H_{54}N_2O_9$. Proton and carbon shifts as well as 2D NMR analysis of **7** in comparison to disorazole F2 (**3**) showed similar shifts and correlations. However, a further methylene group ($\delta_{\text{H-33}}$ 2.08, $\delta_{\text{C-33}}$ 26.7) could be observed, which showed COSY and HMBC correlations to the southern side chain position 30, 31, 32 and 34 ($\delta_{\text{H-30}}$ 3.87, $\delta_{\text{C-30}}$ 78.0; $\delta_{\text{H-31}}$ 5.70, $\delta_{\text{C-31}}$ 136.8; $\delta_{\text{H-32}}$ 5.58, $\delta_{\text{C-32}}$ 129.5; $\delta_{\text{H-34}}$ 1.01, $\delta_{\text{C-34}}$ 14.2). Therefore, compared to **3**, the northern side chain of **7** is prolonged by one methylene unit.

Table S7 NMR data of disorazole F7 (**7**) in methanol-d_4 (^{1}H 500 MHz; ^{13}C 125 MHz).[a]



Pos.	δ_H, m (*J* [Hz])	δ_C, type	COSY	HMBC[b]
1[c]	-	162.3, C	-	-
2	-	134.4, C	-	-
3	8.30, s	146.3, CH	5, 6	2, 4, 5
4	-	163.7, C	-	-
5	6.07, d (11.8)	112.8, CH	3, 6, 7, 8	4, 6, 7
6	6.56, m	138.7, CH	5, 7, 8	4, 7, 8
7	7.14, dd (14.4, 12.4)	130.2, CH	5, 6, 8, 9, 10	5, 6, 9
8	6.79, dd (14.4, 11.1)	136.7, CH	5, 6, 7, 9, 10	5, 6, 7
9	6.47, m	130.1, CH	7, 8, 10, 11	8, 10, 12
10	6.03, t (10.2)	130.2, CH	7, 8, 9, 11, 12	7, 8, 9, 11
11	6.52, m	128.1, CH	9, 10, 12, 13ab	10, 12, 13, 14
12[c]	5.60, m	132.4, CH	11, 13ab	10, 13, 14
13a[c]	2.35, m	29.5, CH_2	11, 12, 13b, 14	11, 12, 14
13b[c]	2.82, m	-	11, 12, 13a, 14	11, 12, 14
14[c]	5.26, m	78.6, CH	13ab	12, 13, 15, 29, 30, 35, 36
15[c]	-	162.2, C	-	-
16	-	133.8, C	-	-
17	8.34, s	146.9, CH	-	16, 18
18	-	164.3, C	-	-
19	2.87, m	38.3, CH_2	20	18, 20, 21

Pos.	δ_H, m (*J* [Hz])	δ_C, type	COSY	HMBC[b]
20	4.30, dd (11.1, 5.7)	70.5, CH	19, 21, 22	18, 19, 21, 22
21	5.83, dd (14.9, 5.7)	136.9, CH	20, 22, 23, 24	19, 20, 22, 23, 24
22	6.49, m	125.8, CH	20, 21, 23, 24	19, 20, 23, 24
23	6.34, t (11.1)	126.3, CH	21 (wk), 22, 24, 25, 26 (wk)	22, 24, 26
24	5.99, t (11.1)	130.4, CH	21, 22, 23, 26	21, 22, 25
25	6.44, m	128.3, CH	23, 25, 26, 27ab	23, 24, 26, 27, 28
26	5.46, m	130.5, CH	24, 25, 27ab	23, 27, 28
27a	2.31, m	29.4, CH_2	25, 26, 27b, 28	25, 26
27b	2.84, m	-	25, 26, 27a, 28	25, 26, 28
28[c]	5.30, dt (11.5, 2.6)	78.7, CH	27ab	1, 26, 27, 37, 38, 42, 43
29[c]	-	42.8, C	-	-
30[c]	3.87, m	77.95 CH	32, 33 (wk)	14, 29, 31, 32, 35/42, 36/43
31[c]	5.70, m	136.8, CH	30, 32, 33	29 (wk), 30, 32, 33, 34
32[c]	5.58, m	129.5, CH	30, 31, 33	30, 31, 33
33	2.08	26.7, CH_2	30, 31, 32, 34	30, 31, 32, 34
34[c]	1.01, t (7.4)	14.2, CH_3	33	31, 33
35/42[c]	0.98, m	19.7, CH_3	-	14/28, 29/37, 30/38, 36/43
36/43[c]	1.04, dt (3.2, 1.7)	19.5, CH_3	-	14/28, 29/37, 30/38, 35/42
37[c]	-	42.8, C	-	-
38[c]	3.86, m	78.03, CH	40, 41	28, 35/42, 36/43, 37, 39, 40
39[c]	5.70, m	129.8, CH	38, 40, 41	38, 40, 41
40[c]	5.61, m	132.8, CH	38, 39, 41	38, 39, 41
41	1.72, dd (6.2, 1.0)	18.2, CH_3	38, 39, 40	37, 38, 39, 40

[a] NMR data of the major component; major and minor components are present in a ratio of 2:1; [b] Proton showing HMBC correlations to indicated carbons; [c] signal doubling indicates the difference of the major and minor to be the interchange of the side chain at position C-14 and C-28.

3.6.2 Supplementary tables

Table S8 Oligonucleotides used in this study.

No.	Name	Sequence (5' to 3')	Application
1	del427DEF-apr-disD12-F1	CTCCGGCGCTCGATCAGGTCGTCGCTCACCTCCGGGAAGATCTCGCGTGAACGCTCAGTGGAACGAGGTTCAC	construct *Papr-apr-disD*$_{12}$ cassette to replace *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$ by *disD*$_{12}$
2	del427DEF-apr-disD12-R1	TGCCCGGGAAACATGTATGCTTTCATAATCTGTACCTCCTTAAG	
3	del427DEF-apr-disD12-F2	CTTAAGGAGGTACAGATTATGAAAGCATACATGTTTCCCGGGCA	
4	del427DEF-apr-disD12-R2	ATTTGATGCCTGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTCTACCGCGAGAGCGCGCGCATGC	
5	del427D-apr-disD12-F	CGGTCGGCGGTCCCTGGGTCCCCGCCAGCGCTGTCCAGGGCAGCGCCTAGACGCTCAGTGGAACGAGGTTCAC	replace *disD*$_{427}$ by *disD*$_{12}$
6	del427D-apr-disD12-R	TTTTGCAAGCATGACGATGACCTCCGGTCGAAGCTTCGACTCTTTACGGTCTACCGCGAGAGCGCGCGCATGC	
7	del427D-apr-disD12AT-Lin-R	TTTTGCAAGCATGACGATGACCTCCGGTCGAAGCTTCGACTCTTTACGGTCTACCGCCTGGCCCCCCTCGGCGCGG	replace *disD*$_{427}$ by truncated *disD*$_{12}$
8	del427D-apr-disD12AT-R	TTTTGCAAGCATGACGATGACCTCCGGTCGAAGCTTCGACTCTTTACGGTCTAGTCTCGACGGATCTGGGCGTACA	
9	apr-disF427-F	AGGCCTCGGCCGATCTCCTGGACCACCGCCTGCGCGCGCTGTCGCGGTAAACGCTCAGTGGAACGAGGTTCAC	insert *Papr-apr* cassette upstream of *disF*$_{427}$
10	apr-disF427-R	CCTCCGAGACCCCTCGGGTTCCGCGCCTGAGACAACATGAGTTTTGCAAGCATAATCTGTACCTCCTTAAGTC	
11	apr-disF427-delOr-F	CTGCCCCCTACGAGCCGGCGCGCCCGGGCGCCGCGGCGCCTGTCAGGAGGTAAACGCTCAGTGGAACGAGGTTCAC	knockout the Or of *disD*$_{427}$
12	apr-disE427-R	CCTCCCATCCCTGGCGACTGCGACCCCTGCCCCGGAAACATGACGGCCTTCATAATCTGTACCTCCTTAAG	knockout *disE*$_{427}$
13	apr-del427DEF-R	ATTTGATGCCTGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTCATAATCTGTACCTCCTTAAG	knockout *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$
14	Papr-apr-disE427-disF427-F1	CCCCGGCGCTCCACCAGGTGGTCGCGCACGTCCGCGAGGCGCTTTCATGAACGCTCAGTGGAACGAGGTTCAC	construct *Papr-apr-disE*$_{427}$-*disF*$_{427}$ cassette to replace *disE*$_{12}$
15	Papr-apr-disE427-disF427-R1	CGCCTTCGATCTGTCGTCGATGCCATAATCTGTACCTCCTTAAG	
16	Papr-apr-disE427-disF427-F2	CTTAAGGAGGTACAGATTATGGCATCGACGACAGATCGAAGGCG	
17	Papr-apr-disE427-disF427-R2	GACAACATGAGTTTTGCAAGCATAAGCGCCTCCTAGATGAGAGAGC	
18	Papr-apr-disE427-disF427-F3	GCTCTCTCATCTAGGAGGCGCTTATGCTTGCAAAACTCATGTTGTC	
19	Papr-apr-disE427-disF427-R3	GACCCTTGCCCGGGAAACATGTATGCTTTCATCGGTTCTTCCTCGGCCCCCTACGCGCTCCGCTGGGCGATGG	
20	apr-disD12-F	TCGAGAAAGGTCCCTGGATACCGTCGTTCAAGGCGCAGAGCTGGGCATAGACGCTCAGTGGAACGAGGTTCAC	insert *Papr-apr* cassette upstream of *disD*$_{12}$
21	apr-disD12-R	CGTCCCATCCCCTTCGCCTGAGACCCTTGCCCGGGAAACATGTATGCTTTCATAATCTGTACCTCCTTAAGTC	
22	apr-disE12-F	CCCCGGCGCTCCACCAGGTGGTCGCGCACGTCCGCGAGGCGCTTTCATGAACGCTCAGTGGAACGAGGTTCAC	knockout *disE*$_{12}$
23	apr-disD12-R	ATTTGATGCCTGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTCATAATCTGTACCTCCTTAAGTC	knockout *disD*$_{12}$
24	amp-Pvan-His-TEV-disE427-F1	CGCGCCACGCGCGCTCGCCGCTCGCGCGGCCCAGGACGCGAACGCAATGGTTACCAATGCTTAATCAGTGAGG	construct *amp-Pvan-His-TEV* cassette for fusion expression with *disE*$_{427}$
25	amp-Pvan-His-TEV-disE427-R2	GCCCTGAAAATACAGGTTTTCGGTCGTTGGGATATCGTAATCGTGATGGTGATGGTGATGGTAGTACGACATATGCGTTTCCTCGCATCGTG	
26	amp-Pvan-His-TEV-disE427-R3	GTCTCGGGGAACTCGTCGTGAATCTCACGCCTTCGATCTGTCGTCGATGCGCCCTGAAAATACAGGTTTTC	

No.	Name	Sequence (5' to 3')	Application
27	Pvan-disE427-F	CCAATAGGTCGCCGAACCACGATGCGAGGAAACGCATATGGCATCGACGACAGATCGAAGGCG	construct the plasmid p15A-km-tps-Pvan-disE427
28	Pvan-disE427-R	TGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTCTCCTAGATGAGAGAGCCCTGCT	
29	Pvan-disE12-F	CCAATAGGTCGCCGAACCACGATGCGAGGAAACGCATATGACATCGATGGCGCGACACCTGGA	construct the plasmid p15A-km-tps-Pvan-disE12
30	Pvan-disE12-R	TGGAGATCCTTAAGATCCGTAACTTGGGGTTTTCGGATCTCTATGCCCAGCTCTGCGCCTTGA	
31	amp-Ptet-delE427-F	CGCGCCACGCGCGCTCGCCGCTCGCGCGGCCCAGGACGCGAACGCAATGGTTACCAATGCTTAATCAGTGAGG	knockout *disE*$_{427}$
32	amp-Ptet-delE427-R	CCCATCCCTGGCGACTGCGACCCCTGCCCCGGAAACATGACGGCCTTCATAGATCCTTTCTCCTCTTTAGATC	
33	MXAN_0683-F	GTGAATGCGCAAACCAACCCTTGGACCTGATTGAACGCGTCCTGGT	construct the plasmid pR6K-tet-MXAN_0683
34	MXAN_0683-R	TCAACAGGTTGAACTGTTTAAACTGATGAGCACCGTCTCCATCATC	
35	MXAN_1743-F	GTGAATGCGCAAACCAACCCTTGGTGCTCAATGACCCCACCCTCAT	construct the plasmid pR6K-tet-MXAN_1743
36	MXAN_1743-R	TCAACAGGTTGAACTGTTTAAACCGCAGGTACACGATGGCGAAGTC	
37	MXAN_2304-F	GTGAATGCGCAAACCAACCCTTGCTCCAAGACCATGCGCGCAACTA	construct the plasmid pR6K-tet-MXAN_2304
38	MXAN_2304-R	TCAACAGGTTGAACTGTTTAAACGAACTGGTTGCCGATGCACTGAC	
39	MXAN_3943-F	GTGAATGCGCAAACCAACCCTTGGATCGAGGTCATCTTCAAGGGTG	construct the plasmid pR6K-tet-MXAN_3943
40	MXAN_3943-R	TCAACAGGTTGAACTGTTTAAACGAAGGGGAAGTAGTGGAACGGAT	
41	MXAN_4127-F	GTGAATGCGCAAACCAACCCTTGGAGATTGTCGCGTTGCTCCACGA	construct the plasmid pR6K-tet-MXAN_4127
42	MXAN_4127-R	TCAACAGGTTGAACTGTTTAAACTGCTCAAGGCGGCAGGCGAGTTC	
43	MXAN_4919-F	GTGAATGCGCAAACCAACCCTTGATACCACACGCTTCTCGTCCGCA	construct the plasmid pR6K-tet-MXAN_4919
44	MXAN_4919-R	TCAACAGGTTGAACTGTTTAAACGGACGGTGACGCTCCTCAGGAAG	
45	MXAN_7298-F	GTGAATGCGCAAACCAACCCTTGCCATGCTGAACATGTGGCTGGTC	construct the plasmid pR6K-tet-MXAN_7298
46	MXAN_7298-R	TCAACAGGTTGAACTGTTTAAACGTCGTGATTTCGAACTGGAGCCG	
47	MXAN_0683-chk-F	TGACGGTTGGGTCTCAAGAGCTG	colony PCR for MXAN_0683 inactivation
48	MXAN_0683-chk-R	AGGTATGCCGAGCGGAAGAACTC	
49	MXAN_1743-chk-F	CCTTGGAGACACACGCTCATGCC	colony PCR for MXAN_1743 inactivation
50	MXAN_1743-chk-R	CAAGGTCATGCGGATTCCACTGC	
51	MXAN_2304-chk-F	GATGAACACCCCATGCCCAGCAC	colony PCR for MXAN_2304 inactivation
52	MXAN_2304-chk-R	CTGCTGCAACCCGTCGTCATCAG	
53	MXAN_3943-chk-F	ACGCCTTGTCAAAACCACCGCCT	colony PCR for MXAN_3943 inactivation
54	MXAN_3943-chk-R	TGATGTTCACCGCGCTGGAGATG	
55	MXAN_4127-chk-F	AACGGACTCAAGGAGACCGCTTC	colony PCR for MXAN_4127 inactivation
56	MXAN_4127-chk-R	CATACGTGCAACATGGCGAAGCC	
57	MXAN_4919-chk-F	CCCAACAAACCGTTGAGGTCAGC	colony PCR for MXAN_4919 inactivation
58	MXAN_4919-chk-R	GAACCCGAAGGAATGCGAACCTG	
59	MXAN_7298-chk-F	TGCCTGTGTCCGTGCAAAAGCAG	colony PCR for MXAN_7298 inactivation
60	MXAN_7298-chk-R	GGGTCCACTCCACTTTCAGGTGC	
61	tet-5out	GAATGGACGATATCCCGCAAGAG	colony PCR for P450 enzyme gene inactivation
62	tet-3out	ACGGGTTGGCATGGATTGTAGGC	

Table S9 *M. xanthus* DK1622 mutants generated in this study.

No.	Name	Description
1	*M. xanthus* DK1622::*dis427*	integration of the disorazole Z biosynthetic gene cluster from *Sorangium cellulosum* strain So ce427 into the genome of *Myxococcus xanthus* DK1622, *km*^{R}
2	*M. xanthus* DK1622::*dis12*	integration of the disorazole A biosynthetic gene cluster from *Sorangium cellulosum* strain So ce12 into the genome of *Myxococcus xanthus* DK1622, *km*^{R}
3	DK1622::*dis427*_Δ*disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$_*disD*$_{12}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$ was replaced by *disD*$_{12}$
4	DK1622::*dis427*_Δ*disD*$_{427}$_*disD*$_{12}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disD*$_{427}$ was replaced by *disD*$_{12}$
5	DK1622::*dis427*_Δ*disD*$_{427}$_AT$_{12}$-Linker$_{12}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disD*$_{427}$ was replaced by the AT domain containing the linker region from *disD*$_{12}$
6	DK1622::*dis427*_Δ*disD*$_{427}$_AT$_{12}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disD*$_{427}$ was replaced by the AT domain from *disD*$_{12}$
7	DK1622::*dis427*_Papr-*disF*$_{427}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the Papr promoter was inserted upstream of *disF*$_{427}$
8	DK1622::*dis427*_Papr-*disF*$_{427}$_ΔOr	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the Papr promoter was inserted upstream of *disF*$_{427}$, the Or domain was deleted
9	DK1622::*dis427*_Δ*disD*$_{427}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disD*$_{427}$ was deleted
10	DK1622::*dis427*_Δ*disE*$_{427}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{427}$ was deleted
11	DK1622::*dis427*_Δ*disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$ was deleted
12	DK1622::*dis12*_Δ*disE*$_{12}$_*disE*$_{427}$-*disF*$_{427}$	integration of the engineered *dis12* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{12}$ was replaced by *disE*$_{427}$-*disF*$_{427}$
13	DK1622::*dis12*_Papr-*disD*$_{12}$	integration of the engineered *dis12* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the Papr promoter was inserted upstream of *disD*$_{12}$
14	DK1622::*dis12*_Papr-*disD*$_{12}$_Δ*disE*$_{12}$	integration of the engineered *dis12* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the Papr promoter was inserted upstream of *disD*$_{12}$, the *disE*$_{12}$ was deleted
15	DK1622::*dis12*_Δ*disD*$_{12}$	integration of the engineered *dis12* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disD*$_{12}$ was deleted
16	DK1622::*dis12*_Δ*disE*$_{12}$-*disD*$_{12}$	integration of the engineered *dis12* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{12}$-*disD*$_{12}$ was deleted
17	DK1622::*dis427*_Pvan-His-TEV-*disE*$_{427}$	integration of the engineered *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} fusion expression of a N-terminal 6xHis tag and a TEV protease cleavage site with *disE*$_{427}$ under the control of a Pvan promoter
18	DK1622::Pvan-*disE*$_{12}$	integration of the *disE*$_{12}$ into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{12}$ is under the control of a Pvan promoter
19	DK1622::Pvan-*disE*$_{427}$	integration of the *disE*$_{427}$ into the genome of *M. xanthus* DK1622, *km*^{R} the *disE*$_{427}$ is under the control of a Pvan promoter
20	DK1622::*dis427*_ΔMXAN0683	integration of the *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the P450 enzyme encoding gene MXAN_0683 was inactivated, *tet*^{R}
21	DK1622::*dis427*_ΔMXAN1743	integration of the *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the P450 enzyme encoding gene MXAN_1743 was inactivated, *tet*^{R}
22	DK1622::*dis427*_ΔMXAN2304	integration of the *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the P450 enzyme encoding gene MXAN_2304 was inactivated, *tet*^{R}
23	DK1622::*dis427*_ΔMXAN4919	integration of the *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the P450 enzyme encoding gene MXAN_4919 was inactivated, *tet*^{R}
24	DK1622::*dis427*_ΔMXAN7298	integration of the *dis427* BGC into the genome of *M. xanthus* DK1622, *km*^{R} the P450 enzyme encoding gene MXAN_7298 was inactivated, *tet*^{R}

3.6.3 Supplementary figures

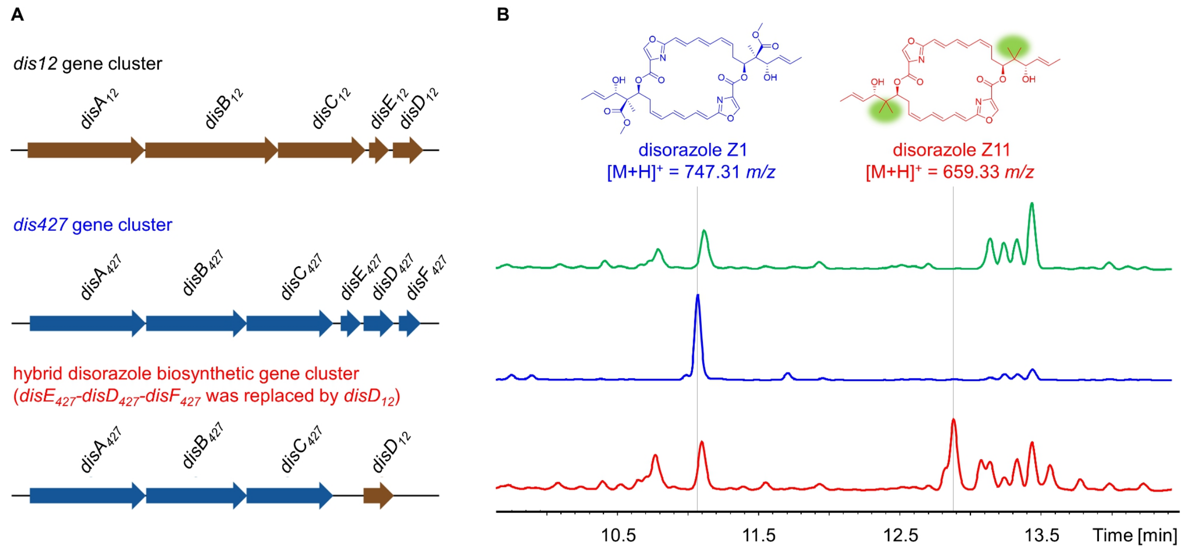


Figure S1 Heterologous expression of a hybrid disorazole Z biosynthetic gene cluster. (A) Scheme of disorazole A biosynthetic gene cluster (*dis12*), disorazole Z biosynthetic gene cluster (*dis427*), and hybrid thereof. (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (green trace), DK1622::*dis427* (blue trace), and DK1622::*dis427*_Δ*disE*$_{427}$-*disD*$_{427}$-*disF*$_{427}$_*disD*$_{12}$ (red trace). The chromatograms are shown as UV absorption at 320 nm.

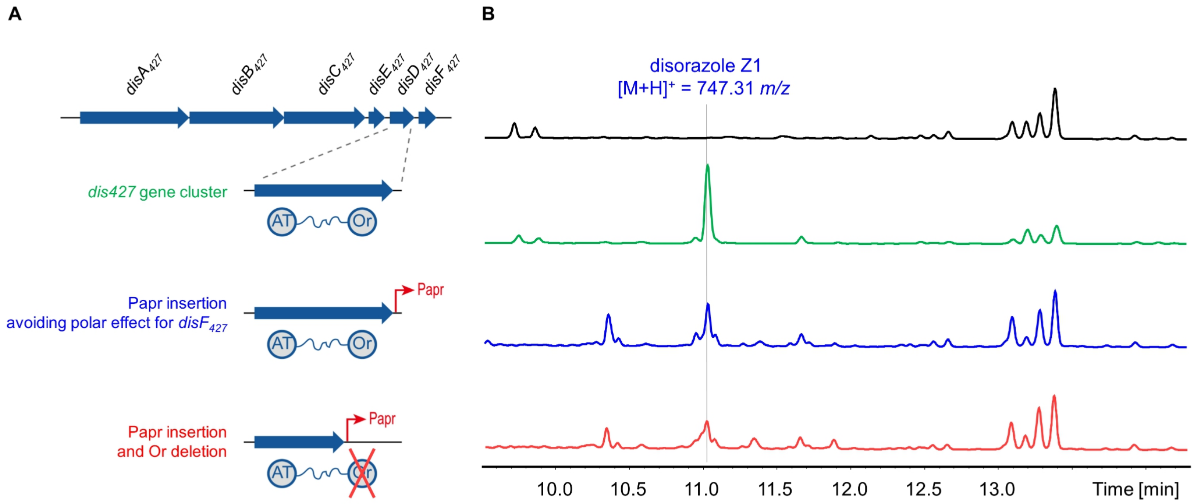


Figure S2 Knock-out of the oxidoreductase domain of *disD*$_{427}$. (A) Scheme of disorazole Z biosynthetic gene cluster (*dis427*) and its engineering. (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (black trace), DK1622::*dis427* (green trace), DK1622::*dis427*_Papr-*disF*$_{427}$ (blue trace), and DK1622::*dis427*_Papr-*disF*$_{427}$_ΔOr (red trace). The chromatograms are shown as UV absorption at 320 nm.

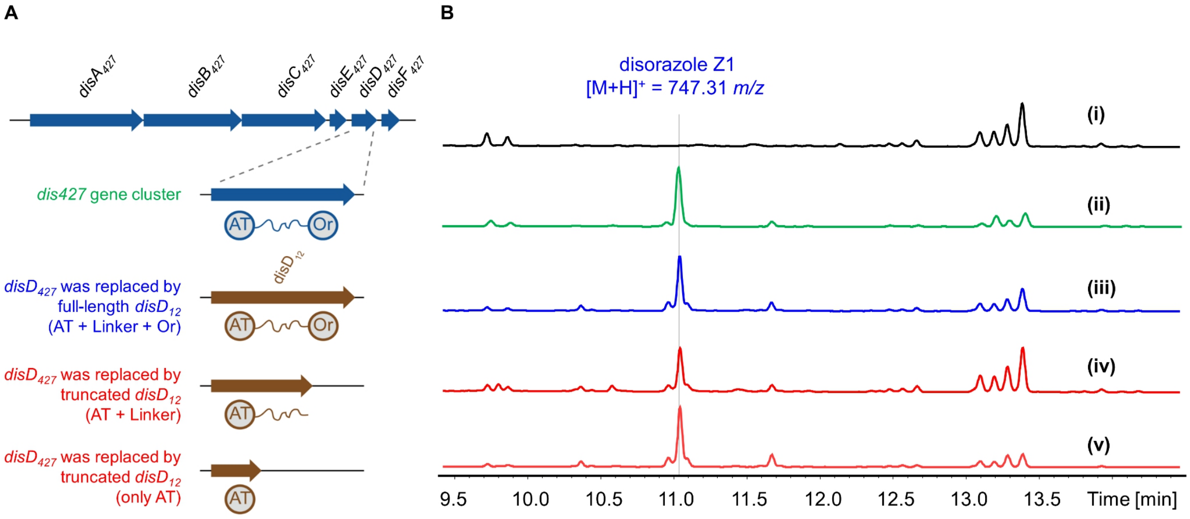


Figure S3 Replacement of *disD*$_{427}$ by full-length or truncated *disD*$_{12}$. (A) Scheme of disorazole Z biosynthetic gene cluster (*dis427*) and its engineering. (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (black trace, i), DK1622::*dis427* (green trace, ii), DK1622::*dis427*_Δ*disD*$_{427}$_*disD*$_{12}$ (blue trace, iii), DK1622::*dis427*_Δ*disD*$_{427}$_AT$_{12}$-Linker$_{12}$ (red trace, iv) and DK1622::*dis427*_Δ*disD*$_{427}$_AT$_{12}$ (red trace, v). The chromatograms are shown as UV absorption at 320 nm.

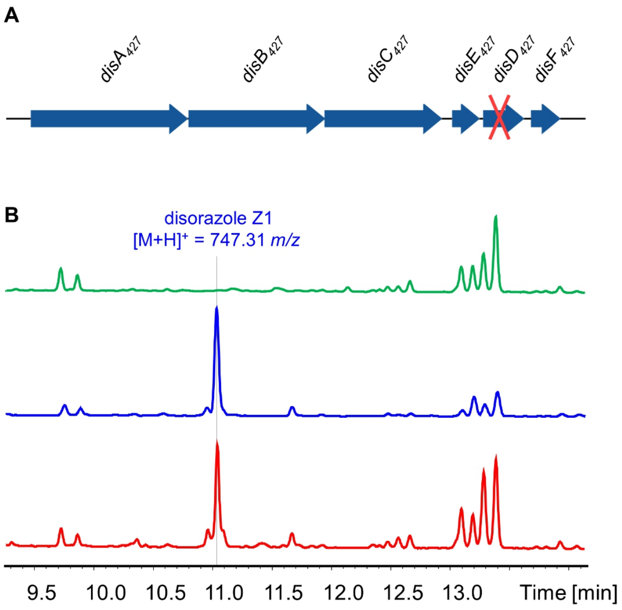


Figure S4 Knock-out of the *disD*$_{427}$. (A) Deletion of the *disD*$_{427}$ gene in disorazole Z biosynthetic gene cluster (*dis427*). (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (green trace), DK1622::*dis427* (blue trace), and DK1622::*dis427*_Δ*disD*$_{427}$ (red trace), chromatograms are shown as UV absorption at 320 nm.

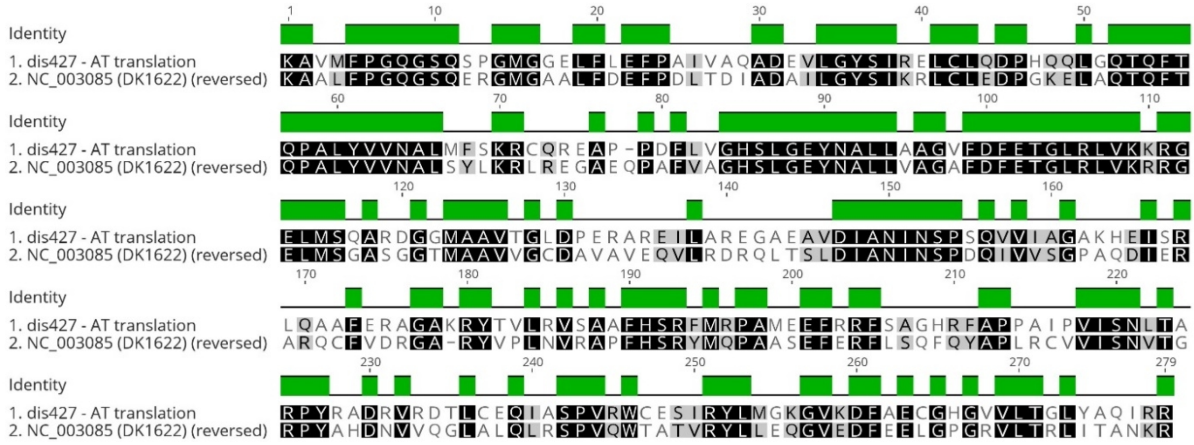


Figure S5 BLAST search of the disorazole Z AT domain against the *M. xanthus* DK1622 genome. The myxovirescin *trans*-AT TaV (MXAN_3942) from *M. xanthus* DK1622 was found to share 55.2% identity and 67.4 similarity with the disorazole AT.

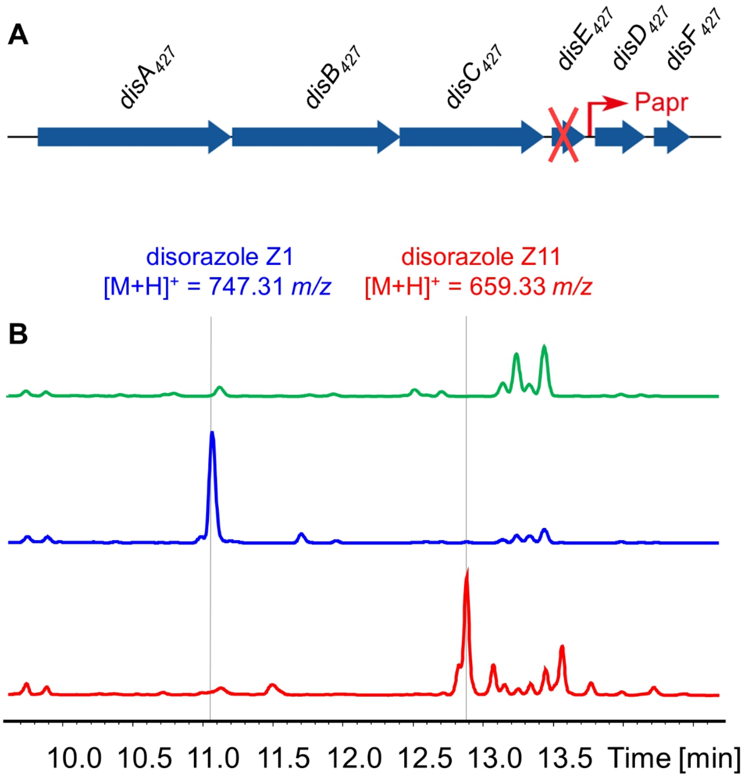


Figure S6 Knock-out of the *disE*$_{427}$. (A) Deletion of the *disE*$_{427}$ gene in disorazole Z biosynthetic gene cluster (*dis427*). (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (green trace), DK1622::*dis427* (blue trace), and DK1622::*dis427*_Δ*disE*$_{427}$ (red trace), chromatograms are shown as UV absorption at 320 nm.

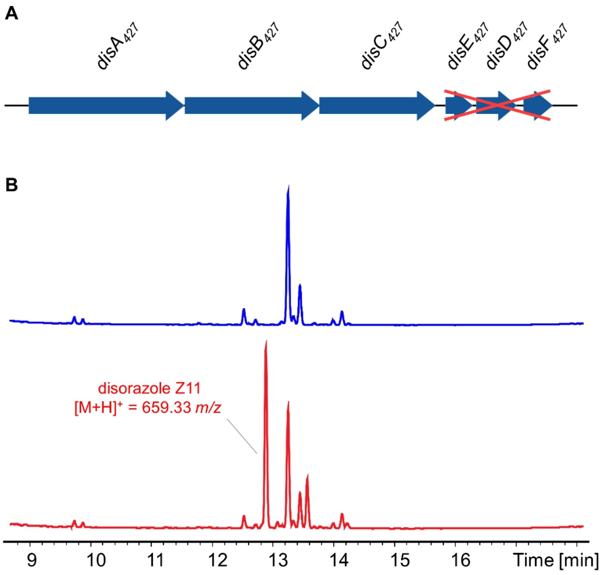


Figure S7 Knock-out of the *disE*427-*disD*427-*disF*427. (A) Deletion of *disE*427, *disD*427, and *disF*427 genes in disorazole Z biosynthetic gene cluster (*dis427*). (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (blue trace) and DK1622::*dis427_ΔdisE*427-*disD*427-*disF*427 (red trace). The chromatograms are shown as UV absorption at 320 nm.

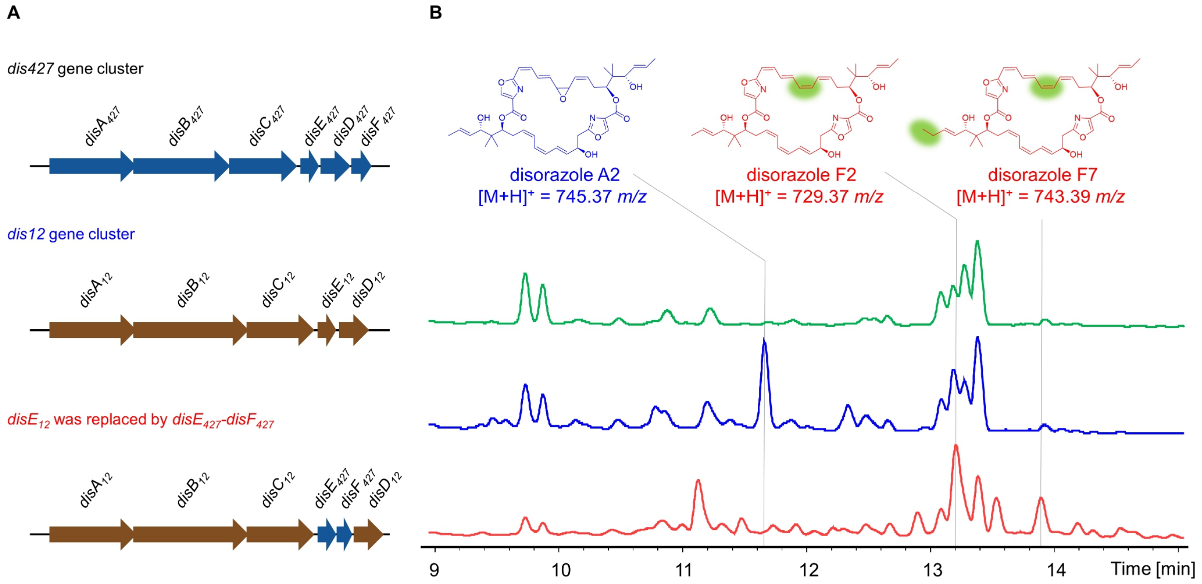


Figure S8 Heterologous expression of a hybrid disorazole A biosynthetic gene cluster. (A) Scheme of disorazole Z biosynthetic gene cluster (*dis427*), disorazole A biosynthetic gene cluster (*dis12*), and hybrid thereof. (B) HPLC-MS analysis of crude extracts from wild-type *M. xanthus* DK1622 (green trace), DK1622::*dis12* (blue trace), and DK1622::*dis12_ΔdisE*12*_disE*427-*disF*427 (red trace). The chromatograms are shown as UV absorption at 320 nm.

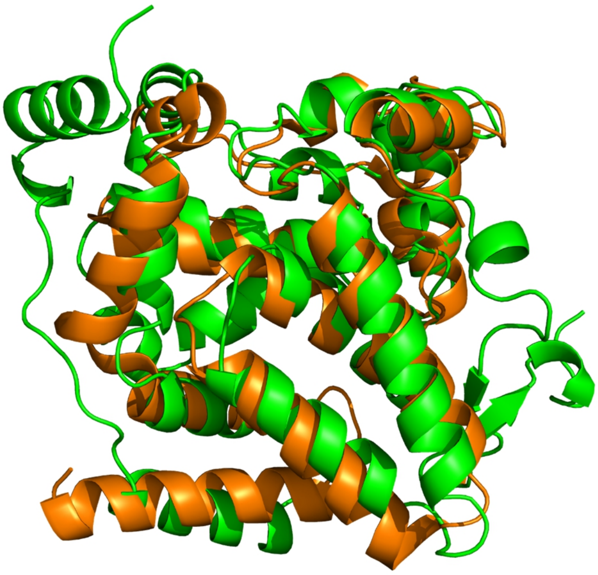


Figure S9 Superimposition of DisE$_{427}$ and Alr8543. Green color: the structure of DisE$_{427}$ predicted by ColabFold;[26] Orange color: the structure of Alr8543 (PDB: 6E12) from Cyanobacteria *Nostoc* sp. PCC 7120.

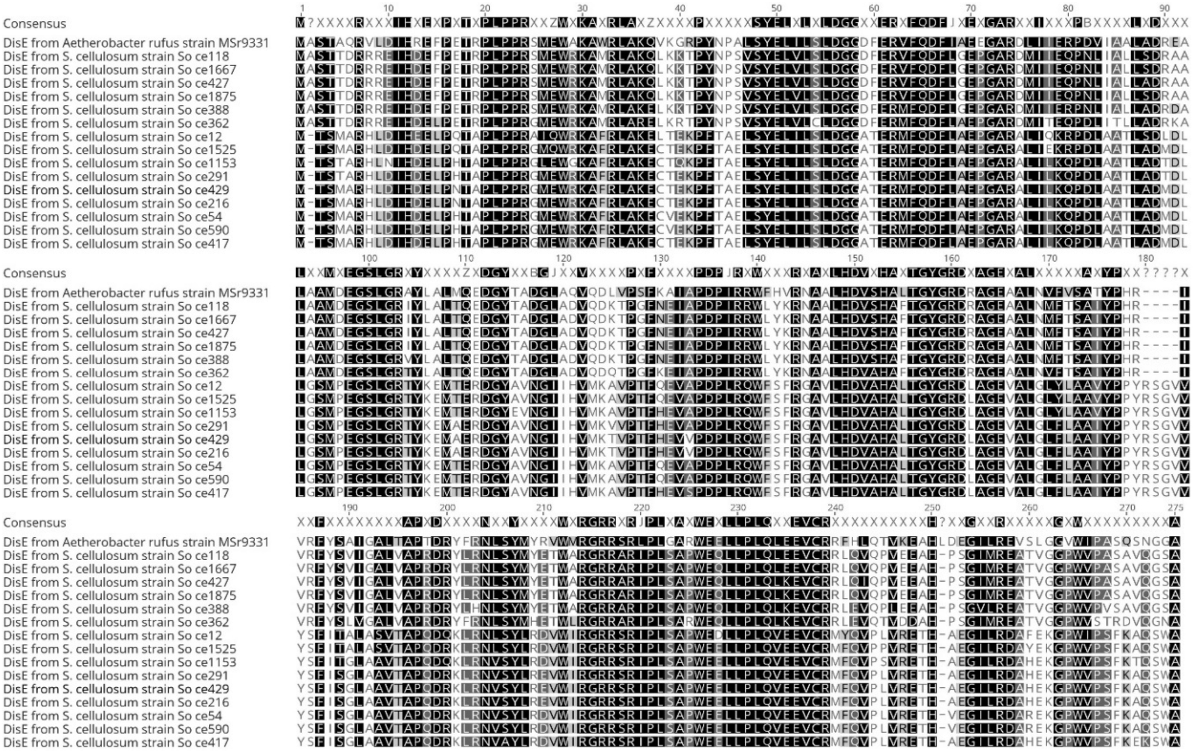
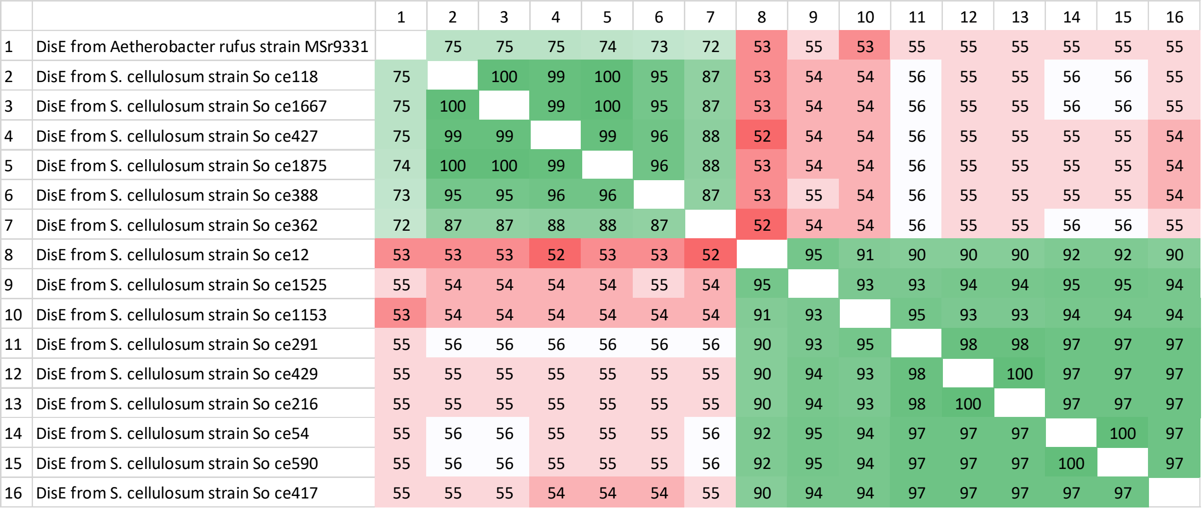


Figure S10 Alignment of 16 DisE proteins from myxobacteria. Multiple alignment was performed using the MAFFT v7.490 integrated in Geneious Prime 2023.0.4 (https://www.geneious.com).[27]



		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	DisE from Aetherobacter rufus strain MSr9331		75	75	75	74	73	72	53	55	53	55	55	55	55	55	55
2	DisE from S. cellulosum strain So ce118	75		100	99	100	95	87	53	54	54	56	55	55	56	56	55
3	DisE from S. cellulosum strain So ce1667	75	100		99	100	95	87	53	54	54	56	55	55	56	56	55
4	DisE from S. cellulosum strain So ce427	75	99	99		99	96	88	52	54	54	56	55	55	55	55	54
5	DisE from S. cellulosum strain So ce1875	74	100	100	99		96	88	53	54	54	56	55	55	55	55	54
6	DisE from S. cellulosum strain So ce388	73	95	95	96	96		87	53	55	54	56	55	55	55	55	54
7	DisE from S. cellulosum strain So ce362	72	87	87	88	88	87		52	54	54	56	55	55	56	56	55
8	DisE from S. cellulosum strain So ce12	53	53	53	52	53	53	52		95	91	90	90	90	92	92	90
9	DisE from S. cellulosum strain So ce1525	55	54	54	54	54	55	54	95		93	93	94	94	95	95	94
10	DisE from S. cellulosum strain So ce1153	53	54	54	54	54	54	54	91	93		95	93	93	94	94	94
11	DisE from S. cellulosum strain So ce291	55	56	56	56	56	56	56	90	93	95		98	98	97	97	97
12	DisE from S. cellulosum strain So ce429	55	55	55	55	55	55	55	90	94	93	98		100	97	97	97
13	DisE from S. cellulosum strain So ce216	55	55	55	55	55	55	55	90	94	93	98	100		97	97	97
14	DisE from S. cellulosum strain So ce54	55	56	56	55	55	55	56	92	95	94	97	97	97		100	97
15	DisE from S. cellulosum strain So ce590	55	56	56	55	55	55	56	92	95	94	97	97	97	100		97
16	DisE from S. cellulosum strain So ce417	55	55	55	54	54	54	55	90	94	94	97	97	97	97	97	

Figure S11 Sequence identities (%) of 16 DisE proteins from myxobacteria.

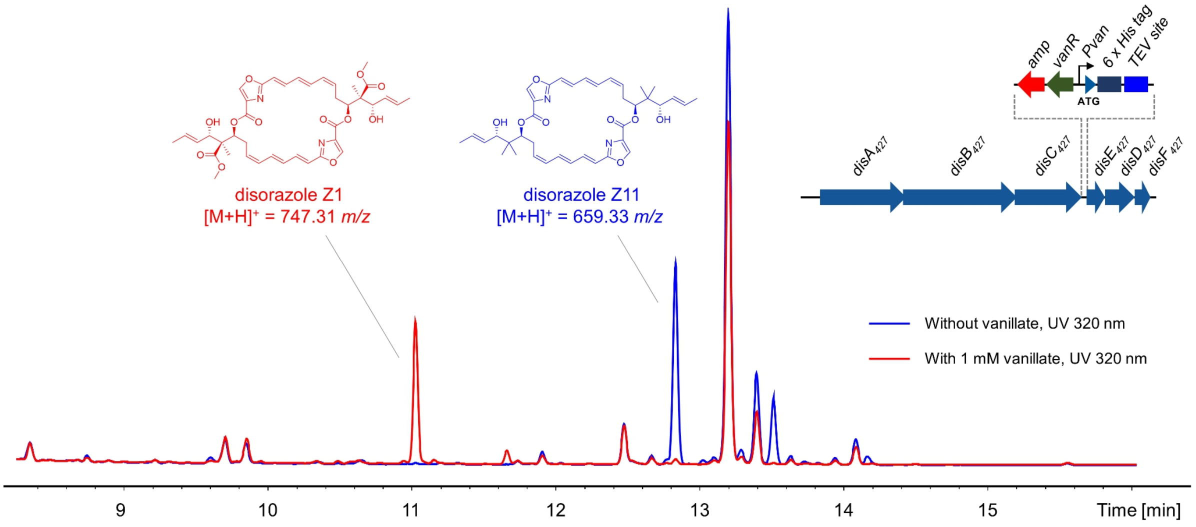
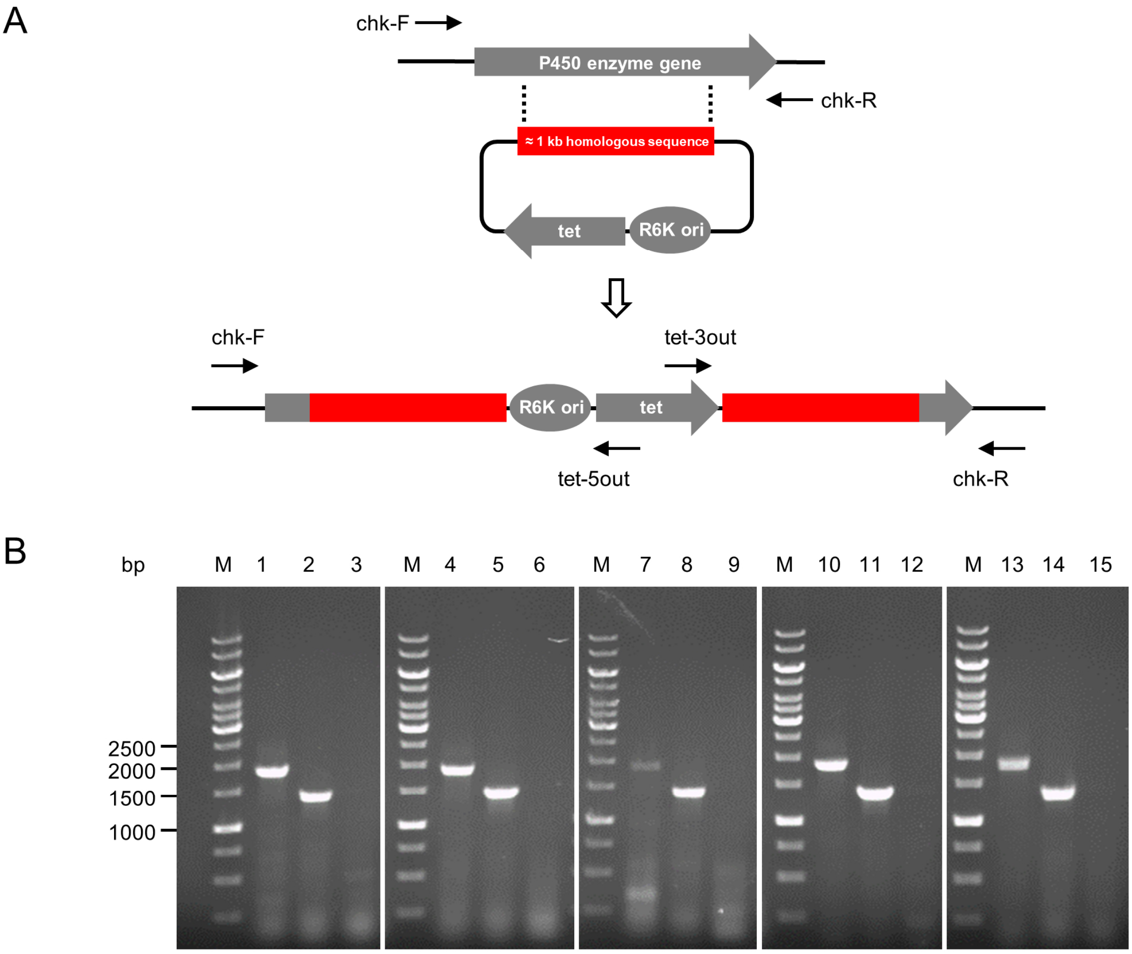


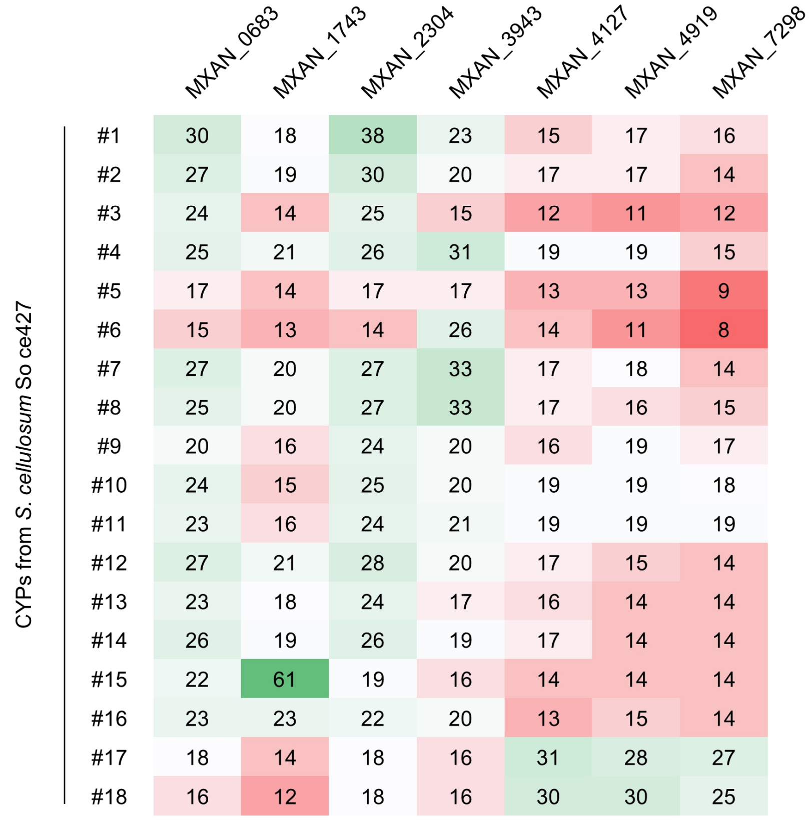
Figure S12 HPLC-MS analysis of crude extracts of DK1622::*dis427*_Pvan-His-TEV-*disE*$_{427}$. The strain DK1622::*dis427*_Pvan-His-TEV-*disE*$_{427}$ containing the engineered disorazole Z biosynthetic gene cluster (as depicted in the upper right corner) was cultivated in the absence of vanillate (blue trace) or with 1 mM vanillate (red trace). The chromatograms are shown as UV absorption at 320 nm. *amp*: ampicillin resistance gene; Pvan: vanillate-inducible promoter; *vanR*: regulatory gene of the Pvan promoter; TEV site: Tobacco Etch Virus protease recognition and cleavage sequence.



C

	Template DNA	Forward primer	Reverse primer	Size of amplicons
1	DK1622::dis$_{427}$_ΔMXAN0683	MXAN_0683-chk-F	tet-5out	1878 bp
2	DK1622::dis$_{427}$_ΔMXAN0683	tet-3out	MXAN_0683-chk-R	1425 bp
3	wild-type *M. xanthus* DK1622	MXAN_0683-chk-F	MXAN_0683-chk-R	1488 bp
4	DK1622::dis$_{427}$_ΔMXAN1743	MXAN_1743-chk-F	tet-5out	1861 bp
5	DK1622::dis$_{427}$_ΔMXAN1743	tet-3out	MXAN_1743-chk-R	1427 bp
6	wild-type *M. xanthus* DK1622	MXAN_1743-chk-F	MXAN_1743-chk-R	1460 bp
7	DK1622::dis$_{427}$_ΔMXAN2304	MXAN_2304-chk-F	tet-5out	1882 bp
8	DK1622::dis$_{427}$_ΔMXAN2304	tet-3out	MXAN_2304-chk-R	1375 bp
9	wild-type *M. xanthus* DK1622	MXAN_2304-chk-F	MXAN_2304-chk-R	1452 bp
10	DK1622::dis$_{427}$_ΔMXAN4919	MXAN_4919-chk-F	tet-5out	1832 bp
11	DK1622::dis$_{427}$_ΔMXAN4919	tet-3out	MXAN_4919-chk-R	1357 bp
12	wild-type *M. xanthus* DK1622	MXAN_4919-chk-F	MXAN_4919-chk-R	1378 bp
13	DK1622::dis$_{427}$_ΔMXAN7298	MXAN_7298-chk-F	tet-5out	1761 bp
14	DK1622::dis$_{427}$_ΔMXAN7298	tet-3out	MXAN_7298-chk-R	1302 bp
15	wild-type *M. xanthus* DK1622	MXAN_7298-chk-F	MXAN_7298-chk-R	1281 bp

Figure S13 Inactivation of P450 enzyme-encoding genes in *M. xanthus* strains. (A) Scheme of gene inactivation by single-crossover homologous recombination; (B) PCR verification of inactivation mutants of *M. xanthus*; (C) The template, primers, and size of amplicons of colony PCR; M: marker, tet: tetracycline resistance gene.



CYPs from *S. cellulosum* So ce427

	MXAN_0683	MXAN_1743	MXAN_2304	MXAN_3943	MXAN_4127	MXAN_4919	MXAN_7298
#1	30	18	38	23	15	17	16
#2	27	19	30	20	17	17	14
#3	24	14	25	15	12	11	12
#4	25	21	26	31	19	19	15
#5	17	14	17	17	13	13	9
#6	15	13	14	26	14	11	8
#7	27	20	27	33	17	18	14
#8	25	20	27	33	17	16	15
#9	20	16	24	20	16	19	17
#10	24	15	25	20	19	19	18
#11	23	16	24	21	19	19	19
#12	27	21	28	20	17	15	14
#13	23	18	24	17	16	14	14
#14	26	19	26	19	17	14	14
#15	22	61	19	16	14	14	14
#16	23	23	22	20	13	15	14
#17	18	14	18	16	31	28	27
#18	16	12	18	16	30	30	25

Figure S14 Comparison of amino acid sequence identities (%) between seven cytochrome P450-dependent enzymes (CYPs) from *M. xanthus* DK1622 and 18 CYPs from *S. cellulosum* So ce427.

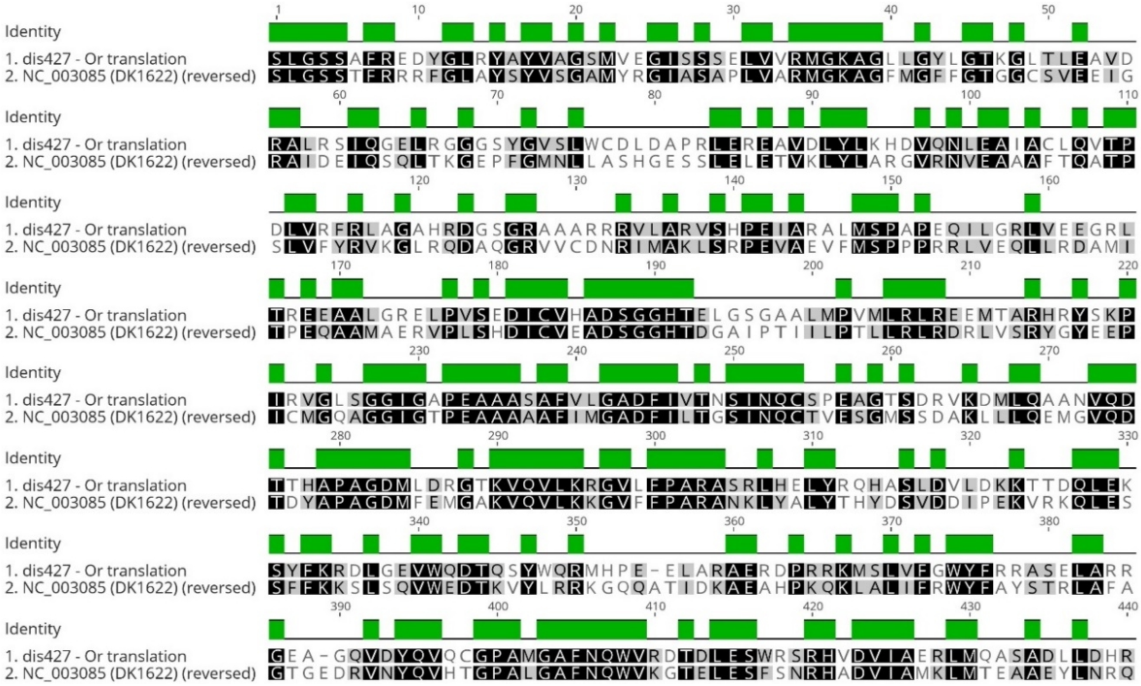


Figure S15 BLAST search of the disorazole Z Or domain against the *M. xanthus* DK1622 genome. The myxovirescin oxygenase TaN (MXAN_3934) from *M. xanthus* DK1622 was found to share a 47.5% identity and a 68.6 similarity with the disorazole Or.

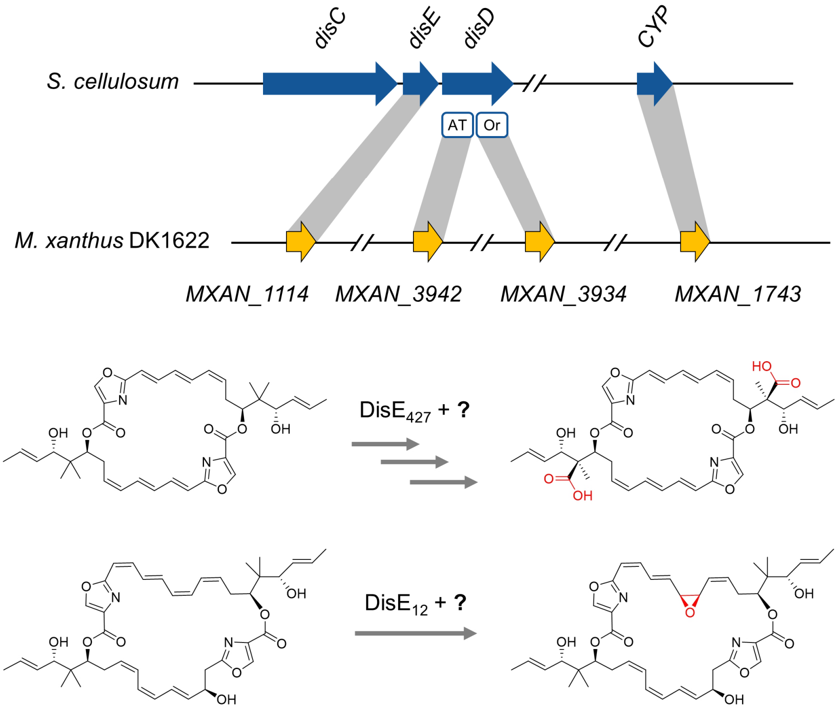


Figure S16 Homologs of the disorazole biosynthetic genes and their potential involvements in the oxidation processes. The symbol “**?**” represents cytochrome P450-dependent enzyme(s), oxidoreductase (Or), or their homologs from *M. xanthus* DK1622.

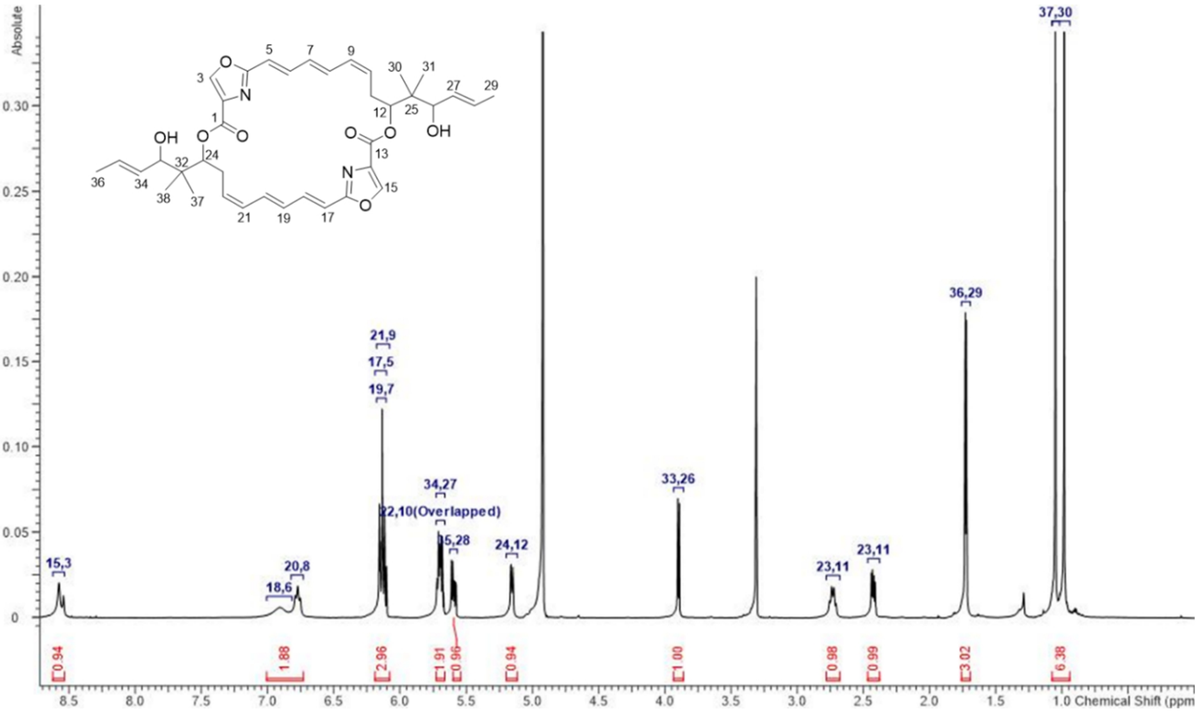


Figure S17 ^{1}H NMR spectrum of disorazole Z11 (**4**) in methanol-d_4 (700 MHz).

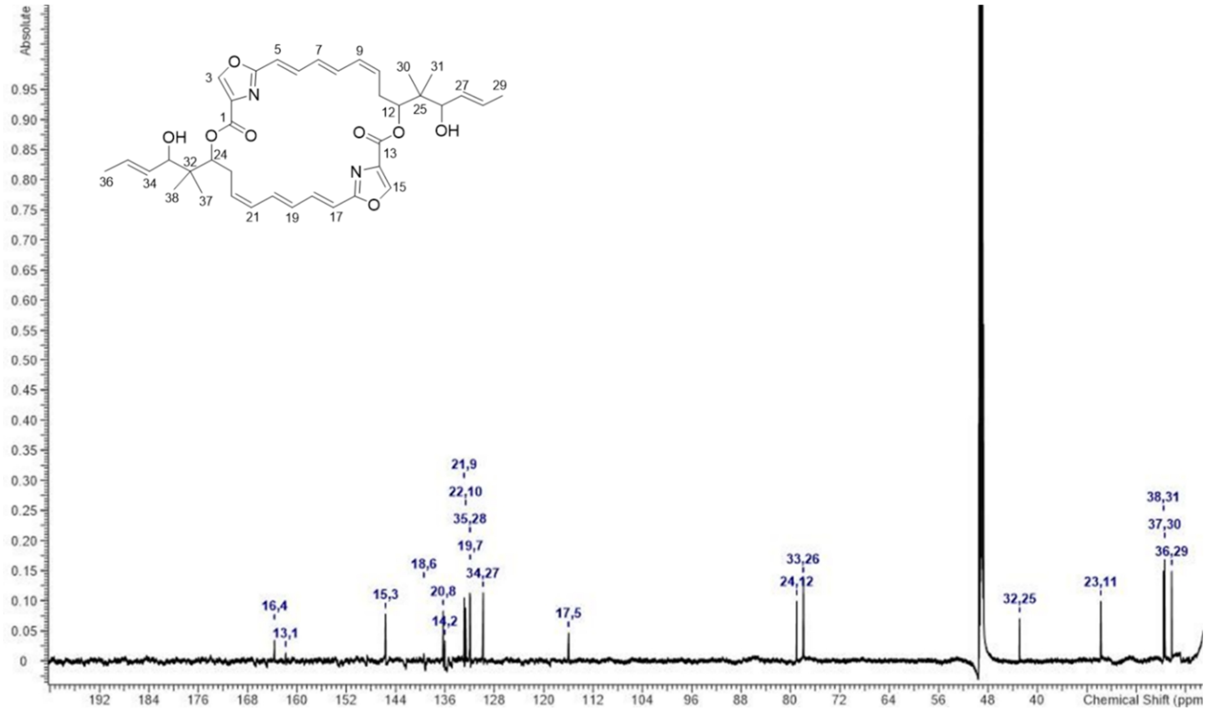


Figure S18 ^{13}C NMR spectrum of disorazole Z11 (**4**) in methanol-d_4 (175 MHz).

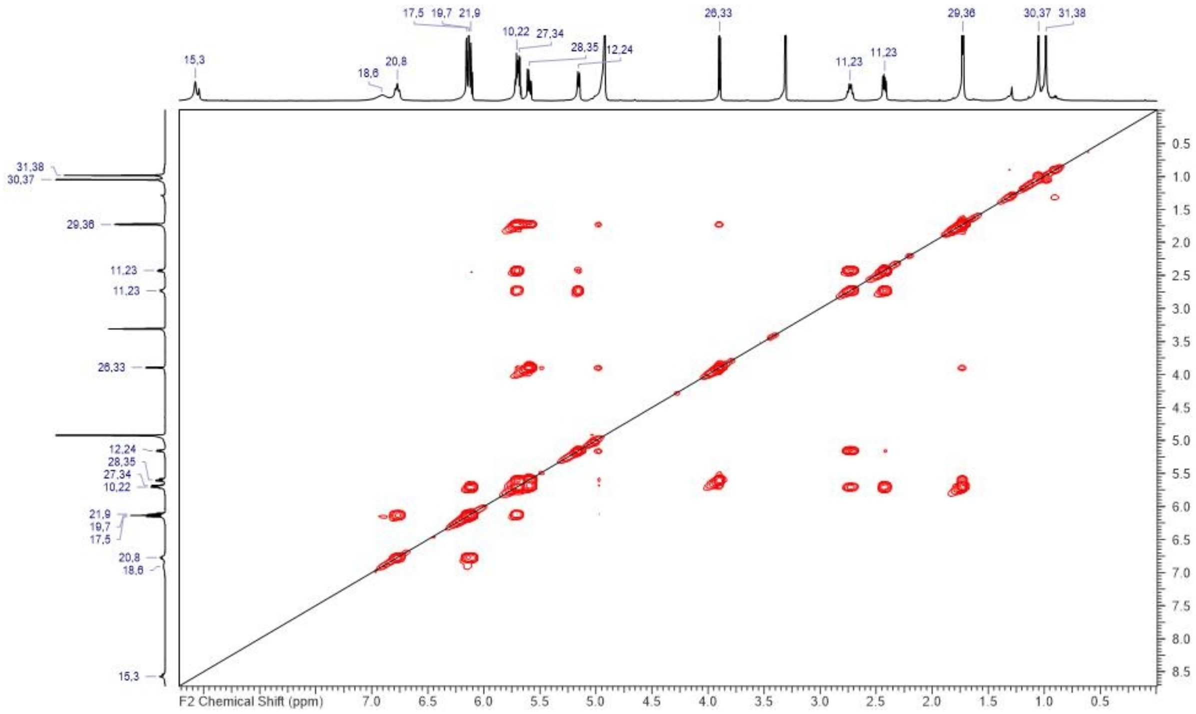


Figure S19 ^{1}H,^{1}H-COSY NMR spectrum of disorazole Z11 (**4**) in methanol-d_4.

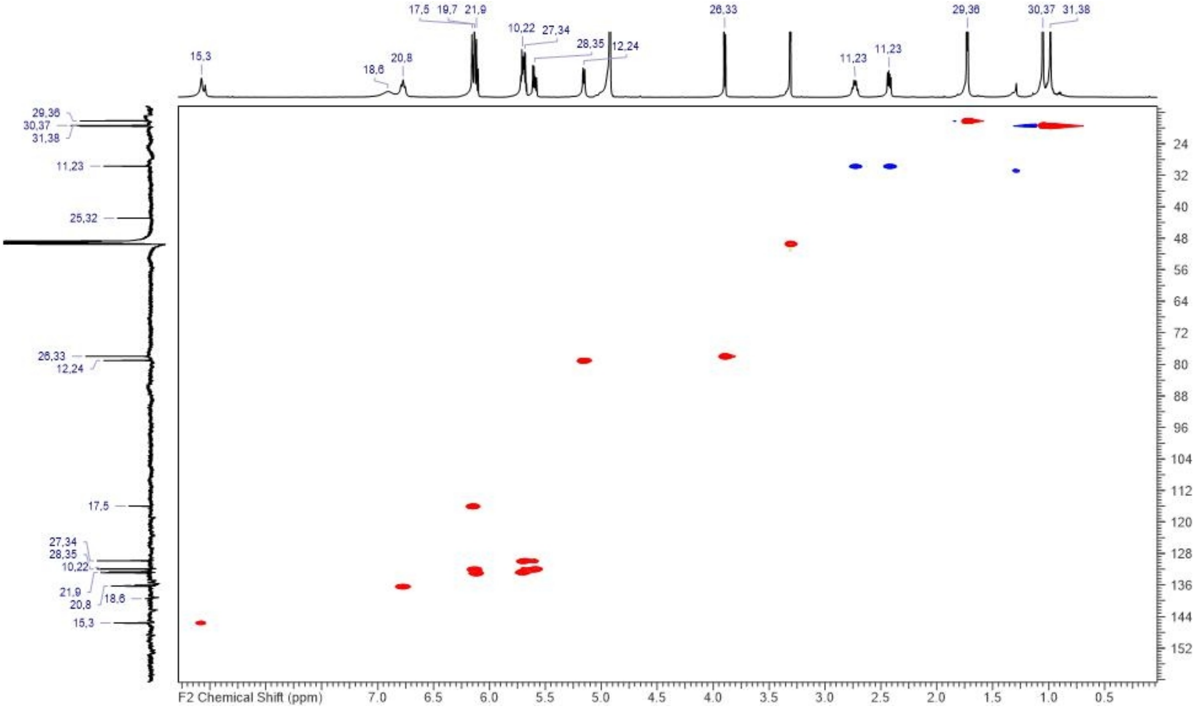


Figure S20 HSQC NMR spectrum of disorazole Z11 (**4**) in methanol-d_4.

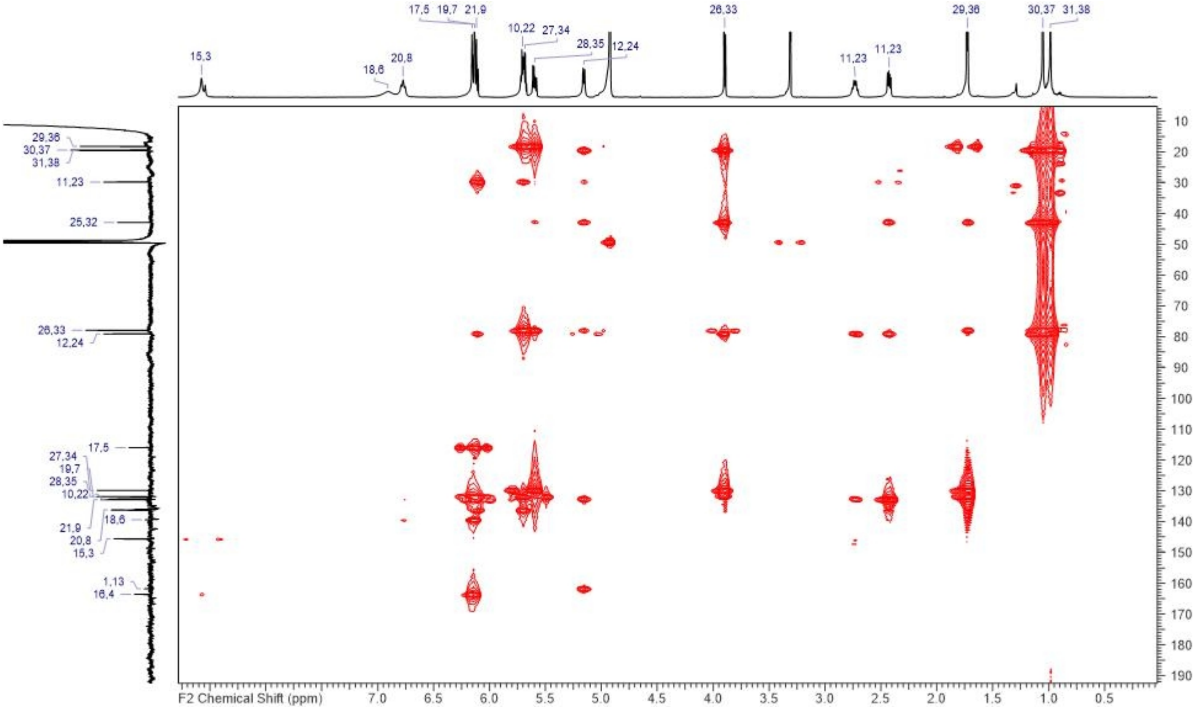


Figure S21 HMBC NMR spectrum of disorazole Z11 (**4**) in methanol-d_4.

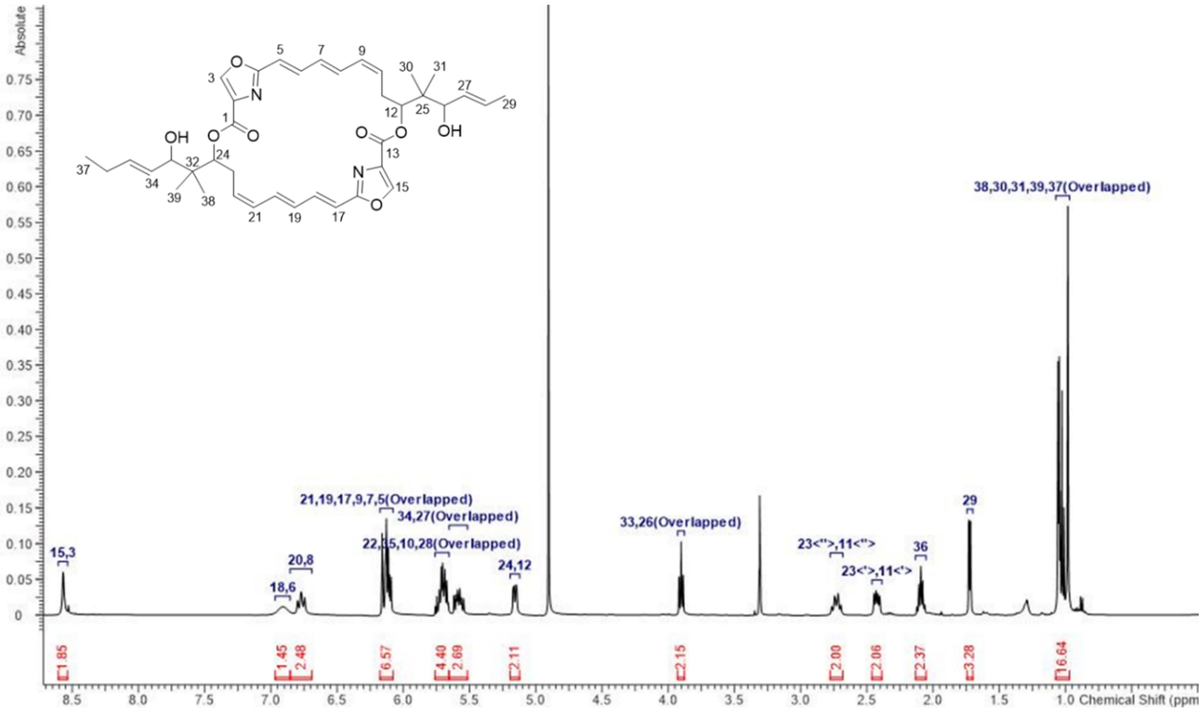


Figure S22 ^{1}H NMR spectrum of disorazole Z12 (**5**) in methanol-d_4 (500 MHz).

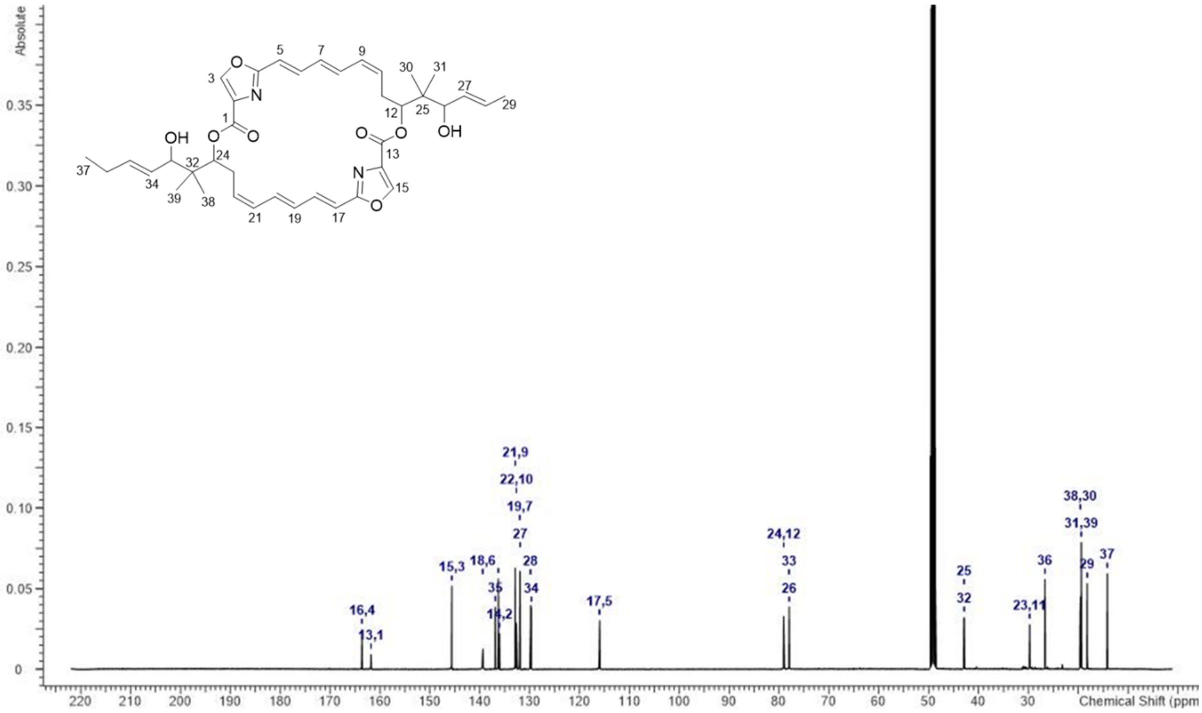


Figure S23 ^{13}C NMR spectrum of disorazole Z12 (**5**) in methanol-d_4 (125 MHz).

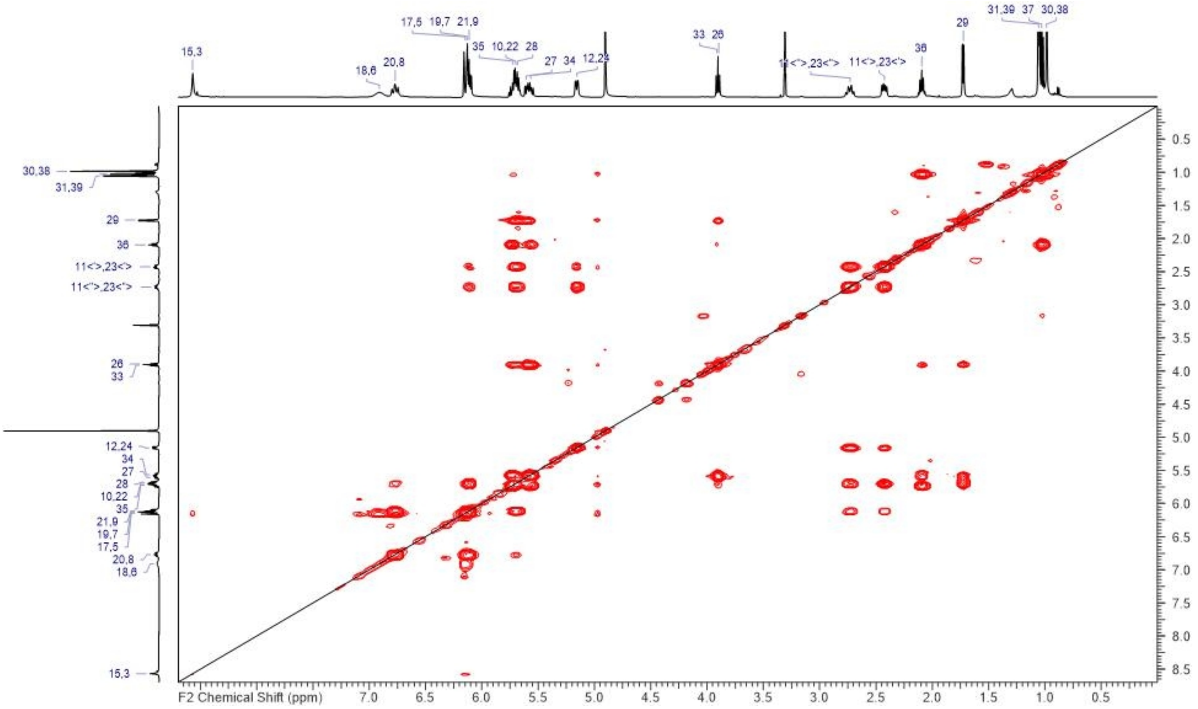


Figure S24 ^{1}H,^{1}H-COSY NMR spectrum of disorazole Z12 (**5**) in methanol-*d*$_4$.

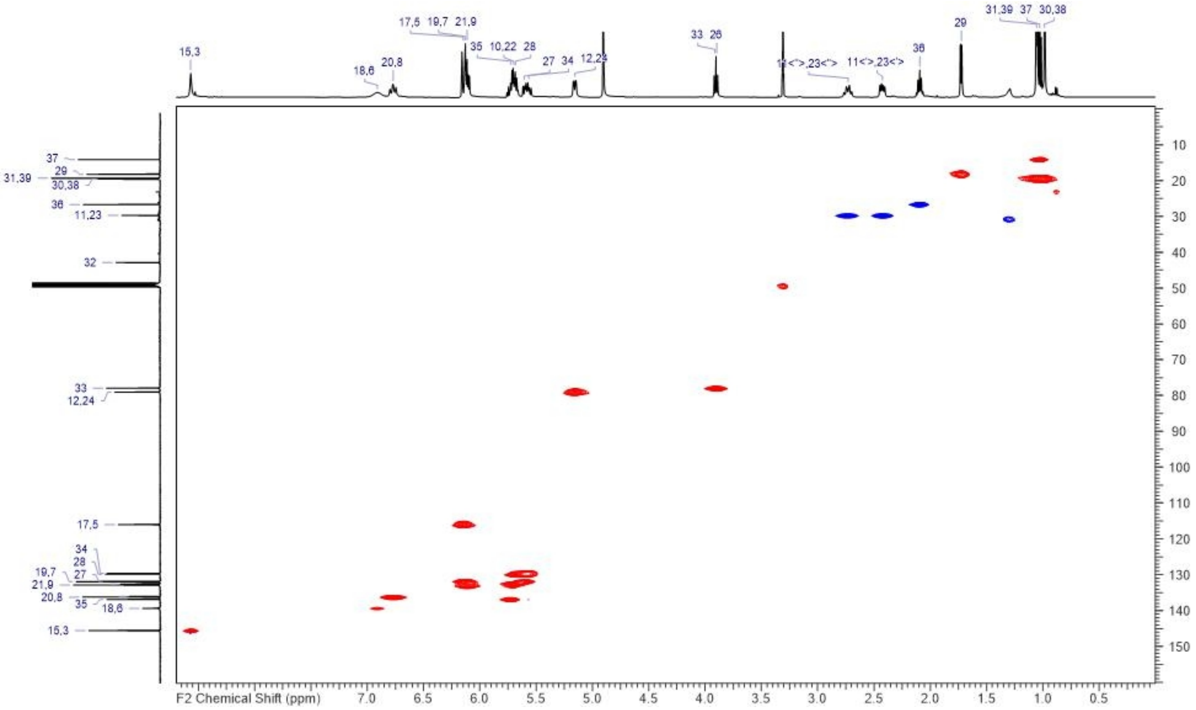


Figure S25 HSQC NMR spectrum of disorazole Z12 (**5**) in methanol-*d*$_4$.

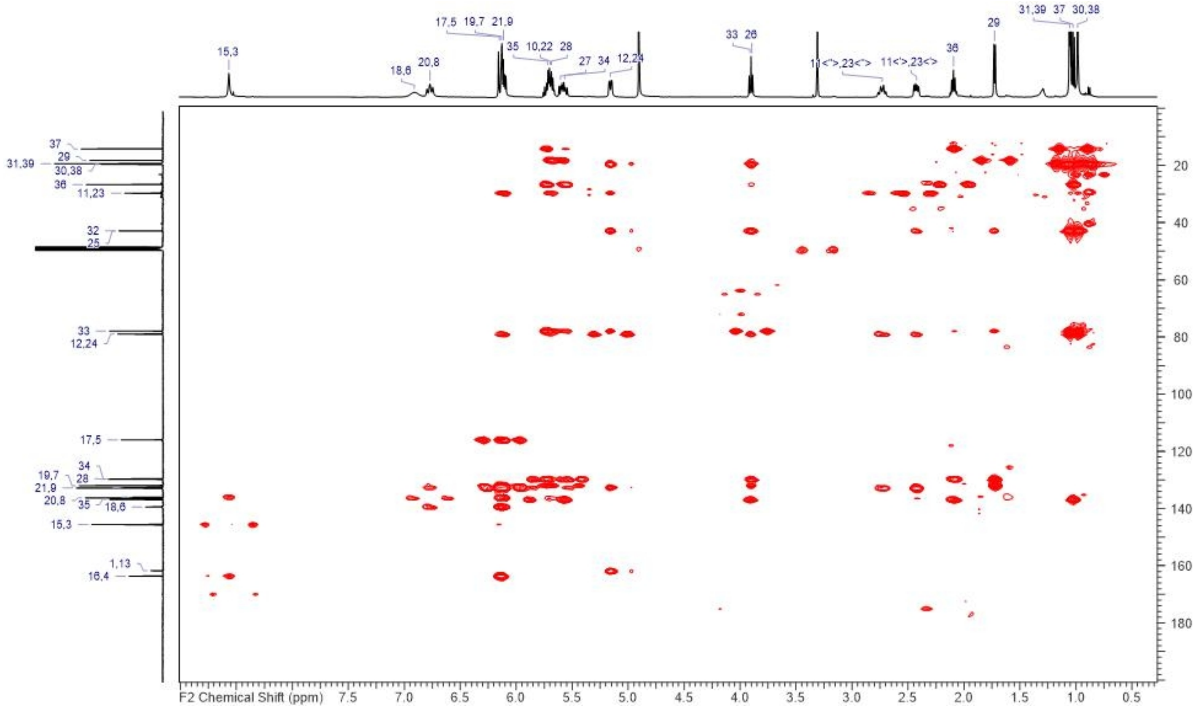


Figure S26 HMBC NMR spectrum of disorazole Z12 (**5**) in methanol-d_4.

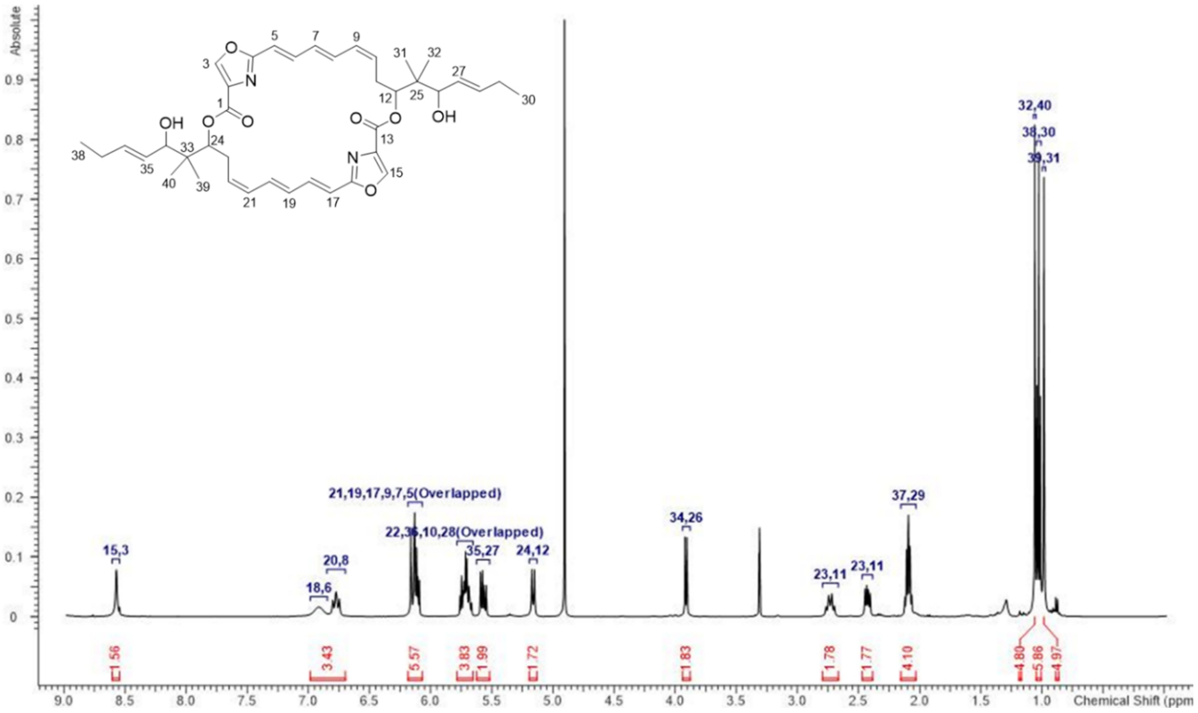


Figure S27 ^{1}H NMR spectrum of disorazole Z13 (**6**) in methanol-d_4 (500 MHz).

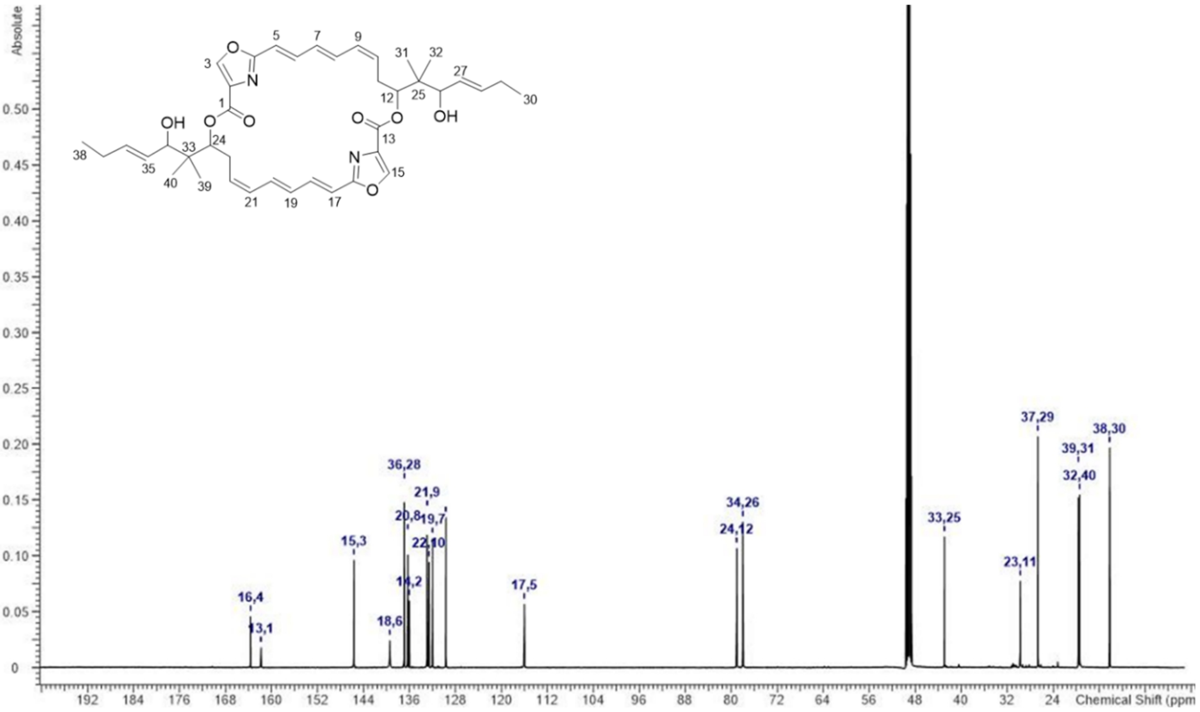


Figure S28 ^{13}C NMR spectrum of disorazole Z13 (**6**) in methanol-d_4 (125 MHz).

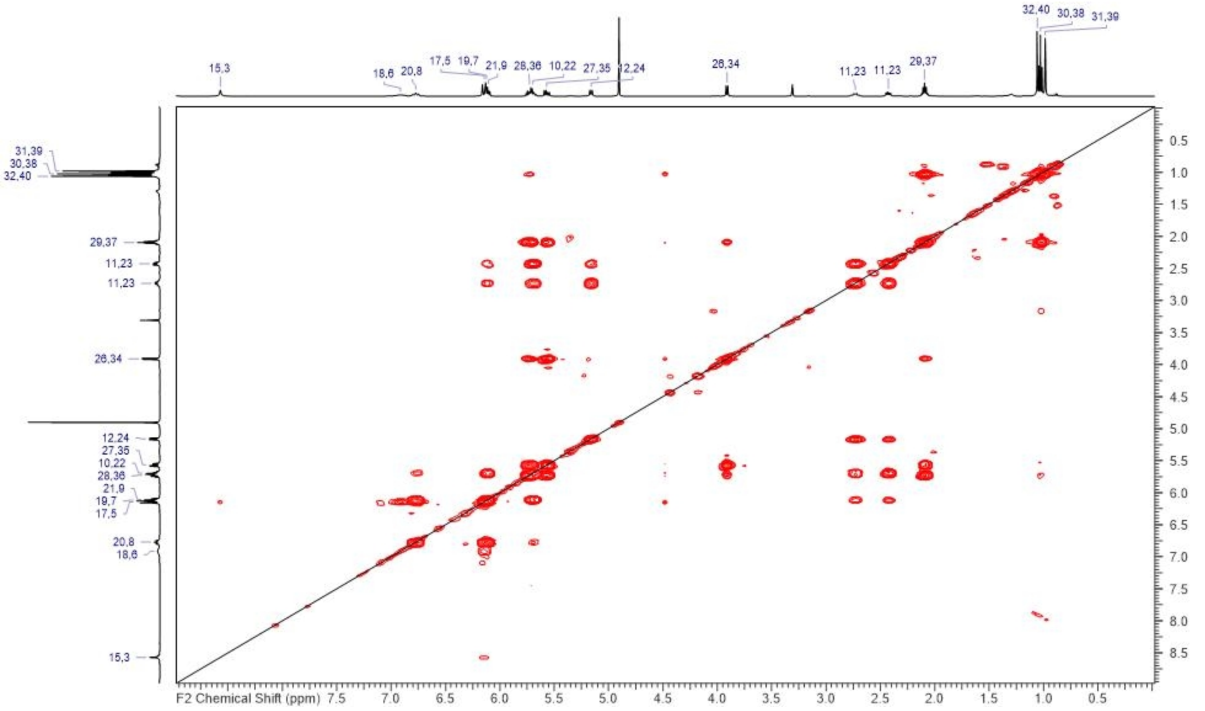


Figure S29 ^{1}H,^{1}H-COSY NMR spectrum of disorazole Z13 (**6**) in methanol-d_4.

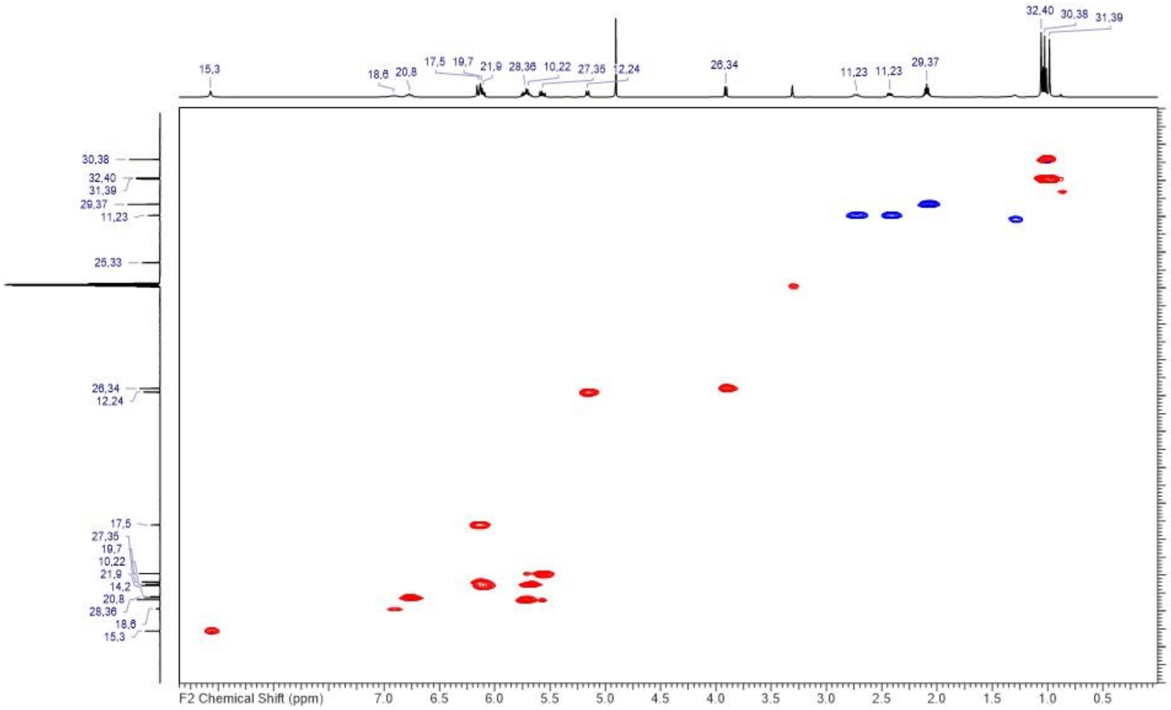


Figure S30 HSQC NMR spectrum of disorazole Z13 (**6**) in methanol-d_4.

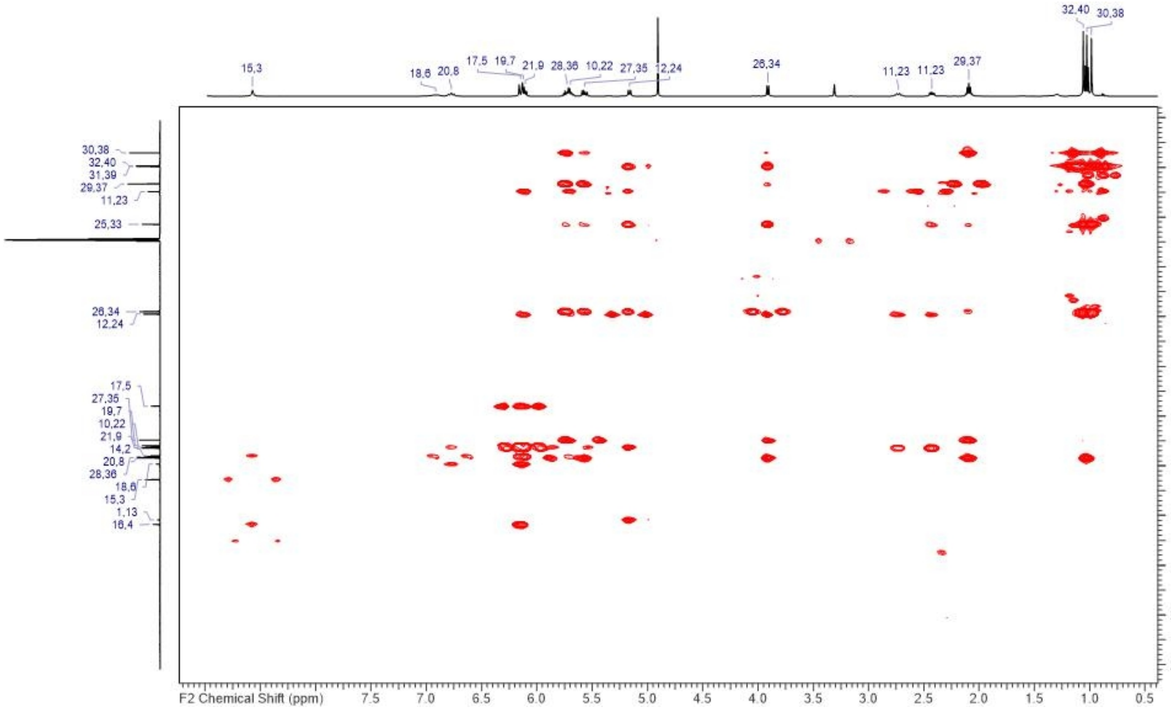


Figure S31 HMBC NMR spectrum of disorazole Z13 (**6**) in methanol-d_4.

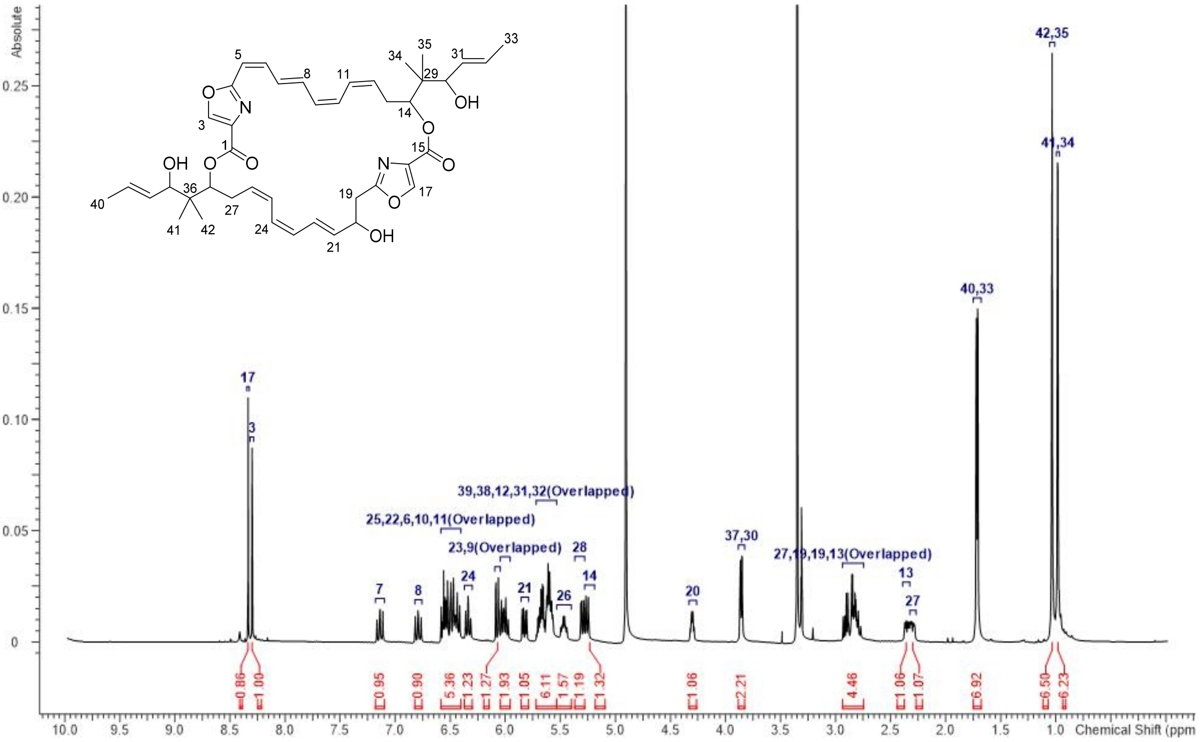


Figure S32 ^{1}H NMR spectrum of disorazole F2 (**3**) in methanol-d_4 (500 MHz).

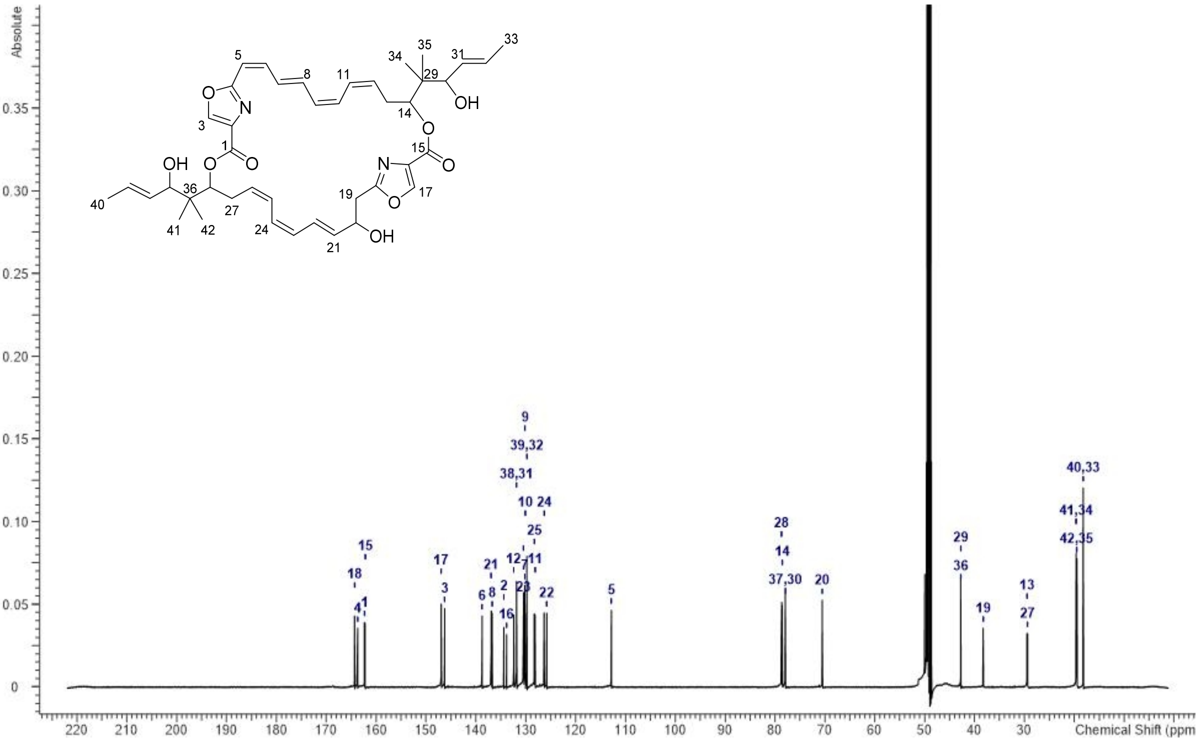


Figure S33 ^{13}C NMR spectrum of disorazole F2 (**3**) in methanol-d_4 (125 MHz).

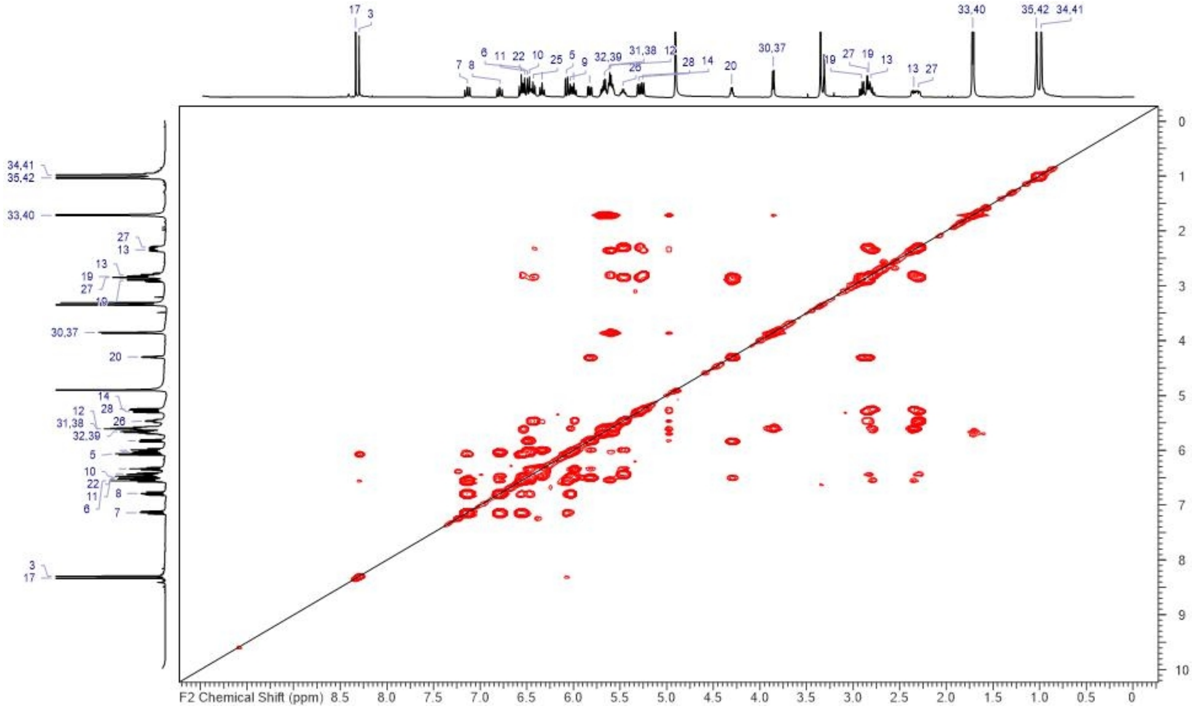


Figure S34 ^{1}H,^{1}H-COSY NMR spectrum of disorazole F2 (**3**) in methanol-*d*$_4$.

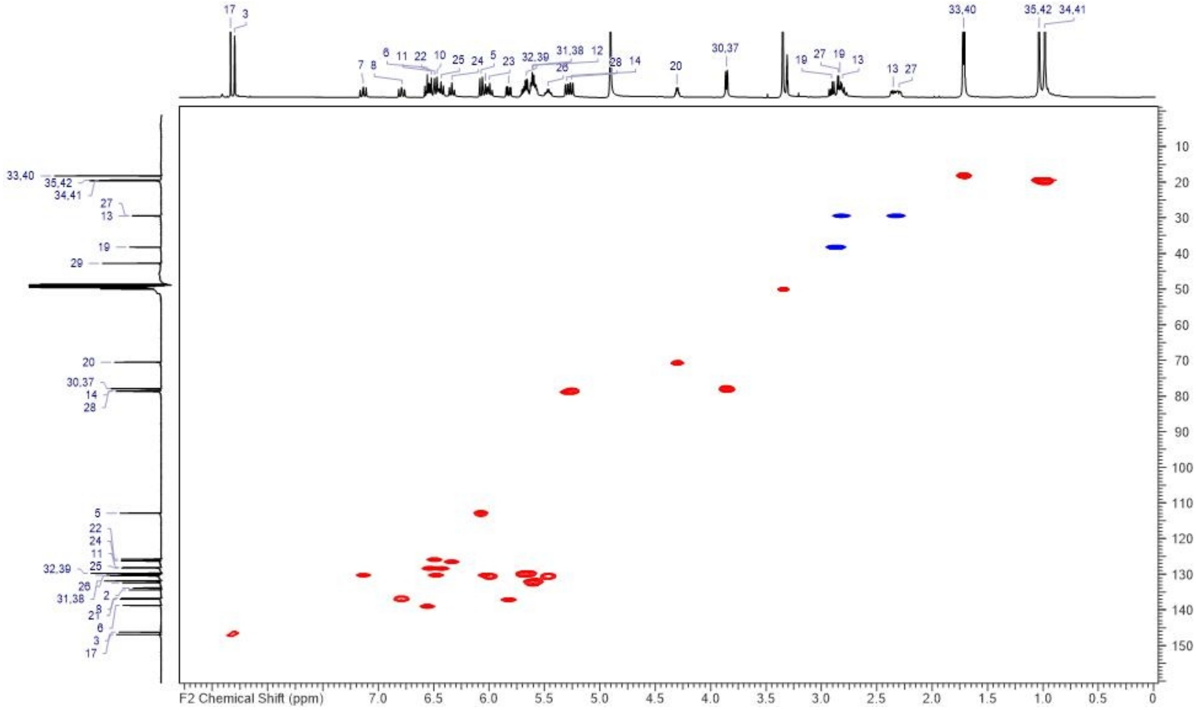


Figure S35 HSQC NMR spectrum of disorazole F2 (**3**) in methanol-*d*$_4$.

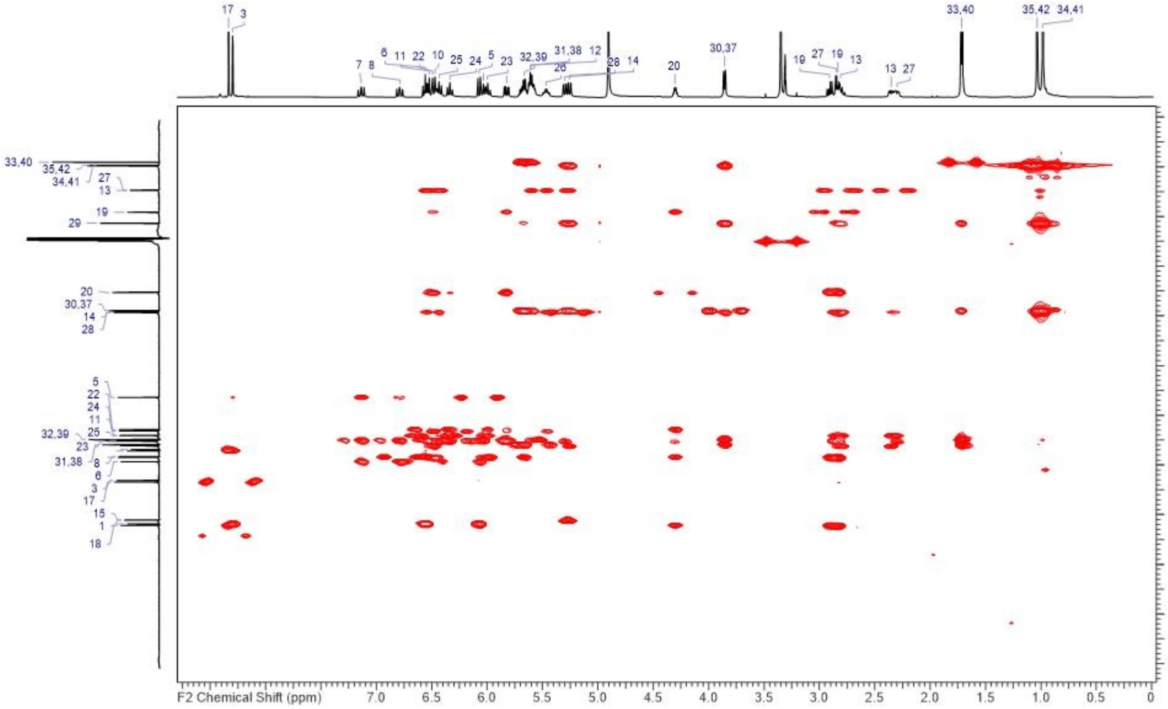


Figure S36 HMBC NMR spectrum of disorazole F2 (**3**) in methanol-d_4.

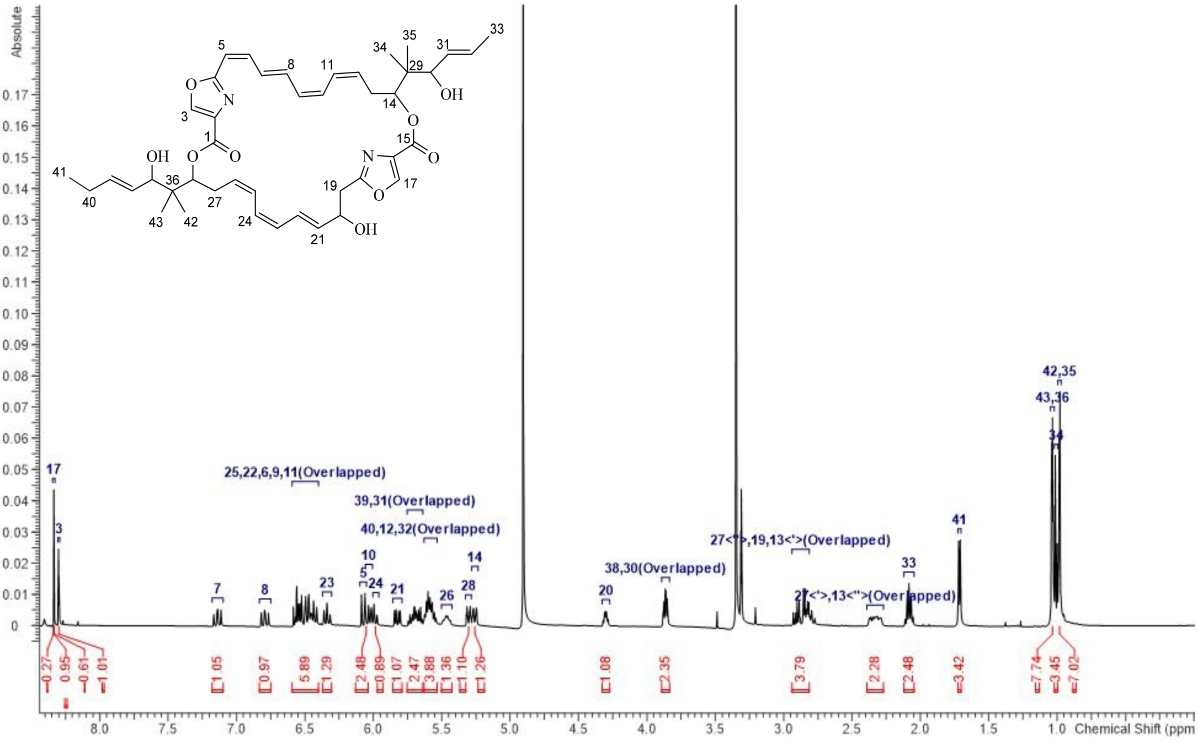


Figure S37 ^{1}H NMR spectrum of disorazole F7 (**7**) in methanol-d_4 (500 MHz).

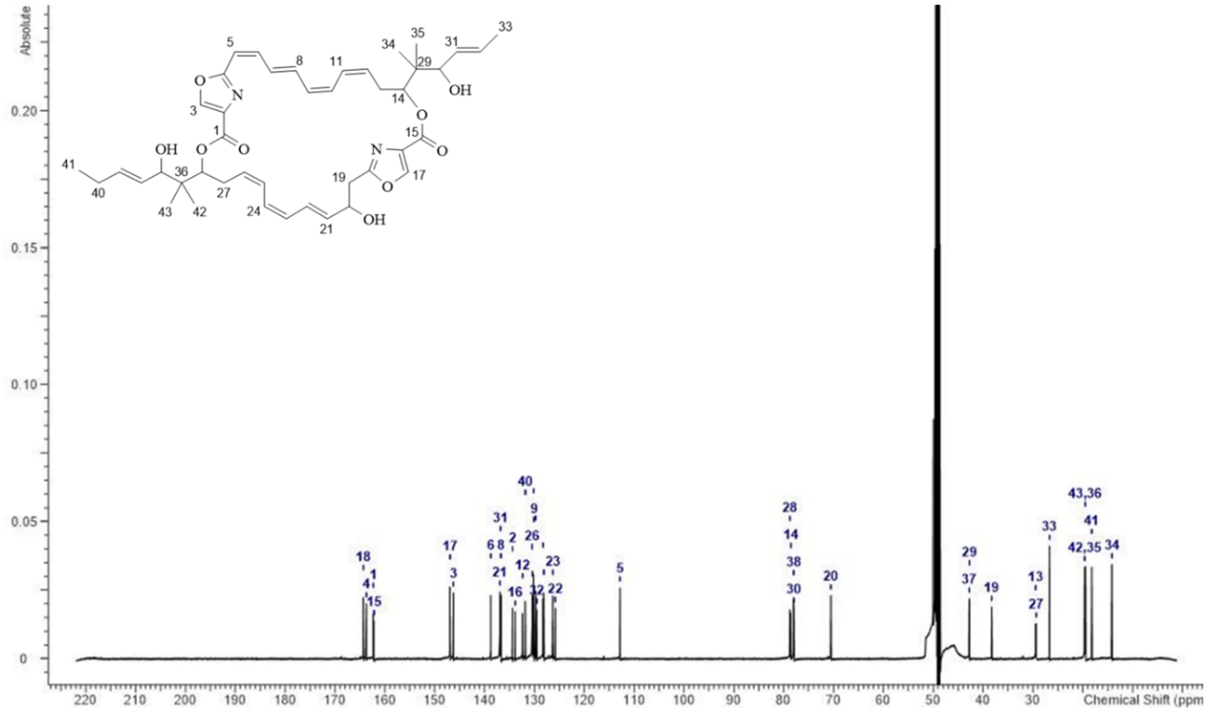


Figure S38 ^{13}C NMR spectrum of disorazole F7 (**7**) in methanol-d_4 (125 MHz).

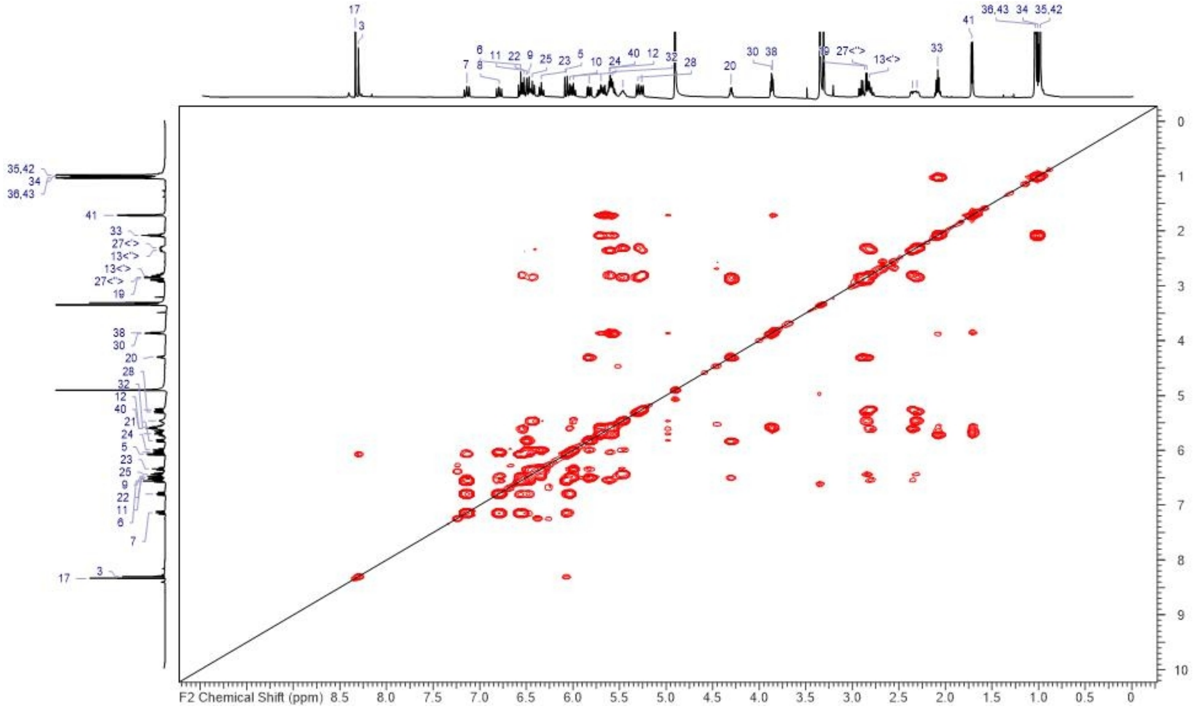


Figure S39 ^{1}H,^{1}H-COSY NMR spectrum of disorazole F7 (**7**) in methanol-d_4.

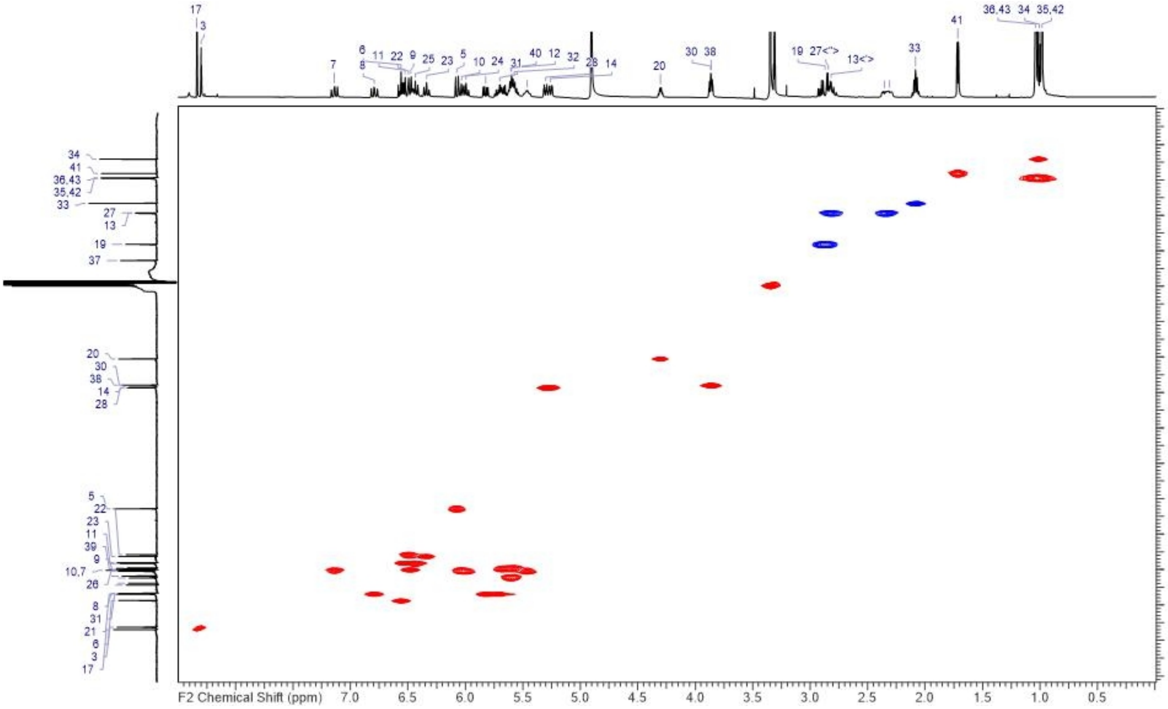


Figure S40 HSQC NMR spectrum of disorazole F7 (**7**) in methanol-*d*4.

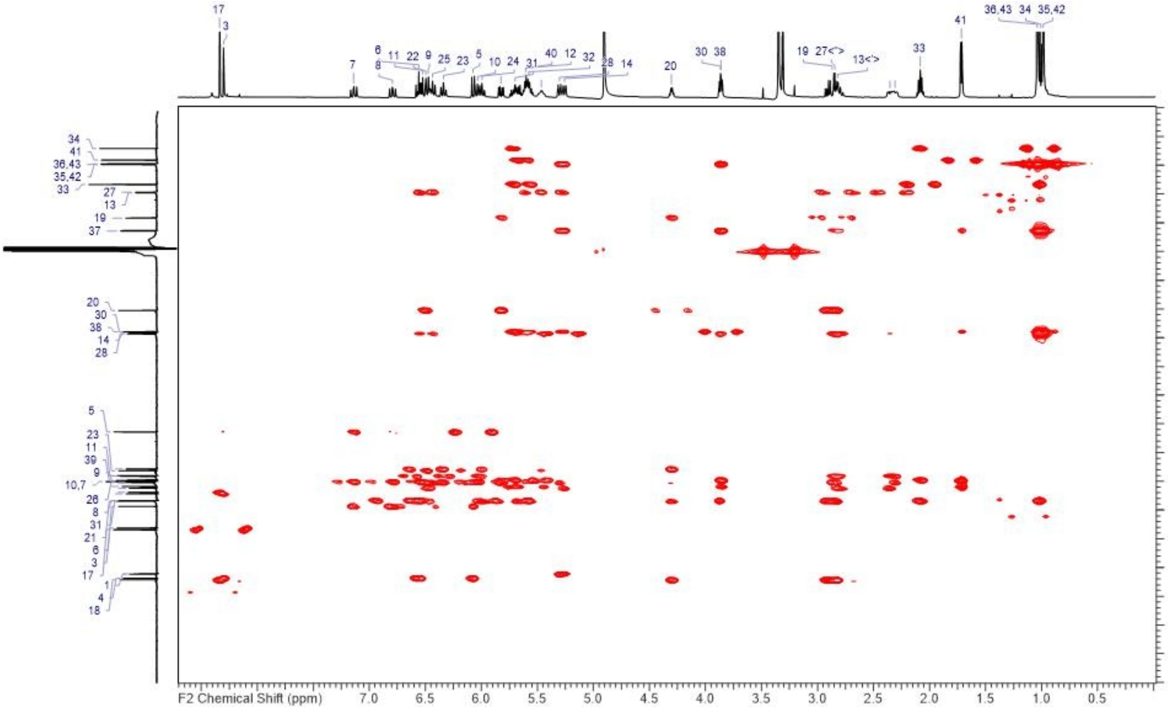


Figure S41 HMBC NMR spectrum of disorazole F7 (**7**) in methanol-*d*4.

3.7 Acknowledgements

The authors would like to express gratitude to Dr. Chengzhang Fu for engaging in helpful discussions, and to Dr. Amninder Kaur for double checking of the NMR data.

3.8 References

[1] C. D. Hopkins and P. Wipf, Isolation, biology and chemistry of the disorazoles: new anti-cancer macrodiolides, *Nat Prod Rep*, **2009**, 26(5), 585-601.

[2] R. Jansen, H. Irschik, H. Reichenbach et al., Disorazoles, Highly Cytotoxic Metabolites from the Sorangicin-Producing Bacterium *Sorangium cellulosum*, Strain So ce12, *Liebigs Ann Chem*, **1994**, 1994(8), 759-773.

[3] Y. Gao, J. Birkelbach, C. Fu et al., The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis and Heterologous Expression, *Microbiol Spectrum*, **2023**, acceped manuscript.

[4] J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2010**, 27(7), 996-1047.

[5] E. J. N. Helfrich and J. Piel, Biosynthesis of polyketides by *trans*-AT polyketide synthases, *Nat Prod Rep*, **2016**, 33(2), 231-316.

[6] Q. Tu, J. Herrmann, S. Hu et al., Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering, *Sci Rep*, **2016**, 6, 21066.

[7] Z.-J. Wang, X. Liu, H. Zhou et al., Engineering of *Burkholderia thailandensis* strain E264 serves as a chassis for expression of complex specialized metabolites, *Front Microbiol*, **2022**, 13, 1073243.

[8] Z.-J. Wang, X. Liu, H. Zhou et al., Engineered Biosynthesis of Complex Disorazol Polyketides in a Streamlined *Burkholderia thailandensis*, *ACS Synth Biol*, **2023**, 12(4), 971-977.

[9] G. I. Belogrudov, P. T. Lee, T. Jonassen et al., Yeast COQ4 encodes a mitochondrial protein required for coenzyme Q synthesis, *Arch Biochem Biophys*, **2001**, 392(1), 48-58.

[10] A. Casarin, J. C. Jimenez-Ortega, E. Trevisson et al., Functional characterization of human COQ4, a gene required for Coenzyme Q10 biosynthesis, *Biochem Biophys Res Commun*, **2008**, 372(1), 35-39.

[11] B. Marbois, P. Gin, M. Gulmezian et al., The yeast Coq4 polypeptide organizes a mitochondrial protein complex essential for coenzyme Q biosynthesis, *Biochim Biophys Acta*, **2009**, 1791(1), 69-75.

[12] R. Schäckel, B. Hinkelmann, F. Sasse et al., The synthesis of novel disorazoles, *Angew Chem Int Ed*, **2010**, 49(9), 1619-1622.

[13] M. Kopp, H. Irschik, S. Pradella et al., Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes, *ChemBioChem*, **2005**, 6(7), 1277-1286.

[14] R. Carvalho, R. Reid, N. Viswanathan et al., The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation, *Gene*, **2005**, 359, 91-98.

[15] V. Simunovic, J. Zapp, S. Rachid et al., Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases, *ChemBioChem*, **2006**, 7(8), 1206-1220.

[16] J. Mistry, S. Chuguransky, L. Williams et al., Pfam: The protein families database in 2021, *Nucleic Acids Res*, **2021**, 49(D1), D412-D419.

[17] S. L. Rea, B. H. Graham, E. Nakamaru-Ogiso et al., Bacteria, yeast, worms, and flies: exploiting simple model organisms to investigate human mitochondrial diseases, *Dev Disabil Res Rev*, **2010**, 16(2), 200-218.

[18] L. C. Blasiak and J. Clardy, Discovery of 3-formyl-tyrosine metabolites from *Pseudoalteromonas tunicata* through heterologous expression, *J Am Chem Soc*, **2010**, 132(3), 926-927.

[19] L. M. Podust and D. H. Sherman, Diversity of P450 enzymes in the biosynthesis of natural products, *Nat Prod Rep*, **2012**, 29(10), 1251-1266.

[20] A. Greule, J. E. Stok, J. J. de Voss et al., Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism, *Nat Prod Rep*, **2018**, 35(8), 757-791.

[21] Y. Khatri, F. Hannemann, O. Perlova et al., Investigation of cytochromes P450 in myxobacteria: excavation of cytochromes P450 from the genome of *Sorangium cellulosum* So ce56, *FEBS Lett*, **2011**, 585(11), 1506-1513.

[22] K. Viehrig, F. Surup, K. Harmrolfs et al., Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin, *J Am Chem Soc*, **2013**, 135(45), 16885-16894.

[23] H. Wang, Z. Li, R. Jia et al., RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression, *Nat Protoc*, **2016**, 11(7), 1175-1190.

[24] F. Yan, D. Auerbach, Y. Chai et al., Biosynthesis and heterologous production of vioprolides: rational biosynthetic engineering and unprecedented 4-methylazetidinecarboxylic acid formation, *Angew Chem Int Ed*, **2018**, 57(28), 8754-8759.

[25] D. Krug and R. Müller, Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products, *Nat Prod Rep*, **2014**, 31(6), 768-783.

[26] M. Mirdita, K. Schütze, Y. Moriwaki et al., ColabFold: making protein folding accessible to all, *Nat Methods*, **2022**, 19(6), 679-682.

[27] K. Katoh and D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol Biol Evol*, **2013**, 30(4), 772-780.

Chapter 4. Genome-guided discovery of the myxobacterial thiolactone-containing sorangibactins

ACS Chemical Biology, published as a Research Article, 2023
DOI: 10.1021/acschembio.3c00063

Yunsheng Gao+,[a,b,c] Christine Walt+,[a,c] Chantal D. Bader,[a,c] and Rolf Müller*[a,b,c]

[a] Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Department of Pharmacy at Saarland University, Campus E8.1, 66123 Saarbrücken, Germany

[b] Helmholtz International Lab for Anti-Infectives, Campus E8.1, 66123 Saarbrücken, Germany

[c] German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany

[+] These authors contributed equally to this work

[*] To whom correspondence should be addressed: rolf.mueller@helmholtz-hips.de

Keywords

myxobacteria, genome mining, heterologous expression, siderophore, homocysteine thiolactone

Author's efforts:

The author designed the research, performed *in silico* analysis and all molecular biology experiments, including promoter engineering, gene knockout, point mutation, and heterologous expression. The author performed feeding experiments, metabolome analysis, and large-scale fermentation. The author conducted compound purification together with Christine Walt. The author wrote the manuscript with inputs from other authors.

Other's efforts:

Christine Walt contributed to compounds purification. She made significant efforts to NMR measurements, *de novo* structure elucidation, and preparing the manuscript. She also contributed to analyzing the MS data and cultivating the original producer strain. Chantal D. Bader contributed to structure elucidation and proofreading of the manuscript. Rolf Müller supervised the research and revised the manuscript.

Signatures:

Yunsheng Gao: Yunsheng Gao
Christine Walt:
Chantal D. Bader:
Rolf Müller:

4.1 Abstract

In this study, an unprecedented myxobacterial siderophore termed sorangibactin was discovered by heterologous expression of a coelibactin-like nonribosomal peptide synthetase (NRPS) gene cluster from the *Sorangiineae* strain MSr11367 in the host *Myxococcus xanthus* DK1622. *De novo* structure elucidation uncovered a linear polycyclic structure consisting of an N-terminal phenol group, an oxazole, tandem *N*-methyl-thiazolidines and an unusual C-terminal γ-thiolactone moiety. Except for the unprecedented oxazoline dehydrogenation to form an oxazole, which we show to be catalyzed by a cytochrome P450-dependent enzyme, other tailoring steps were found necessary for efficient downstream processing. The unusual thioesterase (TE) domain is proposed to select homocysteine or methionine for offloading involving an intramolecular γ-thiolactone formation. Its active site comprises a rare cysteine, which was found essential for product formation by point mutation to alanine or serine, which both abolished its activity. This unusual release mechanism and the resulting rare thiolactone structure can serve as a starting point for detailed biochemical investigations.

4.2 Introduction

Myxobacteria are Gram-negative δ-proteobacteria featuring unique morphological characteristics, complex life cycles, large genomes, and a particularly huge biosynthetic potential for secondary metabolite production.[1] Hundreds of structurally interesting scaffolds and bioactive compounds from myxobacteria were discovered over the past decades.[2] Despite this rich collection of characterized secondary metabolites, genome sequencing as well as biosynthetic gene cluster (BGC) annotation have revealed that the majority of the myxobacterial biosynthetic potential is still untapped.[3-5] Thus, a “gene to compound” strategy can shed light on assigning secondary metabolites to their BGCs and finding unprecedented structures. As myxobacteria, especially the *Sorangiineae* suborder, are usually slow growing and barely tractable for genetic manipulation, heterologous expression is particularly promising to explore them, as it facilitates not only genome mining for natural product discovery but also biosynthetic pathway engineering and detailed biosynthesis studies.[6] One prolific starting point for the examination of cryptic BGCs are so-called broad cosmid libraries,[7] which can comprise the entire genome of a selected bacterial strain in the form of plasmid

DNA and facilitate high-throughput heterologous expression.[8] This approach also enables the unambiguous correlation of BGCs with their products.

One interesting structural class of natural products are bacterial siderophores, which are small-molecular-weight Fe(III)-specific ligands that scavenge iron from their environment and enable uptake into the cell through outer-membrane transporters.[9] Siderophores from myxobacteria belong to two main categories, namely, the catecholate-type such as the myxochelins[10] and hyalachelins[11] and the citrate-hydroxamate-type such as the nannochelins[12]. On the contrary, over 500 congeners of diverse siderophore families have been described from actinobacteria.[13] One particularly puzzling actinobacterial siderophore is coelibactin, the existence of which has only been proposed based on the presence of the respective BGC and for which only a hypothetical chemical structure is found in the literature. Comprehensive bioinformatics analysis of the respective BGC found in the model species *Streptomyces coelicolor* A3(2) enabled a prediction of the core structure based on the predicted sequence of the nonribosomal peptide synthetase (NRPS) gene cluster, yet isolation of coelibactin or derivatives thereof failed for unknown reasons.[14,15]

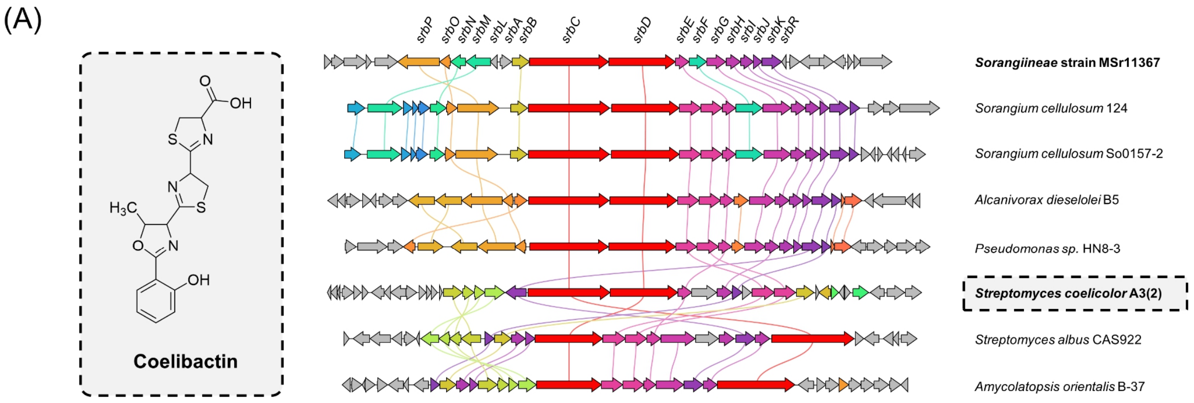
In this study, we report the discovery of a novel myxobacterial secondary metabolite with siderophore-like behavior termed sorangibactin from *Sorangiineae* strain MSr11367 by heterologous expression of a promoter-engineered NRPS gene cluster, which resembles the coelibactin BGC. This coelibactin-like NRPS gene cluster was prioritized due to the following reasons: (1) its bioinformatically predicted structure resembles coelibactin from *Streptomyces coelicolor* A3(2),[14] which raises hope to contribute to solving the puzzle of the coelibactins, (2) its regular operon organization, which is suitable for promoter engineering, and (3) multiple tailoring genes surrounding the core NRPS part, which indicates auxiliary modifications before or after chain release and may endow the final product with structural novelty and diversity.

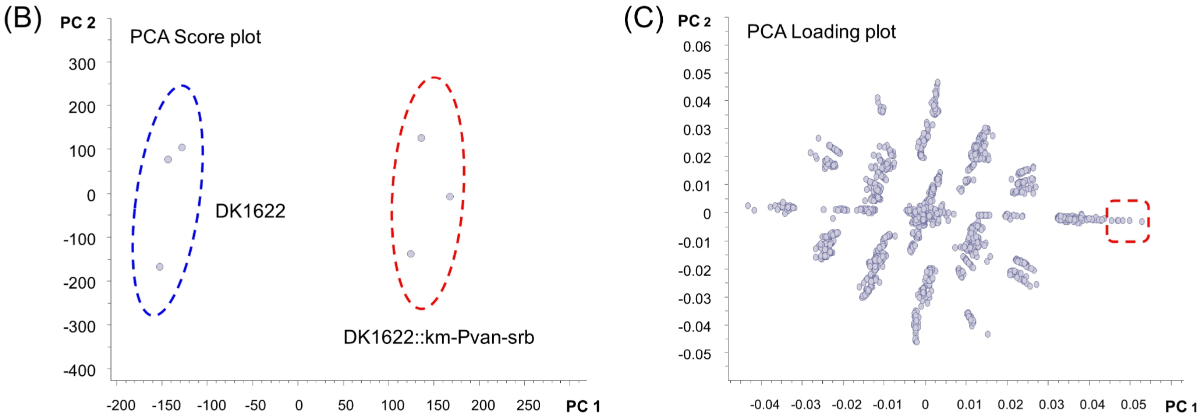
4.3 Results and discussion

4.3.1 Heterologous expression of a coelibactin-like NRPS gene cluster

According to the antiSMASH prediction,[16] the myxobacterial *Sorangiineae* strain MSr11367 exhibits a great biosynthetic potential of 62 biosynthetic gene clusters and thus was prioritized for natural product discovery. The strain was already explored using an OSMAC (one strain many compounds) metabolomics approach in our laboratory, but even altered cultivation conditions did not reflect its full biosynthetic potential making this strain particularly interesting for genome-based discovery. Despite several efforts, MSr11367 proved to be intractable for *in-situ* genetic manipulation. Therefore, a genomic DNA library in the form of cosmids was generated for systematic screening of biosynthetic gene clusters by heterologous expression in the genetically amenable host *Myxococcus xanthus* DK1622.[17] Among the 62 BGCs encoded in the strain's genome, we prioritized a coelibactin-like NRPS gene cluster, that showed widespread occurrence of homologous BGCs in *Sorangium cellulosum*, *Streptomyces*, and *Pseudomonas* strains (Figure 1) presumably indicating a conserved function of the correlating secondary metabolites during evolution. The MSr11367 coelibactin-like gene cluster (sorangibactin BGC) has been deposited in the GenBank under the accession number OQ368584. The respective annotation of the gene cluster is provided in supplementary Table S1.

The unmodified gene cluster was cloned from the cosmid library of the *Sorangiineae* strain MSr11367 and transferred to the host *M. xanthus* DK1622 for expression; however, no heterologous product was detected. Thus, a *km-vanR-Pvan* cassette was inserted between the two main operons (Figure S1), after which the first operon consisting of transporter genes shares a strong constitutive promoter with the kanamycin resistance gene and the second operon consisting of biosynthesis genes is under control of a vanillate-inducible Pvan promoter. Principal component analysis (PCA) of high-performance liquid chromatography-high-resolution mass spectrometry (HPLC-HRMS) datasets revealed several molecular features exclusively present in *M. xanthus* DK1622 mutants containing the promoter-engineered gene cluster and absent in the *M. xanthus* DK1622 wild-type host (Figure 1B−D). The major component (termed sorangibactin A) could also be found in *Sorangiineae* strain MSr11367 extracts in equal concentrations when cultivated under iron-deficient conditions, indicating it as the final product of the respective BGC (Figure 1E). Isolation for structure elucidation was performed utilizing the heterologous host, as besides its less complex metabolome that was beneficial for the purification process, it also shows faster growth rates reducing possible contamination issues during cultivation as observed for the strain MSr11367.





(D)

Name	RT [min]	*m/z* for $[M+H]^+$	Ion formula
Sorangibactins A1 and A2	8.3, 9.1	507.1190	$C_{22}H_{27}N_4O_4S_3$
Sorangibactin B	9.0, 10.2, 10.8	525.1300	$C_{22}H_{29}N_4O_5S_3$
Sorangibactin C	7.4	509.1332	$C_{22}H_{29}N_4O_4S_3$
Sorangibactin D	8.8	376.0422	$C_{16}H_{14}N_3O_4S_2$
Putative sorangibactin associated features	8.2, 8.4, 8.6	378.0580	$C_{16}H_{16}N_3O_4S_2$
	7.2	410.1213	$C_{18}H_{24}N_3O_4S_2$
	7.7	291.0439	$C_{13}H_{11}N_2O_4S$

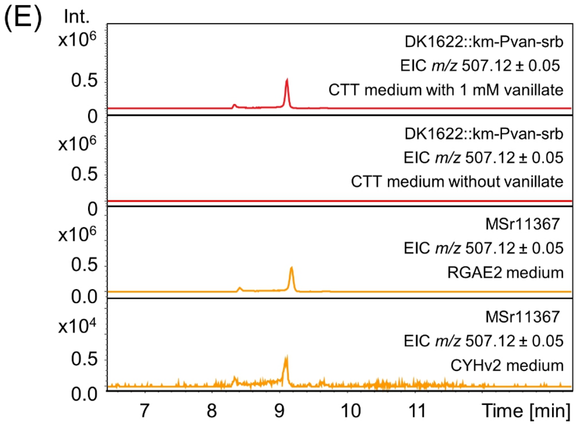
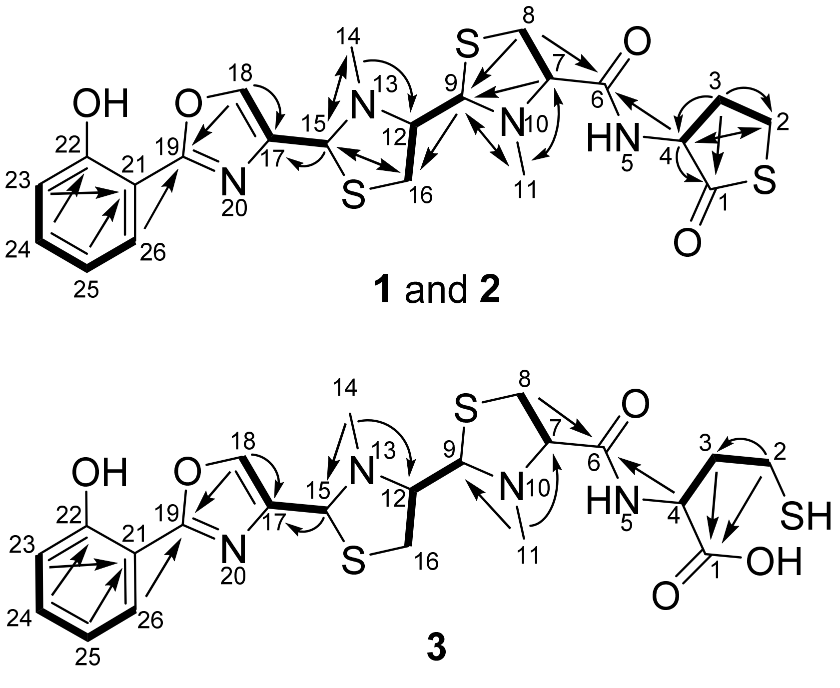


Figure 1 Characterization of the coelibactin-like NRPS gene cluster from the *Sorangiineae* strain MSr11367. (A) Proposed structure of coelibactin from *S. coelicolor* A3(2) and clinker analysis of selected coelibactin BGC homologues from different bacterial species. The BGC from MSr11367 with gene names is shown at the top and has been characterized in this study. Genes with a >30% sequence identity are assigned a unique color and linked. (B) Significant difference between the heterologous expression data set (DK1622::km-Pvan-srb) encircled in red and the wild-type host data set (DK1622) encircled in blue shown by a PCA score plot. (C) Molecular features that contribute to this difference shown by a PCA loading plot. (D) Molecular features that are only found in the heterologous expression data set. (E) HPLC-MS chromatogram of sorangibactin A production by DK1622::km-Pvan-srb (red) and MSr11367 (orange).

4.3.2 Characterization of sorangibactins

HPLC-HRMS measurements exhibit an $[M+H]^+$ peak for the mixture of sorangibactins A1 and A2 at *m/z* 507.1190 matching the molecular formula $C_{22}H_{27}N_4O_4S_3^+$ (*m/z* calcd for $[M+H]^+$ 507.1189, Δ0.2 ppm) featuring 12 double-bond equivalents (DBEs). Combined 1D and 2D NMR spectra uncovered a system comprising four heterocycles and one phenol, starting from the N-terminus with a phenol substituted by an oxazole group in *ortho*-position, which is further connected to two consecutive *N*-methylated thiazolidines linked via an amide bond to a cyclic γ-thiolactone as the C-terminal end (Figure 2). The assignment of the oxazole moiety and the unusual C-terminus were underpinned by ^1H-^{15}N HMBC experiments (see SI) as well as the comparison of the chemical shifts for the homocysteine thiolactone part to already published NPs like thiolactomide.[18] The two signal sets for sorangibactins A1 and A2 were identical besides the signal area of the isomerization of the C-terminal homocysteine thiolactone upon purification. The corresponding methine signal for atom 4 of sorangibactin A1 is located at $\delta(^1H) = 4.09$ ppm and of A2 at $\delta(^1H) = 4.16$ ppm, whereas the adjoining diastereotopic methylene group (atom 3) of sorangibactin A1 shows a signal at $\delta(^1H) = 2.10/1.92$ ppm and A2 features a methylene group without any diastereotopic effect at $\delta(^1H) = 2.04$ ppm. It is worth noting that in the thiolactone part the HMBC correlation between the methylene group (atom 2) at $\delta(^1H) = 3.01$ ppm and the carbonyl group (atom 1) at $\delta(^{13}C) = 179.5$ ppm could not be identified. Nevertheless, the predicted thiolactone ring is underpinned by the COSY and HMBC correlations between atoms 2 and 3 as well as atom 4, by the HMBC correlations from atom 4 and atom 3 to atom 1 as well as the predicted 12 DBEs in agreement with the predicted sum formula.

Several structure parts of sorangibactin show similarities to well described siderophores, however it features some uncommon modifications. The *ortho*-phenol moiety originating from salicylic acid incorporation is also present in yersiniabactin produced by *Yersinia pestis*[19] and predicted for coelibactin from *Streptomyces coelicolor.*[14] Furthermore, yersiniabactin features one thiazolidine moiety and two dihydrothiazoles, whereby the thiazolidine in contrast to sorangibactin is not *N*-methylated and no dihydrothiazoles are present in sorangibactin. The oxazole moiety in sorangibactin is different from the methylated dihydrooxazole predicted for coelibactin. Finally, the main structural difference of sorangibactin to published siderophores is clearly the uncommon homocysteine thiolactone at the C-terminus, reflecting the unprecedented structure of sorangibactin. This structural moiety impeded the purification procedure and also sheds light on possible obstacles that might have hindered structure elucidation of coelibactin: sorangibactins A1 and A2 undergo hydrolysis during the standard purification setup with formic acid as modifier resulting in the C-terminally ring opened sorangibactin B. The α-proton of the acid labile homocysteine thiolactone tends to isomerize even when using the adapted purification workflow leading to a mixture of the two isomers sorangibactin A1 and A2 (see Methods). Even though we could not separate these two isomers, the corresponding NMR data set of the mixture enabled the structural assignment of both of them.



1 and **2**

3

Figure 2 Chemical structures of sorangibactins A1 and A2 (**1** and **2**) and B (**3**) with atom numbering and main COSY (bold) and HMBC (arrows) correlations.

For the stereochemical assignment of sorangibactins A1 and A2 only Marfey’s derivatization assay was applicable, since this molecule is a peptide without any functionalities accessible for Mosher esterification and due to its instabilities not attainable for crystallization experiments as exemplarily performed to solve the stereochemistry of the related siderophore yersiniabactin.[20] Marfey’s analysis however could also only provide limited information for three main reasons: (1) the hydrolysis of thiazolidines is known to lead to the oxidation state of an aldehyde instead of the free acid, which is instable and degrades,[21] (2) the stereo centers at positions 9 and 15 originate from the former carboxyl function of the cysteine before cyclization to the thiazolidine group and thus would lose stereo information again upon hydrolysis, and (3) the sorangibactins A1 and A2 mixture contains both *R*- and *S*-configurations of the α-proton of the C-terminal homocysteine thiolactone, which cannot be separated. Consequently, only positions 7 and 12 could reasonably be accessible by Marfey's analysis, but they were found to isomerize during hydrolysis. HPLC-HRMS analysis displayed a mixture of four enantiomers with a mass of *m/z* 430.14±0.02 corresponding to FDLA-derivatized *N*-methyl cysteine and homocysteine (Figure S2). Hence, we can solely present the flat structure with a bioinformatic prediction of all *S*-configurations due to the absence of an epimerase in the respective modules.

We also obtained a second congener of the sorangibactin family through provoking hydrolysis of the rather labile thioester yielding the ring-opened product sorangibactin B (Figure 2). HPLC-HRMS experiments show an $[M+H]^+$ peak for sorangibactin B at *m/z* 525.1298 with a predicted molecular formula $C_{22}H_{28}N_4O_5S_3^+$ (*m/z* calcd for $[M+H]^+$ 525.1295, Δ0.6 ppm) featuring 11 double-bond equivalents (DBEs). Noteworthily, more than five isomers of sorangibactin B are present in fractions that underwent the hydrolyzing isolation conditions using formic acid during purification. The presumably isomerizing stereocenters are most likely the ones belonging to the homocysteine and the two thiazolidine parts exhibiting 5 stereocenters. Thus, the following structure elucidation belongs to the most abundant isomer unveiling the most intensive NMR correlations. In contrast to sorangibactin A1 and A2 the cyclic homocysteine thiolactone is opened, which shows an impact on the corresponding shifts due to the more shielded free carboxylic acid (atom 1) with a shift at $\delta(^{13}C) = 176.6$ ppm compared to the former thioester function. The HMBC spectrum exhibits two correlations to the free carboxyl group (atom 1) from methylene groups (atoms 2 and 3) at $\delta(^{1}H) = 2.27$ and 1.59 ppm. The more downfield shifted signal can be explained by the close neighborhood to the thiol function. Corresponding COSY correlations from the more shielded methylene group (atom 3) to the homocysteine α-proton (atom 4) at $\delta(^{1}H) = 4.68$ ppm can be observed. Furthermore, HMBC correlations evince from atom 4 as well as the methine group of the adjacent thiazolidine ring (atom 7) at $\delta(^{1}H) = 3.57$ ppm to the connective amide group (atom 6) at $\delta(^{13}C) = 173.0$ ppm. Comparing the remaining data set to sorangibactin A1 and A2, the signals corresponding to the N-terminal part from atom 8 to 26 are similar with ≤0.1 ppm deviation for ^{1}H and ≤2.1 ppm for ^{13}C shifts.

MS^{2} spectral analysis was performed to obtain additional evidence supporting the proposed structures for sorangibactins A1, A2 and B. Just as for the structurally related siderophores yersiniabactin, ulbactin B and escherichelin characteristic fragmentation behavior can also be observed for sorangibactins.[22] The three main fragment ions of sorangibactin B are in accordance with the respective bond cleavages in A1 and A2 (Figure 3A). Fragment ions b, c and d of **1** and **2** correspond to fragment ions b, c+H_2O, and d+H_2O of **3**.

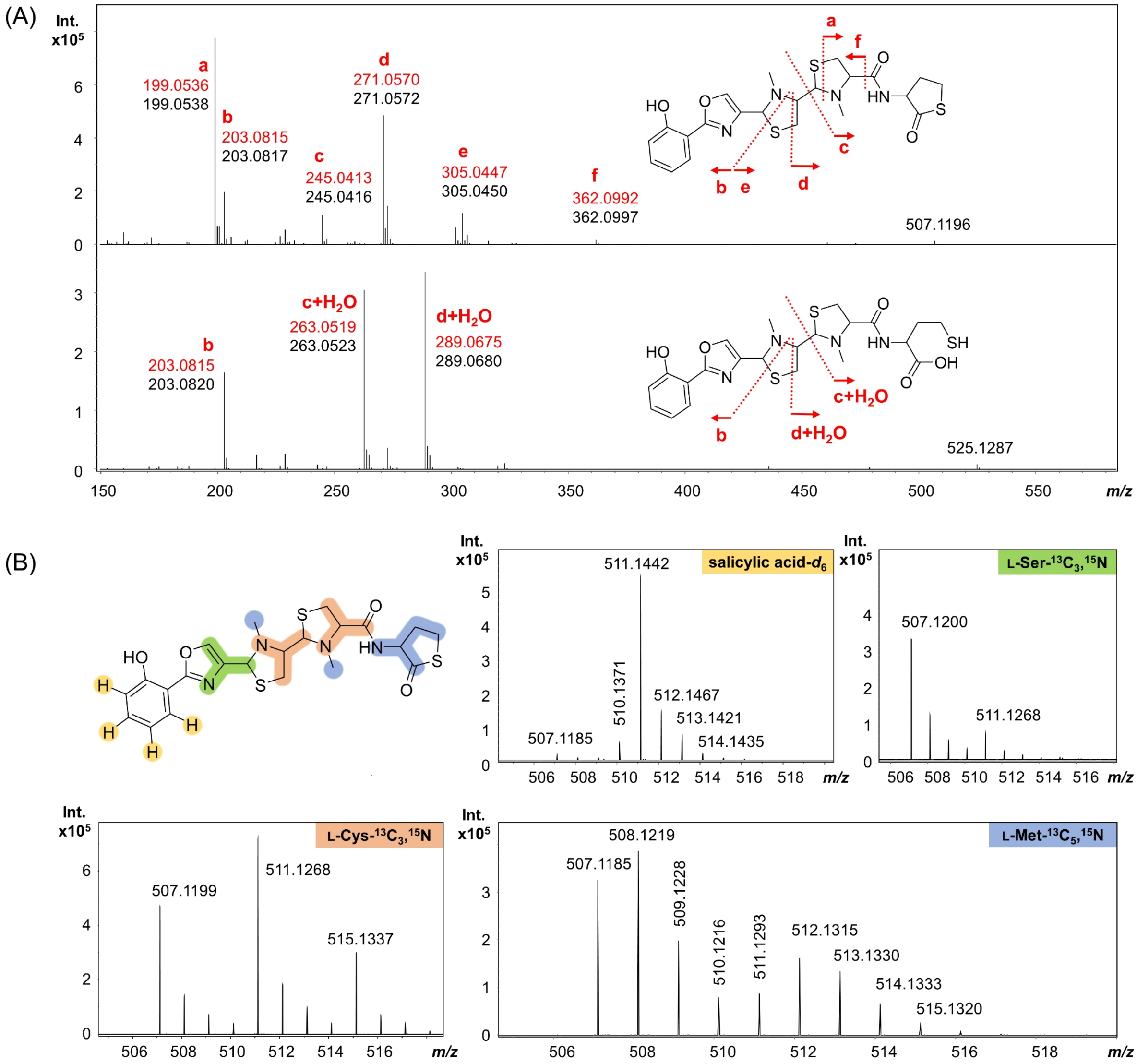


Figure 3 Characterization of sorangibactins by MS^{2} analysis and feeding stable isotopes. (A) Proposed fragmentation pattern of **1** and **2** with abundant fragment ions a−f (upper chromatogram) and of **3** with fragment ions b, c+H_2O and d+H_2O (lower chromatogram). The predicted *m/z* of each fragment is marked in red. (B) Feeding salicylic acid-*d*$_6$ (yellow), L-serine-$^{13}C_3$,^{15}N (green), L-cysteine-$^{13}C_3$,^{15}N (orange), and L-methionine-$^{13}C_5$,^{15}N (blue).

The N-terminal phenol group derived from salicylic acid is further supported by feeding 2-hydroxybenzoic acid-*d*$_6$, resulting in a +4 mass shift of the respective fragment (Figure 3B). Feeding L-serine-$^{13}C_3$,^{15}N and L-cysteine-$^{13}C_3$,^{15}N resulted in the +4 and +8 mass shift, respectively, which indicates the NRPS to incorporate one serine and two cysteines and is consistent with the proposed sorangibactin structure containing one oxazole and two thiazolidines substructures. Interestingly, feeding L-methionine-$^{13}C_5$,^{15}N not only showed a +1 mass shift indicating a methyl transfer, but also a +5, which indicated the incorporated homocysteine originating from methionine at the C-terminus, which is additionally supported by MS^2 analysis. This is consistent with the C-terminal γ-thiolactone structure.

Sorangibactins A1 and A2 showed no obvious bioactivity against our test panel of bacterial and fungal pathogens and selected cell lines (Table S2). However, siderophore-like behavior was observed during UHR-TOF measurements even before the purification of sorangibactin. Besides the double charged species at *m/z* 254.0632 and the $[M+H]^+$ signal at *m/z* 507.1190, the MS spectrum of sorangibactins A1 and A2 exhibits a peak at *m/z* 560.0307. With a mass distance of *m/z* 52.9117 to the $[M+H]^+$ peak it equals $[M\text{-}2H+Fe^{3+}]^+$ with the predicted sum formula of $C_{22}H_{24}FeN_4O_4S_3^+$ (*m/z* calcd for $[M\text{-}2H+Fe^{3+}]^+$ 560.0304, Δ0.5 ppm) and shows a characteristic M-2 iron isotope peak (Figure S3).[23] Additional evidence about sorangibactin being a siderophore is obtained by cultivation of MSr11367 under iron-deficient conditions using RGAE2 medium and iron-rich conditions by supplementing 1 mM $FeCl_3$. As a result, iron limitation during cultivation induced the production of sorangibactin, whereas external addition of iron suppressed the production (Figure S4). This is a well-established indicator for siderophore features of molecules.[24,25] Apart from that, possibilities to determine the exact affinity to iron in terms of binding properties is commonly conducted via competition experiments with a strong binder like EDTA.[26] However, this approach was not possible for the sorangibactins due to their instability and isomerization issues during the purification process.

4.3.3 Proposed biosynthesis of sorangibactins

The sorangibactin BGC spans approximately 32 kb of DNA sequence, from which 17 proteins are annotated (Table S1) and proposed to be involved in biosynthesis (SrbA−SrbK), transportation (SrbL−SrbP) and regulation (SrbR). Besides *in silico* analysis, gene deletion and point mutation by Redαβ recombineering (see Methods and Figure S1) were carried out using the heterologous expression system. Based on those experiments, we present the following biosynthesis hypothesis of the sorangibactins.

Biosynthesis of the sorangibactins was initiated by salicylic acid formation catalyzed by SrbB, which shows 48.5% similarity with Irp9, a bifunctional salicylate synthase from *Yersinia enterocolitica* that can convert chorismate to salicylic acid in yersiniabactin biosynthesis.[27] Deletion of *srbB* abolished heterologous production of sorangibactins in *M. xanthus* DK1622, whereas feeding salicylic acid to the *srbB* knockout mutant restored the production (Figure S5). The salicylic acid was then adenylated and loaded to the first carrier protein of SrbC by SrbK (Figure 4A), a stand-alone A domain-containing protein, which is predicted to use salicylic acid as substrate (Table S3). Deletion of *srbK* completely abolished the heterologous production of sorangibactins in *M. xanthus* DK1622 (Figure S6). SrbC and SrbD are non-ribosomal peptide synthases, which comprise seven and five domains, respectively (Figure 4A). Each adenylation domain was predicted by antiSMASH to use cysteine as a substrate (Table S3); however, our biosynthesis model rather suggests the first adenylation domain in SrbC to activate serine instead, which is further modified to form the oxazole incorporated in the sorangibactin structure. This is consistent with the above-mentioned result from feeding feeding L-serine-$^{13}C_3$,^{15}N (Figure 3B). The terminal thioesterase (TE) domain utilizes a cysteine as the active-site nucleophile instead of commonly used serine (Figure S7), which is not unprecedented but rare.[28] Point mutation of this cysteine to alanine or serine (Figure S8) abolished the heterologous production of sorangibactin A in *M. xanthus* DK1622 (Figure S6), which indicates its essential role. Besides, this TE domain was proposed to select homocysteine or methionine as an intermolecular nucleophile for chain release resulting in one more amino acid incorporation at the C-terminus, which was confirmed by feeding L-methionine-$^{13}C_5$,^{15}N (Figure 3B). Use of a free amino acid for chain release was previously found in threonine-tagged lipopeptide biosynthesis, where the TE domain selects the intermolecular nucleophiles threonine hydroxy or amino groups yielding ester- and amide-linked threonine tags.[29] However, the complete C-terminal homocysteine thiolactone formation in sorangibactin biosynthesis is still unclear and may involve other tailoring enzymes, which is still under investigation.

Besides SrbK and SrbB−SrbD for the core biosynthesis, SrbE−SrbJ are proposed to be tailoring enzymes that work *in trans*. SrbE is a putative oxidoreductase containing a saccharopine dehydrogenase NADP-binding domain (PF03435.17) and SrbF/G are proposed to be methyltransferases. Gene deletion of SrbE/F/G was achieved by replacing them with an ampicillin resistance gene, followed by a Ptet promoter to drive the downstream expression avoiding polar effects (Figure S1), which resulted in the abolishment of sorangibactin A production and indicated their essential role for biosynthesis (Figure S6). SrbH is a cytochrome P450-dependent enzyme, sharing remarkable similarity with the CYP105N1 from *S. coelicolor* A3(2) coelibactin biosynthetic pathway both in sequence (36.2% identity, 51.6% similarity) and structure (RMSD 1.895, Figure S9). Structure characterization of CYP105N1 revealed an unexpected wide open substrate binding pocket, which was proposed to bind the peptidyl carrier protein (PCP)-bound substrate; however, the reaction catalyzed by CYP105N1 is still unclear.[30,31] Deletion of *srbH* abolished sorangibactin A production in *M. xanthus* DK1622, whereas another compound, termed sorangibactin C, was found accumulated (Figure S6). Molecular ion clusters $[M+H]^+$ (*m/z* 509.1332) and $[M+2H]^{2+}$ (*m/z* 255.0716) in ESI-HRMS revealed the elemental formula $C_{22}H_{28}N_4O_4S_3$ of sorangibactin C, which shows highly similar fragmentation pattern with sorangibactin A and contains two more atoms of hydrogen in the first two cyclic moieties (Figure S10). Feeding of L-serine-2,3,3-d_3 revealed that these two atoms of hydrogen originated from serine, and therefore, oxazoline instead of oxazole is proposed for sorangibactin C structure (Figure 4B). SrbH therefore presumably targets the carrier protein-bound substrate and catalyzes oxazoline dehydrogenation to form an oxazole on the assembly line (Figure 4A). However, the accumulation of sorangibactin C could also indicate SrbH to be a post-tailoring enzyme. SrbI shows 47.9% similarity with Irp3 from *Yersinia enterocolitica* (PDB: 5KVQ) and is assigned as *trans*-acting thiazolinyl imine reductase, catalyzing the NADPH-dependent reduction of a CP-tethered thiazoline ring.[32,33] Deletion of *srbI* abolished sorangibactin A production, whereas the shunt product sorangibactin D observed at $[M+H]^+$ *m/z* 376.0422 (calculated $C_{16}H_{14}N_3O_4S_2^+$ 376.0420, Δ0.5 ppm) was accumulated (Figure S11), which indicates that the thiazoline reduction is necessary for downstream processing, otherwise resulting in early hydrolytic release. SrbJ is proposed to belong to the VOC (vicinal oxygen chelate) superfamily, that catalyzes diverse reactions.[34] The function of SrbJ in sorangibactins biosynthesis is unclear but indispensable, as of *srbJ* deletion abolished sorangibactin A production (Figure S6). At this moment, we cannot fully assign each of these biosynthesis enzymes to the structure, as the obvious accumulation of intermediate was not observed in each gene knockout, presumably due to the non-native intermediates stalling the whole machinery or chemical instabilities of the resulting shunt products.

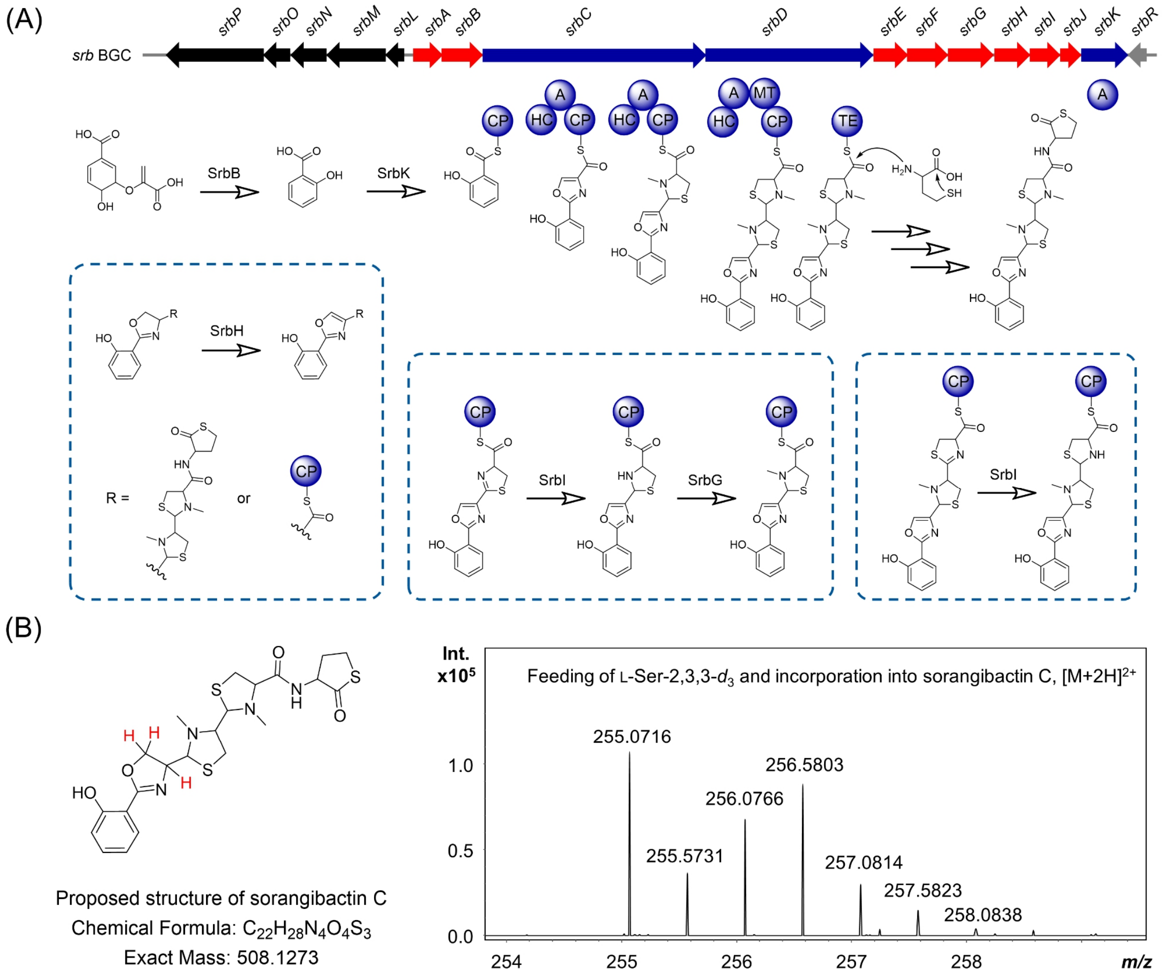


Figure 4 Characterization of the biosynthesis of sorangibactins. (A) Gene cluster organization and proposed biosynthetic pathway. The core biosynthetic genes are marked in blue, the proposed additional biosynthetic genes are marked in red, the proposed transport-related genes and regulatory gene are marked in black and grey, respectively. CP: carrier protein domain, HC: heterocyclization domain, A: adenylation domain, MT: methyltransferase domain, and TE: thioesterase domain. (B) Feeding of L-Ser-2,3,3-*d*3 and incorporation into sorangibactin C, where the proposed deuterium incorporation sites are marked in red.

In addition to the ten core genes described above that were all experimentally confirmed to be directly involved in sorangibactins biosynthesis, seven additional genes were found in the sorangibactin BGC. SrbA is a proposed tailoring enzyme containing an α/β hydrolase domain. Deletion of *srbA* by Redαβ recombineering (see the Methods section and Figure S1) had no obvious influence on heterologous production of sorangibactins in *M. xanthus* DK1622 (Figure S5 and Figure S6), which can either be explained by a functional homologue of *srbA* present in the *M. xanthus* DK1622 chromosome or *srbA* being not involved in sorangibactins biosynthesis. SrbR is an LysR family transcriptional regulator, and SrbL−SrbP are all transporter related. Whether or how the regulator and these transporters are correlated to sorangibactin production, however, was not investigated here.

4.4 Conclusions

Genome mining and subsequent heterologous expression enabled the discovery of a novel myxobacterial natural product termed sorangibactin from the myxobacterial strain MSr11367. This genomics-guided approach highlighted the accessibility of its biosynthetic potential that was not reflected in its metabolome in our previous studies. *De novo* structure elucidation of sorangibactin revealed the presence of an unprecedented γ-thiolactone at the C-terminus, which may help to solve the puzzle of the chemical structure of coelibactin based on their gene cluster similarity via the adapted purification route reducing putative degradation and isomerization issues as well as by providing reference data for comparison to facilitate the structure elucidation. Both biosynthetic pathways contain conserved cytochrome P450 enzymes, and the coelibactin P450 CYP105N1 protein was structurally characterized; however, their functions remained elusive before this study. Here, we discovered the sorangibactin P450-dependent enzyme SrbH, which is proposed to oxidize oxazoline to oxazole, representing the first P450-mediated oxazoline dehydrogenation in NRPS biosynthesis. Furthermore, both gene clusters involve an unusual TE domain containing a rare active-site cysteine instead of the commonly used serine. For the sorangibactin biosynthesis, we propose that homocysteine or methionine is incorporated into the C-terminus by this unusual TE domain, most probably together with other unassigned tailoring enzymes, resulting in γ-thiolactone moiety formation in a manner that is not yet fully understood. Therefore, the bioinformatically predicted coelibactin structure remains to be experimentally characterized. Future studies could address the indispensable SrbE (saccharopine dehydrogenase NADP-binding domain-containing protein, also present in coelibactin biosynthetic pathway with yet unknown function) and SrbJ (glyoxalase-like domain-containing protein, absent in coelibactin biosynthetic pathway but conserved in many other BGC homologues), both of which are rare in natural product biosynthesis and worthy for further biochemical investigations. Homologous gene clusters are widespread not only in myxobacteria but also in *Streptomyces* and *Pseudomonas* strains, which indicates conservative functions across bacteria and potential for further discovery and characterization of sorangibactin analogues.

In summary, this study showcased the benefit of heterologous expression with subsequent gene cluster engineering for the discovery and biosynthesis investigations of myxobacterial natural products. Accessibility of a cosmid library herein will allow further high-throughput screening of cryptic myxobacteria BGCs by heterologous expression and facilitate characterization of their yet untapped biosynthetic potential.

4.5 Methods

4.5.1 Cultivation conditions for bacterial strains

The *Sorangiineae* strain MSr11367 was cultivated in 100 mL RGAE2 medium (3.5 g/L casitone, 1 g/L soytone, 0.5 g/L yeast extract, 2 g/L glucose, 2 g/L soluble starch, 1 g/L sucrose, 1 g/L maltose monohydrate, 2 g/L cellobiose, 0.5 g/L calcium chloride dihydrate, 1 g/L magnesium sulfate heptahydrate, 2.38 g/L HEPES, pH=7.0) or CYHv2 medium (4 g/L soya meal starch, 3 g/L casitone, 2 g/L glucose, 8 g/L soluble starch, 1.5 g/L yeast extract, 1 g/L calcium chloride dihydrate, 1 g/L magnesium sulfate heptahydrate, 11.9 g/L HEPES, 8 mg/L Fe-EDTA, pH=7.2) in a 300 mL flask with shaking at 180 rpm and 30 °C for 14 days. *M. xanthus* DK1622 was used as a heterologous host for gene cluster expression and cultivated in CTT medium (10 g/L casitone, 10 mM Tris-HCl, 8 mM magnesium sulfate, 1 mM potassium phosphate, pH=7.6) at 30 °C. *E. coli* strains GB2005, GB05-red, GBred-gyrA462 were used for gene cluster engineering and cultivated in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH=7.0) at 37 °C. The following antibiotics were supplemented to the medium when the cultivated strain carries the corresponding resistance gene: 10 µg/mL for oxytetracycline, 15 µg/mL for chloramphenicol, 50 µg/mL for kanamycin and 100 µg/mL for ampicillin.

4.5.2 Gene cluster cloning and engineering

The cosmid library for *Sorangiineae* strain MSr11367 was constructed by Explogen LLC (EXG) (https://explogen.com.ua) using genomic DNA and p15A-cm-tet-int vector. After end-sequencing, the cosmid pCos-D05_9223 was found to be able to cover the putative whole gene cluster and used for heterologous expression. For generation of promoter engineered expression construct (pCos-D05_9223-km-Pvan), a *km-vanR-Pvan* cassette amplified by PCR was electroporated into *E. coli* GB05-red harboring plasmid pCos-D05_9223 and expressing Redαβ recombinases for linear-circular homologous recombination (LCHR). Gene knockout constructs (pCos-D05_9223-km-Pvan-delA and pCos-D05_9223-km-Pvan-delAB) were generated in a similar manner except reverse oligo nucleotides containing different downstream homology arms were used. Deletion of *srbE/F/G/H/I/J/K* was based on pCos-D05_9223-km-Pvan. To do this, *amp-Ptet* cassettes amplified by PCR using corresponding oligo nucleotides (Table S4) were used to replace the target gene using LCHR. As a result, the downstream genes will be under control of a Ptet promoter in order to avoid potential polar effect. For point mutation of the TE domain, RedEx method was used,[35] *amp-ccdB* cassettes containing homology arms, mutation sites and PmeI sites were amplified by PCR using corresponding oligo nucleotides (Table S4) and electroporated into *E. coli* GBred-gyrA462 harboring plasmid pCos-D05_9223-km-Pvan for LCHR. The resulting plasmids were digested by PmeI, recircularized by Gibson Assembly and electroporated into *E. coli* GB2005. Correct recombinants were confirmed by restriction digestions and Sanger sequencing.

4.5.3 Heterologous expression in *M. xanthus* DK1622

Expression constructs containing the sorangibactin BGC were transferred to the heterologous host *M. xanthus* DK1622 by electroporation according to a previously established procedure.[36] The gene cluster was then integrated into the *attB* locus on the *M. xanthus* DK1622 chromosome through phage Mx8 integrase mediated site-specific recombination.[37] Then, the mutants were checked by colony PCR using three pairs of primers (Table S4) targeting different positions of the gene cluster to ensure the intact integration. Three independent mutants were inoculated in CTT medium supplemented with appropriate antibiotics and cultivated at 30 °C with shaking until late log-phase. The culture was inoculated 1 to 50 into fresh CTT medium supplemented with 2% XAD-16 resin, appropriate antibiotics and inducers (1 mM vanillate for Pvan promoter and 0.5 μg/mL anhydrotetracycline for Ptet promoter). Cultivation was continued at 30 °C with shaking for 3 days. For comparative analysis, triplicate cultivation of *M. xanthus* DK1622 wild-type strains ware performed in a similar manner except that no antibiotics were added. The cells and adsorber resin were collected by centrifugation and extracted with methanol. The crude extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-HRMS measurement.

4.5.4 HPLC-HRMS measurements

All HRMS measurements were conducted on a maXis 4G UHR-TOF system (Bruker Daltonics) equipped with an ESI source and coupled to a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC, Thermo Fisher) system. Separation is achieved with an Acquity BEH C18 column (100 mm × 2.1 mm, 1.7 μm dp) (Waters) in combination with a Waters VanGuard BEH C18 1.7 μm guard using a linear 5−95% gradient of acetonitrile with 0.1% formic acid in ddH_2O with 0.1% formic acid for 18 min. Detection was conducted by a diode array detector at 200−600 nm, the flow rate adjusted to 0.6 mL/min, the column heated to 45 °C and the HPLC flow split to 75 μL/min before entering the mass spectrometer. Following MS settings were applied for standard measurements: capillary voltage 4000 V, end plate off-set -500 V, nebulizer gas pressure 1 bar, dry gas flow rate 5 L/min, dry gas temperature 200 °C, mass scan range *m/z* 150−2500. MS/MS measurements were carried out in collision-induced dissociation (CID) fragmentation mode with a collision energy of 5 eV. The mass spectrometer was calibrated in Quadratic + HPC calibration mode on the masses of the first isotope signal of sodium formate clusters, which are formed in the ion source after co-injection of a mixture of 1 mM sodium hydroxide in 1:1 water/isopropanol + 0.1% FA. Additional lock mass calibration was performed on *m/z* 622.0277, *m/z* 922.0103 and *m/z* 1221.9960.

4.5.5 Comparative data analysis

Data processing was performed using Bruker DataAnalysis 4.4 and MetaboScape 2021b. For comparative analysis, bucket tables consisting of molecular features (RT-*m/z* pairs) and their corresponding intensity values were generated using the T-ReX 3D algorithm. Principal component analysis (PCA) and t-test were used to find out unique features that are only present in the heterologous expression group, but not in the control group. These unique features were double-checked manually by extracting the corresponding ion chromatograms.

4.5.6 Feeding stable isotope-labeled precursors

The stable isotopes were prepared as stock solutions of 0.2 M and supplemented to the culture at the final concentration of 0.2 mM. To prepare stock solutions, L-cysteine-$^{13}C_3$,^{15}N (Sigma Aldrich, 658057), L-serine-$^{13}C_3$,^{15}N (Sigma Aldrich, 608130), L-serine-2,3,3-d_3 (Cambridge Isotope Laboratories, DLM-582-0.5), L-methionine-d_3 (methyl-d_3) (Campro Scientific GmbH, CS01-182_581) and L-methionine-$^{13}C_5$,^{15}N (Campro Scientific GmbH, CS01-181_212) were dissolved in autoclaved Milli-Q water, 2-hydroxybenzoic acid-d_6 (Campro Scientific GmbH, CS01-182_507) was dissolved in ethanol. The heterologous host *M. xanthus* DK1622::km-Pvan-srb was inoculated in 20 mL CTT medium supplemented with kanamycin at 30 °C with shaking until mid log-phase. After addition of 1 mM vanillate, 2% XAD-16 resin and corresponding stable isotope, cultivation was continued for two more days. The cells and adsorber resin were collected and extracted with methanol. The crude extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-HRMS measurement.

4.5.7 Fermentation and compounds purification

The heterologous producer strain *M. xanthus* DK1622::km-Pvan-srb was inoculated in CTT medium supplemented kanamycin and cultivated at 30 °C with shaking until late log-phase. The culture was inoculated 1 to 50 into 6 × 2 L fresh CTT medium supplemented with 2% XAD-16 resin, 50 µg/mL kanamycin and 1 mM vanillate in 6 × 5 L flaskes. Cultivation was continued at 30 °C with shaking for 4 days. The cells and adsorber resin were collected by centrifugation and extracted stepwise with methanol. The methanol extract was concentrated and partitioned with hexane to remove nonpolar impurity. After evaporation *in vacuo*, the methanol phase was then dissolved in Milli-Q water and extracted twice with equal portions of chloroform. The chloroform phase was evaporated to dryness and redissolved in methanol.

Isolera^{TM} Spektra One (Biotage) system: Normal phase chromatography was performed using silica gel (60 Å, 70−230 mesh, 63−200 µm) packed in a SNAP 50g column as stationary phase and a multistep gradient with hexane, ethyl acetate and methanol as mobile phase using a 15 column volume (CV) gradients and 5 CV for equilibration at a flowrate of 50 mL/min.

Semi-preparative purification of sorangibactins A1 and A2: Respective fractions resulted from Biotage work up containing sorangibactins A1 and A2 were further purified using a

Dionex Ultimate 3000 SL system comprising a SWPS 3000 SL autosampler, P680 pump, TCC100 column oven heated to 45 °C, PDA100 UV-detector at 220 nm and an AFL 3000 fraction collector. Separation was reached using a Waters XBridge Peptide BEH C18 column (250 x 10 mm, 5 µm) in combination with a mobile phase consisting of A = 0.1% acetic acid in ddH_2O and B = 0.1% acetic acid in acetonitrile running a 5−95% B gradient in 16 min with a 4 min equilibration at 5% B with a flow rate of 5 mL/min. The acid labile sorangibactins A1 and A2 were pooled in a flask filled with $NaHCO_3$ for neutralization purpose, which was subsequently disjoint through liquid-liquid partitioning three times using chloroform. In total 2.4 mg sorangibactin A1 and A2 mixture was obtained and stored under nitrogen at -20 °C until further use.

Semi-preparative purification of sorangibactin B: Separation was achieved via a Waters XBridge Peptide BEH C18 column (250 x 10 mm, 5 µm) using formic acid instead of acetic acid and following adapted multi-step gradient: hold at 5% B for 1 min, increasing to 20% B within 0.5 min, 1.5−12.5 min increasing to 80% B, increasing to 95% B within 0.5 min, hold at 95% B for 1.5 min, go back to 5% B within 0.5 min and re-equilibrate for 5 min at a flow rate of 5 mL/min. The obtained fraction was kept at room temperature overnight to achieve conversion to sorangibactin B. Finally, 4.5 mg of sorangibactin B isomer mixture could be obtained.

4.5.8 NMR measurements and structure elucidation

NMR measurements: For structure elucidation of sorangibactins A1 and A2, all ^{1}H-, ^{13}C- and 2D-NMR spectra were acquired in methanol-d_4 on a Bruker Avance III (Ascend) 700 MHz spectrometer equipped with a 5 mm TCI cryoprobe using standard pulse programs. Observed chemical shifts (δ) are listed in ppm, coupling constants (J) in Hz and all spectra were calibrated with respective methanol signals at $\delta(^1H) = 3.31$ ppm and $\delta(^{13}C) = 49.2$ ppm. Sorangibactin B was measured using the same parameters on a Bruker Avance III (Ultrashield) 500 MHz spectrometer.

Structure elucidation of sorangibactins A1 and A2 mixture: ^{1}H-NMR, ^{13}C and HSQC spectra show five aromatic double-bond signals at $\delta(^1H) = 7.89$ (1H, s) $\delta(^{13}C) = 136.4$, $\delta(^1H) = 7.82$ (1H, d, $J = 7.80$ Hz) $\delta(^{13}C) = 127.2$, $\delta(^1H) = 7.36$ (1H, m) $\delta(^{13}C) = 133.7$, $\delta(^1H) = 7.00$ (1H, d, $J = 8.38$Hz) $\delta(^{13}C) = 118.2$ and $\delta(^1H) = 6.95$ (1H, m) $\delta(^{13}C) = 120.7$ ppm.

Furthermore, five methine signals can be found at $\delta(^{1}H) = 5.01$ (1H, s) $\delta(^{13}C) = 70.1$, $\delta(^{1}H) = 4.21$ (1H, m) $\delta(^{13}C) = 76.1$, $\delta(^{1}H) = 4.09$ (1H, dd, $J = 6.94$, 3.66 Hz) $\delta(^{13}C) = 63.1$, $\delta(^{1}H) = 3.59$ (1H, t, $J = 7.95$Hz) $\delta(^{13}C) = 81.2$ as well as $\delta(^{1}H) = 3.29$ (1H, m) $\delta(^{13}C) = 75.6$ ppm. Additionally, two methyl signals are located at $\delta(^{1}H) = 2.57$ (3H, s) $\delta(^{13}C) = 41.5$ and $\delta(^{1}H) = 2.47$ (3H, s) $\delta(^{13}C) = 40.8$, one methylene signal at $\delta(^{1}H) = 3.01$ (2H, m) $\delta(^{13}C) = 24.9$ and three diastereotopic methylene signals at $\delta(^{1}H) = 3.15$, 3.02 (2H, m) $\delta(^{13}C) = 35.3$, $\delta(^{1}H) = 3.29$, 3.16 (2H, m) $\delta(^{13}C) = 33.8$ and $\delta(^{1}H) = 2.10$, 1.92 (2H, m) $\delta(^{13}C) = 24.7$ ppm. Sorangibactin A1 consists of five ring systems, starting from an N-terminal with a salicylic acid, which carbonyl group is fused into an oxazole connected to two consecutive *N*-methylated thiazolidines linked via an amide bond to a cyclic homocysteine thiolactone C-terminus. The salicylic acid moiety is assigned by corresponding COSY correlations between the four aromatic double bond signals at $\delta(^{1}H) = 7.82$, 7.36, 7.00 and 6.95 ppm in combination with HMBC correlations to the quaternary carbon with a downfield shift at $\delta(^{13}C) = 158.5$ ppm due to the hydroxyl substituent and a quaternary carbon, which is deshielded with a shift of $\delta(^{13}C) = 112.4$ ppm. The remaining aromatic signal exhibits a characteristic downfield shift at $\delta(^{1}H) = 7.89$ ppm with distinct HMBC correlations to two quaternary carbons at $\delta(^{13}C) = 162.9$ and 144.5 ppm pointing towards a heterocyclic ring with two double bonds to maintain aromatization. The ^{1}H-^{15}N-HMBC spectrum uncovers the corresponding correlation to a deshielded nitrogen atom with a downfield shift at $\delta(^{15}N) = 239.9$ ppm characteristic for oxygen containing heteroaromatic systems as oxazole moieties. The linkage between salicylic acid and oxazole is the deshielded quaternary carbon at $\delta(^{13}C) = 162.9$ ppm due to HMBC correlations from the salicylic acid signal at $\delta(^{1}H) = 7.82$ ppm. This carbon atom represents the fused carbonyl group of the salicylic acid. The connection to the adjacent ring is represented by the corresponding COSY and HMBC correlations from the methine signal at $\delta(^{1}H) = 5.01$ ppm to the oxazole moiety. Corresponding 2D NMR correlations underpin this ring to be an *N*-methylated thiazolidine cycle. The diastereotopic methylene signal could be observed at $\delta(^{1}H) = 3.29$ and 3.16 ppm as well as the methyl signal at $\delta(^{1}H) = 2.57$ ppm with HMBC correlations to the methine signal at $\delta(^{1}H) = 5.01$ ppm and the one at $\delta(^{1}H) = 3.29$ ppm, respectively. Further on, the COSY spectrum exhibits cross peaks for the latter methine signal to the methine signal of the subsequent *N*-methylated thiazolidine ring at $\delta(^{1}H) = 4.21$ ppm. The corresponding shifts of this ring are more shielded in comparison to the previous are explained by its greater distance to the deshielding oxazole moiety. This is supported by the ^1H-^{15}N HMBC spectrum with corresponding cross peak signals for the nitrogen atom in the third ring at $\delta(^{15}N) = 59.8$ ppm and the one in the fourth cycle at $\delta(^{15}N) = 56.7$ ppm. Finally, the fourth and fifth ring system are linked via an amide group with a shift of $\delta(^{13}C) = 167.9$ in line with HMBC correlations from the thiazolidine signals at $\delta(^1H) = 3.59$ and 3.15, 3.02 ppm on the one side and the methine signal at $\delta(^1H) = 4.09$ ppm from the last ring on the other side. This moiety features COSY correlations from a diastereotopic signal at $\delta(^1H) = 2.10$, 1.92 ppm to the mentioned methine group as well as to another methylene group at $\delta(^1H) = 3.01$ ppm. Moreover, corresponding HMBC correlations from this methine and diastereotopic methylene signals to the deshielded quaternary carbon with a shift of $\delta(^{13}C) = 179.5$ ppm indicate a cyclic homocysteine thiolactone C-terminus, which is in line with already published homocysteine thiolactones like thiolactomide [18]. The respective deviations between our measurements in methanol-d_4 and the reference data of thiolactomide in dimethylsulfoxid-d_6 are <0.5 ppm for ^1H- and <6 ppm for ^{13}C-shifts. However the quaternary carbon of the thioester shows a greater difference with 26.3 ppm, which can be explained by the different surrounding functions of shielding thiazolidins and deshielding alkyl chains. Further, the missing correlation in the ^1H-^{15}N-HMBC spectrum might be due to a possibly unfavorable conformation for the detection of the $^3J(^1H$-$^{15}N)$-coupling from the diastereotopic methylene group to the nitrogen atom of the amide bond.

The two signal sets for sorangibactins A1 and A2 were identical except for signals corresponding to the C-terminal homocysteine thiolactone due to isomerization upon purification. The methine signal for sorangibactin A1 can be found at $\delta(^1H) = 4.09$ ppm and for A2 at $\delta(^1H) = 4.16$ ppm, whereby the adjoining diastereotopic methylene group of sorangibactin A1 is located at $\delta(^1H) = 2.10$,1.92 ppm and A2 shows a methylene group signal without any diastereotopic effect at $\delta(^1H) = 2.04$ ppm.

Structure elucidation of sorangibactin B: explained in the section of **4.3.2**.

4.5.9 Marfey's analysis for the determination of stereochemistry

For the determination of stereocenters 7 and 12 of the *N*-methylated thiazolidines as well as stereo center 4 of the cyclic homocysteine thiolactone moiety, an adapted Marfey's derivatization protocol already described for the thiamyxins was applied after acidic hydrolysis to *N*-methyl cysteine and homocysteine, respectively.[38] About 0.1 mg of sorangibactins A1 and A2 mixture was hydrolyzed in 100 µL of 3 N HCl at 110 °C for 45 min. After evaporation of the solvent at 110 °C, two 50 µL aliquots of hydrolyzed sorangibactins A1 and A2 in ddH_2O were supplemented with 60 µL of 1 N $NaHCO_3$ each and 20 µL of the derivatization reagent (1% 1-fluroro-2,4-dinitrophenyl-5-leucine-amide solution in acetone [D-FDLA and L-FDLA]). The obtained mixture was reacted at 40 °C and 700 rpm for 2 h. Addition of 60 µL 3 N HCl stopped the reaction, and the solution was diluted with 150 µL of methanol and acetonitrile to be analyzed by HPLC-HRMS with maXis4G system coupled to a Dionex Ultimate 3000 SL system. The measurement was done in positive ESI mode after separation by a Waters Acquity BEH C18 column (50 x 2.1 mm, 1.7 µm) with following multi-step gradient (A = 0.1% formic acid in ddH_2O and B = 0.1% formic acid in acetonitrile): 5−10% B in 1 min, 10−35% B in 14 min, 35−55% in 7 min, 55−80% in 3 min, hold at 80% for 1 min and re-equilibration at 5% B for 5 min. The flow was split to 75 µL/min before entering the MS. UV data were recorded at 340 nm simultaneously with the MS detection in centroid mode ranging from 150−1000 *m/z*.

4.5.10 Bioactivity evaluation of sorangibactin A

All microorganisms used for the biological assays were obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures) or were part of our in-house strain collection. All microbial strains, except for *M. smegmatis* which was inoculated into M7H9 medium (4.7 g/L Middlebrook 7H9 Broth Base (Sigma M0178-500), 4 mL glycerol, 100 mL BD BBL™ Middlebrook OADC Enrichment) and fungi which were inoculated into Myc 2.0 medium (10 g/L Bacto peptone, 10 g/L yeast extract, 20 mL glycerol, pH 6.3), were inoculated into MHB II medium (BD 212322) to achieve a final inoculum of 10^5 colony-forming units (CFU)/mL. Sorangibactin A was prepared as a DMSO stock. Serial dilutions in the growth medium were prepared in a sterile 96-well plate and the strain suspensions were added. Growth inhibition was assessed after incubation for 24 h at RT, 30 °C or 37 °C. Minimum inhibitory concentrations (MIC) are defined as the lowest compound concentration where no visible growth is observed.

The HepG2 cell line was obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures). Cells were seeded into a sterile CellBIND® 96-well plate with DMEM (Gibco 41965039) supplemented with 10% FBS (PAN-Biotech P40-37500) to obtain a final concentration of 10^5 cells/mL. For equilibration, the cells were incubated at 37 °C with 5% CO_2 for 2 h. The prepared sorangibactin A stock in DMSO was added to a separate sterile flat-bottom 96-well plate, serially diluted, and transferred from the lowest to highest concentration to the cell plate. After static incubation at 37 °C with 5% CO_2 for 5 days, the cells were stained with thiazolyl blue tetrazolium bromide (MTT-reagent) at a concentration of 500 µg/mL for 2 h at 37 °C with 5% CO_2. The cells converted the reagent to violet crystals. After discarding the media, hydrochloride-isopropanol (1:250) solution was added to dissolve the crystals. OD was measured at 570 nm and 630 nm as a reference. IC_{50} was assessed by analyzing the nonlinear curve regression.

4.6 Supporting information

4.6.1 Supplementary tables

Table S1 Annotation of the sorangibactin biosynthetic gene cluster. BLAST was performed using the blastp algorithm with default parameters in the NCBI non-redundant protein sequences database.

Name	Length (aa)	Deduced function	Closest homologue	Coverage/identity (%)
SrbA	317	alpha/beta hydrolase	PYQ66785.1	89/44.70
SrbB	472	salicylate synthase	WP_020740538.1	93/61.88
SrbC	2146	non-ribosomal peptide synthetase	WP_020740537.1	98/65.29
SrbD	1814	non-ribosomal peptide synthetase	KYF49860.1	99/68.72
SrbE	371	saccharopine dehydrogenase NADP-binding domain-containing protein	WP_020740533.1	100/61.89
SrbF	482	bifunctional Gfo/Idh/MocA family oxidoreductase/class I SAM-dependent methyltransferase	WP_020740532.1	80/49.21
SrbG	516	class I SAM-dependent methyltransferase	AGP40737.1	99/54.26
SrbH	394	cytochrome P450	KYG09918.1	99/74.10
SrbI	358	thiazolinyl imide reductase	KYF49855.1	97/74.86
SrbJ	237	VOC family protein	WP_044986807.1	95/67.11
SrbK	538	AMP-binding protein	WP_044986805.1	99/73.64
SrbL	190	TonB-dependent siderophore receptor	WP_245591859.1	55/60.58
SrbM	675	TonB-dependent siderophore receptor	WP_020740545.1	100/66.62
SrbN	407	hypothetical protein	KYF49844.1	90/62.77
SrbO	297	ABC transporter ATP-binding protein	WP_020740540.1	95/79.79
SrbP	1130	ABC transporter permease	KYG09926.1	99/68.62
SrbR	93	LysR family transcriptional regulator	WP_089851301.1	83/45.45

Table S2 Bioactivity results of sorangibactin A (**1** and **2**).

Microbial strains	**MIC [μg/ml]**
E. coli acrB [JW0451-2]	64
P. aeruginosa PA14 [DSM-19882]	>64
E. coli WT [BW25113]	>64
B. subtilis [DSM-10]	64
S. aureus [Newman]	64
A. baumannii [DSM-30008]	>64
C. freundii [DSM-30039]	>64
M. hiemalis [DSM-2656]	32
P. anomala [DSM-6766]	64
C. albicans [DSM-1665]	>64
C. neoformans [DSM-11959]	64
M. smegmatis [MC2155]	>64
Cell line	**IC_{50} [μg/mL]**
HepG2	8.952 ± 5.65

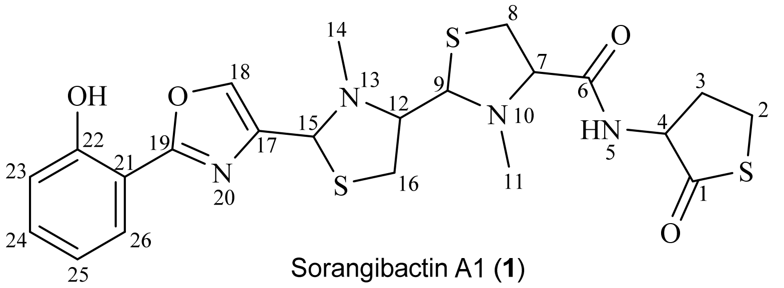
Table S3 Substrate specificities of adenylation domains in *srb* gene cluster. Prediction was performed by NRPSpredictor2 incorporated in antiSMASH.

Domain	**Specificity-conferring code**	**Predicted substrate**	**Actual substrate**
SrbC-A1	DLYNLGLIHK	Cys	Ser
SrbC-A2	DLFNLSMIWK	Cys	Cys
SrbD-A	DLYNLSLIWK	Cys	Cys
SrbK-A	ALPAQGVLNK	Sal	Sal

Table S4 Oligo nucleotides used in this study.

Name	Sequence (5' to 3')	Application
cos-chk-F	TCGTGCTGCTGTCATTGGTGTGC	end-sequencing of cosmid
cos-chk-R	ATAGCGCCCCTGTGTGTTCTCGT	
km-vanR-Pvan-F	ACGAGGCACGCGCTCGCTGCCACCAGGGCAGCACAGAAGACAAGTCTCATCATAATCTGTACCTCCTTAAGTCAGAAGAACTCGTCAAGAAGGC	promoter engineering
km-vanR-Pvan-R	GCTGCGCCTCGAGCCTCCTGCCCCTGCGCGGACACGGCACACAGGATTCCCATATGCGTTTCCTCGCATCG	
km-vanR-Pvan-delA-R	ACGTACGTCGGCCTATCGGACATACCGTCCTCAGCGTACTGTCGATATTTCATATGCGTTTCCTCGCATCG	knockout of *srbA*
km-vanR-Pvan-delAB-R	AGCACGCGGGCGATCGTGCCCCGCACATCCAGGAAACTATCGTGACTCGGCATATGCGTTTCCTCGCATCG	knockout of both *srbA* and *srbB*
amp-Ptet-delE-F	TCATCGCCGTGCTCGGTGCTTACGGCGATGTGGGGCGCCAGGCCACGCGGTTACCAATGCTTAATCAGTGAGG	knockout of *srbE*
amp-Ptet-delE-R	CGTCTCGACCAGCACGGAGATCCCTCGTTCGAGCTGCTCGCCGATGCTCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delF-F	GCATCGGCGAGCAGCTCGAACGAGGGATCTCCGTGCTGGTCGAGACGCCGTTACCAATGCTTAATCAGTGAGG	knockout of *srbF*
amp-Ptet-delF-R	TCGAGGGTCCGGCACCACGCGGCGAGCAAGGCCGAGAAGTCGATGATCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delG-F	TCATCGACTTCTCGGCCTTGCTCGCCGCGTGGTGCCGGACCCTCGATCCGTTACCAATGCTTAATCAGTGAGG	knockout of *srbG*
amp-Ptet-delG-R	TCGATGCCCTCGAAGTCGTTGAACGGATACGGTCGGGGGGAGTCCGTCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delH-F	CTCGACCAAAAGGATTCGACCCATGACGGACTCCCCCCGACCGTATCCGTTTACCAATGCTTAATCAGTGAGG	knockout of *srbH*
amp-Ptet-delH-R	AAGGTAAATGCGGCCAAAGGCCGTGCCCGCGACGAGGACGCGACGTTTCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delI-F	AACGTCGCGTCCTCGTCGCGGGCACGGCCTTTGGCCGCATTTACCTTGCCTTACCAATGCTTAATCAGTGAGG	knockout of *srbI*
amp-Ptet-delI-R	TGGTGAATGTCGCGAACCCAGAGCAAAACGTGGCTACAGGACGTGGTCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delJ-F	CCACGTTTTGCTCTGGGTTCGCGACATTCACCAGGCGGTGCGCGACTACCTTACCAATGCTTAATCAGTGAGG	knockout of *srbJ*
amp-Ptet-delJ-R	CGGTACGTCGCCGCGAACTCCGCAGGCCAATTTACGAAGCCTTCGAGCATAGATCCTTTCTCCTCTTTAGATC	
amp-Ptet-delK-F	ATTGGCCTGCGGAGTTCGCGGCGACGTACCGCGAACTCGGCTACTGGCGTTTACCAATGCTTAATCAGTGAGG	knockout of *srbK*
amp-Ptet-delK-R	CCGCGAGGCTGCGCGGCGATCCGTGCGGCCAGCTCGCGCTTGTTGACCTTAGATCCTTTCTCCTCTTTAGATC	
TE-C1586A-F	CGCAGCGCCTCGTGGACGAGGGCTTCGAAGAGTTCCAGCTCATTGGCTATGCGCTCGGCGGGCTGCTGGCCATGTTTAAACTTTGTTCAAAAAAAA	TE point mutation
TE-C1586S-F	CGCAGCGCCTCGTGGACGAGGGCTTCGAAGAGTTCCAGCTCATTGGCTATTCGCTCGGCGGGCTGCTGGCCATGTTTAAACTTTGTTCAAAAAAAA	
TE-C1586-R	CCGCGCTCGAGCAGACGGCGCGCCACCTCGATGGCCAGCAGCCCGCCGAGGTTTAAACTTTGTTTATTTTTCT	
srb-chk01-F	CCAAGTACAAGCAGCTCGTCGCA	colony PCR
srb-chk01-R	GCCGTCGTCGTCGATGAACATCA	
srb-chk02-F	CTACCGCATCCAAGTCGAAGACC	
srb-chk02-R	CGATGCAGGATCGTGAGGTTCAC	
srb-chk03-F	TGGCTTCGATTACCGCACGTTGG	
srb-chk03-R	GGCGTCCTTGGGTATCGAATACG	

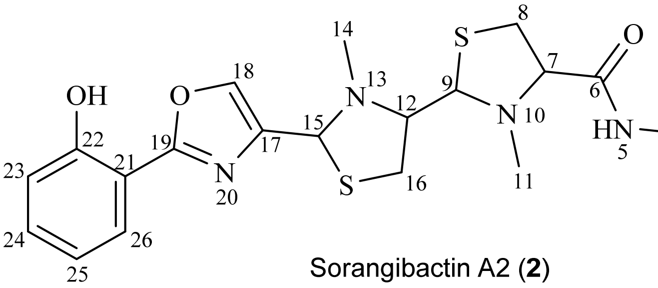
Table S5 NMR spectroscopic data of sorangibactin A1 (**1**) in methanol-d_4 at 700/175 MHz.



Sorangibactin A1 (**1**)

#	δ ^{13}C [ppm]	δ ^{1}H [ppm] mult (*J* [Hz])	δ ^{15}N [ppm]	COSY	HMBC	^{1}H-^{15}N HMBC
1	179.5	-	-	-	-	-
2	24.9	3.01, m	-	3	3,4	-
3	24.7	2.10,1.92, m	-	2,4	1,2,4	-
4	63.1	4.09, dd (6.94,3.66)	-	3	1,2,6	-
5	-	-	-	-	-	-
6	167.9	-	-	-	-	-
7	81.2	3.59, t (7.95)	-	8	6,8,9,11	-
8	35.3	3.15,3.02, m	-	7	6,7,9	10
9	76.0	4.21, m	-	12,16	11,16	-
10	-	-	56.7	-	-	-
11	40.8	2.47, s	-	-	7,9	10
12	75.7	3.29, m	-	9,16	9,16	-
13	-	-	59.8	-	-	-
14	41.9	2.57, s	-	-	12,15	13
15	70.1	5.01, s	-	18	12,14,16,17,18	20
16	33.8	3.29,3.16, m	-	9,12	14,15	13
17	144.5	-	-	-	-	-
18	136.4	7.89, s	-	15	17,19	20
19	162.9	-	-	-	-	-
20	-	-	239.9	-	-	-
21	112.4	-	-	-	-	-
22	158.5	-	-	-	-	-
23	118.2	7.00, d (8.38)	-	24	21,22,25	19,20
24	133.7	7.36, m	-	23,25	22,26	19,20
25	120.7	6.95, m	-	24,26	21,22,23,24,26	-
26	127.2	7.82, d (7.80)	-	25	19,22,24	-

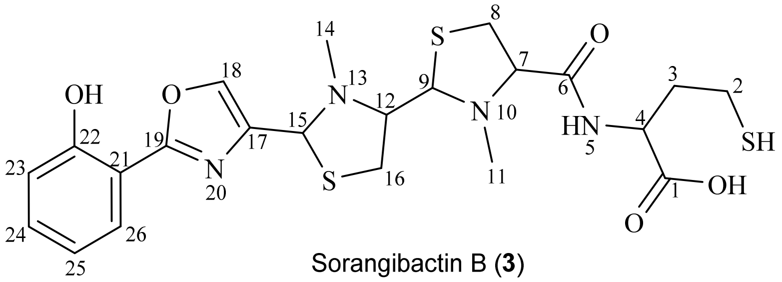
Table S6 NMR spectroscopic data of sorangibactin A2 (**2**) in methanol-d_4 at 700/175 MHz.



Sorangibactin A2 (**2**)

#	δ ^{13}C [ppm]	δ ^{1}H [ppm] mult (*J* [Hz])	δ ^{15}N [ppm]	COSY	HMBC	^{1}H-^{15}N HMBC
1	179.5	-	-	-	-	-
2	24.9	3.01, m	-	3	4	-
3	23.9	2.04, m	-	2,4	1,2,4	-
4	63.1	4.16, dd (6.79,4.29)	-	3	1,2,6	-
5	-	-	-	-	-	-
6	167.9	-	-	-	-	-
7	81.2	3.59, t (7.95)	-	8	6,8,9,11	-
8	35.3	3.15,3.02, m	-	7	6,7,9	10
9	76.0	4.21, m	-	12,16	11,16	-
10	-	-	56.7	-	-	-
11	40.8	2.47, s	-	-	7,9	10
12	75.7	3.29, m	-	9,16	9,16	-
13	-	-	59.8	-	-	-
14	41.9	2.57, s	-	-	12,15	13
15	70.1	5.01, s	-	18	12,14,16,17,18	20
16	33.8	3.29,3.16, m	-	9,12	14,15	13
17	144.5	-	-	-	-	-
18	136.4	7.89, s	-	15	17,19	20
19	162.9	-	-	-	-	-
20	-	-	239.9	-	-	-
21	112.4	-	-	-	-	-
22	158.5	-	-	-	-	-
23	118.2	7.00, d (8.38)	-	24	21,22,25	19,20
24	133.7	7.36, m	-	23,25	22,26	19,20
25	120.7	6.95, m	-	24,26	21,22,23,24,26	-
26	127.2	7.82, d (7.80)	-	25	19,22,24	-

Table S7 NMR spectroscopic data of sorangibactin B (**3**) in methanol-d_4 at 500/125 MHz.



Sorangibactin B (**3**)

#	δ ^{13}C [ppm]	δ ^{1}H [ppm] mult (*J* [Hz])	COSY	HMBC
1	176.6	-	-	-
2	35.2	2.27, t (7.47)	3	3,4
3	26.3	1.59, m	2,4	1
4	51.0	4.68, m	3	6
5	-	-	-	-
6	173.0	-	-	-
7	76.3	3.57, m	8	6
8	33.7	3.25,3.07, m	7	6,7
9	78.1	4.31, m	12	12,14
10	-	-	-	-
11	42.5	2.56, m	-	7,9
12	75.2	3.37, m	9,16	-
13	-	-	-	-
14	42.5	2.64, m	-	12,15
15	69.9	5.06, m	18	17,18
16	34.6	3.23,3.08, m	12	12
17	142.5	-	-	-
18	136.4	7.88, m	15	17,19
19	161.8	-	-	-
20	-	-	-	-
21	110.9	-	-	-
22	157.2	-	-	-
23	118.2	7.01, m	24	21,22
24	133.9	7.38, m	23,25	21,22
25	120.8	6.97, m	24,26	21,23
26	127.2	7.83, m	25	19,22,24

4.6.2 Supplementary figures

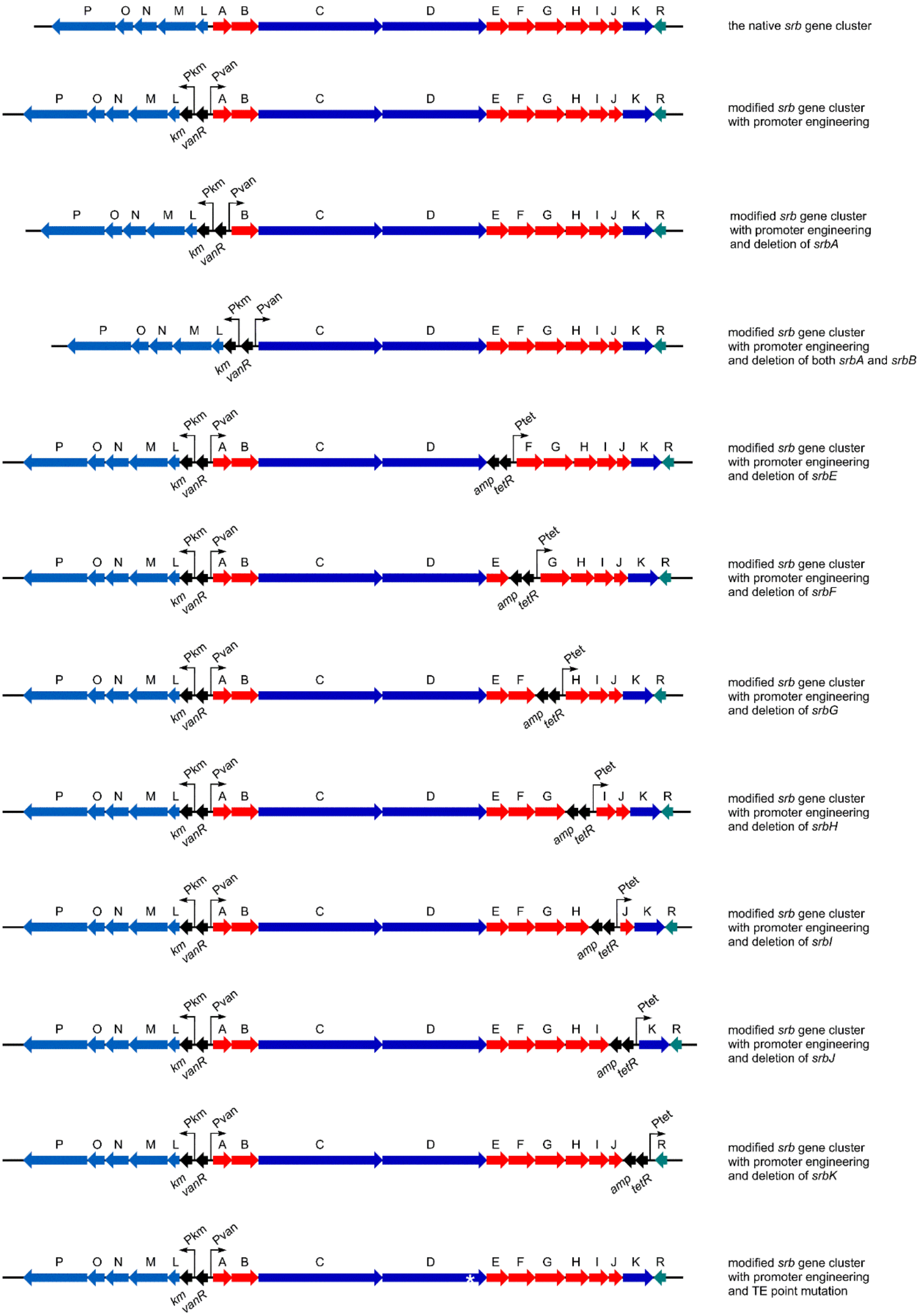


Figure S1 Expression constructs used in this study.

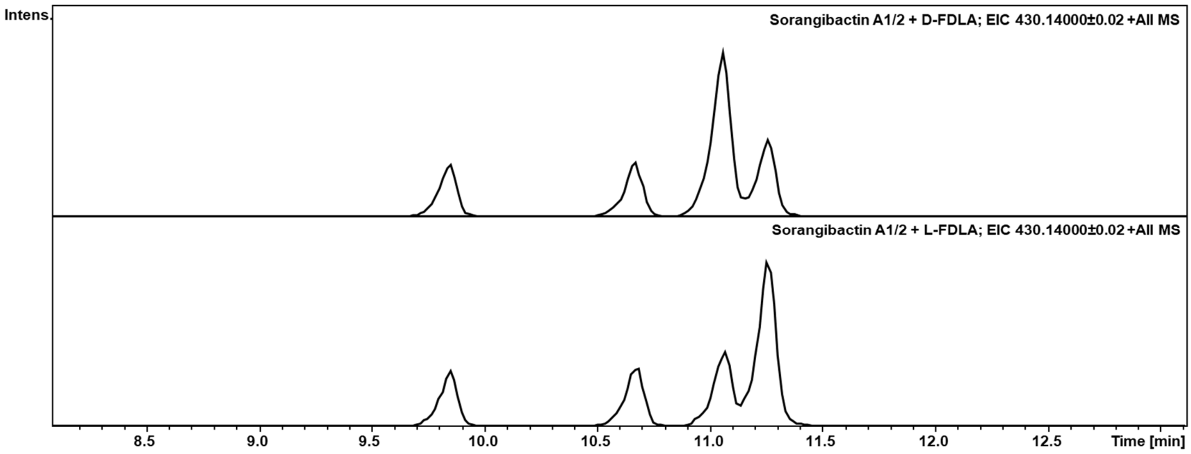


Figure S2 HPLC-MS measurements of Marfey's derivatization of sorangibactin A (**1** and **2**). Upper chromatogram: derivatized with D-FDLA; Lower chromatogram: derived with L-FDLA; The traces are shown as EIC 430.14 ± 0.02 *m/z* for derivatized *N*-methyl cysteine and homocysteine.

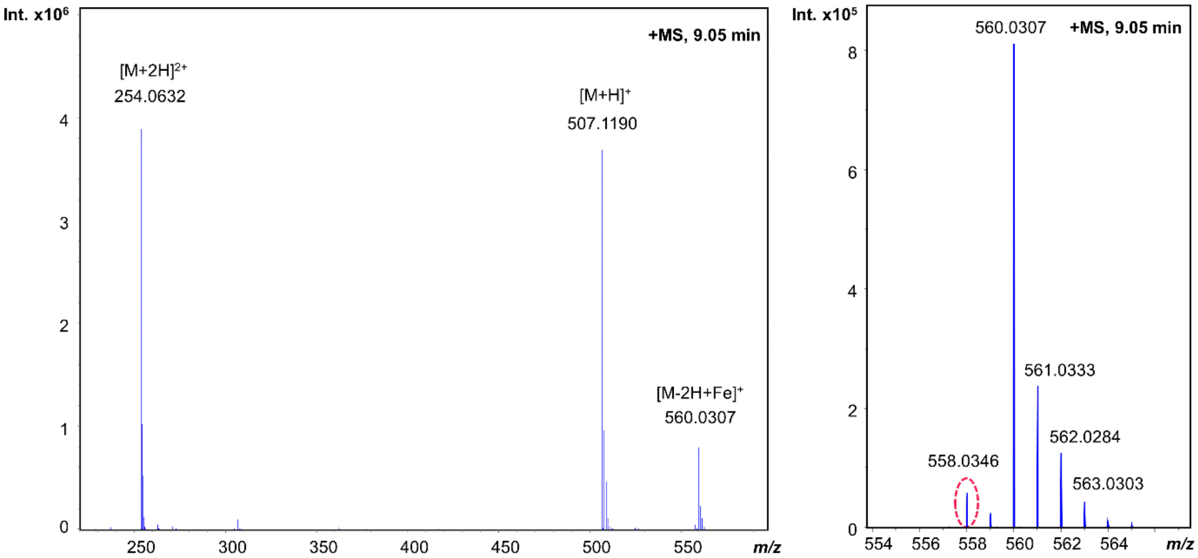


Figure S3 HRMS spectrum of sorangibactin A. UHR-TOF measurements of sorangibactin A containing extract showing the siderophore-like behavior by insource iron chelating: double charged species at *m/z* 254.0632, single charged species at *m/z* 507.1189 and [M-2H+Fe^{3+}]$^+$ species at *m/z* 560.0307 with a characteristic M-2 iron isotope peak at *m/z* 558.0346.

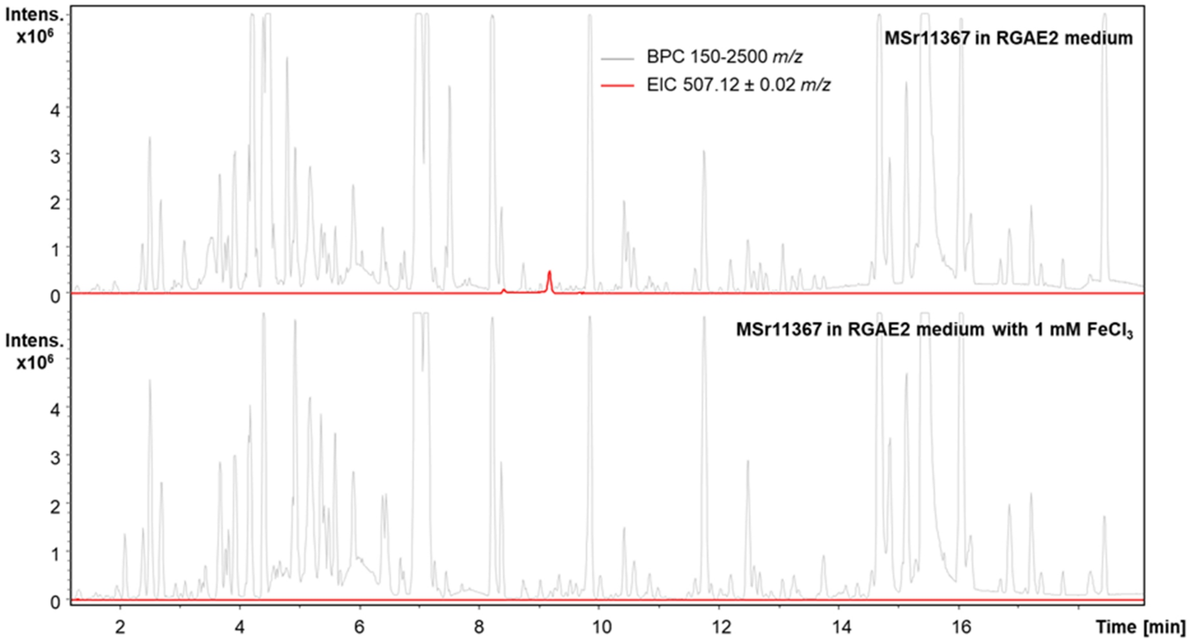


Figure S4 Production of sorangibactin A by MSr11367 with and without iron addition.

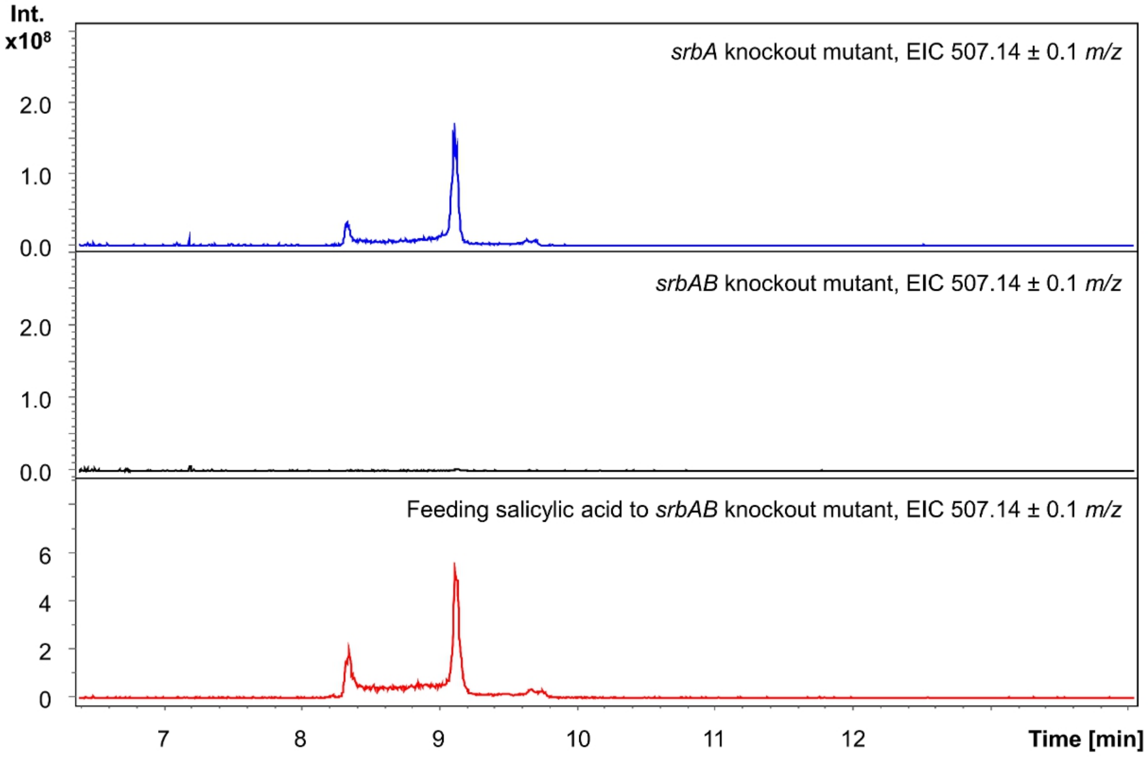


Figure S5 HPLC-MS measurements of *srbA*, *srbAB* knockout mutants and feeding salicylic acid to *srbAB* knockout mutant.

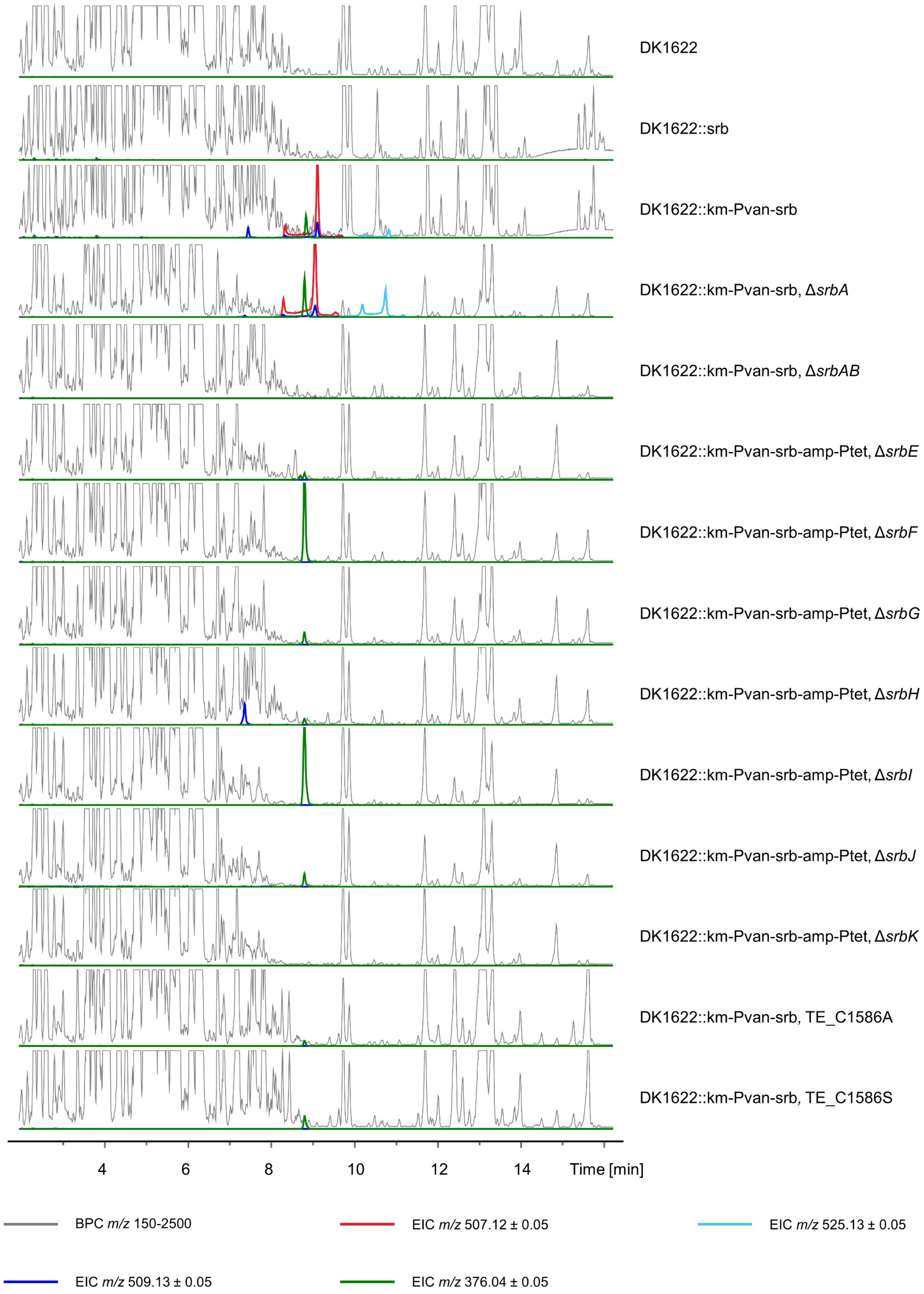


Figure S6 HPLC-MS analysis of crude extrudes from *M. xanthus* DK1622 and mutants containing the corresponding sorangibactin expression constructs.

Figure S7 Alignment of TE domains. The sorangibactin TE domain is aligned with coelibactin TE, yersiniabactin TE and 10 structure available TEs from PDB database. The catalytic motif (normally GxSxG) is indicated in red dashed box.

Figure S8 Point mutation of C1586A and C1586S of the sorangibactin TE domain.

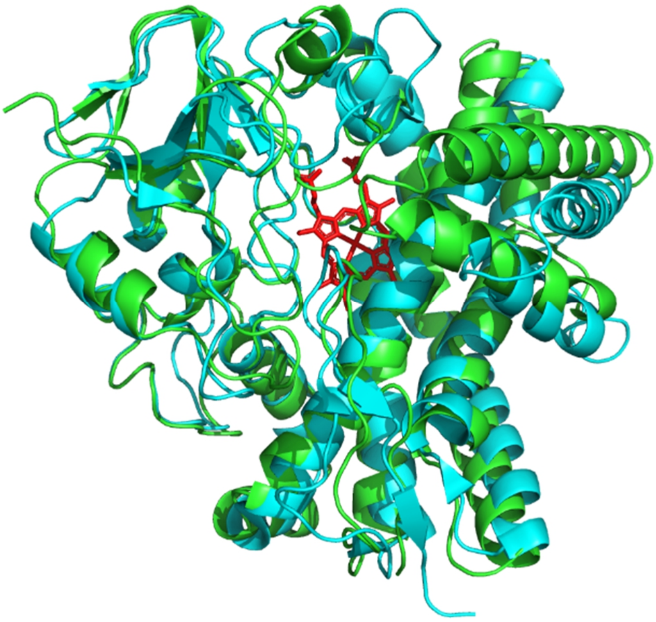


Figure S9 Superimposition of SrbH and CYP105N1. Green color: structure of SrbH predicted by ColabFold;[39] Cyan color: CYP105N1 from *Streptomyces coelicolor* A3(2) (PDB 3TYW); Heme is shown as red stick.

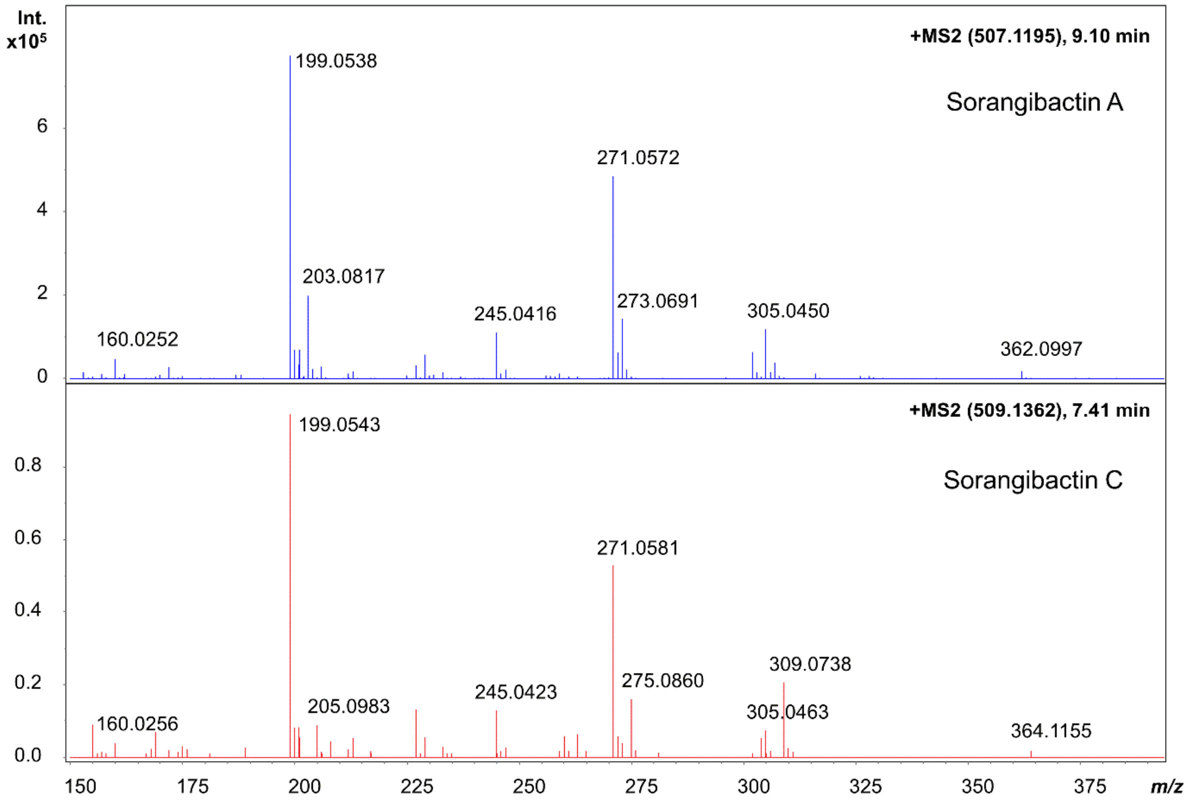
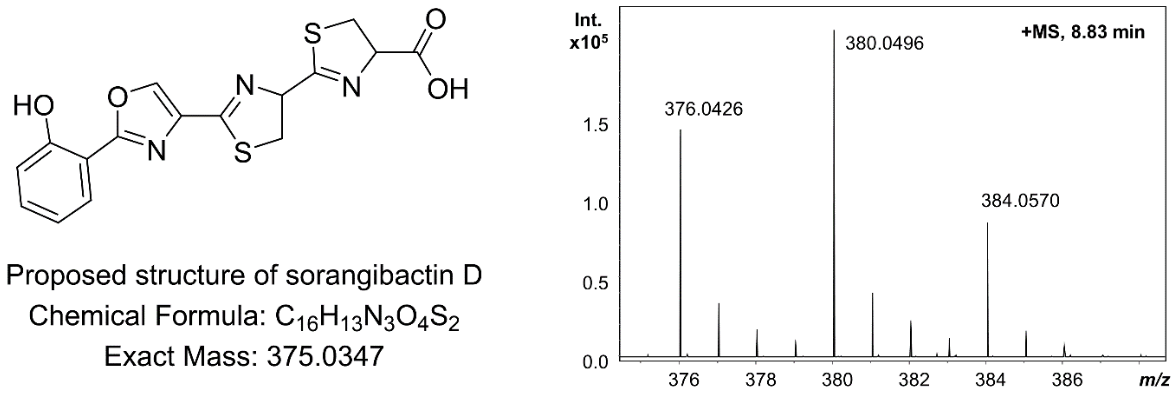


Figure S10 HRMS^2 of sorangibactin C in comparison with sorangibactin A.



Proposed structure of sorangibactin D
Chemical Formula: $C_{16}H_{13}N_{3}O_{4}S_{2}$
Exact Mass: 375.0347

Figure S11 Proposed structure and incorporation of L-Cys-$^{13}C_3$,^{15}N of sorangibactin D.

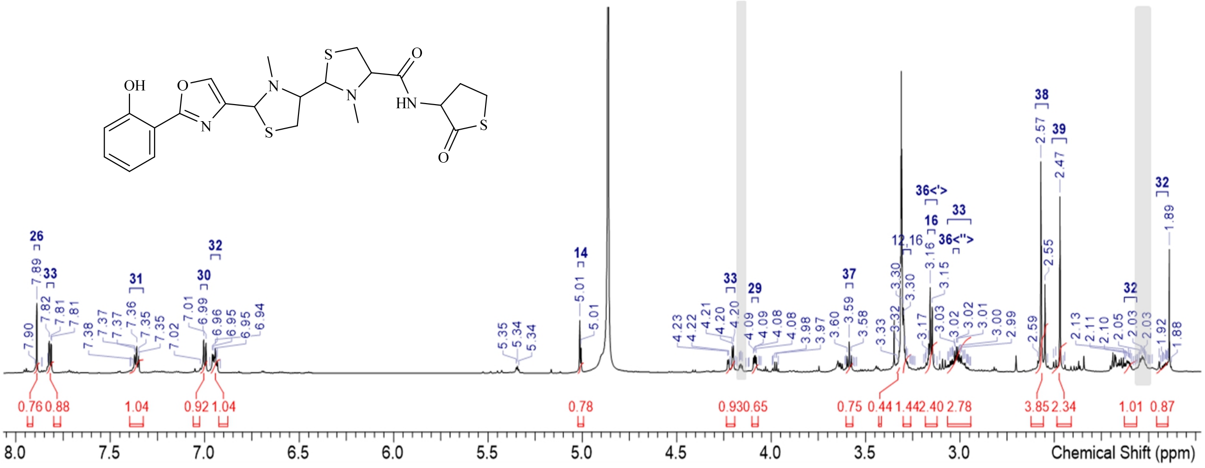


Figure S12 ^{1}H-spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 700 MHz. Signals corresponding to sorangibactin A2 (**2**) are highlighted in gray.

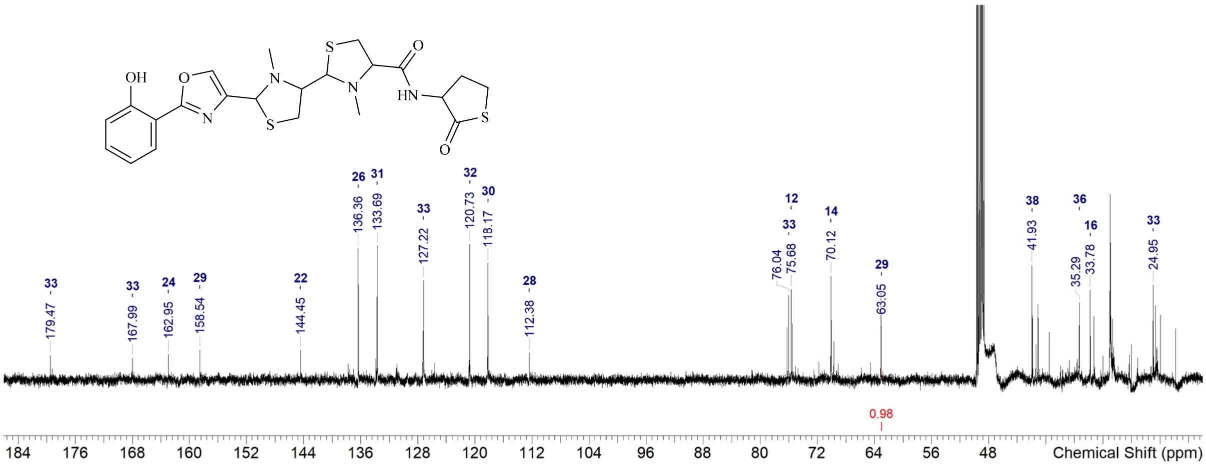


Figure S13 ^{13}C-spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 175 MHz.

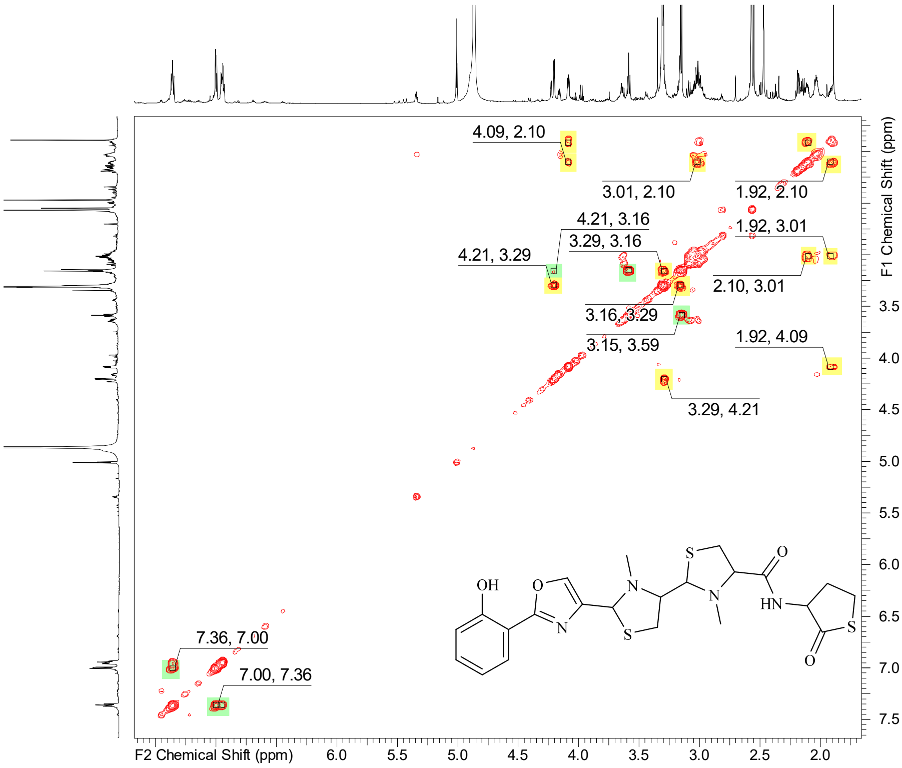


Figure S14 COSY spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 700 MHz.

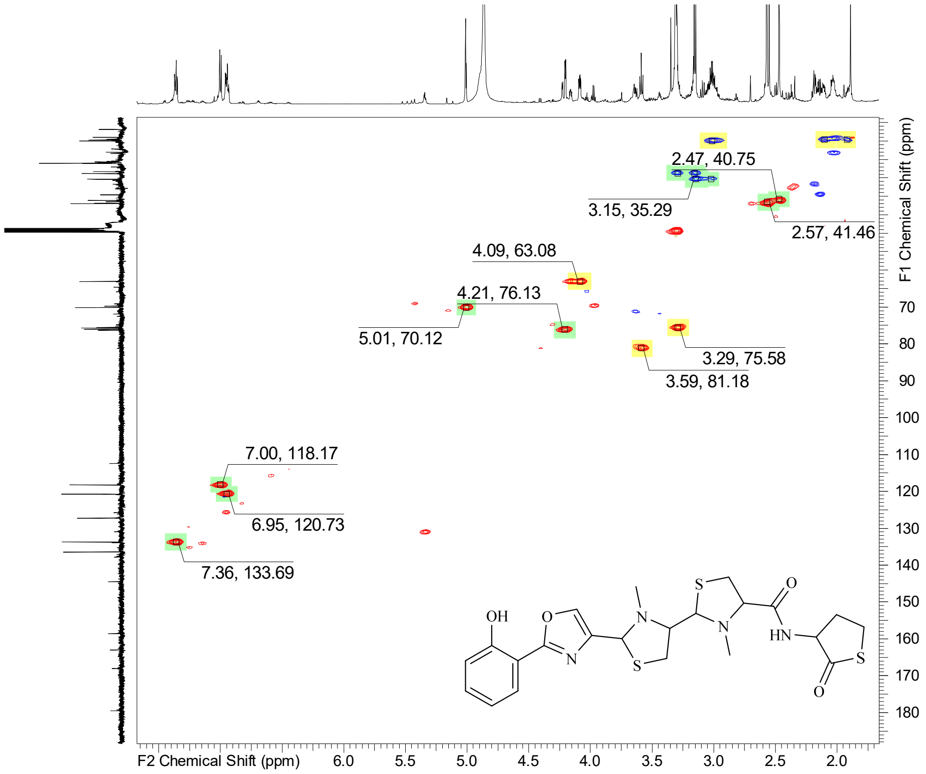


Figure S15 HSQC spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

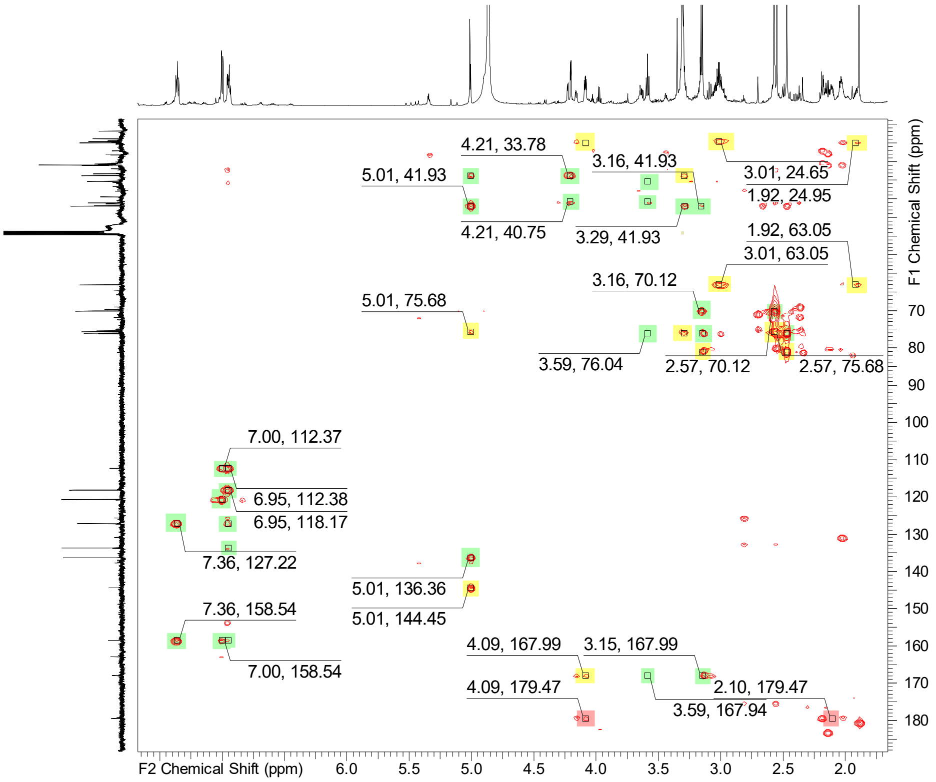


Figure S16 HMBC spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

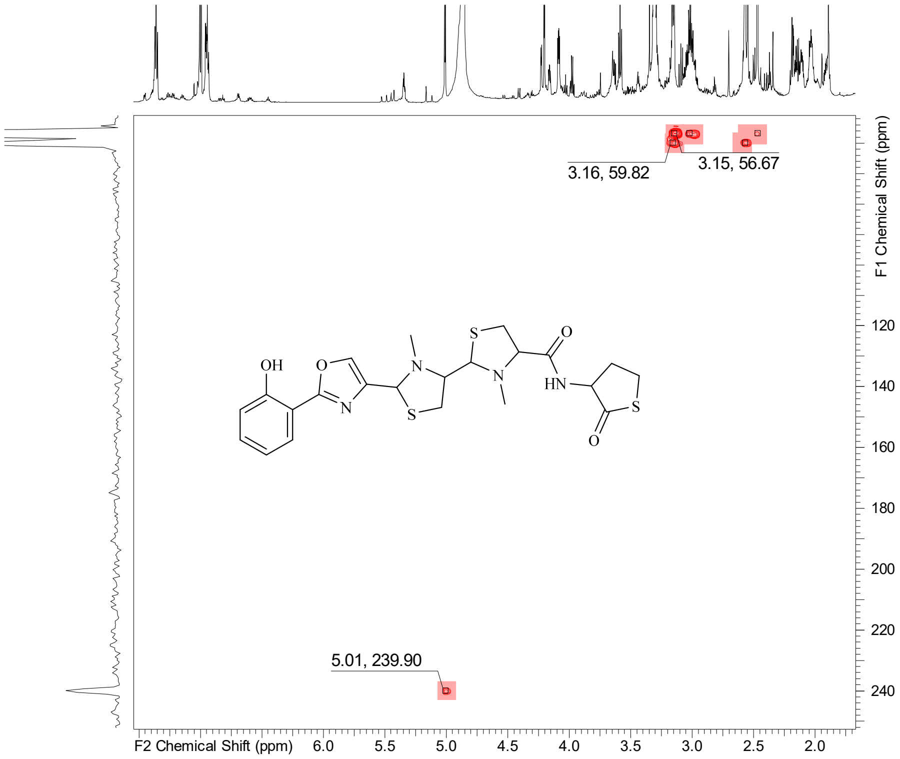


Figure S17 ^{1}H-^{15}N-HMBC spectrum of sorangibactin A1 (**1**) in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{15}N).

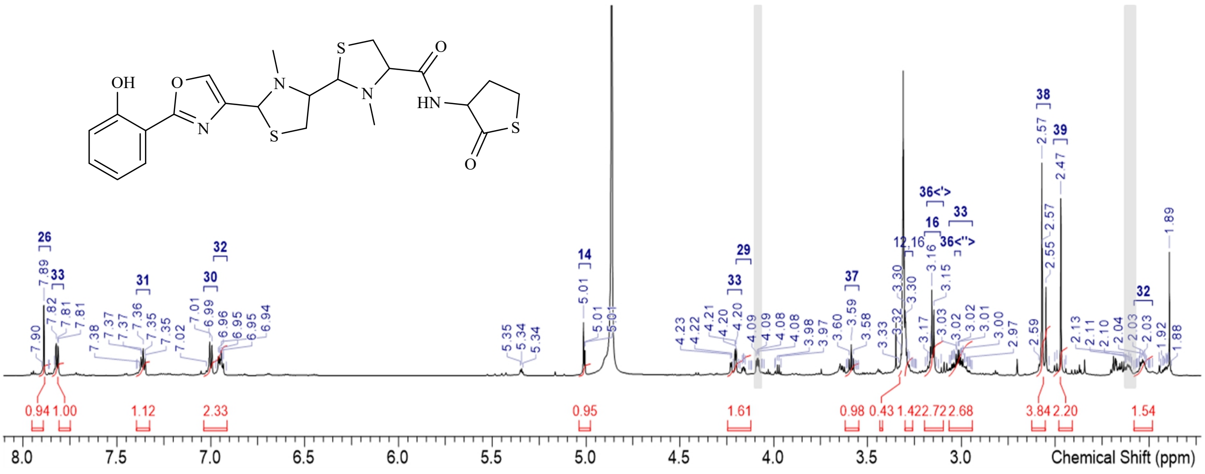


Figure S18 ^{1}H-spectrum of sorangibactin A2 (**2**) in methanol-d_4 at 700 MHz. Signals corresponding to sorangibactin A1 (**1**) are highlighted in gray.

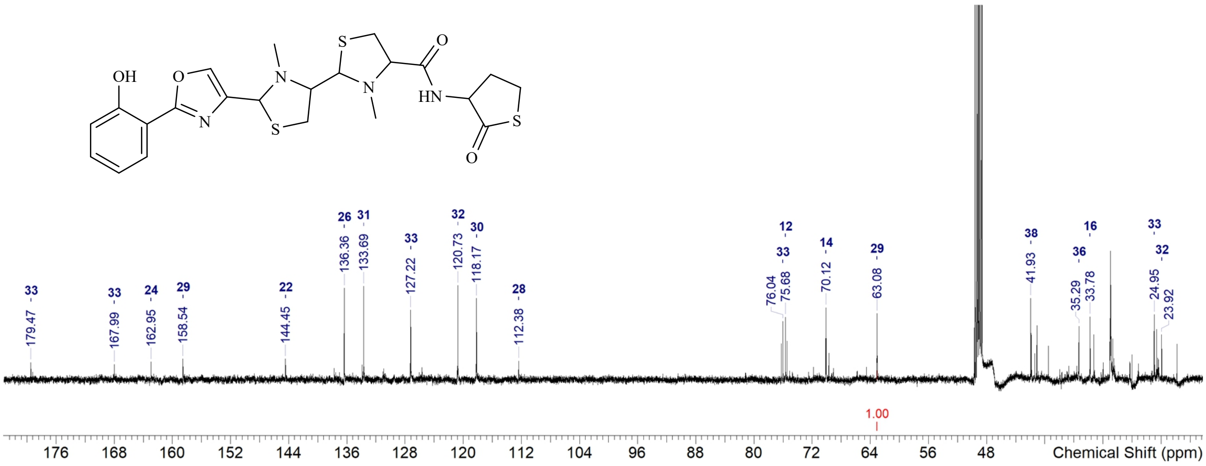


Figure S19 ^{13}C-spectrum of sorangibactin A2 (**2**) in methanol-d_4 at 175 MHz.

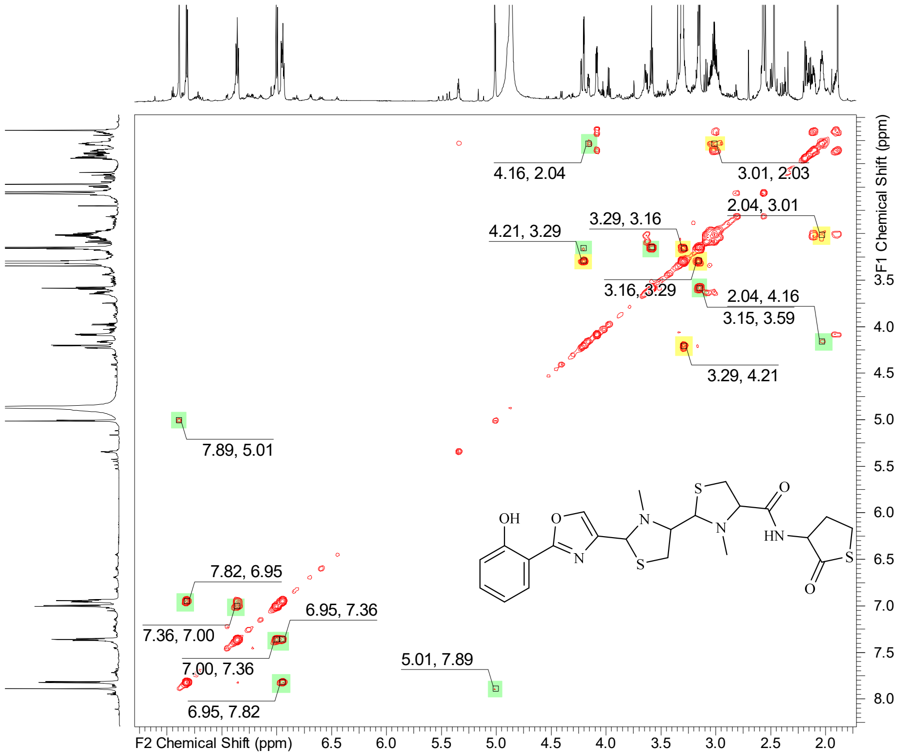


Figure S20 COSY spectrum of sorangibactin A2 (**2**) in methanol-*d*4 at 700 MHz.

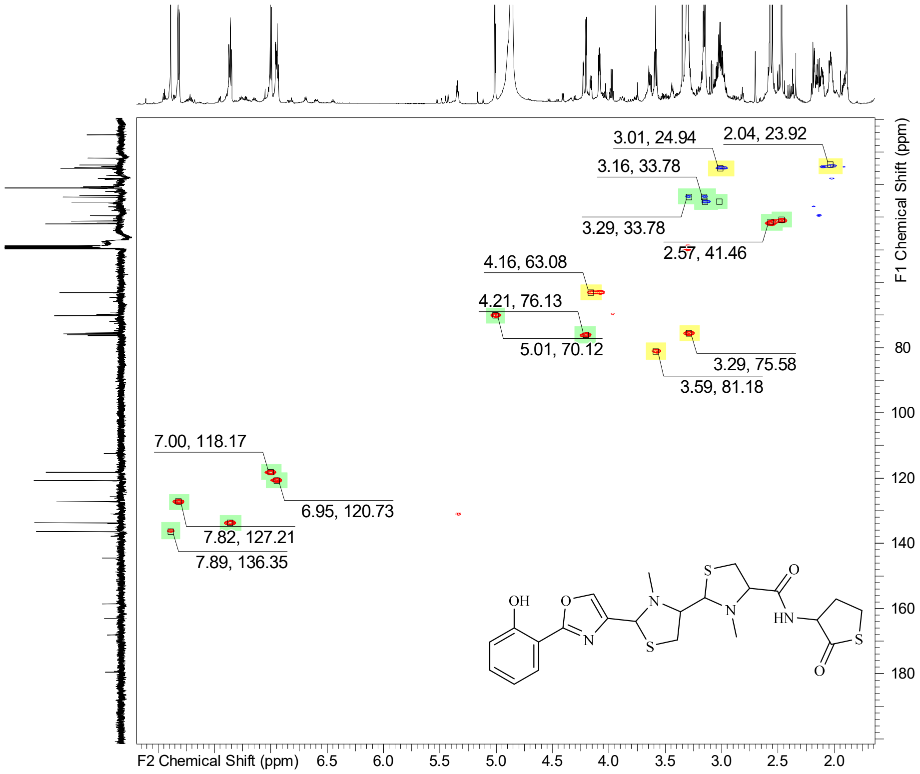


Figure S21 HSQC spectrum of sorangibactin A2 (**2**) in methanol-*d*4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

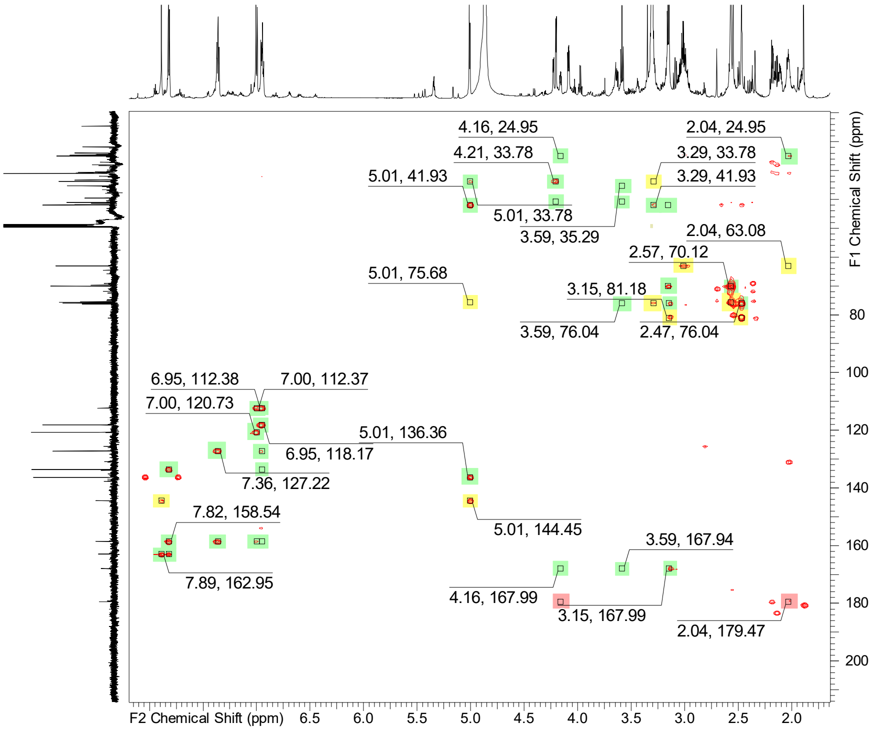


Figure S22 HMBC spectrum of sorangibactin A2 (**2**) in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

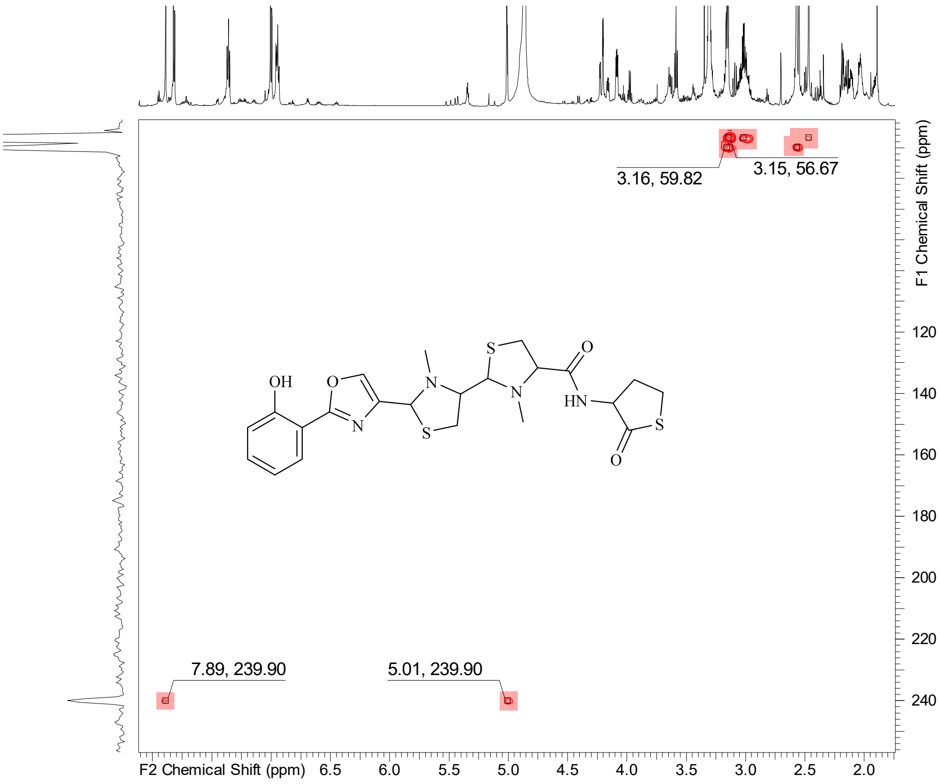


Figure S23 ^{1}H-^{15}N-HMBC spectrum of sorangibactin A2 (**2**) in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{15}N).

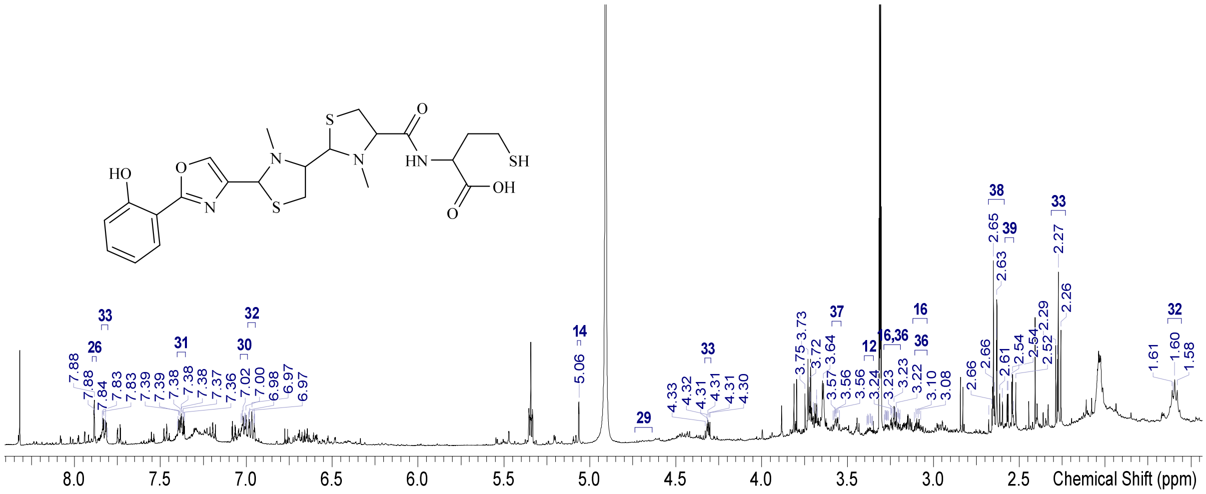


Figure S24 ^{1}H-spectrum of sorangibactin B (**3**) in methanol-*d*$_4$ at 500 MHz.

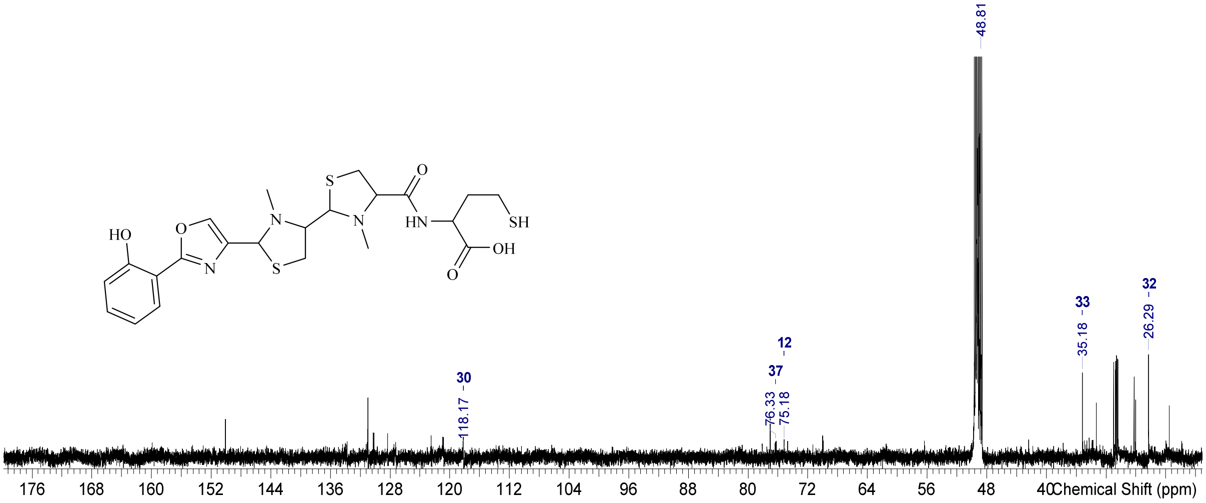


Figure S25 ^{13}C-spectrum of sorangibactin B (**3**) in methanol-*d*$_4$ at 125 MHz.

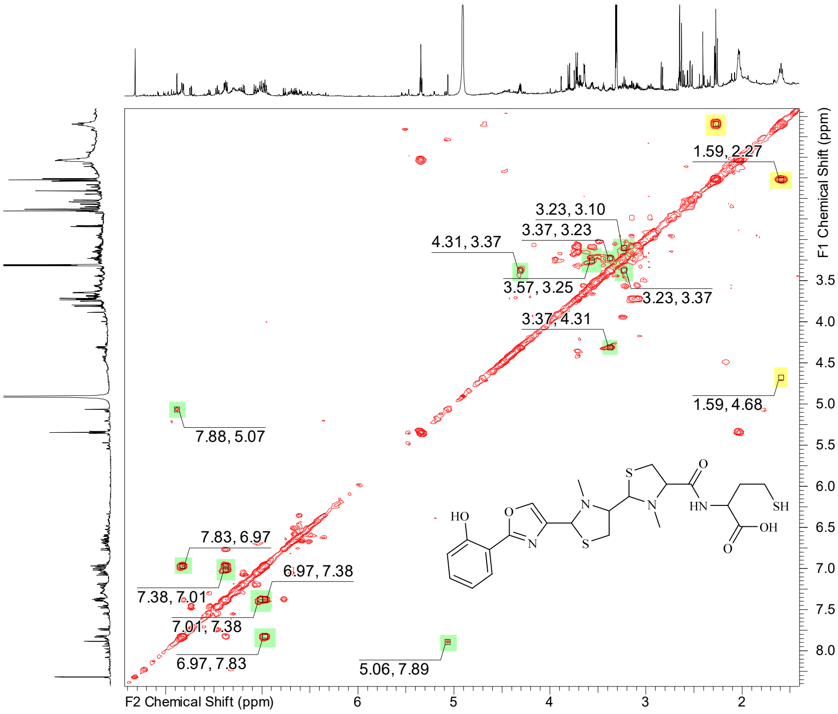


Figure S26 COSY spectrum of sorangibactin B (**3**) in methanol-d_4 at 500 MHz.

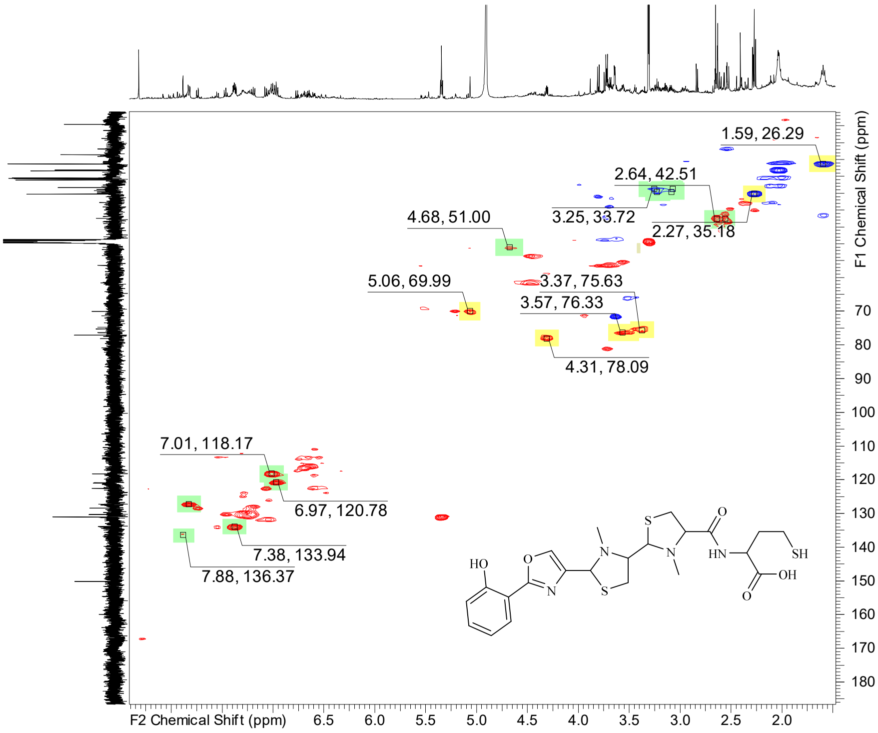


Figure S27 HSQC spectrum of sorangibactin B (**3**) in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C). Unpicked surrounding signals are most likely resulting from the presence of many different isomers in the same sample upon the acidic purification procedure required for the ring opened target compound of **3**.

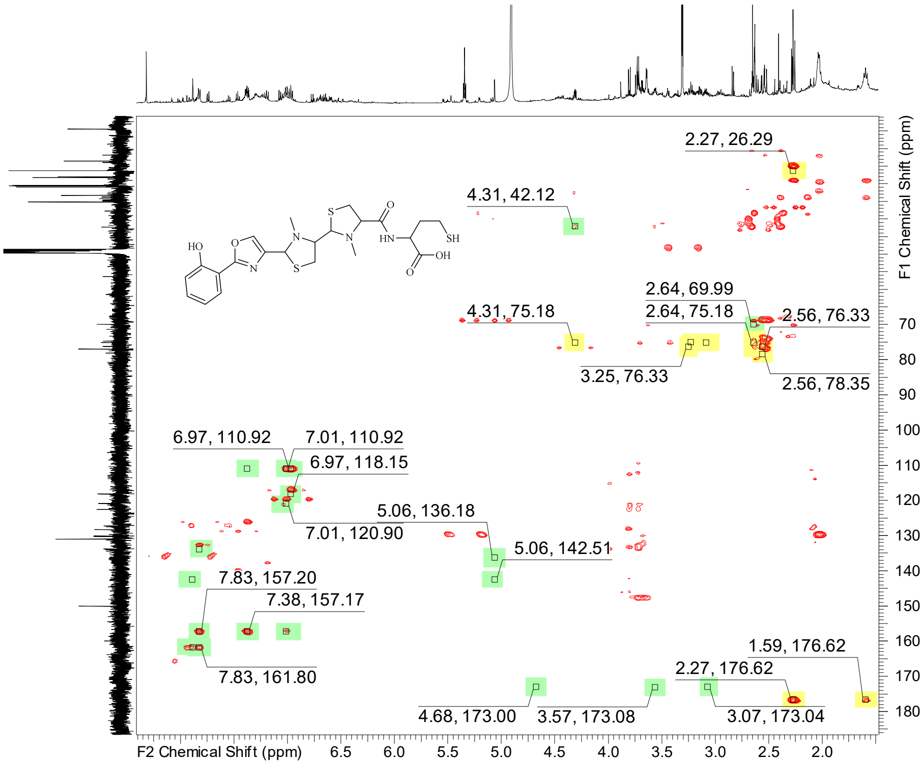


Figure S28 HMBC spectrum of sorangibactin B (**3**) in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).

4.7 Acknowledgements

The authors thank Dr. Ronald Garcia for providing the myxobacterial strain, Viktoria George and Alexandra Amann for bioactivity tests, Dr. Daniel Krug and Dr. Nicolas A. Frank for helpful discussion, and Dr. Fabian Panter for assistance with metabolome analysis and compound purification.

4.8 References

[1] K. Gerth, S. Pradella, O. Perlova et al., Myxobacteria: proficient producers of novel natural products with various biological activities—past and future biotechnological aspects with the focus on the genus *Sorangium*, *J Biotechnol*, **2003**, 106(2-3), 233-253.

[2] J. Herrmann, A. A. Fayad, R. Müller, Natural products from myxobacteria: novel metabolites and bioactivities, *Nat Prod Rep*, **2017**, 34(2), 135-160.

[3] K. D. Bauman, K. S. Butler, B. S. Moore et al., Genome mining methods to discover bioactive natural products, *Nat Prod Rep*, **2021**, 38(11), 2100-2129.

[4] F. Panter, D. Krug, S. Baumann et al., Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria, *Chem Sci*, **2018**, 9(21), 4898-4908.

[5] A. Gavriilidou, S. A. Kautsar, N. Zaburannyi et al., Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes, *Nat Microbiol*, **2022**, 7(5), 726-735.

[6] L. Huo, J. J. Hug, C. Fu et al., Heterologous expression of bacterial natural product biosynthetic pathways, *Nat Prod Rep*, **2019**, 36(10), 1412-1436.

[7] J. Clos and D. Zander-Dinse, Cosmid Library Construction and Functional Cloning, *Methods Mol Biol*, **2019**, 1971, 123-140.

[8] S. C. Wenzel and R. Müller, Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways, *Curr Opin Biotechnol*, **2005**, 16(6), 594-606.

[9] R. C. Hider and X. Kong, Chemistry and biology of siderophores, *Nat Prod Rep*, **2010**, 27(5), 637-657.

[10] B. Kunze, N. Bedorf, W. Kohl et al., Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacterales). Production, isolation, physico-chemical and biological properties, *J Antibiot (Tokyo)*, **1989**, 42(1), 14-17.

[11] S. Nadmid, A. Plaza, G. Lauro et al., Hyalachelins A-C, unusual siderophores isolated from the terrestrial myxobacterium *Hyalangium minutum*, *Org Lett*, **2014**, 16(16), 4130-4133.

[12] B. Kunze, W. Trowitzsch-Kienast, G. Höfle et al., Nannochelins A, B and C, new iron-chelating compounds from *Nannocystis exedens* (myxobacteria). Production, isolation, physico-chemical and biological properties, *J Antibiot (Tokyo)*, **1992**, 45(2), 147-150.

[13] W. Wang, Z. Qiu, H. Tan et al., Siderophore production by actinobacteria, *BioMetals*, **2014**, 27(4), 623-631.

[14] S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga et al., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, **2002**, 417(6885), 141-147.

[15] Timothy C. Johnstone and Elizabeth M. Nolan, Beyond iron: non-classical biological functions of bacterial siderophores, *Dalton Trans*, **2015**, 44(14), 6320-6339.

[16] K. Blin, S. Shaw, A. M. Kloosterman et al., antiSMASH 6.0: improving cluster detection and comparison capabilities, *Nucleic Acids Res*, **2021**, 49(W1), W29-W35.

[17] Y.-J. Yang, R. P. Singh, X. Lan et al., Genome Editing in Model Strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System, *Biomolecules*, **2018**, 8(4), 137.

[18] J.-P. Jang, M. C. Kwon, T. Nogawa et al., Thiolactomide: A New Homocysteine Thiolactone Derivative from *Streptomyces* sp. with Neuroprotective Activity, *J. Microbiol. Biotechnol.*, **2021**, 31(12), 1667-1671.

[19] Hubert Haag, Klaus Hantke, Hartmut Drechsel et al., Purification of yersiniabactin: a siderophore and possible virulence factor of *Yersinia enterocolitica*, *Microbiology*, **1993**, 139(9), 2159-2165.

[20] M. C. Miller, S. Parkin, J. D. Fetherston et al., Crystal structure of ferric-yersiniabactin, a virulence factor of *Yersinia pestis*, *J Inorg Biochem*, **2006**, 100(9), 1495-1500.

[21] T. H. Fife, R. Natarajan, C. C. Shen et al., Mechanism of thiazolidine hydrolysis. Ring opening and hydrolysis of 1,3-thiazolidine derivatives of *p*-(dimethylamino)cinnamaldehyde, *J Am Chem Soc*, **1991**, 113(8), 3071-3079.

[22] M. Schulz, V. Gaitanoglou, O. Mantel et al., Metabolomics Study on Pathogenic and Non-pathogenic *E. coli* with Closely Related Genomes with a Focus on Yersiniabactin and Its Known and Novel Derivatives, *Metabolites*, **2020**, 10(6), 221.

[23] R. Hermenau, J. L. Mehl, K. Ishida et al., Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria, *Angew Chem Int Ed*, **2019**, 58(37), 13024-13029.

[24] J. B. Neilands, Siderophores: structure and function of microbial iron transport compounds, *J Biol Chem*, **1995**, 270(45), 26723-26726.

[25] S. W. Wilhelm and C. G. Trick, Iron-limited growth of cyanobacteria: Multiple siderophore production is a common response, *Limnol Oceanogr*, **1994**, 39(8), 1979.

[26] Robert D. Perry, Paul B. Balbo, Heather A. Jones et al., Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation, *Microbiology*, **1999**, 145(5), 1181-1190.

[27] O. Kerbarh, A. Ciulli, N. I. Howard et al., Salicylate biosynthesis: Overexpression, purification, and characterization of Irp9, a bifunctional salicylate synthase from *Yersinia enterocolitica*, *J Bacteriol*, **2005**, 187(15), 5061-5066.

[28] R. F. Little and C. Hertweck, Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis, *Nat Prod Rep*, **2022**, 39(1), 163-205.

[29] T. Thongkongkaew, W. Ding, E. Bratovanov et al., Two Types of Threonine-Tagged Lipopeptides Synergize in Host Colonization by Pathogenic *Burkholderia* Species, *ACS Chem Biol*, **2018**, 13(5), 1370-1379.

[30] Y.-R. Lim, M.-K. Hong, J.-K. Kim et al., Crystal structure of cytochrome P450 CYP105N1 from *Streptomyces coelicolor*, an oxidase in the coelibactin siderophore biosynthetic pathway, *Arch Biochem Biophys*, **2012**, 528(2), 111-117.

[31] B. Zhao, S. C. Moody, R. C. Hider et al., Structural analysis of cytochrome P450 105N1 involved in the biosynthesis of the zincophore, coelibactin, *Int J Mol Sci*, **2012**, 13(7), 8500-8513.

[32] K. M. Meneely and A. L. Lamb, Two structures of a thiazolinyl imine reductase from *Yersinia enterocolitica* provide insight into catalysis and binding to the nonribosomal peptide synthetase module of HMWP1, *Biochemistry*, **2012**, 51(44), 9002-9013.

[33] K. M. Meneely, T. A. Ronnebaum, A. P. Riley et al., Holo Structure and Steady State Kinetics of the Thiazolinyl Imine Reductases for Siderophore Biosynthesis, *Biochemistry*, **2016**, 55(38), 5423-5433.

[34] P. He and G. R. Moran, Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily, *J Inorg Biochem*, **2011**, 105(10), 1259-1272.

[35] C. Song, J. Luan, R. Li et al., RedEx: a method for seamless DNA insertion and deletion in large multimodular polyketide synthase gene clusters, *Nucleic Acids Res*, **2020**, 48(22), e130.

[36] O. Perlova, J. Fu, S. Kuhlmann et al., Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*, *Appl Environ Microbiol*, **2006**, 72(12), 7485-7494.

[37] V. Magrini, C. Creighton, P. Youderian, Site-specific recombination of temperate *Myxococcus xanthus* phage Mx8: genetic elements required for integration, *J Bacteriol*, **1999**, 181(13), 4050-4061.

[38] P. A. Haack, K. Harmrolfs, C. D. Bader et al., Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus Inhibitors, *Angew Chem Int Ed*, **2022**, 61(52), e202212946.

[39] M. Mirdita, K. Schütze, Y. Moriwaki et al., ColabFold: making protein folding accessible to all, *Nat Methods*, **2022**, 19(6), 679-682.

Chapter 5. Final discussion

5.1 Genome mining through heterologous expression of myxobacterial gene clusters

5.1.1 Opportunities and challenges of genome mining

In the genomic era, advancements in genome sequencing and gene cluster annotation have greatly facilitated the process of genome mining, allowing subsequent prioritization of promising gene clusters. These valuable gene cluster resources represent a promising starting point for the discovery of novel (bio)chemistry.

Gene clusters of interest can be identified by using the BLAST algorithm to search for conserved gene(s) (see Chapter 1). We were able to identify a putative enediyne gene cluster (termed Endi8315) from the *Cystobacterineae* strain MCy8315 by searching for the enediyne warhead gene cassette in our in-house myxobacterial genome database (Figure 1, Figure 2, and Table 1), despite the fact that myxobacteria were not previously known to produce the enediyne family of antitumor antibiotics.

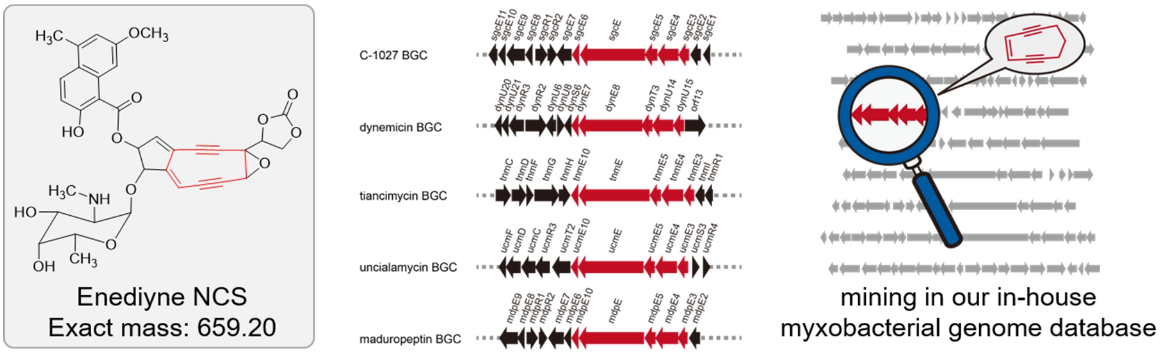


Figure 1 Scheme of targeted genome mining to identify a myxobacterial enediyne gene cluster.

By conducting a search for pentapeptide repeat protein (PRP) genes in proximity to or within a gene cluster, we prioritized a PRP-related gene cluster (termed PRP9280, Figure 2 and Table 1) with promising potential in the production of novel topoisomerase inhibitors.[1] By searching for genes encoding tRNA-dependent cyclodipeptide synthases (CDPS) responsible for diketopiperazine formation, we identified a CDPS gene cluster (termed CDPS11579, Figure 2 and Table 1) from the *Nannocystineae* strain MNa11579 that resembles the BGC of bicyclomycin,[2,3] an clinically useful antibiotic. By searching for gene homologs of *polH* encoding for a radical *S*-adenosyl methionine (SAM) enzyme that is conserved in the biosynthetic pathways of *C*-nucleoside antibiotics,[4] we identified two types of nucleoside gene clusters (termed Nuc1182 and Nuc4204, Figure 2 and Table 1) from *Sorangiineae* strains. By searching for LxmK homologs,[5] we identified a RiPP gene cluster (termed RiPP1654, Figure 2 and Table 1) belonging to the recently discovered class V lanthipeptide BGC family from the *S. cellulosum* strain So ce1654.

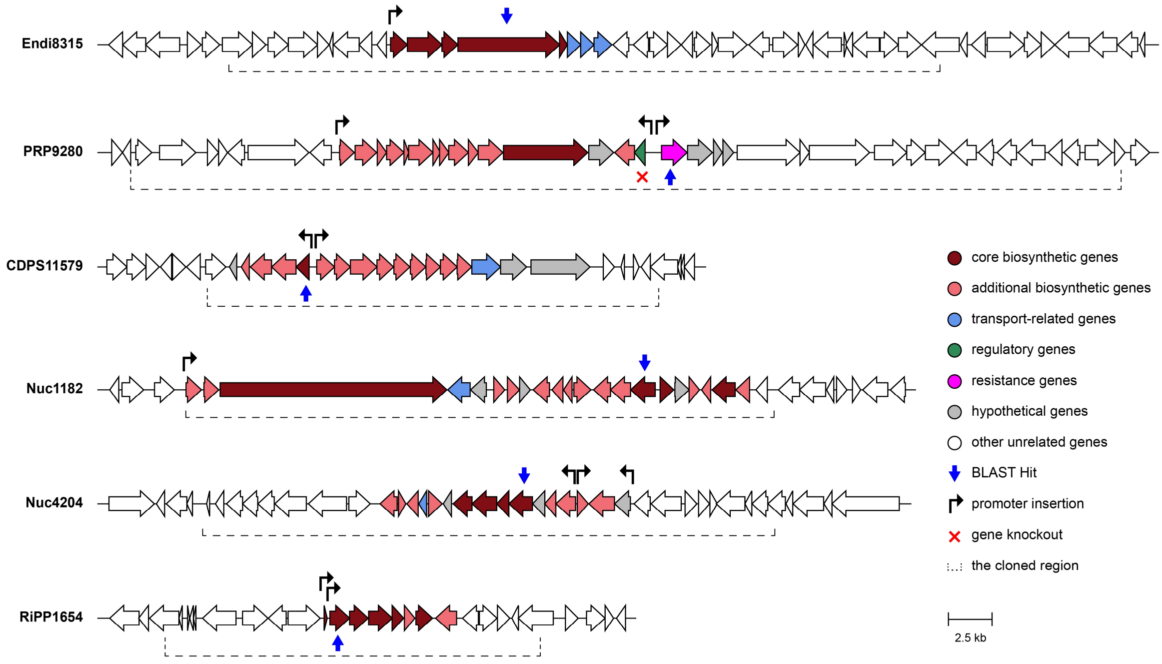


Figure 2 Map of identified gene clusters through targeted genome mining (No. 1−6 in Table 1).

The identification of biosynthetic pathways for known natural products can be accomplished through insertional mutagenesis, retrobiosynthesis and comparative analysis. Using these approaches, the disorazole A and disorazole Z biosynthetic gene clusters were identified.[6-8] By conducting a search for the conserved hypothetical protein DisE, we were able to discover additional 14 disorazole biosynthetic gene clusters. These clusters can be classified into two distinct categories based on phylogenetic analysis of their DisE protein sequences, with an exception of the one from *A. rufus* MSr9331 (SBSr003) (Figure 3 and Chapter 3).

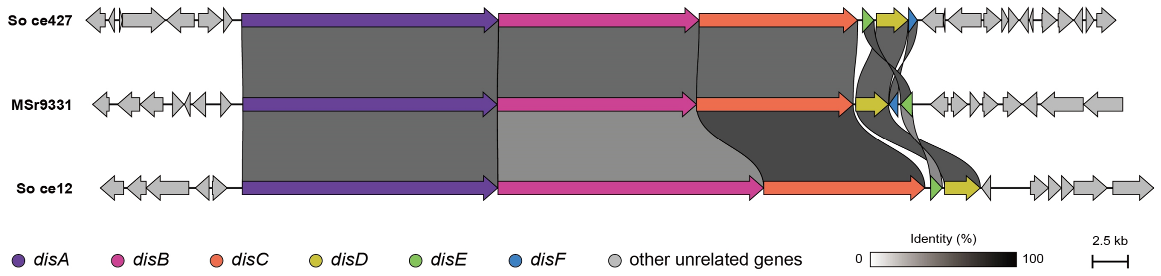


Figure 3 Map and comparison of three representative disorazole gene clusters (No. 7−9 in Table 1).

When prioritizing biosynthetic gene clusters for heterologous expression, in addition to considering the similarity to known BGCs, other factors also need to be taken into account, such as the integrity/size/boundary of the gene cluster, the gene/operon organization and tailoring genes. Based on these rules, two NRPS/PKS gene clusters were prioritized for direct cloning from the above mentioned myxobacterium MNa11579 (No. 10−11 in Figure 4). With the accessibility of cosmid library for *Sorangiineae* strain MSr11367 and *Corallococcus coralloides* MCy6431, high-through screening of uncharacterized gene clusters becomes possible. Thus, additional 9 cryptic gene clusters (No. 12−20 in Figure 4) were prioritized following the rules mentioned above for heterologous expression.

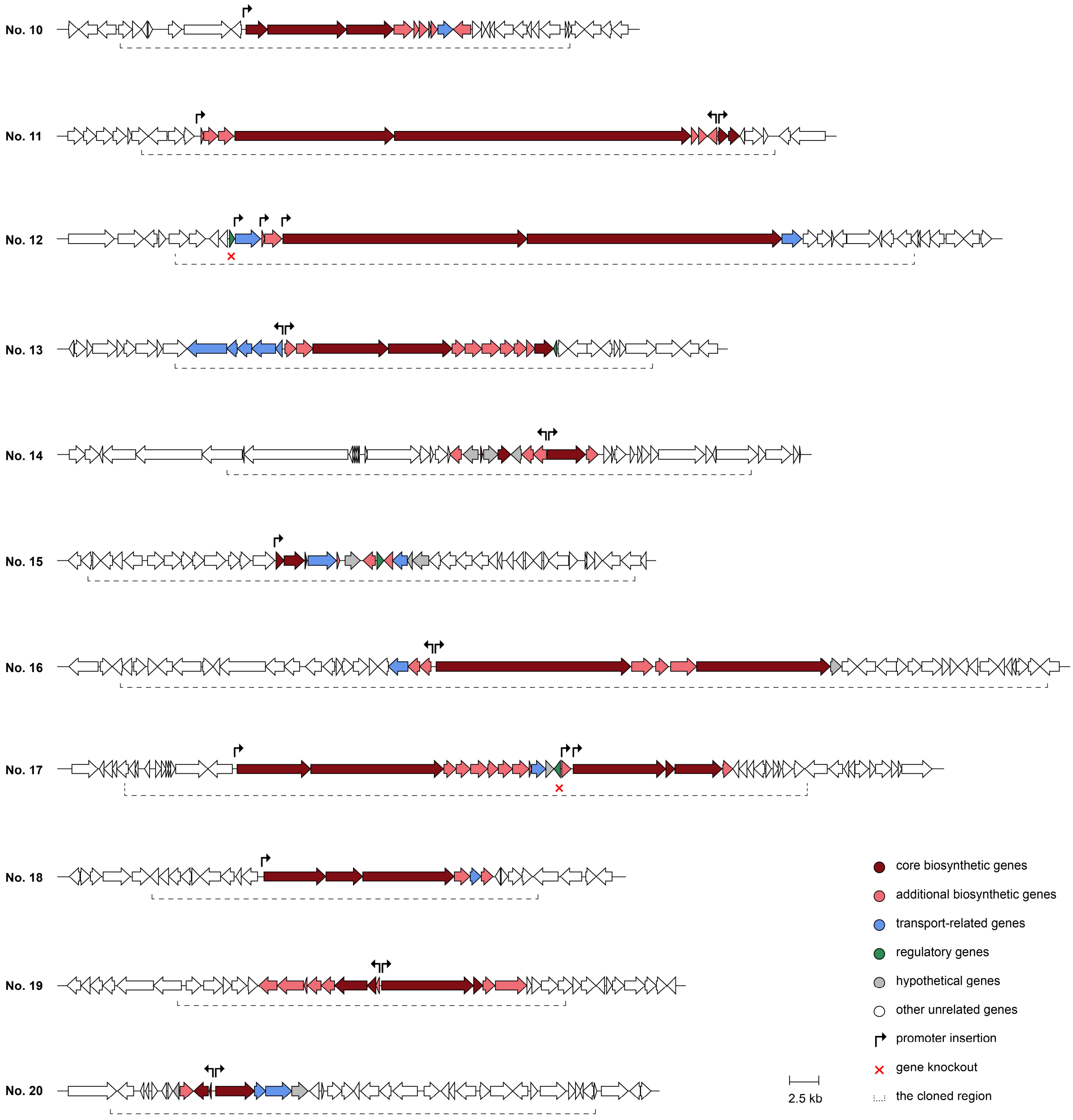


Figure 4 Map of cryptic gene clusters prioritized for heterologous expression (No. 10−20 in Table 1).

Table 1 Overview of myxobacterial gene clusters characterized by heterologous expression in *M. xanthus* DK1622 in this study.

No.	Type	Strain	Size	Cloning and modification(s) strategies	Remarks & Results
1	enediyne	*Cystobacterineae* strain MCy8315 (CcG34)	40 kb	direct cloning, promoter engineering	This BGC contains a conserved warhead gene cassette. No heterologous products were observed.
2	PRP related (NRPS)	*Myxococcus fulvus* MCy9280 (SBMx132)	57 kb	direct cloning, promoter engineering, knockout of regulator	This BGC is partially similar to ketomemicin and amipurimycin gene clusters.[9,10] No heterologous products were observed.
3	diketopiperazine	*Nannocystineae* strain MNa11579 (SBNc019)	26 kb	direct cloning, promoter engineering	This BGC is similar to the bicyclomycin CDPS gene cluster.[2] No heterologous products were observed.
4	nucleoside	*Sorangium cellulosum* So ce1182	34 kb	direct cloning, promoter engineering	This BGC contains *polH*-like, *polJ*-like, and *polA*-like genes,[4] and NRPS genes. No heterologous products were observed.
5	nucleoside	*Byssovorax cruenta* MSr4204 (Byc1)	33 kb	direct cloning, promoter engineering	This BGC contains *polH*-like, *polJ*-like, and *polA*-like genes,[4] and a *pumJ*-like gene.[11] No heterologous products were observed.
6	RiPP	*Sorangium cellulosum* So ce1654	8 kb	direct cloning, promoter engineering	A putative class V lanthipeptide BGC.[5] No heterologous products were observed.
7	disorazole (*trans*-AT PKS/NRPS)	*Sorangium cellulosum* So ce427	48 kb	direct cloning, promoter engineering, gene swapping, gene knockout, point mutations	Successful expression can be achieved using the native gene cluster. Intermediates and derivatives were produced and characterized.
8	disorazole (*trans*-AT PKS/NRPS)	*Aetherobacter rufus* MSr9331 (SBSr003)	47 kb	direct cloning, promoter engineering	*A. rufus* MSr9331 was not known to produce disorazole. No heterologous products were observed.
9	disorazole (*trans*-AT PKS/NRPS)	*Sorangium cellulosum* So ce12	52 kb	subcloning, gene swapping, gene knockout	Heterologous expression was achieved.[12] Intermediates and derivatives were produced and characterized in this study.
10	PKS/NRPS	*Nannocystineae* strain MNa11579 (SBNc019)	39 kb	direct cloning, promoter engineering	No heterologous products were observed.
11	*trans*-AT PKS/NRPS	*Nannocystineae* strain MNa11579 (SBNc019)	54 kb	direct cloning, promoter engineering	No heterologous products were observed.
12	NRPS	*Sorangiineae* strain MSr11367 (SBSr044)	53 kb	stitching two cosmids, promoter engineering, knockout of regulator	This BGC contains diazeniumdiolate biosynthetic genes.[13,14] No heterologous products were observed.
13	NRPS	*Sorangiineae* strain MSr11367 (SBSr044)	32 kb	cosmid library, promoter engineering, gene knockout, point mutations	Successful expression can be achieved using the promoter-engineered gene cluster, resulting in the discovery of sorangibactins,[15] which were also found to be produced by the original producer strain.
14	lanthipeptide	*Sorangiineae* strain MSr11367 (SBSr044)	10 kb	cosmid library, promoter engineering	No heterologous products were observed.
15	lasso peptide	*Sorangiineae* strain MSr11367 (SBSr044)	5−15 kb	cosmid library, promoter engineering	No heterologous products were observed.
16	PKS/NRPS	*Sorangiineae* strain MSr11367 (SBSr044)	52 kb	stitching two cosmids, promoter engineering	A homologous BGC was found in *S. cellulosum* So ce26. No heterologous products were observed.

No.	Type	Strain	Size	Cloning and modification(s) strategies	Remarks & Results
17	NRPS	*Sorangiineae* strain MSr11367 (SBSr044)	48 kb	stitching two cosmids, promoter engineering, knockout of regulator	Proposed shunt products (Figure 5A) were produced and purified.
18	PKS/NRPS	*Sorangiineae* strain MSr11367 (SBSr044)	25 kb	cosmid library, promoter engineering	No heterologous products were observed.
19	PKS/NRPS	*Corallococcus coralloides* MCy6431 (Ccc1071)	30 kb	cosmid library, promoter engineering	Similar to the marineosin gene cluster.[16] Proposed shunt products (Figure 5B−C) were produced and purified.
20	lanthipeptide	*Corallococcus coralloides* MCy6431 (Ccc1071)	10 kb	cosmid library, promoter engineering	No heterologous products were observed.

As mentioned in the previous chapters, heterologous expression represents a useful approach for functional characterization of secondary metabolite biosynthetic gene clusters, allowing for the discovery of novel natural products and investigation of their biosynthetic pathways. Moreover, with advancements in molecular cloning and engineering, manipulation and heterologous expression of large biosynthetic gene clusters are no longer decisive bottlenecks. However, one of the biggest challenges remains the low success rate of obtaining the desired products after heterologous expression of cryptic biosynthetic gene clusters. Despite extensive efforts, including the utilization of different hosts, promoter engineering, regulator manipulation, medium screening, and various combinations of these strategies, successful heterologous expression of cryptic gene clusters still remains a challenging and often tedious trial-and-error process. As shown in Table 1, for most of gene clusters, we were unable to identify their corresponding products after heterologous expression of the promoter-engineered gene clusters in the host *M. xanthus* DK1622.

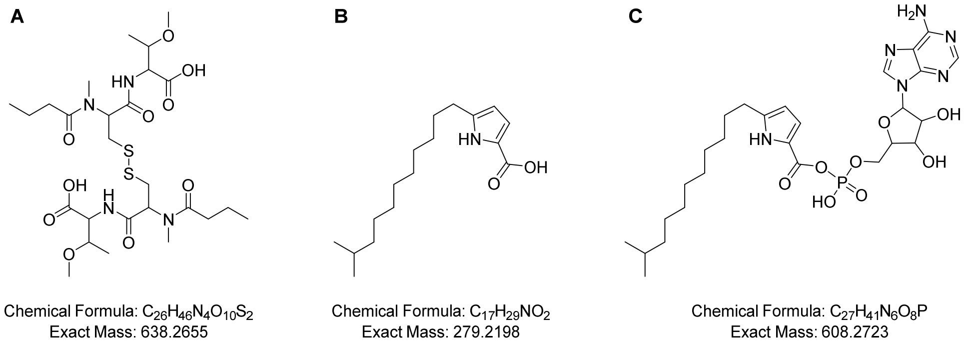


Figure 5 Proposed structure of the heterologous products of BGC No. 17 (A) and BGC No. 19 (B and C).

5.1.2 Factors affecting successful heterologous expression and potential solutions

1) In some cases, some biosynthetic genes are located outside the core gene cluster region, and therefore, successful heterologous expression requires the host organism with suitable surrogates. In disorazole biosynthesis, enzymes responsible for epoxide formation, methylation of the C6-OH, and oxidation of the geminal dimethyl group are encoded by genes situated outside the core biosynthetic gene clusters. In the sorangibactin biosynthetic gene cluster, genes encoding for P450 redox partners (iron-sulfur-containing ferredoxin and FAD-containing ferredoxin reductase) are absent. As functional pairing of the P450 enzyme (CYP) and redox partners is crucial for electron transfer from NAD(P)H to heme,[17] it is necessary to consider a heterologous host containing appropriate P450 redox partners for expression of the CYP-containing biosynthetic pathways. Missing tailoring genes and splitting biosynthetic genes across different genomic loci are not rare for myxobacteria. For instance, the first two modules, the glycosyltransferase, and a P450 enzyme are encoded by genes located outside the core ajudazol biosynthetic gene cluster from *Cystobacter* sp. SBCb004.[18] Similarly, the tailoring genes encoding for a glycosyltransferase and an acyltransferase involved in disciformycin biosynthesis are thought to be located elsewhere in the genome of the *Pyxidicoccus fallax* strain AndGT8.[19] To address this issue, a meticulous examination of the module organization and integrity/boundary of the BGC is required. Additionally, comparative analysis of the selected BGC with its homologs from alternative strains, and using phylogenetically closely-related hosts for heterologous expression hold promise for resolving this challenge.

2) It is possible that some cryptic biosynthetic gene clusters are non-functional in their current stage of the evolutionary process. In theory, natural product biosynthesis is an energy-intensive process, and natural selection may favor the loss or inactivation of biosynthetic gene clusters that are no longer beneficial for the organism, leading to the accumulation of non-functional gene clusters; in addition, enzymes may evolve to bind different substrates or acquire new functions during the cycles of inactivation and reactivation. In practice, active gene clusters are more likely to be successfully expressed in a heterologous host even without modifications compared to cryptic gene clusters. For example, *S. cellulosum* So ce12 and So ce427 are capable of producing disorazole A and disorazole Z, respectively, and successful

heterologous expression of the disorazole gene clusters from both strains can be easily achieved; however, the disorazole gene cluster from the *A. rufus* strain MSr9331, which does not produce disorazole naturally, could not be successfully expressed even after promoter engineering. Moreover, many compounds that were initially discovered through heterologous expression were later found to be naturally produced by the original strains. For example, the sorangibactins, discovered by heterologous expression, were later found to be produced by *Sorangiineae* strain MSr11367 under iron-limited conditions. While it is challenging to predict whether a cryptic gene cluster is functional or not at this moment, once this is possible, for example through developing computational tools and algorithms, a higher success rate of heterologous expression will be definitely achieved. Currently, at least at the gene transcription level, with the aid of transcriptome analysis, we can obtain valuable insights into the functional expression of the gene cluster identified through genome mining.

3) The limited toolset available for myxobacteria presents challenges for successful heterologous expression. Host selection is crucial, but currently there are limited heterologous hosts available for myxobacteria.[20] *M. xanthus* DK1622 is by far the most versatile host for heterologous expression of myxobacterial gene clusters; however, further developments are still needed. For example, the efficiency of transformation using electroporation decreases dramatically for large constructs (>60 kb), hindering the heterologous expression of gene clusters that exceed 60 kb in size.[21] Even if mutants can be obtained, they occasionally lose partial regions of the BGC, especially when expressing large NRPS/PKS gene clusters with reparative sequences, e.g., the disorazole gene cluster, regardless of whether transpositional or site-specific integration methods are used. As a result, these mutants are unable to produce the products, despite confirmation of the presence of the gene cluster in the mutants through colony PCR; however, the intact integration cannot be ensured solely based on colony PCR results. Noteworthily, the fast-growing myxobacterial host *Corallococcus macrosporus* GT-2 showed an advantageous production of myxochromide compared to *M. xanthus*;[22] however, no further developments or applications were reported. Non-myxobacterial hosts such as *Streptomyces*, *Pseudomonads*, *Escherichia*, and *Burkholderia* have also been explored for heterologous expression of myxobacterial gene clusters; however, the success rate and production titer vary on a case-by-case basis. The successful/efficient heterologous expression is closely correlated with the phylogenetic proximity between the heterologous host

and the original producer strain, and *M. xanthus* DK1622 is currently the most suitable host for clusters originating from *Cystobacterineae*; however, the development of heterologous hosts from the suborders *Nannocystineae* and *Sorangiineae* is still required. Therefore, akin to the engineering of *Streptomyces* chassis, optimization of a panel of myxobacterial hosts that cover the three suborders could drive the successful heterologous expression of myxobacterial gene clusters.[23-26] In addition to the selection of an appropriate host, ensuring proper transcription is also pivotal to successful heterologous expression,[27] therefore, promoter sets that are compatible with the selected host are also needed for refactoring the gene clusters.[28] This is no longer an issue for *M. xanthus* since two inducible promoter systems have been established and a range of constitutive promoters with varying strength have been characterized.[29,30]

4) Undesired mutations introduced during cloning and engineering could inactivate the gene cluster. Unlike homologous expression, where the biosynthetic pathways are activated *in situ*, heterologous expression requires cloning the entire gene cluster into a plasmid that is typically maintained in *E. coli* or *S. cerevisiae*. As a result, mutations and intramolecular rearrangement may arise during propagation and modification of these exogenous genes, particularly when repeating sequences are present or their products are toxic to the heterologous host. In addition, when PCR is used for cloning and engineering, mutations may arise due to the impurities of the DNA oligonucleotides or the non-fidelity of the DNA polymerase. Illumina sequencing could be used for trouble shooting; however, this approach is still costly and thus cannot be implemented for all the constructs generated during high-throughput screening of cryptic gene clusters. Therefore, it is advantageous to use the direct cloning approach, high-purity oligos, high-fidelity polymerases, a lower cultivation temperature, inducible promoters, et al., to prevent undesired mutations.

5) The methods used for cultivation, extraction, measurement, and analysis can significantly impact the production and detection of secondary metabolites,[31,32] and this is equally applicable to heterologous expression. Even if a functional gene cluster is completely integrated into a suitable heterologous host, the production of desired compounds still depends on appropriate fermentation conditions. For example, the heterologous production of vioprolide A and C, which contain the 4-methylazetidinecarboxylic moiety, could only be

achieved by supplementing the CTT medium with vitamin B12, a critical cofactor for *S*-adenosylmethionine (SAM) biosynthesis.[33] To avoid overlooking any heterologous products, it is recommended to optimize or customize extraction and detection methods for specific compounds according to their predicted molecular weight, chemical polarity, characteristic functional groups, and other relevant properties. Typically, heterologous products are identified by comparing the metabolic profiles of the mutant strain containing the exogenous gene cluster with that of the wild-type host, which is commonly accomplished through the manual comparison of base peak chromatograms (BPC) and/or UV traces. However, low-intensity or overlapping signals are often overlooked using this approach. As a consequence, computational tools are needed for downstream processing of the HPLC-HRMS datasets.[34] Feature extractions (including retention time, *m/z*, and intensity values) using software such as the MetaboScape® (Bruker) and subsequent statistical analyses, such as principal component analysis (PCA) and t-test, can help to find out compounds that are exclusively present in the dataset of the mutant strains, even in tiny amounts.[1,35]

6) Successful expression finally depends on the proper identification and characterization of the heterologous products. Simply observing new peaks does not necessarily mean the identification of the true heterologous products, as these peaks might represent intermediates, derivatives, shunt products, or artificial compounds. Generation of these unexpected products can arise from factors such as different precursor pools, incomplete expression of the gene cluster, crosstalk between biosynthetic pathways, or substrate promiscuity of biosynthetic enzymes. In addition to careful dereplication against medium blank and wild-type host, it is recommended to doublecheck the original producer strain, perform a MS/MS-based molecular networking analysis, and feed stable-isotope-labeled precursors if necessary. This process enables prioritization of the detected heterologous products for further purification and structure characterization.

In summary, the successful heterologous expression of cryptic gene clusters requires a combination of the *in silico* prioritization, gene cluster manipulation, host selection, optimized cultivation and detection methods, and product(s) identification. By following these strategies, researchers might increase the chances of successfully expressing cryptic gene clusters and unlocking bacterial potential for the production of novel natural products.

5.2 Insights into the biosynthesis of disorazole and sorangibactin

5.2.1 The disorazole biosynthesis

The identification and heterologous expression of disorazole gene clusters, manipulation of biosynthetic pathways, and characterization of intermediates/derivatives provide insights into the understanding of the biosynthesis, especially the tailoring modifications involved in epoxy group and methyl ester group formation (Figure 6). Based on evidence from feeding substrates to the corresponding gene-knockout mutants, these tailoring steps are proposed to occur after the release of the disorazole bis-lactones.

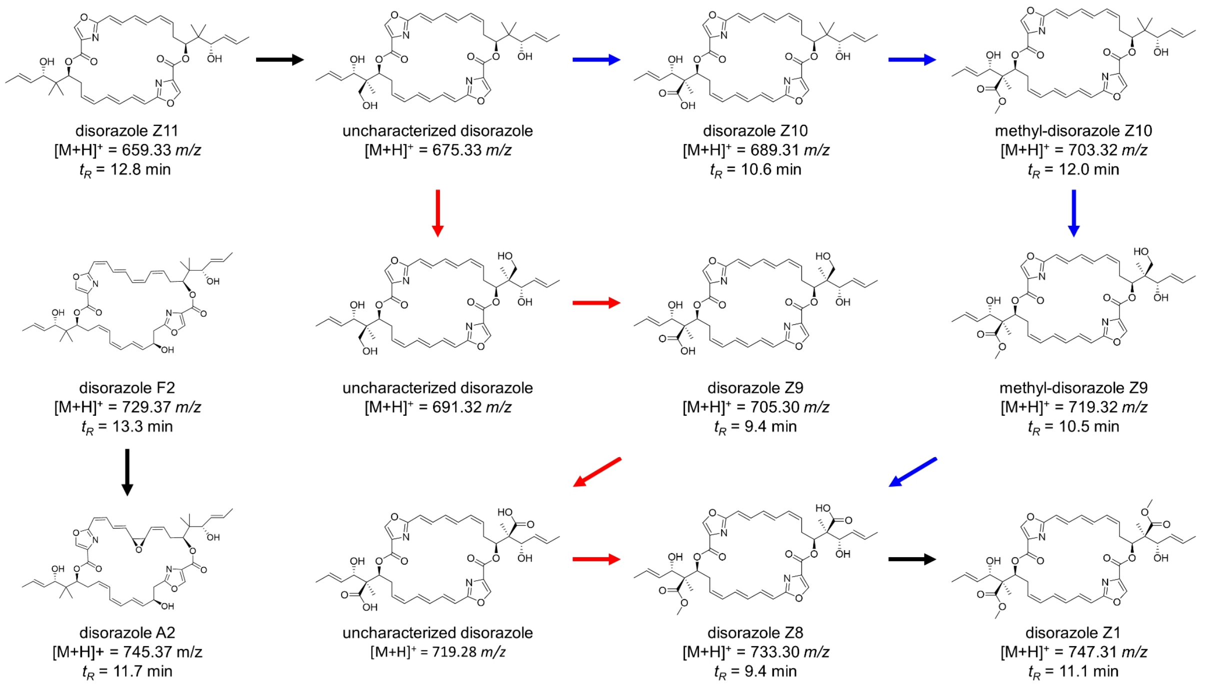


Figure 6 Post-modifications involved in disorazole biosynthesis. In the disorazole Z biosynthesis, stereospecific oxidation of the geminal dimethyl groups and subsequent *O*-methylations might take place initially on one side (indicated by blue arrows) or simultaneously on both sides (indicated by red arrows) of the symmetric disorazole Z11 substrate.

Both *in vivo* and *in vitro* studies have demonstrated that DisF catalyzes the *O*-methylation step necessary for the formation of methyl ester groups in disorazole Z biosynthesis. The accumulation of disorazole Z10 as the primary product instead of the uncharacterized disorazole ($[M+H]^+$ = 719.28 *m/z*) resulting from the deletion of *disF*$_{427}$ indicates that stereospecific oxidation of the geminal dimethyl groups and subsequent *O*-methylations might take place initially on one side of the symmetric disorazole Z11 substrate (Figure 6). Moreover, extensive *in vivo* studies have confirmed the involvement of $DisE_{12}$ in epoxidation and $DisE_{427}$ in oxidations of the geminal dimethyl group in disorazole A and disorazole Z biosynthesis, respectively. However, despite numerous efforts, complete *in vitro* reconstitution of the post-modification pathways has not yet been achieved, mainly because of the unavailability of functional DisE proteins and their associated oxidase partners.

The heterologous production of disorazole Z12, Z13 and F7 suggests that the biosynthesis of disorazole can be initiated using not only an acetate starter but also a propionate starter. This may help to solve the puzzle of the initiation mechanism in disorazole biosynthetic pathways that lack the starter module. Further biosynthesis studies, for example, investigating the function of non-extending modules and tandem HC domains, could be performed using the established heterologous systems.

5.2.2 The sorangibactin biosynthesis

The successful heterologous expression of a coelibactin-like NRPS gene cluster, and the subsequent characterization of sorangibactins, unambiguously established the correlation between a previously cryptic gene cluster and its products. By employing a combination of *in silico* analysis, gene knockout experiments, and feeding stable-isotope-labeled precursors, we were able to propose a concise biosynthetic pathway for sorangibactin production. However, some genes (e.g. *srbE*, *srbF*, and *srbJ*) initially proposed to encode tailoring enzymes could not be assigned to specific steps in the biosynthetic pathway. Their deletion did not result in the accumulation of intermediates, but completely abolished the production of sorangibactin A, indicating their crucial role in sorangibactin biosynthesis. It is plausible that these tailoring enzymes act on the PCP-bound substrates and that non-native intermediates stall the entire biosynthetic machinery, as previously demonstrated in the pyochelin biosynthesis.[36-38] Polar effects can be excluded because gene knockouts were accompanied by inserting a Ptet promoter to drive the downstream gene expression; in addition, when these genes were complemented *in trans*, the production of sorangibactin A was restored (Figure 7), which further confirmed the involvement of these genes in sorangibactin biosynthesis. At this moment, the function of these genes in sorangibactin biosynthesis is still unknown and would require *in vitro* studies.

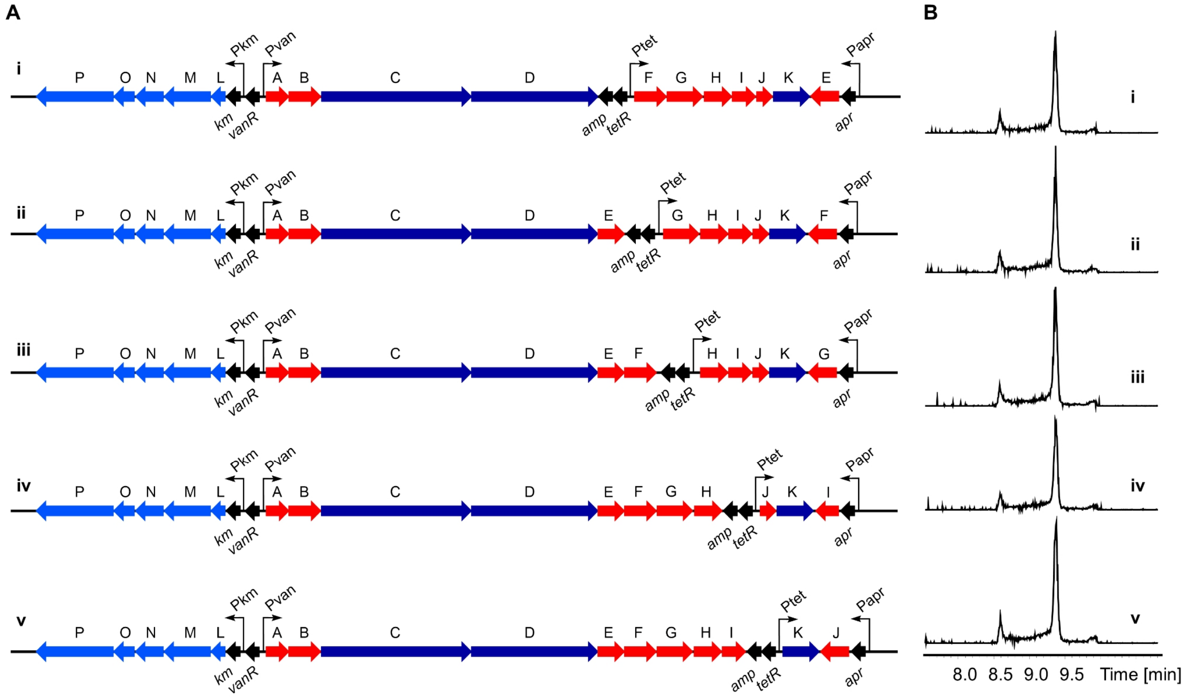


Figure 7 Gene knockouts and subsequent complements in sorangibactin biosynthetic pathway. (A) Scheme of expression constructs. (B) HPLC-MS analysis of crude extracts from *M. xanthus* mutants containing the corresponding expression constructs, chromatograms are shown as EIC of *m/z* 507.19 ± 0.1 for sorangibactin A.

The most intriguing aspect of sorangibactin biosynthesis is the formation of its unprecedented C-terminal moiety. To the best of my knowledge, the formation of a five-membered thiolactone ring has only been reported in thiolactomycin biosynthesis, where the involvement of a P450-dependent enzyme has been proposed.[39] However, in our case, the terminal TE domain containing an unusual active-site cysteine is believed to be involved. A similar TE domain from the fungal KU42 gene cluster has been shown to select cysteine or homocysteine and catalyze transthioesterification for chain release.[40] Whether a similar process is involved in sorangibactin biosynthesis and the subsequent rare thiolactone formation still require further biochemical investigations.

5.3 Rational engineering for natural products production

Engineering biosynthetic pathways could not only provide valuable insights into the biosynthesis of natural products but also facilitate improved production for downstream application or generation of non-natural compounds for structure-activity-relationship (SAR) studies. In this context, lessons and challenges regarding rational engineering, encountered in manipulation of disorazole and sorangibactin biosynthetic pathways, are discussed below.

5.3.1 PKS engineering – the disorazole biosynthetic pathways

Disorazole A and disorazole Z differ significantly in the size of their core bis-lactones and post-modifications, which are reflected in their biosynthetic genes (*disB*, *disE*, and *disF*). Therefore, we sought to hybrid these two biosynthetic pathways by swapping these genes to generate non-natural disorazoles (Figure 8); however, no desired heterologous products were observed. It is possible that disorazole biosynthetic pathways have highly strict substrate-specificity or that the functional interaction was lost in the hybrid system.

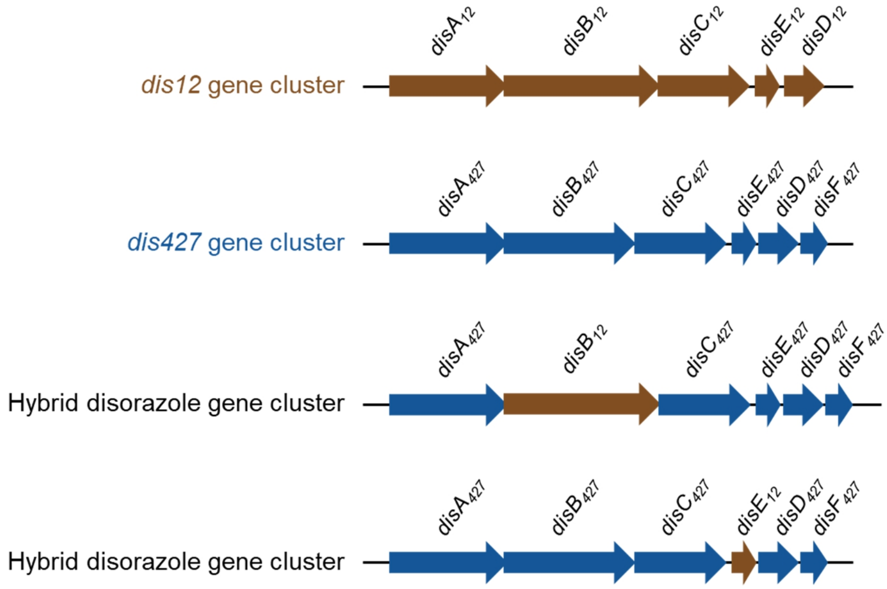


Figure 8 Engineering the disorazole biosynthetic pathways through gene swapping.

In comparison to gene swapping, site-directed mutagenesis is a minimally invasive approach that could minimize deleterious effects of the manipulation.[41] Structural and functional studies of the disorazole *gem*-dimethylating methyltransferase *in vitro* revealed that A326 acts as a gatekeeping residue controlling mono- or di-methylation; whereas, H288 is the key active site residue.[42] Therefore, point mutation of A326F and H288N were performed in the disorazole Z gene cluster (Figure 9) in order to generate *C*-desmethyl-disorazole Z.[43] For the point mutations of A326F, disorazole Z1 was still produced, along with a derivative with the mass that is 14 Da smaller compared to disorazole Z1, indicating the production of the *C*-desmethyl-disorazole Z. Due to the low yield and signal overlap with disorazole Z1 during semi-prep HPLC, this compound could not be purified in enough amounts for structure determination by NMR. For the point mutation of H288N, neither disorazole Z1 nor the expected derivative was produced. This could be attributed to the inactivation of the *C*-methyltransferase, leading to the formation of a non-native ACP-bound intermediate that cannot be processed by the downstream biosynthetic pathway. In addition to the methyltransferase domain, we also performed point mutations of the NRPS oxidase domain, which was revealed to use a conserved active-site tyrosine for deprotonation,[44,45] in order to generate the oxazoline-containing disorazole. However, the point mutation of Y1824F (Figure 9) completely abolished disorazole production, probably due to a significant difference in the geometry of the oxazoline-containing intermediate compared to the oxazole counterpart, rendering it incompatible with the downstream domains.

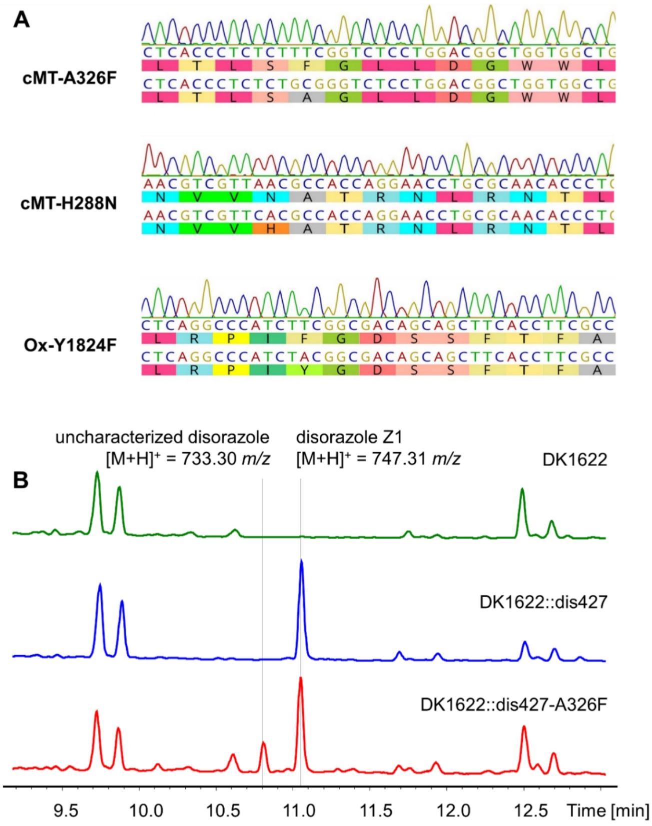


Figure 9 Engineering the disorazole biosynthetic pathways through site-directed mutagenesis. (A) Alignments of Sanger sequencing results of point-mutation constructs with the *dis427* gene cluster. (B) HPLC-MS analysis of crude extracts from *M. xanthus* DK1622 wild type and mutants containing the native and engineered *dis427* gene clusters. The chromatograms are shown as UV absorption at 320 nm.

The success rate is limited for obtaining the desired disorazole analogues through pathway engineering, even when minimally invasive mutagenesis technique is implemented. This is probably because the disorazole biosynthetic pathways have strict substrate specificities and protein-protein interactions. In addition, the yield of disorazole production in the heterologous host *M. xanthus* DK1622 remains low even after promoter engineering, not to mention that the yield of non-natural products would be even lower after pathway engineering, which hindered the identification and characterization.

5.3.2 NRPS engineering – the sorangibactin biosynthetic pathway

In the sorangibactin biosynthesis, SrbB is shown to catalyze salicylic acid formation. Deletion of *srbB* abolished sorangibactin production, which can be restored by providing exogenous salicylic acid. This suggests that the heterologous host *M. xanthus* DK1622 is incapable of synthesizing salicylic acid, and feeding of salicylic acid analogues to the *srbB*-knockout mutant might be a feasible way to generate sorangibactin analogues, which may help to overcome the instability of sorangibactin and simplify its purification. Moreover, halogenated analogues with improved bioactivities could be produced using precursor-directed mutabiosynthesis, as shown for amychelin.[46]

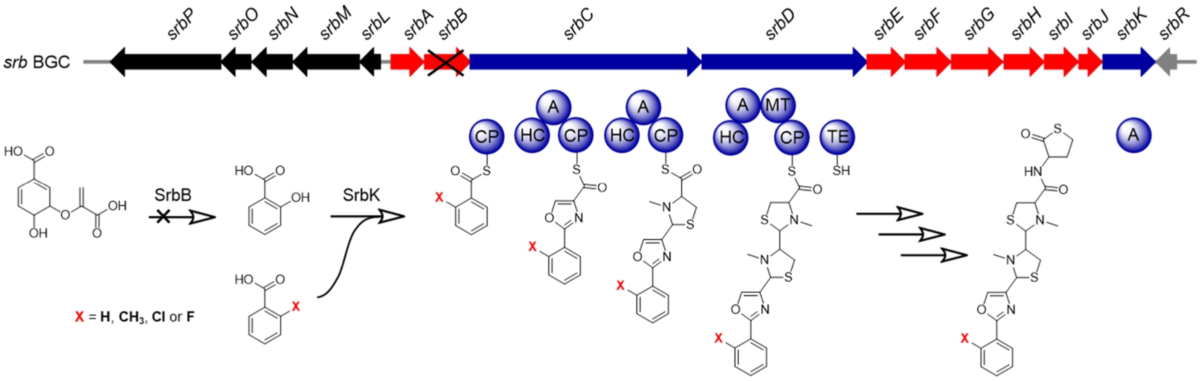


Figure 10 Scheme of precursor-directed mutabiosynthesis of sorangibactin analogues.

However, feeding of benzoic acid, 2-methyl-benzoic acid, 2-chloro-benzoic acid or 2-fluoro-benzoic acid to the *srbB*-knockout mutant did not result in the production of the corresponding sorangibactin analogues, which suggests that the biosynthetic machinery displays a high selectivity for the 2-OH group of salicylic acid (Figure 10). Therefore, to create non-natural sorangibactins using precursor-directed mutabiosynthesis, it may be necessary to use salicylic acid analogues that retain the 2-OH group.

5.3.3 Structure biology for rational engineering of biosynthetic pathways

Currently, molecular engineering of large gene clusters is no longer a significant challenge, and customized gene clusters can even be synthesized *de novo*. However, the greatest obstacle remains the rational engineering and design of functional biosynthetic pathways. Structure biology of biosynthetic machinery is playing an increasingly significant role in guiding pathway engineering, avoiding the time-consuming trial-and-error processes.[47] By examining the structure of catalytic domains or multi-domain complexes, regions that may be modified to alter their activity, specificity or stability could be identified.[48,49] Besides, by examining the structure of protein targets, binding sites could be identified, aiding in the understanding of the mode of action and optimization of the structure of natural product for drug development. For instance, cryo-EM structure-guided biosynthetic engineering was used to generate non-natural darobactins that exhibit superior antibiotic activities.[50] In addition, conducting structural investigations on domain-domain/substrate interactions and conformational dynamics can provide a detailed understanding of the catalytic mechanism and serve as a guidance for rational engineering of megasynthetases.[47,51]

5.4 Summary and perspectives

The increasing collection of genetic information, the accumulation of knowledge about natural product biosynthesis, the development of bioinformatics tools, and the application of artificial intelligence provide unprecedented opportunities for natural product discovery through genome mining.[52] In addition, the development of molecular cloning and engineering techniques have facilitated heterologous expression, enabling the characterization of cryptic gene clusters, the manipulation of known biosynthetic pathways, and even the construction of artificial metabolic networks in an advantageous host.[53,54] As such, disorazoles, sorangibactins, and many other studies have highlighted the promising renaissance of natural product research in the genomic era.[55]

However, the successful heterologous expression of cryptic gene clusters and obtaining (sufficient quantities of) desired products after pathway engineering still remain challenging tasks, with low success rates. Further studies, such as improving the accuracy of *in silico* predictions, developing advantageous hosts, and rationally redesigning of functional gene clusters, can contribute to the sustainable discovery and diversification of natural products through heterologous expression. This would be followed by the optimization of heterologous production, solving the bottleneck of compound supply for downstream studies and pharmaceutical applications.

5.5 References

[1] F. Panter, D. Krug, S. Baumann et al., Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria, *Chem Sci*, **2018**, 9(21), 4898-4908.

[2] S. Meng, W. Han, J. Zhao et al., A Six-Oxidase Cascade for Tandem C-H Bond Activation Revealed by Reconstitution of Bicyclomycin Biosynthesis, *Angew Chem Int Ed*, **2018**, 57(3), 719-723.

[3] J. B. Patteson, W. Cai, R. A. Johnson et al., Identification of the Biosynthetic Pathway for the Antibiotic Bicyclomycin, *Biochemistry*, **2018**, 57(1), 61-65.

[4] E. A. Lilla and K. Yokoyama, Carbon extension in peptidylnucleoside biosynthesis by radical SAM enzymes, *Nat Chem Biol*, **2016**, 12(11), 905-907.

[5] M. Xu, F. Zhang, Z. Cheng et al., Functional Genome Mining Reveals a Class V Lanthipeptide Containing a D-Amino Acid Introduced by an $F_{420}H_2$-Dependent Reductase, *Angew Chem Int Ed*, **2020**, 59(41), 18029-18035.

[6] Y. Gao, J. Birkelbach, C. Fu et al., The Disorazole Z Family of Highly Potent Anticancer Natural Products from *Sorangium cellulosum*: Structure, Bioactivity, Biosynthesis and Heterologous Expression, *Microbiol Spectrum*, **2023**, acceped manuscript.

[7] R. Carvalho, R. Reid, N. Viswanathan et al., The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation, *Gene*, **2005**, 359, 91-98.

[8] M. Kopp, H. Irschik, S. Pradella et al., Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes, *ChemBioChem*, **2005**, 6(7), 1277-1286.

[9] J. Kawata, T. Naoe, Y. Ogasawara et al., Biosynthesis of the Carbonylmethylene Structure Found in the Ketomemicin Class of Pseudotripeptides, *Angew Chem Int Ed*, **2017**, 56(8), 2026-2029.

[10] W.-J. Kang, H.-X. Pan, S. Wang et al., Identification of the Amipurimycin Gene Cluster Yields Insight into the Biosynthesis of C9 Sugar Nucleoside Antibiotics, *Org Lett*, **2019**, 21(9), 3148-3152.

[11] M. Sosio, E. Gaspari, M. Iorio et al., Analysis of the Pseudouridimycin Biosynthetic Pathway Provides Insights into the Formation of C-nucleoside Antibiotics, *Cell Chem Biol*, **2018**, 25(5), 540-549.e4.

[12] Q. Tu, J. Herrmann, S. Hu et al., Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering, *Sci Rep*, **2016**, 6, 21066.

[13] R. Hermenau, K. Ishida, S. Gama et al., Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system, *Nat Chem Biol*, **2018**, 14(9), 841-843.

[14] R. Hermenau, J. L. Mehl, K. Ishida et al., Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria, *Angew Chem Int Ed*, **2019**, 58(37), 13024-13029.

[15] Y. Gao, C. Walt, C. D. Bader et al., Genome-Guided Discovery of the Myxobacterial Thiolactone-Containing Sorangibactins, *ACS Chem Biol*, **2023**, 18(4), 924-932.

[16] S. M. Salem, P. Kancharla, G. Florova et al., Elucidation of final steps of the marineosins biosynthetic pathway through identification and characterization of the corresponding gene cluster, *J Am Chem Soc*, **2014**, 136(12), 4565-4574.

[17] W. Zhang, L. Du, F. Li et al., Mechanistic Insights into Interactions between Bacterial Class I P450 Enzymes and Redox Partners, *ACS Catal*, **2018**, 8(11), 9992-10003.

[18] H. Zeng, J. Birkelbach, J. Hoffmann et al., Expanding the Ajudazol Cytotoxin Scaffold: Insights from Genome Mining, Biosynthetic Investigations, and Novel Derivatives, *J Nat Prod*, **2022**, 85(11), 2610-2619.

[19] F. Surup, K. Viehrig, K. I. Mohr et al., Disciformycins A and B: 12-Membered Macrolide Glycoside Antibiotics from the Myxobacterium *Pyxidicoccus fallax* Active against Multiresistant Staphylococci, *Angew Chem Int Ed*, **2014**, 53(49), 13588-13591.

[20] J. J. Hug and R. Müller, Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products, *Comprehensive Natural Products III (Third Edition)*, **2020**, 6, 149-216.

[21] J. Fu, S. C. Wenzel, O. Perlova et al., Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition, *Nucleic Acids Res*, **2008**, 36(17), e113.

[22] O. Perlova, K. Gerth, S. Kuhlmann et al., Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate *Corallococcus macrosporus* GT-2, *Microb Cell Fact*, **2009**, 8, 1.

[23] M. Myronovskyi, B. Rosenkränzer, S. Nadmid et al., Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters, *Metab Eng*, **2018**, 49, 316-324.

[24] M. Myronovskyi and A. Luzhetskyy, Heterologous production of small molecules in the optimized *Streptomyces* hosts, *Nat Prod Rep*, **2019**, 36(9), 1281-1294.

[25] Y. Ahmed, Y. Rebets, M. R. Estévez et al., Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters, *Microb Cell Fact*, **2020**, 19(1), 5.

[26] J. Liu, X. Wang, G. Dai et al., Microbial chassis engineering drives heterologous production of complex secondary metabolites, *Biotechnol Adv*, **2022**, 59, 107966.

[27] D. C. Stevens, T. P. A. Hari, C. N. Boddy, The role of transcription in heterologous expression of polyketides in bacterial hosts, *Nat Prod Rep*, **2013**, 30(11), 1391-1411.

[28] M. Myronovskyi and A. Luzhetskyy, Native and engineered promoters in natural product discovery, *Nat Prod Rep*, **2016**, 33(8), 1006-1019.

[29] A. A. Iniesta, F. García-Heras, J. Abellón-Ruiz et al., Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-β-D-thiogalactopyranoside or vanillate, *J Bacteriol*, **2012**, 194(21), 5875-5885.

[30] W.-F. Hu, L. Niu, X.-J. Yue et al., Characterization of Constitutive Promoters for the Elicitation of Secondary Metabolites in Myxobacteria, *ACS Synth Biol*, **2021**, 10(11), 2904-2909.

[31] C. D. Bader, M. Neuber, F. Panter et al., Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria, *Anal Chem*, **2020**, 92(23), 15403-15411.

[32] C. D. Bader, P. A. Haack, F. Panter et al., Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems, *J Nat Prod*, **2021**, 84(2), 268-277.

[33] F. Yan, D. Auerbach, Y. Chai et al., Biosynthesis and heterologous production of vioprolides: rational biosynthetic engineering and unprecedented 4-methylazetidinecarboxylic acid formation, *Angew Chem Int Ed*, **2018**, 57(28), 8754-8759.

[34] D. Krug, G. Zurek, B. Schneider et al., Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionisation-time-of-flight mass

spectrometry and compound-based principal component analysis, *Anal Chim Acta*, **2008**, 624(1), 97-106.

[35] N. S. Cortina, D. Krug, A. Plaza et al., Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome, *Angew Chem Int Ed*, **2012**, 51(3), 811-816.

[36] L. E. Quadri, T. A. Keating, H. M. Patel et al., Assembly of the *Pseudomonas aeruginosa* nonribosomal peptide siderophore pyochelin: In vitro reconstitution of aryl-4, 2-bisthiazoline synthetase activity from PchD, PchE, and PchF, *Biochemistry*, **1999**, 38(45), 14941-14954.

[37] H. M. Patel and C. T. Walsh, In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and *N*-methyltransferase activities, *Biochemistry*, **2001**, 40(30), 9023-9031.

[38] C. Reimmann, H. M. Patel, L. Serino et al., Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*, *J Bacteriol*, **2001**, 183(3), 813-820.

[39] X. Tang, J. Li, B. S. Moore, Minimization of the Thiolactomycin Biosynthetic Pathway Reveals that the Cytochrome P450 Enzyme TlmF Is Required for Five-Membered Thiolactone Ring Formation, *ChemBioChem*, **2017**, 18(12), 1072-1076.

[40] M.-C. Tang, C. R. Fischer, J. V. Chari et al., Thioesterase-Catalyzed Aminoacylation and Thiolation of Polyketides in Fungi, *J Am Chem Soc*, **2019**, 141(20), 8198-8206.

[41] S. Kushnir, U. Sundermann, S. Yahiaoui et al., Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library, *Angew Chem Int Ed*, **2012**, 51(42), 10664-10669.

[42] J. L. Meinke, M. R. Mehaffey, D. T. Wagner et al., Structural and Functional Studies of a *gem*-Dimethylating Methyltransferase from a *trans*-Acyltransferase Assembly Line, *ACS Chem Biol*, **2018**, 13(12), 3306-3314.

[43] C. P. Bold, D. Lucena-Agell, M. Á. Oliva et al., Synthesis and Biological Evaluation of C(13)/C(13')-Bis(desmethyl)disorazole Z, *Angew Chem Int Ed Engl*, **2023**, 62(5), e202212190.

[44] B. Pang, Y. Chen, F. Gan et al., Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases, *J Am Chem Soc*, **2020**, 142(25), 10931-10935.

[45] C. M. Fortinez, K. Bloudoff, C. Harrigan et al., Structures and function of a tailoring oxidase in complex with a nonribosomal peptide synthetase module, *Nat Commun*, **2022**, 13(1), 548.

[46] F. Xie, S. Dai, Y. Zhao et al., Generation of Fluorinated Amychelin Siderophores against *Pseudomonas aeruginosa* Infections by a Combination of Genome Mining and Mutasynthesis, *Cell Chem Biol*, **2020**, 27(12), 1532-1543.

[47] K. D. Patel, M. R. MacDonald, S. F. Ahmed et al., Structural advances toward understanding the catalytic activity and conformational dynamics of modular nonribosomal peptide synthetases, *Nat Prod Rep*, **2023**, Advance Article, DOI: 10.1039/d3np00003f.

[48] K. J. Weissman, The structural biology of biosynthetic megaenzymes, *Nat Chem Biol*, **2015**, 11(9), 660-670.

[49] M. Grininger, Enzymology of assembly line synthesis by modular polyketide synthases, *Nat Chem Biol*, **2023**, 19(4), 401-415.

[50] C. E. Seyfert, C. Porten, B. Yuan et al., Darobactins Exhibiting Superior Antibiotic Activity by Cryo-EM Structure Guided Biosynthetic Engineering, *Angew Chem Int Ed*, **2023**, 62(2), e202214094.

[51] S. H. Mishra, A. K. Kancherla, K. A. Marincin et al., Global protein dynamics as communication sensors in peptide synthetase domains, *Sci Adv*, **2022**, 8(28), eabn6549.

[52] F. I. Saldívar-González, V. D. Aldas-Bulos, J. L. Medina-Franco et al., Natural product drug discovery in the artificial intelligence era, *Chem Sci*, **2022**, 13(6), 1526-1546.

[53] D. A. Yee, K. Niwa, B. Perlatti et al., Genome mining for unknown-unknown natural products, *Nat Chem Biol*, **2023**, 19(5), 633-640.

[54] J. Zhang, L. G. Hansen, O. Gudich et al., A microbial supply chain for production of the anti-cancer drug vinblastine, *Nature*, **2022**, 609(7926), 341-347.

[55] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, The re-emergence of natural products for drug discovery in the genomics era, *Nat Rev Drug Discov*, **2015**, 14(2), 111-129.