

# **Biosynthesis of Sporothriolides and Sporochartines in Fungi**

Von der Naturwissenschaftlichen Fakultät  
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades  
**Doktor der Naturwissenschaften (Dr. rer. nat.)**  
genehmigte Dissertation

von  
**Dongsong Tian, Master (China)**

2021

Referent: Prof. Dr. Russell J. Cox

Korreferent: Prof. Dr. Marc Stadler

Tag der Promotion: 14.09.2021

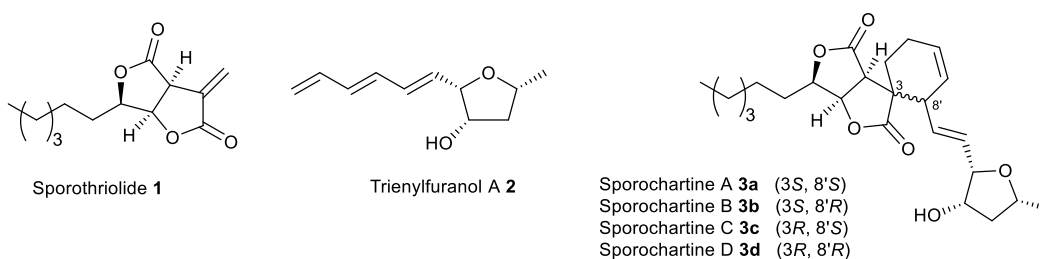
## Abstract

**Key words:** Alkyl citrates, biosynthesis, sporothriolides, sporochartines, polyene

The project focused on understanding the biosynthesis of sporothriolides, sporochartines and trienylfuranol A. Gene cluster identification, gene knock out, heterologous expression and protein *in vitro* assays were used during the investigation.

Alkyl citrate biosynthetic gene clusters of the antifungal metabolite sporothriolide **1** were identified from the genomes of the ascomycetes: *Hypomontagnella monticulosa* MUCL 54604, *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471. A transformation protocol was established, and genes encoding a fatty acid synthase subunit and a citrate synthase were simultaneously knocked out which led to the loss of sporothriolide and sporochartine production. Heterologous expression of the *spo* genes in *Aspergillus oryzae* then led to the production of intermediates and shunts and delineation of a new fungal biosynthetic pathway originating in fatty acid biosynthesis. Finally, a hydrolase was revealed by *in vitro* studies likely contributing towards self-resistance of the producer organism. *In vitro* reactions showed that the sporochartines are derived from non-enzymatic Diels-Alder cycloaddition of **1** and trienylfuranol A **2** during the fermentation and extraction process.

Several hrPKS gene clusters were identified as the potential polyene BGC for trienylfuranol A **2** through multiple bioinformatic analysis, however metabolites produced from the PKS in heterologous expression belong to either different polyene type compounds or pyrone derivatives. Based on these results, a highly unusual epoxidation/decarboxylation mechanism was proposed to be involved during trienylfuranol A **2** biosynthesis, and a new pyrone BGC likely to encode the biosynthesis of a large class of bioactive compounds related to islandic acid **161** was identified.



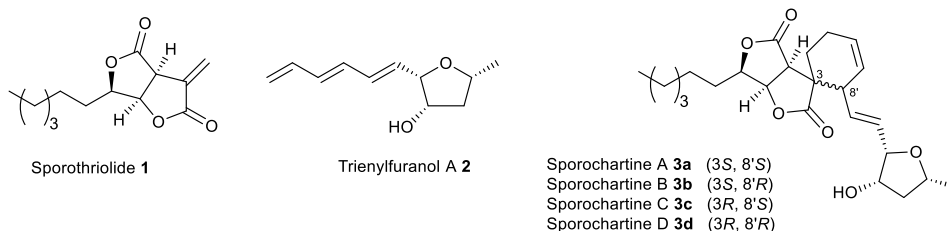
## Zusammenfassung

**Schlagwörter:** Alkyl citrates, biosynthesis, sporothriolides, sporochartines, polyene

Das Projekt fokussiert darauf, die Biosynthese der pilzlichen Sekundärmetabolite Sporothriolides, Sporochartines und Trienyfuranol A zu verstehen. Methoden zur Identifizierung von Genclustern, Gen-Knockout, heterologe Expression und *in vitro* Proteinassays wurden während der Untersuchung angewendet.

Die Alkylcitrat-Biosynthesecluster (*spo*) des antifungalen Metabolits Sporothriolide **1** wurden in den Genomen der Ascomyceten *Hypomontagnella monticulosa* MUCL 54604, *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471 identifiziert. Ein Transformationsprotokoll wurde etabliert und Gene, die für eine Fettsäuresynthaseuntereinheit und Citratsynthase kodieren, wurden gleichzeitig ausgeschaltet. Dies führte zu einem Verlust der Sporothriolide- und Sporochartine-Produktion. Heterologe Expression der Gene aus dem *spo* Gencluster in *Aspergillus oryzae* führte dann zur Produktion von Intermediaten und Abzweigungsprodukten. Die Identifizierung dieser Produkte ermöglichte die Beschreibung eines neuen Biosyntheseweges in Pilzen, der von der Fettsäurebiosynthese abgeleitet ist. Darüberhinaus wurde die Funktion eines im Biosynthesecluster kodierten Enzyms mit Hilfe von *in vitro* Untersuchungen als Hydrolase aufgeklärt, welche vermutlich zur Selbstresistenz des Produzenten beiträgt. *In vitro* Reaktionen von Sporothriolide **1** und Trienyfuranol A **2** zeigten, dass Sporochartine mittels nicht-enzymatischer Diels-Alder-Reaktion während des Fermentationsprozesses entstehen.

Im zweiten Teil der Arbeit wurde die Biosynthese von Trienyfuranol A **2** untersucht. Mehrere Kandidaten-Gencluster mit einer reduzierenden Polyketidsynthase als zentrales Gen wurden mit Hilfe von bioinformatischen Analysen identifiziert und mittels heterologer Expression untersucht. Jedoch besaß keine der produzierten Substanzen strukturelle Ähnlichkeit mit Trienyfuranol A **2**. Stattdessen handelte es sich um andere Arten von Polyen-Strukturen oder Pyron-Derivate. Unabhängig von diesen Ergebnissen wurde eine Biosyntheseweg für Trienyfuranol A vorgeschlagen, der auf einem ungewöhnlichen Epoxidierungs-Decarboxylierungs-Mechanismus basiert, und ein neues Pyron-BGC wurde identifiziert, das wahrscheinlich die Biosynthese einer großen Klasse bioaktiver Verbindungen ähnlich zu Islandic acid **161** kodiert.





## Acknowledgement

I appreciate Prof. Dr. Russell Cox for providing me with the great opportunity to join the fantastic group in Hannover. Thank you for the nice discussions and helpful supervisions of my PhD project. Your professional knowledge, passion, inspiration, patience and encouragement are very impressive.

I would like to thank Prof. Dr. Jamal Ouazzani who is our great collaborator in France, and the talk with you in Hannover was very nice. In addition, I would like to thank Prof. Dr. Marc Stadler from HZI-Braunschweig and Dr. Mark W Sumarah from Canada for the kind gift of fungi strains and useful suggestions. Thank you Dr. Daniel Wibberg and Prof. Dr. Jörn Kalinowski from Bielefeld, your excellent work on the gDNA and RNA sequencing made a good start for my project. Thank you to the great technical team members in OCI and BMWZ, especially to Katja Körner who is always enthusiastic to help others, and to Jörg Fohrer as well as Linn Haase from our great NMR department.

I would like to thank Prof. Dr. Marc Stadler and Prof. Dr. Jakob Franke for being my PhD thesis examiners. And Thanks Henrike and Dr. Kevin Becker for taking their time to read my thesis and giving me helpful suggestions.

Moreover, I would like to thank all the Cox group members. Thank you Dr. Eric Kuhnert, who is always happy to discuss the scientific questions and give me a hand to solve my life puzzles, but next time please use pure Chinese. Thank you Sen, Chongqing, Jin and Yunlong, the together travel, cooking, Mahjong gaming, swimming and basketball playing in the spare time fully charged my battery. Except, I would like to thank Elizabeth, Carsten, Lukas, Lei, Vjaceslavs, Mary, Katharina, Henrike, Maurice, Haili, Jing, Francesco, Steffen and Eman. Also, special thanks to my project collaborator Dr. Tian Cheng for the warm reception in Braunschweig to me and my girlfriend. If there are some names I forget to mention here, please also accept my thanks.

Thanks to the China Scholarship Council to cover my living expense in Germany.

Particularly, I would like to thank my parents Ming-An (明安) and Zi-Mei (自梅), my sister Jin-Ling (金灵), and my old brother Dong-Sheng (东升), who covered my back throughout my study from college to the PhD. In addition, thanks to Chun-Xia (春霞), my girlfriend, you are the backstage hero. It would be impossible to submit the thesis without your supports.

I enjoyed the time in the last four years, I think I will often recall this memorable experience as it means a lot to me.

## Abbreviations and Units

ACP	acyl carrier protein	KI	ketosteroid isomerase-like
<i>att</i>	site-specific attachment	LCMS	liquid chromatography mass spectrometry Spectrometry
antiSMASH	antibiotics & Secondary Metabolite Analysis	MPT	malonyl/palmitoyl transferase
<i>bf</i>	byssochlamic acid	MS	mass spectrometry
BGC	biosynthetic gene cluster	mRNA	messenger RNA
BLAST	basic local alignment search tool	MeOD	deuterated methanol
AT	acetyltransferase	MeOH	methanol
bp	base pair	NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
cDNA	complementary DNA	NEL	normalised expression level
C-MeT	C-methyltransferase	NMR	nuclear magnetic resonance
CoA	coenzyme A	NOESY	nuclear overhauser effect spectroscopy
COSY	correlation spectroscopy	nrPKS	non-reducing PKS
carb	carbenicillin	NRPS	non-ribosomal peptide synthetase
cam	chloramphenicol	ORF	open reading frame
CS	citrate synthase	<i>ory</i>	oryzine
CDCl <sub>3</sub>	deuterated chloroform	PCR	polymerase chain reaction
DA(ase)	Diels Alder(ase)	PEG	polyethylene glycol
DAD	diode array detector	PKS	polyketide synthase
ddH <sub>2</sub> O	double distilled H <sub>2</sub> O	ppm	parts per million
DH	dehydratase	prPKS	partially reducing PKS
DESeq	differential expression sequence	PUFA	polyunsaturated fatty acid
DNA	deoxyribonucleic acid	PTM	polycyclic tetramate macrolactam
EDTA	ethylenediaminetetraacetic acid	<i>plf</i>	piliformic acid
ER	enoyl reductase	<i>P<sub>amyB</sub></i>	amyB promoter
eGFP	enhanced green fluorescent protein	P450	cytochrome P450
ESI	electrospray ionization	PEBP	phosphatidylethanolamine-binding proteins PPPRPRprotein
EIC	extracted ion chromatogram	PPTase	phosphopantetheinyltransferase
ELSD	evaporative light scattering detector	<i>P<sub>gpdA</sub></i>	gpdA promoter
FPLC	fast protein liquid chromatography	RT-PCR	reverse transcription PCR
FAS	fatty acid synthase	RT	retention time
FMO	FAD-dependent monooxygenase	RNA	ribonucleic acid
FAD	flavin adenine dinucleotide	rpm	revolutions per minute
gDNA	genomic DNA	SAM	S-adenosyl methionine
HMBC	heteronuclear multiple bond correlation	SM	secondary metabolites
HPLC	high performance liquid chromatography	SDR	short chain dehydrogenase/reductase
hrPKS	highly reducing polyketide synthase	SQHKS	squalestatin hexaketide synthase
<sup>1</sup> H NMR	proton NMR	SQTKS	squalestatin tetraketide synthase
HRMS	high resolution mass spectrometry	<i>spo</i>	sporothiolide
HSQC	heteronuclear single quantum coherence	TIC	total ion current
<i>hph</i>	hygromycin B resistance	TAE	tris-acetate-EDTA
IPTG	isopropyl-β-D-thiogalactopyranoside	TE	thiolesterase
ITS	internal transcribed spacer	TMS	tetramethylsilane
Kan	kanamycin	UV	ultra violet
kb	kilo base pairs	UPLC	ultra-performance liquid chromatography
KR	ketoacyl reductase	WT	wild type
KS	ketoacyl synthase	yFAS	yeast FAS
KO	knockout	YHR	yeast homologous recombination

## Contents

Abstract.....	I
Zusammenfassung.....	II
Acknowledgement.....	III
Abbreviations and Units.....	IV
Contents.....	V
<b>1 Introduction.....</b>	<b>1</b>
1.1 Natural Products from Fungi.....	1
1.1.1 Fungal Fatty Acid Biosynthesis.....	2
1.1.2 Fungal Polyketide Biosynthesis.....	4
1.2 The Biosynthesis of Maleidrides and Alkyl Citrates.....	6
1.2.1 Maleidrides.....	6
1.2.2 Squalestatin.....	7
1.2.3 Oryzine.....	8
1.2.4 Hexylcitric Acid Derivatives.....	9
1.2.5 Piliformic Acid.....	10
1.3 Techniques Used in Fungal Biosynthesis Investigations.....	11
1.3.1 Isotopic Labelling.....	11
1.3.2 Genome Sequencing.....	13
1.3.3 Gene Knockout.....	14
1.3.4 Heterologous expression.....	16
1.4 Overall Aims.....	18
<b>2 Biosynthetic Studies of Sporothriolide.....</b>	<b>19</b>
2.1 Introduction.....	19
2.2 Project Aims.....	20
2.3 Results.....	20
2.3.1 Sporothriolide Production from <i>H. monticulosa</i> , <i>H. spongiphila</i> and <i>H. submonticulosa</i> .....	20
2.3.1.1 Producing and Non-producing Conditions.....	21
2.3.1.2 Time Course Study of Sporothriolide Production.....	22
2.3.2 The Identification of Multiforisin H.....	23
2.3.3 Acetate Feeding Experiments of Sporothriolide.....	26
2.3.4 Genome and Transcriptome Analysis.....	27
2.3.4.1 Whole Genome Sequencing and antiSMASH Analysis.....	27
2.3.4.2 Sporothriolide BGC Analysis.....	28

2.3.4.3	Transcriptomic Analysis.....	31
2.3.4.4	MultiGeneBlast.....	34
2.3.5	Gene Knockout in <i>H. spongiphila</i> CLL 205.....	35
2.3.5.1	Fungal Transformation of <i>H. spongiphila</i> CLL 205.....	35
2.3.5.1a	Antibiotics Screening and Protoplast Preparation.....	36
2.3.5.1b	Transformation with pTH-GS- <i>eGFP</i> .....	36
2.3.5.2	Knockout of <i>spofasA/spoE</i> .....	37
2.3.5.2a	Vector Construction for <i>spofasA/spoE</i> Knockout.....	37
2.3.5.2b	<i>spofasA/spoE</i> Knockout Workflow.....	38
2.3.5.2c	<i>spofasA/spoE</i> Knockout Transformant Analysis.....	39
2.3.5.3	Attempted Knockout of Other Genes.....	40
2.3.5.3a	<i>spoG</i> , <i>spoH</i> and <i>spoK</i> Knockout Vector Constructions and Transformation.....	40
2.3.5.3b	<i>spoG</i> , <i>spoH</i> and <i>spoK</i> Knockout Transformant Analysis.....	41
2.3.6	Heterologous Expression in <i>A. oryzae</i> NSAR1.....	43
2.3.6.1	Expression Vectors and Cloning Strategies.....	43
2.3.6.2	Gene Combinations and Plasmid Constructions.....	44
2.3.6.2a	Overview of Constructed Plasmids.....	44
2.3.6.2b	Gene Combinations and <i>A. oryzae</i> NSAR1 Transformation.....	48
2.3.6.3	Expression of the <i>spo</i> Genes for Early Steps.....	49
2.3.6.4	Co-expression of Early Step Genes with Later Tailoring Genes.....	50
2.3.6.4a	Co-expression of <i>spofasA</i> , <i>spofasB</i> , <i>spoE</i> , <i>spoL</i> and <i>spo</i> .....	50
2.3.6.4b	Co-expression of <i>spofasA</i> , <i>spofasB</i> , <i>spoE</i> , <i>spoL</i> , <i>spoK</i> and <i>spoG</i> .....	51
2.3.6.4c	Co-expression of <i>spofasA</i> , <i>spofasB</i> , <i>spoE</i> , <i>spoL</i> , <i>spoK</i> , <i>spoG</i> and <i>spoH</i> .....	56
2.3.6.4d	Co-expression of <i>spofasA</i> , <i>spofasB</i> , <i>spoE</i> , <i>spoL</i> , <i>spoK</i> , <i>spoG</i> and <i>spoJ</i> .....	57
2.3.6.4e	Co-expression of <i>spofasA</i> , <i>spofasB</i> , <i>spoE</i> , <i>spoL</i> , <i>spoK</i> , <i>spoG</i> , <i>spoH</i> and <i>spoJ</i> .....	58
2.3.7	<i>In Vitro</i> Activity Assay with SpoG and SpoI.....	62
2.3.7.1	Expression, Purification and Activity Assay of SpoG.....	62
2.3.7.2	Expression, Purification and Activity Assay of SpoI.....	66
2.4	Discussion.....	68
2.4.1	Production and Labelling Experiment.....	68

2.4.2	Bioinformatic Analysis.....	69
2.4.3	Bipartite Knockout.....	69
2.4.4	Heterologous Expression and Protein <i>In Vitro</i> Assay.....	69
2.4.5	Biosynthesis of Other Alkyl Citrates.....	73
2.5	Conclusion and Prospect.....	76
<b>3</b>	<b>Biosynthetic Studies of Sporochartine.....</b>	<b>78</b>
3.1	Introduction.....	78
3.1.1	Examples of Diels-Alder [4+2] Cycloaddition in Natural Products Biosynthesis.....	78
3.1.2	Enzymatic and Non-enzymatic Diels-Alder [4+2] Reactions.....	80
3.2	Project Aims.....	81
3.3	Results.....	81
3.3.1	Sporochartine Production.....	81
3.3.2	Time Course Study of Sporochartine Production.....	82
3.3.3	Acetate Feeding Experiment.....	83
3.3.4	<i>In Vitro</i> Spontaneous Diels-Alder Cycloaddition for Sporochartine.....	85
3.4	Discussion.....	87
3.5	Conclusion and Prospect.....	88
<b>4</b>	<b>Biosynthetic Studies of Trienylfuranol A.....</b>	<b>89</b>
4.1	Introduction.....	89
4.1.1	Depudecin.....	89
4.1.2	Aureonitol.....	91
4.1.3	Polyenoic Acid.....	92
4.1.4	Bacterial Ene-diyne.....	93
4.2	Project Aims.....	95
4.3	Results.....	96
4.3.1	Trienylfuranol A Production.....	96
4.3.2	Acetate Feeding Experiment.....	98
4.3.3	Mining of Polyketide Polyene hrPKS.....	101
4.3.3.1	Genome Screening for hrPKS.....	101
4.3.3.2	hrPKS Domain Analysis and Transcriptomic Analysis.....	101
4.3.4	Potential Polyene hrPKS BGC Analysis.....	104
4.3.4.1	'Cluster A' Analysis.....	104
4.3.4.2	'Cluster B' Analysis.....	106
4.3.5	Gene Knockout and Heterologous Expression of 'Cluster A'.....	108
4.3.5.1	Gene Knockout of <i>hspPKSI</i> .....	108

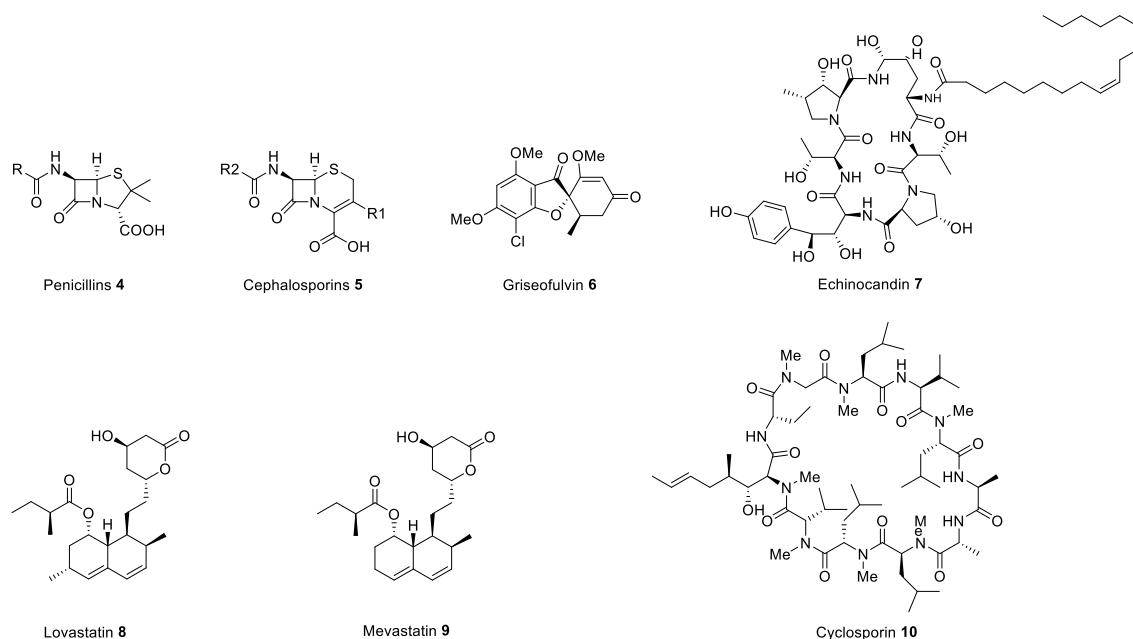
4.3.5.2	Heterologous Expression of <i>hmPKS1</i> and <i>hmCAR5</i> in <i>A. oryzae</i> .....	109
4.3.5.2a	Gene Combinations, Vector Constructions and <i>A. oryzae</i> NSAR1 Transformation.....	110
4.3.5.2b	Expression of <i>hmPKS1</i> in <i>A. oryzae</i> .....	111
4.3.5.2c	Co-expression of <i>hmPKS1</i> with <i>hmCAR5</i> .....	114
4.3.6	Heterologous Expression of ‘Cluster B’.....	114
4.3.6.1	Gene Cloning and Vector Constructions.....	114
4.3.6.2	Expression of <i>hmPKS2</i> and <i>hmPKS3</i> in <i>A. oryzae</i> .....	116
4.4	Discussion.....	118
4.4.1	Isotopic Labelling Study.....	118
4.4.2	Genome Mining and Transcriptomic Analysis.....	119
4.4.3	Biosynthetic Study of The hrPKS BGC ‘Cluster A’.....	119
4.4.4	Biosynthetic Study of The hrPKS BGC ‘Cluster B’.....	122
4.5	Conclusion and Prospect.....	126
<b>5</b>	<b>Overall Conclusion and Outlook</b> .....	<b>128</b>
<b>6</b>	<b>Experimental</b> .....	<b>131</b>
6.1	Biology.....	131
6.1.1	DNA and RNA Extraction and Sequencing.....	131
6.1.2	Strains and Transformation.....	132
6.1.2.1	<i>E. coli</i> Transformation.....	133
6.1.2.2	<i>S. cerevisiae</i> Transformation (Yeast Homologous Recombination).....	133
6.1.2.3	<i>A. oryzae</i> Transformation.....	133
6.1.2.4	<i>Hypomontagnella spongiphila</i> Transformation.....	134
6.1.3	Primer Sets and Cloning.....	135
6.1.4	Components of Buffers, Solutions, Liquid Medium and Agar.....	136
6.2	Chemistry.....	138
6.2.1	Fermentation and Extraction of Compounds.....	138
6.2.1.1	Small Scale.....	138
6.2.1.2	Large Scale.....	139
6.2.2	Analytical LCMS.....	139
6.2.3	Preparative LCMS.....	139
6.2.4	HRMS.....	140
6.2.5	Nuclear Magnetic Resonance (NMR) Analysis.....	140
<b>References</b>	.....	<b>142</b>
<b>7</b>	<b>Appendix</b> .....	<b>149</b>

# 1 Introduction

## 1.1 Natural Products from Fungi

Fungi represent an incredibly rich and rather overlooked reservoir of natural products. Research on fungal metabolites dates back to the 1870s, when pigments synthesized in conspicuous mushroom fruiting bodies attracted the attention of organic chemists. The 20th century witnessed the discovery, isolation and chemical characterization of a vast diversity of natural products from fungi.<sup>1</sup> At the same time, the variety of fungal species and the diversity of their habitats, allow the conclusion that fungi continue to be a rich source of new metabolites.

Fungi produce a wide variety of molecules referred to as secondary metabolites (SM), *e.g.*, polyketides, non-ribosomal peptides, terpenes and alkaloids.<sup>2</sup> While not directly involved in fundamental metabolic processes of growth and energy generation, SM display an array of biological activities that contribute to the survival of the producing organism in an occupied ecological niche, such as mediating communication within one species or between different species defence against competitors, nutrient acquisition, and even symbiotic interactions.<sup>3</sup>



**Figure 1.1** Examples of fungal secondary metabolites.

Not only does the role of SM make them interesting to study, but many SM, including penicillins, statins, and cyclosporins, have been found to have medical applications.<sup>4</sup> Among fungal natural products, particular interest is given to antimicrobials, due to the reduction in effectiveness of existing antibiotics used to treat bacterial infections, which is seen as a major threat to global

health security.<sup>5</sup> Penicillins **4** (Figure 1.1) and cephalosporins **5** are  $\beta$ -lactam antibiotics and represent the most widely used antimicrobials in the world: cephalosporins accounting for 28% and penicillins for 19% of the global market of antibiotics in 2009.<sup>6</sup> Some antifungals are also produced by fungi, like griseofulvin **6** and echinocandin **7**, which have been used as medicines.<sup>7</sup> Another critical medical application that fungal SM are known for is that of cholesterol-lowering agents, such as lovastatin **8** which is primarily produced by *Aspergillus terreus*, and mevastatin **9** found in *Penicillium citrinum*.<sup>8,9</sup> Fungi could also produce some other SM with immunosuppressant activity. A typical example is the non-ribosomal peptide (NRP) cyclosporin **10** produced by *Tolypocladium inflatum* and widely used to avoid organ rejection in transplant surgery.<sup>10</sup>

### 1.1.1 Fatty Acid Biosynthesis

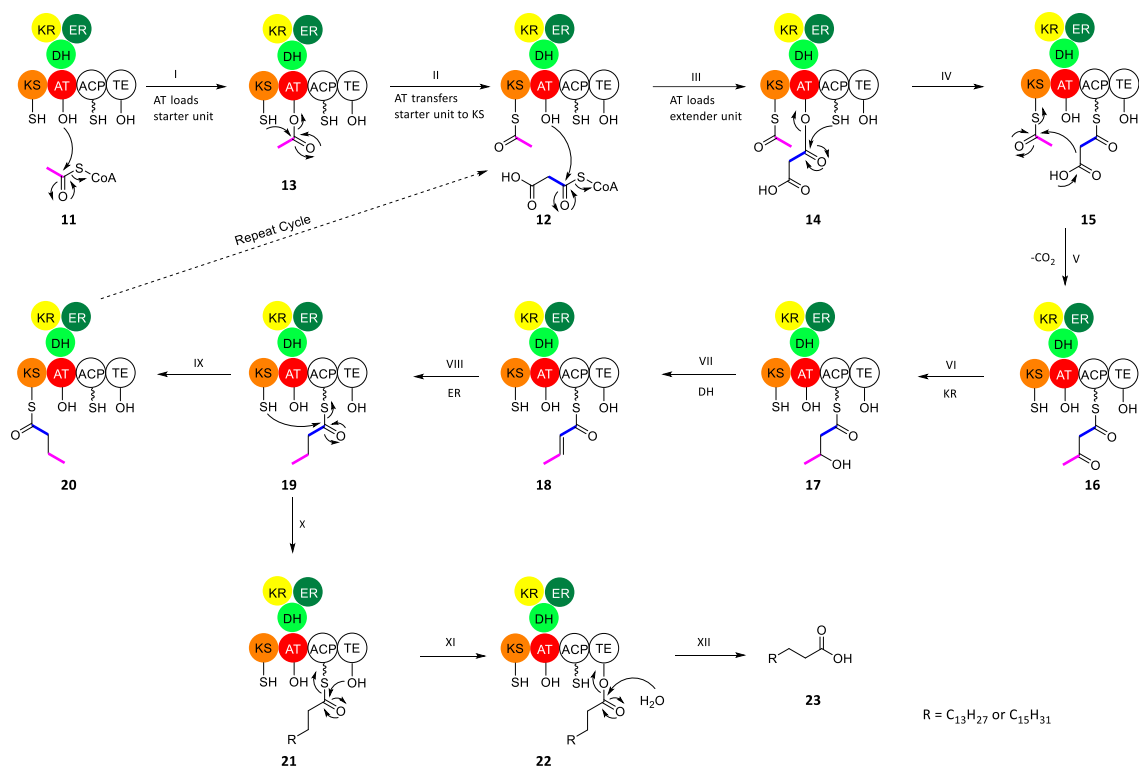
Fatty acid biosynthesis is a central metabolic pathway that entails the iterative elongation of fatty acid chains through a set of chemical reactions conserved in all kingdoms of life. Despite the fundamentally different FAS architectures of bacteria, plants, fungi and vertebrates, all of them integrate all necessary enzymatic activities together with acyl carrier protein (ACP) domains used for covalent substrate shuttling from one active site to the other.<sup>11,12</sup>

The acetyl primer **11** and malonyl elongation substrates **12** are loaded from coenzyme A (CoA) to ACP by acetyltransferase (AT) and malonyl/palmitoyl transferase (MPT) and condensed to acetoacetyl-ACP in a decarboxylative reaction catalyzed by ketoacyl synthase (KS) (Scheme 1.1, steps I – V, intermediates **13** – **16**). In three subsequent reaction steps VI – VIII (intermediates **17** – **19**), the  $\beta$ -carbon groups are processed by ketoacyl reductase (KR), dehydratase (DH), and enoyl reductase (ER), which results in fully saturated acyl-ACP **19** that can serve directly as a primer (Step IX, intermediate **20**) for the next condensation reaction. In each reaction cycle, the growing acyl-chain is elongated by two carbon units until it reaches a length of 16 to 18 carbon atoms **21**. The fully saturated carbon backbone is then released as a free acid by a thiolesterase (TE) domain (Scheme 1.1, steps X – XII, intermediates **21** – **23**).<sup>11</sup>

Fatty acid synthases (FAS) are classified into two main groups: type I and type II systems. Type I FAS are multi-domain proteins in which catalytic domains are covalently linked, which represent independent biosynthetic factories because they integrate all necessary enzymatic activities in one megasynthase protein.<sup>11</sup> Type I FAS are found in fungi and animals. In contrast, type II FAS are a complex of non-covalently linked mono-functional proteins, each enzymatic activity is catalyzed by a unique protein in the dissociated system. Type II FAS are found in bacteria, plants, and parasites. It is usually believed that type I FAS is a more efficient biosynthetic



machine because the enzymatic activities are fused into a single polypeptide and the intermediates do not diffuse from the complex.<sup>13</sup> However type II systems offer other advantages because the dissociated acyl ACP intermediates can react with many other cellular catalytic systems.

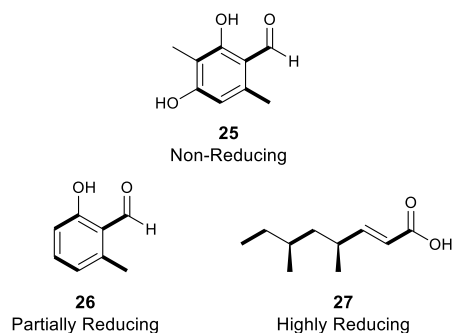


Early fungal FAS was encoded by a single gene, but it split into two separate FAS genes (subunits  $\alpha$  and  $\beta$ ) during fungal evolution (Figure 1.2).<sup>11</sup> In the linear domain organization of modern yeast FAS (yFAS), subunit  $\beta$  contains the AT, ER, DH and the majority of the split MPT domain. The  $\alpha$ -chain comprises the ACP, KR, KS, PPTase (phosphopantetheinyltransferase) and smaller part of the MPT domain. The two chains assemble into a heterododecameric complex with a barrel shape. Three full sets of enzymatic domains for fatty acid biosynthesis are located in each of the two reaction chambers which are defined by the  $\beta$ -chain (Figure 1.2). Also, three mobile ACP domains are double tethered to the central hub and the reaction chamber walls (Figure 1.2). Functionally, ACP can be compared to a mobile arm, which supplies substrates to productive sites of an assembly line.<sup>11</sup>

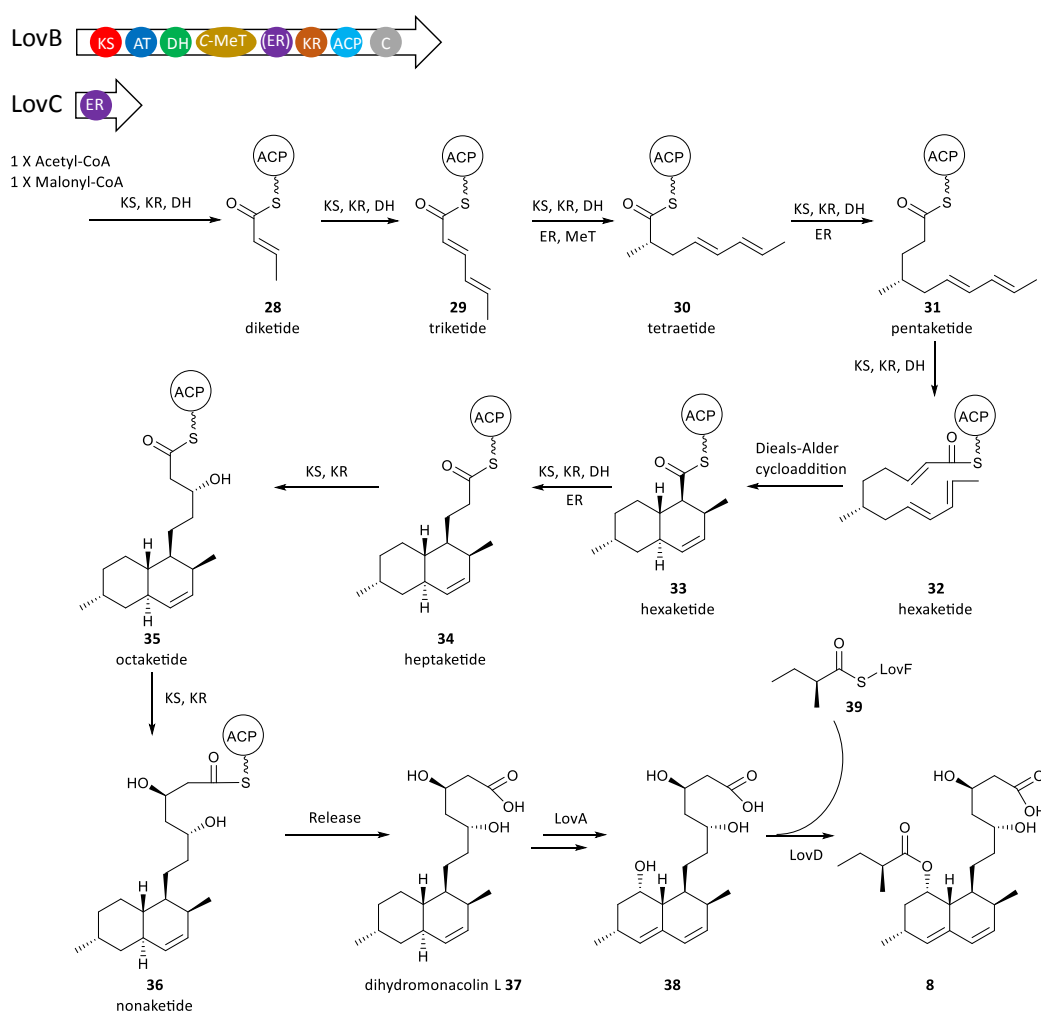
In the barrel-like architecture of yFAS, the enzymatic domains are architecturally arranged and concentrated to minimize diffusion distances of the consecutive step of a fatty acid synthesis cycle. Naturally occurring yFAS only efficiently produce a single type of product, saturated fatty acids.<sup>11</sup>



core nonaketide assembly. The ER domain in LovB is inactive, but a *trans*-acting ER (LovC) is functional and interacts with LovB. LovC accepts three intermediates (**30**, **31**, and **33**) and catalyses the reduction steps during the biosynthesis of lovastatin **8**. Another PKS (LovF) produces a diketide intermediate **39**, which is attached to **38** as the final step for lovastatin formation (Scheme 1.2).



**Figure 1.3** Fungal polyketides of nrPKS, prPKS and hrPKS.



**Scheme 1.2** The biosynthesis of lovastatin **8** as an example of iterative fungal hrPKS, adapted from Kennedy *et al.*, 1999.<sup>17,18</sup> ER domain in brackets means inactive.

## 1.2 The Biosynthesis of Maleidrides and Alkyl Citrates

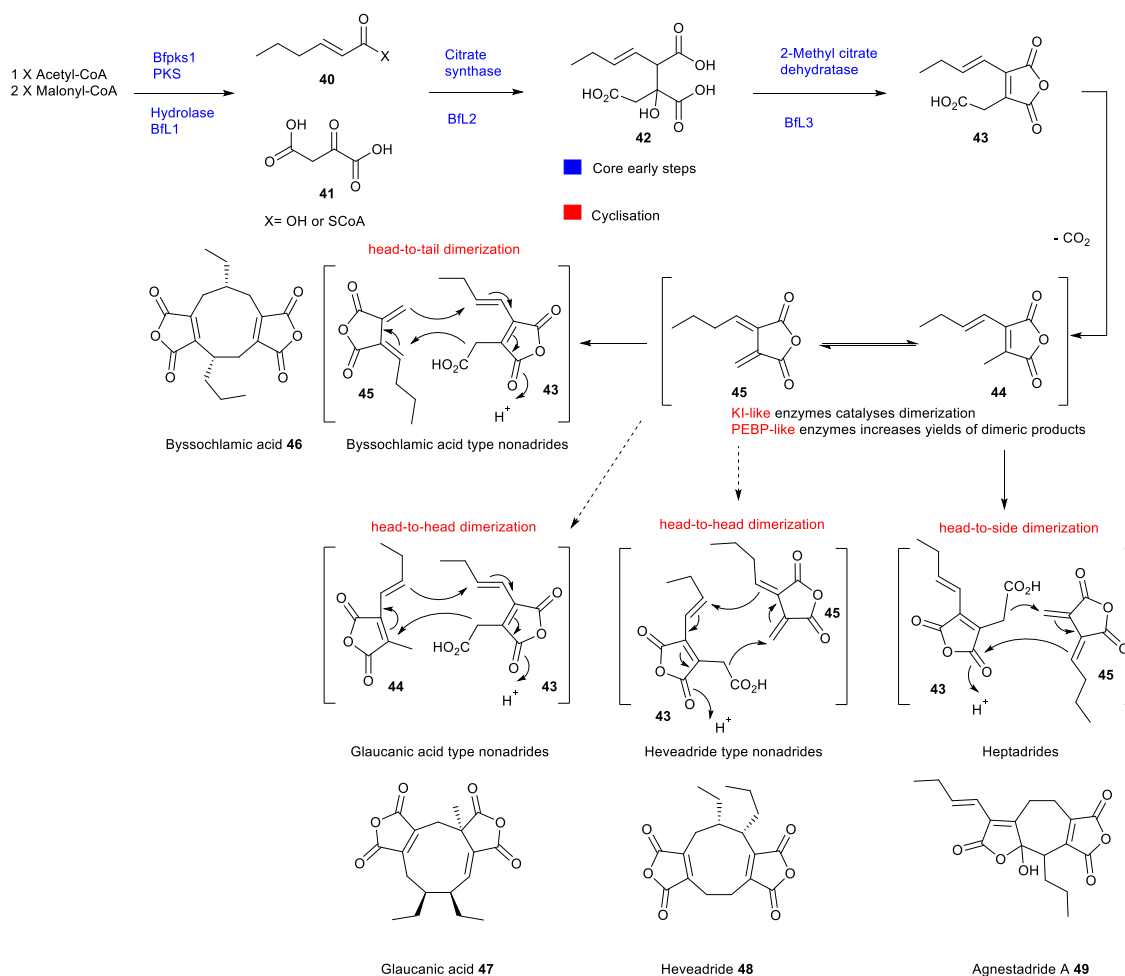
### 1.2.1 Maleidrides

Maleidrides are carbocyclic compounds with one or two maleic anhydride moieties. A well-studied example is byssochlamic acid **46**, which was first isolated from the fungus *Byssochlamys fulva*.<sup>19,20</sup> The Cox group<sup>21</sup> sequenced the genome of **46** producer *B. fulva*. Then bioinformatic analysis was performed to identify a likely maleidride BGC, which was validated by knockout (KO) and heterologous expression experiments.

The putative byssochlamic acid BGC contains four core genes encoding: a hrPKS (Bfpks1), a hydrolase (BfL1), a citrate synthase (BfL2) and a methylcitrate dehydratase (BfL3). Then the transcriptomic analysis of the organism under byssochlamic acid producing and non-producing conditions confirmed the maleidride BGC boundary.<sup>21</sup>

Various combinations of gene sets were constructed for heterologous expression experiments in *Aspergillus oryzae* NSAR1. Co-expression of the hrPKS *bfpks1*, citrate synthase *bfl2*, methylcitrate dehydratase *bfl3* and hydrolase *bfl1* led to the production of both **43** and **44** (Scheme 1.3).

Co-expression of the two ketosteroid isomerase (KI)-like genes (*bfl6* and *bfl10*) with the four core genes (*bfpks1*, *bfl1*, *bfl2*, and *bfl3*) led to the production of byssochlamic acid **46** and the decarboxylated intermediate **44**, as well as the low titre of agnestadride A **49** and the intermediate **43** (Scheme 1.3). It indicated the two KI-like genes catalyze the dimerization of monomers (**43** and **45**) to form more complicated scaffolds. However, heterologous expression studies showed that single use of either KI is not sufficient to catalyze any dimerization. More interestingly, experiments showed that the two PEBP (phosphatidylethanolamine-binding proteins) enzymes (BfL5 and BfL9) appear to be involved in the dimerization, because higher titres of byssochlamic acid **46** and heptadride **49** were observed when co-expression the PEBP genes with the four core genes and KI-like genes in *A. oryzae*.



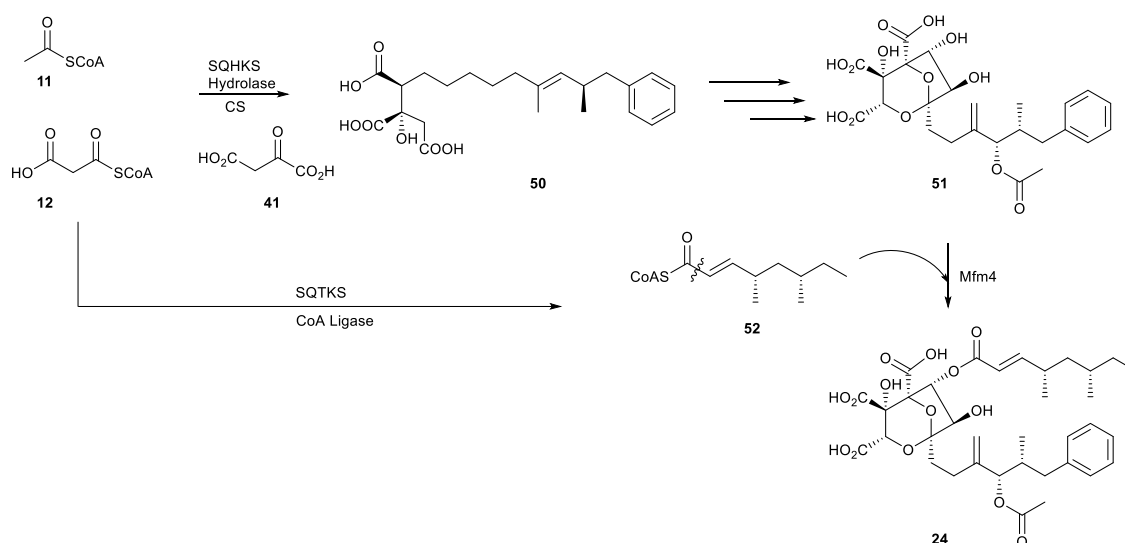
Scheme 1.3 The proposed biosynthesis of maleidrides by Williams *et al.*, 2016.<sup>21</sup>

## 1.2.2 Squalestatin

Squalestatin S1 **24**, isolated from *Phoma* sp., is a potent and selective inhibitor of squalene synthase.<sup>22</sup> For the biosynthesis research of squalestatin S1 **24** (Scheme 1.4), detailed molecular studies have revealed that a dedicated SQHKS (squalestatin hexaketide synthase) produces a carbon skeleton that is then condensed with oxaloacetate by citrate synthase (CS) to give an early alkyl citrate intermediate **50** that is further oxidatively processed to **51**, then **51** is coupled with a tetraketide **52** that is assembled by SQTKS (squalestatin tetraketide synthase), to afford squalestatin S1 **24**.<sup>14</sup>

Byssochlamic acid **46** and squalestatin S1 **24** share similar early steps in the biosynthetic pathways, such as the condensation of polyketide and oxaloacetate that catalysed by the key enzyme CS to produce alkyl citrate intermediates (Scheme 1.3 – 1.4). But the following tailoring steps are diverse. For instance, these reactions include dehydration, decarboxylation and

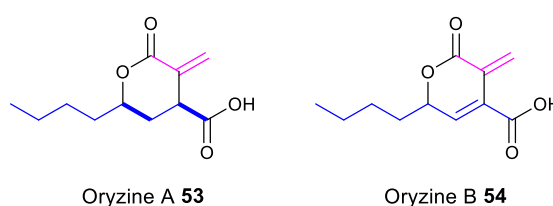
dimerization required for byssochlamic acid **46** formation, and multiple oxidations involved during squalestatin S1 **24** biosynthesis.



**Scheme 1.4** The proposed biosynthesis of squalestatin S1 **24**, adapted from Lebe *et al.*, 2019.<sup>14</sup>

### 1.2.3 Oryzine

Oryzines A **53** and B **54** (Figure 1.4) are two secondary metabolites that were isolated from *Aspergillus oryzae*.<sup>23</sup> They belong to the alkyl citrate type compounds based on the combination of a C<sub>8</sub> unit and a C<sub>3</sub> unit. Therefore, at least a CS is required to perform the key alkyl citrate backbone construction. Similar early step genes also exist in the byssochlamic acid **46** and squalestatin S1 **24** BGC (Scheme 1.3 – 1.4).



**Figure 1.4** Structures of oryzine A **53** and oryzine B **54**.

A putative oryzine gene cluster was found by CS (BfL2) homology search of *A. oryzae* genome. In summary, the gene cluster (Figure 1.5) encodes two fungal FAS subunits (*oryfasA* and *oryfasB*); a citrate synthase (*oryE*); a methylcitrate dehydratase (*oryR*); a decarboxylase (*oryM*); an alpha-ketoglutarate-dependent dioxygenase (*oryG*) and two lactonases (*oryH* and *oryL*). In addition, three transporters (*oryC*, *oryF*, *oryN*), a transcriptional regulator (*oryO*), a putative dehydrogenase (*oryD*), an acyl-CoA ligase (*oryP*) and a P450 (*oryQ*) can be found in this gene cluster. However,

no experimental evidence could be used to describe the oryzine biosynthetic pathway details before our work.<sup>23</sup>

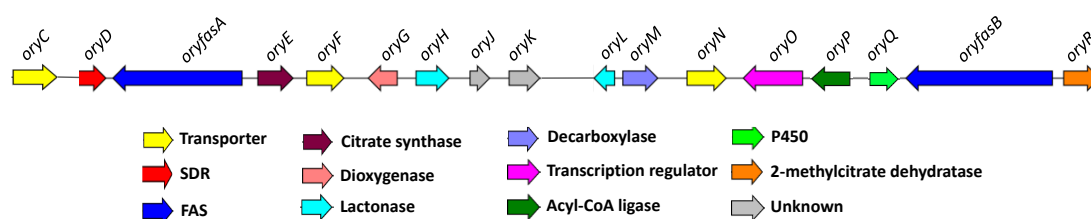
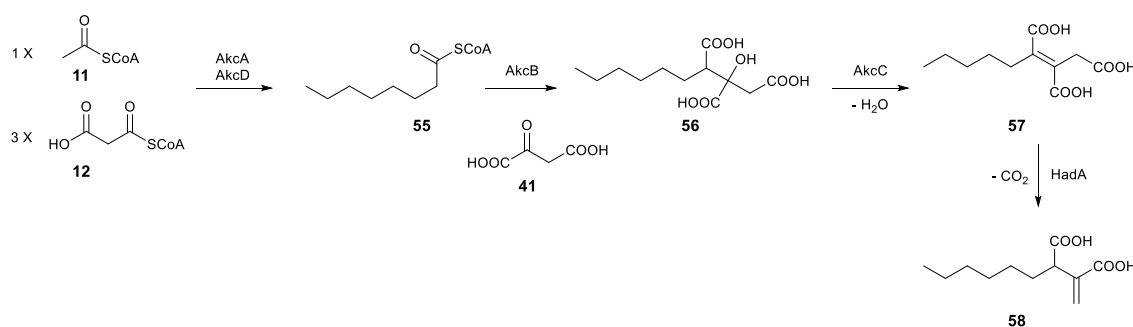


Figure 1.5 Putative oryzine BGC from *A. oryzae* RIB 40.<sup>23</sup>

### 1.2.4 Hexylcitric Acid Derivatives

Recently, another alkyl citrate gene cluster was identified from the filamentous fungus *Aspergillus niger* by bioinformatic analysis.<sup>24</sup> This BGC encodes FAS subunit alpha (*akcA*), FAS subunit beta (*akcD*), citrate synthase (*akcB*), transcriptional regulator (*akcR*), 2-methylcitrate dehydratase (*akcC*) as well as other co-localized functional genes.<sup>24</sup> Eleven hexylcitric acids were generated at  $\text{g}\cdot\text{L}^{-1}$  level through the overexpression of the transcriptional regulator *akcR* and the hexylaconitic acid decarboxylase gene *hadA* (outside the alkyl citrate gene cluster), which are defined as artificial production (with genetic manipulation) compared with the previously reported titre ( $\text{mg}\cdot\text{L}^{-1}$  level) of natural production.

The early steps in the proposed pathway (Scheme 1.5) are similar to the early steps of byssochlamic acid **46** and squalestatin S1 **24** biosynthesis (Scheme 1.3 – 1.4). Hexylcitric acid **56** is generated by the condensation of fatty acid unit **55** and oxaloacetate **41**. Normally, a dedicated citrate synthase catalyses this reaction, the AkcB takes the role in this pathway. The next step is the dehydration by 2-methylcitrate dehydratase homolog (AkcC) to make **57**, then decarboxylation by HadA to produce **58**. However, the stereochemical courses of citrate synthase (AkcB) and 2-methylcitrate dehydratase (AkcC) are unknown in the proposed pathway due to the lack of NMR spectra evidence of these intermediates.<sup>24</sup>



Scheme 1.5 Proposed biosynthetic pathway for the hexylcitric acid and hexylaconitic acid by Palys *et al.*, 2020.<sup>24</sup>

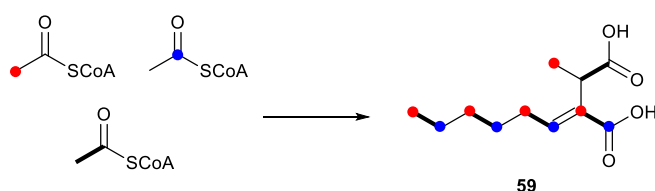
### 1.2.5 Piliformic Acid

Piliformic acid **59** was isolated as a secondary metabolite from several closely related fungi of xylariaceous genera.<sup>25,26</sup> Piliformic acid **59** requires a C<sub>8</sub> unit (octanoate chain) and a C<sub>3</sub> moiety for the construction of the scaffold.

The octanoate chain could be assembled from a FAS. Alternatively, a dedicated PKS may build a carbon chain for secondary metabolite biosynthesis. Extensive isotopic labelling investigations were carried out by the O'Hagan group.<sup>27</sup> Through the observations of the stereochemical location of deuterium in the octanoate chain, the origin of the C<sub>8</sub> unit was concluded to be from a FAS. This can be determined because the stereochemical course of ER in fungal FAS (*pro-R* labelled) and fungal PKS (*pro-S* labelled) are opposite.<sup>27</sup>

However, whether the octanoate is biosynthesised *de novo* for secondary metabolism, which means from a specific short-chain FAS, or possibly octanoate is a result of  $\beta$ -oxidation of the higher fatty acids which are synthesised by a FAS of primary metabolism was not known. O'Hagan group designed the labelling experiment by supplementing isotopically labelled [3-<sup>13</sup>C] decanoate.<sup>28</sup> Then, the incorporation patterns of the carbons from the C-1 acetate were observed, which could be well explained that the  $\beta$ -oxidation of the labelled material yields [1-<sup>13</sup>C] acetate and then incorporation *de novo* for piliformic acid **59** synthesis (Scheme 1.6). In the contrast, the only [1-<sup>13</sup>C] octanoate generated from the  $\beta$ -oxidation of [3-<sup>13</sup>C] decanoate was not observed. These results nicely proved the C<sub>8</sub> unit of piliformic acid is derived from a FAS with a sole function of octanoate production for **59** biosynthesis.

In addition, the C<sub>3</sub> unit is indicated to be derived from the citric acid cycle intermediate oxaloacetate by the efficient incorporation of labelled succinate.<sup>27</sup> In a word, the manner of natural products biosynthesis study before the genomic age was achieved predominately by utilizing isotopic labelling.



Scheme 1.6 The labelling patterns for piliformic acid **59**.<sup>27,28</sup>

Based on the knowledge of byssochlamic acid **46** and squalestatin S1 **24** biosynthesis, the condensation of the C<sub>8</sub> and C<sub>3</sub> moieties in piliformic acid **59** is probably also catalysed by a



dedicated CS. However, the dedicated BGC and biosynthetic pathway for piliformic acid were still unknown before our work.

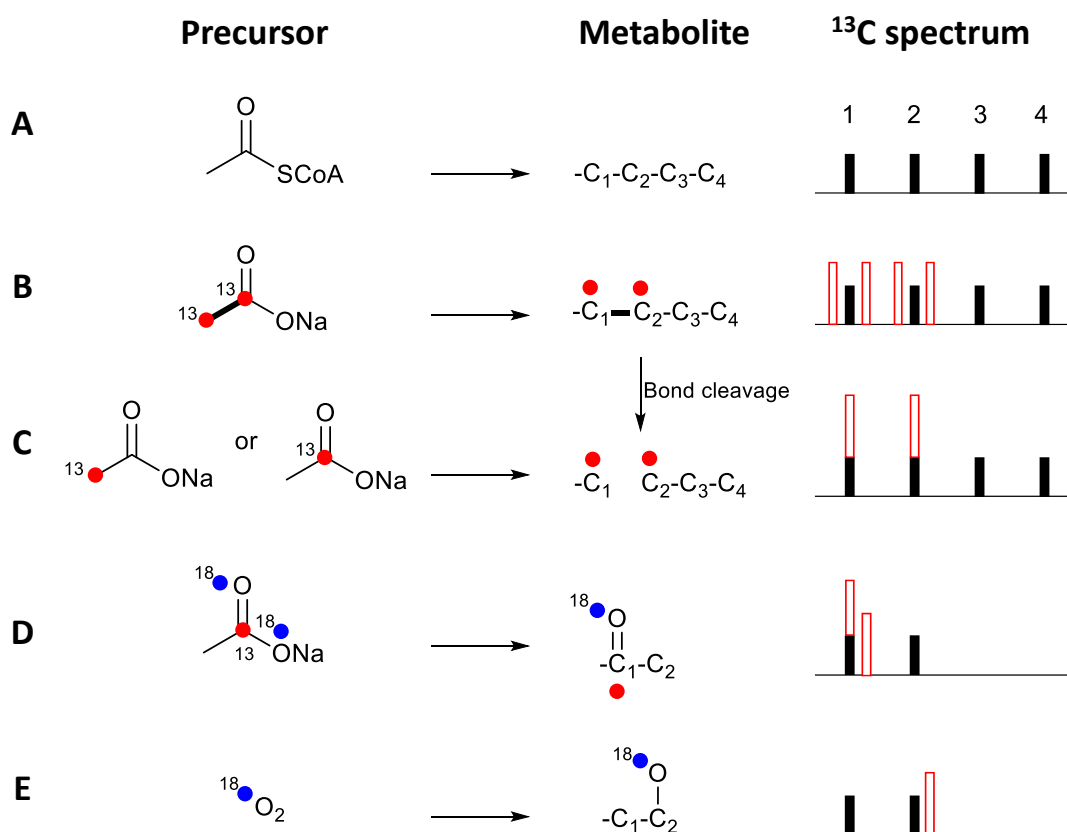
### 1.3 Techniques Used in Fungal Biosynthesis Investigations

#### 1.3.1 Isotopic Labelling

The long and successful history of isotopic labelling experiments has tremendously changed and deepened our understanding of natural products biosynthesis. Even now, it continues to provide important insights into the biosynthetic pathways of secondary metabolites.<sup>29–31</sup>

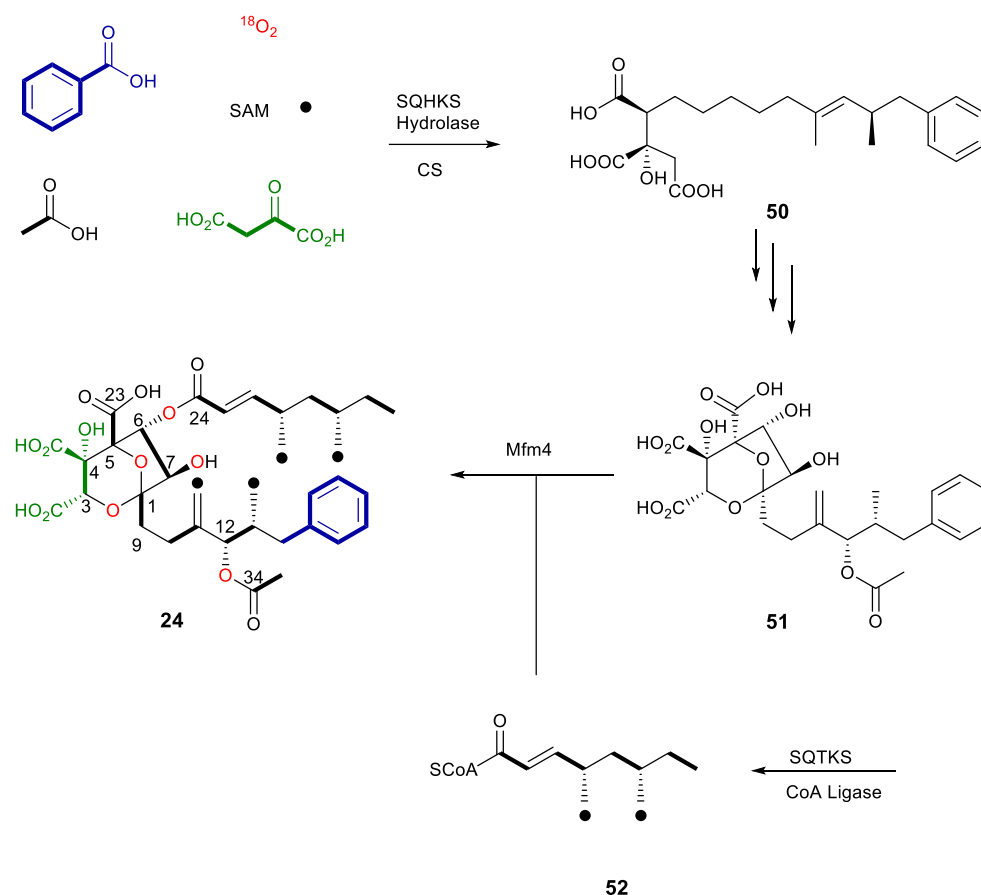
The  $^{13}\text{C}$  incorporations are useful for understanding fatty acids and polyketides biosynthesis. Incorporation of the double-labelled acetate ( $[1,2-^{13}\text{C}_2]$ ) into the metabolites, a  $^{13}\text{C}$ - $^{13}\text{C}$  coupling will be observed in the  $^{13}\text{C}$  NMR if the labelled acetate bond remains intact throughout the biosynthetic pathway. And the  $^1J_{\text{CC}}$  coupling value at C-1 and C-2 are identical (Figure 1.6B). However, the intensity of the carbon signal will be enhanced when incorporating the single-labelled acetate ( $[1-^{13}\text{C}]$ ,  $[2-^{13}\text{C}]$ ) into the metabolites, as well as when the incorporation of the double-labelled acetate bond break, because of an increase of the  $^{13}\text{C}$  content at a particular carbon (Figure 1.6C).<sup>32</sup>

Moreover, elucidating the origins of hydrogen and oxygen are also vital to understand the mechanism of chemical steps involved in biosynthesis. When incorporation of  $^{18}\text{O}$  is achieved by using a precursor where the  $^{18}\text{O}$  is directly linked to a  $^{13}\text{C}$  ( $[1-^{13}\text{C}, ^{18}\text{O}_2]$  acetate) or by growth in an  $^{18}\text{O}_2$  environment, the presence of the  $^{18}\text{O}$  alpha to  $^{13}\text{C}$  can be detected in the  $^{13}\text{C}$  NMR spectrum by an upfield shift of the signals (Figure 1.6D – E). The substitution of a deuterium alpha or beta to  $^{13}\text{C}$  can also result in a similar upfield shift in  $^{13}\text{C}$  NMR, but additional with the multiplet from the spin-spin coupling ( $^1J_{\text{CD}}$ ).<sup>32</sup>



**Figure 1.6** Illustration of feeding labelled precursors and subsequent signal of incorporated atoms in <sup>13</sup>C NMR spectra, adapted from Simpson, 1987.<sup>32</sup>

A good example of isotopic labelling practice is in the study of squalastatin S1 **24** biosynthesis.<sup>14,33</sup> First, <sup>13</sup>C labelled acetate feeding experiment showed the origin of 1 polyketides, from a benzoate-primed hexaketide and a dimethylated tetraketide (Scheme 1.7).<sup>33</sup> Then the incorporation of <sup>18</sup>O ([1-<sup>13</sup>C, <sup>18</sup>O<sub>2</sub>] acetate or <sup>18</sup>O<sub>2</sub> as a precursor) in squalastatin S1 **24** revealed that oxygen atoms (red colour) at C-1, C-3, C-5, C-6, C-7 and C-12 are inserted by molecular oxygen not acetate, therefore a few oxidation steps are anticipated to be involved in the biosynthetic pathway. In contrast, the C-24 and C-34 carbonyl oxygen atoms are derived from acetate.<sup>33</sup>



**Scheme 1.7** The labelling studies of squalenolide S1 **24**, adapted from Jones *et al.*, 1992.<sup>33</sup>

### 1.3.2 Genome Sequencing

The development of genome sequencing accelerates the biosynthetic study of microbial natural products. Especially, researchers realize the potential and capability of microorganisms to produce more amounts of SM than previously acknowledged.<sup>2,34</sup>

For example, Galagan and co-workers<sup>35</sup> reported a high-quality draft sequence (approximately 40-megabases) of the *Neurospora crassa* genome in 2003. Notably, this is the first fungal genome that was sequenced. Since then, the ‘1000 Fungal Genome Project’ funded by the Joint Genome Institute (JGI) to sequence 1000 fungal genomes from across the Fungal Tree of Life.<sup>36</sup> This project has significantly increased the amount of sequenced fungal genomes.

Recently, the Cox group<sup>37</sup> sequenced the genome of thirteen taxonomically well-defined fungi from Hypoxylaceae (Xylariales, Ascomycota) family and one Xylariaceae by using combinations of Illumina and Oxford nanopore technologies or PacBio sequencing. These high quality genome sequences not only satisfy the taxonomic purposes in mycology but also provide opportunities for the study of fungal evolution, host-fungus interactions, as well as the biosynthesis of secondary

metabolites. For instance, more than 750 biosynthesis gene clusters have been found from the thirteen sequenced genomes. These gene clusters include various types of secondary metabolic pathways, such as polyketides, terpenes, peptides, meroterpenoids and alkaloids. And the discovery and characterization of cytochalasan<sup>38</sup> and azaphilone<sup>39</sup> gene clusters resulted from these genome sequences.

### 1.3.3 Gene Knockout

Targeted knockout (KO) can result in two outcomes. Either the KO with the designed target, or ectopic intergration occurs elsewhere in the genome (Figure 1.7). Both possibilities result in incorporation of the selectable marker. Often ectopic intergration greatly exceeds targeted incorporation, meaning that tedious screening is required to find the desired KO transformant.<sup>40</sup>

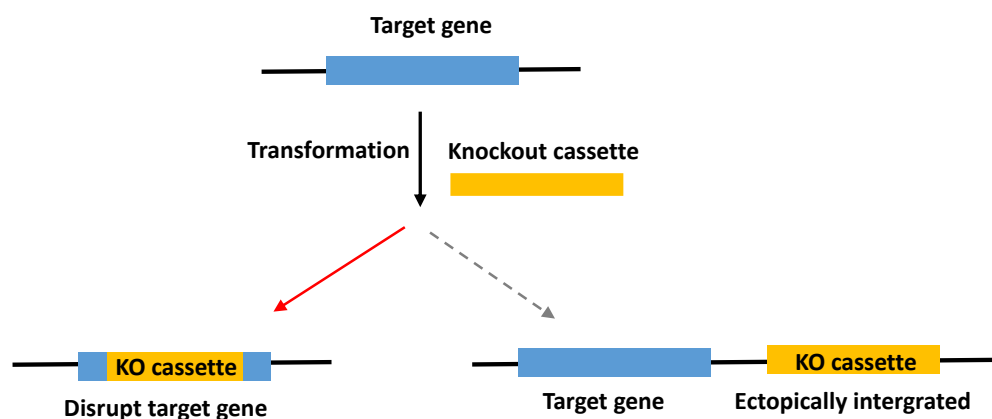


Figure 1.7 Integration of a knockout cassette.

Fairhead and co-workers<sup>41</sup> developed a split-marker technology, also called the bipartite method, to overcome this problem for *S. cerevisiae*. This method has also been applied successfully to diverse filamentous fungi.

Split-marker technology requires a mixture of two DNA fragments comprising overlapping sequences of a selectable marker gene. Only by homologous recombination of three crossing-over events can generate a functional marker gene, which allows producing an intact gene targeting cassettes for gene substitution in fungal transformation. These two fragments can be easily obtained from PCR amplification (Figure 1.8).<sup>42</sup>

Nielsen and co-workers<sup>42</sup> showed that bipartite knockout results in a higher frequency of correct targeting events compared to that classical transformation of a continuous gene targeting cassettes. The bipartite method is extremely flexible and can be easily applied in genome manipulations, like promoter replacements and GFP tagging. However, this technology also has a disadvantage.

Compared with other traditional transformation methods, the split-marker systems (bipartite method) dramatically reduces the frequency of fungal transformation. For example in *Magnaporthe grisea*, Jeong and co-workers<sup>43</sup> showed that the number of transformant obtained from the split marker was smaller than classical methods under the same transformation condition, which is >120 and five to 20 respectively. Although the frequency was reduced by about 85 – 96%, they could get two positive transformants with target gene-substitution from bipartite knockout, but none from the traditional approach.

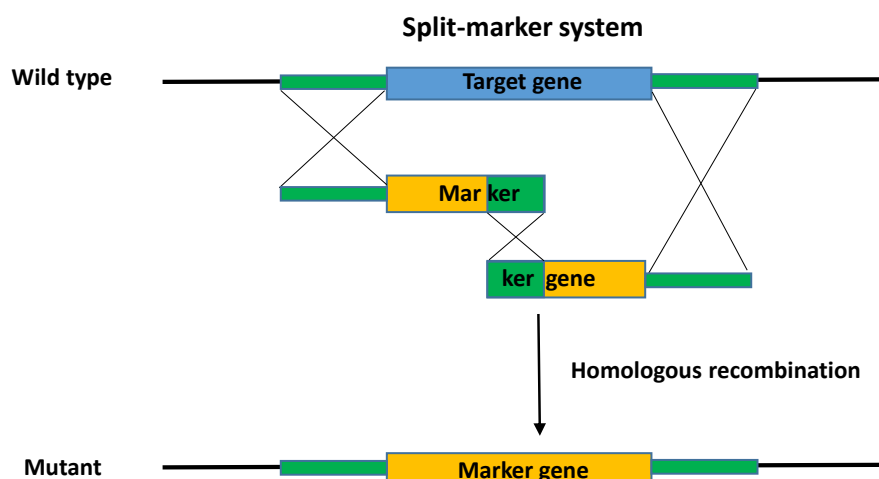
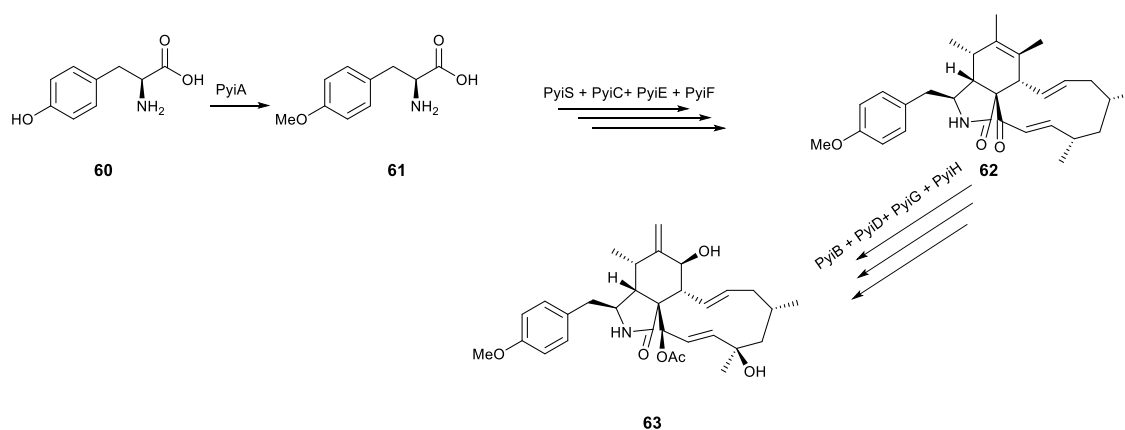


Figure 1.8 Strategy of bipartite knockout, adapted from Nielsen *et al.*, 2006.<sup>42</sup>

Also in our group, there are successful examples of gene disruption by using the bipartite method, for example in the biosynthetic investigation of cytochalasan H **63** in *Magnaporthe grisea* (Scheme 1.8).<sup>38</sup> Through the gene inactivation of the functional genes (*O*-methyltransferase, *trans*-enoyl reductase, *O*-acetyltransferase, oxidoreductase and P450, respectively), the late-stage biosynthetic pathway of **63** was fully elucidated and reveals that *O*-methyltyrosine **61** is the true precursor for **63**.



Scheme 1.8 The proposed biosynthesis of cytochalasan H **63** by Wang *et al.*, 2019.<sup>38</sup>

### 1.3.4 Heterologous Expression

Heterologous expression not only can be used to discover the secondary metabolite pathways but also enable the improvement of the yields of natural products. Also, this strategy provides the opportunity in synthetic biology study to produce novel compounds. A good heterologous host features two factors at least: easy cultivation and genetic manipulation. For example, both bacteria (*E. coli* and *Streptomyces* sp) and fungi (*Aspergillus oryzae*, *Aspergillus nidulans*, and *Saccharomyces cerevisiae*) are used as heterologous expression platforms.<sup>44</sup>

For instance, the first successful heterologous expression of the polycyclic tetramate macrolactam (PTM) ikarugamycin BGC in *E. coli* opened one way to investigate cryptic iPKS/NRPS biosynthetic pathways found in other bacteria.<sup>45</sup> However, it is not possible for bacteria to process eukaryotic introns and bacteria often possess a significant codon bias, therefore bacteria are not the ideal hosts for fungal gene expression.

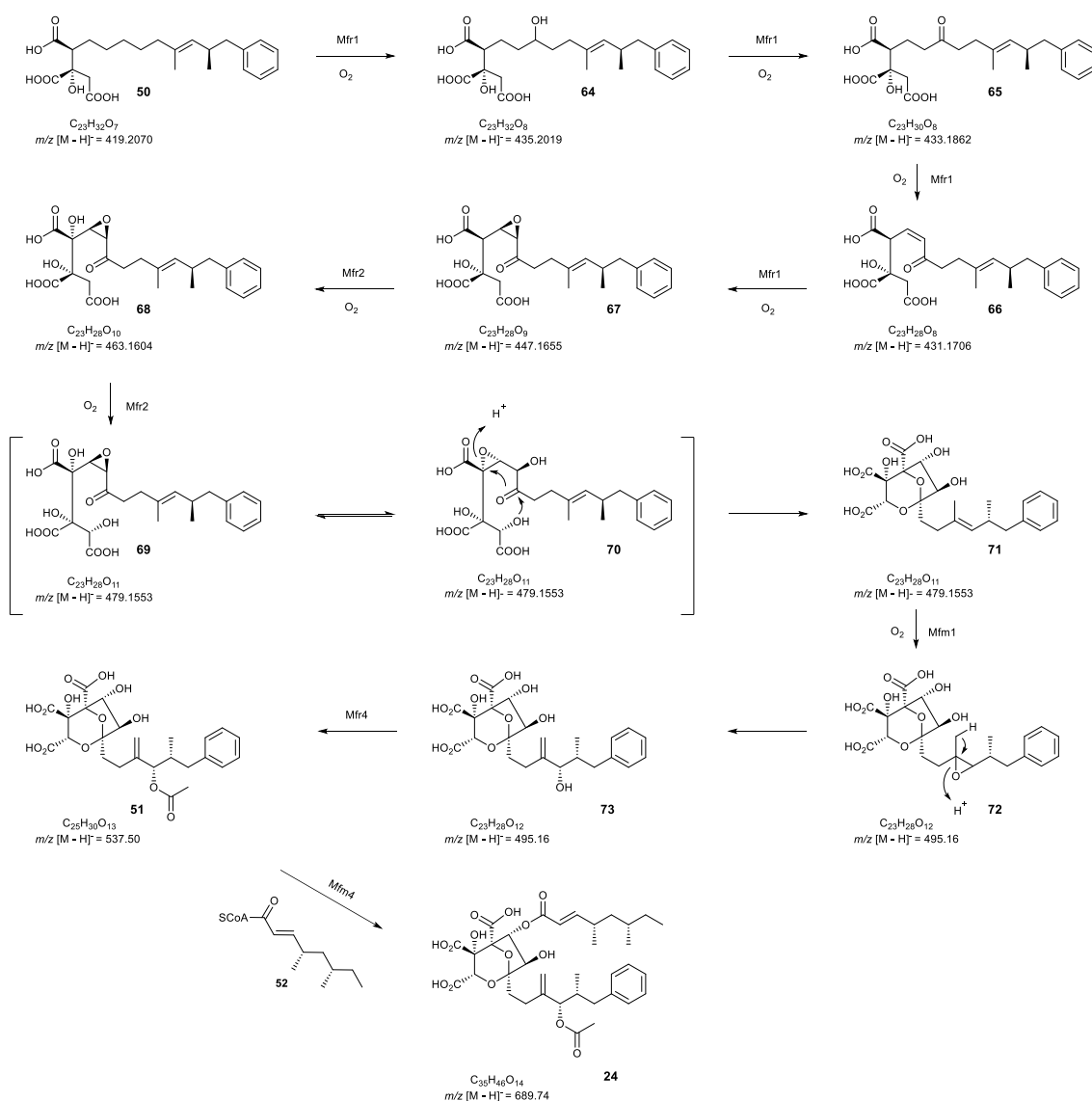
Yeast (*S. cerevisiae*) is a suitable host for some fungal BGC expression. Recently, the bostrycoidin (a red aza-anthraquinone pigment) gene cluster was successfully expressed in an engineered *S. cerevisiae*.<sup>46</sup> Firstly, the primary metabolism of the *S. cerevisiae* was optimized for higher flux towards the acetyl- and malonyl-CoA pathways which render a higher concentration of precursor for the bostrycoidin polyketide construction. In addition, a PPTase native to the original producer was cloned into the expression host and co-expressed with the PKS to activate the ACP domain, required for polyketide backbone assembly. Finally, the maximum titer (2.2 mg·L<sup>-1</sup>) of bostrycoidin production was achieved after 2 days of galactose induction.

The filamentous fungus *A. oryzae* has been utilized to express fungal gene clusters for many years, and there are many successful examples in our group. For instance, the biosynthetic pathway elucidation of the maleidride byssochlamic acid, the meroterpenoid xenovulene A, the polyketides squalestatin S1 and sorbicillinoids.<sup>14,21,47,48</sup> The well-established genetic manipulation approaches in *A. oryzae* make it a great success as a heterologous expression system to study secondary metabolite biosynthesis and production.<sup>49</sup>

Heterologous expression can also be used to solve tricky biosynthetic problems. For example, heterologous expression in *A. oryzae* was used to characterise the function of two putative nonheme-iron-dependent enzymes (Mfr1 and Mfr2) involved in the post tailoring steps of squalestatin S1 **24** biosynthesis (Scheme 1.9).<sup>14</sup> Co-expression of the oxidase (Mfr1), PKS (Sqhks), hydrolase (Mfm8) and citrate synthase (Mfr3) led to the observation of several oxidised congeners (**64** *m/z* 435 [M–H]<sup>-1</sup>, **65** *m/z* 433 [M–H]<sup>-1</sup>, **66** *m/z* 431 [M–H]<sup>-1</sup> and **67** *m/z* 447 [M–H]<sup>-1</sup>) in LCMS chromatogram. **50** is a substrate for the stepwise oxidations by Mfr1, first make

alcohol **64**, then ketone **65**, unsaturated ketone **66**, and further oxidise to give the epoxide **67**. When additionally expressed the oxidase Mfr2 to the system, Liquid Chromatography Mass Spectrometry (LCMS) results showed the oxidised metabolites of **68**  $m/z$  463  $[M - H]^{-1}$  and **71**  $m/z$  479  $[M - H]^{-1}$ . There is a proposed mechanism of acetal formation and epoxide opening during the conversion of **70** to **71**.

Subsequently, the copper-dependent oxygenase Mfm1 introduces a hydroxyl group (**73**) required for later acetylation to give **51** that serves as the substrate for the final acylation reaction catalysed by Mfm4 to afford squalestatin S1 **24** (Scheme 1.9).



Scheme 1.9 The proposed oxidative cascade during the biosynthesis of squalestatin S1 **24** by Lebe *et al.*, 2019.<sup>14</sup>

## 1.4 Overall Aims

The overall aim of this project is to focus on understanding the biosynthesis of alkyl citrate compounds such as sporothriolide **1** and polyketide polyenes such as trienylfuranol A **2**, as well as the adduct sporochartine **3**.

Extensive strategies of isotopic labelling, genome and transcriptome sequencing and bioinformatic analysis, gene knockout, heterologous expression and *in vitro* studies will be performed to delineate the biosynthetic pathway of sporothriolide and trienylfuranol A. In addition, the study will support the proposed biosynthesis of alkyl citrates of piliformic acid and oryzines.

For sporochartine which is proposed to be a consequence of Diels-Alder cycloaddition of sporothriolide and trienylfuranol A, the particular aim is to answer the question of whether it's an enzymatic (Diels-Alderase) catalysis result or a chemically spontaneous reaction.

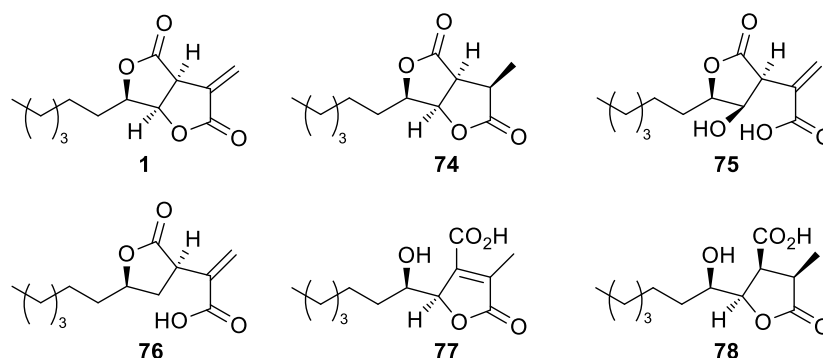


## 2 Biosynthetic Studies of Sporothriolides

### 2.1 Introduction

Sporothriolide **1** was first isolated and elucidated by Krohn and co-workers<sup>50</sup> from *Sporothrix* sp. (strain No. 700). Lately, the Stadler<sup>51</sup> group published their work about the secondary metabolites from *Hypoxylon monticulosum* MUCL 54604 (now referred to as *Hypomontagnella monticulosa*).<sup>37,52</sup> They found that this organism mainly produces sporothriolide **1** together with other analogues (Figure 2.1). Shortly after, the Ouazzani<sup>53</sup> group isolated sporothriolide and the structurally more complicated sporochartine from *Hypoxylon monticulosum* CLL 205 (now referred to as *Hypomontagnella spongiphila*) isolated from a marine sponge.<sup>37,52</sup> The crystal structure of sporothriolide **1** was first reported by the Ye<sup>54</sup> group, obtained from the endophyte *Nodulisporium* sp. A21.

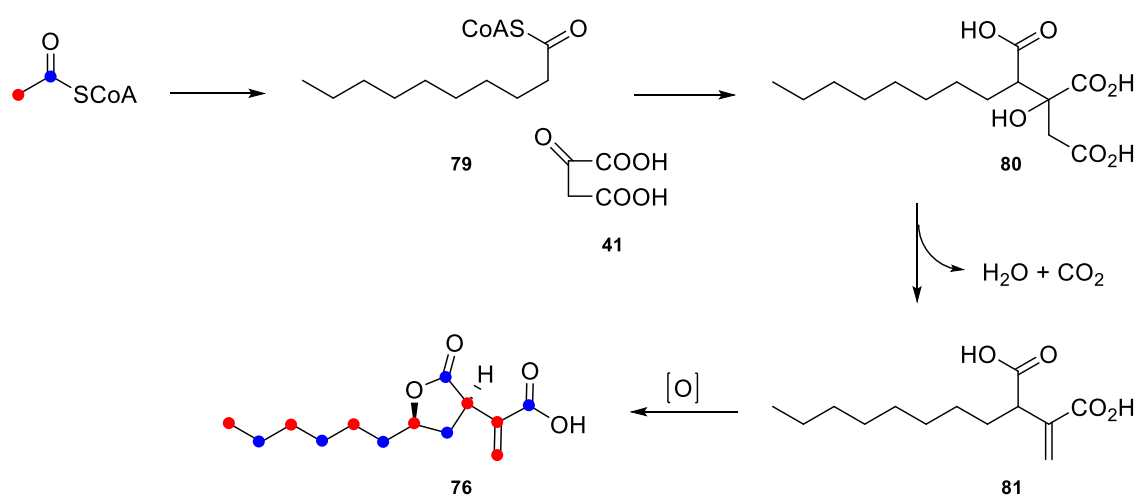
Bioactivity studies have shown that sporothriolide **1** is a potent antifungal agent with  $EC_{50}$  of  $11.6 \pm 0.8 \mu\text{M}$  against the phytopathogenic fungus *Rhizoctonia solani*,<sup>55</sup> while the  $EC_{50}$  of positive control carbendazim was  $9.6 \pm 0.7 \mu\text{M}$ .



**Figure 2.1** Structures of sporothriolide **1**, dihydrosporothriolide **74**, sporothric acid **75**, deoxysporothric acid **76**, isosporothric acid **77** and dihydroisosporothric acid **78**.

Sporothriolides probably belong to the alkyl citrate family of metabolites, similar to piliformic acid and the oryzines (Section 1.2.3 – 1.2.5). These compounds are generated from the condensation of fatty acids with a decarboxylated Krebs cycle intermediate, commonly oxaloacetate. Isotopic labelling experiments for deoxysporothric acid **76** using  $[1-^{13}\text{C}]$  and  $[2-^{13}\text{C}]$  acetate reported by the Ye<sup>54</sup> group showed that the labelling pattern is consistent with the hypothesis of a fatty acid or polyketide origin. A speculative biosynthetic pathway is shown in Scheme 2.1. First, the  $\text{C}_{10}$  fatty acid chain **79** is condensed with  $\text{C}_4$  oxaloacetate **41**, followed by decarboxylation and dehydration to result in intermediate **81**, that undergoes lactonization to form deoxysporothric acid **76**.

Although sporothriolide **1** has been known as potent antifungal agents for almost two decades, their detailed biosynthesis was still unknown before our work.



**Scheme 2.1** Labelling patterns and hypothetical biosynthetic pathways for deoxysporothric acid **76**, adapted from Cao *et al.*, 1999.<sup>54</sup>

## 2.2 Project Aims

Although the labelling patterns for deoxysporothric acid **76** have been reported, the labelling pattern of sporothriolide **1** was unknown. It will be interesting to compare the difference and similarity of these two metabolites. The biosynthetic gene cluster for sporothriolide is also unknown at the start of this project.

Genome sequencing of the producing fungi, supported by transcriptomic studies will be performed to predict the putative biosynthetic gene cluster (BGC) of sporothriolide. Transcriptomic analysis of *H. monticulosa* under sporothriolide producing and non-producing conditions will be conducted. Results will be used to identify intron positions and translational start and stop positions. Gene knockout will be used to further confirm the correct BGC for sporothriolide. Heterologous expression will be used to reconstitute the biosynthesis of sporothriolide in *A. oryzae*. Meanwhile, the function of interesting enzymes will be investigated by expression and *in vitro* assay.

## 2.3 Results

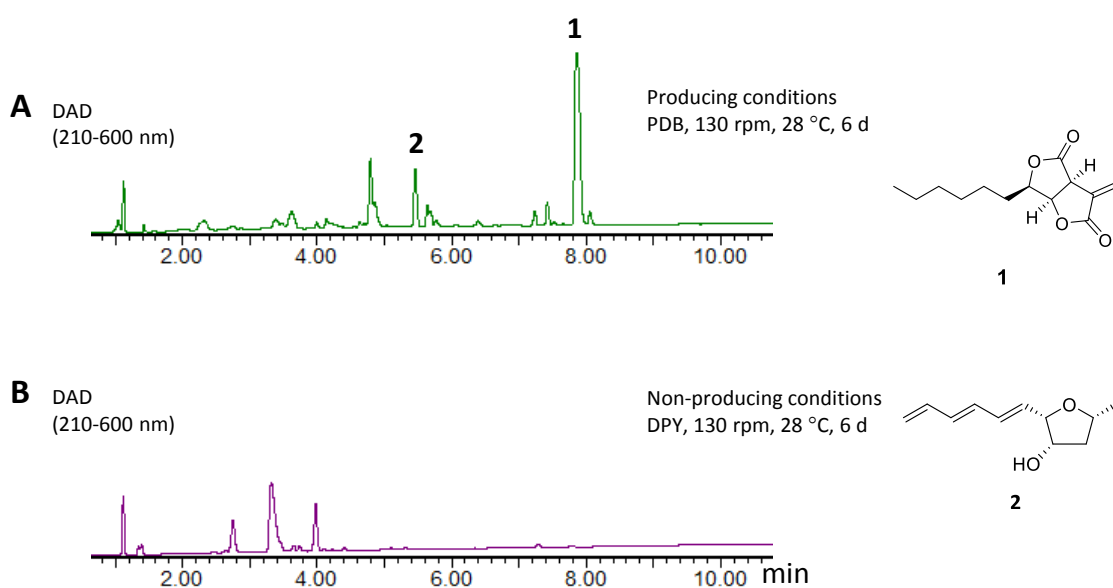
### 2.3.1 Sporothriolide Production from *H. monticulosa*, *H. spongiphila* and *H. submonticulosa*

*H. monticulosa* and *H. spongiphila* are known to produce sporothriolide.<sup>51,53</sup> Here, we cultivated *Hypoxylon submonticulosum* (now referred to as *Hypomontagnella submonticulosa*) and found that it was able to produce sporothriolide metabolites as effectively as the other two fungi (Section

4.3.1). *H. monticulosa*, *H. spongiphila*, and *H. submonticulosa* were obtained as gifts from the Stadler,<sup>51</sup> Ouazzani<sup>53</sup> and Sumarah<sup>56</sup> groups, respectively.

### 2.3.1.1 Producing and Non-Producing Conditions

In order to investigate the expression level of the biosynthetic gene cluster, producing and non-producing fermentation conditions were studied. The cultivation condition of *H. monticulosa* was originally reported in YMG medium (Table 6.5),<sup>51</sup> the *H. spongiphila* and *H. submonticulosa* were previously grown in PDB medium (Table 6.5).<sup>53,56</sup> The organisms were grown in various liquid media in shake culture. At the end of the fermentation, the cultures were extracted into ethyl acetate. The concentrated organic extracts were examined by LCMS. The difference between PDB and YMG fermentation conditions of *H. monticulosa* was compared in our experiment, but the chemical profile of the secondary metabolites from two cultivation media showed no obvious variation. Finally, we chose ‘PDB medium, 130 rpm, 28 °C, 6 day’ as sporothriolide producing conditions for *H. monticulosa* and *H. spongiphila* (Figure 2.2). And the *H. submonticulosa* producing conditions are ‘PDB medium, 100 rpm, 25 °C, 6 day’ (see LCMS chromatograph in Section 4.3.1).

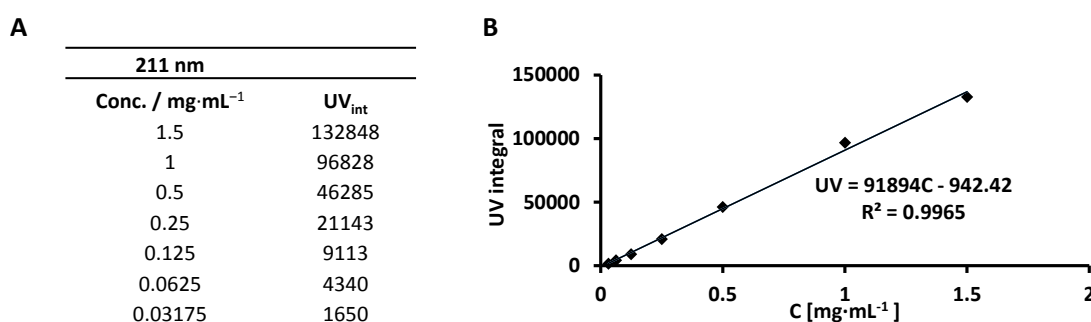


**Figure 2.2** Diode array detector (DAD) chromatograms of *H. monticulosa* MUCL 54604 extracts: **A**, under producing conditions (PDB medium, 130 rpm, 28 °C, 6 d); and **B**, under non-producing conditions (DPY medium, 130 rpm, 28 °C, 6 d).

We tested DPY, LB and MMK2 (Table 6.5) media as potential non-producing fermentation conditions. The results showed that no sporothriolide metabolites were produced in MMK2 and DPY media at least up to day 6 of the fermentation. Because more mycelia was harvested in DPY medium, ‘DPY medium, 130 rpm, 28 °C, 6 day’ was used as the non-producing conditions for *H. monticulosa* and *H. spongiphila* (Figure 2.2).

### 2.3.1.2 Time Course Study of Sporothriolide Production

In order to quantify the production of compounds, the calibration curve for sporothriolide **1** was made based on a UV integration method. A dilution series of **1** (1.5, 1.0, 0.5, 0.25, 0.125, 0.0625, 0.03175 mg·mL<sup>-1</sup>) was measured and the corresponding signals for the extracted wavelength at 211 nm ( $\lambda_{\max}$ ) were integrated (Figure 2.3A). Integrated values were then plotted against the sample concentration and fitted into a straight line (Figure 2.3B). The relationship is linear within the 0.03 – 1.5 mg·mL<sup>-1</sup> concentration range. The equation was applied to quantify **1** across different samples.



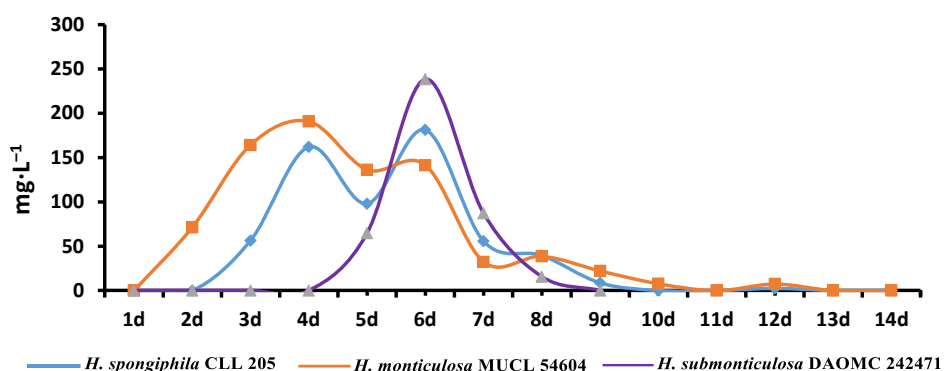
**Figure 2.3 A**, the integrated values of different dilutions of sporothriolide **1** solutions at 211 nm from LCMS analysis; **B**, calibration curve for sporothriolide **1** quantification.

Timecourse experiments for the three *Hypomontagnella* wild type strains were conducted to monitor the production of **1**. *H. monticulosa* and *H. spongiphila* were grown respectively in a flask containing 1 L PDB medium at 28 °C, 130 rpm for 14 days and 10 mL aliquots were taken daily. *H. submonticulosa* was grown in a flask containing 1 L PDB medium at 25 °C, 100 rpm for 9 days and 10 mL aliquots were taken daily. Aliquots were extracted with equal amounts of ethyl acetate and the organic phase was evaporated under vacuum (Section 6.2.1). Crude extracts were analysed by LCMS (Section 6.2.2) and compound titres were calculated using the previously described equation (Table 2.1). Production of **1** was visualized graphically (Figure 2.4).

The data shows that all three fungi can produce high amounts of sporothriolide **1**. The *H. spongiphila* and *H. submonticulosa* strains had the biggest production of **1** at the 6<sup>th</sup> day with 181 mg·L<sup>-1</sup> and 238 mg·L<sup>-1</sup>, respectively. However, the highest production amount of *H. monticulosa* was at the 4<sup>th</sup> day with 190 mg·L<sup>-1</sup>. All strains displayed a dramatic production decrease of sporothriolide at the 7<sup>th</sup> day, which indicated a fast degradation and a short life cycle of **1**. Meanwhile, sporothriolide **1** and cometabolites **74** – **78** were purified, characterised by NMR and compared to literature (Chapter 7).

	<i>H. monticulosa</i>	<i>H. spongiphila</i>	<i>H. submonticulosa</i>
Day	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>
1	0	0	0
2	71.3	0	0
3	164.2	56.4	0
4	190.7	162.1	0
5	136.2	98.0	64.5
6	141.5	181.3	238.3
7	32.4	55.6	87.0
8	38.7	39.6	15.6
9	22.2	8.6	0
10	7.7	0	-
11	0	0	-
12	7.4	2.4	-
13	0	0	-
14	0	0	-

**Table 2.1** Temporal production of sporothriolide **1** from *H. monticulosa* MUCL 54604, *H. spongiphila* CLL 205, and *H. submonticulosa* DAOMC 242471 cultivated under producing conditions.



**Figure 2.4** Kinetic production curve of sporothriolide **1**.

### 2.3.2 The Identification of Multiforisin H

During the investigation of metabolites from *H. spongiphila* CLL 205, we isolated a known compound multiforisin H **82** (Figure 2.5) which was initially found from Ascomycete *Gelasinospora* species.<sup>57</sup> The previous bioactivity study revealed multiforisin H **82** is a potent immunosuppressive agent, with IC<sub>50</sub> values of 1.8 ug·mL<sup>-1</sup> and 0.9 ug·mL<sup>-1</sup> against Con A- (T-cells) and LPS-induced proliferations of mouse splenic lymphocytes, respectively.<sup>57</sup>

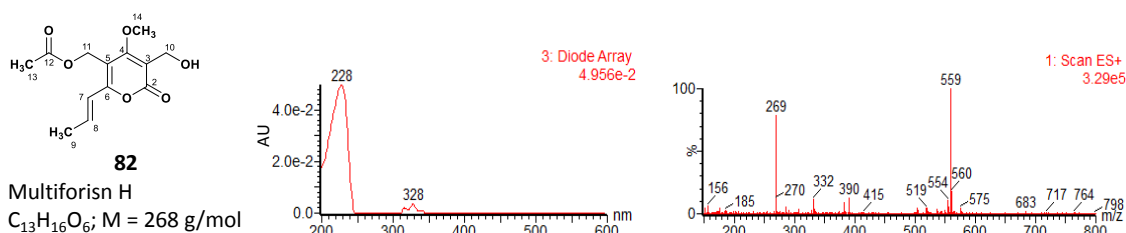
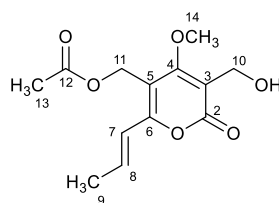


Figure 2.5 UV and mass spectra (ES<sup>+</sup>) of compounds **82**.

By analysing the <sup>1</sup>H and <sup>13</sup>C NMR spectra, we observed the mix signals of sporothriolide **1** and multiforisin H **82**. Then, manually omitted the sporothriolide **1** NMR chemical shifts, the rest signals on the NMR spectra (Figure 2.6 – 2.7) were well assigned to multiforisin H **82**. And the NMR, UV and mass data of multiforisin H **82** obtained from our experiments were fit well with the literature (Table 2.2, Figure 2.5).<sup>57</sup> Although it has been almost 20 years since multiforisin metabolites have been discovered from fungi, no biosynthetic information was known before our study.



Multiforisin H **82**

pos.	$\delta_C$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_C$ / ppm literature <sup>57</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>57</sup>
<b>2</b>	164.9	-	164.7	-
<b>3</b>	110.9	-	110.8	-
<b>4</b>	168.9	-	168.8	-
<b>5</b>	108.0	-	107.9	-
<b>6</b>	158.1	-	158.0	-
<b>7</b>	119.4	6.39, 1H, dq (15.3, 1.7)	119.3	6.39, 1H, dq (15.1, 1.7)
<b>8</b>	138.4	6.86, 1H, dq (15.3, 7.1)	138.2	6.86, 1H, dq (15.1, 7.0)
<b>9</b>	19.1	1.96, 3H, dd (7.1, 1.7)	18.9	1.97, 3H, dd (7.0, 1.7)
<b>10</b>	56.0	4.58, 2H, s	55.9	4.59, 2H, s
<b>11</b>	56.3	4.96, 2H, s	56.2	4.97, 2H, s
<b>12</b>	170.9	-	170.8	-
<b>13</b>	21.0	2.07, 3H, s	20.9	2.08, 3H, s
<b>14</b>	63.3	4.07, 3H, s	63.2	4.07, 3H, s

Table 2.2 <sup>1</sup>H NMR (400 MHz) data and <sup>13</sup>C NMR (100 MHz) data for **82** in CDCl<sub>3</sub>. Literature<sup>57</sup> data were measured in CDCl<sub>3</sub>.

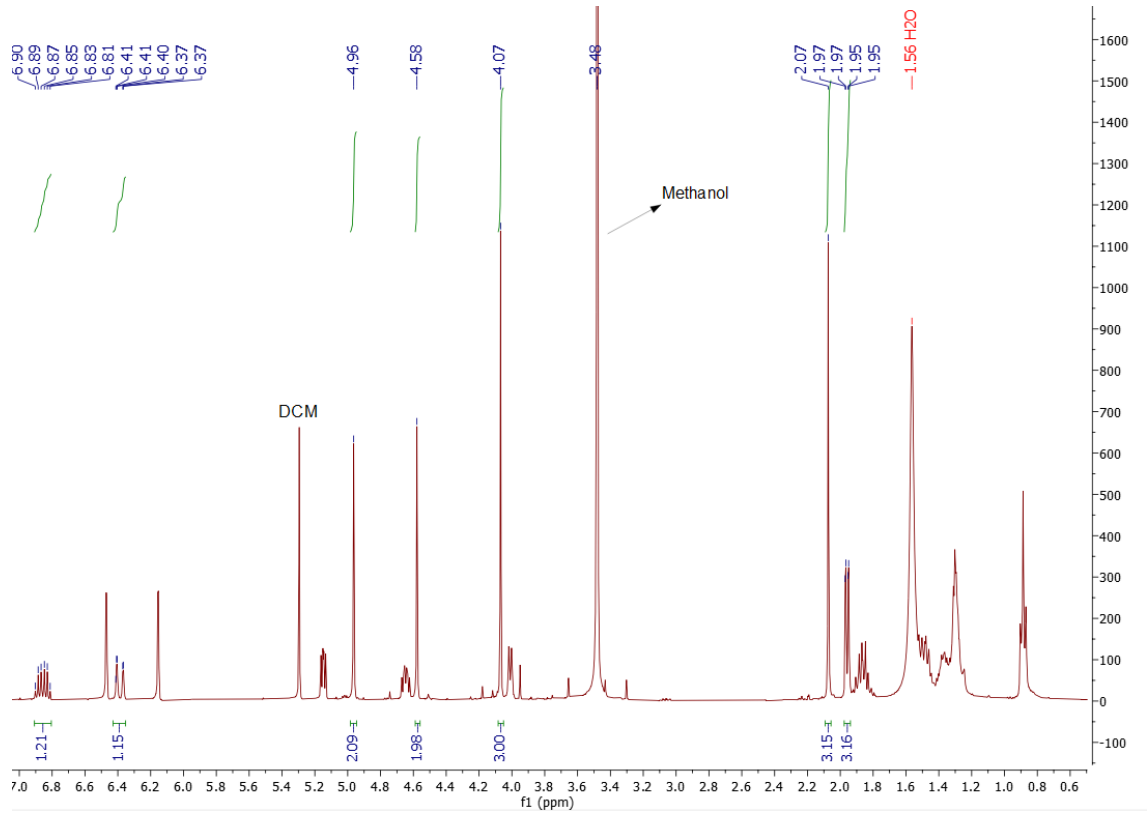


Figure 2.6  $^1\text{H}$  NMR of compound **82** (mix with sporothriolide **1**).

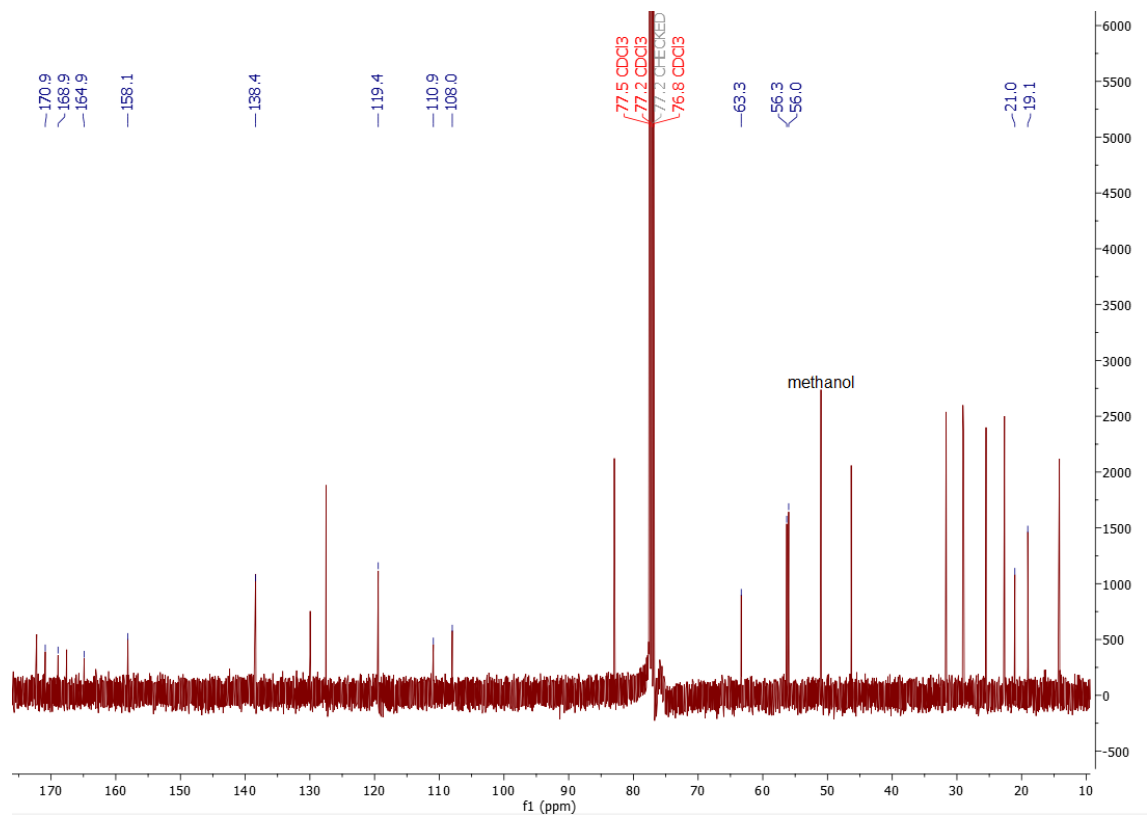


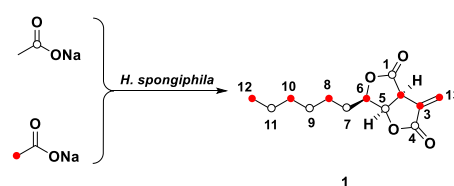
Figure 2.7  $^{13}\text{C}$  NMR of compound **82** (mix with sporothriolide **1**).

### 2.3.3 Acetate Feeding Experiments of Sporothriolide

Piliformic acid **59** is an alkyl citrate derivative, its biosynthesis was proposed to be derived from the condensation of a fatty acid and oxaloacetate **41** by extensive isotopic labelling studies.<sup>27,28</sup> Sporothriolide appears to follow this alkyl citrate biosynthetic pathway from the labelling pattern of deoxysporothric acid **76** (Scheme 2.1). Other well-known examples like the maleidrides byssochlamic acid **46** also share these similar core early steps (Section 1.2.1), but a highly reduced polyketide was used to condense with oxaloacetate.

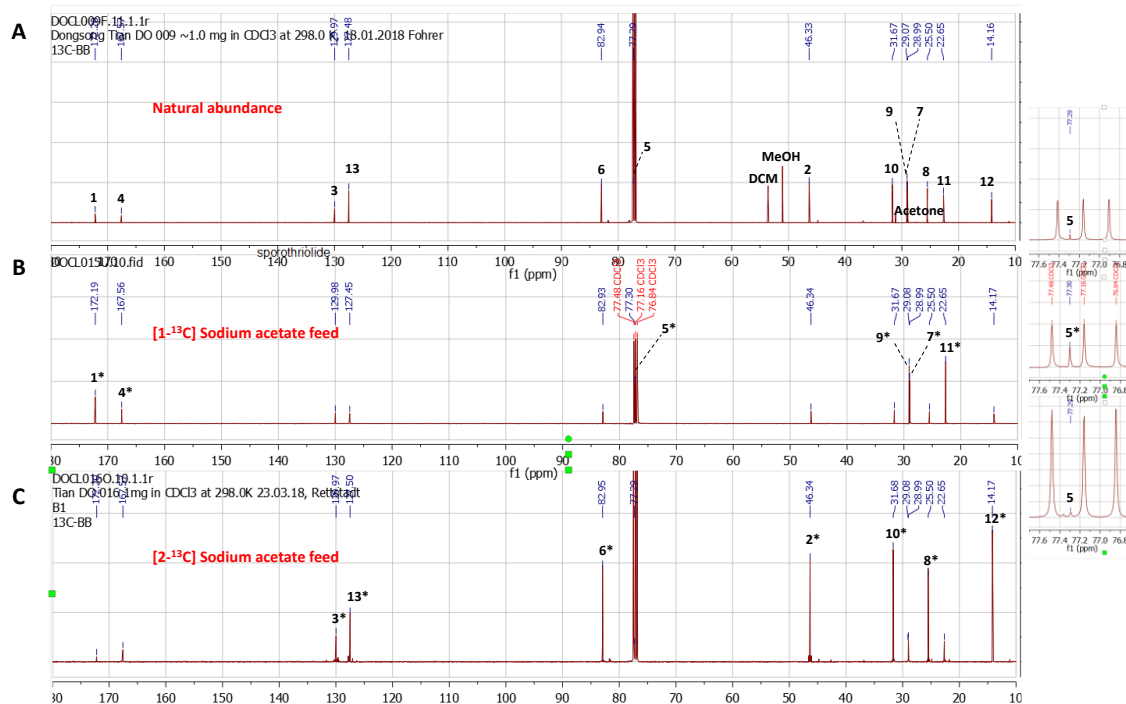
To study the labelling pattern of **1**, feeding experiments with labelled sodium acetate were conducted based on the previously calculated production kinetics. Therefore, 415 mg of [1-<sup>13</sup>C] and [2-<sup>13</sup>C] sodium acetate dissolved in 3 mL ddH<sub>2</sub>O were separately supplemented to 3 days old *H. spongiphila* cultures grown in 500 mL PDB medium (five 500 mL flasks each contain 100 mL medium) at 28 °C and 130 rpm. Feeding was repeated on day 4 and 5 to reach a final concentration of labelled acetate of 10 mM. Cultures were harvested on day 6. Extraction, purification and structure elucidation of compounds were achieved as described in Section 6.2. <sup>13</sup>C NMR spectra were recorded for labelled sporothriolide **1**. Peak enhancement was estimated by calculating the ratio between the normalised peak intensity of each signal of the labelled compound and the normalised signal intensity of each carbon at natural abundance. C-12 was used as a reference for [1-<sup>13</sup>C] sodium acetate feeding experiments, C-11 was used as a reference for [2-<sup>13</sup>C] sodium acetate feeding experiments. (Table 2.3, Figure 2.8).

Position	1- <sup>13</sup> C-Sodium Acetate Incorporation / Fold	2- <sup>13</sup> C-Sodium Acetate Incorporation / Fold
1	6.5	1
2	1	3.5
3	1.5	2.5
4	4.5	0.5
5	2.5	0.5
6	0.5	3
7	3	1
8	0.5	3
9	3.5	0.5
10	1	4
11 (reference)	5	1
12 (reference)	1	8
13	0.5	2



**Table 2.3** NMR signal peak enhancement of labelled sporothriolide **1**. A significant incorporation threshold fold is set  $\geq 2$ .





**Figure 2.8**  $^{13}\text{C}$  NMR spectra of sporothriolide **1**. **A**, Natural abundance; **B**,  $[1-^{13}\text{C}]$  sodium acetate feeding experiment; **C**,  $[2-^{13}\text{C}]$  sodium acetate feeding experiment. The enhanced peaks are marked with asterisk.

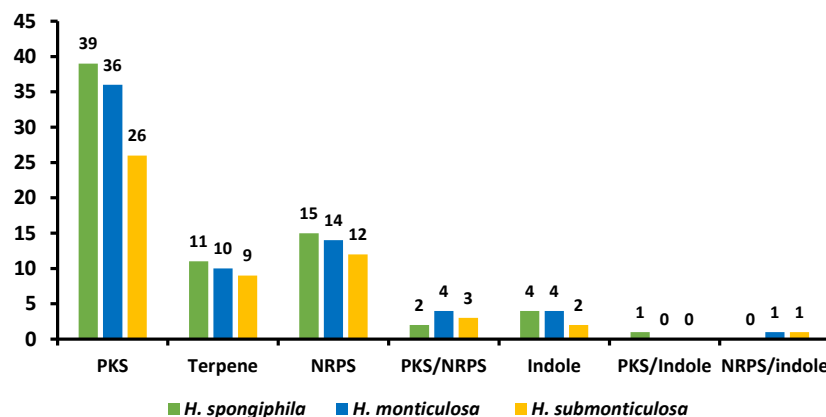
## 2.3.4 Genome and Transcriptome Analysis

### 2.3.4.1 Whole Genome Sequencing and antiSMASH Analysis

Genomic DNA from *Hypomontagnella monticulosa* MUCL 54604, *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471 was extracted and submitted to the CeBiTec (Center for Biotechnology) University of Bielefeld for Oxford nanopore/Illumina sequencing (Section 6.1.1). Three high-quality draft genomes were obtained (Table 2.4). Gene prediction and annotation were performed by Augustus<sup>58</sup> and GeneMark.<sup>59</sup>

Strain	<i>H. monticulosa</i> MUCL 54604	<i>H. spongiphila</i> CLL 205	<i>H. submonticulosa</i> DAOMC 242471
Estimated Genome size / bp	42,889,121 (Oxford nanopore/Illumina)	42,321,440 (Oxford nanopore/Illumina)	41,374,079 (Illumina)
Scaffolds/contigs	30	16	123
N <sub>50</sub> value / bp	3,439,634	5,056,634	657,615
Annotated genes	11,204	12,622	10,988

**Table 2.4** Oxford Nanopore/Illumina-Sequencing results for *H. spongiphila*, *H. monticulosa* and *H. submonticulosa*.



**Figure 2.9** Bar chart of predicted number of biosynthetic gene clusters found in the genomes of *H. spongiphila* genome (green), *H. monticulosa* genome (blue) and *H. submonticulosa* genome (yellow) using antiSMASH fungal v 5.1.2.<sup>60</sup>

The complete genome sequence of *H. monticulosa* MUCL 54604, *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471 were submitted to the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH)<sup>60</sup> fungal version 5.1.2 for the prediction of putative gene clusters. A total of 72, 69 and 53 secondary metabolite BGC were predicted in the *H. spongiphila*, *H. monticulosa* and *H. submonticulosa* genomes, respectively. These BGC are matched with several encoding genes, such as PKS (polyketides synthase), NRPS (nonribosomal peptide synthetase), terpenes and alkaloids (Figure 2.9). The antiSMASH prediction showed that PKS gene clusters are the most common in all three genomes, and the second most common are NRPSs (Figure 2.9). However, secondary metabolites gene clusters with fatty acid synthases were not predicted by antiSMASH.

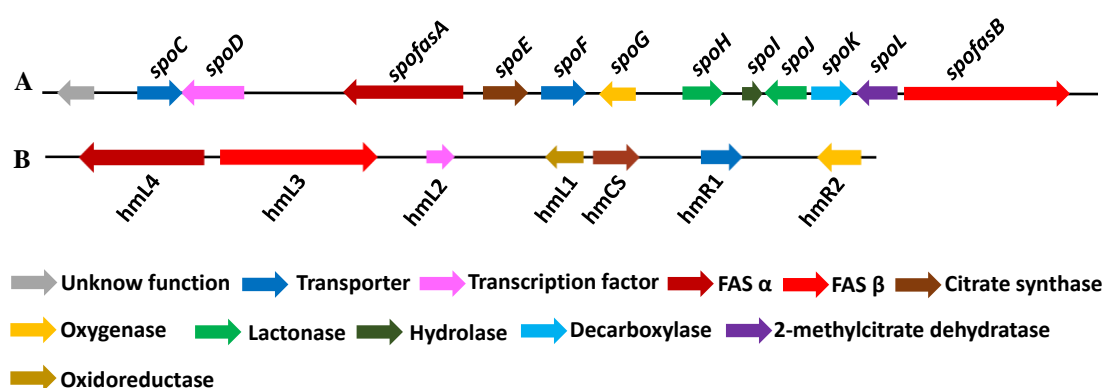
### 2.3.4.2 Sporothriolide BGC Analysis

Sporothriolide **1** is proposed to share similar early steps with byssochlamic acid **46** biosynthesis (Section 1.2.1, Scheme 1.3). Their backbones are constructed by condensation of polyketide or fatty acid chain with oxaloacetate moiety through a citrate synthase-like enzyme.

Manual homology searches against the genome using citrate synthase gene *bfl2* (GenBank ANF07286.1) from the byssochlamic acid BGC, rapidly identified a target BGC based around fungal FAS (fatty acid synthase) genes co-located with CS (citrate synthase) genes in all three organisms (Figure 2.10 – 2.11). This BGC also encodes the expected oxygenase and decarboxylase catalysts (Figure 2.10A, Table 2.5). BLASTp<sup>61</sup> and PHYRE2<sup>62</sup> platforms were used to annotate and predict the putative function of all proteins encoded by the BGC (Table 2.5). In addition to the expected citrate synthase *spoE*, and two fungal FAS subunits (FAS $\alpha$  and FAS $\beta$ ) *spofasA* and *spofasB*, the BGC encodes: a methylcitrate dehydratase (*spoL*); a decarboxylase (*spoK*) similar to *cis*-aconitate decarboxylase,<sup>63</sup> a dioxygenase (*spoG*); and two putative

lactonases (*spoH* and *spoJ*). In addition, the BGC also encodes two transporters (*spoC* and *spoF*), along with a transcriptional regulator (*spoD*) and *spoI* with unknown function (GenBank MT889334, Table 7.1). The three clusters are highly homologous, containing the same genes in the same order and orientations (Figure 2.11).

There exist two more FAS gene clusters in the genome. One is most likely specific for the production of primary metabolite fatty acids as it is not clustered with functional tailoring genes. The second is a secondary metabolite FAS gene cluster, but with only a citrate synthase and an oxidase nearby (Figure 2.10B). Obviously, these two are not the complex BGCs expected for sporothriolide **1** biosynthesis.

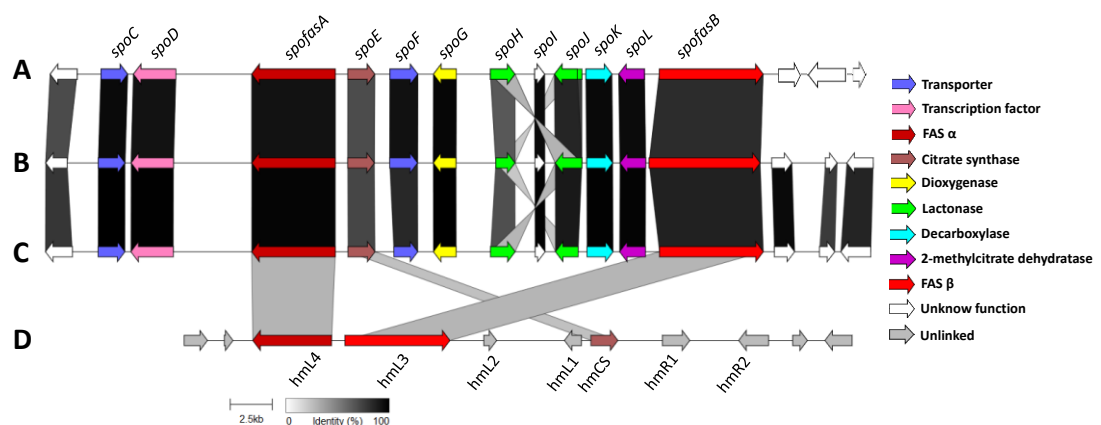


**Figure 2.10** Two gene clusters from *H. monticulosa* MUCL 54604 which contain fatty acid synthases and citrate synthase. **A**, the *spo* BGC; **B**, BGC with unknown function.

Clinker & clustermap,<sup>64</sup> a Python based tool, which was recently developed to generate accurate, interactive, publication-quality gene cluster comparison figures were used to illustrate the clusters (Figure 2.11). We processed this software for the comparison of *spo* (sporothriolide) cluster from *H. monticulosa*, *H. spongiphila* and *H. submonticulosa* organisms, and they showed a high similarity level indicating a conserved relationship (Figure 2.11A – 2.11C). Also, similarities of FAS and CS genes were displayed between two fatty acid synthase gene clusters (Figure 2.11C – 2.11D).

Gene (locus_tag)	Gene	AA	Putative Function	BLASTp <sup>a</sup> , PHYRE2 <sup>b</sup>	Predicted Cofactor
HmMg6350	-	319	Unknown	Amino-acid permease <sup>a</sup> , Membrane protein <sup>b</sup>	/
HmMg6351	<i>spoC</i>	508	Transporter	Citrinin biosynthesis cluster MFS transporter <sup>a</sup>	/
HmMg6352	<i>spoD</i>	819	Transcription factor	Transcriptional regulatory protein <sup>a</sup>	/
HmMg6353	<i>spofasA</i>	1619	Fatty acid synthase subunit alpha	Fatty acid synthase subunit alpha <sup>a</sup>	/
HmMg6354	<i>spoE</i>	460	Citrate synthase	Citrate synthase <sup>a</sup>	/
HmMg6355	<i>spoF</i>	493	Transporter	Efflux pump <sup>a</sup>	/
HmMg6356	<i>spoG</i>	373	Dioxygenase	Sulfonate dioxygenase <sup>a</sup> , Oxidoreductase <sup>b</sup>	Alpha-ketoglutarate
HmMg6357	<i>spoH</i>	443	Lactonase	Gluconolactonase <sup>a</sup> , Hydrolase <sup>b</sup>	/
HmMg6358	<i>spoI</i>	184	Putative hydrolase	Unknown <sup>a, b</sup>	/
HmMg6359	<i>spoJ</i>	441	Lactonase	Gluconolactonase <sup>a</sup> , Hydrolase <sup>b</sup>	/
HmMg6360	<i>spoK</i>	508	Decarboxylase	Aconitate decarboxylase <sup>a</sup> , Isomerase <sup>b</sup>	/
HmMg6361	<i>spoL</i>	491	Dehydratase	2-Methylcitrate dehydratase <sup>a</sup>	/
HmMg6362	<i>spofasB</i>	2060	Fatty acid synthase subunit beta	Fatty acid synthase subunit beta <sup>a</sup>	/
HmMg6363	-	408	Unknown	Cytosol aminopeptidase <sup>a</sup> , Ribosomal protein <sup>b</sup>	/
HmMg6364	-	238	Unknown	Unknown <sup>a</sup> , Sulfotransferase <sup>b</sup>	/
HmMg6365	-	516	Unknown	Methionyl-tRNA formyltransferase <sup>a</sup>	/

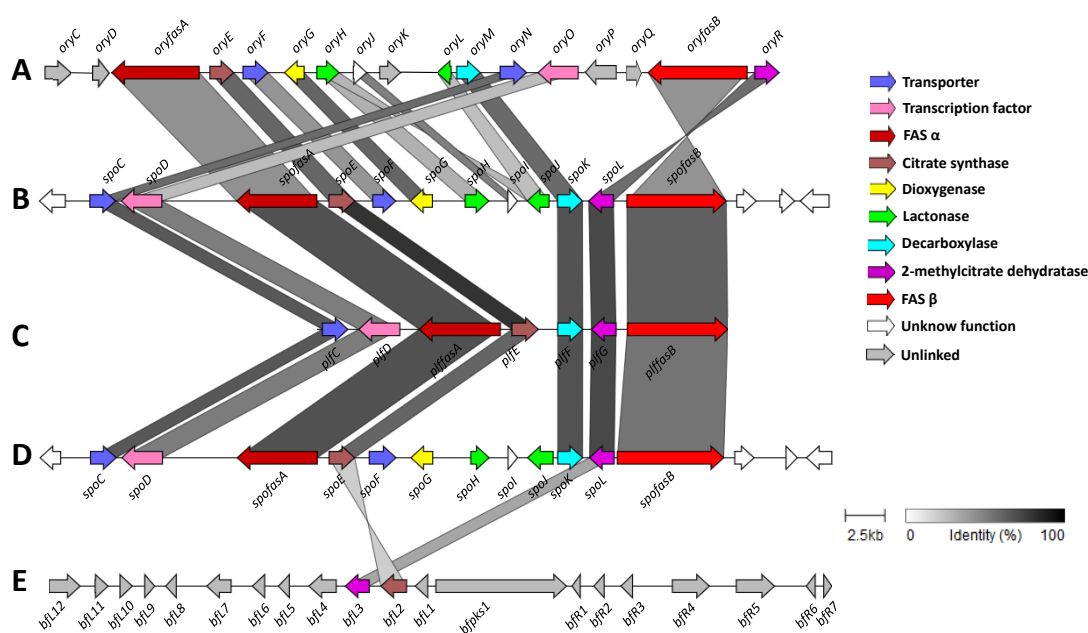
**Table 2.5** Annotation of sporothriolide BGC and surrounding genes from *H. monticulosa* MUCL 54604 genome using BLASTp<sup>61</sup> and PHYRE-2.<sup>62</sup>



**Figure 2.11** Clinker & clustermap<sup>64</sup> comparison of putative sporothriolide BGC and the unknown function BGC (contains FAS and CS): **A**, *spo* cluster from *H. submonticulosa* DAOMC 242471; **B**, *spo* cluster from *H. spongiphila* CLL 205; **C**, *spo* cluster from *H. monticulosa* MUCL 54604; **D**, unknown function BGC from *H. monticulosa* MUCL 54604.

Homology searches were also manually conducted by Dr. Eric Kuhnert within the genomes of other Hypoxylaceae and *Xylaria hypoxylon* obtained from an associated study<sup>37</sup> using the citrate synthase SpoE and the fatty acid synthase subunits (SpofasA, SpofasB) as templates. Results

show that only *X. hypoxylon* contained a cluster with a similar organisation to the *spo* cluster, which however lacked dioxygenase and lactonase genes (Figure 2.12C).



**Figure 2.12** Clinker & clustermap<sup>64</sup> comparison of alkyl citrates BGC and maleidrides BGC: **A**, putative oryzine BGC from *A. oryzae* RIB 40;<sup>23</sup> **B and D**, *spo* cluster from *H. monticulosa* MUCL 54604; **C**, putative piliformic acid BGC of *Xylaria hypoxylon*; **E**, byssochlamic acid BGC from *Byssochlamys fulva* IMI 40021.<sup>21</sup>

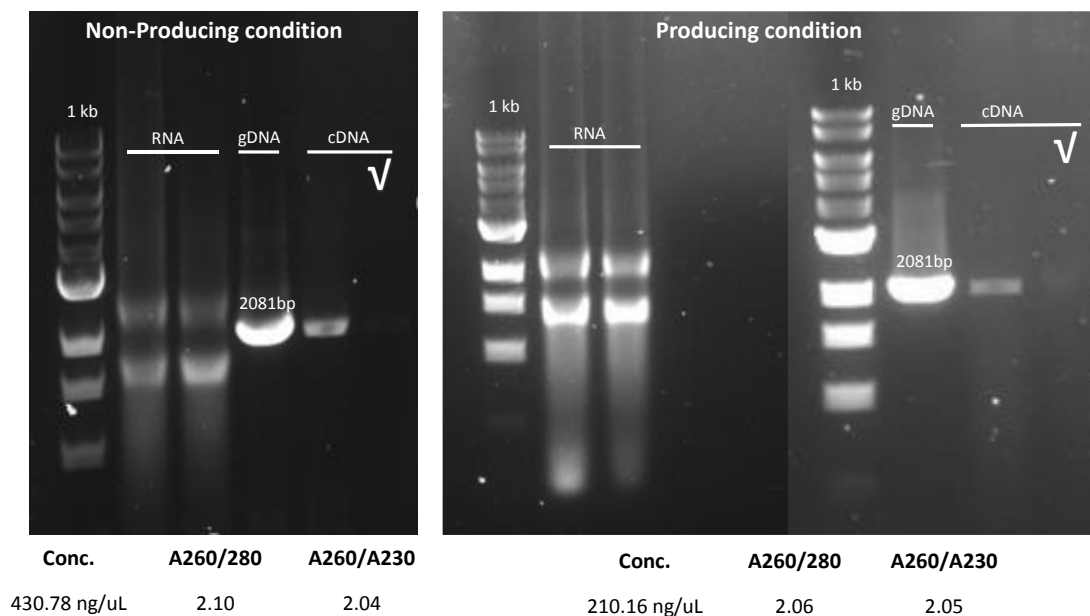
The *X. hypoxylon* gene cluster encodes: a citrate synthase (*plfE*); two fungal FAS subunits (*plffasA* and *plffasB*); a methylcitrate dehydratase (*plfG*) and a decarboxylase (*plfF*) similar to the *cis*-aconitate decarboxylase.<sup>63</sup> As *X. hypoxylon* is a known producer of piliformic acid **59** (2-hexylidene-3-methylsuccinic acid),<sup>25</sup> we hypothesise that the compound is the most likely product of the *plf* cluster.

In addition, a synteny analysis between the *spo* BGC, the oryzine (*ory*) BGC from *Aspergillus oryzae*, the byssochlamic acid (*bf*) BGC from *Byssochlamys fulva* and the putative piliformic acid (*plf*) BGC from *Xylaria hypoxylon* was also performed (Figure 2.12). The *spo* cluster, *ory* cluster and *plf* cluster share more homologous genes than with the *bf* cluster based on the clinker & clustermap results, which fit with the structural diversity between alkyl citrates and maleidrides.

### 2.3.4.3 Transcriptomic Analysis

In order to get further insight into the proposed sporothriolide BGC and elaborate on the gene cluster boundary, *H. monticulosa* MUCL 54604 was cultivated under producing (PDB medium) and non-producing (DPY medium) conditions (Section 2.3.1.1). Mycelia were collected and used to prepare total RNA using the Quick-RNA Fungal/Bacterial Miniprep Kit (Zymo Research, Germany), see details in Section 6.1.1. Genomic DNA (gDNA) contamination of the extracted RNA was checked by PCR amplification of “Polymerase II subunit” (housekeeping gene) with

forward primer P1 and reverse primer P2 (Table 6.2) binding to intron region, cDNA was used as a template which was obtained from reverse transcription of mRNA. The PCR amplification with gDNA template was used as a positive control with a length of 2081 bp (Figure 2.13).



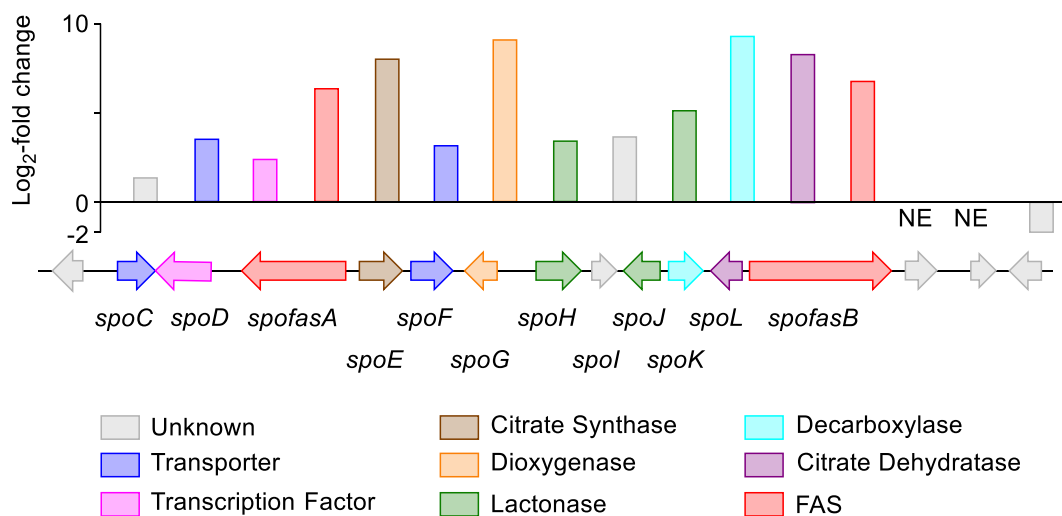
**Figure 2.13** RNA extraction from producing and non-producing conditions of *H. monticulosa* MUCL 54604.

Three replicates of ideal quality RNA samples (gDNA free) for each condition were submitted to CeBiTec for cDNA sequencing (Section 6.1.1). Data was used to perform a differential expression sequence (DESeq) analysis.<sup>65</sup> Mean normalized expression levels (NEL) from producing conditions (A) and non-producing conditions (B) were used to calculate the  $\log_2$ -fold change (B/A) to visualize differences in expression levels for each gene in the BGC.

If a  $\log_2$ -fold change of  $> 2$  was observed, genes are regarded as differentially expressed. Expression levels of the *spo* genes and surrounding genes are listed in Table 2.6, Figure 2.14. Differential gene expression analysis of the predicted functional genes from *spoC* to *spofasB* show a strong upregulation under producing conditions (Figure 2.14), while genes outside this region show either low expression (Table 2.6, Figure 2.14) or no change in transcription level. Based on this analysis, the sporothriolide 1 gene cluster boundary is set which including 12 genes from *spoC* to *spofasB*.

Gene (locus_tag)	Gene	NEL A (non-producing)	NEL B (producing)	Log <sub>2</sub> -fold change (B/A)
HmMg6349	-	40.42	54.97	0.44
HmMg6350	-	92.24	246.13	1.42
HmMg6351	<i>spoC</i>	1176.34	11594.43	3.30
HmMg6352	<i>spoD</i>	35.24	174.58	2.31
HmMg6353	<i>spofasA</i>	9.33	771.14	6.37
HmMg6354	<i>spoE</i>	18.66	3936.16	7.72
HmMg6355	<i>spoF</i>	153.39	1263.85	3.04
HmMg6356	<i>spoG</i>	32.13	18022.20	9.13
HmMg6357	<i>spoH</i>	66.33	528.71	2.99
HmMg6358	<i>spoI</i>	5350.01	55269.24	3.37
HmMg6359	<i>spoJ</i>	69.44	2463.48	5.15
HmMg6360	<i>spoK</i>	2.07	1441.91	9.44
HmMg6361	<i>spoL</i>	3.11	820.59	8.04
HmMg6362	<i>spofasB</i>	21.76	2108.77	6.60
HmMg6363	-	0	3.72	-
HmMg6364	-	-	-	-
HmMg6365	-	3.11	0.88	-1.82

**Table 2.6** Normalized expression level (NEL, BaseMean) for genes of the *spo* cluster and adjacent genes from *H. monticulosa* MUCL 54604 strain. Data calculated with DESeq.



**Figure 2.14** Transcriptomic analysis of the sporothriolide BGC from *H. monticulosa* MUCL 54604. Log<sub>2</sub>-fold changes are calculated between producing and non-producing conditions. NE, no expression.

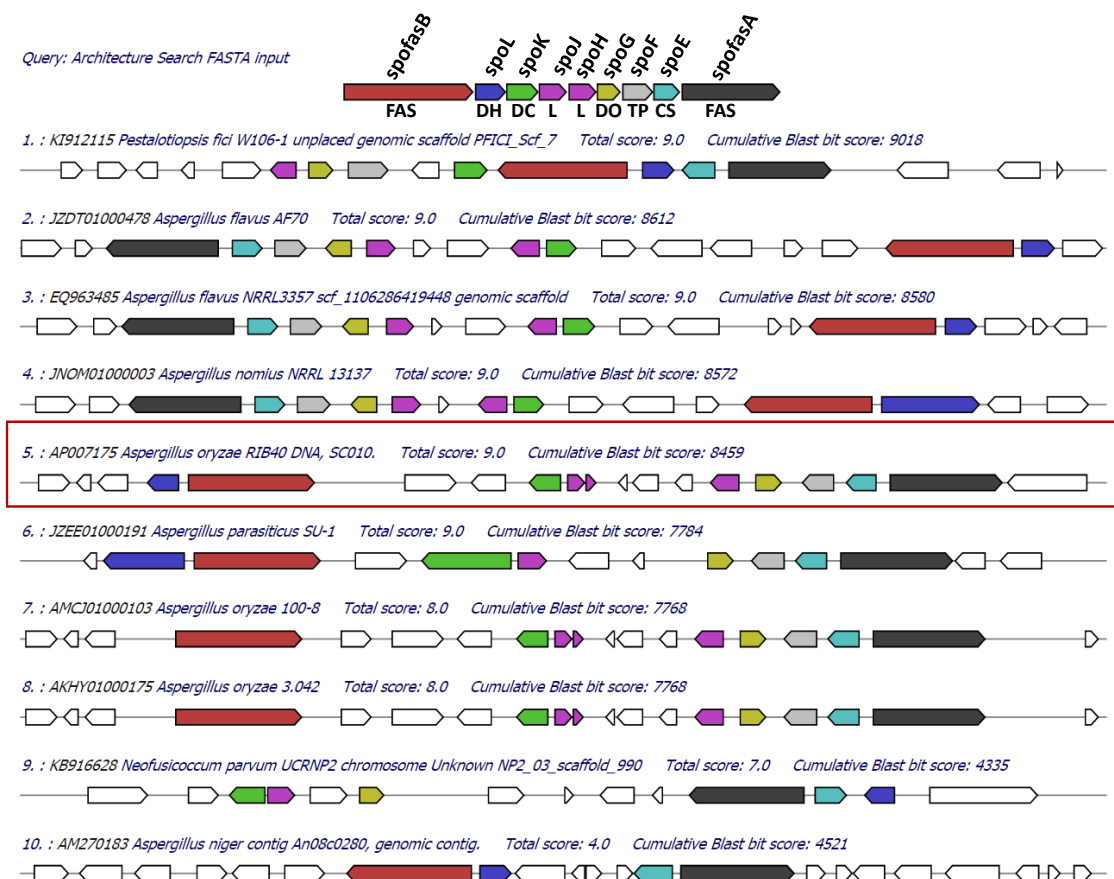
Other very useful information obtained from the transcriptomic analysis includes the exact intron and exon positions of the expressed genes, as well correcting the genome sequencing errors and re-annotating genes function. By using the transcriptome data we re-annotated the *spo* gene cluster, and the final corrected amino acid sequences of protein (GenBank MT889334) are shown in Chapter 7, Table 7.1.

#### 2.3.4.4 MultiGeneBlast

In order to identify homologous *spo* clusters in other fungal genomes, a MultiGeneBlast<sup>66</sup> analysis was conducted (<http://multigeneblast.sourceforge.net/>), using templates SpofasA (FAS  $\alpha$ ), SpofasB (FAS  $\beta$ ), SpoE (Citrate synthase), SpoF (Transporter), SpoG (Dioxygenase), SpoH (Lactonase), SpoJ (Lactonase), SpoK (Decarboxylase) and SpoL (2-Methyl citrate dehydratase) on the genome *H. monticulosa* MUCL 54604 as MultiGeneBlast architecture search. Due to an outdated version, only genomes released in NCBI GenBank until November 2015 were considered.

The results showed that such FAS-based pathways are relatively common in fungi but have remained unrecognised to date: modern bioinformatic tools such as antiSMASH<sup>60</sup> do not yet recognize such BGCs, probably because of the high number of 'primary-metabolism' related steps. At least five homologous clusters encoding fungal FAS  $\alpha$  and  $\beta$  components, and citrate synthase (*spoE*), dehydratase (*spoL*), decarboxylase (*spoK*), dioxygenase (*spoG*) and one or more lactonases (*spoHJ*) were detected. The best hits of the search revealed the presence of various similar clusters in *Pestalotiopsis fici* and others predominantly in *Aspergillus* species (Figure 2.15), and similar pathways will likely appear elsewhere. The genus *Aspergillus* includes the previously published oryzine cluster (*ory*) from *A. oryzae* (Figure 2.12 and Figure 2.15).





**Figure 2.15** MultiGeneBlast (architecture search) of the sporothriolide biosynthetic enzymes of *H. monticulosa* MUCL 54604. There are six hits with 'Total score 9.0', and the putative biosynthetic gene cluster of the oryzines was included (red frame).

### 2.3.5 Gene Knockout in *H. spongiphila* CLL 205

Gene knockout is an ideal methodology not only to verify the correct cluster of natural products biosynthesis but also to reveal the function of each gene through individual inactivation. Among them, the bipartite knockout (Figure 1.8) is one of the most useful strategies in fungi. This involves co-transformation with two overlapping fragments, creating a functional fungal selection marker by homologous recombination of two split pieces (Section 1.3.3).

#### 2.3.5.1 Fungal Transformation of *H. spongiphila* CLL 205

Because the hypoxylon species had never been genetically transformed before our work, the development of an efficient fungal transformation protocol was an early aim. This was done by trialling and testing various protoplast preparation conditions, different concentrations and constituents of transformation solutions, as well as several plate-spreading methods. Finally, we established an effective fungal transformation protocol (Section 6.1.2) for *H. spongiphila*. In addition, this protocol is suitable for *H. monticulosa* and *H. submonticulosa* transformations.

### 2.3.5.1a Antibiotics Screening and Protoplast Preparation

A wide range of genes have been used as selectable markers for fungi, especially the *hph* gene (hygromycin B resistance) which is widely used as a fungal selection system.<sup>38</sup> To test the sensitivity of *H. spongiphila* to hygromycin, this fungus was subcultured on DPY agar containing different concentrations (5, 10, 25, 50, 100, 150, 200  $\mu\text{g}\cdot\text{mL}^{-1}$ ) of hygromycin. Observation of the colonies at 28 °C for 10 days showed that *H. spongiphila* was able to grow at a concentration of 5, 10, 25, 50 and 100  $\mu\text{g}\cdot\text{mL}^{-1}$  hygromycin B, but is inhibited at concentrations of 150 and 200  $\mu\text{g}\cdot\text{mL}^{-1}$  (Figure 2.16A). Therefore, we selected 150  $\mu\text{g}\cdot\text{mL}^{-1}$  of hygromycin as the selection method in the *H. spongiphila* transformation.

Protoplast preparation is a vital first step for fungal transformation. Protoplasts quality and quantity dominantly determine the efficiency of transformation and hence knockout efficiency. From the developed protocol (Section 6.1.2), a decent amount of protoplasts ( $1.22 \times 10^9 \cdot \text{mL}^{-1}$ ) of *H. spongiphila* was obtained (Figure 2.16B).

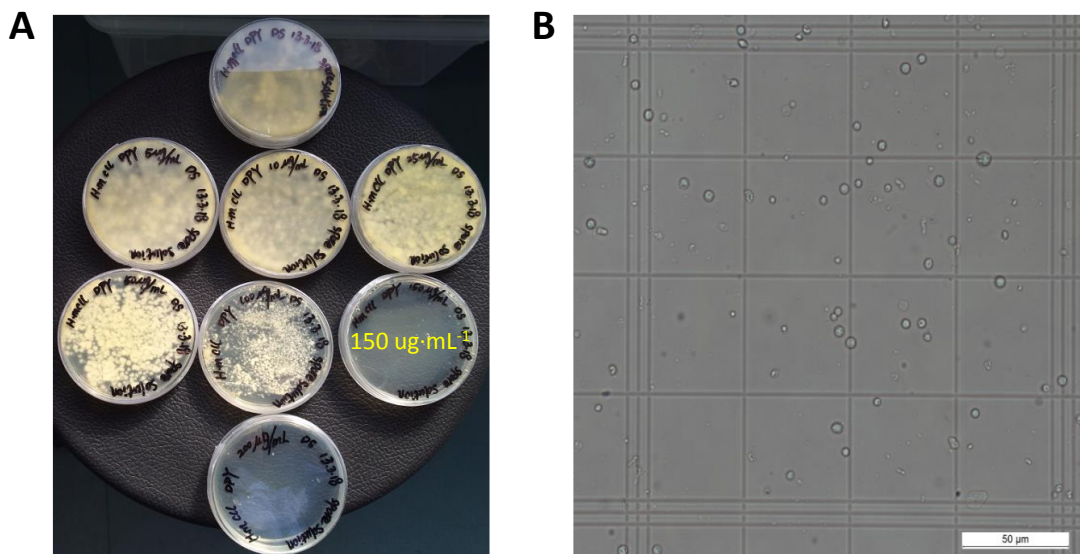
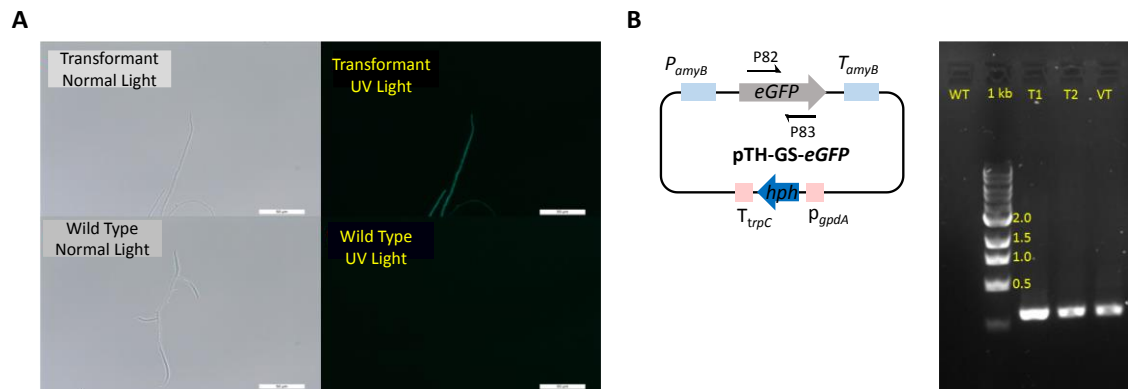


Figure 2.16 A, hygromycin sensitive assay; B, protoplasts of *H. spongiphila* CLL 205.

### 2.3.5.1b Transformation with pTH-GS-eGFP

The pTH-GS-eGFP<sup>38</sup> vector contains an *hph* gene (resistance cassette) for selection with hygromycin B, and the *eGFP* gene encoding an enhanced green fluorescent protein (eGFP) that exhibits a visible green fluorescence when exposed under inducing light. The *eGFP* gene is expressed under the control of the *P<sub>amyB</sub>* promoter which is induced by starch or maltose (Figure 2.17B), while *hph* is expressed from the *P<sub>gpdA</sub>* promoter, a strong constitutive promoter from *Aspergillus nidulans*.

The established fungal transformation method (Section 6.1.2) for *H. spongiphila* was tested with the pTH-GS-*eGFP* vector, this was conducted by the master student Haoxuan Zeng. Nine colonies were obtained from three rounds of hygromycin selection plates and then cultivated in DPY liquid medium (Table 6.5). DPY medium contains starch as the carbon source for the induction of  $P_{amyB}$  promoter. Microscopic analysis of three transformants showed that they displayed green fluorescence (Figure 2.17A). In addition, PCR (P82 + P83, Table 6.2) amplified the *eGFP* gene of three transformants, in contrast, the wild type (WT) did not show a PCR product (Figure 2.17B).



**Figure 2.17** A, cell image of *H. spongiphila* wild type and transformant (supported by Haoxuan Zeng); B, agarose gel electrophoresis of *eGFP* gene from wild type and transformant. WT: wild type; T: transformant; VT, vector.

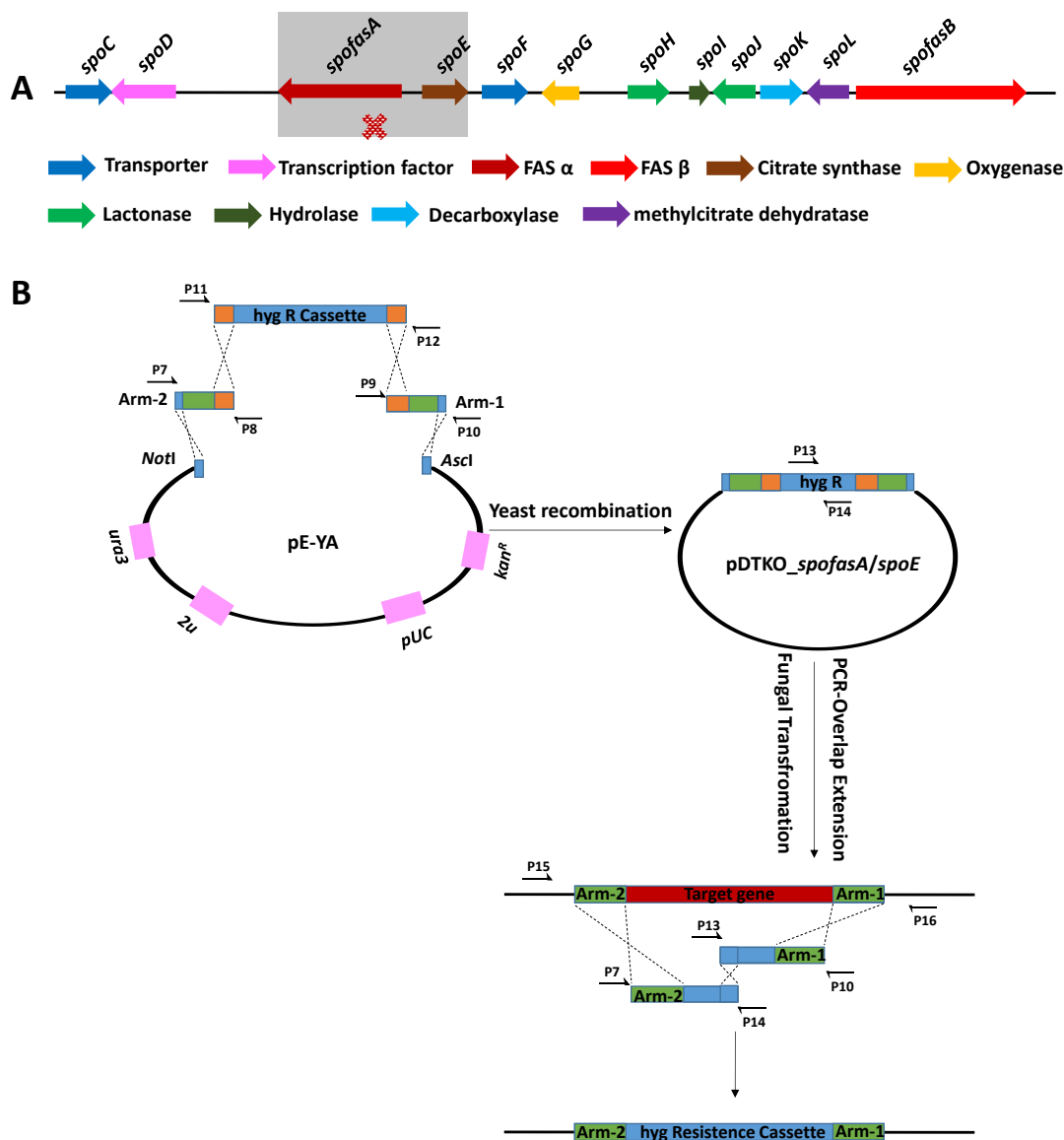
### 2.3.5.2 Knockout of *spofasA/spoE*

#### 2.3.5.2a Vector Construction for *spofasA/spoE* Knockout

As homologues of both genes were located in the genomes (Figure 2.10) potentially complementing gene loss, the *spoE* (CS) and *spofasA* genes in *H. spongiphila* were then deleted simultaneously.

A vector based on the pE-YA plasmid including a 500 bp upstream fragment of *spofasA* and a 500 bp downstream fragment of *spoE* separated by a hygromycin resistance gene (*hph*) as selection marker was constructed (Figure 2.18). Homologous flanking arm-1 (500 bp upstream fragment of *spofasA*) and arm-2 (500 bp downstream fragment of *spoE*) were amplified by PCR from gDNA of *H. spongiphila* CLL205 using the primer sets P7 + P8 and P9 + P10 (Table 6.2). The pTH-GS-*eGFP* vector was used as the template to amplify the hygromycin resistance cassette (*hph*) with the primer sets P11 + P12 (Table 6.2). The pE-YA empty vector was linearized with the restriction enzymes *AscI* and *NotI* (New England BioLabs). Recombination of the fragments was achieved by Yeast Homologous Recombination (Section 6.1.2.2). Plasmids were purified from yeast using the Zymoprep™ Yeast Plasmid Miniprep II kit (Zymo Research), transformed into Top10 *E. coli* cells for amplification, screened by colony PCR (Table 6.2) and purified using the NucleoSpin™ Plasmid kit (Machery-Nagel).

Two overlapping fragments of the constructed KO cassette (1916 bp and 3280 bp) were PCR amplified by OneTaq® 2X Master Mix (New England BioLabs, USA) using the primer sets P7 + P14 and P13 + P10 (Table 6.2). Fragments were purified with the GenElute™ PCR Clean-Up Kit (Sigma-Aldrich) and used for the transformation of *H. spongiphila* protoplasts.



**Figure 2.18** A, gene cluster for sporothriolide 1; B, the bipartite knockout (Figure 1.8) method to replace the targeted genes with a hygromycin resistance cassette *via* homologous recombination. Red: Target gene sequence (*spofasA* and *spoE* from *H. spongiphila*); Blue: Hygromycin resistance cassette, containing the  $P_{gdpA}$  promoter, the hygromycin resistance gene (*hph*) and the  $T_{trpC}$  terminator.

### 2.3.5.2b *spofasA/spoE* Knockout Workflow

The protocol details of *H. spongiphila* transformation is in Section 6.1.2. The first selection plates with hygromycin B were incubated at 28 °C for 8 to 10 days until colonies were observed. Fifty colonies (collected from 3 rounds of fungal transformation) were transferred to new DPY plates with hygromycin B for another 3 to 4 days growing, this is the second-round selection. Forty-six

well-growing colonies were then again transferred to new DPY plates with hygromycin B for the third-round selection. Finally, 46 viable transformants were placed on normal DPY agar and incubated for 5 to 6 days, followed by cultivating in producing conditions for chemical metabolite analysis (Figure 2.19). In general, one whole round of knockout procedures will take one month. However, each round efficiency is impossible to estimate in advance. Therefore the fungal transformation was performed consecutively for 3 to 5 rounds, with an aim to isolate *ca* 50 transformants after the 3<sup>rd</sup> selection round.

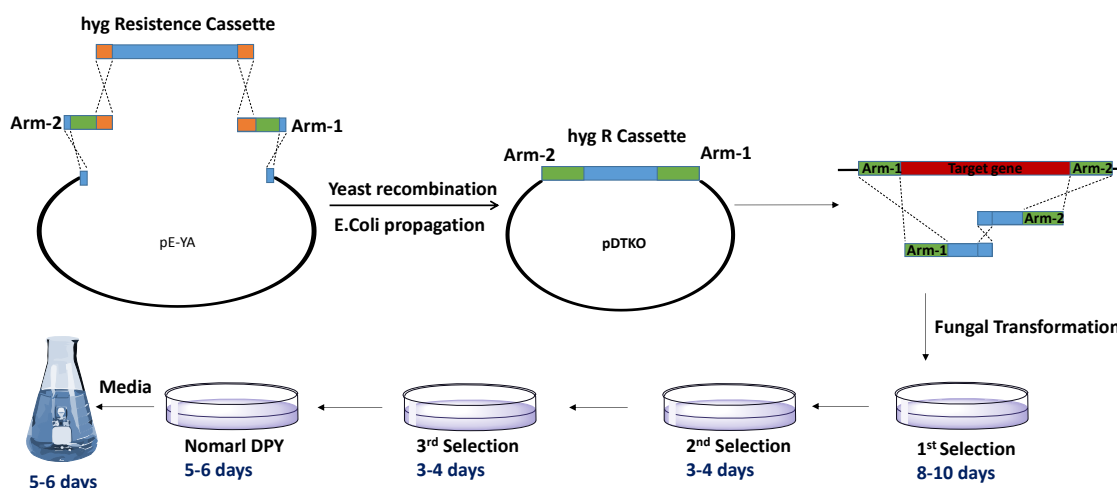


Figure 2.19 Working flow of the bipartite knockout of *H. spongiphila* CLL 205.

### 2.3.5.2c *spofasA/spoE* Knockout Transformant Analysis

Forty-six transformants with hygromycin resistance were generated after three rounds of fungal transformation. All 46 transformants were cultivated under sporothriolide **1** producing conditions and compared to the wild type (WT) strain (Figure 2.20). One of these produced neither sporothriolide **1** nor the related spochartine or congeners **75** – **78**.

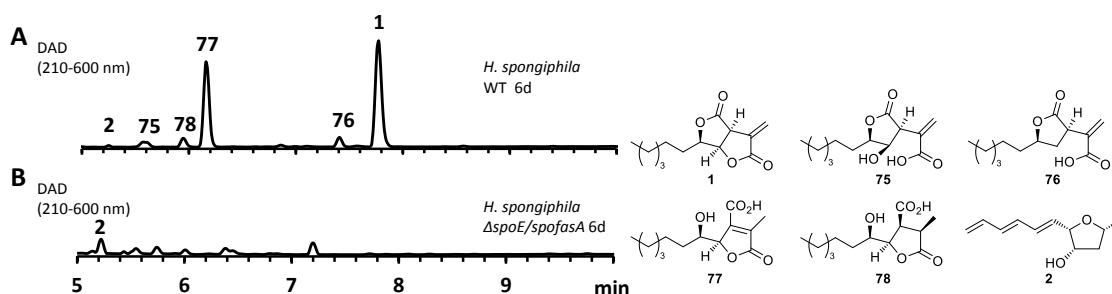


Figure 2.20 High Performance Liquid Chromatography (HPLC) analysis of crude extracts from *H. spongiphila* CLL 205 grown under producing conditions of **1**: **A**, Diode Array Detector (DAD) chromatogram of wild type (WT); **B**, DAD chromatogram of  $\Delta spoE/spofasA$ .

PCR analysis of this specific transformant was performed. The transformant and WT were used for the extraction of gDNA. Amplification of the first fragment (Figure 2.21) with

primers P15 + P14 (Table 6.2), P15 binding to outside region of target gene while P14 binding to the *hph* gene. Meanwhile, the second insertion fragment (Figure 2.21) was amplified by using primers P13 + P16, as in the previous case P16 binding to the outside region of the target gene while P13 binding to the *hph* gene. As shown in gel results, the incorporation of the hygromycin resistance gene at the target position was confirmed. The *spoE* and *spofasA* were successfully disrupted (Figure 2.21).

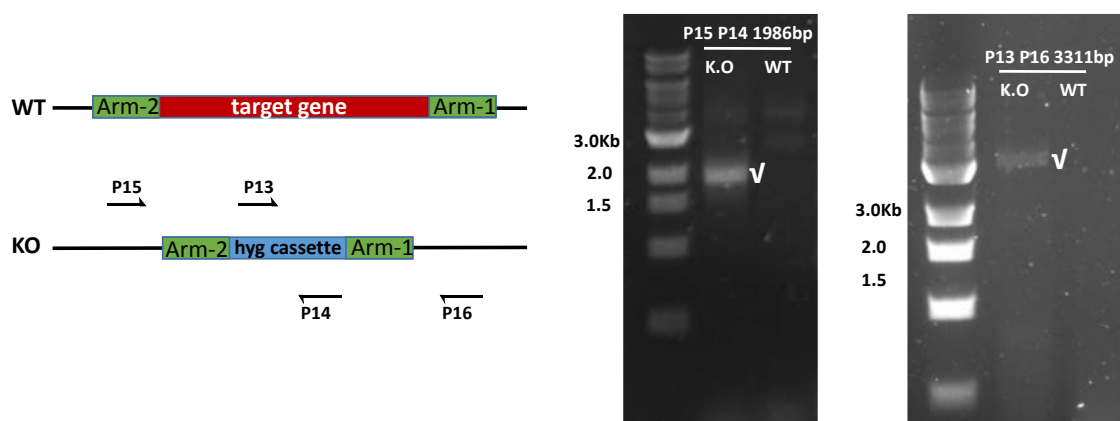


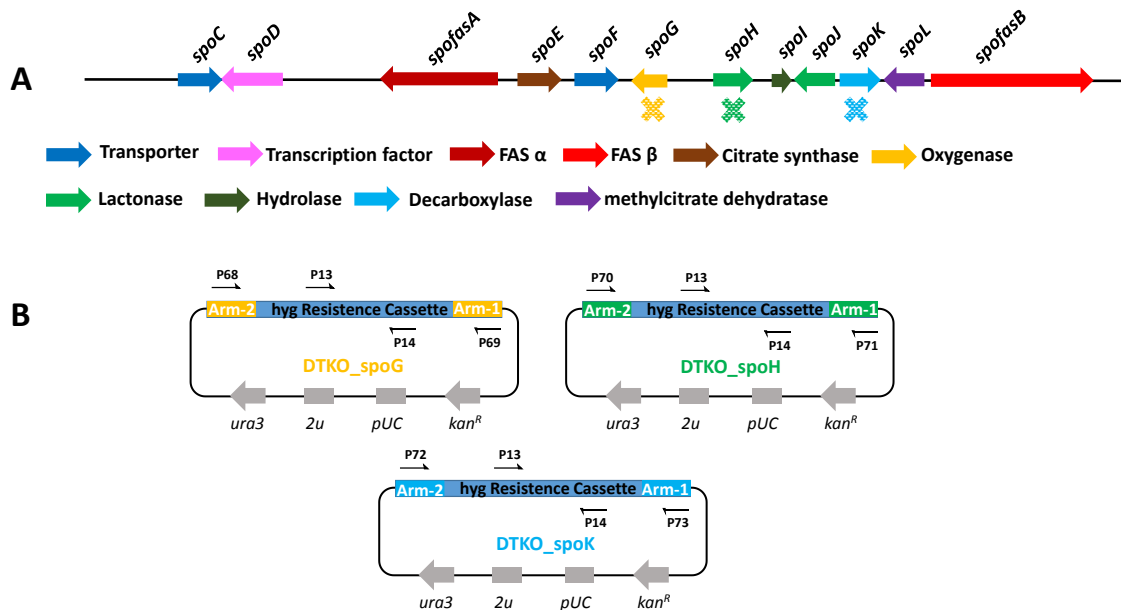
Figure 2.21 Verification of positive *H. spongiphila*  $\Delta spoE/spofasA$  transformant using PCR.

### 2.3.5.3 Attempted Knockout of Other Genes

The correct gene cluster for sporothriolide biosynthesis was proved through gene knockout of *spofasA/spoE* abolishing target metabolite production. In addition, the cluster also includes *spoG*, *spoH*, and *spoK* encoding dioxygenase, lactonase and decarboxylase, respectively. We designated the following individual knockout experiments targeting tailoring genes *spoG*, *spoH*, and *spoK* in sporothriolide **1** biosynthesis, in order to observe their functions by analysing metabolic intermediates after gene deactivation.

#### 2.3.5.3a *spoG*, *spoH* and *spoK* Knockout Vector Constructions and Transformation

KO vectors were constructed as previously described for the *spofasA/spoE* knockout experiment (Figure 2.18). The bipartite fragments for the knockout experiment were amplified by using primers P68 + P14 and P13 + P69 for *spoG* (Figure 2.22, Table 6.2), P70 + P14 and P13 + P71 for *spoH*, as well as P72 + P14 and P13 + P73 for *spoK*.



**Figure 2.22** **A**, Gene cluster of sporothriolide; **B**, constructed plasmids for bipartite knockout of *spoG*, *spoH* and *spoK*, respectively.

Then fragments were purified and used directly for  $\text{CaCl}_2/\text{PEG}$  mediated protoplast transformation of *H. spongiphila*. Finally, 36 transformants from *spoG* KO, 33 transformants targeting *spoK* KO, and 50 transformants targeting *spoH* KO were obtained after three rounds of fungal transformation (Table 2.7).

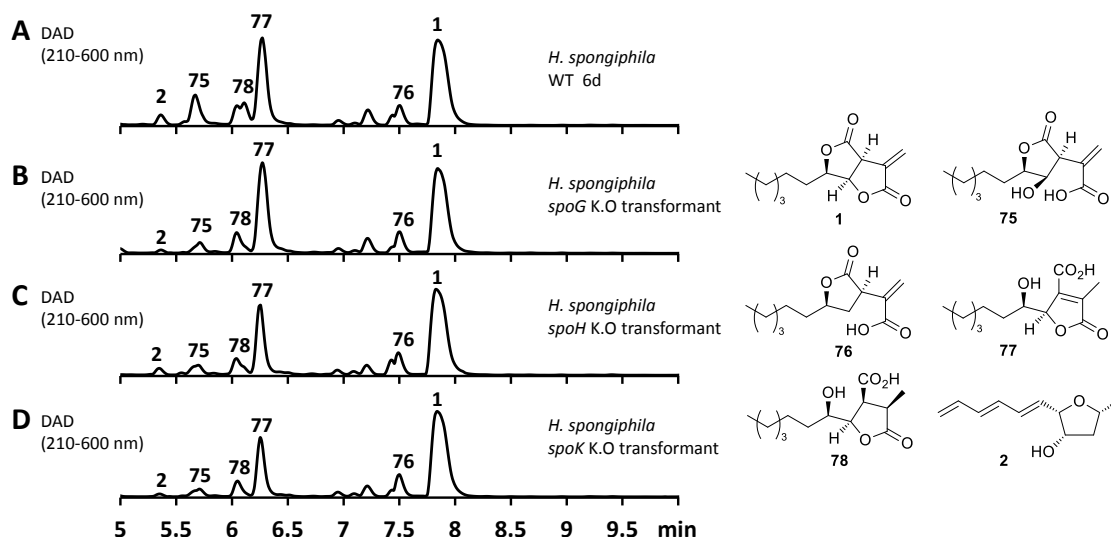
Target genes	1 <sup>st</sup> Round	2 <sup>nd</sup> Round	3 <sup>rd</sup> Round	Total
<i>spoG</i>	2	9	25	36
<i>spoH</i>	8	25	17	50
<i>spoK</i>	4	14	15	33

**Table 2.7** Transformants obtained from three rounds fungal transformation of *spoG*, *spoH* and *spoK* knockout, respectively.

### 2.3.5.3b *spoG*, *spoH* and *spoK* Knockout Transformant Analysis

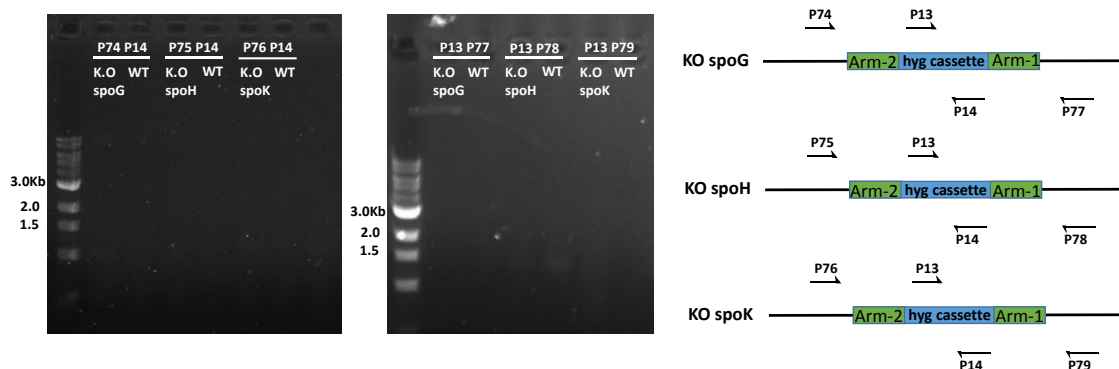
All transformants generated from KO were cultivated under sporothriolide **1** producing conditions, PDB liquid medium at 130 rpm 28 °C 6 day, then HPLC chromatograms of the extracts were compared with an extract of the WT strain (Figure 2.23). Unfortunately, sporothriolide metabolites **75** – **78** were observed in all transformants extracts.





**Figure 2.23** HPLC analysis of crude extracts from *H. spongiphila* grown under producing conditions of **1**: **A**, Diode Array Detector (DAD) chromatogram of wild-type (WT); **B**, DAD chromatogram of *spoG* KO transformant; **C** DAD chromatogram of *spoH* KO transformant; **D**, DAD chromatogram of *spoK* KO transformant.

Genomic DNA of transformant from KO *spoG*, *spoK*, and *spoH* (Figure 2.24) were extracted for genetic testing. Forward primers P74 for *spoG*, P75 for *spoH*, and P76 for *spoK* binding to the upstream of flanking arm-2 and reverse primer P14 binding to the inside of the hygromycin cassette were employed for the first fragment amplification (Figure 2.24). In the same manner, forward primer P13 binding to inside of the hygromycin cassette and reverse primers P77 for *spoG*, P78 for *spoH*, and P79 for *spoK* binding to the downstream of flanking arm-1 were applied for the second fragment insertion verification (Figure 2.24). However, no PCR products were observed on the gel. This suggested that bipartite substrates were ectopically integrated within the genome somewhere else, not in the corresponding position of target *spoG*, *spoH*, and *spoK* genes. From the first KO it is clear that the efficiency of the bipartite KO is low in *H. spongiphila* (1/46). It is therefore not surprising that KO in the other case was not successful.



**Figure 2.24** Genetic analysis of *spoG*, *spoH*, *spoK* KO transformant. No positive knockout results from the gel.



### 2.3.6 Heterologous Expression in *A. oryzae* NSAR1

The first knockout experiment (Section 2.3.5.2) supported the correct BGC for sporothriolide **1**. Although the employment of gene knockout in *Hypomontagnella* species is deficient for enzyme function investigation, heterologous expression (Section 1.4.4) of the sporothriolide BGC in quadruply auxotrophic *A. oryzae* NSAR1 may open ways to study the biosynthetic steps of sporothriolide **1**.

#### 2.3.6.1 Expression Vectors and Cloning Strategies

*A. oryzae* NSAR1 is auxotrophic in arginine ( $\Delta argB$ ), sulfate assimilation ( $sC^-$ ), adenine ( $adeA^-$ ) and ammonium ( $niaD^-$ ) metabolism. In addition, the natural sensitivity towards the antibiotics bleomycin ( $ble^R$ ) and glufosinate-ammonium (Basta;  $bar^R$ ) enables the use of two more dominant selectable markers.<sup>49</sup> Two fungal transformation vectors were used in this work (Figure 2.25), each using single complementation with one of two different selection markers ( $argB$  and  $adeA$ ). Each expression vector has four gene cloning sites in total with promoter/terminator  $P/T_{amyB}$ ,  $P/T_{adh}$ ,  $P/T_{gpdA}$  and  $P/T_{eno}$  in the pTAYAGS series of fungal expression vectors (Figure 2.25).<sup>49,67</sup>

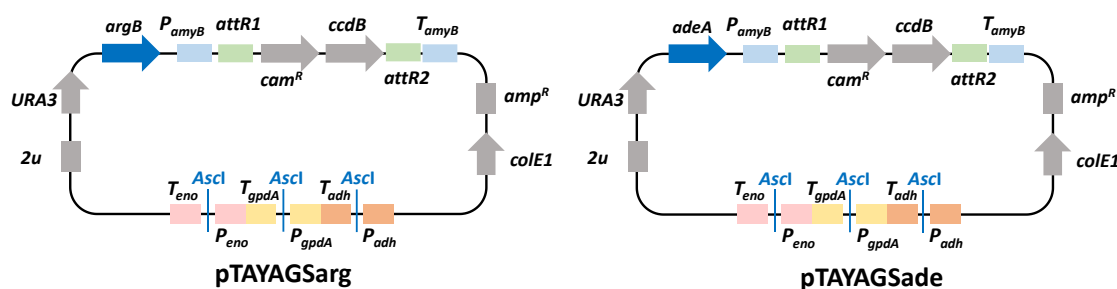
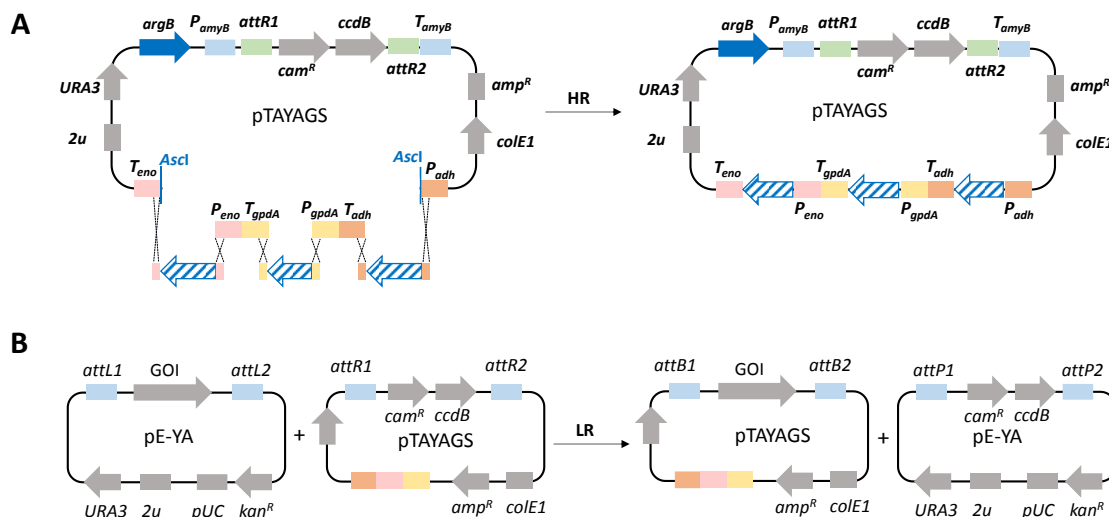


Figure 2.25 pTAYAGS vectors used in this heterologous expression work.

Promoter/terminator  $P/T_{adh}$ ,  $P/T_{gpdA}$  and  $P/T_{eno}$  sequences flanking  $AscI$  sites are designed for gene insertion by homologous recombination in *Saccharomyces cerevisiae* (Section 6.1.2.2). The desired vector is cut in three pieces by the  $AscI$  restriction endonuclease and co-transformed with up to three DNA fragments (Figure 2.26A). Each DNA fragment includes a homologous overlap sequence (at least 30 bp) introduced by PCR amplification. The  $2\mu$  origin of replication ( $ori$ ) and  $ura3$  gene (encoding an orotidine 5'-phosphate decarboxylase) are used to select uracil auxotroph *S. cerevisiae*. Origin of replication ( $colE1$ ) and  $amp^R$  gene (resistance to ampicillin) are utilised for propagation in *E. coli*.

The fourth cloning site is located between promoter/terminator  $P/T_{amyB}$  which flank the recombination sequences  $attR1$  and  $attR2$  to constitute a Gateway *in vitro* recombination cassette.<sup>49,67</sup> This is composed of the  $ccdB$  gene encoding the CcdB killer protein, and the  $cam^R$  gene encoding chloramphenicol resistance (Figure 2.26A). The interchange of DNA fragments

flanked with *attR* sites and *attL* sites are achieved *in vitro* by a Gateway cloning kit (LR recombinase, Invitrogen). Normally, large genes are assembled from multiple parts through homologous recombination in *S. cerevisiae* in the entry vector pE-YA<sup>67</sup> (with *attL* sites). This gene is then cloned into the target pTAYAGS destination vectors by Gateway cloning (Figure 2.26B).<sup>49</sup>



**Figure 2.26** **A**, *in vivo* homologous recombination in pTAYAGS vector; **B**, Gateway cloning by LR recombination between pE-YA vector (with GOI) and pTAYAGS vector leading to expression pTAYAGS plasmids. HR, homologous recombination.

### 2.3.6.2 Gene Combinations and Plasmid Constructions

#### 2.3.6.2a Overview of Constructed Plasmids

Based on the transcriptomic analysis in Section 2.3.4.3, we could define the *spo* cotranscribed region as running from *spoC* to *spofasB*, including 12 genes in total (Figure 2.14). These include genes encoding: two fungal FAS subunits (*spofasA* and *spofasB*); two transporters (*spoC* and *spoF*); a transcriptional regulator (*spoD*); a citrate synthase (*spoE*); a dioxygenase (*spoG*); two lactonases (*spoH* and *spoJ*); a decarboxylase (*spoK*); a methylcitrate dehydratase (*spoL*), and *spoI* with unknown function (GenBank MT889334, Table 7.1). For heterologous expression, the transcription factor (*spoD*) and transporters (*spoC* and *spoF*) were omitted since they are supposed to be only useful for the gene cluster expression regulation within the original host *H. monticulosa*, thus they are insignificant for the secondary metabolites biosynthetic pathway. In addition, the functional unknown *spoI* was not considered in our heterologous expression work. Finally, eight functional genes from the *spo* BGC were selected for expression in *A. oryzae* NSAR1. These are *spofasA*, *spofasB*, *spoE*, *spoG*, *spoH*, *spoJ*, *spoK*, and *spoL*.

Genomic DNA from *H. monticulosa* was isolated from 5 days old liquid cultures using the GeneElute™ Plant Genomic DNA Miniprep Kit (Sigma Life Science, USA) following the

manufacturer's instructions. For RNA extraction, *H. spongiphila* and *H. monticulosa* were grown under sporothriolide producing conditions for 3 days in PDB medium. RNA was obtained by using the method in Section 6.1.1 and transcribed into cDNA using the High Capacity RNA-to-cDNA™ kit (Applied Biosystems by Thermo Fisher Scientific, USA).

The transcriptomic analysis made solid supports for the determination of gene borders, as well as intron and exon positions (Table 7.1). Genes of interest were either amplified from *H. monticulosa* cDNA or gDNA using the primer sets listed in Table 2.8 depending on the investigated combination of genes (for primer details see Chapter 6, Table 6.2).

Two destination vectors (pTAYAGSarg and pTAYAGSade) were employed as a basis for plasmid construction. Genes were either inserted *via* yeast homologous recombination (YHR) or Gateway cloning (Figure 2.26 – 2.27). Before vector assembly, the pTAYAGS and pE-YA plasmids were digested with restriction endonucleases (*AscI* for pTAYAGS; *AscI* and *NotI* for pE-YA). Targeted genes and digested destination vectors were then together transformed into yeast cells for YHR (see Section 6.1.2.2 for details). To transfer genes from the pE-YA to the pTAYAGS vectors, a Gateway cloning kit (LR Clonase II Enzym-Mix, Invitrogen, USA) was applied following the manufacturer's guidelines. A detailed list of the constructed vectors used in this work can be found in Figure 2.27.

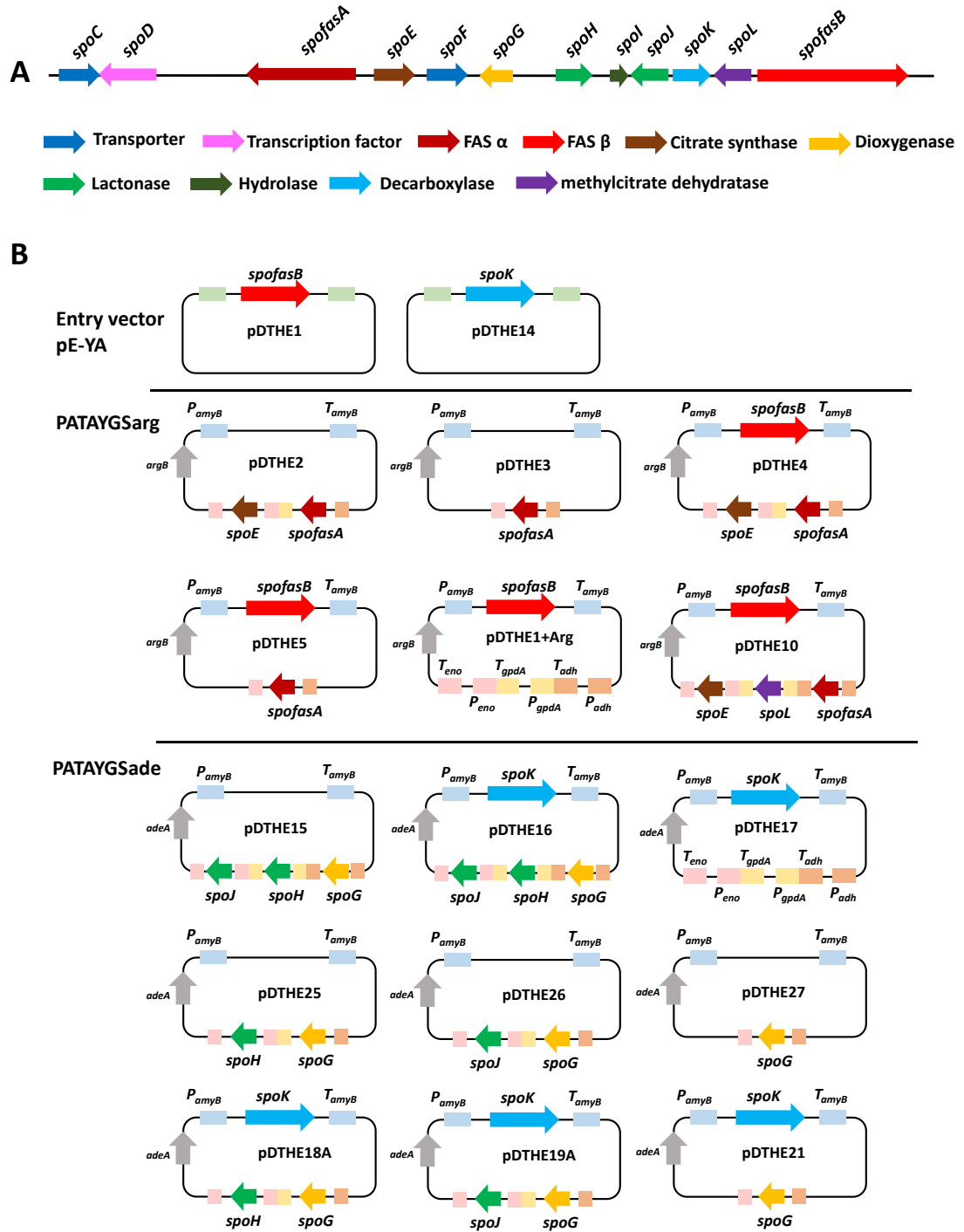


Figure 2.27 A, gene cluster of sporothriolide; B, constructed plasmids for *A. oryzae* heterologous expression studies.

## 2 Biosynthetic Studies of Sporothriolide

Primer no.	Template	Target vector	Purpose	PCR condition*
P17+P18	gDNA of <i>H. monticulosa</i>	pDTHE1	FAS $\beta$ ( <i>spofasB</i> ), fragment 1	Q5, 65°C
P19+P20	cDNA of <i>H. monticulosa</i>	pDTHE1	FAS $\beta$ ( <i>spofasB</i> ), fragment 2	Q5, 55°C
P19+P6	<i>E. coli</i> transformants with pDTHE1	pDTHE1	Colony PCR	OneTaq
P43+P44	cDNA of <i>H. monticulosa</i>	pDTHE14	Decarboxylase ( <i>spoK</i> )	Q5, 65°C
P5+P6	<i>E. coli</i> transformants with pDTHE14	pDTHE14	Colony PCR	OneTaq
P23+P24	gDNA of <i>H. monticulosa</i>	pDTHE2	FAS $\alpha$ ( <i>spofasA</i> ), fragment 1	Q5, 65°C
P21+P22	cDNA of <i>H. monticulosa</i>	pDTHE2	FAS $\alpha$ ( <i>spofasA</i> ), fragment 2	Q5, 65°C
P25+P26	cDNA of <i>H. monticulosa</i>	pDTHE2	Citrate synthase ( <i>spoE</i> )	Q5
P50+P22	<i>E. coli</i> transformants with pDTHE2	pDTHE2	Colony PCR	OneTaq
P60+P61	<i>E. coli</i> transformants with pDTHE2	pDTHE2	Colony PCR	OneTaq
P21+P22	cDNA of <i>H. monticulosa</i>	pDTHE3	FAS $\alpha$ ( <i>spofasA</i> ) fragment 1	Q5, 65°C
P23+P27	gDNA of <i>H. monticulosa</i>	pDTHE3	FAS $\alpha$ ( <i>spofasA</i> ) fragment 2	Q5, 65°C
P50+P22	<i>E. coli</i> transformants with pDTHE3	pDTHE3	Colony PCR	OneTaq
P21+P28	pDTHE2	pDTHE10	FAS $\alpha$ ( <i>spofasA</i> )	Q5, 65°C
P25+P26	cDNA of <i>H. monticulosa</i>	pDTHE10	Citrate synthase ( <i>spoE</i> )	Q5
P29+P30	gDNA of <i>H. monticulosa</i>	pDTHE10	Dehydratase ( <i>spoL</i> ) fragment 1	Q5
P31+P32	gDNA of <i>H. monticulosa</i>	pDTHE10	Dehydratase ( <i>spoL</i> ) fragment 2	Q5
P19+P49	<i>E. coli</i> transformants with pDTHE10	pDTHE10	Colony PCR	OneTaq
P52+P22	<i>E. coli</i> transformants with pDTHE10	pDTHE10	Colony PCR	OneTaq
P56+P57	<i>E. coli</i> transformants with pDTHE10	pDTHE10	Colony PCR	OneTaq
P60+P61	<i>E. coli</i> transformants with pDTHE10	pDTHE10	Colony PCR	OneTaq
P33+P34	gDNA of <i>H. monticulosa</i>	pDTHE15	Lactonase ( <i>spoH</i> ) fragment 1	Q5, 65°C
P35+P36	gDNA of <i>H. monticulosa</i>	pDTHE15	Lactonase ( <i>spoH</i> ) fragment 2	Q5
P37+P38	gDNA of <i>H. monticulosa</i>	pDTHE15	Lactonase ( <i>spoH</i> ) fragment 3	Q5
P39+P40	cDNA of <i>H. monticulosa</i>	pDTHE15	Lactonase ( <i>spoJ</i> )	Q5
P41+P42	cDNA of <i>H. monticulosa</i>	pDTHE15	Dioxygenase ( <i>spoG</i> )	Q5
P52+P53	<i>E. coli</i> transformants with pDTHE15	pDTHE15	Colony PCR	OneTaq
P56+P57	<i>E. coli</i> transformants with pDTHE15	pDTHE15	Colony PCR	OneTaq
P60+P61	<i>E. coli</i> transformants with pDTHE15	pDTHE15	Colony PCR	OneTaq
P41+P42	pDTHE15	pDTHE25	Dioxygenase ( <i>spoG</i> )	Q5
P33+P45	pDTHE15	pDTHE25	Lactonase ( <i>spoH</i> )	Q5
P50+P53	<i>E. coli</i> transformants with pDTHE25	pDTHE25	Colony PCR	OneTaq
P56+P61	<i>E. coli</i> transformants with pDTHE25	pDTHE25	Colony PCR	OneTaq
P41+P46	pDTHE15	pDTHE26	Dioxygenase ( <i>spoG</i> )	Q5
P39+P40	pDTHE15	pDTHE26	Lactonase ( <i>spoJ</i> )	Q5
P50+P57	<i>E. coli</i> transformants with pDTHE26	pDTHE26	Colony PCR	OneTaq
P60+P61	<i>E. coli</i> transformants with pDTHE26	pDTHE26	Colony PCR	OneTaq
P41+P47	pDTHE15	pDTHE27	Dioxygenase ( <i>spoG</i> )	Q5
P50+P61	<i>E. coli</i> transformants with pDTHE27	pDTHE27	Colony PCR	OneTaq
pDTHE4 (LR clone of pDTHE1 + pDTHE2)				
pDTHE5 (LR clone of pDTHE1 + pDTHE3)				
pDTHE1 + Arg (LR clone of pDTHE1 + pTAYAGSarg)				
pDTHE16 (LR clone of pDTHE14 + pDTHE15)				
pDTHE17 (LR clone of pDTHE14 + pTAYAGSade)				
pDTHE18A (LR clone of pDTHE14 + pDTHE25)				
pDTHE19A (LR clone of pDTHE14 + pDTHE26)				
pDTHE21 (LR clone of pDTHE14 + pDTHE27)				

**Table 2.8** Primer sets used in constructed plasmids for heterologous expression. \* Deviating PCR annealing temperatures (standard is 60 °C) are stated under PCR condition. Primers sequence see Table 6.2.

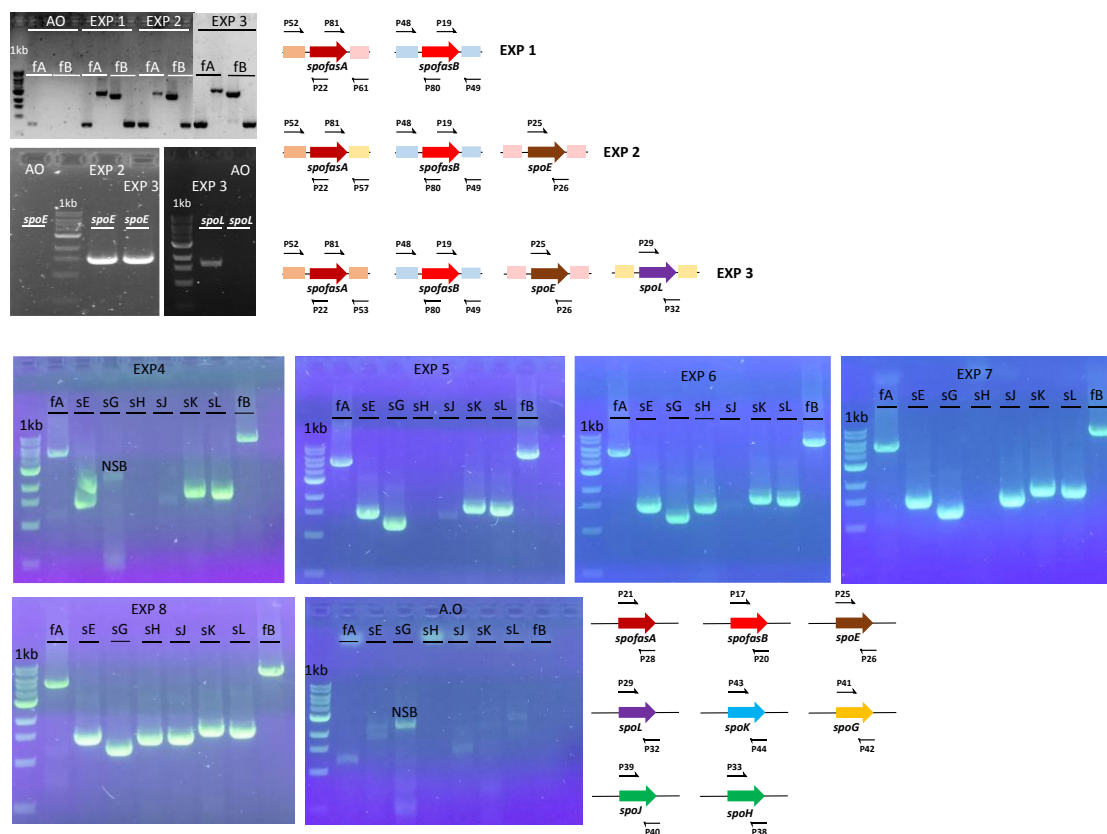
### 2.3.6.2b Gene Combinations and *A. oryzae* NSAR1 Transformation

In total, eight different (Table 2.9) gene-expression combinations were designed, and the corresponding vectors were co-transformed in *A. oryzae* NSAR1 by using a CaCl<sub>2</sub>/PEG mediated protoplast protocol (Section 6.1.2.3).

The obtained transformants (exact numbers see Table 2.9) were genetically checked by PCR reaction (Figure 2.28) with appropriate primers (Chapter 6, Table 6.2). For the large piece of *spofasA* and *spofasB*, parts of the head and tail were amplified by using primers binding to the promoter and terminator regions (EXP1 – 3), and for the short piece of genes, the full length was examined. Agarose gel electrophoresis showed that all genes from the positive transformants were successfully integrated into the *A. oryzae* (Figure 2.28). All transformants from each individual experiment were cultured in *P<sub>amyB</sub>* induction medium DPY at 28 °C for 5 – 7 days and 110 rpm, mycelia and culture media were extracted separately and analysed by LCMS (Section 6.2).

EXP	<i>spo</i>								Trans.	Products
	<i>fasA</i>	<i>E</i>	<i>G</i>	<i>H</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>fasB</i>		
	FASA	CS	DO	Lact	Lact	DC	DH	FASB		
<b>1</b>	✓	-	-	-	-	-	-	✓	<b>3/9</b>	-
<b>2</b>	✓	✓	-	-	-	-	-	✓	<b>3/10</b>	-
<b>3</b>	✓	✓	-	-	-	-	✓	✓	<b>2/10</b>	-
<b>4</b>	✓	✓	-	-	-	✓	✓	✓	<b>2/5</b>	<b>83</b>
<b>5</b>	✓	✓	✓	-	-	✓	✓	✓	<b>2/7</b>	<b>75 – 76, 83, 88 – 91</b>
<b>6</b>	✓	✓	✓	✓	-	✓	✓	✓	<b>2/9</b>	<b>75 – 76, 83, 88 – 89</b>
<b>7</b>	✓	✓	✓	-	✓	✓	✓	✓	<b>2/8</b>	<b>75 – 76, 83, 88 – 89</b>
<b>8</b>	✓	✓	✓	✓	✓	✓	✓	✓	<b>1/6</b>	<b>1, 76, 90 – 92</b>

**Table 2.9** Combinations of *spo* genes expressed in *A. oryzae* NSAR1. Trans: Transformants. The number 3/9 means that there were nine transformants obtained in EXP1 in total, and three of them are positive.



**Figure 2.28** Genetic analysis of *A. oryzae* NSAR1 transformant: EXP1 (*spofasA* + *spofasB*); EXP2 (*spofasA* + *spofasB* + *spoE*); EXP3 (*spofasA* + *spofasB* + *spoE* + *spoL*); EXP4 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK*); EXP5 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG*); EXP6 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*); EXP7 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*); EXP8 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*). EXP ID see Table 2.9. fA: *spofasA*; fB: *spofasB*; sE: *spoE*; sG: *spoG*; sH: *spoH*; sJ: *spoJ*; sK: *spoK*; sL: *spoL*. NSB: Non-specific binding.

### 2.3.6.3 Expression of the *spo* Genes for Early Steps

The fatty acid biosynthetic pathway is an attractive target for the production of chemicals and transportation fuels with different chain lengths.<sup>68</sup> The fatty acid synthase in sporothriolide BGC is expected to assemble a C<sub>10</sub> alkyl chain which potentially contains an alkene for later oxidation. Citrate synthase SpoE is proposed to be responsible for the condensation of the alkyl chain with oxaloacetate to form alkyl citrates and 2-methylcitrate dehydratase SpoL is proposed to eliminate one molecule of water. These steps are similar to the early steps of byssochlamic acid **46** biosynthesis (Scheme 1.3).

Three plasmids (pDTHE5\_EXP1, pDTHE4\_EXP2 and pDTHE10\_EXP3; Figure 2.29) were individually transformed into *A. oryzae*, then all transformants were cultured and extracted for LCMS analysis. The LCMS results showed that no related new metabolites were observed except the C<sub>16</sub>/C<sub>18</sub> fatty acids produced from host *A. oryzae* (EXP1 – 3, Figure 2.30).

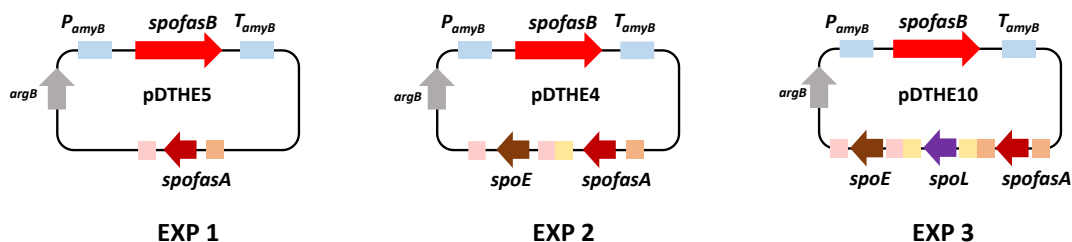


Figure 2.29 Plasmids used in expression work of EXP1, EXP2 and EXP3. EXP ID see Table 2.9.

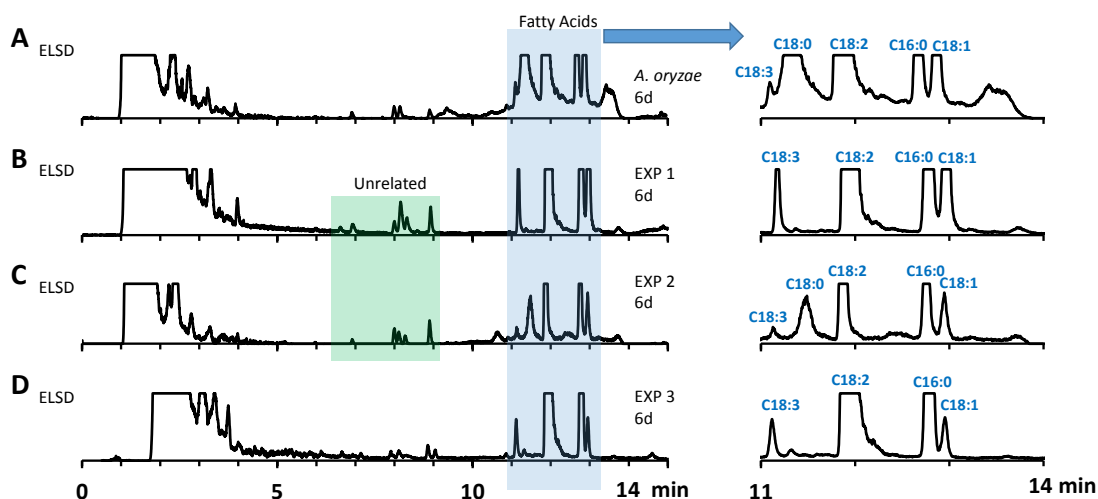


Figure 2.30 HPLC analysis of crude extracts (from cell) from *A. oryzae* transformant. **A**, ELSD chromatogram of untransformed *A. oryzae* NSAR1; **B**, ELSD chromatogram of EXP1 (*spofasA* + *spofasB*) transformant; **C**, ELSD chromatogram of EXP2 (*spofasA* + *spofasB* + *spoE*) transformant; **D**, ELSD chromatogram of EXP3 (*spofasA* + *spofasB* + *spoE* + *spoL*) transformant. EXP ID see Table 2.9.

### 2.3.6.4 Co-expression of Early Step Genes with Later Tailoring Genes

#### 2.3.6.4a Co-expression of *spofasA*, *spofasB*, *spoE*, *spoL* and *spoK*

SpoK is predicted as a decarboxylase. There is also a decarboxylation step after dehydration in the byssochlamic acid **46** pathway (Scheme 1.3), but in this case, it is spontaneous.<sup>21</sup> In order to investigate the function of the SpoK decarboxylase in sporothriolide **1** biosynthesis, we cloned *spoK* into the pTAYAGSade vector pDTHE17 and co-transformed it with vector pDTHE10 containing *spofasA*, *spofasB*, *spoE* and *spoL* (EXP4, Figure 2.31).

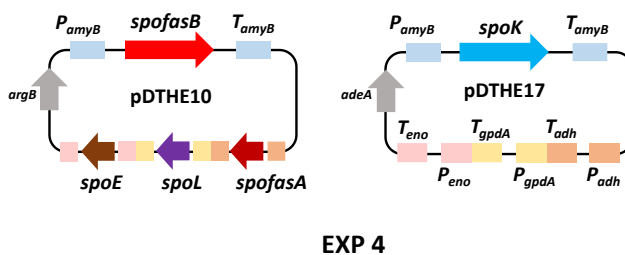
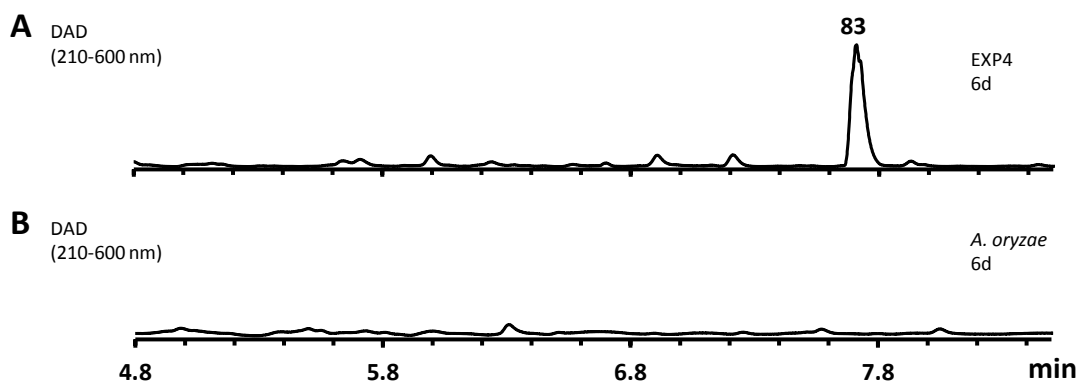


Figure 2.31 Plasmids used in expression work of EXP4. EXP ID see Table 2.9.

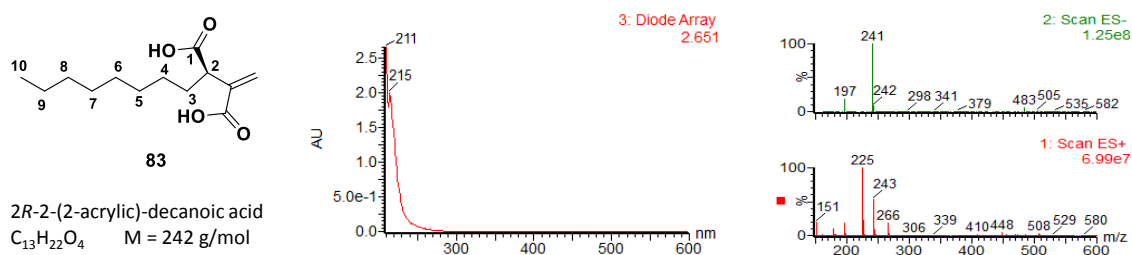


All transformants from EXP4 were cultivated and extracted for LCMS analysis. From the results, a new peak with a molecular weight of 242 (ESI-MS  $m/z$  241  $[M - H]^-$ , 243  $[M + H]^+$ ) was observed in transformant but not in untransformed *A. oryzae* extracts (Figure 2.32 – 2.33).



**Figure 2.32** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, Diode Array Detector (DAD) chromatogram of EXP4 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK*) transformant; **B**, DAD chromatogram of untransformed *A. oryzae* NSAR1. EXP ID see Table 2.9.

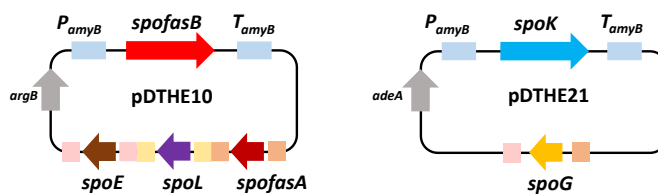
A scale-up fermentation (1L DPY liquid medium) of the producing transformant was carried out and extracts were subsequently employed for preparative LCMS (Section 6.2). The purification of **83** was achieved ( $4 \text{ mg}\cdot\text{L}^{-1}$ ; Table 7.4). The structure of **83** (Figure 2.33) was determined as the known compound 2*R*-2-(2-acrylic)-decanoic acid<sup>54</sup> by extensive 1D and 2D NMR analysis (see Chapter 7 for details). The *R* absolute configuration at C-2 was confirmed by comparing the optical rotation with literature.<sup>54</sup>



**Figure 2.33** UV and mass spectra (ES<sup>+</sup> and ES<sup>-</sup>) of **83**.

### 2.3.6.4b Co-expression of *spofasA*, *spofasB*, *spoE*, *spoL*, *spoK* and *spoG*

Oxidation of **83** obtained from EXP4 is necessary for sporothriolide **1** construction. Thus the dioxygenase SpoG was assumed to catalyse the oxidation step(s) during sporothriolide **1** biosynthesis. We cloned *spoG* in the pTAYAGSade plasmid (pDTHE21) containing *spoK*, and it was co-transformed to *A. oryzae* with pDTHE10 containing *spofasA*, *spofasB*, *spoE* and *spoL* (EXP5, Figure 2.34).

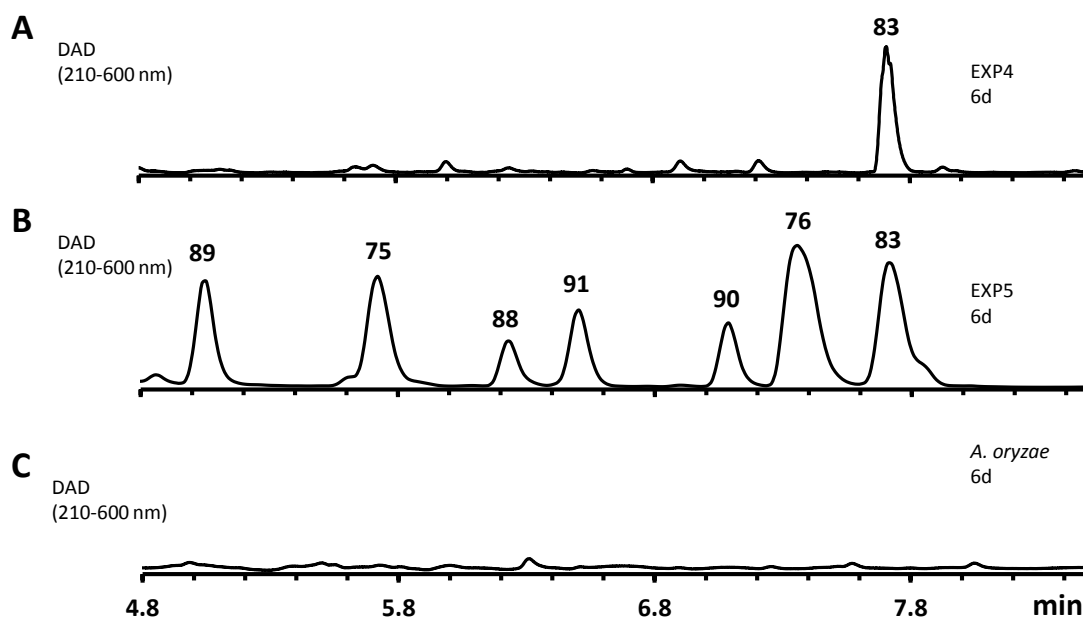


## EXP 5

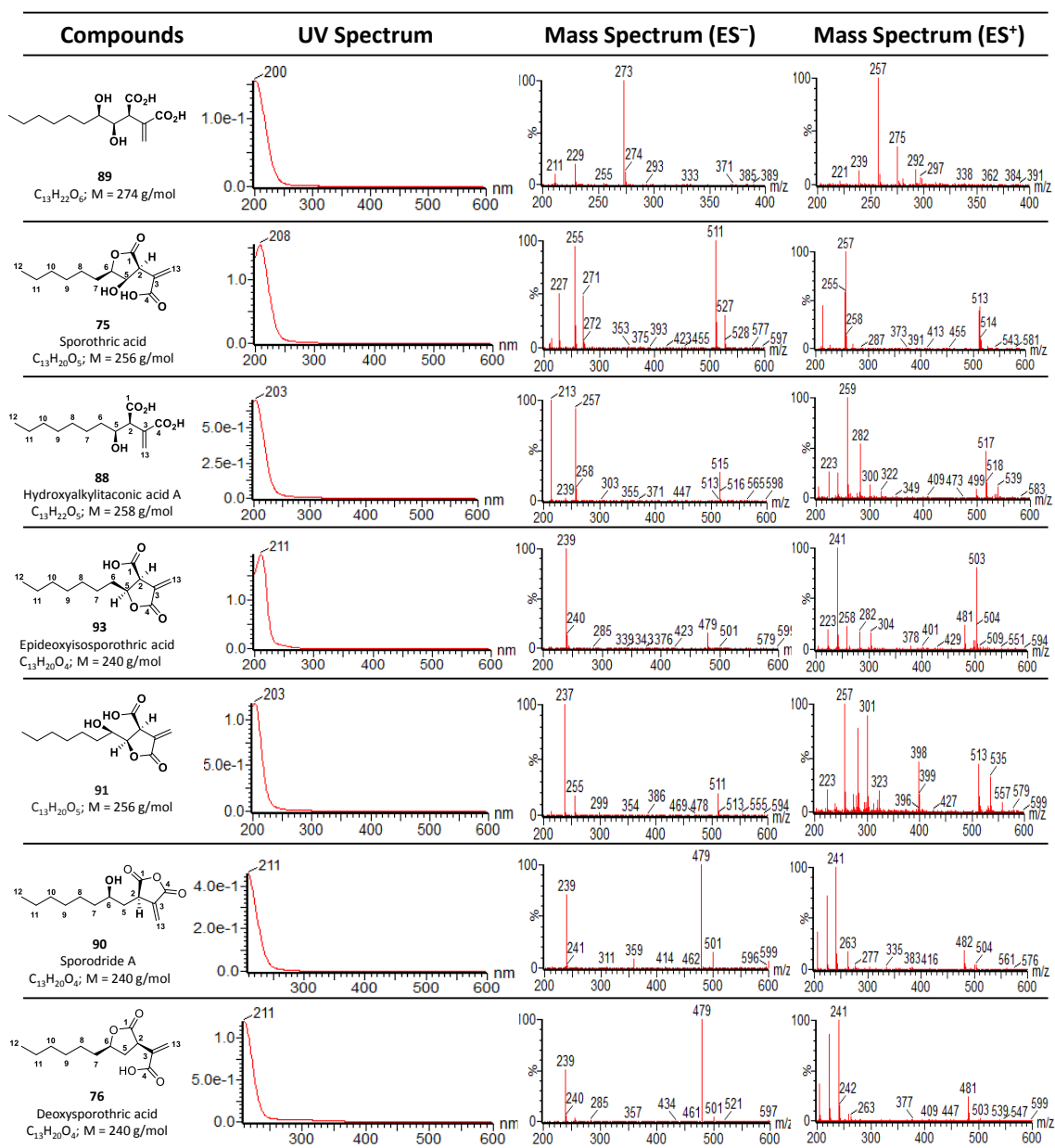
**Figure 2.34** Plasmids used in expression work of EXP5. EXP ID see Table 2.9.

Cultivation of all obtained transformants in DPY induction medium was then performed. Organic extracts of the cells and culture media were examined by LCMS (Figure 2.35). The addition of the putative dioxygenase *spoG* to the previous set of genes in *A. oryzae* results in the formation of a broad range of new compounds in addition to **83** (EXP5, Figure 2.35, Table 2.10). Four of the observed products (**75**, **76**, **88**, **90**) were purified and their structures were deduced by NMR (see Chapter 7 for spectra details).

Compound **83** (Figure 2.33) was observed in EXP5 as same as in EXP4. Compound **76** (Table 2.10) was characterised as deoxysporothric acid,<sup>53</sup> a known compound reported from *H. spongiphila* CLL 205. Compound **75** (Table 2.10) was shown to be a known compound sporothric acid<sup>51</sup> which was first found in *H. monticulososa* MUCL 54604, it was characterized by a lactone ring established between C-6 and the carboxylic acid group of the fatty acid chain.



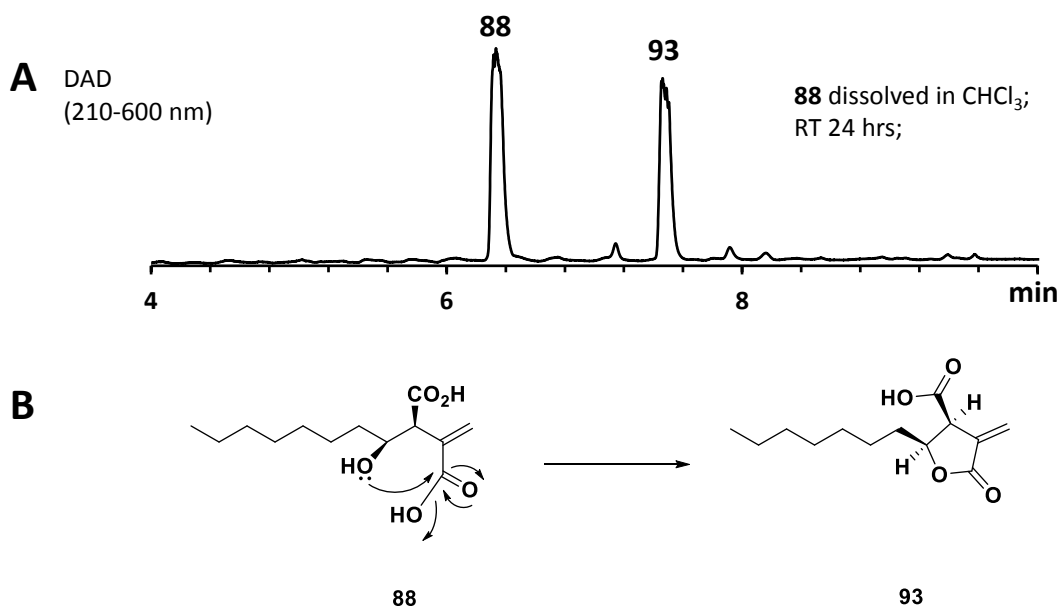
**Figure 2.35** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, Diode Array Detector (DAD) chromatogram of EXP4 (*spoFasA* + *spoFasB* + *spoE* + *spoL* + *spoK*) transformant; **B**, Diode Array Detector (DAD) chromatogram of EXP5 (*spoFasA* + *spoFasB* + *spoE* + *spoL* + *spoK* + *spoG*) transformant; **C**, DAD chromatogram of untransformed *A. oryzae* NSAR1. EXP ID see Table 2.9.



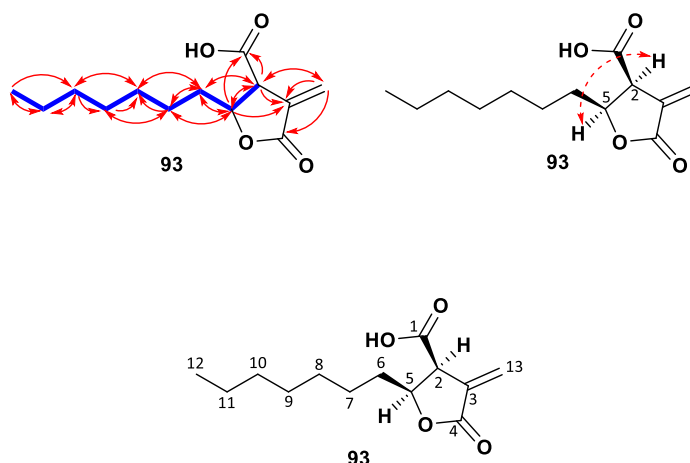
**Table 2.10** UV and mass spectra (ES<sup>+</sup> and ES<sup>-</sup>) of compounds from EXP5. See Figure 2.33 for compound **83**.

Compounds **88**, **89**, **90** and **91** were all new (Table 2.10). Compound **90** (Table 2.10) was designated as sporodride A to serve as the first sporothriolide anhydride structure, the details of structure elucidation and NMR, HR-ESI-MS spectra for **90** are given in Chapter 7.

Compound **88** (Table 2.10) was identified as the C-5 hydroxylated product of **83**. It was interesting that **88** always cyclized spontaneously to **93** in a short time (Figure 2.36), although the NMR measuring was operated right away. We could obtain the clean <sup>1</sup>H NMR of **88** (Chapter 7), but the <sup>13</sup>C NMR of **88** was a mixture of **88** and **93** (Figure 2.38), the later-measured 2D NMR measurement resulted in pure **93** (Chapter 7), which suggested a full conversion from **88** to **93**.



**Figure 2.36** **A**, HPLC chromatogram (DAD) of compound **88** transformed to compound **93** spontaneously; **B**, the proposed mechanism.



**Figure 2.37** Key HMBC,  $^1\text{H}$ - $^1\text{H}$  COSY and NOESY correlations of **93**.

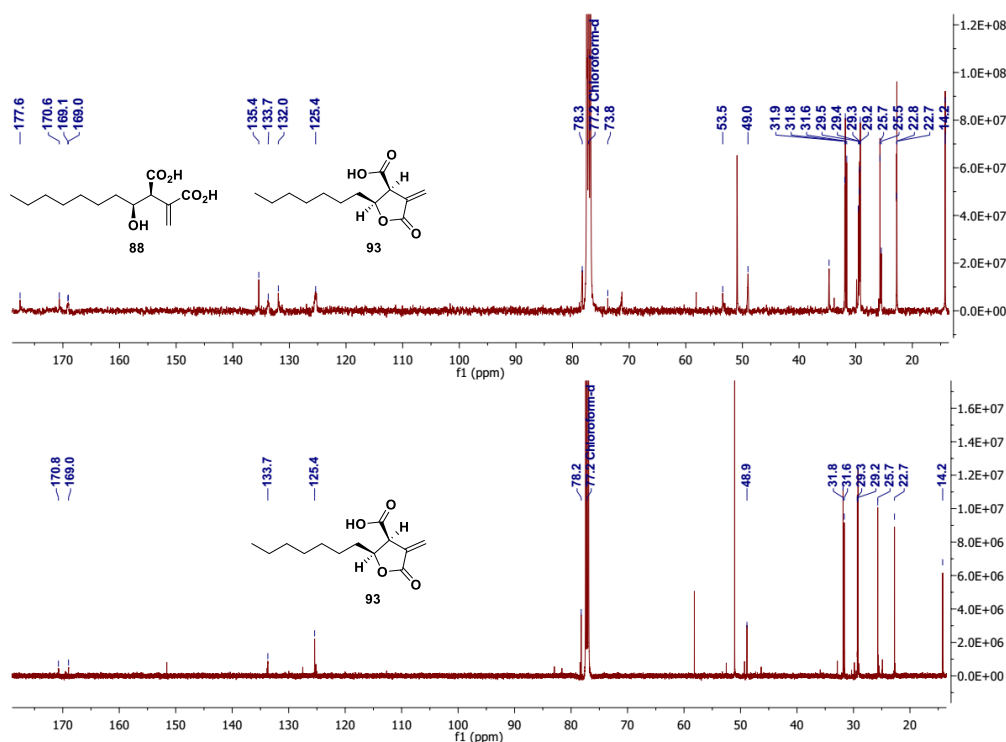
Compound **88** was isolated as a white powder with molecular formula  $\text{C}_{13}\text{H}_{22}\text{O}_5$  (calc.  $[\text{M} - \text{H}]^-$  HRMS 257.1389, measured 257.1386), indicating 3 degrees of unsaturation. Compound **93** is a white powder with molecular formula  $\text{C}_{13}\text{H}_{20}\text{O}_4$  (calc.  $[\text{M} - \text{H}]^-$  HRMS 239.1283, measured 239.1282), indicating **93** is a dehydration product from **88**. Extensive analysis by 1D and 2D NMR (Table 2.11) of **93** showed that the structure is quite similar with deoxyisoporphoric acid,<sup>54</sup> the difference is a loss of one methyl and one aromatic methine, but the addition of one aromatic quaternary carbon and one aromatic methylene. The molecular weight of **93** is also identical with deoxyisoporphoric acid, indicating the C-2/C-3 alkene shifted outside the lactone ring led to the C-3/C-13 alkene **93**. HMBC correlations from H-13 to C-2, C-3, and C-4, from H-2 to C-1, C-3, C-5, C-6, and C-13, as well correlations from H-2, H-5 and H-13 to C-3 support this speculation (Figure 2.37). The  $^1\text{H}$ - $^1\text{H}$  COSY correlations of H-5, H-6, and H-7 make a further confirmation

(Figure 2.37). The relative configuration of **93** was assigned by NOESY correlations of H-2 and H-5, suggesting these two protons are located on the same face (Figure 2.37).



pos.	88		93			
	$\delta_C$ /ppm	$\delta_H$ /ppm (J/Hz)	$\delta_C$ /ppm	$\delta_H$ /ppm (J/Hz)	HMBC	$^1H$ - $^1H$ COSY
1	177.6	-	170.8	-	-	-
2	53.5	3.54, 1H, d (6.8)	48.9	4.00, 1H, dt (2.2, 7.7)	1, 3, 5, 6, 13	5
3	135.4	-	133.7	-	-	-
4	169.1	-	169.0	-	-	-
5	73.8	4.13, 1H, m	78.2	4.65, 1H, m	1, 2, 3, 6, 7	2, 6
6	34.7	1.43-1.51, 2H, m	31.6	1.71, 2H, m	2, 5, 7, 8	5, 7
7	25.5	1.21-1.35, 2H, m	25.7	1.41, 1H, m 1.55, 1H, m	5, 6, 8, 9	6, 8
8	29.4	1.21-1.35, 2H, m	29.2	1.29, 2H, m	6, 7, 10	7, 9
9	29.5	1.21-1.35, 2H, m	29.3	1.32, 2H, m	7, 8	8, 10
10	31.9	1.21-1.35, 2H, m	31.8	1.26, 2H, m	8, 9, 11	9, 11
11	22.8	1.21-1.35, 2H, m	22.7	1.28, 2H, m	10, 12	10, 12
12	14.2	0.87, 3H, t (6.5)	14.2	0.88, 3H, t (7.2)	10, 11	11
13	132.0	6.56, 1H, s 5.98, 1H, s	125.4	6.43, 1H, d (2.1) 5.87, 1H, d (2.1)	2, 3, 4	-

**Table 2.11**  $^1H$  NMR (400 MHz) data and  $^{13}C$  NMR (100 MHz) data for **88** in  $CDCl_3$ .  $^1H$  NMR (500 MHz) data and  $^{13}C$  NMR (125 MHz) data for **93** in  $CDCl_3$ .



**Figure 2.38**  $^{13}C$  NMR of compound **88** and **93**.

Compound **88** was the precursor of lactone **93**, thus a reasonable scaffold for **88** was proposed. The carbon and proton signals for **88** could be easily assigned based on **93** (Table 2.11). Also, the relative configuration of **93** should keep consistent with **88**. Lactone **93** was designated as epideoxyisosporthric acid as it's a double bond isomer of deoxyisosporthric acid. Linear diacid **88** was defined as hydroxyoctaylitaconic acid A.

#### 2.3.6.4c Co-expression of *spofasA*, *spofasB*, *spoE*, *spoL*, *spoK*, *spoG* and *spoH*

Two enzymes predicted as lactonases are encoded by the *spo* BGC (Figure 2.27A), and their roles played in sporothriolide **1** biosynthesis were still not clear. Since there are two furan rings in the sporothriolide **1** structure, we speculated about the function of these lactonases. For example, each enzyme could be responsible for one lactone ring closure, or one lactonase catalyses two ring formations and the another lactonase performs hydrolysis of the product. Because sporothriolide is a potent antifungal agent,<sup>55</sup> the existence of a self-resistance gene in *spo* BGC is possible.

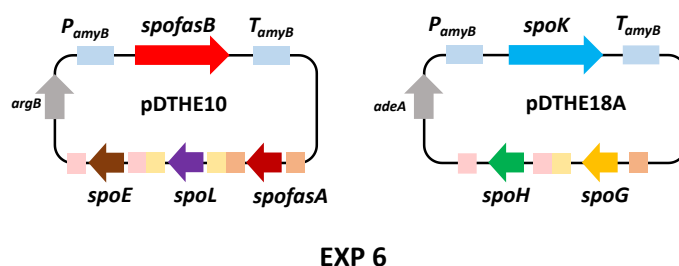
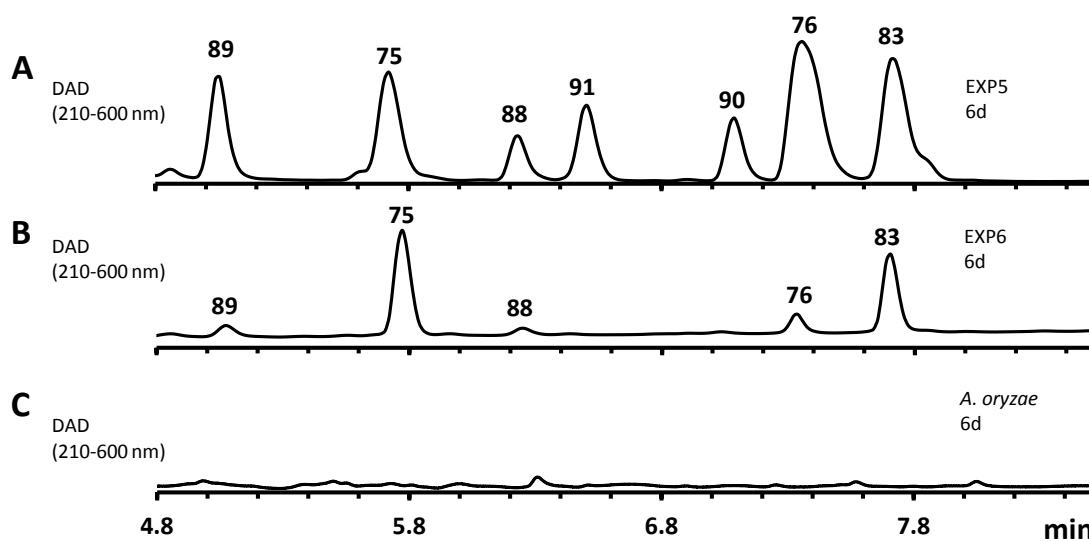


Figure 2.39 Plasmids used in expression work of EXP6. EXP ID see Table 2.9.

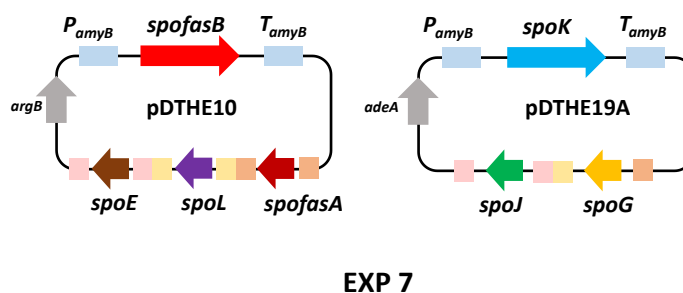
To test the ability of lactonase SpoH, we cloned *spoH* into the plasmid pDTHE18A containing *spoK* and *spoG*. It was co-transformed with plasmid pDTHE10 (*spofasA*, *spofasB*, *spoE*, *spoL*) into *A. oryzae*. After three rounds of fungal transformation (EXP6, Figure 2.39), the obtained transformants were cultivated in DPY medium at 28 °C for 6 days and 110 rpm. The cells and culture media were extracted separately and extracts were submitted for LCMS analysis. Several intermediates were observed in LCMS chromatograms, such as **76**, **83**, **88**, **89** and the predominant product **75**, whereas sporothriolide **1** could not be detected (Figure 2.40). Therefore, the function of SpoH was unable to be deduced from this EXP6, at least it cannot convert **89** or **75** to sporothriolide.



**Figure 2.40** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, Diode Array Detector (DAD) chromatogram of EXP5 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG*) transformant; **B**, Diode Array Detector (DAD) chromatogram of EXP6 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*) transformant; **C**, DAD chromatogram of untransformed *A. oryzae* NSAR1. EXP ID see Table 2.9.

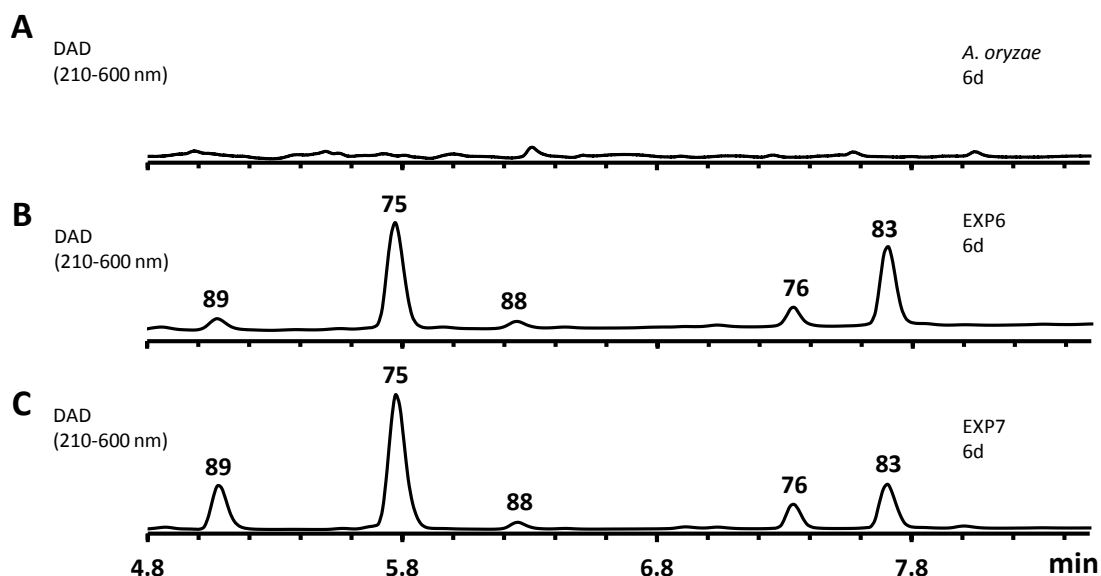
#### 2.3.6.4d Co-expression of *spofasA*, *spofasB*, *spoE*, *spoL*, *spoK*, *spoG* and *spoJ*

Another lactonase, encoded by *spoJ* is also found in the *spo* BGC. To test the function of *spoJ*, it was cloned into the plasmid pDTHE19A containing *spoK* and *spoG*. It was co-transformed to *A. oryzae* with plasmid pDTHE10 containing *spofasA*, *spofasB*, *spoE* and *spoL*. After three rounds of fungal transformation (EXP7, Figure 2.41), the obtained transformants were cultured in DPY medium and then extracted for LCMS analysis (Figure 2.42).



**Figure 2.41** Plasmids used in expression work of EXP7. EXP ID see Table 2.9.

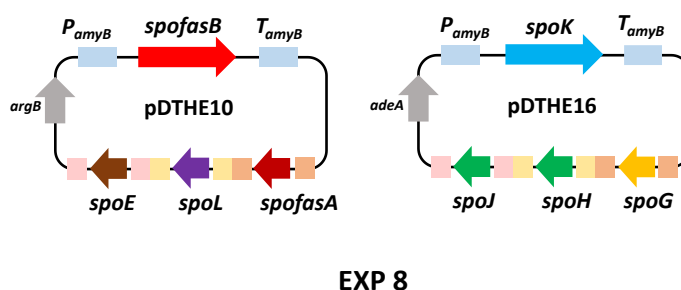
The metabolic profile (Figure 2.42) in EXP7 was almost identical to that observed in EXP6 (Figure 2.40). Intermediates **76**, **83**, **88**, **89** and predominant product **75** were observed, however sporothriolide **1** was not observed (Figure 2.42). Similar to SpoH (EXP6), lactonase SpoJ (EXP7) appears not to accept **89** or **75** as a substrate to catalyse ring closure to sporothriolide **1**.



**Figure 2.42** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, DAD chromatogram of untransformed *A. oryzae* NSAR1; **B**, Diode Array Detector (DAD) chromatogram of EXP6 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*) transformant; **C**, Diode Array Detector (DAD) chromatogram of EXP7 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoJ*) transformant. EXP ID see Table 2.9.

### 2.3.6.4e Co-expression of *spofasA*, *spofasB*, *spoE*, *spoL*, *spoK*, *spoG*, *spoH* and *spoJ*

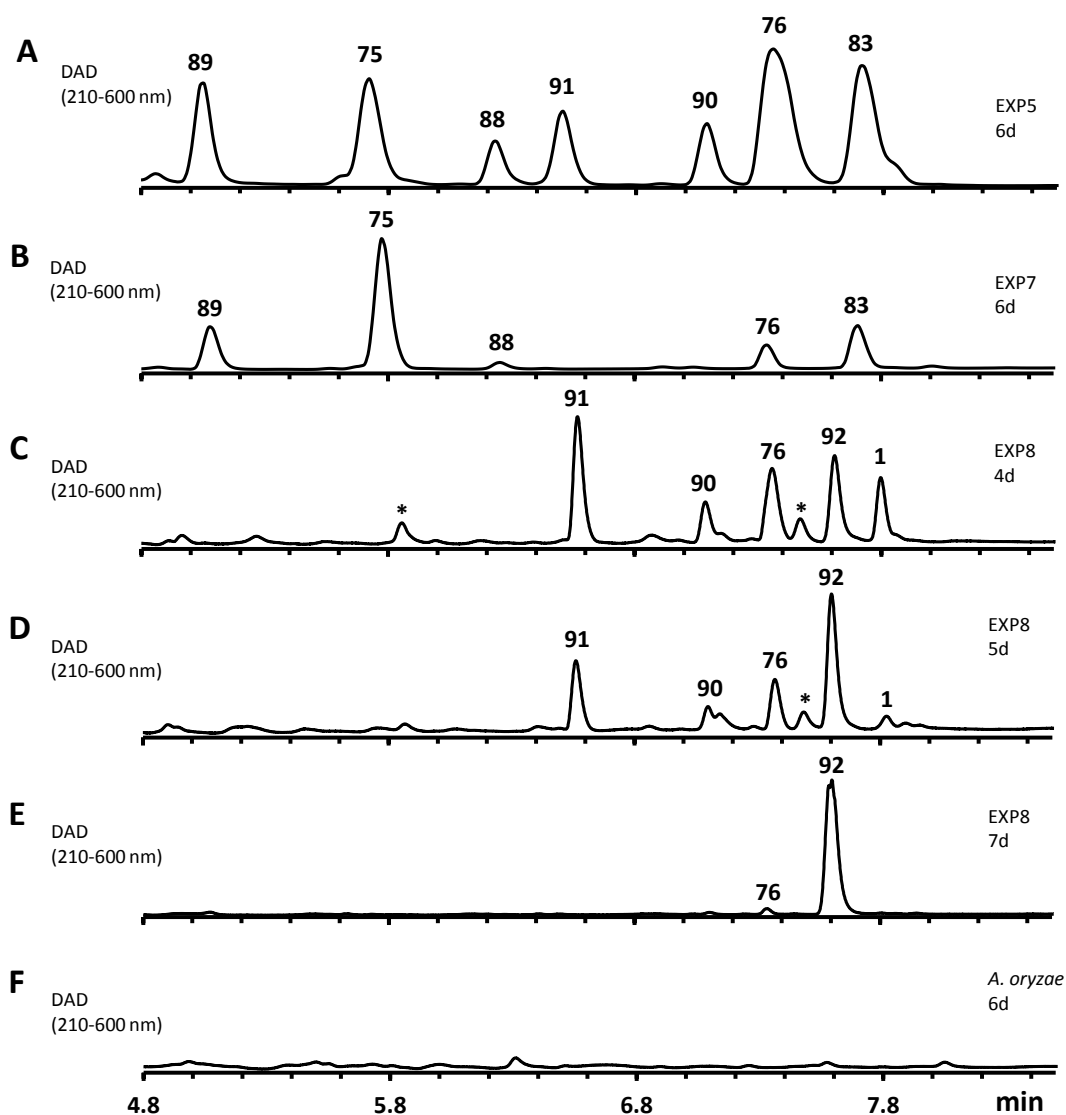
All eight genes (*spofasA*, *spofasB*, *spoE*, *spoL*, *spoK*, *spoG*, *spoH* and *spoJ*) from *spo* BGC (Figure 2.27A) were co-expressed in *A. oryzae*, and plasmids pDTHE10 and pDTHE16 were used for three repeated fungal transformation experiment (EXP8, Figure 2.43). Obtained transformants were cultured in DPY medium under the same conditions used previously. The extracts of the media were then analysed by LCMS. However, heterologous expression of all eight genes led to the observation of a range of new metabolites in *A. oryzae* including sporothriolide **1** after 4 days (Figure 2.44C).



**EXP 8**

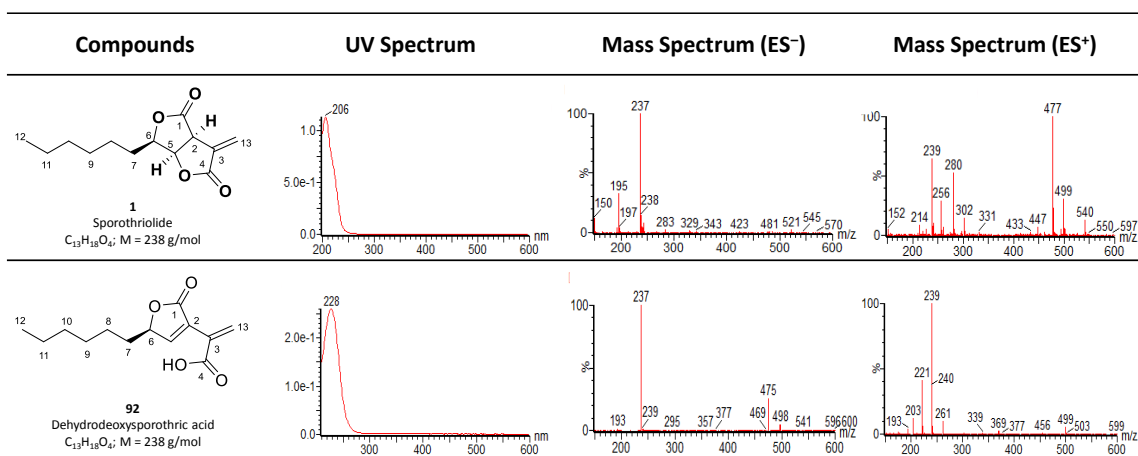
**Figure 2.43** Plasmids used in expression work of EXP8. EXP ID see Table 2.9.





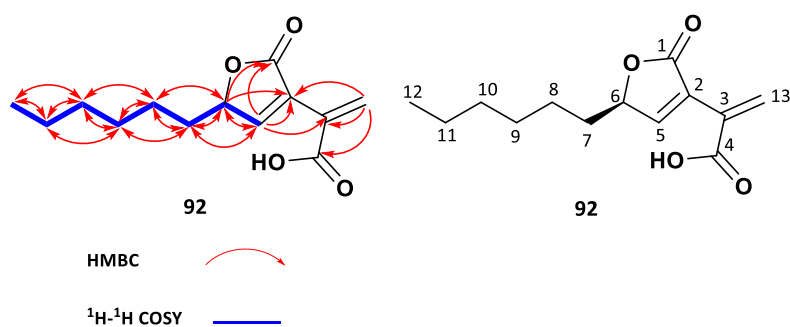
**Figure 2.44** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, Diode Array Detector (DAD) chromatogram of EXP5 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG*) transformant; **B**, Diode Array Detector (DAD) chromatogram of EXP7 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoJ*) transformant; **C-E**, Diode Array Detector (DAD) chromatogram of EXP8 (*spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH* + *spoJ*) transformant at 4d, 5d, and 7d, respectively; **F**, DAD chromatogram of untransformed *A. oryzae* NSAR1. EXP ID see Table 2.9. \*Unrelated compounds.

After longer fermentation up to 7 days, these were converted to **92** (Figure 2.44E) which has the same molecular formula (Table 2.12) as sporothriolide ( $C_{13}H_{18}O_4$ ;  $M = 238$  g/mol), but a slightly different retention time (See Table 7.4 for compounds details).



**Table 2.12** UV and mass spectra (ES<sup>+</sup> and ES<sup>-</sup>) of compounds from EXP8. See Table 2.9 for other compounds.

Purification of **92** (*ca* 8 mg·L<sup>-1</sup>) was achieved and full 1D and 2D NMR analysis was performed (Table 2.13; See full spectra in Chapter 7). Compound **92** was isolated as a white powder. The molecular formula C<sub>13</sub>H<sub>18</sub>O<sub>4</sub> is in accordance with its HRMS data, indicating 5 degrees of unsaturation. The 1D and 2D NMR data (Table 2.13) revealed that the scaffold of compound **92** shares a high similarity with deoxysporothric acid **76**, except the absence of one aliphatic methylene group and one aliphatic methine group, but the presence of two sp<sup>2</sup>-hybridized carbons. Together with the molecular weight of **92** is 2 Da less than **76**, which suggests dehydrogenation. The <sup>1</sup>H-<sup>1</sup>H COSY correlations of H-5, H-6 and H-7, as well as the HMBC correlations from H-5 to C-1, C-2, C-3, C-6, and C-7, and the correlations from H-6 and H-13 to C-2 indicate the double bond located at C-2 and C-5 (Figure 2.45). Thus the planar structure of **92** was solved.



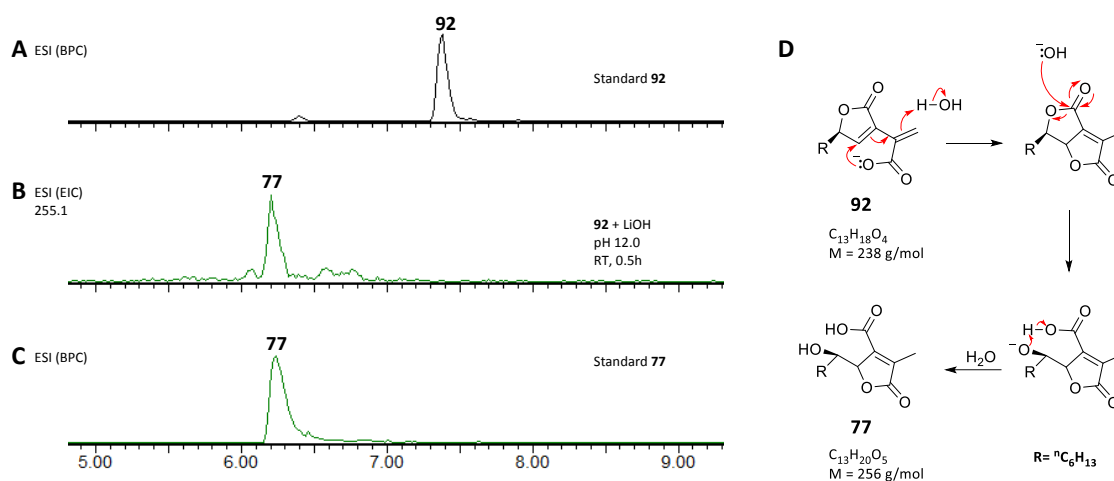
**Figure 2.45** Key HMBC and <sup>1</sup>H-<sup>1</sup>H COSY correlations of **92**.



pos.	1		92		HMBC	<sup>1</sup> H- <sup>1</sup> H COSY
	$\delta_C$ /ppm	$\delta_H$ /ppm (J / Hz)	$\delta_C$ /ppm	$\delta_H$ /ppm (J / Hz)		
1	167.6	-	171.7	-	-	-
2	46.3	4.01, 1H, dt (6.8, 2.0)	124.9	-	-	-
3	130.0	-	128.6	-	-	-
4	172.2	-	170.4	-	-	-
5	77.3	5.15, 1H, dd (6.8, 4.7)	153.4	7.95, 1H, d (1.7)	1, 2, 3, 6, 7	6
6	82.9	4.65, 1H, ddd (7.9, 6.3, 4.7)	80.8	4.99, 1H, m	1, 2, 5, 7, 8,	5, 7
7	29.0	1.88, 2H, m	33.4	1.79, 1H, m 1.69, 1H, m	5, 6, 8, 9	6, 8
8	25.5	1.49, 2H, m	25.1	1.46, 2H, m	6, 7, 9, 10	7, 9
9	29.1	1.38, 2H, m	29.1	1.34, 2H, m	7, 8, 10, 11	8, 10
10	31.7	1.30, 2H, m	31.7	1.28, 2H, m	8, 9, 11, 12	9, 11
11	22.6	1.30, 2H, m	22.6	1.29, 2H, m	9, 10, 12	10, 12
12	14.2	0.89, 3H, t (7.0)	14.1	0.88, 3H, t (7.0)	10, 11	11
13	127.5	6.47, 1H, d (2.1) 6.16, 1H, d (2.1)	133.7	7.19, 1H, s 6.81, 1H, s	2, 3, 4	-

**Table 2.13** <sup>1</sup>H NMR (500 MHz) data and <sup>13</sup>C NMR (125 MHz) data for **1** in CDCl<sub>3</sub>. <sup>1</sup>H NMR (600 MHz) data and <sup>13</sup>C NMR (150 MHz) data for **92** in CDCl<sub>3</sub>.

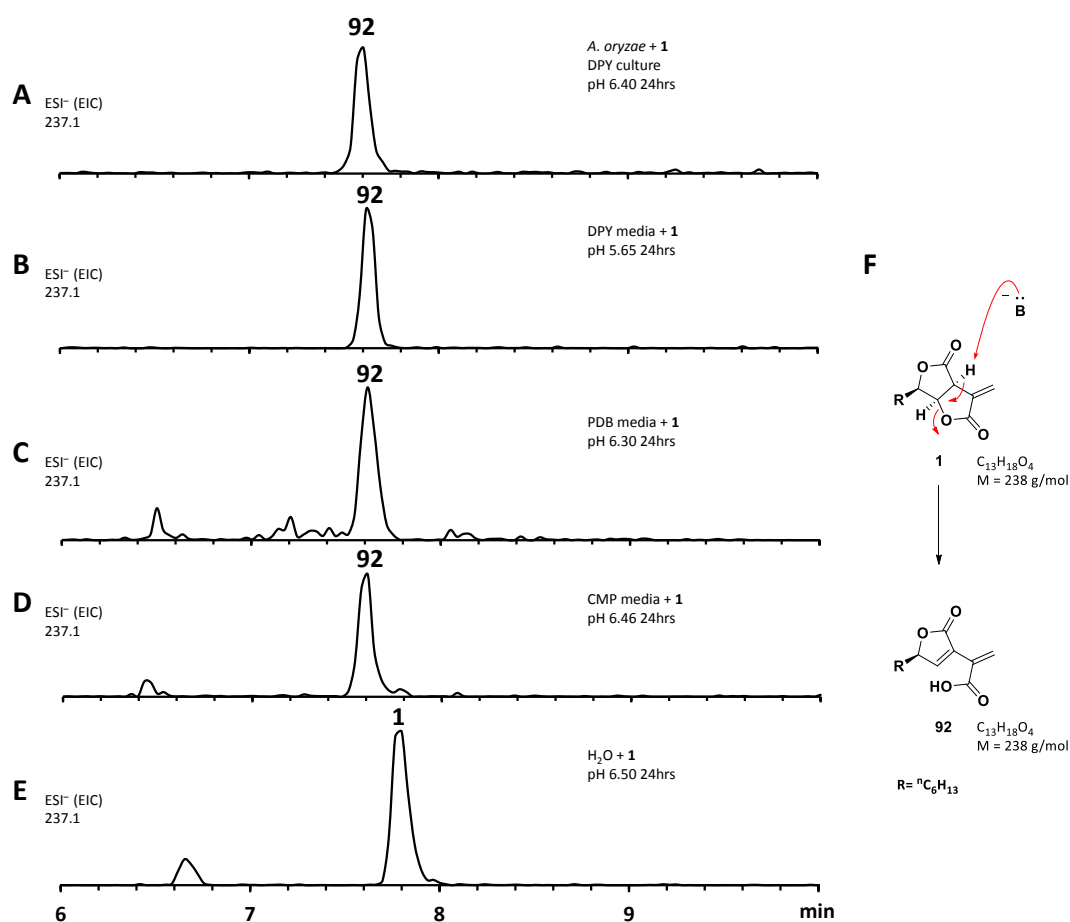
The absolute configuration of **92** was determined to be 6*R* by chemical conversion. Under LiOH (1 M) base condition, **92** was chemically transformed to isosporothric acid **77** rapidly (Figure 2.46). Lactone **92** was a new compound until our work, thus we designated **92** as dehydrodeoxysporothric acid.



**Figure 2.46** Structure transformation of **92**. **A** and **C**, ESI<sup>-</sup> spectra (BPC, Base Peak Chromatograms) of standard **92** and **77**; **B**, Extracted Ion Chromatograms (255.1 [M-H]<sup>-</sup>) of 1 mg **92** dissolved in 2 mL methanol then treated with 0.3 mL LiOH (1 M) to pH 12, RT for 0.5 h; **D**, proposed mechanism from **92** to **77**. R = <sup>n</sup>C<sub>6</sub>H<sub>13</sub>.

The structure of **92** showed it to be the monocyclic analogue of **1**, presumably either resulting from an eliminative ring-opening of **1** itself or being a biosynthetic precursor for sporothriolide **1**. Therefore we designed an experiment to test the stability of sporothriolide **1** (Figure 2.47).

Sporothriolide **1** (0.5 mg) was fed to *A. oryzae* NSAR1 grown in 20 mL DPY culture, or incubated with 20 mL of various media (DPY, PDB, CMP, water) under normal fermentation conditions (28 °C, 110 rpm) for 24 h. Mixtures were extracted with equal amounts of ethyl acetate and analysed by LCMS. Degradation of **1** to **92** was observed under all conditions except in water (Figure 2.47). Additionally, the pH value of each mixture was measured to investigate the possibility of pH-dependent elimination. Values between 5.65 and 6.50 indicate that the pH of the solutions has likely no influence on the degradation and that medium-specific ingredients are likely responsible for the conversion of **1** to **92**.



**Figure 2.47** Extracted Ion Chromatograms (237.1 [M-H]<sup>-</sup>) from the sporothriolide **1** stability assay with **1** being incubated under different conditions. **A**, with *A. oryzae* NSAR1 DPY culture, pH 6.40, 24 hrs; **B**, with DPY media, pH 5.65, 24 hrs; **C**, with PDB media, pH 6.30, 24 hrs; **D**, with CMP media, pH 6.46, 24 hrs; **E**, with water, pH 6.50, 24 hrs; **F**, proposed mechanism from **1** to **92**. R = <sup>n</sup>C<sub>6</sub>H<sub>13</sub>.

### 2.3.7 *In Vitro* Activity Assay with SpoG and SpoI

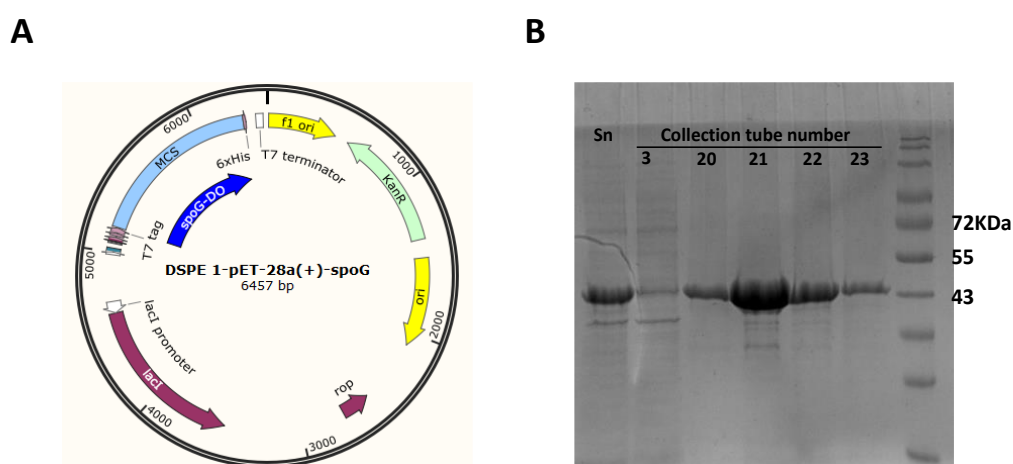
#### 2.3.7.1 Expression, Purification and Activity Assay of SpoG

As neither the presence of **87** nor **89** could be unambiguously proved in the extracts of the *A. oryzae* transformant due to their instability or quick conversion, the function of SpoG was further evaluated *in vitro*.

For expression of SpoG in *Escherichia coli* BL21 (DE3),<sup>48</sup> the expression plasmid pET-28a (+) (Novagen) was digested with *Bam*HI and *Xho*I (New England BioLabs) restriction enzymes. *H. monticulosa* MUCL 54604 cDNA was used as the DNA template to amplify the SpoG coding sequence with the primer pair P64 + P65 (Table 6.2). T4 ligase (New England BioLabs) was utilized for ligation of the restriction digested vector and PCR fragments. Transformation of competent cells was performed based on a standard *E. coli* transformation protocol (Section 6.1.2.1).

A pre-culture was grown overnight in 10 mL LB medium containing 50  $\mu\text{g}\cdot\text{mL}^{-1}$  kanamycin at 37 °C 200 rpm. 900  $\mu\text{L}$  of the pre-culture was used to inoculate three flasks containing 50 mL LB medium with 50  $\mu\text{g}\cdot\text{mL}^{-1}$  kanamycin. Cells were grown at 37 °C and 200 rpm until an OD600 value between 0.4-0.6 was reached. Isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG, 0.25 M stock) was added to a final concentration of 0.3 mM, and cells were incubated for another 16 h at 25 °C and 160 rpm. Cells were harvested by centrifugation (8000  $\times$  g, 3 min) at 4 °C and resuspended in loading buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 10% glycerol (v/v)) and lysed by sonication for 10 min (Pulse 10 s / 10 s). Cell debris was removed from the total lysate by centrifugation (10.000  $\times$  g, 20 min, 4 °C).

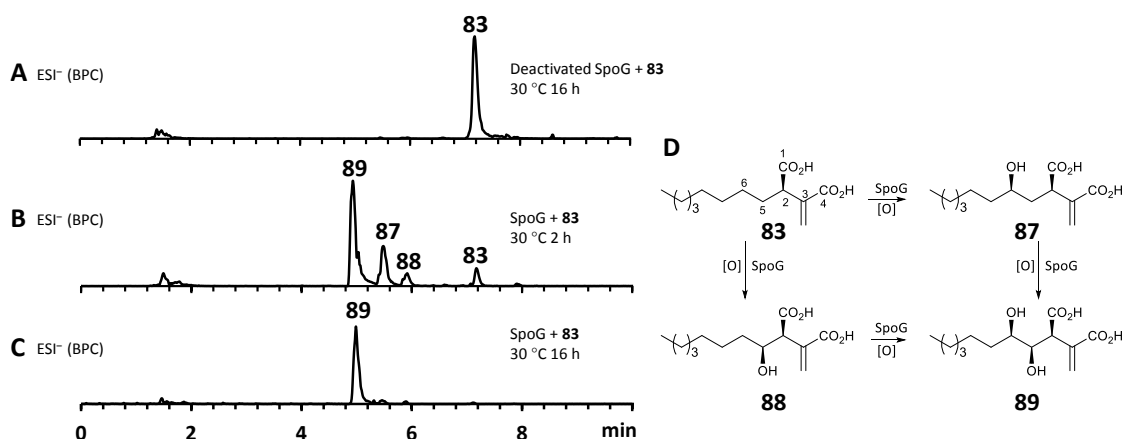
SpoG containing a his<sub>6</sub>-tag (44.7 KDa) was purified by FPLC (Fast protein liquid chromatography) as a soluble protein. The loaded Ni-NTA column with bound protein was eluted with elution buffer (loading buffer + 500 mM imidazole). The buffer was exchanged to storage buffer (50 mM Tris-HCl pH 7.5, 20% glycerol (v/v)) by ultrafiltration with a molecular weight cut-off of 30 KDa. The purity of the protein was assessed by SDS-PAGE (Figure 2.48).



**Figure 2.48** **2.45 A**, plasmid used in SpoG expression; **B**, SDS-PAGE of SpoG after purification. 12% SDS gel run at 180 V for 1h. Sn, centrifuge supernatant of the lysed *E. coli* cells.

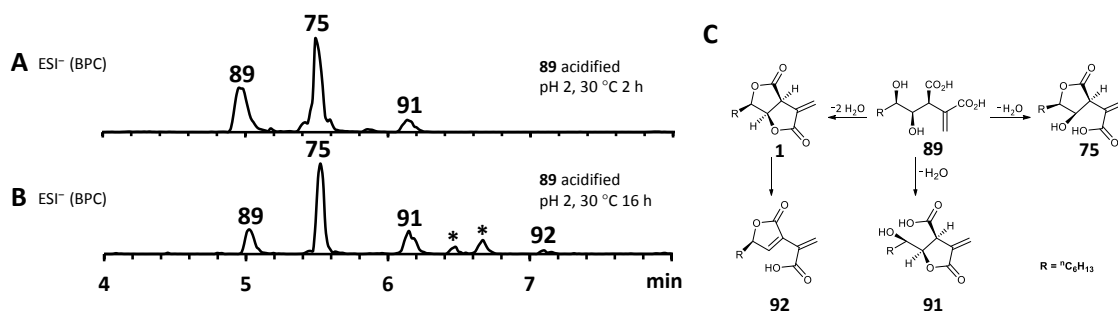
Sequence analysis indicated that SpoG requires  $\alpha$ -ketoglutarate as a cofactor. *In vitro* assays were conducted by incubating SpoG (50  $\mu\text{M}$ ) with intermediate **83** (2.5 mM), tris buffer (50 mM, pH

7.5; Table 6.3), ascorbate (4 mM),  $\alpha$ -ketoglutarate (4 mM), and  $\text{FeSO}_4$  (0.2 mM) at 30 °C for 2 h, the total volume of reaction mixture is 50  $\mu\text{L}$ . After that, 100  $\mu\text{L}$  of chloroform was added into the reaction mixture to precipitate the protein, The mixture was then vortexed for 1 min and centrifuged at 15,000  $\times$  g for 3 min, then the top layer of supernatant (aqueous phase) was directly subjected for analytical LCMS. The negative control was conducted by using deactivated SpoG (boiled at 95 °C, 20 min) under the same conditions as above. LCMS (Figure 2.49) showed the formation of three products, the main compound of which was identical with **89** (by mass, UV and retention time) from the heterologous expression experiments (Table 2.10). In addition, **88** was formed as a minor product alongside another compound with identical mass likely representing its proposed regioisomer **87** (Figure 2.49B). Extended incubation time up to 16 h resulted in a complete conversion of **87** and **88** into **89** indicating that **87** and **88** are precursors of **89** and that SpoG catalyses two consecutive rounds of hydroxylation (Figure 2.49C).



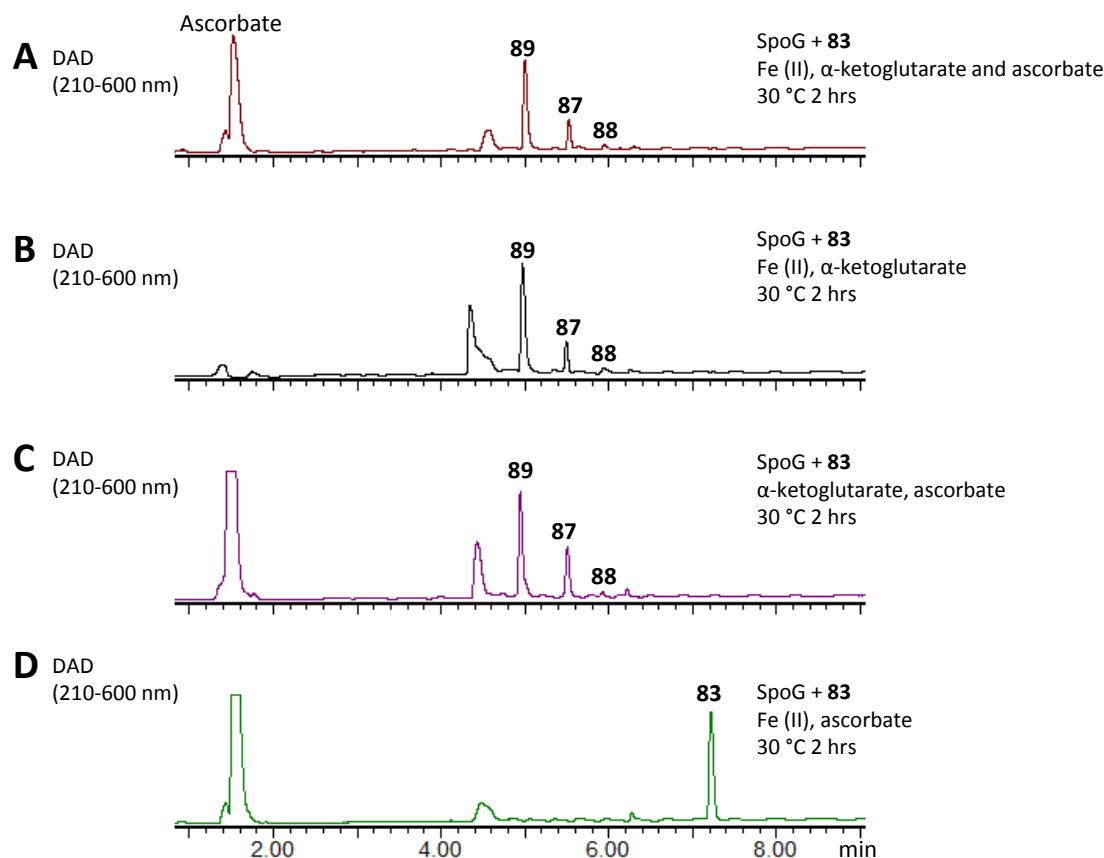
**Figure 2.49** *In vitro* assay of SpoG using purified proteins. **A**, ESI<sup>-</sup> trace of boiled SpoG (50  $\mu\text{M}$ ) incubated with **83** (2.5 mM), tris buffer (50 mM, pH 7.5), ascorbate (4 mM),  $\alpha$ -ketoglutarate (4 mM), and  $\text{FeSO}_4$  (0.2 mM) at 30 °C, 16 h; **B**, ESI<sup>-</sup> trace of SpoG incubated with **83** under the same conditions for 2 h; **C**, ESI<sup>-</sup> trace of SpoG incubated with **83** under the same conditions for 16 h; **D**, proposed chemical steps catalysed by SpoG.

Compound **75**, which occurred in the heterologous host (EXP5, Figure 2.35), was not observed in the *in vitro* experiments, suggesting that *A. oryzae* can catalyse this transformation. As previous reports demonstrated that lactonization is pH-dependent,<sup>69</sup> *in vitro* generated **89** was acidified to pH 2 leading to the partial conversion of **89** into **75** and minor quantities of **91** after 2 h of incubation (Figure 2.50A). In addition to **75** and **91**, very small amounts of **92** (derived from **1**) could be found after increased incubation periods (Figure 2.50B).



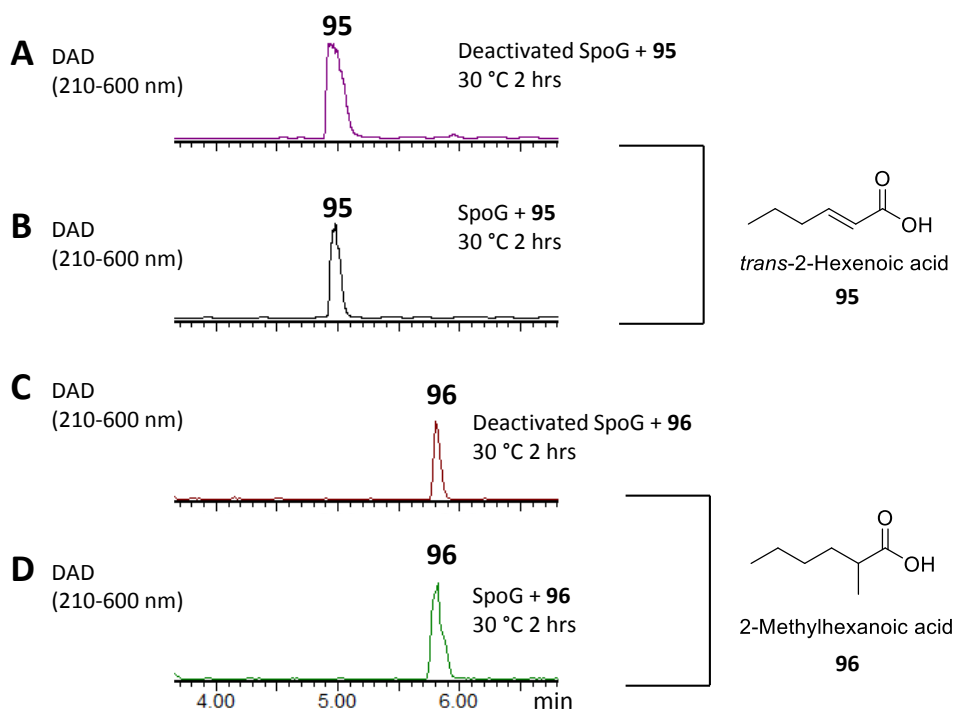
**Figure 2.50** Lactonization under acidic conditions. **A**, ESI<sup>-</sup> spectrum of **89** (obtained directly *via* incubation of **83** with SpoG) acidified with formic acid to pH 2, at 30 °C for 2 h; **B**, as **A**, 30 °C for 16 h; **C**, proposed chemical steps during the acidification. \*Unknown degradation compounds with the same mass data (ESI-MS  $m/z$  661 [M-H]<sup>-</sup>, 663 [M+H]<sup>+</sup>). R = <sup>n</sup>C<sub>6</sub>H<sub>13</sub>.

The cofactor dependence of SpoG was also studied by excluding ascorbate,  $\alpha$ -ketoglutarate, and FeSO<sub>4</sub> individually. The reaction conditions and extraction method are the same as described above. Results (Figure 2.51) showed that  $\alpha$ -ketoglutarate is essential for turnover *in vitro*, but iron and ascorbate can be omitted.



**Figure 2.51** The cofactor dependence of SpoG (50  $\mu$ m) incubated with **83** (2.5 mM), tris buffer (50 mM, pH 7.5), ascorbate (4 mM),  $\alpha$ -ketoglutarate (4 mM), and FeSO<sub>4</sub> (0.2 mM) at 30 °C 2 hrs. DAD chromatograms of extracts from assays: **A**, including Fe (II),  $\alpha$ -ketoglutarate and ascorbate; **B**, excluding ascorbate; **C**, excluding Fe (II); **D**, excluding  $\alpha$ -ketoglutarate.

Additionally, the alternative monocarboxylic acid substrates *trans*-2-hexenoic acid **95** and 2-methylhexanoic acid **96** were tested with SpoG. Reaction composition, conditions and extraction were as previously described. Results showed that SpoG does not accept these alternative substrates (Figure 2.52).



**Figure 2.52** The substrate promiscuity study of SpoG. **A**, DAD chromatograms of boiled SpoG (50  $\mu$ m) incubated with *trans*-2-hexenoic acid **95** (2.5 mM), tris buffer (50 mM, pH 7.5), ascorbate (4 mM),  $\alpha$ -ketoglutarate (4 mM), and  $\text{FeSO}_4$  (0.2 mM) at 30 °C, 2 h; **B**, DAD chromatograms of SpoG (50  $\mu$ m) incubated with with **95**, condition as previous; **C**, DAD chromatograms of boiled SpoG (50  $\mu$ m) incubated with with 2-Methylhexanoic acid **96**, condition as previous; **D**, DAD chromatograms of SpoG (50  $\mu$ m) incubated with with **96**, condition as previous.

### 2.3.7.2 Expression, Purification and Activity Assay of SpoI

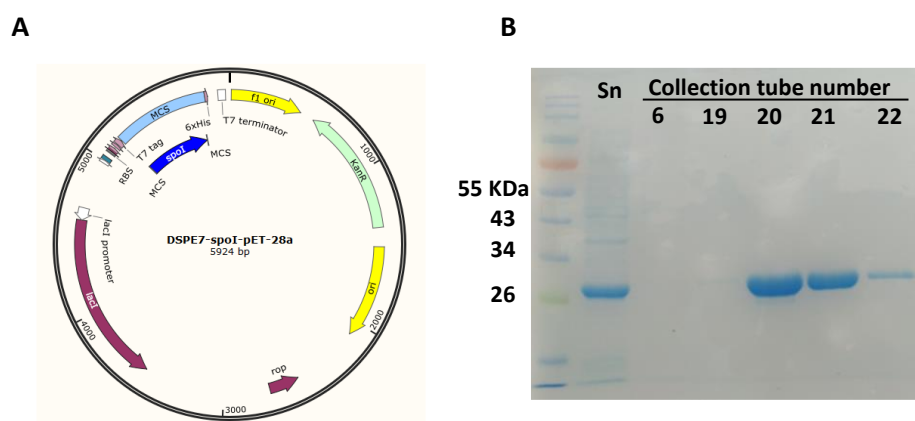
As the biosynthesis of **1** can be fully explained by the activity of the investigated proteins, the role of the unknown protein SpoI remained obscure. The function of *spoI* was especially intriguing as it shows high expression levels under both producing and non-producing conditions for **1** (Section 2.3.4.3, Table 2.6). BLASTp<sup>61</sup> analysis of SpoI showed it is a member of the cupin-domain-containing proteins. It is 35.7% identical to VirC which is involved in the fungal trichoxide biosynthetic pathway,<sup>70</sup> although VirC is also of unknown function. Structural analysis using PHYRE-2<sup>62</sup> also showed no relationship to known functionally characterized enzymes.

Thus we planned to express the SpoI protein and test its function *in vitro*. For expression of SpoI in *E. coli* BL21 (DE3), the expression plasmid pET-28a (+) was digested with *Bam*HI and *Eco*RI (New England BioLabs) restriction enzymes. *H. monticulosa* MUCL 54604 cDNA was used as the DNA template to amplify the SpoI coding sequence with the



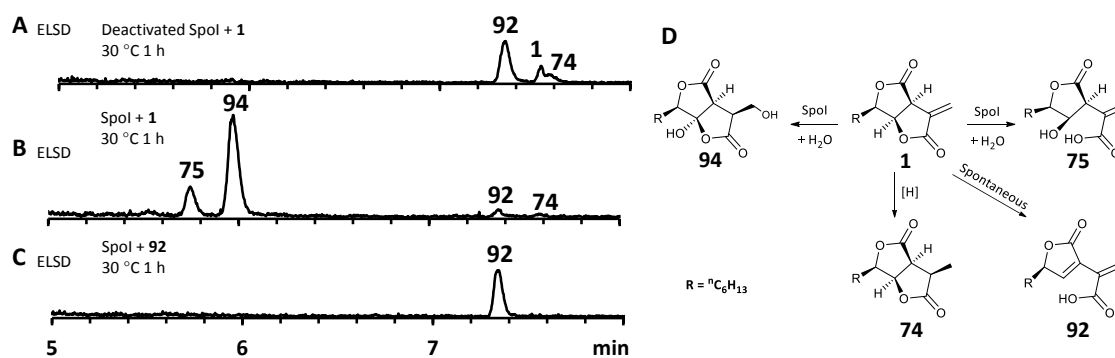
primer pair P66 + P77 (Table 6.2). T4 ligase (New England BioLabs) was utilized for ligation of the restriction digested vector and PCR fragments. Transformation of competent cells was performed based on a standard *E. coli* transformation protocol.

Expression of SpoI was achieved as described in Section 2.3.7.1 (same as SpoG), Purification of the protein in soluble form was achieved by FPLC, except the molecular weight cut-off used is 10 KDa (SpoI with his<sub>6</sub>-tag: 24.1 kDa). The purity of the protein was assessed by SDS-PAGE (Figure 2.53).



**Figure 2.53** **A**, plasmid used in SpoI expression; **B**, SDS-PAGE of SpoI after FPLC purification. 12% SDS gel run at 180 V for 50 min. Sn, centrifuge supernatant of the lysed *E. coli* cells.

*In vitro* assays were conducted by incubating SpoI (50  $\mu$ M) with sporothriolide **1** (2.5 mM) in PBS buffer (50  $\mu$ L, pH 7.5; Table 6.3) at 30  $^{\circ}$ C for 1 h. After that, 50  $\mu$ L of acetonitrile was added into the reaction mixture to precipitate the protein. The mixture was then vortexed for 1 min and centrifuged at 15,000  $\times$  g for 3 min, then the supernatant was directly subjected for analytical LCMS. The negative control was conducted by using deactivated SpoI (boiled at 95  $^{\circ}$ C, 20 min) under the same condition as above (Figure 2.54).



**Figure 2.54** *In vitro* assay of SpoI using purified proteins. **A**, ELSD chromatogram of boiled SpoI (50  $\mu$ M) incubated with **1** (2.5 mM) in PBS buffer (pH 7.5) at 30  $^{\circ}$ C for 1 h; **B**, ELSD chromatogram of SpoI (50  $\mu$ M) incubated with **1** (2.5 mM) in PBS buffer (pH 7.5) at 30  $^{\circ}$ C for 1 h; **C**, ELSD chromatogram of SpoI (50  $\mu$ M) incubated with **92** (2.5 mM) in PBS buffer (pH 7.5) at 30  $^{\circ}$ C for 1 h; **D**, proposed chemical steps catalysed by SpoI. R =  $^{13}$ C<sub>6</sub>H<sub>13</sub>.

Incubation of SpoI with **1** resulted in the formation of small amounts of **75** and a new compound **94** (Figure 2.54B) which share the same molecular formula (HRMS  $m/z$  255.1217 [M – H]<sup>-</sup>; calc. for C<sub>13</sub>H<sub>19</sub>O<sub>5</sub>, 255.1232). As the formation of **92** could be observed during the experiment it was not clear if **92** was used as substrate instead of **1**. To test this, **92** was incubated with SpoI which did not lead to product formation demonstrating that **1** is indeed the substrate of SpoI (Figure 2.54C). Isolation of **94** by upscaling the *in vitro* reaction proved unsuccessful, thus the structure could not be elucidated by NMR. As **94** and **75** share the same molecular formula C<sub>13</sub>H<sub>20</sub>O<sub>5</sub>, it was initially assumed the SpoI catalyses hydrolysis, which leads in the case of **75** to a ring opening of one of the lactones. However, the lack of **91** and **77** allows only one possible structure for **94** (See HR-ESI-MS and UV spectra in Chapter 7), where the exomethylene group was hydroxylated (Figure 2.54D). As **94** was never observed in the wild type, it is likely to be quickly degraded *in vivo*.

## 2.4 Discussion

### 2.4.1 Production and Labelling Experiment

Based on a combination of morphological and chemotaxonomic studies,<sup>52</sup> *Hypomontagnella* is regarded as a new genus segregated from *Hypoxylo*, with *Hypomontagnella monticulosa* as type species.<sup>37,52</sup> Sporothriolide-related compounds can presently only be obtained from the genus *Hypomontagnella* as a major component of their secondary metabolites.

The acetate labelling patterns in sporothriolide **1** are consistent with the hypothesis of fatty acid or polyketide origin. Carbons C-3, C-4, and C-13 are derived from a decarboxylated Krebs cycle intermediate, probably oxaloacetate (Table 2.3) hinting at the existence of a decarboxylase. Generally, citrate synthase is responsible for the condensation of polyketides or fatty acids with oxaloacetate to construct the alkyl citrates, such as during the biosynthesis of maleidrides byssochlamic acid **46**.<sup>21</sup> Carbon C-6, derived from C-2 of acetate, is oxygenated indicating the involvement of an oxygenase during biosynthesis (Table 2.3). The alkyl chain in sporothriolide is most probably assembled from fatty acid biosynthesis because it was not highly modified or methylated, similar to the case of piliformic acid **59**.<sup>27,28</sup> Moreover, sporothriolide **1** and deoxysporothric acid **76** have identical labelling patterns as well as similar structures,<sup>54</sup> which indicates a close relationship in their biosynthetic pathways meaning that they should have the same core early steps, and likely very similar BGCs.

### 2.4.2 Bioinformatic Analysis

Genome and RNA sequencing of the producing fungi were performed (Section 2.3.4). Based on manual searches using the citrate synthase *bfL2* (ANF07286.1) and 2-methylcitrate dehydratase *bfL3* (ANF07285.1) involved in byssochlamic acid **46** biosynthesis, a highly conserved alkyl citrate gene cluster (Figure 2.10 – 2.11) was rapidly identified which we regarded as the candidate BGC for sporothriolide **1**. Gene expression levels under producing and non-producing conditions were analysed from transcriptome data. Hence, the gene cluster boundary was defined, as well as the intron and exon positions (Table 7.1). There are 12 genes contained in the gene cluster, which are the expected citrate synthase (CS) *spoE*, and two fungal FAS subunits (FAS $\alpha$  and FAS $\beta$ ) *spofasA* and *spofasB*. In addition, the BGC encodes: a methylcitrate dehydratase (*spoL*); a decarboxylase (*spoK*); a dioxygenase (*spoG*); and two putative lactonases (*spoH* and *spoJ*). Moreover, the cluster encodes two transporters (*spoC* and *spoF*), along with a transcriptional regulator (*spoD*) and *spoI* with unknown function (Figure 2.14).

### 2.4.3 Bipartite Knockout

Gene knockout is a very useful tool to investigate the biosynthetic steps of natural products. Bipartite fragments targeting genes *spofasA/spoE* were transformed in *H. spongiphila* with a specifically developed protocol (Section 6.1.2.4). Finally, 46 transformants with hygromycin resistance were obtained after three rounds, however only one colony was proved to be a true knockout by checking its metabolites and genetic analysis (Figure 2.20 – 2.21). This was the first time to apply the genetic manipulation tools in *H. spongiphila*. However, utilizing this strategy to achieve knockout of other tailoring genes including *spoG*, *spoH*, and *spoK* was unsuccessful because of the low efficiency of K.O in *H. spongiphila* (Figure 2.23 – 2.24). In total, 119 transformants were obtained from attempted *spoG*, *spoH*, and *spoK* knockout but all proved to be false positive, resulting from ectopic integration.

### 2.4.4 Heterologous Expression and Protein *In Vitro* Assay

Heterologous expression of the *spo* BGC in *A. oryzae* NSAR1 and protein *in vitro* assay were employed to study the biosynthetic pathway of sporothriolide **1**. The results of EXP4 (expression of *spofasA* + *spofasB* + *spoE* + *spoL* + *spoK*; Figure 2.32) indicated a biosynthetic pathway (Scheme 2.2) in which the fungal FAS proteins produce a dedicated decanoic acid **84**, probably as a CoA thiolester. Fungal FAS possess a bifunctional malonyl-palmitoyl transferase (MPT) domain which loads malonyl CoA and off-loads an acyl CoA product.<sup>12</sup> Decanoyl CoA **84** is likely to be reacted with oxaloacetate by citrate synthase SpoE, in analogy to other known pathways,<sup>14, 21</sup> to give an octanyl citrate **85** which is then dehydrated to give octanyl aconitate **86**,

and decarboxylated to the first observable intermediate, octanyl itaconic acid **83**, a known compound found from the endophytic fungus *Pestalotiopsis theae*.<sup>54</sup>

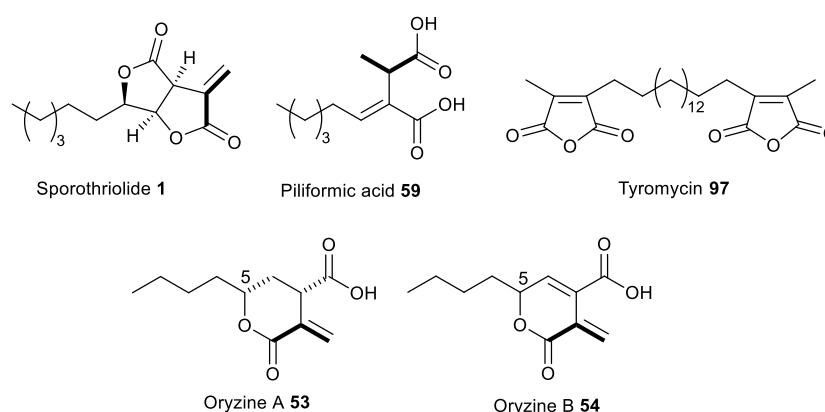
From EXP5 (expression of *spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG*; Figure 2.35), we proposed that intermediate **83** is the substrate of dioxygenase SpoG (Scheme 2.2). Due to the presence of a hydroxyl group at C-5, we assumed that **75** is derived from a double hydroxylated intermediate **89** (Table 2.10), which undergoes spontaneous lactonization and thus putatively represents a pathway intermediate. The main peak corresponding with the molecular formula C<sub>13</sub>H<sub>22</sub>O<sub>6</sub> (calcd. for C<sub>13</sub>H<sub>21</sub>O<sub>6</sub>, 273.1338, measured 273.1339) for **89** (see Chapter 7 for HR-ESI-MS spectrum) was observed in the crude extract, but could not be purified to support its identity as the proposed intermediate. Compound **76** is structurally related to **75**, but lacks the hydroxyl group at C-5. It is most likely a shunt metabolite derived from the putative monohydroxylated intermediate **87**. Compound **90** is also monohydroxylated at C-6, but instead of the lactone ring, it contains a maleic anhydride. Likewise, **90** is probably also derived from the putative intermediate **87** through a spontaneous reaction of the carboxylic acid moieties. Another monohydroxylated intermediate is **88**, hydroxylated at C-5. Isolation of **91** was not successful, and LCMS chromatograms showed that **91** was partially converted to isosporothric acid **77** in a short time (see Chapter 7). As **91** shares the same molecular formula and very similar retention time with **77**, its structure was deduced as the isomeric form of **77**, which itself is likely formed by spontaneous lactonization between C-5 and the carboxylic acid moiety derived from oxaloacetate of the putative intermediate **89**.

The results in EXP6 (expression of *spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH*; Figure 2.40) and EXP7 (expression of *spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoJ*; Figure 2.42) indicate that separate usage of either lactonase SpoH or SpoJ is insufficient to drive the pathway for sporothriolide formation. The results from EXP8 (expression of *spofasA* + *spofasB* + *spoE* + *spoL* + *spoK* + *spoG* + *spoH* + *spoJ*; Figure 2.44) showed that only when both lactonases are expressed together is the final pathway product **1** observed, which quickly converts into **92** when exposed to longer fermentation periods. Besides, small titers of **91** (Figure 2.44) were observed at early fermentation time points, which disappeared at the end of the fermentation period, indicating that **91** may also be a pathway intermediate that is converted into **1**. Either **75** or **91**, or both could be intermediates during the conversion of **89** to **1**. However, our *in vivo* experiments were unable to resolve this question. We propose that the lactonases SpoH and SpoJ form a bifunctional heterodimer to lactonize **89** into the final pathway product **1** as neither appears to catalyse the formation of **1** in isolation (Scheme 2.2).



spontaneous lactonization between C-6 hydroxyl functionality and the carboxyl group of the fatty acid derived chain seems to be favoured in this case. Also, results of *in vitro* assays with functional unknown SpoI showed it performs effective hydrolysis of sporothriolide **1**, indicating *spoI* serves as a potential self-resistance gene in *spo* BGC.

The sporothriolide pathway shows several points of novelty. First, a dedicated fungal FAS system produces decanoate rather than the usual octadecanoate. This differs from other known pathways such as those involved in maleidride<sup>21</sup> and squalastatin<sup>14</sup> biosynthesis which begin with a dedicated PKS (Scheme 1.3 – 1.4). FAS systems that selectively form short chains should find utility in the production of biofuels, for example, where C<sub>8</sub> and C<sub>10</sub> lipids are advantageous, and extensive efforts are underway to engineer short-chain synthases for this outcome.<sup>72</sup> Fungal FAS appear to have already evolved the ability to selectively produce these valuable shorter compounds, such as decanoate in sporothriolide **1** and octanoate in piliformic acid **59** and oryzines **53** – **54** (Figure 2.55), for example.



**Figure 2.55** Structures of lactone and maleic anhydride metabolites from fungi. Bold bonds show oxaloacetate-derived carbons where known.

The early steps (Scheme 2.2) of the *spo* pathway then follow the well-known primary metabolic steps from acetyl CoA to itaconic acid, but modified to accept a longer alkyl unit. The first four steps of the pathway thus closely mirror primary metabolic steps. The presence of the decarboxylase encoded by *spoK* differentiates the pathway from those proceeding towards the maleidrides where a different mode of decarboxylation appears to be coupled with multimerization catalysed by dedicated KI and PEBP enzymes (Scheme 1.3),<sup>21</sup> and from the squalastatins<sup>14</sup> where no dehydration or decarboxylation reactions take place, but where much more extensive backbone oxygenation leads to more complex cyclisation modes (Scheme 1.4).

The oxygenase SpoG catalyses the hydroxylations required to form the *bis*-lactone, and this moves the pathway more conclusively towards secondary metabolism. These hydroxylated intermediates are clearly able to undergo many spontaneous cyclizations, but

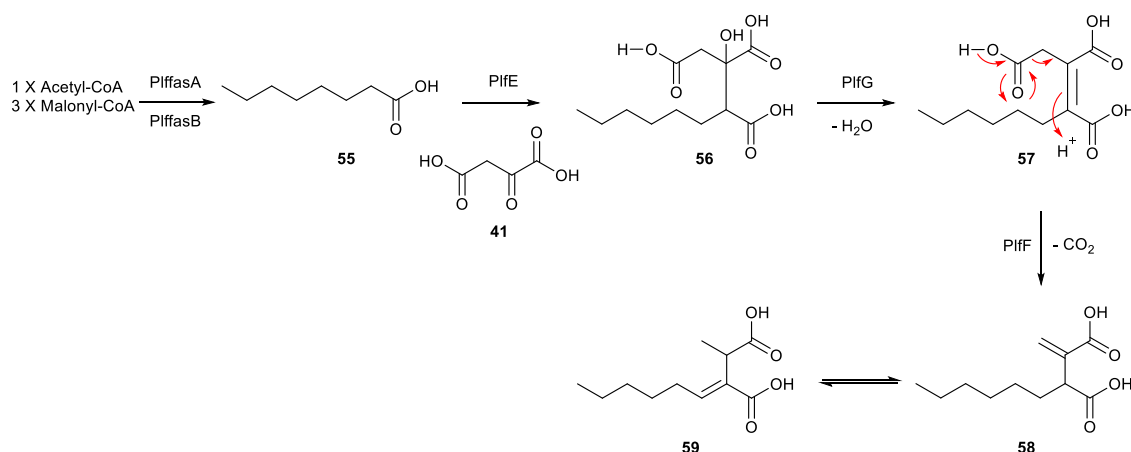
the presence of both lactonases SpoH and SpoJ directs the pathway to sporothriolide **1**. The presence and selectivity of the SpoG hydroxylase presumably controls the formation of related classes of metabolites in this family. For example in the case of the furofurandiones such as **1** double hydroxylation is required.

Attempts to express SpoH and SpoJ and study them *in vitro* in analogy to SpoG were not successful, leading to insoluble and inactive proteins in *E. coli* and non-viable colonies in *S. cerevisiae*. Either **75** or **91**, or both could be intermediates during the conversion of **89** to **1** (Scheme 2.2). However, our *in vivo* experiments were unable to resolve this question and lack of soluble protein obviated *in vitro* experiments. We propose that the lactonases SpoH and SpoJ form a bifunctional heterodimer to lactonize **89** into the final pathway product **1** as neither appears to catalyse the formation of **1** in isolation (Scheme 2.2). It also appears that the selectivity of the lactonase components (SpoHJ) may be able to control the formation of  $\gamma$ -lactones (*e.g.* **1**) vs  $\delta$ -lactones (*e.g.* **53** and **54**; Figure 2.55).

The identification of SpoI as a hydrolase involved in the putative self-resistance mechanism of the producer organism against **1** adds to the increasing knowledge about self-resistance genes in fungi<sup>73</sup> and will enable the identification of similar enzymes in other biosynthetic pathways. Knowledge of the overall pathway may also help to elucidate the BGC of more elusive compounds such as maleic anhydride compounds tyromycin<sup>74</sup> **97** (Figure 2.55) where no biosynthetic information currently exists.

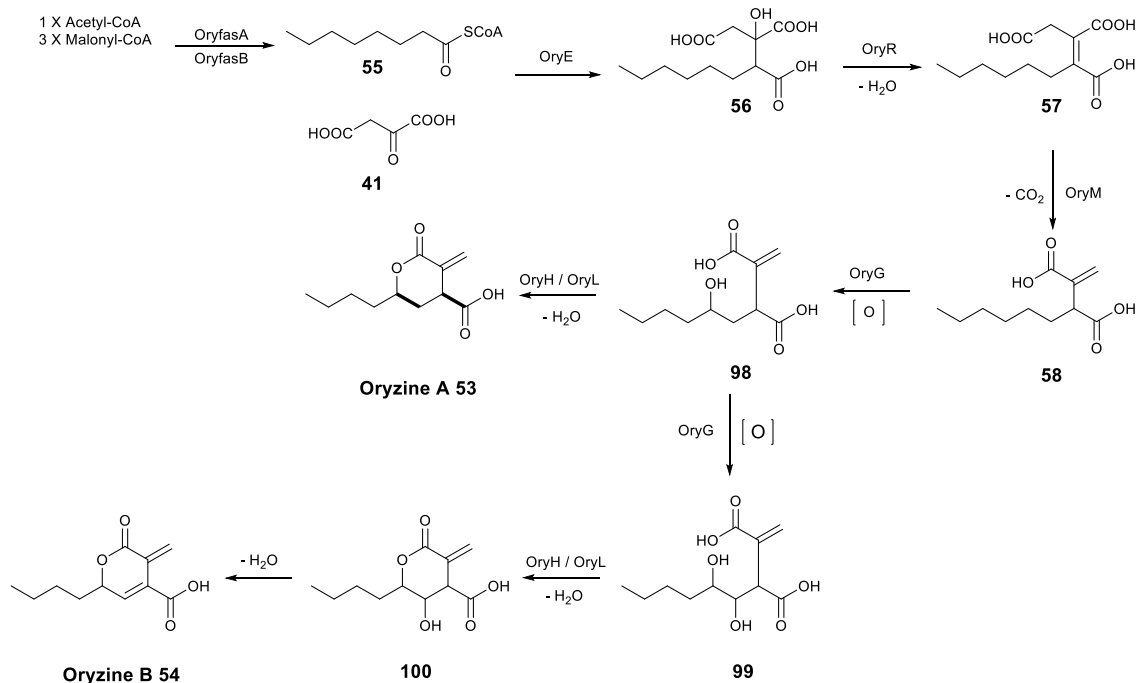
#### 2.4.5 Biosynthesis of Other Alkyl Citrates

Based on the delineated *spo* pathway, a hypothetical biosynthesis for piliformic acid **59** was suggested in Scheme 2.3. The *plf* BGC is the only alkyl citrate BGC present in the piliformic acid producer *X. hypoxylon*, so while no physical evidence links the BGC to piliformic acid **59**, it is the only likely possibility. The fungal FAS subunits PlffasA and PlffasB probably assemble the octanoate chain **55**, following the condensation with oxaloacetate **41** by CS (PlfE), then dehydration by PlfG and decarboxylation by PlfF are very likely to synthesize the alkene **58**. The last step is the isomerisation of **58** to **59**.



**Scheme 2.3** The proposed biosynthesis of piliformic acid **59**.

In addition, a biosynthetic pathway for oryzine could be speculated from the research of sporothriolide **1** (Scheme 2.4). The oryzine pathway shares identical early steps with piliformic acid **59**, but the presence and selectivity of the dioxygenase (OryG) most likely drives the pathway to the oryzine formation. Compounds such as oryzine A **53** (Scheme 2.4) require only a single hydroxylation to yield intermediate **98**, then lactonization (by OryH / OryL) to afford **53**. However, oryzine B **54** may arise from the double hydroxylated intermediate **99**, then underwent lactonization (by OryH / OryL) and dehydration to yield the final product.



**Scheme 2.4** The proposed biosynthesis of oryzine.

Finally, the similarities and differences of biosynthetic pathways of maleidrides (exemplified by byssochlamic acid) and alkyl citrates (exemplified by sporothriolide, oryzine and piliformic acid) could be concluded here. The early steps of sporothriolide **1** biosynthesis should be similar with



byssochlamic acid **46** and oryzines **53** – **54** (Scheme 2.2 – 2.4). Byssochlamic acid **46** requires a triketide from highly reducing polyketide synthase *bfpksI* (Scheme 1.3), but the octanoate for oryzines **53** – **54**, piliformic acid **59** and decanoate for sporothriolide **1** are derived from fatty acid synthases. For byssochlamic acid **46** biosynthesis, a hydrolase encoded by *bfL1* is required to release the polyketide from the PKS. While it is not necessary in the case of fungal FAS which release CoA thioesters directly, and no genes in the oryzines **53** – **54** and sporothriolide **1** clusters are homologous to *bfL1*.

The citrate synthase (SpoE) is proposed to catalyse the condensation of decanoate and oxaloacetate, then dehydration by the methylcitrate dehydratase (SpoL), which are similar to the early steps proposed during oryzines **53** – **54**, piliformic acid **59** and byssochlamic acid **46** biosynthesis (Scheme 2.5). For byssochlamic acid **46**, the subsequent reaction is the synthesis of maleic anhydride then spontaneous decarboxylation to give monomers **43**, **44**, and **45** for various dimerization.<sup>21</sup> However, sporothriolide **1** may lose a carboxyl functionality earlier, which could be catalysed by a decarboxylase (SpoK), to prevent anhydride formation like oryzines **53** – **54** (OryM in *ory* BGC; Figure 2.12) and piliformic acid **59** (PlfF in *plf* BGC; Figure 2.12). Compounds such as oryzine A **53** (Scheme 2.5) requires only a single hydroxylation at C-5 performed by OryG (68% identity, 77% similarity),<sup>23</sup> while oryzine B **54** (Scheme 2.5) may arise from double hydroxylation and 3,4-dehydration. Piliformic acid **59** requires no hydroxylation. Its 2,5 unsaturation most likely arises by isomerisation of **104**, for example, consistent with the lack of a *spoG* homolog in its BGC.



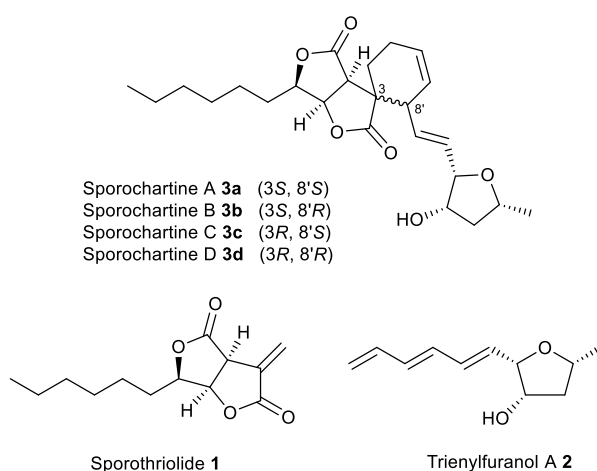
Bipartite knockout of *spoE* and *spofasA* simultaneously in *H. spongiphila* CLL 205 abolished the production of **1**, thus verified the correct BGC. Then, the sporothriolide biosynthetic pathway was reconstituted in *A. oryzae* NSAR1 by heterologous expression and several new intermediates were obtained. SpoG, a non-heme iron-dependent dioxygenase was studied *in vitro* and *in vivo*, it moves the pathway more conclusively towards secondary metabolism. Lactonases SpoH and SpoJ were proposed to form a bifunctional heterodimer to lactonize **89** into the final pathway product **1** as neither could separately catalyse the formation of **1**. In addition, a self-resistance gene *spoI* was proved to efficiently convert the antifungal sporothriolide **1** to **75** and **94**, protecting the producer organism from its toxic product when it accumulates in the cells. These results provide a full biosynthetic pathway for **1** production, and the order of enzymes encoded by the sporothriolide BGC is delineated.

In the future, sporothriolide **1** as the first fully elucidated biosynthetic example of alkyl citrate will shed light on the biosynthetic investigation of other alkyl citrates, such as piliformic acid **59** and oryzines **53** – **54**. More meaningfully, understanding the selectivity of the lactonases controlling the formation of  $\gamma$ -lactones (*e.g.* sporothriolide **1**) and  $\delta$ -lactones (*e.g.* oryzines **53** – **54**) would provide opportunities to engineer these biocatalysts to expand novel natural products' discovery.

## 3 Biosynthetic Studies of Sporochartine

### 3.1 Introduction

The sporochartines **3a** – **3d** (Figure 3.1) are a family of compounds with a spiro cyclohexane-furan scaffold which were first found in *H. monticulosa* CLL 205.<sup>53,75</sup> Sporochartine is proposed to be derived from a Diels-Alder cycloaddition of sporothriolide **1** and trienylfuranol A **2**.<sup>53,75</sup> Trienylfuranol A **2** (Figure 3.1), is a polyene polyketide, first reported by the Sumarah group<sup>56</sup> from an endophytic fungus *Hypoxylon submonticulosum* DAOMC 242471 (now referred to as *Hypomontagnella submonticulosa*).<sup>37,52</sup>



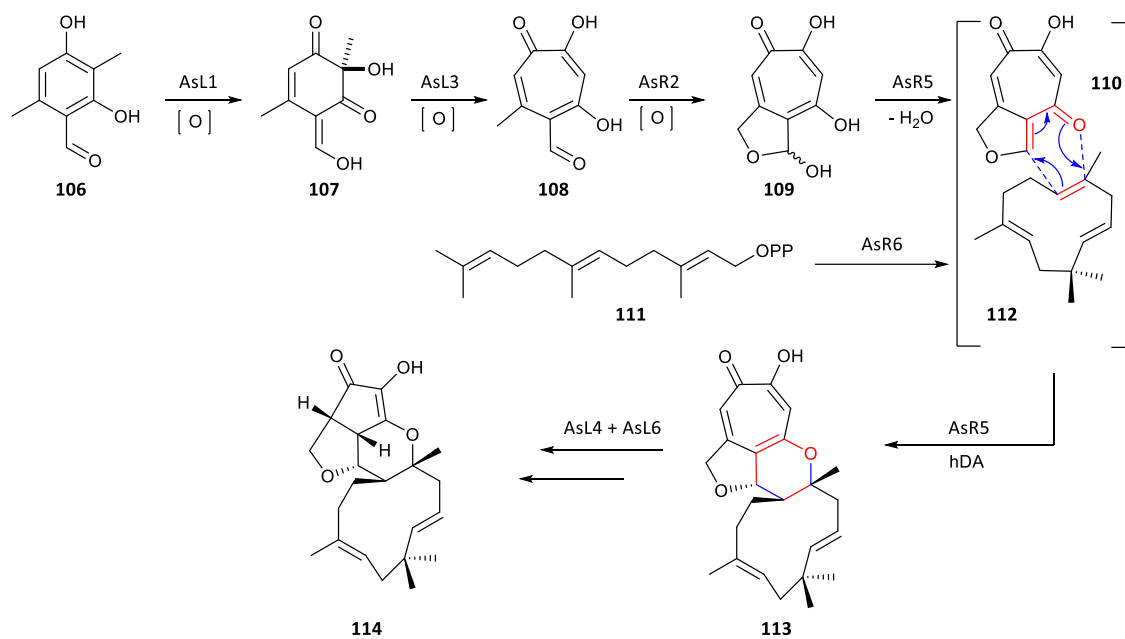
**Figure 3.1** Structures of sporochartines **3a** – **3d**, sporothriolide **1** and trienylfuranol A **2**.

#### 3.1.1 Examples of Diels-Alder [4+2] Cycloaddition in Natural Products Biosynthesis

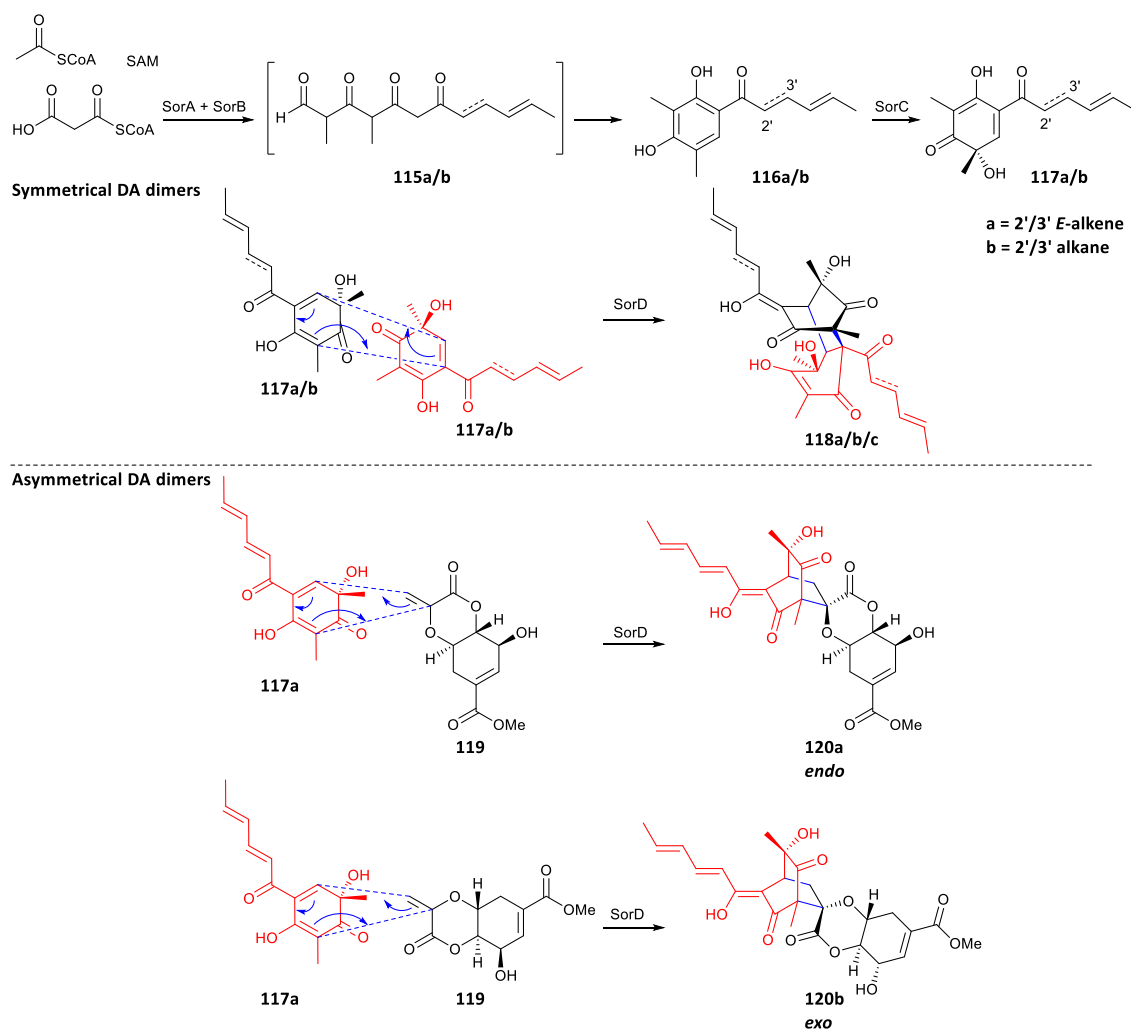
The Cox group has previously investigated Diels-Alder (DA) reactions involved in natural products biosynthesis, such as xenovulene A (Scheme 3.1),<sup>47</sup> sorbicillinoids<sup>48</sup> (Scheme 3.2) and cytochalasans.<sup>76</sup>

During xenovulene A **114** biosynthesis (Scheme 3.1), AsR5 catalyses the intermolecular hetero Diels-Alder [4+2] cycloaddition of tropolone **110** and humulene **112** to form **113**, and then AsL4 and AsL6 function as oxidative-ring-contracting enzymes to achieve the final xenovulene A **114**. AsR5 appears to be a multifunctional enzyme that works in the production of the required enone **110** by the elimination of water of **109**.<sup>47</sup>

### 3 Biosynthetic Studies of Sporochartine



**Scheme 3.1** The proposed biosynthetic pathway to xenovulene A **114**.<sup>47</sup>



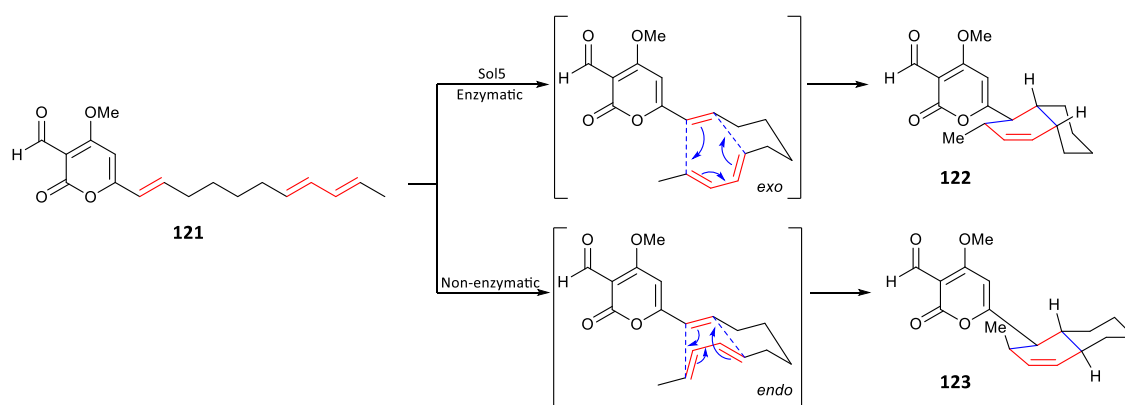
**Scheme 3.2** Biosynthesis of the sorbicillinoid in *Trichoderma reesei* QM6a.<sup>48</sup>

Another example is the FAD-dependent monooxygenase (FMO) SorD which acts during sorbicillinoid biosynthesis (Scheme 3.2). The multifunctional SorD can execute different types of dimerization reactions including ‘symmetrical’ Diels-Alder reaction of two moieties **117a/b** to give **118a/b/c**, and the ‘asymmetrical’ Diels-Alder reaction between **117a** and scytolide **119** to form **120a/b**. In addition, the studies showed that SorD plays a role in the ‘symmetrical’ Michael-addition dimerization to make other complicated scaffolds.<sup>48</sup>

### 3.1.2 Enzymatic and Non-enzymatic Diels-Alder [4+2] Reactions

Although there are over 400 natural products proposed to use the [4+2] DA cycloaddition reactions during their biosynthetic pathways, the biologically characterised Diels-Alderases (DAases) are few until now.<sup>77–80</sup> On the one hand, the identification of DAases by bioinformatic is difficult because of the poor sequence homology and the absence of specific catalytic motifs or cofactors. Most DA cycloadditions are capable of performing at sufficient rates spontaneously without enzyme catalysis, but non-enzymatic DA products are often produced with low stereoselectivity.

For example, the intramolecular DA reactions operating during solanapyrone biosynthesis can be either enzymatic or non-enzymatic (Scheme 3.3). The product ratio of **122** (solanapyrone A; *exo*) : **123** (solanapyrone D; *endo*) resulting from the Sol5 (flavin-dependent oxidase) catalysis is 87 : 13. However, under non-enzymatic conditions, the reaction gives the *exo* **122** and the *endo* **123** products in a 3 : 97 ratio. The result showed that Sol5 processes in an *exo*-selective manner in the conversion of **121** to solanapyrones **122** and **123**.<sup>81,82</sup> In addition, the reaction rates of DA [4+2] cycloadditions with the Sol5 enzyme is 4.1 times as fast as the non-enzymatic.



**Scheme 3.3** Diels-Alder [4+2] reactions in solanapyrones biosynthesis.<sup>81,82</sup>

## 3.2 Project Aims

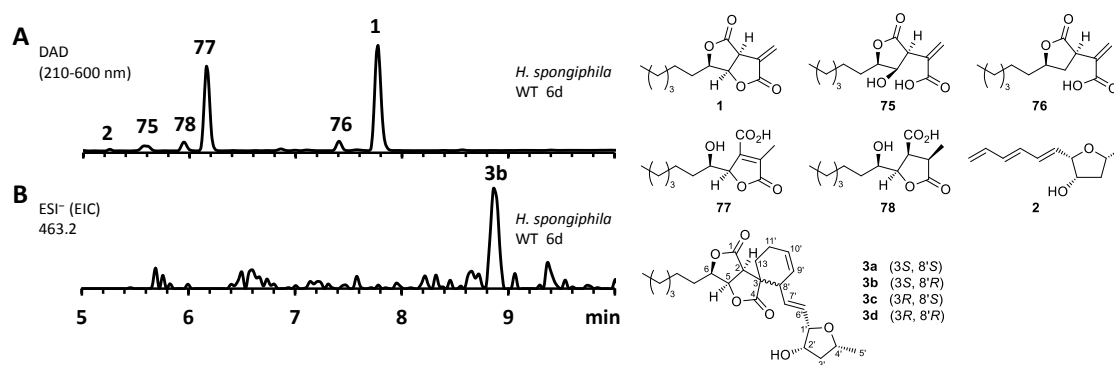
This project aims to answer whether sporochartine is formed through DA [4+2] cycloaddition of spirothriolide **1** and trienylfuranol A **2**, and, if so, whether the DA reactions are enzymatic or spontaneous.

## 3.3 Results

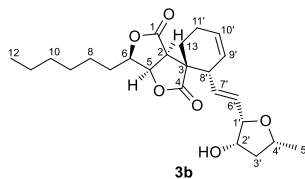
### 3.3.1 Sporochartine Production

In our work, sporochartine is identified from extracts of all three fungi compared with the retention time and mass fragmentation of standard compounds which were supplied by the Ouazzani group.<sup>53,75</sup> However, the production of sporochartine is poorly reproducible and in low titre, and only sporochartine B **3b** can be observed in extracted ion chromatograms (Figure 3.2). Multiple liquid media and solid agar (PDB/PDA, LB, DPY, YMG, MMK2, MOF), as well as various fermentation conditions, were tested for sporochartine production and optimisation but no efficient conditions were found.

Through large-scale fermentation (1 L) of *H. spongiphila* CLL 205 in PDB medium (130 rpm, 28 °C, 6 days), we isolated 1.2 mg sporochartine B **3b** from the extracts (Figure 3.3) and then performed 1D and 2D NMR experiments to verify the structure of **3b** (see Chapter 7 for spectra details). The NMR data (Table 3.1) was also compared with the literature.<sup>53,75</sup>

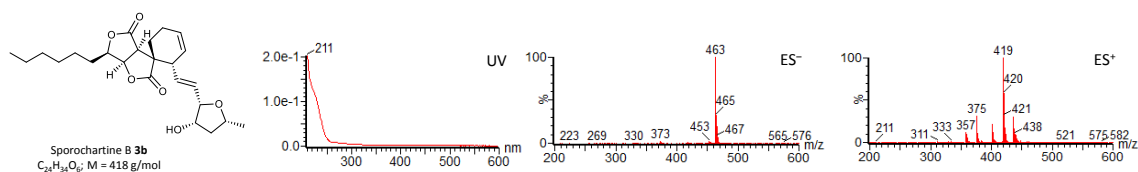


**Figure 3.2** A, Diode Array Detector (DAD) chromatogram of *H. spongiphila* CLL 205 WT grown under producing conditions of **1**; B, extracted ion chromatogram (EIC of **3b**, ESI<sup>-</sup>, 463.2, M + HCOO<sup>-</sup>) of WT extract.



pos.	$\delta_C$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_C$ / ppm Literature <sup>53, 75</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>53, 75</sup>
1	173.0	-	173.1	-
2	47.2	3.31, 1H, d (5.9)	47.2	3.30, 1H, d (5.8)
3	50.8	-	51.0	-
4	178.7	-	178.7	-
5	78.7	5.14, 1H, dd (4.1, 6.0)	78.7	5.13, 1H, dd (4.3, 5.9)
6	81.1	4.40, 1H, m	81.7	4.39, 1H, m
7	28.9	1.76, 1H, m 1.85, 1H, m	28.9	1.76, 1H, m 1.85, 1H, m
8	25.3	1.45, 2H, m	25.3	1.45, 2H, m
9	29.1	1.34, 2H, m	29.1	1.34, 2H, m
10	31.7	1.28, 2H, m	31.7	1.29, 2H, m
11	22.8	1.29, 2H, m	22.9	1.29, 2H, m
12	14.2	0.88, 3H, m	14.3	0.88, 3H, t (6.9)
13	26.9	2.03, 1H, m 2.12, 1H, m	26.9	2.04, 1H, m 2.13, 1H, m
1'	82.8	4.19, 1H, m	82.8	4.19, 1H, m
2'	73.8	4.28, 1H, m	73.9	4.27, 1H, m
3'	42.4	1.59, 1H, m 2.40, 1H, m	42.4	1.59, 1H, m 2.39, 1H, m
4'	74.1	4.06, 1H, m	74.2	4.07, 1H, m
5'	22.4	1.33, 3H, d (6.2)	22.5	1.34, 3H, d (6.1)
6'	130.1	5.75, 1H, dd (4.0, 15.5)	130.1	5.76, 1H, dd (3.9, 15.4)
7'	130.8	5.84, 1H, ddd (1.6, 9.0, 15.5)	130.9	5.82, 1H, ddd (1.5, 8.8, 15.4)
8'	46.9	3.23, 1H, m	47.0	3.23, 1H, brm
9'	124.7	5.54, 1H, m	124.8	5.54, 1H, brd (10.9)
10'	129.9	5.95, 1H, m	130.0	5.95, 1H, brd (10.9)
11'	22.6	2.25, 1H, m 2.80, 1H, m	22.7	2.25, 1H, m 2.79, 1H, m

**Table 3.1**  $^1\text{H}$  NMR (500 MHz) data and  $^{13}\text{C}$  NMR (125 MHz) data for sporochartine **B 3b** in  $\text{CDCl}_3$ . Literature<sup>53,75</sup> data was measured at 500 MHz in  $\text{CDCl}_3$ .

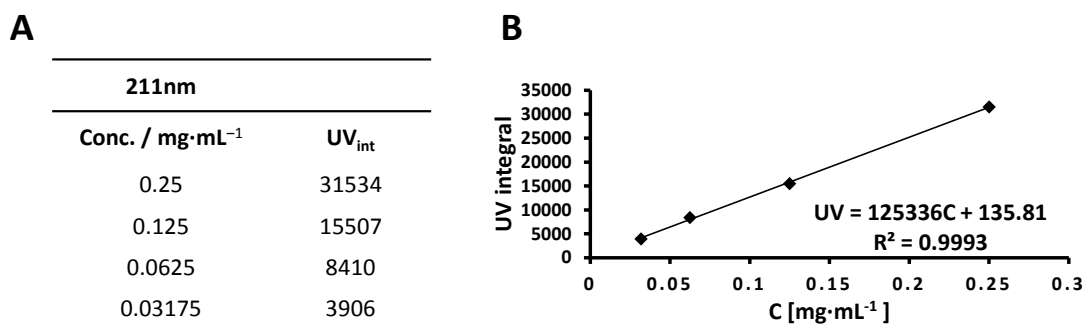


**Figure 3.3** UV and mass spectra ( $\text{ES}^+$  and  $\text{ES}^-$ ) of sporochartine **B 3b**.

### 3.3.2 Time Course Study of Sporochartine Production

To quantify the production of compounds, the calibration curve (Figure 3.4) for sporochartine **B 3b** was made based on a UV integration method (Section 2.3.1.2). Timecourse experiments for the three *Hypomontagnella* wild type strains were conducted to monitor the production of **3b**, the procedure details are listed in Section 2.3.1.2. Compound titres were calculated using the equation  $C = (\text{UV}_{\text{int}} - 135.81)/125336$  (Table 3.2). Production kinetics of **3b** were visualized graphically (Figure 3.5).

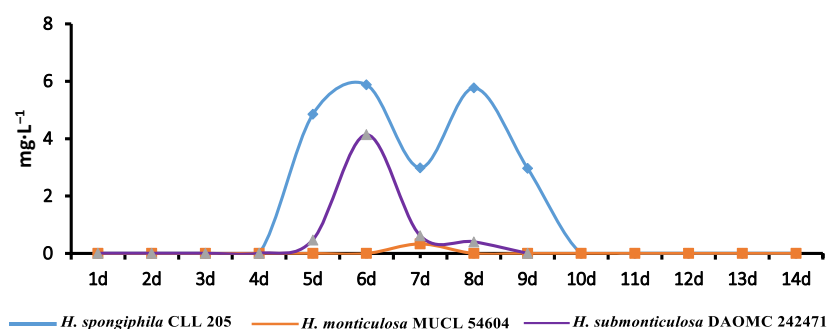




**Figure 3.4** A, the integrated values of different dilutions of sporochartine B **3b** solutions at 211 nm from LCMS analysis; B, calibration curve for sporochartine B **3b** quantification.

	<i>H. monticulosa</i>	<i>H. spongiphila</i>	<i>H. submonticulosa</i>
Day	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	4.8	0.5
6	0	5.9	4.1
7	0.3	3.0	0.6
8	0	5.8	0.4
9	0	3.0	0
10	0	0	-
11	0	0	-
12	0	0	-
13	0	0	-
14	0	0	-

**Table 3.2** Temporal production of sporochartine B **3b** from *H. monticulosa* MUCL 54604, *H. spongiphila* CLL 205, and *H. submonticulosa* DAOMC 242471 cultivated under producing conditions.



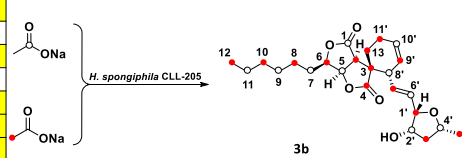
**Figure 3.5** Kinetic production curve of sporochartine B **3b**.

### 3.3.3 Acetate Feeding Experiment

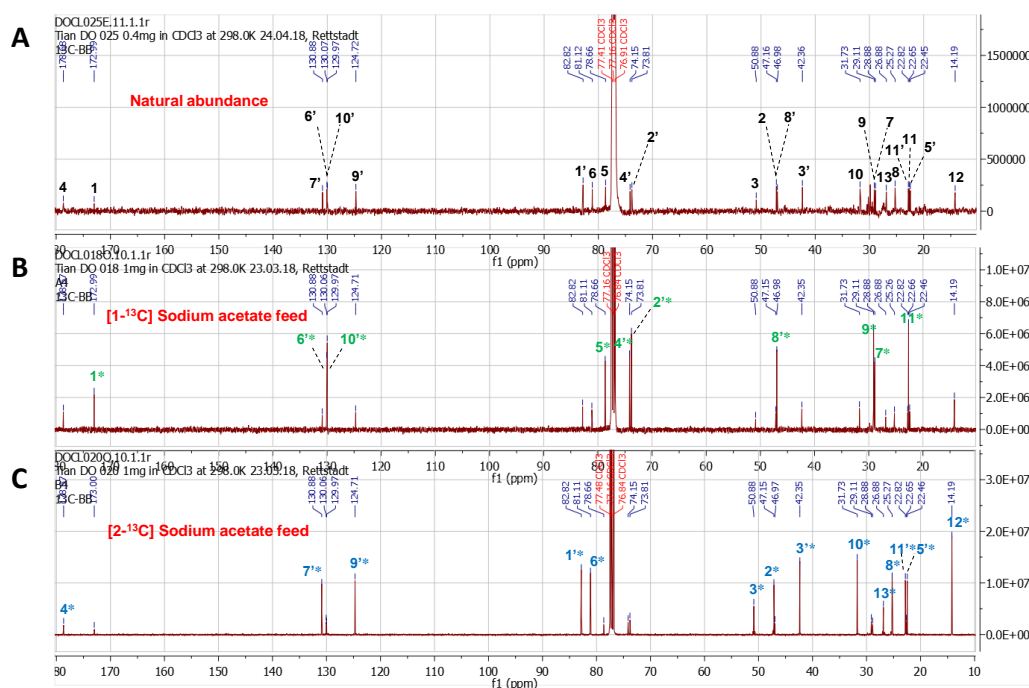
Sporochartine B **3b** is hypothesised to be constructed through the Diels-Alder cycloaddition of sporothriolide **1** and trienylfuranol A **2**.<sup>53,75</sup> Trienylfuranol A **2** is proposed to be of polyketide origin, thus **3b** should be constructed from the same acetate building blocks as **1** and **2**.

To study the labelling pattern of **3b**, [1-<sup>13</sup>C] and [2-<sup>13</sup>C] labelled sodium acetates were utilized in separate feeding experiments (Section 2.3.3). <sup>13</sup>C NMR spectra were recorded for labelled sporochartine B **3b**. Peak enhancement was estimated by calculating the ratio between the normalised peak intensity of each signal of the labelled compound and the normalised signal intensity of each carbon at natural abundance. C-12 and C-11 were used as references for [1-<sup>13</sup>C] and [2-<sup>13</sup>C] sodium acetate feeding experiments, respectively (Figure 3.6, Table 3.3).

Position	1- <sup>13</sup> C-Sodium Acetate Incorporation / Fold	2- <sup>13</sup> C-Sodium Acetate Incorporation / Fold
1	2.9	1.2
2	0.4	3.5
3	0.6	4.5
4	1.3	2
5	1.7	0.7
6	0.6	5
7	2	0.8
8	0.4	4.6
9	3	1.1
10	0.6	5.9
11 (reference)	2.9	1
12 (reference)	1	9.5
13	0.4	2.5
1'	0.5	4.3
2'	2.8	1.2
3'	0.5	5.4
4'	2.3	1.1
5'	0.5	4.6
6'	2	1
7'	0.5	4.7
8'	2.5	1
9'	0.5	4.8
10'	2.6	0.8
11'	0.4	4.3



**Table 3.3** NMR signal peak enhancement of labelled sporochartine B **3b**. A significant incorporation threshold fold is set  $\geq 1.5$  fold.

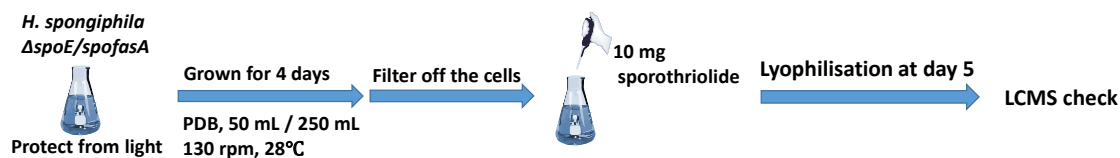


**Figure 3.6** <sup>13</sup>C NMR spectra of sporochartine B **3b**. **A**, Natural abundance; **B**, [1-<sup>13</sup>C] sodium acetate feeding experiment; **C**, [2-<sup>13</sup>C] sodium acetate feeding experiment. The enhanced peaks are marked with an asterisk.

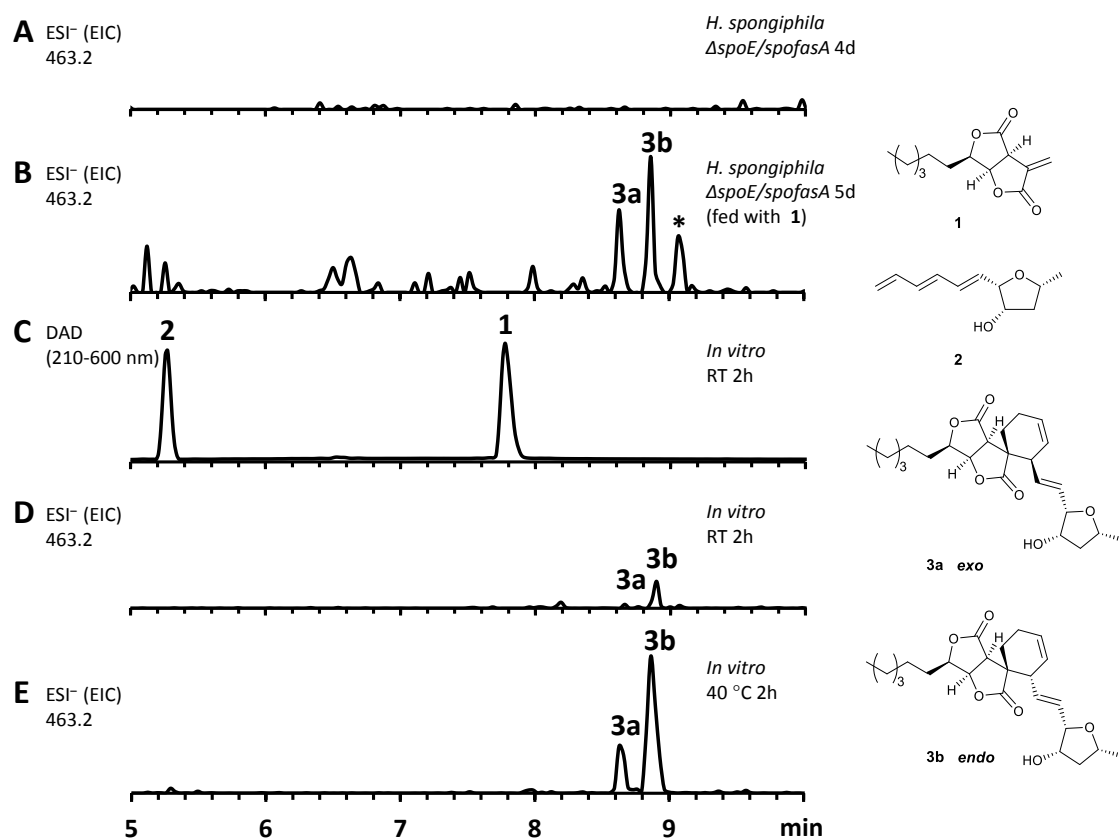
### 3.3.4 *In Vitro* Spontaneous Diels-Alder Cycloaddition for Sporochartine

From our previous bioinformatic analysis of the *spo* BGC, no gene encoding a likely DAase is present in the BGC. Thus, we speculated the Diels-Alder reaction of **1** and **2** to form sporochartine is probably non-enzymatic. In addition, sporochartines **3a** – **3d** occur as a mixture of various stereoisomers which fit with the non-enzymatic Diels-Alder reaction outcomes.<sup>75</sup> Therefore, various *in vitro* conditions were investigated to mimic the putative reaction between **1** and **2**.

In the *H. spongiphila* *spofasA/spoE* KO mutant, the titre of trienylfuranol A **2** is slightly increased (e.g. 40 mg·L<sup>-1</sup> at day 6; Figure 2.20, Table 4.1). To test if sporochartine can be formed spontaneously under fermentation conditions in the medium (Figure 3.7), the *H. spongiphila*  $\Delta$ *spoE/spofasA* KO mutant was grown under producing conditions for four days to accumulate **2**. The cells were then filtered off and the supernatant was supplemented with 10 mg of **1**. Control was conducted in parallel, in which the filtered supernatant was not supplied with **1**. Afterwards, the supernatant was incubated under fermentation conditions for 24 h before being lyophilized. The lyophilisates were partially dissolved in methanol and directly subjected to LCMS analysis. Extracted ion chromatogram searches with  $m/z$  463.2 [M + HCOO<sup>-</sup>]<sup>-</sup> showed that sporochartine A **3a** and sporochartine B **3b** were detected when **1** was fed to the medium (Figure 3.8), but lacked in the control, consistent with the hypothesis that **1** can spontaneously react with **2** in the medium.



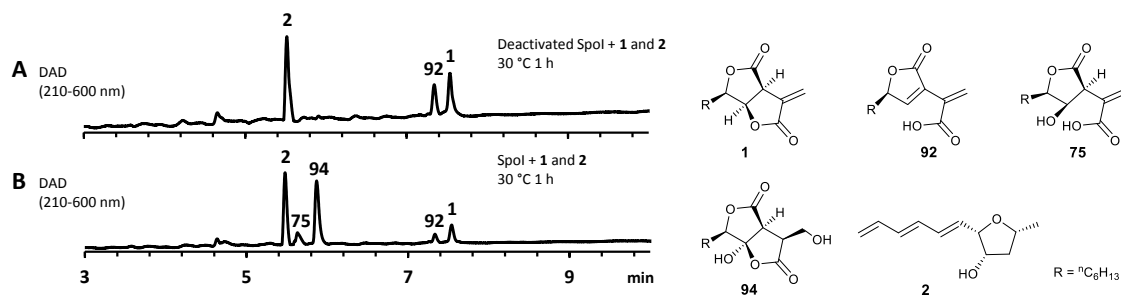
**Figure 3.7** The mimic reaction between sporothriolide **1** and trienylfuranol A **2** under *in vitro* fermentation conditions.



**Figure 3.8** *In vitro* assay for spontaneous sporochartines (**3a** and **3b**) production. **A**, Extracted ion chromatogram (EIC of **3a/3b**, ESI<sup>-</sup>, 463.2, M + HCOO<sup>-</sup>) of crude extract from *H. spongiphila* CLL 205  $\Delta$ *spoE/spofasA*; **B**, EIC of **3a/3b** from *H. spongiphila* CLL 205  $\Delta$ *spoE/spofasA* fed with **1**; **C,D** UV/vis spectrum and EIC (**3a/3b**) of *in vitro* reaction between **2** and **1** in ethyl acetate at RT, 2h; **E**, EIC (**3a/3b**) of *in vitro* reaction between **1** and **2** in ethyl acetate at 40 °C, 2 h. \*Unrelated peak.

We then tested if the standard extraction conditions (ethyl acetate and evaporation of the organic phase at 40 °C) can influence the formation of sporochartine. Sporothriolide **1** (1.5 mg) and trienylfuranol A **2** (1 mg) were dissolved together in 1 mL ethyl acetate and incubated under nitrogen in the dark at either room temperature or 40 °C. LCMS was used to monitor the *in vitro* reaction. After 2 hours at room temperature, sporochartine A **3a** (minor) and sporochartine B **3b** (major) were observed by LCMS (Figure 3.8). The reaction was continued overnight but did not result in increased concentrations of **3a/3b**. The reaction was then heated to 40 °C for two hours, under these conditions the reaction accelerated to give the same compounds in higher titre (Figure 3.8).

Although the previous *in vitro* experiments of SpoI showed it serves as a self-resistance gene in *spo* BGC (Figure 2.54), we also assayed SpoI *in vitro* to test whether it can be a possible DAase. Therefore, we mixed the purified SpoI protein, sporothriolide **1** and trienylfuranol A **2** in PBS buffer. However, the result (Figure 3.9) was consistent with our previous finding (Figure 2.54), that SpoI efficiently hydrolyzes sporothriolide **1** and doesn't accept the substrate trienylfuranol A **2**.



**Figure 3.9** *In vitro* assay of SpoI using trienylfuranol A **2** and sporothriolide **1**: **A**, DAD chromatogram of deactivated SpoI (50  $\mu$ M) incubated with **1** (2.5 mM) and **2** (2.5 mM) in PBS buffer (pH 7.5) at 30  $^{\circ}$ C for 1 h; **B**, DAD chromatogram of SpoI (50  $\mu$ M) incubated with **1** (2.5 mM) and **2** (2.5 mM) in PBS buffer (pH 7.5) at 30  $^{\circ}$ C for 1 h.

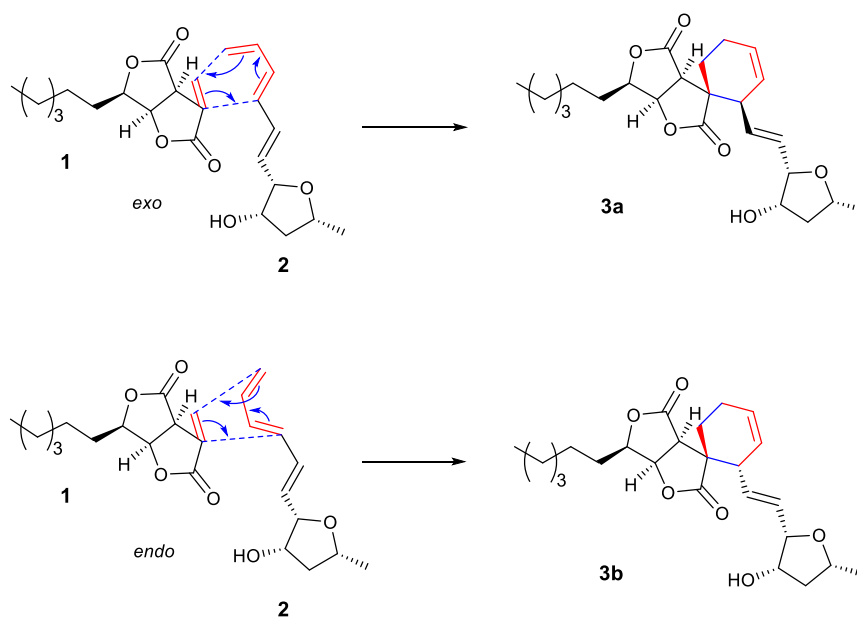
### 3.4 Discussion

Repeating the literature<sup>53,75</sup> production of sporochartines **3a** – **3d** in *H. spongiphila*, we only obtained a small amount of sporochartine B **3b**, verified by NMR (Chapter 7). The time course study showed that *H. spongiphila* and *H. submonticulosa* can produce *ca* 4 – 6 mg sporochartine B **3b** in 1 L fermentation media (Table 3.2). But the **3b** titre in *H. monticulosa* is pretty low, less than 0.5 mg·L<sup>-1</sup>. Thus, the Stadler group's previous research<sup>51</sup> that failed to detect sporochartine metabolites from *H. monticulosa* does make sense.

The acetate labelling patterns of sporochartine B **3b** (Table 3.3) are consistent with the incorporation of sporothriolide **1** (Table 2.3). In addition, the labelling patterns at the right part of **3b** fit with the hypothesis of polyketide origin of trienylfuranol A **2**. The labelling experiment results indicate that sporochartine formation is a consequence of **1** and **2** cycloaddition.

Sporochartines **3a** – **3b** are formed as mixtures of diastereomers<sup>75</sup> at the cyclohexene moiety and our *in vitro* experiment (Figure 3.8) showed that the *endo* DA product **3b** always occurs in higher amounts than *exo* adduct **3a**. This can be explained by the preference of DA reactions for *endo* product formation over *exo* (Scheme 3.4). In addition, bioinformatic analysis of the *spo* BGC showed no gene encoding a DAase likely to be involved in the DA cycloaddition of **1** and **2**. Hence, it can be concluded that the observed sporochartine is probably formed during the organism fermentation process through non-enzymatic DA cycloaddition of sporothriolide **1** and trienylfuranol A **2**, and that this effect is enhanced through the extraction procedure (Figure 3.8).

However, besides the discovery of enzymatic (SorD, Section 3.1.1) DA reactions during the sorbicillinoid biosynthesis,<sup>48</sup> the non-enzymatic DA reactions are also observed.<sup>83</sup>



Scheme 3.4 Diels-Alder [4+2] reactions to form sporochartine, adapted from Ouazzani *et al.*, 2017.<sup>75</sup>

### 3.5 Conclusion and Prospect

Results of acetate labelling experiments (Table 3.3) indicate that sporochartine originates from the DA [4+2] cycloaddition of sporothriolide **1** and trienylfuranol A **2**. Additionally, our further experiments (Figure 3.8) showed that unlike in many other fungal natural products,<sup>47,48,76</sup> the DA reaction required to form the sporochartine A **3a** and sporochartine B **3b** is non-enzymatic. Furthermore, no genes encoding proteins likely to be involved in the biosynthesis of trienylfuranol A **2**, such as a PKS, or other oxidative proteins, were found within or near the *spo* gene cluster. Therefore, the biosynthesis of **2** remains cryptic for now, but experiments in this area will form part of our future investigations.

## 4. Biosynthetic Studies of Trienylfuranol A

### 4.1 Introduction

As the previous biosynthetic studies of sporochartine **3** (Figure 4.1) showed in Chapter 3, sporochartine **3** is formed through the DA cycloaddition of trienylfuranol A **2** and sporothriolide **1**. The investigation of sporothriolide biosynthesis was described in Chapter 2. However, no biosynthetic information about trienylfuranol A **2** was known before our study except the report of compound isolation and characterization.

Trienylfuranol A **2** and polyene analogues trienylfuranone A **124** and trienylfuranone B **125** (Figure 4.1) were found by the Sumarah<sup>56</sup> group from the plant endophytic fungus *Hypoxyton submonticulosum* (now referred to as *Hypomontagnella submonticulosa*).<sup>37,52</sup> There exist some structurally similar fungal compounds such as depudecin **126** and aureonitol **127**. In addition, polyenes are known precursors of the bacterial enediynes for which biosynthetic information is well known.

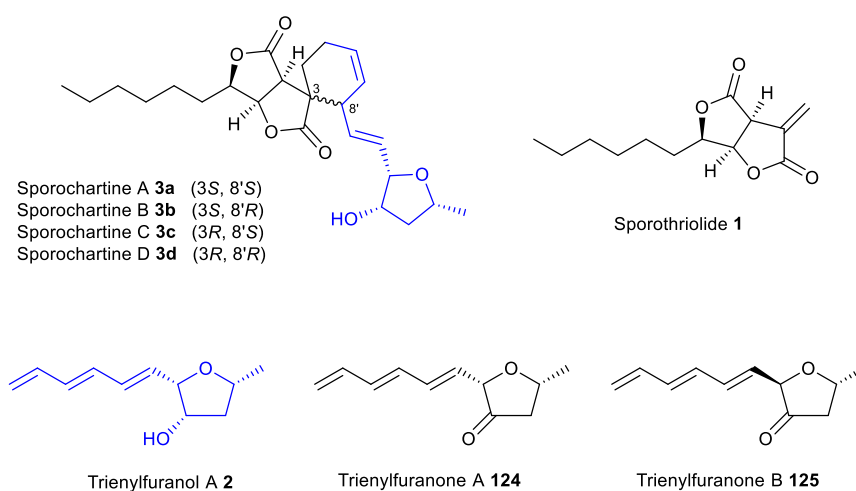


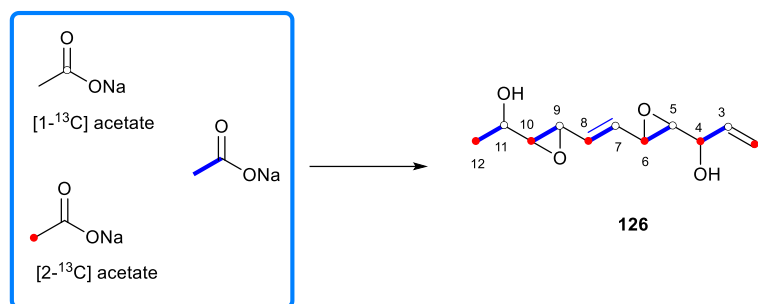
Figure 4.1 Structures of trienylfuranol A **2** and trienylfuranone A **124** and trienylfuranone B **125**.

#### 4.1.1 Depudecin

Depudecin **126** was first identified from the pathogenic fungus *Alternaria brassicicola* as an eleven carbon linear polyketide.<sup>84</sup> Isotopic labelling experiments (Scheme 4.1) indicated the backbone is built from six acetates, with a decarboxylation at C-2.<sup>85</sup> The acetate C-2 originated C-4, C-6 and C-10 positions in depudecin **126** are linked with oxygen, meaning oxygenases must be involved during biosynthesis.

The gene cluster (Figure 4.2) for depudecin **126** biosynthesis was confirmed by employing a gene KO strategy.<sup>86</sup> The depudecin BGC encodes a PKS (*DEP5*), two FAD-dependent

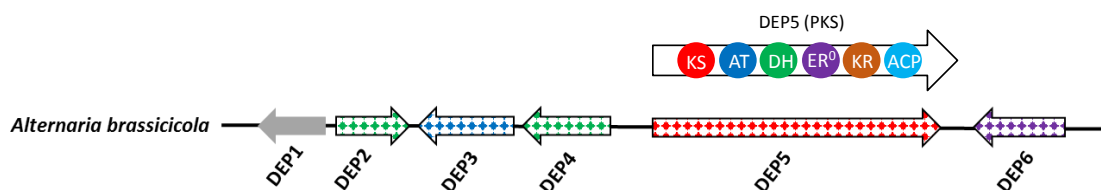
monooxygenases (FMO; *DEP2* and *DEP4*), a transporter (*DEP3*), a transcriptional regulator (*DEP6*) and a protein of unknown function (*DEP1*). Although KO experiments showed that the *DEP2*, *DEP4*, *DEP5* and *DEP6* genes are essential for depudecin formation, no intermediates were obtained so that the pathway is still unknown. A good explanation may be the lability of these polyene compounds that leads to degradation.



**Scheme 4.1** The labelling pattern of depudecin **126** by using [1-<sup>13</sup>C], [2-<sup>13</sup>C] and [1,2-<sup>13</sup>C<sub>2</sub>] acetate, adapted from Tanaka *et al.*, 2000.<sup>85</sup>

Bioinformatic analysis of the PKS (*DEP5*) identified KS, AT, DH, ER<sup>0</sup>, KR and ACP domains (Figure 4.2). The domain architecture of *DEP5* fits with the normal fungal hrPKS,<sup>11,87</sup> and the ER domain is non-functional which corresponds to the polyene polyketide metabolite. In other cases where an ER<sup>0</sup> domain is present, the BGC often encodes *trans*-acting ER proteins such as LovC<sup>17,18</sup> and TenC<sup>88</sup> involved during lovastatin and tenellin biosynthesis, respectively. However, the depudecin BGC does not encode a *trans*-acting ER consistent with the intermediacy of a polyene.

Trienylfuranol A **2** and depudecin **126** are both C<sub>11</sub> chains that contain several double bonds oxygenated carbon atoms. Therefore, their biosynthesis may share similarities to some extent.



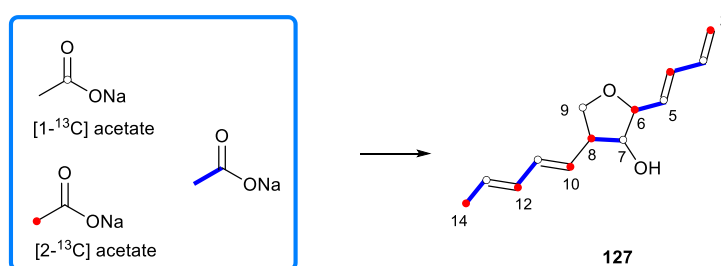
Gene	Amino acids	Proposed protein function
DEP1	580	Unknown
DEP2	646	FAD-dependent monooxygenase
DEP3	699	Major facilitator superfamily transporter
DEP4	708	FAD-dependent monooxygenase
DEP5	2451	Polyketide synthetase
DEP6	695	Transcription factor

**Figure 4.2** Depudecin **126** biosynthetic gene cluster in *Alternaria brassicicola*, adapted from Wight *et al.*, 2009.<sup>86</sup> ER<sup>0</sup>, non-functional ER domain.



### 4.1.2 Aureonitol

Aureonitol **127** was first isolated and elucidated from *Chaetomium coarctatum*.<sup>89</sup> Extensive isotopic labelling experiments showed the polyketide chain is built from seven acetates (Scheme 4.2), the loss of one carbon at C-2 is consistent with a decarboxylation. In addition, the bond break of the intact acetate at C-9/C-10 indicates that there is a carbon-carbon rearrangement during biosynthesis. Moreover, the oxygenated C-6 and C-8 carbons which are derived from the C-2 of acetate, suggest the involvement of at least two oxygenation reactions during the biosynthesis.<sup>90</sup>



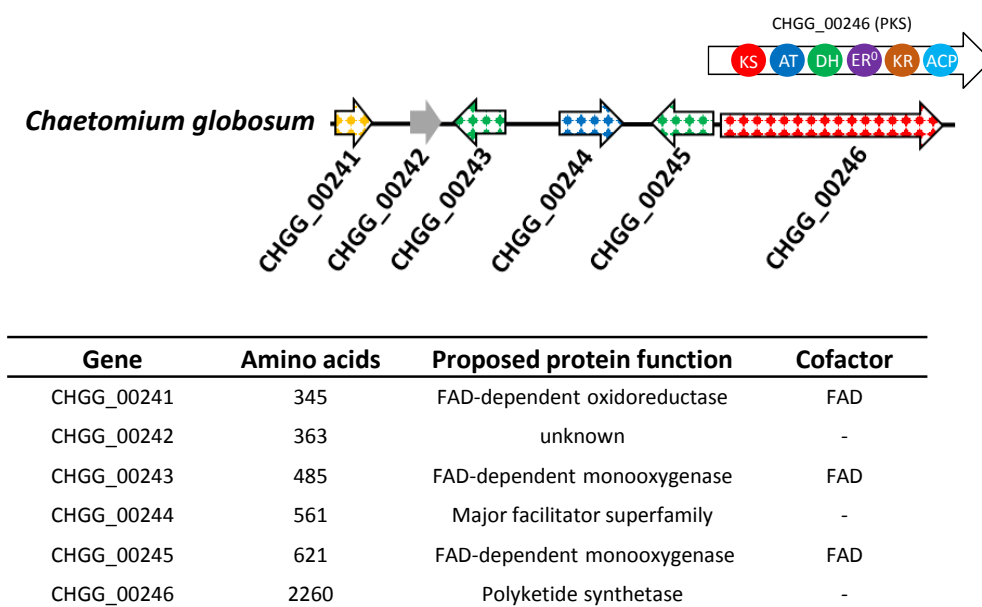
**Scheme 4.2** The labelling pattern of aureonitol **127** by using [1-<sup>13</sup>C], [2-<sup>13</sup>C] and [1,2-<sup>13</sup>C<sub>2</sub>] acetate, adapted from Seto *et al.*, 1979.<sup>90</sup>

Through genome sequencing and gene cluster mining of the producer *Chaetomium globosum*, a hrPKS gene cluster was identified (Figure 4.3).<sup>91</sup> In addition to the PKS (*CHGG\_00246*), the gene cluster encodes two FMO (*CHGG\_00243* and *CHGG\_00245*), a P450 monooxygenase (*CHGG\_00240*), a dehydrogenase (*CHGG\_00239*), an oxidoreductase (*CHGG\_00241*) and a transporter (*CHGG\_00244*). Later, gene KO experiments were performed to confirm this gene cluster and supported the essential role of *CHGG\_00241*, *CHGG\_00243* and *CHGG\_00245* during aureonitol **127** biosynthesis.<sup>91</sup>

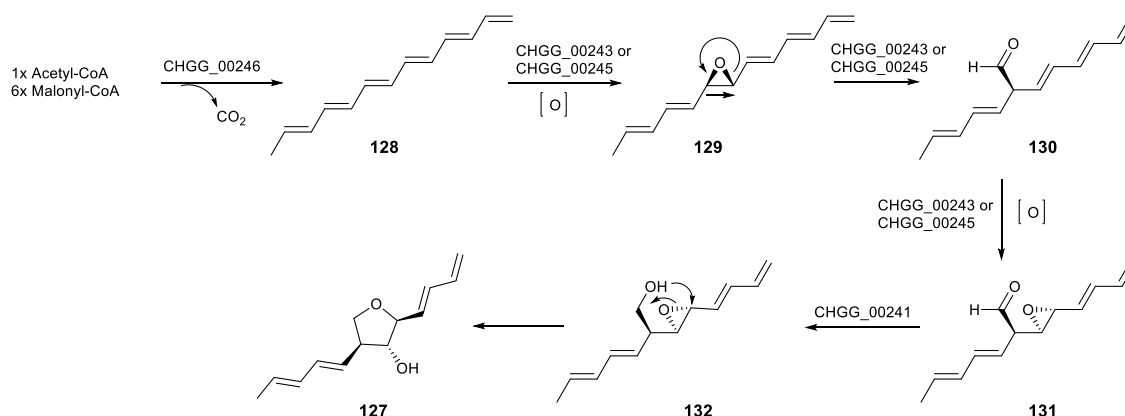
The domain structure of the PKS (*CHGG\_00246*) is highly similar to polyene PKS (DEP5). It consists of KS, AT, DH, ER<sup>0</sup>, KR and ACP domains (Figure 4.3). Like the depudecin PKS (DEP5), the aureonitol PKS (*CHGG\_00246*) belongs to the normal fungal hrPKS type.<sup>11,87</sup> Once again, the lack of a *trans*-ER is consistent with a polyene intermediate.

Nakazawa and co-workers<sup>91</sup> hypothesized a biosynthetic pathway for the C<sub>13</sub> polyketide aureonitol **127**. The polyketide polyene backbone **128** (Scheme 4.3) is probably assembled by an hrPKS (*CHGG\_00246*). Then, an FMO (*CHGG\_00243* or *CHGG\_00245*) is assumed to execute the epoxidation to yield **129**. A following epoxide rearrangement likely occurs to make aldehyde **130**. During strobilurin biosynthesis, there is a similar oxidative rearrangement step which is catalyzed by the FMO (Str9).<sup>92</sup> The intermediate **130** could be oxidized by an FMO (*CHGG\_00243* or *CHGG\_00245*) at positions C-6 and C-7 to form **131**. Subsequently, an oxidoreductase (*CHGG\_00241*) possibly reduces aldehyde **131** to alcohol **132**. Finally, furan ring

formation yields aureonitol **127**. However, the mechanism of how the polyene precursor **128** is produced cannot be proposed before our study.



**Figure 4.3** Aureonitol **127** biosynthetic gene cluster in *Chaetomium globosum*, adapted from Nakazawa *et al.*, 2013.<sup>91</sup> ER<sup>0</sup>, non-functional ER domain.



**Scheme 4.3** Proposed biosynthetic pathway for aureonitol **127**, adapted from Nakazawa *et al.*, 2013.<sup>91</sup>

Structures of aureonitol **127** and trienylfuranol A **2** all have the furan rings with polyene substituents, therefore their biosynthesis should reasonably share some similarities. But, whether there exists a similar oxidative rearrangement as in the aureonitol **127** case during trienylfuranol A **2** biosynthesis, is not known.

#### 4.1.3 Polyenoic Acid

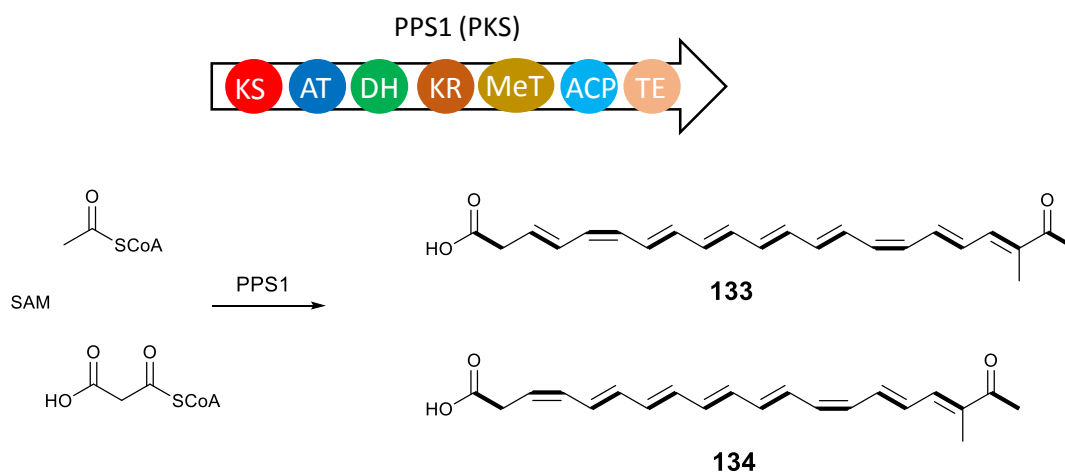
Recently, a unique polyene PKS (PPS1) was identified by the Hoffmeister group from a taxonomically yet unidentified agaricomycete fungus (BY1).<sup>93</sup> Heterologous expression of PPS1 in *Aspergillus niger* obtained two polyene compounds, nonaenoic acid **133** and octaenoic acid

**134** (Scheme 4.4). The  $^{13}\text{C}$  labelling studies confirmed the polyketide origin for these two polyenes.<sup>94</sup>

The domain architecture of PPS1 is shown in Scheme 4.4, in which the KS, AT, DH, KR, C-MeT, ACP and TE domains are identified. The TE domain located at the C-terminus of PPS1 makes it different to the normal fungal hrPKS, such as the DEP5 (depudecin **126** in Figure 4.2) and CHGG\_00246 (aureonitol **127** in Figure 4.3). In addition, the C-MeT domain in PPS1 is functional and there are methylations in polyketide **133** and **134**. In contrast, the C-MeT domains in PKS DEP5 and CHGG\_00246 are not present, and no methylations are observed in the structures of depudecin **126**, aureonitol **127** and trienylfuranol A **2**.

In addition, no decarboxylation or oxidation is observed in **133** and **134**. However, during depudecin **126**, aureonitol **127** and trienylfuranol A **2** biosynthesis, decarboxylation and oxidation reactions are all required. Although polyenoic acids **133** and **134** are produced by a fungus, their PKS domain architecture and biosynthesis differ from those that make depudecin **126**, aureonitol **127** and trienylfuranol A **2**.

Interestingly, Hoffmeister and co-workers found that the expression of PPS1 is significantly upregulated when the fungal mycelium is injured. The injury-induced *de novo* synthesis of the toxic polyene pigments **133** and **134** is regarded as a fungal response strategy.<sup>95</sup>

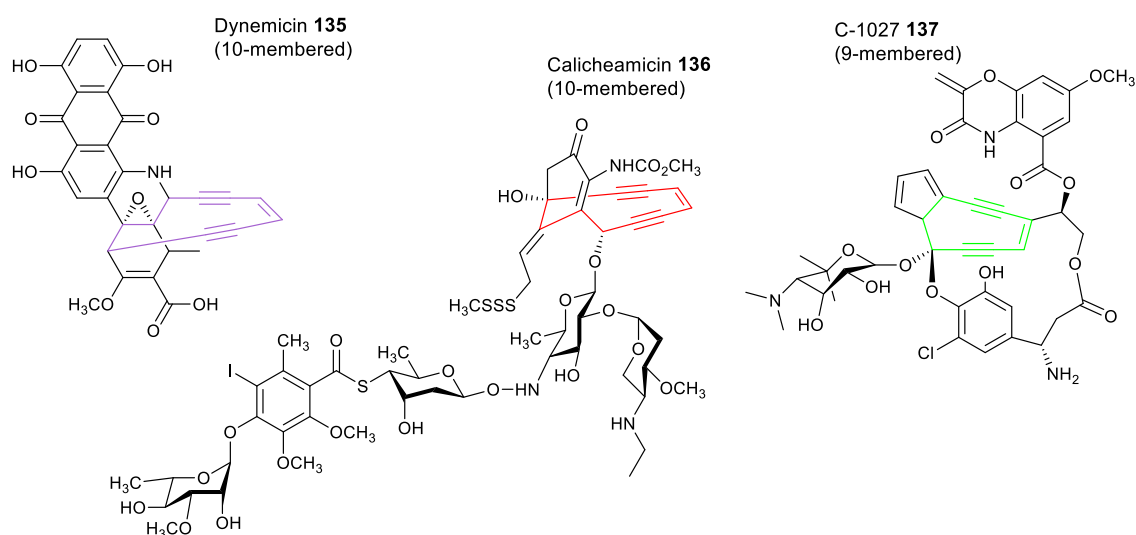


**Scheme 4.4** The hrPKS (PPS1) derived from agaricomycete fungus (BY1) produces two polyene compounds, adapted from Brandt *et al.*, 2017.<sup>93</sup> The bold bonds represent the intact acetate units.

#### 4.1.4 Bacterial Eneidiynes

The eneidiynes are a family of natural products isolated from bacteria. The molecular architecture is a 9- or 10-membered ring containing two acetylenic groups conjugated to a double bond. These structures are exemplified by demycin **135**, calicheamicin **136** and C-1027 **137** (Figure 4.4).<sup>96</sup>

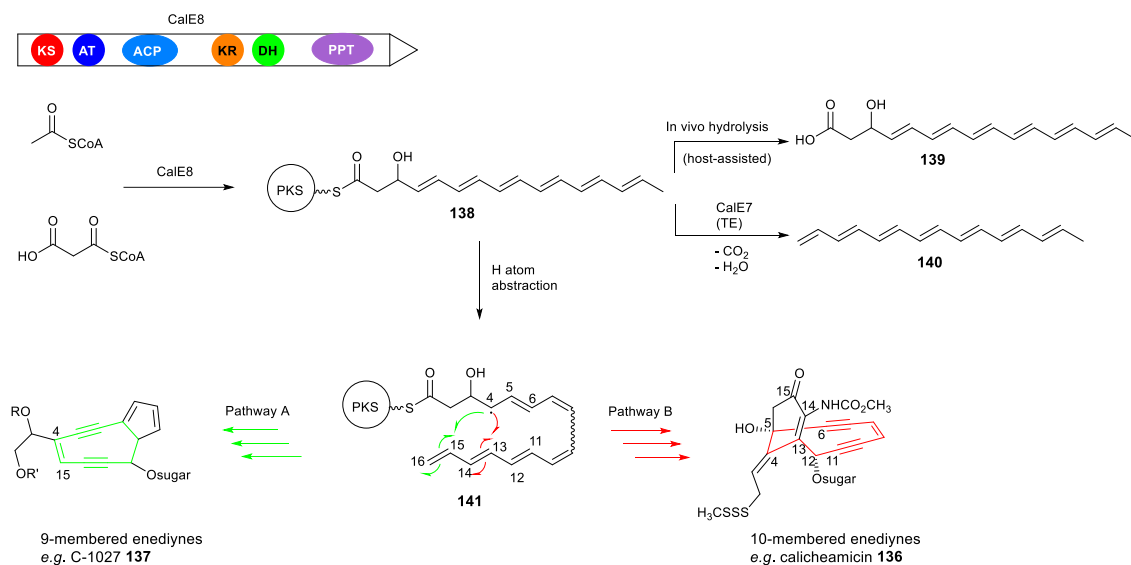
Bioactivity studies show that enediynes are potent antibiotics and they also possess strong cytotoxicity against tumour cells.<sup>97</sup> Because of their unprecedented scaffolds, enediynes have generated high interest to decipher their biosynthesis. Early studies employing isotopic labelling experiments proved acetate as the carbon source for backbone construction.<sup>98</sup> Entering the genomic age, the gene clusters responsible for enediyne biosynthesis have been increasingly found. These BGC encode an iterative type I PKS.<sup>99–101</sup>



**Figure 4.4** Enediynes antibiotics: dynemicin **135**, calicheamicin **136**, C-1027 **137**.

A gene KO strategy was applied to verify the role of the PKS during enediyne biosynthesis.<sup>100</sup> Heterologous expression of *pks* and a thiolesterase gene (exemplified by *sgcE/sgcE10* and *ncsE/ncsE10*) in *Escherichia coli* by the Shen group<sup>96</sup> produced the first isolable and characterized intermediate, the polyene 1,3,5,7,9,11,13-pentadecaheptaene **140**. The mechanism was proposed to include a decarboxylation/dehydration of the  $\beta$ -hydroxy octaketide **139** to make **140** (Scheme 4.5). The polyketide polyene **140** was regarded as a highly significant core precursor for 9- or 10-membered enediynes construction at the early biosynthesis investigations.

However, the Townsend group's research indicated that heptaene **140** is just a shunt product of the enzyme-bound octaketide **65** (Scheme 4.5).<sup>102,103</sup> Although heptaene **140** is a common product of enediyne PKS/TE pairs from all subfamilies, it is not a precursor of the enediynes natural products. Subsequently, the Townsend group<sup>102,103</sup> observed the PKS-bound **138** *in vitro*, and their research indicated that the release of the free  $\beta$ -hydroxy acid **139** from the corresponding PKS-bound **138** is host assisted (*in vivo* hydrolysed by *E. coli*) rather than programmed. Therefore, they concluded that enzyme-bound polyene octaketide **138** is the true biosynthetic intermediate to enediynes as well as the branch point for a divergence to different enediynes subclasses (Scheme 4.5).<sup>102,103</sup>



**Scheme 4.5** Proposed biosynthesis of enediynes by Belecki *et al.*, 2013.<sup>102,103</sup>

A proposed mechanism is shown in Scheme 4.5. The abstraction of a hydrogen atom at the terminal C-16 methyl of **138** yields the resonance equivalent **141**, which can direct to two divergent pathways A and B by an accessory enzyme catalysing regiospecific cyclization. The assumed pathway A is adapted to describe the buildup of the core architecture of 9-membered enediyne (*e.g.* C-1027 **137**). However, the hypothesized cyclization manner in pathway B could be used to form 10-membered enediyne (*e.g.* calicheamicin **136**).<sup>102,103</sup>

Bioinformatic analysis of the PKS (CalE8) identified KS, AT, ACP, KR, DH and PPTase domains (Scheme 4.5). However, the domain architecture of enediyne PKS (CalE8) is very unusual. For example, the ACP domain is located between the AT and KR domains that are similar to the domain structure of polyunsaturated fatty acid (PUFA) synthases.<sup>104</sup> In addition, a PPTase domain is positioned at the C-terminal region. And the self-phosphopantetheinylation makes the enediyne PKS extremely unique.<sup>96</sup>

Thus, although the polyketide polyene **138** was proved to be the key precursor for enediyne biosynthesis, the domain architecture of bacterial enediyne PKS is quite different from that of the fungal hrPKS, such as DEP5 (depudecin **126** in Figure 4.2) and CHGG\_00246 (aureonitol **127** in Figure 4.3). Therefore, the polyene biosynthesis in enediyne is probably not similar to the logic of depudecin **126** and aureonitol **127**.

## 4.2 Project Aims

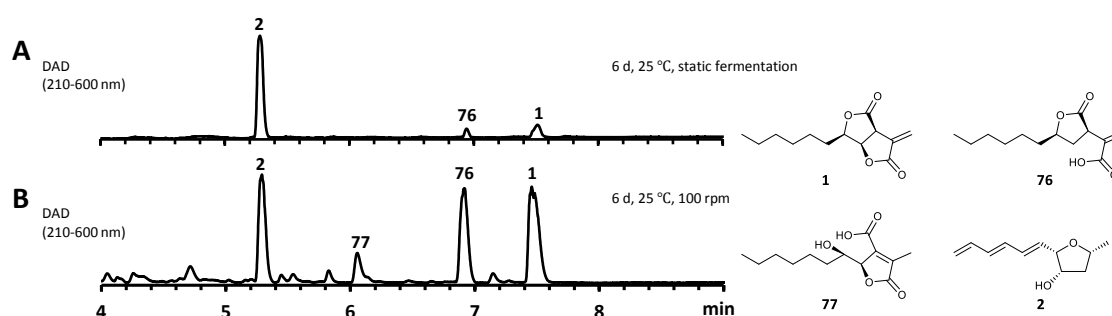
This project will focus on the delineation of trienylfuranol A **2** biosynthesis. Isotopic labelling experiments will be utilized to get the first insight about trienylfuranol A **2** scaffold construction. In particular, we want to know if there has been any rearrangement during biosynthesis and from

which end of the chain decarboxylation has occurred. Blast searches of the three *Hypomontagnella* genomes, utilizing the protein sequences of DEP5 (depudecin PKS) and CHGG\_00246 (aureonitol PKS) as queries, will be done to obtain the hits reservoir of hrPKS. The C-MeT and ER domain alignments will be performed to pick out the potential polyene hrPKS. Linking the genome data with transcriptomic analysis will also be done. Then, detailed gene annotation will be made to the top hits gene clusters. Gene knockout in *H. spongiphila* and heterologous expression in *A. oryzae* will be performed for the potential polyene hrPKS gene clusters. Isolation and characterization of the compounds will be done if possible. Then we will try to uncover the BGC and biosynthetic pathway of trienylfuranol A **2**.

### 4.3 Results

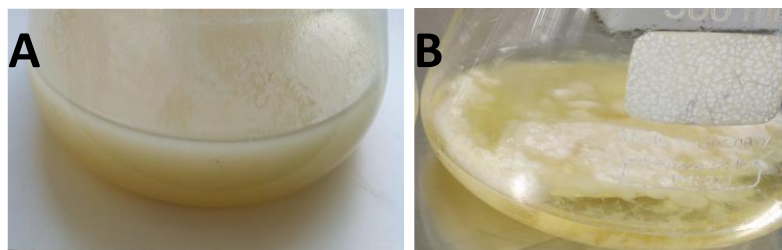
#### 4.3.1 Trienylfuranol A Production

The literature<sup>56</sup> production conditions for trienylfuranol A **2** are cultivation of *H. submonticulosa* in PDB medium at 25 °C, in static conditions, for 6 weeks. To decrease the fermentation time, we developed producing conditions in which static or shaken (100 rpm) fermentation for only 6 days (Figure 4.5) can yield a reasonable amount of **2**. Sporothriolide **1** and analogues **76** and **77** are also observed from *H. submonticulosa*, however, the previous study from the Sumarah group did not report this.



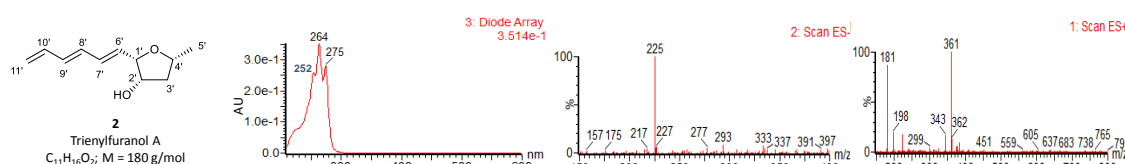
**Figure 4.5** The production of trienylfuranol A **2**. **A**, DAD chromatogram of *H. submonticulosa* DAOMC 242471 WT grown under PDB medium, 6 d, 25 °C, still; **B**, DAD chromatogram of *H. submonticulosa* DAOMC 242471 WT under PDB medium, 6 d, 25 °C, 100 rpm.

The *H. submonticulosa* growth in shaken conditions (Figure 4.6A) after 6 days resulted in a homogeneous culture, which is similar to the morphology of *H. monticulosa* and *H. spongiphila* under producing conditions (28 °C, 130 rpm, 6 days). Under static fermentation conditions (Figure 4.6B), the fungus forms a mat that floats on the top of the liquid media. In addition, the morphological characters of three fungi grown on PDA solid agar show some differences, on which the *H. monticulosa* and *H. submonticulosa* formed sporulating regions after 14 days while marine-derived *H. spongiphila* remained sterile.<sup>37</sup>



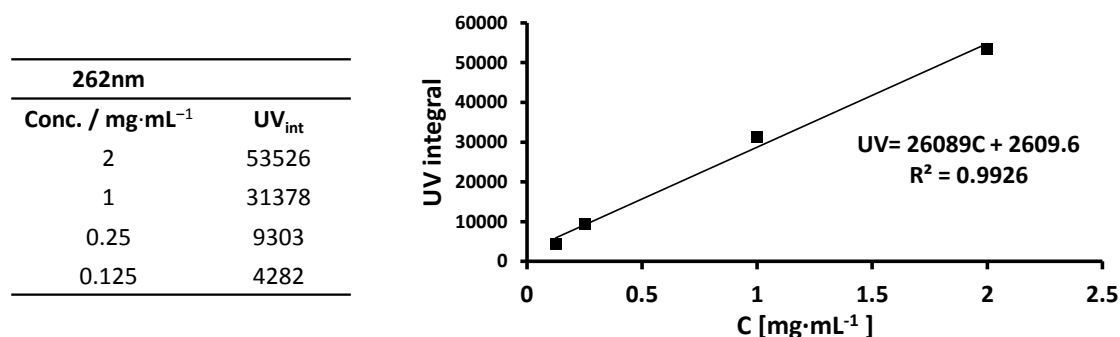
**Figure 4.6** Culture morphology of *Hypomontagnella submonticulosa*. **A**, after 6 days of growth at 25 °C in shaken (100 rpm) conditions; **B**, after 6 days of growth at 25 °C in static conditions.

Through large-scale fermentation (1 L) of *H. submonticulosa* DAOMC 242471 in PDB medium (static conditions, 25 °C, 6 days), we isolated 23 mg of trienylfuranol A **2** (Figure 4.7) from the extracts and then performed 1D and 2D NMR experiments to confirm its structure (see Chapter 7 for details). The shaken conditions are used as the producing conditions in the time course study of *H. submonticulosa* (Section 2.3.1.2, Section 3.3.2, Section 4.3.1), because of the higher titre of sporothriolide **1**, and its analogues are also produced under these conditions (Figure 4.5B).



**Figure 4.7** UV and mass spectra (ES<sup>+</sup> and ES<sup>-</sup>) of trienylfuranol A **2**.

To quantify the production of compounds, a calibration curve for trienylfuranol A **2** was made based on a UV integration method. A dilution series of **2** (2.0, 1.0, 0.25, 0.125 mg·mL<sup>-1</sup>) was measured and the corresponding signals for the extracted wavelength at 262 nm ( $\lambda_{\max}$ ) were integrated (Figure 4.8A). Integrated values were then plotted against the sample concentration and fitted into a straight line (Figure 4.8B). The relationship is linear within the 0.125 – 2.0 mg·mL<sup>-1</sup> concentration range. The equation was applied to quantify **2** across different samples.



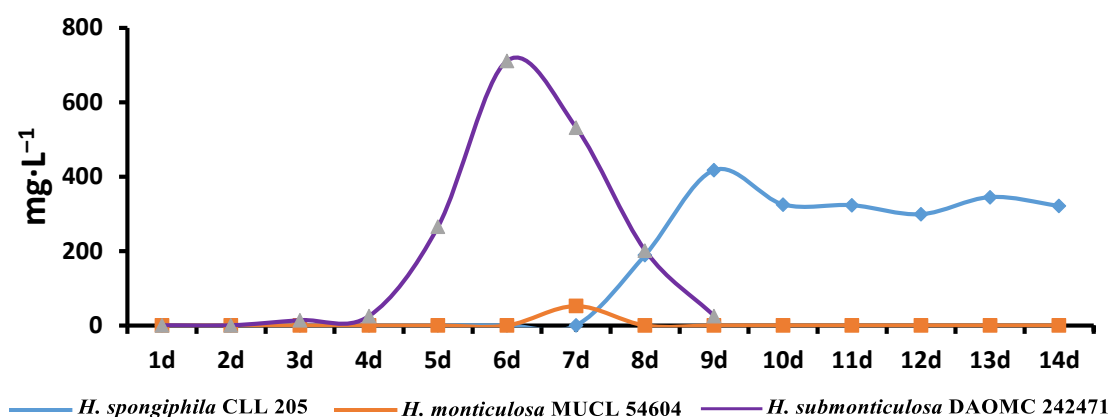
**Figure 4.8 A**, the integrated values of different dilutions of trienylfuranol A **2** solutions at 262 nm from LCMS analysis; **B**, calibration curve for trienylfuranol A **2** quantification.

Timecourse experiments for the three *Hypomontagnella* wild type strains were conducted to monitor the production of **2**, the procedure details refer to Section 2.3.1.2. Compound titres were

calculated using the equation  $C = (UV_{\text{int}} - 2609.6)/26089$  (Table 4.1). Production kinetics of **2** were visualized graphically (Figure 4.9).

	<i>H. monticulosa</i>	<i>H. spongiphila</i>	<i>H. submonticulosa</i>
Day	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>	Conc. / mg·L <sup>-1</sup>
1	0	0	0
2	0	0	0
3	0	0	13.7
4	0	0	24.9
5	0	0	265.5
6	0	0	711.6
7	52.2	0	532.1
8	0	188.7	201.2
9	0	417.2	25.1
10	0	325.0	-
11	0	323.0	-
12	0	298.8	-
13	0	344.9	-
14	0	321.2	-

**Table 4.1** Temporal production of trienylfuranol A **2** from *H. monticulosa* MUCL 54604, *H. spongiphila* CLL 205, and *H. submonticulosa* DAOMC 242471 cultivated under producing conditions.

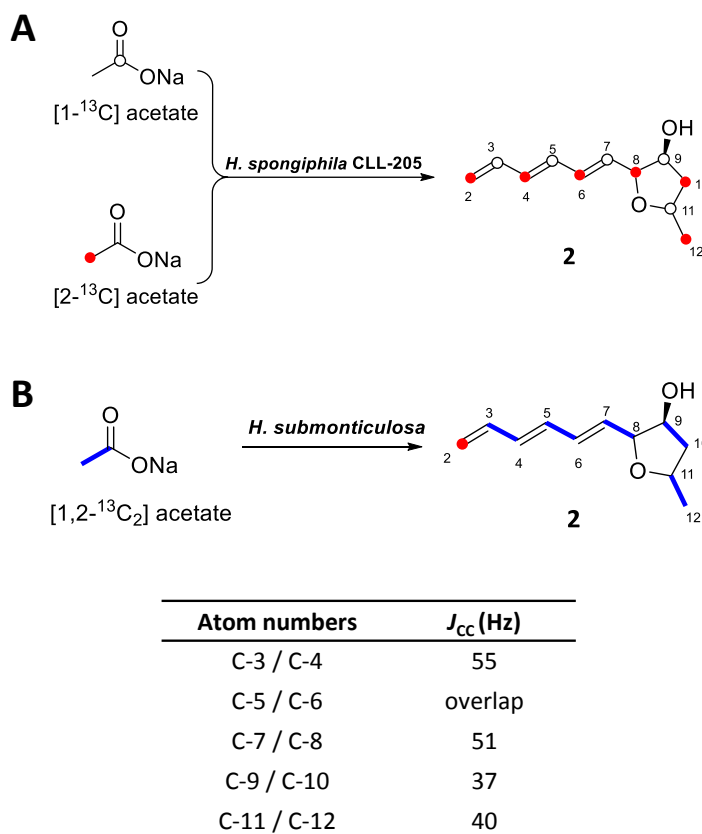


**Figure 4.9** Kinetic production curve of trienylfuranol A **2**.

#### 4.3.2 Acetate Feeding Experiment

During the sporochartine biosynthetic studies (Chapter 3), the labelling patterns of **3b** (Table 3.3, Figure 4.10A) fit with the hypothesis of a polyketide origin for trienylfuranol A **2**. In addition, carbon C-8 derived from C-2 acetate is oxygenated indicating the involvement of oxidation during biosynthesis. The results also showed that the trienylfuranol A **2** backbone is built in an order of one C-1 acetate carbon linked with one C-2 acetate carbon, which means the carbon-carbon rearrangement steps observed during aureonitol **127** biosynthesis<sup>91</sup> do not occur in **2**. However, the loss of C-1 acetate-derived carbon at C-2 or at C-12 opens the question of which side is the ‘head’ or ‘tail’ during the polyketide chain construction (Figure 4.10A).

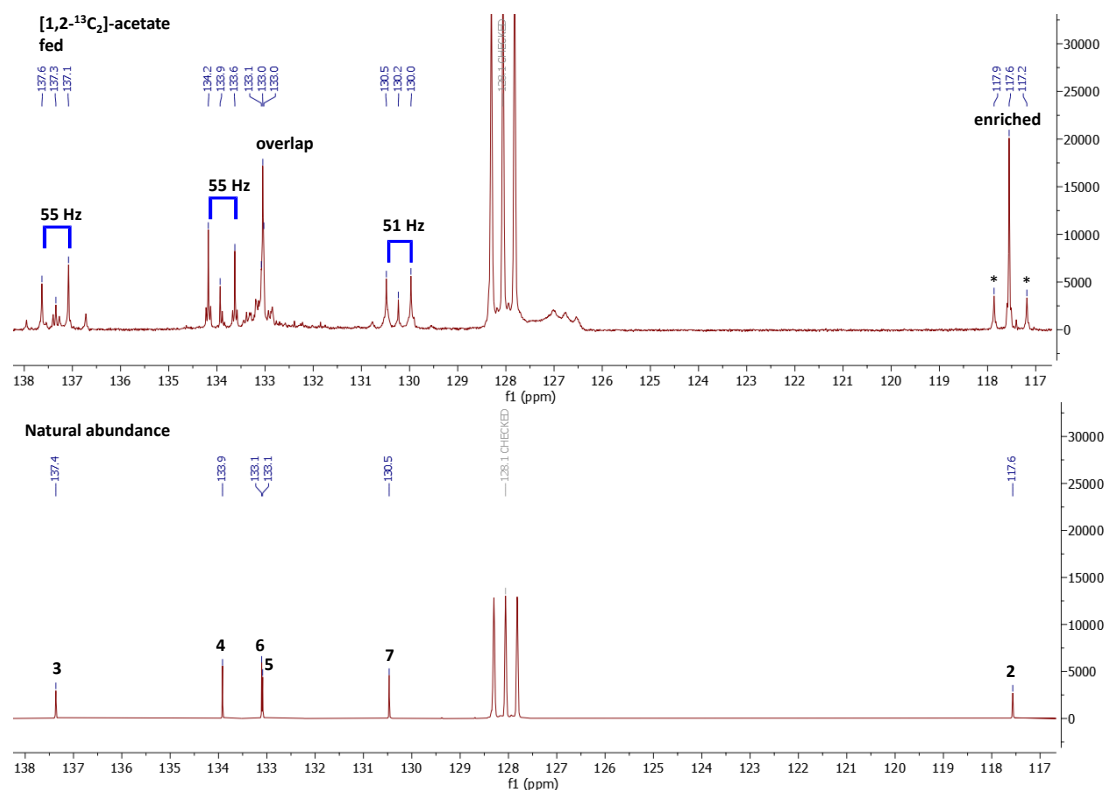




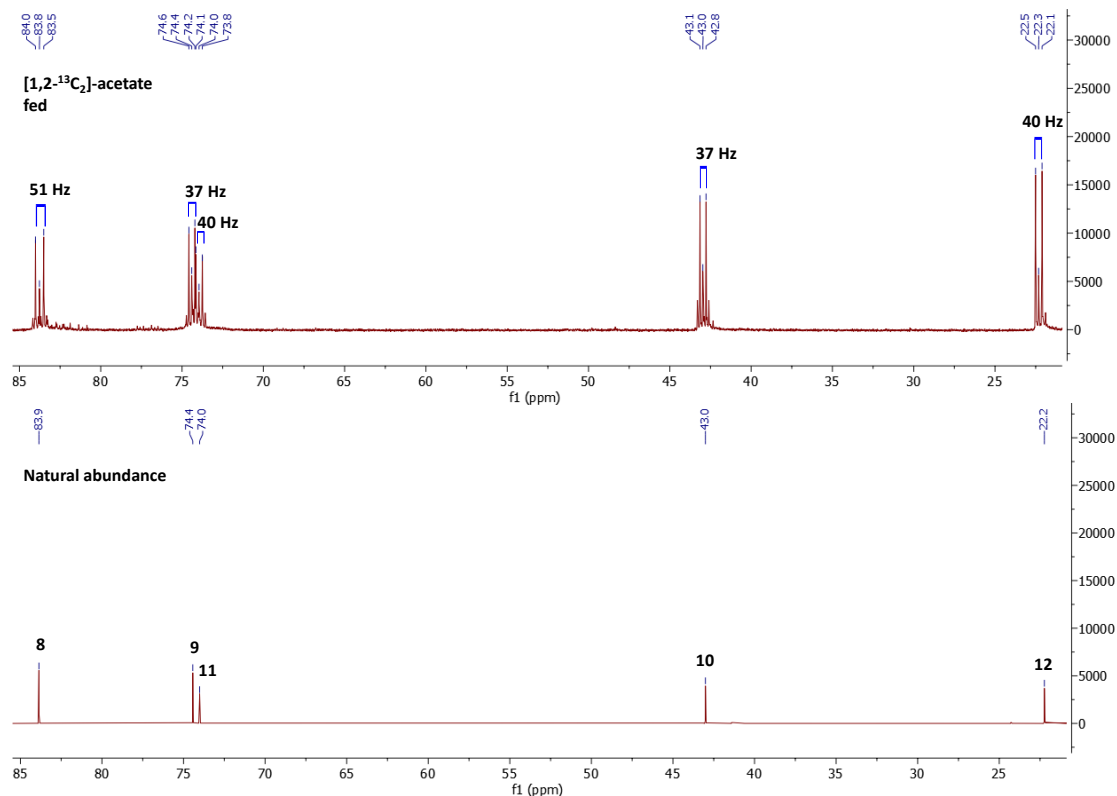
**Figure 4.10** **A**, the labelling patterns of **2** from  $[1-^{13}\text{C}]$  and  $[2-^{13}\text{C}]$  sodium acetate feeding experiments (Table 3.3); **B**, the labelling patterns and one bond  $^{13}\text{C}$ - $^{13}\text{C}$  spin-spin couplings constants observed in trienylfuranol A **2** derived from  $[1,2-^{13}\text{C}_2]$  acetate feeding experiment.

If the  $[1,2-^{13}\text{C}_2]$  acetates are used as the carbon source, a  $^{13}\text{C}$ - $^{13}\text{C}$  spin-spin coupling ( $^1J_{CC}$ ) could be observed in  $^{13}\text{C}$  NMR when the double labelled acetates remain intact during the construction of the scaffold. Alternately, if the acetate unit is broken, then the peak intensity of the broken carbon will be enriched without showing  $^1J_{CC}$  coupling.<sup>32</sup>

To investigate the polyketide chain extension direction during trienylfuranol A **2** biosynthesis,  $[1,2-^{13}\text{C}_2]$  sodium acetates were utilized in the feeding experiment, the procedures refer to the previous description in Section 2.3.3. Labelled **2** was extracted and purified (Section 6.2.1).  $^{13}\text{C}$  NMR spectra revealed all carbons from C-3 to C-12 display a doublet with coupling constants from 37 and 55 Hz, except the overlap signals of C-5 and C-6, which are characteristics for intact acetate units (Figure 4.10 – 4.12). However, the resonance for C-2 is enriched but uncoupled (the extra  $^1J_{CC}$  coupling of C-2/C-3 was observed), showing it is a C-2 carbon from a cleaved acetate unit. In summary, the  $^{13}\text{C}$  incorporation patterns of trienylfuranol A **2** reveal it assembled from a polyene hrPKS, and there are several oxidative reactions involved during the biosynthesis.



**Figure 4.11** Natural abundance <sup>13</sup>C NMR spectrum (chemical shift between 113 and 142 ppm, 100 MHz in C<sub>6</sub>D<sub>6</sub>) of trienylfuranol A **2** compared to <sup>13</sup>C NMR spectrum of the feeding experiment ([1,2-<sup>13</sup>C<sub>2</sub>] sodium acetate). Asterisk: the <sup>1</sup>J<sub>CC</sub> coupling of C-2/C-3.

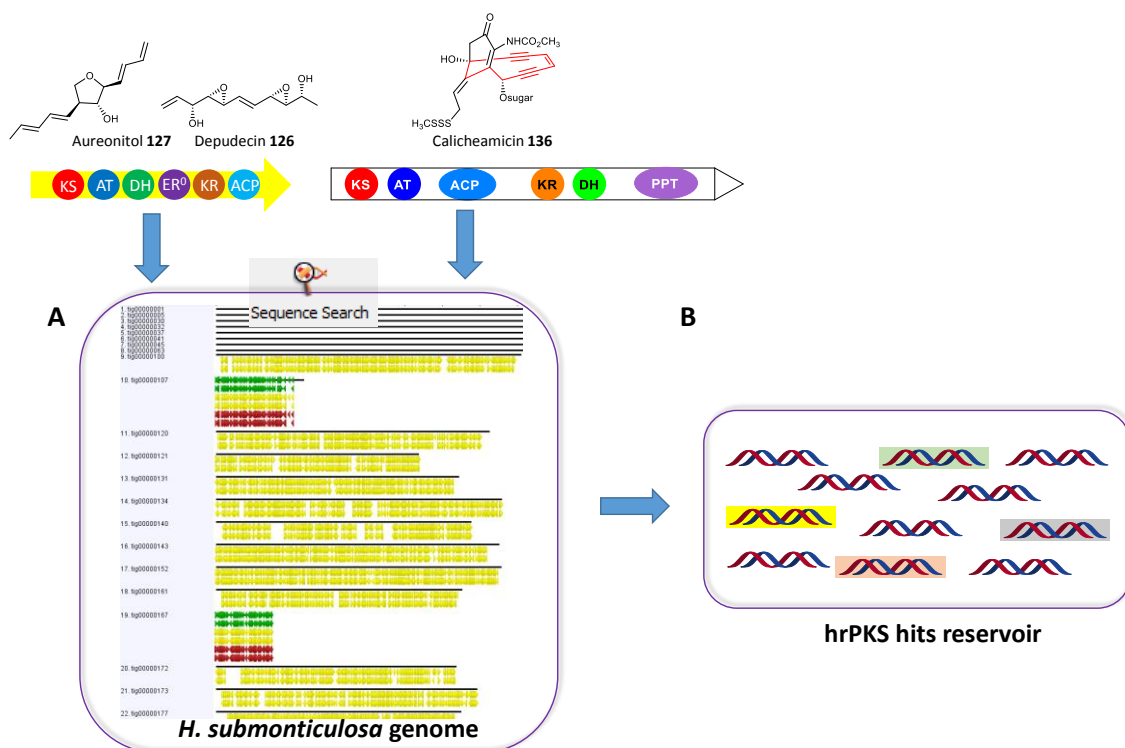


**Figure 4.12** Natural abundance <sup>13</sup>C NMR spectrum (chemical shift between 20 and 90 ppm, 100 MHz in C<sub>6</sub>D<sub>6</sub>) of trienylfuranol A **2** compared to <sup>13</sup>C NMR spectrum of the feeding experiment ([1,2-<sup>13</sup>C<sub>2</sub>] sodium acetate).

### 4.3.3 Mining of Polyketide Polyene hrPKS

#### 4.3.3.1 Genome Screening for hrPKS

Based on the  $^{13}\text{C}$  isotopic labelling study, trienylfuranol A **2** is probably built by a polyene hrPKS. Therefore, the first step is to screen the *Hypomontagnella* genomes and find all the potential polyene hrPKS gene clusters. The sequences of DEP5 (depudecin **126** PKS), CHGG\_00246 (aureonitol **127** PKS) and CalE8 (enediynes calicheamicin **136** PKS) were utilized as queries to blast the *H. submonticulosa* genome (Figure 4.13). Because the domain architecture of the bacterial enediynes PKS (CalE8) is distinct from that of fungal hrPKS, there was no significant similarity found from the *H. submonticulosa* genome. The PKS hits lists created by the blast searches of DEP5 and CHGG\_00246 were combined, and the nrPKS and PKS-NRPS (polyketide synthase-nonribosomal peptide synthetase) were manually excluded. In addition, the PKS that only exist in *H. submonticulosa* but not in *H. monticulosa* or *H. spongiphila*, were also excluded. A total of eighteen hrPKS gene clusters was obtained (Table 4.2).



**Figure 4.13** Genome screening of potential polyene hrPKS in *H. submonticulosa*. AT, acyltransferase; KS, ketosynthase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; ER<sup>0</sup>, non-functional ER domain.

#### 4.3.3.2 hrPKS Domain Analysis and Transcriptomic Analysis

From the structural point of view (depudecin **126**, aureonitol **127** and trienylfuranol A **2**), the C-MeT and ER domains in polyene hrPKS should be non-functional or missing. This was found to be the case for both the depudecin **126** and aureonitol **127** PKS. For instance, DEP5 (depudecin

PKS) and CHGG\_00246 (aureonitol PKS) lack C-MeT domains (Figure 4.2 – 4.3; Table 4.2). And the ER domains in the two PKS are inactive because the NADPH cofactor binding sites (LxHx(G/A)xGGVG) are incomplete based on the ER sequence alignments (Figure 7.86).

To evaluate the status of the C-MeT and ER domain in the obtained eighteen hrPKS and to pick out the potential polyene hrPKS from the *Hypomontagnella* genomes, detailed domain analysis was performed. For example, the prediction of PKS domain architecture by Interpro<sup>105</sup> and SMART<sup>106</sup> (Table 4.2), as well as the C-MeT and ER domain sequence alignments (Figure 7.86 – 7.87). In the ER domain alignments (Clustal Omega (1.2.4)<sup>107</sup>), the intact NADPH cofactor binding sites (LxHx(G/A)xGGVG)<sup>108</sup> were regarded as the basis of function. While the ‘GAGTG’ conserved cofactor motif and His-Glu dyad<sup>109</sup> were used to predict the C-MeT domain function.

In Table 4.2, seventeen PKSs have the ER domain (PKS12 is exceptional), as determined *via* PKS domain architecture analysis. And the ER domains in 16 PKS (PKS1 – 11, and PKS14 – 18) are probably inactive based on the sequence alignments shown in Figure 7.86. In addition, the C-MeT domain is present in 13 PKS (Table 4.2) by PKS architecture analysis. The C-MeT domain is missing in PKS4 – 5, PKS9 – 10, and PKS13 (Table 4.2). The sequence alignments of these thirteen C-MeT domains revealed that seven of them (from PKS1 – 3, PKS6 – 8, and PKS11) are possibly non-functional since their conserved cofactor motif are incomplete (Figure 7.87).

PKS	Cluster	Gene <i>H. monticulosa</i>	Domain Structure	NEL A (non-producing)	NEL B (producing)	Log <sub>2</sub> -fold change (B/A)	Identity / Similarity (%), to DEP5	Identity / Similarity (%), to CHGG_00246
1	Cluster A	HmMg6165	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	0.10	1488.35	13.86	60.8 / 73.9	41.8 / 55.3
2	Cluster B	HmMg10163	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	60.11	86.35	0.52	27.9 / 42.4	23.4 / 37.0
3	Cluster B	HmMg10159	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	101.57	290.21	1.51	22.2 / 37.0	18.6 / 31.9
4	Cluster C	HmMg3162	KS AT DH ER <sup>0</sup> KR ACP	0.10	2.16	4.43	29.3 / 44.9	25.1 / 38.8
5	Cluster D	HmMg1580	KS AT DH ER <sup>0</sup> KR ACP	0.10	3.11	4.96	27.6 / 44.0	23.1 / 37.5
6	Cluster E	HmMg5986	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	99.51	1445.88	3.86	24.5 / 40.0	20.3 / 33.8
7	Cluster F	HmMg1069	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	2.07	11.40	2.46	21.3 / 36.2	18.5 / 28.1
8	Cluster G	HmMg3307	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	19.69	47.37	1.27	19.4 / 33.4	16.3 / 27.6
9	Cluster H	HmMg1003	KS AT DH ER <sup>0</sup> KR ACP	NE	NE	NE	25.2 / 40.0	21.0 / 33.1
10	Cluster I	HmMg6736	KS AT DH ER <sup>0</sup> KR ACP	20.73	1.07	-4.27	25.6 / 40.5	21.9 / 33.7
11	Cluster J	HmMg7628	KS AT DH ER <sup>0</sup> KR ACP C-MeT <sup>0</sup>	67.37	116.57	0.79	26.9 / 43.3	21.4 / 34.5
12	Cluster J	HmMg7624	KS AT DH C-MeT <sup>+</sup> KR ACP	NE	NE	NE	22.4 / 36.2	17.2 / 30.9
13	Cluster K	HmMg8513	KS AT DH ER <sup>0</sup> KR ACP	NE	NE	NE	26.4 / 42.3	20.4 / 34.2
14	Cluster L	HmMg4170	KS AT DH C-MeT <sup>0</sup> ER <sup>0</sup> KR ACP	49.75	99.82	1.00	25.5 / 40.9	21.6 / 34.1
15	Cluster M	HmMg1180	KS AT DH C-MeT <sup>+</sup> ER <sup>0</sup> KR ACP	123.33	172.28	0.48	22.4 / 36.4	19.0 / 30.5
16	Cluster N	HmMg1129	KS AT DH C-MeT <sup>+</sup> ER <sup>0</sup> KR ACP	NE	NE	NE	22.7 / 39.5	19.2 / 33.1
17	Cluster O	HmMg4795	KS AT DH C-MeT <sup>+</sup> ER <sup>0</sup> KR ACP	NE	NE	NE	23.3 / 39.1	20.3 / 34.6
18	Cluster P	HmMg3964	KS AT DH C-MeT <sup>+</sup> ER <sup>0</sup> KR ACP	NE	NE	NE	24.9 / 38.8	20.7 / 33.3
-	Depudecin	DEP5	KS AT DH ER <sup>0</sup> KR ACP	-	-	-	-	40.3 / 53.5
-	Aureonitol	CHGG_00246	KS AT DH ER <sup>0</sup> KR ACP	-	-	-	40.3 / 53.5	-

**Table 4.2** The overview of hits PKS from *H. monticulosa*. The domain structure (Green colour: active; Red colour: inactive; Black colour: function not analysed), transcriptome data (no expression and downregulation are highlighted in grey boxes; upregulation is highlighted in pink boxes), and identity/similarity (Calculation by using Blosum62 with threshold 1) to known polyene PKS are shown in this table. ER<sup>0</sup>, non-functional ER domain; C-MeT<sup>0</sup>, non-functional C-MeT domain; NEL: normalized expression level; NE: no expression.

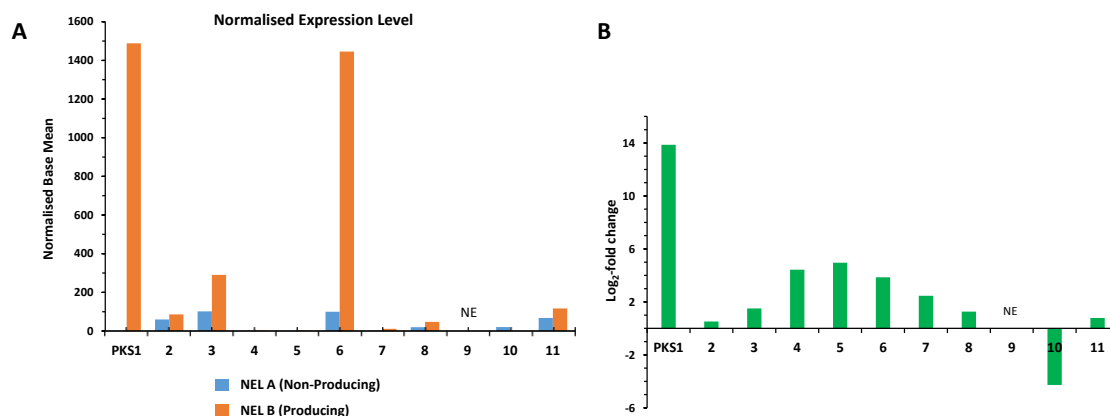
Finally, eleven potential polyene hrPKS (PKS1 – 11 in Table 4.2) with missing or likely non-functional ER and C-MeT domains were selected from the *Hypomontagnella* genomes. The PKS genes from *H. monticulosa* were exemplified in Table 4.2.

During the *spo* BGC prediction (Section 2.3.4), transcriptome data from *H. monticulosa* MUCL played a role to reveal the gene expression levels under producing and non-producing conditions. Genes located inside the *spo* BGC show a strong upregulation under producing conditions (Figure 2.14), while genes outside this region show either low expression (Table 2.6, Figure 2.14) or no changes in transcription level. By analysing the LCMS chromatograms of *H. monticulosa* MUCL 54604 extracts (Figure 2.2), we found that the production of trienylfuranol A **2** adapts to the producing and non-producing conditions of sporothriolide **1**. Therefore, the transcriptome data that we obtained from sporothriolide producing and non-producing conditions should be also useful for analysis of the expression of potential trienylfuranol A biosynthetic genes.

We made a comprehensive expression level comparison of the top 11 potential polyene hrPKS by analysing the transcriptome data from *H. monticulosa* MUCL (Table 4.2, Figure 4.14). Results showed that *PKS1* from ‘Cluster A’ (Table 4.2) not only has the largest NEL (normalised expression level; BaseMean) value but also has the most significant log<sub>2</sub>-fold change between producing and non-producing conditions. In a word, the *PKS1* in ‘Cluster A’ is highly upregulated under producing conditions, as compared to non-producing conditions. In addition, compared to the other 11 hrPKS (PKS1 – 11; Table 4.2) with depudecin **126** and aureonitol **127** PKS, *PKS1* in ‘Cluster A’ (Table 4.2) shows the highest identity and similarity among the hrPKS lists, respectively (60.8% / 73.9% to DEP5, and 41.8% / 55.3% to CHGG\_00246).

‘Cluster B’ harbours two hrPKS (*PKS2* and *PKS3*; Table 4.2), and the NEL values (Table 4.2, Figure 4.14) of these two hrPKS are both relatively high and the identity/similarity of *PKS3* to polyene PKS DEP5 and CHGG\_00246 are 27.9% / 42.4% and 23.4% / 37.0%, respectively. The hrPKS expression levels (NEL values) of ‘Cluster C’ (*PKS4*; Table 4.2) and ‘Cluster D’ (*PKS5*; Table 4.2, Figure 4.14) are very low, but these two PKS are relatively closer to the polyene PKS of depudecin **126** and aureonitol **127**, as compared with *PKS2* and *PKS3* in ‘Cluster B’ (see detailed values in Table 4.2). The *PKS6* in ‘Cluster E’ (Table 4.2, Figure 4.14) was highly expressed and strongly upregulated (Log<sub>2</sub> fold change is 3.86) from non-producing to producing conditions, but it seems not as close to DEP5 and CHGG\_00246 as compared to *PKS2*, *PKS4* and *PKS5*, the details were shown in Table 4.2.

In summary, results of the extensive bioinformatic analysis suggested ‘Cluster A’ encodes a possible trienylfuranol A **2** PKS. However this data is not conclusive, we need more detailed gene annotation and analysis of these gene clusters.



**Figure 4.14** The transcriptomic analysis of 11 potential polyene hrPKS in *H. monticulosa* MUCL (Data based on Table 4.2). **A**, bar chart of the normalised expression level of the hrPKS, non-visible bar means either no expression or very low NEL BaseMeans (Table 4.2); **B**, bar chart of Log<sub>2</sub>-fold changes of the hrPKS. NE, no expression.

### 4.3.4 Potential Polyene hrPKS BGC Analysis

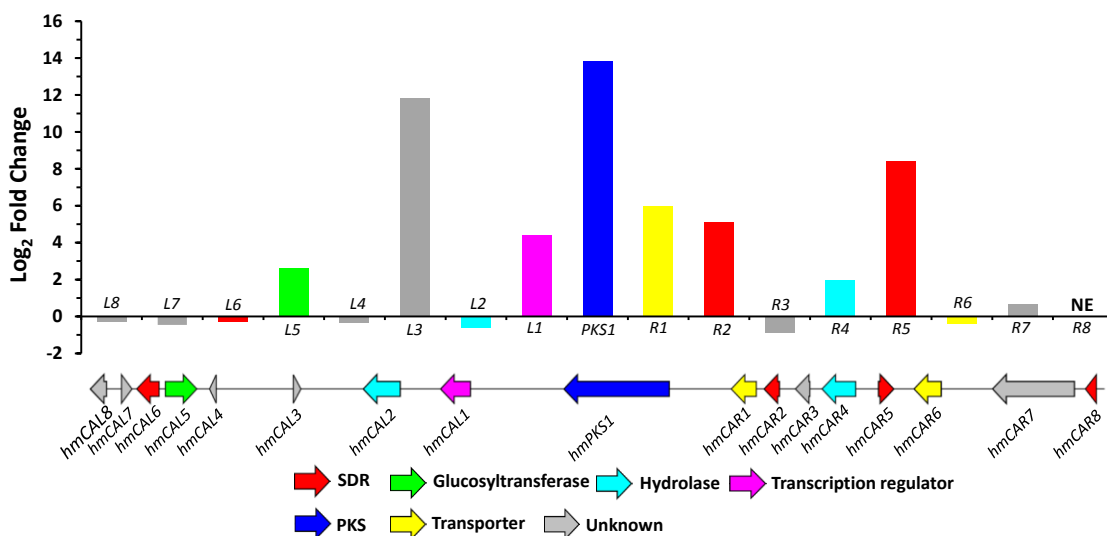
#### 4.3.4.1 ‘Cluster A’ Analysis

Comprehensive gene cluster annotation serves as the first step for understanding the function of a BGC. BLASTp, PHYRE2 and InterPro platforms were used to annotate and predict the function of all proteins encoded by ‘Cluster A’ from *H. monticulosa* MUCL 54604 (Table 4.3). In addition to the expected hrPKS (*hmPKS1*), the genes of the BGC were found to encode: four SDR (short-chain dehydrogenase/reductase) *hmCAL6*, *hmCAR2*, *hmCAR5* and *hmCAR8*; a glucosyltransferase (*hmCAL5*); two putative hydrolases (*hmCAL2* and *hmCAR4*); a transcriptional regulator (*hmCAL1*); and two transporters (*hmCAR1* and *hmCAR6*). In addition, the cluster encodes six proteins of unknown function (*hmCAL8*, *hmCAL7*, *hmCAL4*, *hmCAL3*, *hmCAR3* and *hmCAR7*).

By analysing the transcriptome data, the gene’s borders, as well as intron and exon positions from ‘Cluster A’, were corrected and putative gene functions were re-annotated (Table 7.2, Table 4.3). Additionally, each gene’s expression level under trienylfuranol A **2** non-producing and producing conditions were assessed. Expression level analysis showed that genes from *hmCAL3* to *hmCAR5* exhibit a strong upregulation under producing conditions except for *hmCAL2* and *hmCAR3* (Figure 4.15), while genes outside this region show either down-regulation or low expression in transcription level. Based on the analysis, the ‘Cluster A’ boundary was initially set from *hmCAL3* to *hmCAR5*. Genes highlighted in the grey boxes in Table 4.3 are therefore probably not involved in the gene cluster. In addition, similar to the expression level of *hmPKS1*, the expression level of *hmCAR5* (SDR) also shows a big NEL value as well as log<sub>2</sub>-fold change (Table 4.3, Figure 4.15).

Gene (locus_tag)	Gene	AA	Putative Function	BLASTp <sup>a</sup> , PHYRE2 <sup>b</sup> , InterPro <sup>c</sup>	Predicted Cofactor	NEL A (non-producing)	NEL B (producing)	Log <sub>2</sub> -fold change (B/A)
HmM_Brakerg10432	hmCAL8	342	Unknown	RNA polymerase subunit <sup>b, c</sup>	/	60.11	49.27	-0.29
HmM_Brakerg10433	hmCAL7	90	Unknown	Cytochrome subunit <sup>b, c</sup>	/	26.95	19.61	-0.46
HmM_Brakerg10434	hmCAL6	514	SDR	Dihydroorotate dehydrogenase <sup>a</sup> , FMN-linked oxidoreductase <sup>b</sup>	FMN	139.92	115.78	-0.27
HmM_Brakerg10435	hmCAL5	646	Glucosyltransferase	Glucosyltransferase <sup>a</sup> , transferase <sup>b</sup>	/	3.11	19.26	2.63
HmM_Brakerg10436	hmCAL4	152	Unknown	Unknown <sup>a, b, c</sup>	/	72.55	56.63	-0.36
HmM_Brakerg10437	hmCAL3	156	Unknown	Allergen homolog, <sup>a</sup> protease inhibitor <sup>b</sup>	/	0.1	367.75	11.84
HmM_Brakerg10438	hmCAL2	842	Putative hydrolase	Glycosidase, <sup>a</sup> hydrolase <sup>b</sup>	/	53.89	35.29	-0.61
HmM_Brakerg10439	hmCAL1	638	Transcriptional regulator	Transcription activator DEP6, <sup>a</sup> transcription activator <sup>b</sup>	/	7.25	156.53	4.43
HmM_Brakerg10440	hmPKS1	2375	Polyketide synthase	Polyketide synthase <sup>a, b, c</sup>	/	0.1	1488.35	13.86
HmM_Brakerg10441	hmCAR1	513	Transporter	Transporter <sup>a, b</sup>	/	0.1	6.26	5.97
HmM_Brakerg10442	hmCAR2	273	SDR	Short-chain dehydrogenase, <sup>a</sup> oxidoreductase <sup>b</sup>	NAD(P)	3.11	110.09	5.15
HmM_Brakerg10443	hmCAR3	267	Unknown	Unknown <sup>a, b, c</sup>	/	6.22	3.32	-0.90
HmM_Brakerg10444	hmCAR4	734	Putative hydrolase	Unknown, <sup>a, c</sup> hydrolase <sup>b</sup>	/	61.15	245.20	2.00
HmM_Brakerg10445	hmCAR5	347	SDR	Dehydrogenase, <sup>a</sup> oxidoreductase <sup>b</sup>	NAD(P)	4.15	1419.11	8.42
HmM_Brakerg10446	hmCAR6	585	Transporter	Transporter <sup>a, b</sup>	/	43.53	33.28	-0.39
HmM_Brakerg10447	hmCAR7	1872	Unknown	Vacuolar membrane, <sup>a</sup> signaling protein <sup>b</sup>	/	13.47	21.40	0.67
HmM_Brakerg10448	hmCAR8	252	SDR	Short-chain dehydrogenase, <sup>a</sup> oxidoreductase <sup>b</sup>	NAD(P)	-	-	-

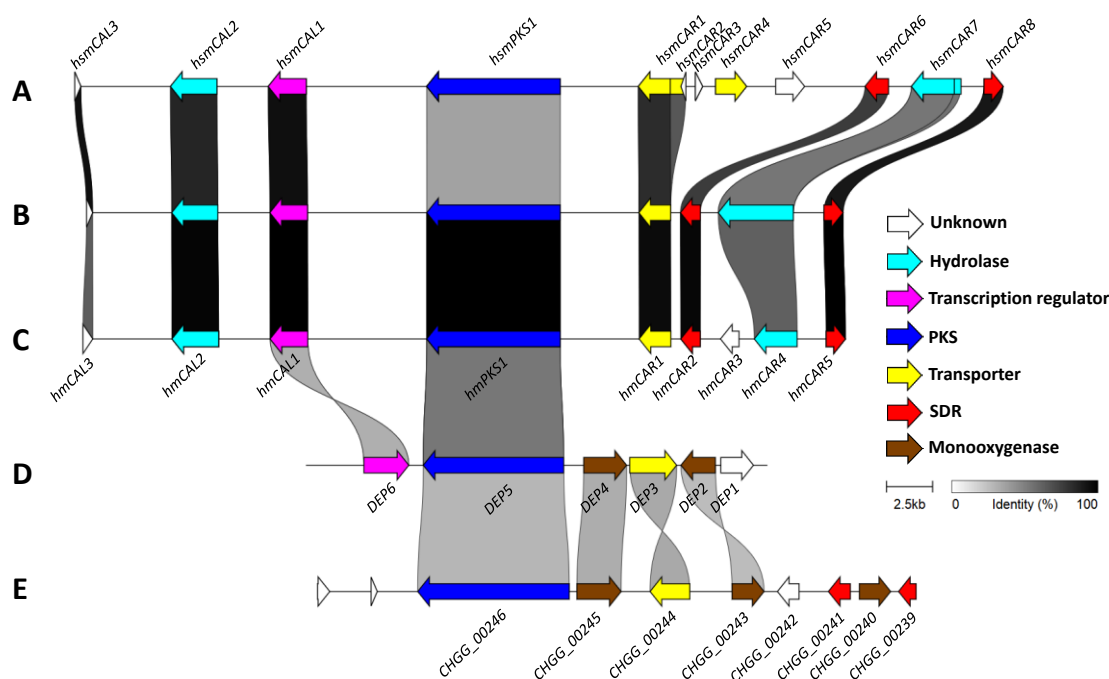
**Table 4.3** Annotation of potential polyene hrPKS BGC ‘Cluster A’ from *H. monticulosa* MUCL 54604. Normalized expression level (NEL, BaseMean) for genes in ‘Cluster A’, data calculated with DESeq. Note: The Gene locus tag shown here is varied from Table 4.2, ‘HmMg6165’ in Table 4.2 correspond to ‘HmM\_Brakerg10440’ here.



**Figure 4.15** Transcriptome data of the potential polyene hrPKS BGC ‘Cluster A’ from *H. monticulosa* MUCL 54604. Bar chart of Log<sub>2</sub>-fold change of the genes from ‘Cluster A’. NE, no expression.

‘Cluster A’ is also found in the *H. spongiphila* and *H. submonticulosa* genomes, and we performed a clinker & clustermap<sup>64</sup> comparison among the three clusters as well as with the depudecin **126** BGC and with the aureonitol **127** BGC. From the results (Figure 4.16), the ‘Cluster A’ in three organisms show high similarity levels except for the insertion of four additional genes *hsmCAR2*, *hsmCAR3*, *hsmCAR4* and *hsmCAR5* (Figure 4.16A) in *H. submonticulosa*. The gene annotation through BLASTp, PHYRE2 and InterPro showed that *hsmCAR4* encodes a transporter, but the other three are function unknown (Table 4.4). In addition, depudecin **126** BGC and the

aureonitol **127** BGC show homologies at the PKS, transporter, transcriptional regulator and monooxygenase (Figure 4.16D – 4.16E). It is noteworthy that there is no oxygenase encoded by ‘Cluster A’, however oxidation is regarded as an indispensable step for the formation of trienylfuranol A **2** which based on the  $^{13}\text{C}$  labelling study (Section 4.3.2).



**Figure 4.16** Clinker & clustermap comparison of potential polyene BGC ‘Cluster A’ and other polyenes BGC. **A**, ‘Cluster A’ (*hsmCA*) in *H. submonticulosa*; **B**, ‘Cluster A’ (*hspCA*) in *H. spongiphila*; **C**, ‘Cluster A’ (*hmCA*) in *H. monticulosa*; **D**, the depudecin<sup>86</sup> BGC in *Alternaria brassicicola*; **E**, the aureonitol<sup>91</sup> BGC in *Chaetomium globosum*.

Gene (locus_tag)	Gene	AA	Putative Function	BLASTp <sup>a</sup> , PHYRE2 <sup>b</sup> , InterPro <sup>c</sup>	Predicted Cofactor
HsMg_8438_t	<i>hsmCAR2</i>	63	Unknown	Unknown <sup>a, b, c</sup>	/
HsMg9082	<i>hsmCAR3</i>	132	Unknown	Unknown <sup>a, b, c</sup>	/
HsMg9083	<i>hsmCAR4</i>	561	Transporter	Transporter, <sup>a, b</sup> major facilitator superfamily <sup>c</sup>	/
HsMg_8441_t	<i>hsmCAR5</i>	382	Unknown	Unknown <sup>a, b, c</sup>	/

**Table 4.4** The function annotation of the *hsmCAR2*, *hsmCAR3*, *hsmCAR4* and *hsmCAR5* from *H. submonticulosa* ‘Cluster A’ (*hsmCA*).

#### 4.3.4.2 ‘Cluster B’ Analysis

Based on the online platform of BLASTp, PHYRE2 and InterPro analysis, the hrPKS BGC ‘Cluster B’ (Table 4.5) was found to encode: an epimerase (*hmCBL5*); two transcriptional regulators (*hmCBL4* and *hmCBR4*); three SDR (*hmCBL3*, *hmCBR5* and *hmCBR7*); a putative hydrolase (*hmCBL1*); two PKS (*hmPKS2* and *hmPKS3*); an *O*-methyltransferase (*hmCBR3*); an *O*-acetyltransferase (*hmCBR1*); two P450 monooxygenases (*hmCBR2* and *hmCBR6*); and four



function unknown genes (*hmCBL6*, *hmCBL2*, *hmCBR8* and *hmCBR9*). Transcriptome data was used to correct each gene's border, as well as intron and exon positions (Table 7.3).

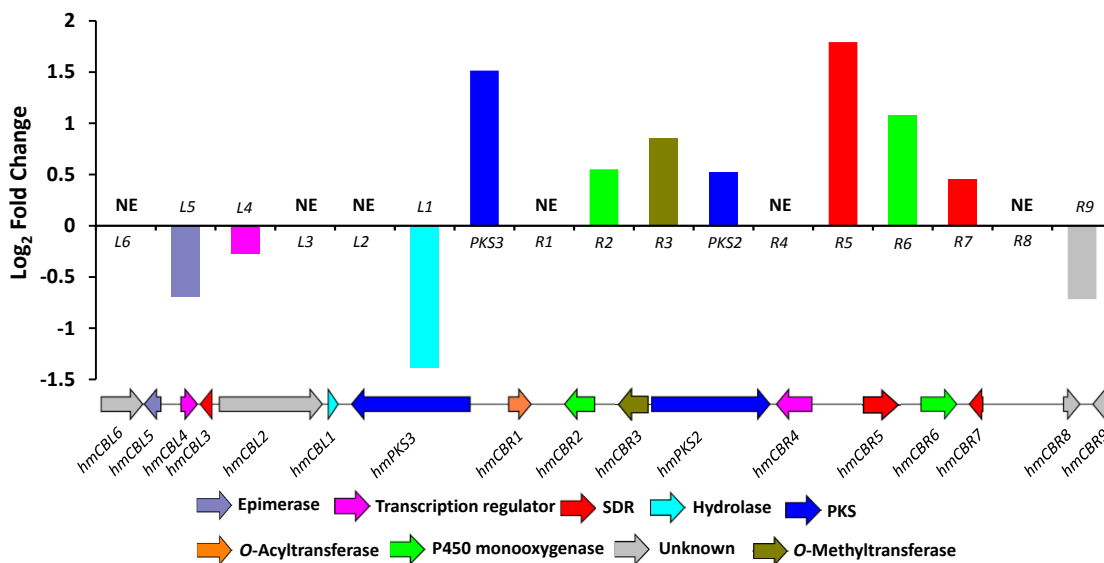
Gene (locus_tag)	Gene	AA	Putative Function	BLASTp, <sup>a</sup> PHYRE2 <sup>b</sup>	Predicted Cofactor	NEL A (non-producing)	NEL B (producing)	Log <sub>2</sub> -fold change (B/A)
HmM_Brakerg3645	<i>hmCBL6</i>	906	Unknown	DNA ligase, <sup>a</sup> ligase <sup>b</sup>	/	-	-	-
HmM_Brakerg3646	<i>hmCBL5</i>	343	Epimerase	Glucose epimerase, <sup>a</sup> isomerase <sup>b</sup>	/	462.24	286.16	-0.69
HmM_Brakerg3647	<i>hmCBL4</i>	364	Transcriptional regulator	Unknown, <sup>a</sup> transcription regulator <sup>b</sup>	/	32.13	26.61	-0.27
HmM_Brakerg3648	<i>hmCBL3</i>	186	SDR	Cytochrome oxidase assembly protein, <sup>a</sup> oxidoreductase <sup>b</sup>	/	-	-	-
HmM_Brakerg3649	<i>hmCBL2</i>	2248	Unknown	DNA polymerase <sup>a, b</sup>	/	-	-	-
HmM_Brakerg3650	<i>hmCBL1</i>	163	Putative hydrolase	Unknown, <sup>a</sup> hydrolase <sup>b</sup>	/	1.04	0.40	-1.39
HmM_Brakerg3651	<i>hmPKS3</i>	2504	HRPKS	Highly reducing PKS <sup>a</sup>	/	101.57	290.21	1.51
HmM_Brakerg3652	<i>hmCBR1</i>	487	O-acetyltransferase	O-acetyltransferase <sup>a, b</sup>	/	-	-	-
HmM_Brakerg3653	<i>hmCBR2</i>	537	P450	Cytochrome P450 monooxygenase <sup>a, b</sup>	/	100.53	147.30	0.55
HmM_Brakerg3654	<i>hmCBR3</i>	427	O-methyltransferase	O-methyltransferase, <sup>a</sup> transferase <sup>b</sup>	/	169.97	307.35	0.85
HmM_Brakerg3655	<i>hmPKS2</i>	2591	HRPKS	Solanapyrone / Prosolanapyrone PKS <sup>a</sup>	/	60.11	86.35	0.52
HmM_Brakerg3656	<i>hmCBR4</i>	654	Transcriptional regulator	Solanapyrone Sol4, <sup>a</sup> transcription activator <sup>b</sup>	/	-	-	-
HmM_Brakerg3657	<i>hmCBR5</i>	526	SDR	Oxidoreductase <sup>a, b</sup>	FAD	203.14	703.65	1.79
HmM_Brakerg3658	<i>hmCBR6</i>	680	P450	Cytochrome P450 monooxygenase <sup>a, b</sup>	/	23.84	50.29	1.08
HmM_Brakerg3659	<i>hmCBR7</i>	274	SDR	Oxidoreductase <sup>a, b</sup>	/	386.59	529.38	0.45
HmM_Brakerg3660	<i>hmCBR8</i>	279	Unknown	Unknown <sup>a, b</sup>	/	-	-	-
HmM_Brakerg3661	<i>hmCBR9</i>	333	Unknown	Unknown, <sup>a</sup> membrane protein <sup>b</sup>	/	9.33	5.68	-0.72

**Table 4.5** Annotation of potential polyene BGC 'Cluster B' from *H. monticulosa* MUCL 54604. Normalized expression level (NEL, BaseMean) for genes in 'Cluster B', data calculated with DESeq. Note: The Gene locus tag shown here is varied from Table 4.2, 'HmMg10159' and 'HmMg10163' in Table 4.2 correspond to 'HmM\_Brakerg3651' and 'HmM\_Brakerg3655' here, respectively.

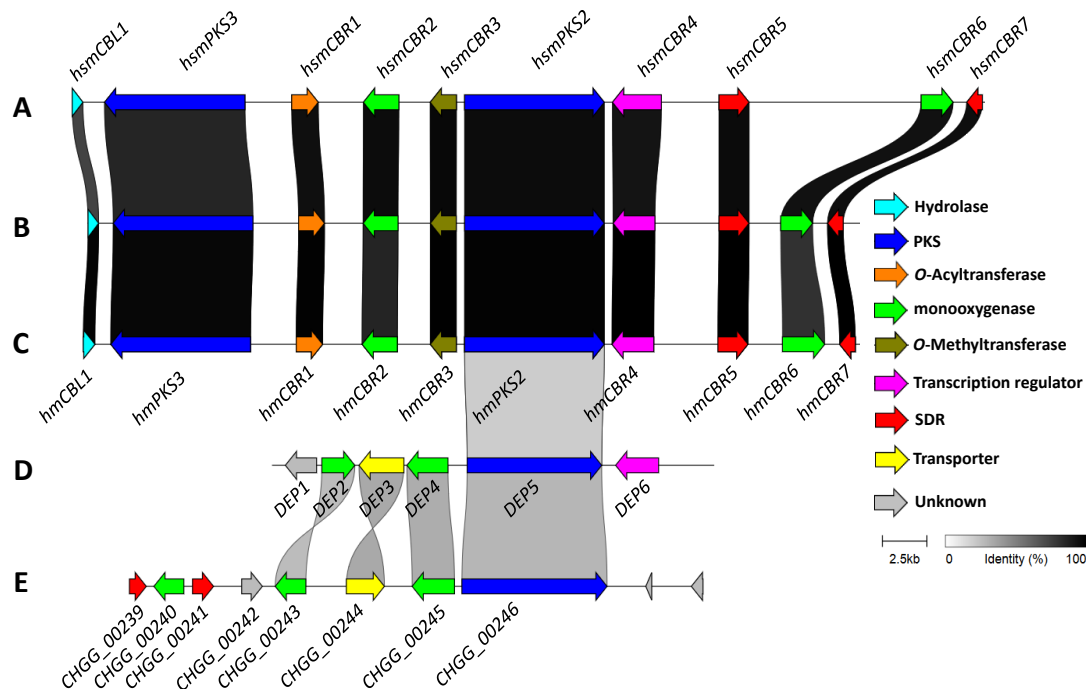
To understand the expression level of functional genes in 'Cluster B' and set the BGC boundaries, we analyzed the transcriptome data from *H. monticulosa*. Results showed that genes from *hmPKS3* to *hmCBR7* are upregulated under producing conditions except for *hmCBR1* and *hmCBR4* (Table 4.5, Figure 4.17). In contrast, genes outside this region showed either down-regulation or no expression. Genes highlighted in the grey boxes in Table 4.5 are therefore probably not involved in 'Cluster B'

A clinker & clustermap comparison of 'Cluster B' from three *Hypomontagnella* organisms was performed (Figure 4.18). Results showed that these three BGC are highly conserved. In addition, the depudecin **126** BGC and aureonitol **127** BGC were compared with 'Cluster B' (Figure 4.18D – 4.18E), but the homologies only display at the hrPKS (HmPKS2, DEP5 and CHGG\_00246). Although there are monooxygenases (HmCBR2 and HmCBR6) encoded by 'Cluster B', no significant similarity is observed with the oxygenases from the depudecin **126** BGC or aureonitol **127** BGC (Figure 4.18).

Notably, 'Cluster B' contains two hrPKSs (HmPKS2 and HmPKS3; Table 4.5), however just one hrPKS (DEP5 and CHGG\_00246, respectively) is used during the backbone construction of depudecin **126** and aureonitol **127**.



**Figure 4.17** Transcriptome data of the potential polyene BGC ‘Cluster B’ from *H. monticulosa* MUCL 54604. Bar chart of Log<sub>2</sub>-fold changes of genes from ‘Cluster B’. NE, no expression.



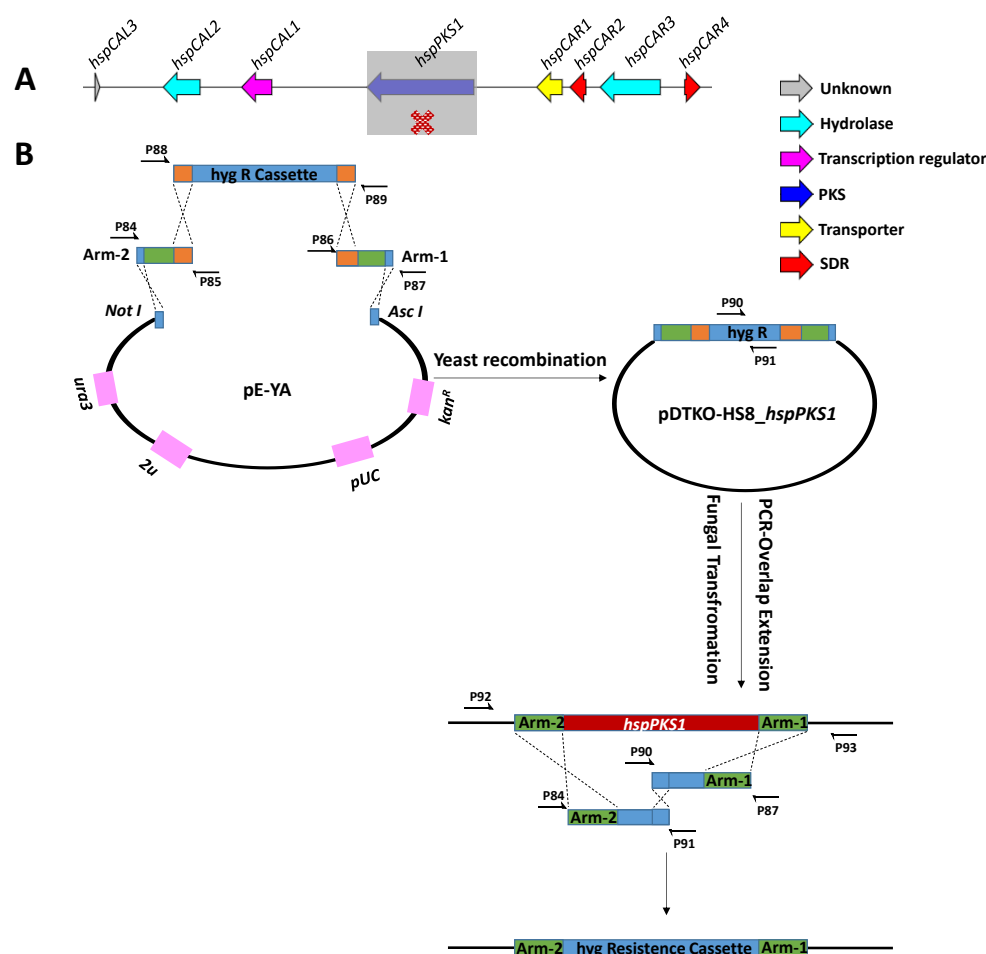
**Figure 4.18** Clinker & clustermap comparison of potential polyene BGC ‘Cluster B’. **A**, ‘Cluster B’ (*hsmCB*) in *H. submonticulosa*; **B**, ‘Cluster B’ (*hspCB*) in *H. spongiphila*; **C**, ‘Cluster B’ (*hmcCB*) in *H. monticulosa*; **D**, the depudecin<sup>86</sup> BGC in *Alternaria brassicicola*; **E**, the aureonitol<sup>91</sup> BGC in *Chaetomium globosum*.

### 4.3.5 Gene Knockout and Heterologous Expression of ‘Cluster A’

#### 4.3.5.1 Gene Knockout of *hspPKS1*

In order to verify whether ‘Cluster A’ is the correct BGC for trienylfuranol A **2** or not, we designed the KO experiment. Similar to the *spoE/spofasA* knockout performed in *H. spongiphila* (Section 2.3.5), the KO vector targeting *hspPKS1* in ‘Cluster A’ from *H. spongiphila* was

constructed as shown in Figure 4.19. The vector construction details are the same as those previously described in Section 2.3.5.2a and the primers are listed in Table 6.2. Two overlapping fragments were amplified and co-transformed into *H. spongiphila* protoplasts. For the fungal transformation workflow details, see Section 2.3.5.2b. The vector construction, fungal transformation as well as transformant analysis were conducted by the project collaborator Dr. Tian Cheng. Before this thesis submission, the transformant analysis work is still ongoing. However, our previous experiment (Section 2.3.5) taught us that KO experiments are rarely successful in this organism.



**Figure 4.19** **A**, potential polyene BGC 'Cluster A' in *H. spongiphila*; **B**, the bipartite knockout method to replace the targeted genes with a hygromycin resistance cassette *via* homologous recombination. Red: Target gene sequence (*hspPKS1* from *H. spongiphila*); Blue: Hygromycin resistance cassette, containing the  $P_{gdpA}$  promoter and the hygromycin resistance gene (*hph*).

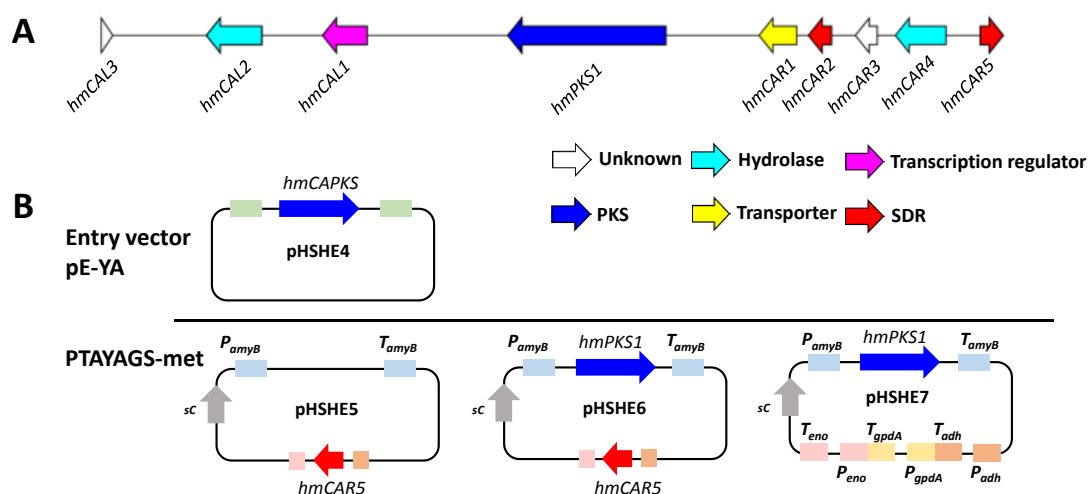
#### 4.3.5.2 Heterologous Expression of *hmPKS1* and *hmCAR5* in *A. oryzae*

The heterologous expression (Section 1.4.4) strategy plays an important role in studying the biosynthesis of natural products, especially the role in the investigation of specific enzyme functions. The elucidation of the sporothriolide **1** biosynthesis pathway is a good example (Section 2.3.6).

Based on the transcriptome analysis, ‘Cluster A’ from *H. monticulosa* MUCL 54604 (Table 4.3) encodes two putative hydrolases (*hmCAL2* and *hmCAR4*), a transcriptional regulator (*hmCAL1*), a hrPKS (*hmPKS1*), a transporter (*hmCAR1*), two SDR (*hmCAR2* and *hmCAR5*) and two proteins of unknown function (*hmCAL3* and *hmCAR3*). Because the PKS (*hmPKS1*) and the SDR (*hmCAR5*) are the two most strongly expressed genes from ‘Cluster A’ (Figure 4.15), we firstly selected them for heterologous expression in *A. oryzae* NSAR1.

#### 4.3.5.2a Gene Combinations, Vector Constructions and *A. oryzae* NSAR1 Transformation

Genes of interest (intron-free) were amplified from *H. monticulosa* gDNA using the primer sets listed in Table 4.6 depending on the investigated combination of genes (primer details see Table 6.2). The pTAYAGS-met vector was employed as a basis for plasmid construction (Figure 4.20). For the strategies of gene cloning and vector constructions refer to ‘Section 2.3.6.1 – 2.3.6.2’.



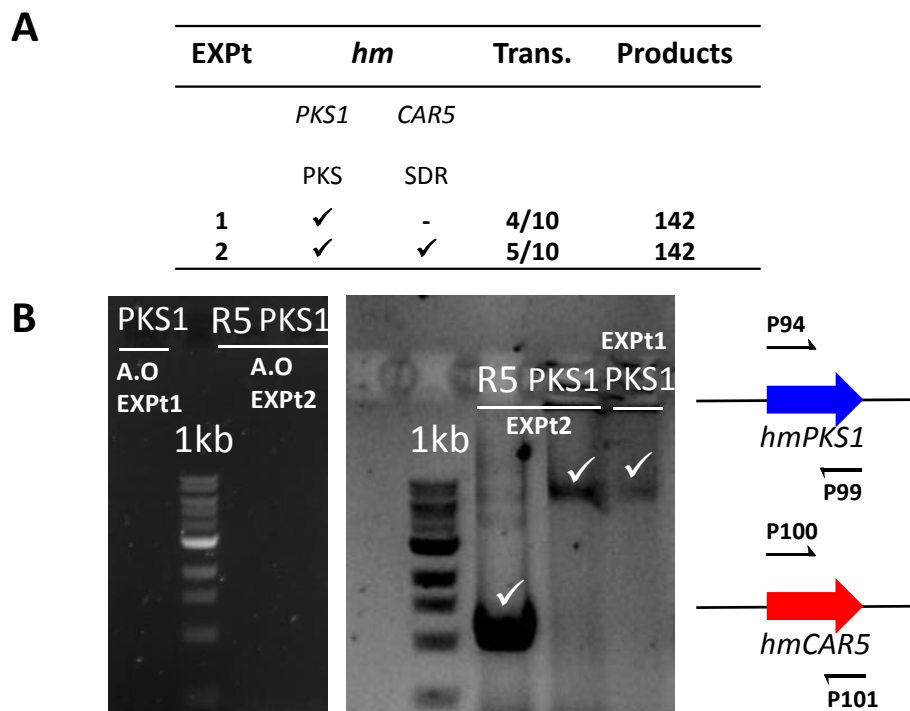
**Figure 4.20** A, hrPKS ‘Cluster A’ (*hmCA*) in *H. monticulosa*; B, constructed plasmids for *A. oryzae* heterologous expression studies.

Primer no.	Template	Target vector	Purpose	PCR condition*
P94+P95	gDNA of <i>H. monticulosa</i>	pHSHE4	<i>hmPKS1</i> , fragment 1	Q5
P96+P97	gDNA of <i>H. monticulosa</i>	pHSHE4	<i>hmPKS1</i> , fragment 2	Q5, 69°C
P98+P99	gDNA of <i>H. monticulosa</i>	pHSHE4	<i>hmPKS1</i> , fragment 2	Q5
P94+P99	<i>E. coli</i> transformants with pHSHE4	pHSHE4	Colony PCR	OneTaq
P100+P101	gDNA of <i>H. monticulosa</i>	pHSHE5	<i>hmCAR5</i>	Q5
P100+P101	<i>E. coli</i> transformants with pHSHE5	pHSHE5	Colony PCR	OneTaq
pHSHE6 (LR clone of pHSHE4 + pHSHE5)				
pHSHE7 (LR clone of pHSHE4 + pTAYAGS-met)				

**Table 4.6** Primer sets used in constructed plasmids for heterologous expression. \* Deviating PCR annealing temperatures (standard is 60 °C) are stated under PCR condition. Primers sequence see Table 6.2.

Two different gene-expression combinations were designed (Figure 4.21), and the corresponding vectors were transformed in *A. oryzae* NSAR1. The obtained transformants (exact number see Figure 4.21A) were genetically checked by PCR. Results showed that the genes were successfully

integrated into the *A. oryzae* positive transformant (Figure 4.21B). All transformants were cultured in DPY medium at 28 °C for 5 – 7 days and 110 rpm, mycelia and culture media were extracted separately and submitted for LCMS analysis.

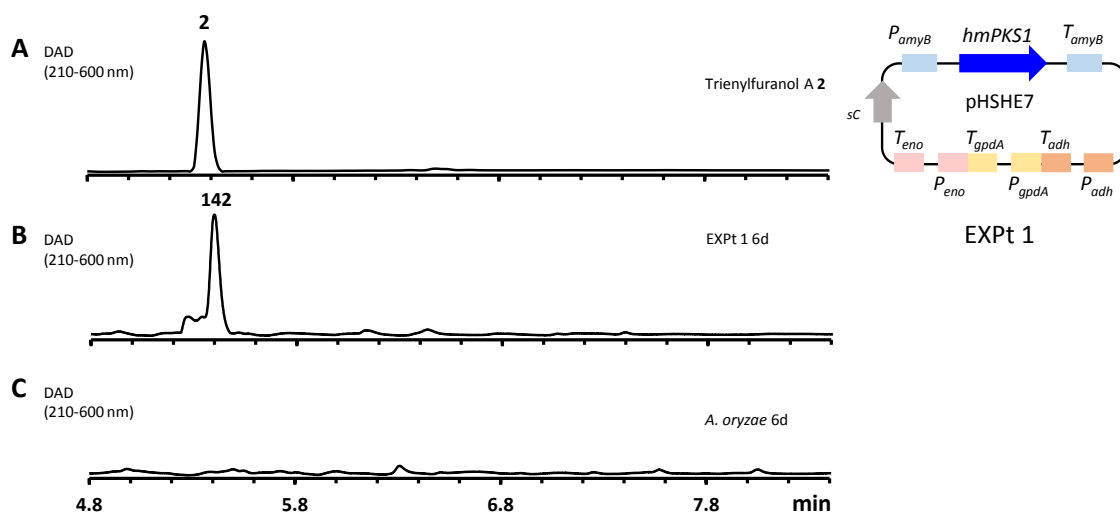


**Figure 4.21** **A**, combinations of genes expressed in *A. oryzae* NSAR1; **B**, genetic analysis of *A. oryzae* NSAR1 transformant: EXPt1 (*hmPKS1*), EXPt2 (*hmPKS1* + *hmCAR5*). Primers sequence see Table 6.2.

#### 4.3.5.2b Expression of *hmPKS1* in *A. oryzae*

The plasmid pHSHE7 containing *hmPKS1* gene was transformed in *A. oryzae*, and the transformants obtained from EXPt1 (Figure 4.21A) were cultured and extracted for LCMS analysis. From the results, a new peak with a molecular weight of 162 (ESI-MS  $m/z$  161 [M – H]<sup>-</sup>, 163 [M + H]<sup>+</sup>) was observed in transformant but not in untransformed *A. oryzae* extracts (EXPt1; Figure 4.22, Table 4.7), and the retention time of this new peak is close to that of trienylfuranol A **2** (Table 7.4). The UV absorption of **142** is characteristic of a tetraene ( $\lambda_{\max}$ : 284 nm, 296 nm, 309 nm) compared with the **2** as a triene ( $\lambda_{\max}$ : 252 nm, 264 nm, 275 nm).

A scale-up fermentation in 1 L DPY liquid medium of the producing transformant was carried out and extracts were subsequently employed for preparative LCMS. The purification of **142** was achieved (2 mg·L<sup>-1</sup>; Table 7.4), and the NMR and HRMS experiments were performed (spectra details see Chapter 7).



**Figure 4.22** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, DAD chromatogram of trienylfuranol A 2; **B**, DAD chromatogram of EXPt1 (*hmpKS1*) transformant; **C**, DAD chromatogram of untransformed *A. oryzae* NSAR1.

Compounds	UV Spectrum	Mass Spectrum (ES <sup>-</sup> )	Mass Spectrum (ES <sup>+</sup> )
<p><b>2</b> Trienylfuranol A C<sub>11</sub>H<sub>16</sub>O<sub>2</sub>; M = 180 g/mol</p>	<p>3: Diode Array 3.514e-1</p>	<p>2: Scan ES<sup>-</sup></p>	<p>1: Scan ES<sup>+</sup></p>
<p><b>142</b> Tetraenylmethoxirane A C<sub>11</sub>H<sub>14</sub>O; M = 162 g/mol</p>	<p>3: Diode Array 2.265</p>	<p>2: Scan ES<sup>-</sup></p>	<p>1: Scan ES<sup>+</sup></p>

**Table 4.7** UV and mass spectra (ES<sup>+</sup> and ES<sup>-</sup>) of compounds **2** and **142**.

Compound **142** was isolated as a yellow oil with the molecular formula C<sub>11</sub>H<sub>14</sub>O (calc. [M + H]<sup>+</sup> HRMS 163.1123, measured 163.1123), indicating 5 degrees of unsaturation. Extensive analysis by 1D and 2D NMR (Table 4.8, Figure 4.23) of **142** showed that the structure belongs to the polyene type. From the <sup>13</sup>C NMR and <sup>13</sup>C (DEPT 135) NMR data analysis, seven olefinic methines and one olefinic methylene with carbon chemical shifts from δ<sub>C</sub> 117.5 to 137.5 ppm were observed. The <sup>1</sup>H-<sup>1</sup>H COSY correlations of H-2 and H-3, H-3 and H-4, H-4 and H-5, H-5 and H-6, H-6 and H-7, H-7 and H-8, as well as correlations of H-8 and H-9 suggested a polyene backbone of **142**. In addition, the intensive HMBC correlations from H-2 to C-3 and C-4, from H-3 to C-4, from H-4 to C-3 and C-5, from H-5 to C-4, C-6 and C-7, from H-6 to C-5 and C-7, from H-7 to C-5 and C-6, from H-8 to C-6, C-7 and C-10, as well as correlations from H-9 to C-7 and C-10 further afforded the tetra-ene structures. Additionally, the C-10 and C-11 have chemical shifts of δ<sub>C</sub> 70.4 ppm and 76.3 ppm respectively, which indicate connection of oxygen atoms. The <sup>1</sup>H-<sup>1</sup>H COSY correlations of H-9 to H-10, H-10 to H-11 and H-9, combined with the

HRMS data further proved that an epoxide group is located at the C-10/C-11 position. C-12 is a terminal methyl group that was assigned by the HMBC correlations of H-12 to C-10 and C-11, as well as  $^1\text{H}$ - $^1\text{H}$  COSY correlations of H-11 and H-12. Therefore, the planar structure of **142** was established (Figure 4.23).

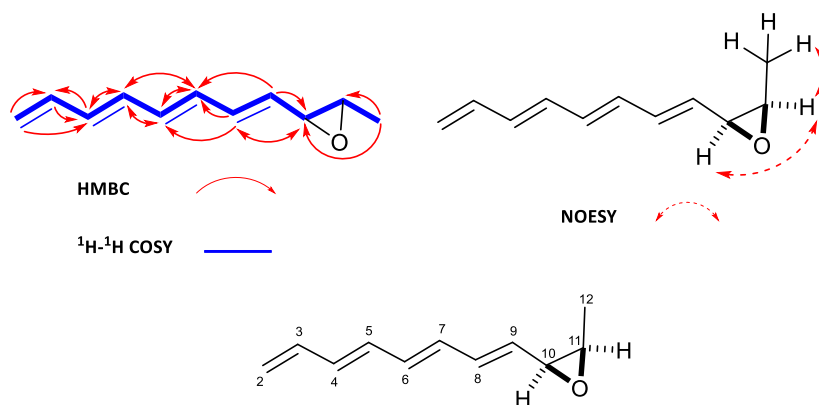
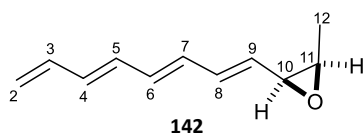


Figure 4.23 Key HMBC,  $^1\text{H}$ - $^1\text{H}$  COSY and NOESY correlations of **142**.

The protons of H-8 and H-9 are configured *anti* which is supported by the coupling constant (15.4 Hz). However, other alkene configurations were not distinguishable because of signal overlap. Based on the geometry of trienylfuranol A **2**, we assumed all of the other double bonds possess *E* configuration. Furthermore, the NOESY correlations of H-10 and H-11 suggests these two protons are located on the same face (Figure 4.23). Thus, the relative configuration was solved. The **142** has not been reported before our work, and we designated **142** as tetraenylmethoxriane A.

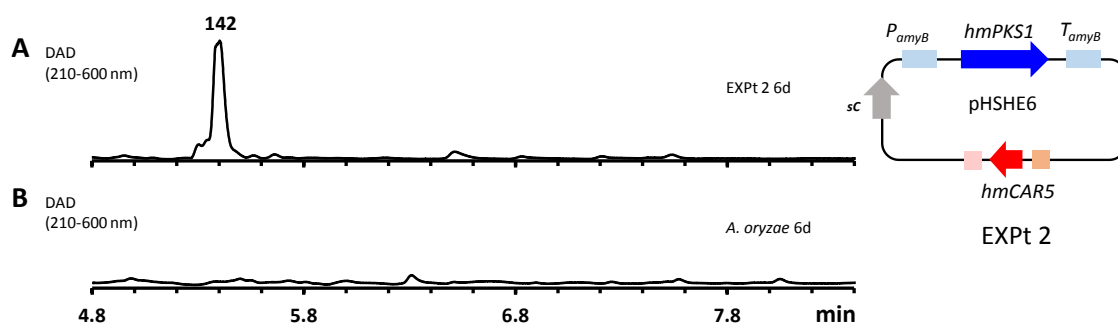


pos.	<b>142</b>		
	$\delta_{\text{C}} / \text{ppm}$	$\delta_{\text{H}} / \text{ppm} (J / \text{Hz})$	
<b>2</b>	117.5	5.13, 1H, dd (16.9, 1.5) 5.00, 1H, dd (10.1, 1.5)	3, 4
<b>3</b>	137.5	6.31, 1H, m	4
<b>4</b>	134.1	6.14, 1H, m	3, 5
<b>5</b>	133.2	6.11, 1H, m	4, 6, 7
<b>6</b>	133.6	6.14, 1H, m	5, 7
<b>7</b>	133.2	6.11, 1H, m	5, 6
<b>8</b>	132.6	6.20, 1H, ddd (15.4, 10.1, 1.3)	6, 7, 10
<b>9</b>	132.6	5.56, 1H, dd (15.4, 6.7)	7, 10
<b>10</b>	76.3	3.81, 1H, ddd (6.7, 3.9, 1.2)	8
<b>11</b>	70.4	3.58, 1H, qd (6.4, 3.9)	-
<b>12</b>	17.7	0.99, 3H, d (6.4)	10, 11

Table 4.8  $^1\text{H}$  NMR (600 MHz) data and  $^{13}\text{C}$  NMR (150 MHz) data for **142** in  $\text{C}_6\text{D}_6$ .

### 4.3.5.2c Co-expression of *hmPKS1* with *hmCAR5*

To test the function of the highly expressed SDR (*hmCAR5*) in ‘Cluster A’, the plasmid pHSHE6 containing *hmPKS1* and *hmCAR5* was transformed in *A. oryzae* (EXPt2; Figure 4.21). The obtained transformants were cultivated and extracted for LCMS analysis. However, except **142**, no new peaks were observed in the chromatograms (Figure 4.24).



**Figure 4.24** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, DAD chromatogram of EXPt2 (*hmPKS1* + *hmCAR5*) transformant; **C**, DAD chromatogram of untransformed *A. oryzae* NSAR1.

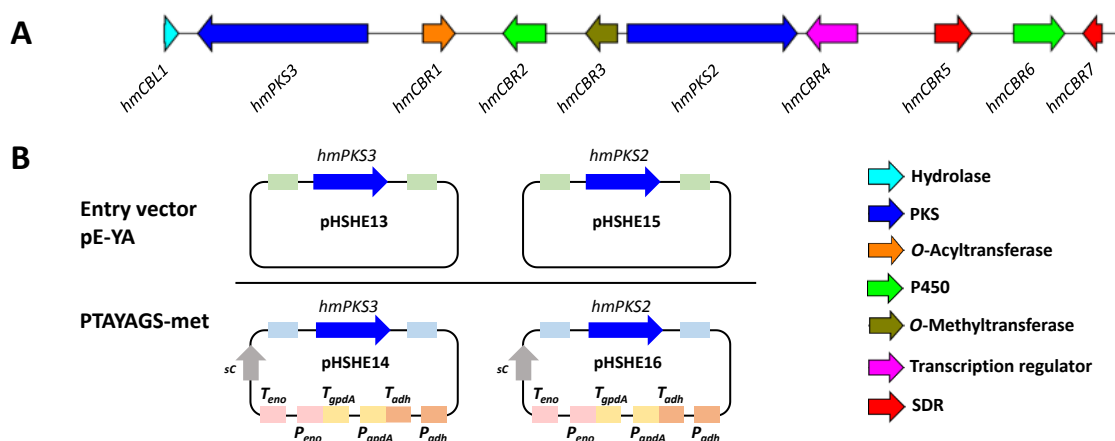
### 4.3.6 Heterologous Expression of ‘Cluster B’

‘Cluster B’ from *H. monticulosa* encodes two PKS (*hmPKS2* and *hmPKS3*; Table 4.2 and Table 4.5) and also contains several tailoring genes such as P450 monooxygenase (*hmCBR2* and *hmCBR6*), *O*-methyltransferase (*hmCBR3*) and *O*-acetyltransferase (*hmCBR1*). It would be interesting to investigate which natural products are produced by heterologous expressing ‘Cluster B’ in *A. oryzae*.

#### 4.3.6.1 Gene Cloning and Vector Constructions

Two PKS genes (*hmPKS2* and *hmPKS3*) were amplified from the *H. monticulosa* gDNA using the primers skipping over the predicted intron positions (Table 4.9, primers details see Chapter 6, Table 6.2). The pTAYAGS-met vector was employed as a basis for plasmid construction (Figure 4.25), the strategies of gene cloning and vector construction refer to Section 2.3.6.1 – 2.3.6.2.



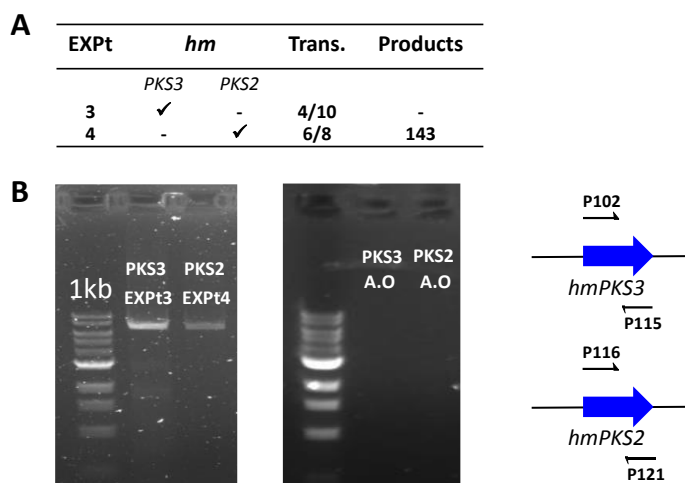


**Figure 4.25** A, the hrPKS 'Cluster B' (*hmCB*) in *H. monticulosa*; B, constructed plasmids for *A. oryzae* heterologous expression studies.

Primer no.	Template	Target vector	Purpose	PCR condition*
P102+P103	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 1	Q5
P104+P105	Order sequence_10159 use	pHSHE13	<i>hmPKS3</i> , fragment 2	Q5
P106+P107	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 3	Q5
P108+P109	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 4	Q5
P110+P111	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 5	Q5
P112+P113	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 6	Q5
P114+P115	gDNA of <i>H. monticulosa</i>	pHSHE13	<i>hmPKS3</i> , fragment 7	Q5
P110+P111	<i>E.coli</i> transformants with pHSHE13	pHSHE13	Colony PCR	Q5
P116+P117	Order sequence_10163 use	pHSHE15	<i>hmPKS2</i> , fragment 1	Q5
P118+P119	gDNA of <i>H. monticulosa</i>	pHSHE15	<i>hmPKS2</i> , fragment 2	Q5
P120+P121	gDNA of <i>H. monticulosa</i>	pHSHE15	<i>hmPKS2</i> , fragment 3	Q5
P116+P117	<i>E.coli</i> transformants with pHSHE15	pHSHE15	Colony PCR	OneTaq
pHSHE14 (LR clone of pHSHE13 + pTAYAGS-met)				
pHSHE16 (LR clone of pHSHE15 + pTAYAGS-met)				

**Table 4.9** Primer sets used in constructed plasmids for heterologous expression. \* Deviating PCR annealing temperatures (standard is 60 °C) are stated under PCR condition. Primers sequence see Table 6.2.

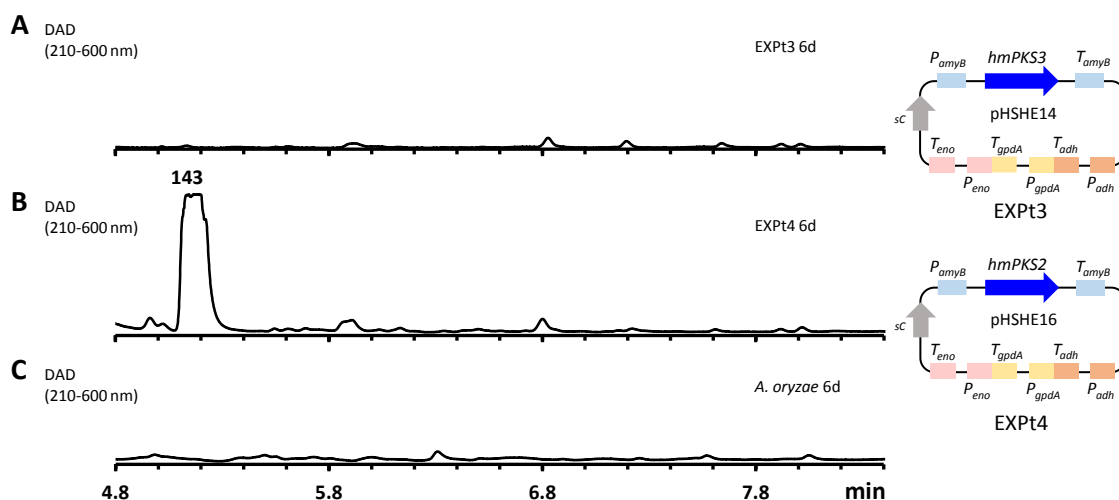
Two different expression vectors were designed (Figure 4.25), each harbouring one PKS gene. The corresponding vectors were transformed in *A. oryzae* NSAR1. The obtained transformants (exact number see Figure 4.26A) were genetically checked by PCR (Figure 4.26B). All transformants were then cultured in DPY medium at 28 °C for 5 – 7 days and 110 rpm, mycelia and culture media were extracted separately and submitted for LCMS analysis.



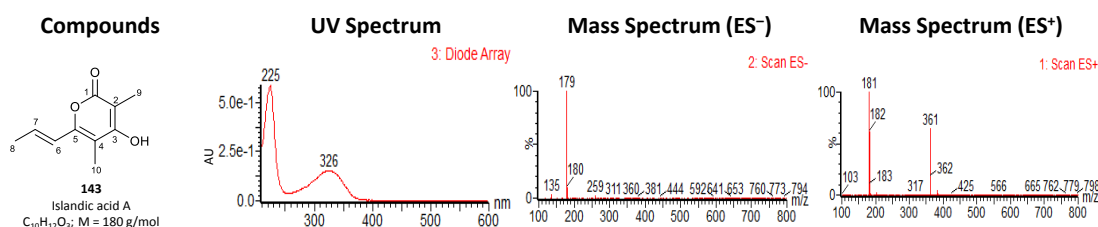
**Figure 4.26** **A**, combinations of genes expressed in *A. oryzae* NSAR1; **B**, genetic analysis of *A. oryzae* NSAR1 transformant: EXPt3 (*hmPKS3*); EXPt4 (*hmPKS2*).

#### 4.3.6.2 Expression of *hmPKS2* and *hmPKS3* in *A. oryzae*

The plasmids pHSHE14 (containing *hmPKS3*; Table 4.2 and Table 4.5) and pHSHE16 (containing *hmPKS2*; Table 4.2 and Table 4.5) were separately transformed in *A. oryzae*. The transformants from EXPt3 (*hmPKS3*; Figure 4.26) and EXPt4 (*hmPKS2*; Figure 4.26) were cultured and extracted for LCMS analysis, respectively. Only in LCMS results from EXPt4 transformant, a new peak **143** with a molecular weight of 180 (ESI-MS  $m/z$  179 [M – H]<sup>-</sup>, 181 [M + H]<sup>+</sup>) was observed (Figure 4.27 – 4.28). Although **143** shares the same molecular weight with trienylfuranol A **2** and their retention time is close to each other (Table 7.4), the UV absorption of **143** ( $\lambda_{\max}$ : 225 nm, 326 nm; Figure 4.28) does not exhibit a character of polyenes compared with **2** and **142** (Table 4.7). The purification of **143** (50 mg·L<sup>-1</sup>; Table 7.4) was performed by preparative LCMS and the structure was elucidated by extensive NMR and HRMS spectra (see Chapter 7 for details).



**Figure 4.27** HPLC analysis of crude extracts (from media) from *A. oryzae* transformant. **A**, DAD chromatogram of EXPt3 (*hmPKS3*) transformant; **B**, DAD chromatogram of EXPt4 (*hmPKS2*) transformant; **C**, DAD chromatogram of untransformed *A. oryzae* NSAR1.



**Figure 4.28** UV and mass spectra ( $ES^+$  and  $ES^-$ ) of compounds **143**.

Compound **143** was isolated as a colourless powder with molecular formula  $C_{10}H_{12}O_3$  (calc.  $[M + H]^+$  HRMS 181.0865, measured 181.0875), indicating 5 degrees of unsaturation. The HSQC correlations (Figure 4.29) assigned all the protons to the corresponding carbons. The comprehensive analysis of the HSQC,  $^1H$  and  $^{13}C$  NMR spectra (Figure 4.29, Table 4.10) suggested the existence of one carbonyl group, six olefinic carbons and two of them are olefinic methines, as well as three methyl groups. The HMBC correlations from H-9 ( $\delta_H$  1.93, s) to C-1 ( $\delta_C$  167.6), C-2 ( $\delta_C$  99.9) and C-3 ( $\delta_C$  167.8), and from H-10 ( $\delta_H$  2.00, s) to C-3 ( $\delta_C$  167.8), C-4 ( $\delta_C$  108.6) and C-5 ( $\delta_C$  153.6), suggested a 2,4-dimethyl-3-hydroxypyranone nucleus. The COSY relationships of H-6 ( $\delta_H$  6.42, dq,  $J = 15.4, 1.7$  Hz) and H-7 ( $\delta_H$  6.61, dq,  $J = 15.4, 6.9$  Hz) and H-8 ( $\delta_H$  1.92, m) provided a propene side chain. The HMBC correlations of the olefinic protons H-6 and H-7 to C-5 indicated the side chain to be linked to the C-5 position. The  $6E$  geometry was identified by the H-6/H-7 coupling (15.4 Hz).

Compound **143** has not been reported before our work. We designated **143** as islandic acid A, because of its structural similarity to the known metabolite islandic acid **161** (Section 4.4.4).

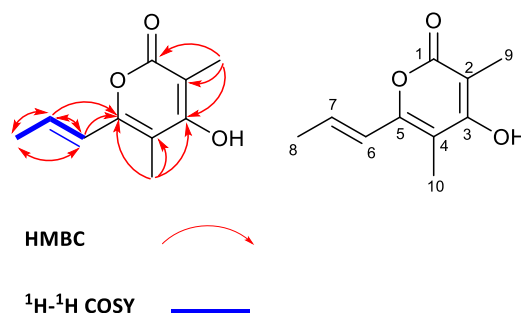
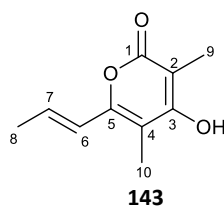


Figure 4.29 Key HMBC and  $^1\text{H}$ - $^1\text{H}$  COSY correlations of **143**.



pos.	<b>143</b>			
	$\delta_{\text{C}}$ / ppm	$\delta_{\text{H}}$ / ppm (J / Hz)	HMBC	$^1\text{H}$ - $^1\text{H}$ COSY
1	167.6	-	9	-
2	99.9	-	9	-
3	167.8	-	9, 10	-
4	108.6	-	10	-
5	153.6	-	6, 7, 10	-
6	121.3	6.42, 1H, dq (15.4, 1.7)	7, 8	7
7	134.3	6.61, 1H, dq (15.4, 6.9)	6, 8	6, 8
8	18.7	1.92, 3H, m	6, 7	7
9	9.0	1.93, 3H, m	-	-
10	9.4	2.00, 3H, s	-	-

Table 4.10  $^1\text{H}$  NMR (400 MHz) data and  $^{13}\text{C}$  NMR (100 MHz) data for **143** in  $\text{CD}_3\text{OD}$ .

## 4.4 Discussion

### 4.4.1 Isotopic Labelling Study

The time course study showed that *H. submonticulosa* can produce *ca.* 700 mg trienylfuranol A **2** in 1 L fermentation media (Table 4.1) under shaking conditions. The production titres of **2** in *H. spongiphila* and *H. monticulosa* are also considerable, which is *ca* 400  $\text{mg}\cdot\text{L}^{-1}$  and 50  $\text{mg}\cdot\text{L}^{-1}$ , respectively.

Results from the  $[1,2\text{-}^{13}\text{C}_2]$  acetate labelling experiment revealed that trienylfuranol A **2** is composed of five intact acetate units and a C-2 atom from a cleaved acetate. And only the peak intensity of C-2 (Figure 4.10) is increased compared to unlabelled **2** from  $[1,2\text{-}^{13}\text{C}_2]$  labelling experiment, indicating the carboxyl group at C-2 is decarboxylated during **2** formation.

In summary, trienylfuranol A **2** is shown to be assembled from at least six intact acetates, but the loss of a C-1 acetate-derived carbon at C-2, and oxygenation at C-8 suggest the involvement of decarboxylation and oxidation during biosynthesis. Interestingly, the  $^{13}\text{C}$  incorporation patterns of the  $\text{C}_{11}$  trienylfuranol A **2** are quite similar to that of the  $\text{C}_{11}$  depudecin **126** (Scheme 4.1) and the  $\text{C}_{13}$  aureonitol **127** (Scheme 4.2), but no carbon-carbon rearrangement was observed as during aureonitol **127** biosynthesis. These three compounds (**126**, **127** and **2**) all belong to polyketide polyene derivatives, and their biosynthesis may share similarities to some extent based on  $^{13}\text{C}$  incorporation patterns.

#### 4.4.2 Genome Mining and Transcriptomic Analysis

By utilizing multiple bioinformatic analysis, two top hits potential polyene hrPKS BGC ‘Cluster A’ and ‘Cluster B’ were identified from the *Hypomontagnella* genomes (Table 4.2). ‘Cluster A’ (Table 4.3) was found to encode some functional enzymes such as hrPKS (PKS1), SDR, glucosyltransferase. Through transcriptomic analysis, the expression level of PKS1 in ‘Cluster A’ reveals a significant upregulation from non-producing to producing conditions (Table 4.2, Figure 4.14). However, the lack of the essential oxygenase makes ‘Cluster A’ a less likely possibility for the biosynthesis of trienylfuranol A **2**.

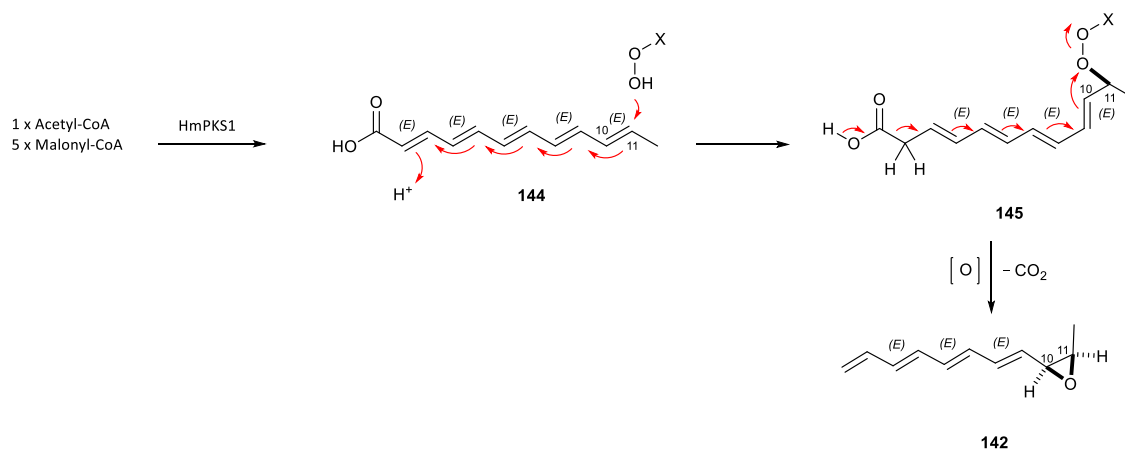
In contrast, there are a few oxygenases encoded by ‘Cluster B’ (Table 4.5), for instance, P450 monooxygenases (HmCBR2 and HmCBR6). But the presence of two PKS (*hmPKS2* and *hmPKS3*; Table 4.2, Table 4.5, and Figure 4.18) in this gene cluster makes ‘Cluster B’ mysterious.

#### 4.4.3 Biosynthetic Study of The hrPKS BGC ‘Cluster A’

To confirm whether polyene hrPKS BGC ‘Cluster A’ is responsible for trienylfuranol A **2** biosynthesis, a specific KO experiment targeting *hspPKS1* is still ongoing. The heterologous expression is also a very useful strategy in the natural products biosynthesis study (Section 1.4.4). When *hmPKS1* alone was expressed (EXPt1; Figure 4.21 – 4.22) in *A. oryzae*, a  $\text{C}_{11}$  polyene **142** (Table 4.7, Figure 4.22 and Figure 4.23) was obtained as an oxidised outcome of the *EEEEEE* polyene hexaketide **144** which derived from the hrPKS assembling pipeline (Scheme 4.6). In this experiment (EXPt1) we expressed PKS1 (Table 4.2) only. Although a polyene product was observed it was epoxidized, indicating the involvement of an *A. oryzae* oxygenase. In addition, the structure of **142** indicates the ER domain and C-MeT domain in PKS1 are inactive, which fits with our previous PKS1 domain analysis results (Table 4.2, Figure 7.86 – 7.87).

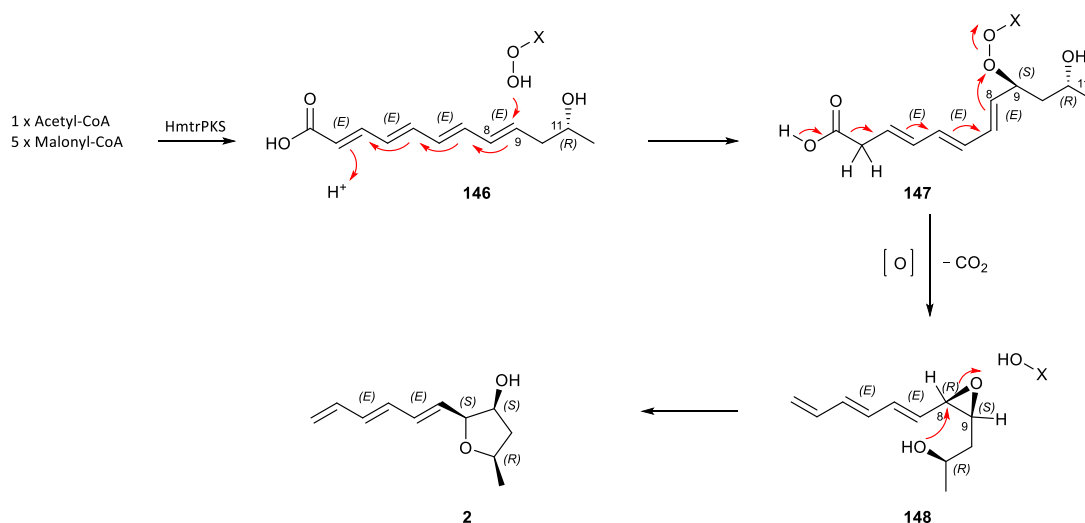
Here, we proposed a unique epoxidation/decarboxylation catalysed by the *A. oryzae* host to explain the conversion of **144** to **142** with the *cis* H-10/H-11 geometry (Scheme 4.6). An FMO or P450 iron peroxide species could execute the key epoxidation/decarboxylation during the

biosynthesis.<sup>110</sup> For **142**, it is supposed that an as-yet unidentified native oxygenase in *A. oryzae* is responsible for the oxidation of the intermediate **144**. There have been observed previous examples of the unexpected oxidations of biosynthetic intermediates by expression host *A. oryzae*.<sup>111</sup>



**Scheme 4.6** The proposed pathway to **142**. X = Flavin species, or Fe-heme.

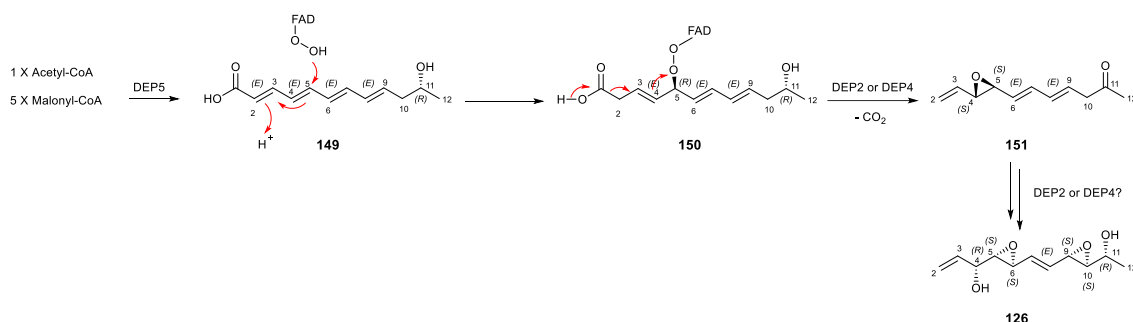
Inspired by the proposed biosynthetic pathway for the C<sub>11</sub> polyene **142**, the biosynthesis of the C<sub>11</sub> trienylfuranol A **2** was also reasonably suggested. The plausible biosynthetic pathway (Scheme 4.7) of trienylfuranol A **2** probably starts with an *EEEE* polyene hexaketide **146** with a hydroxyl group at C-11,<sup>112–117</sup> then following an epoxidation/decarboxylation mechanism reaction which could convert **146** to **147** with the H-8/H-9 in a *cis* geometry, which are similar to **142** biosynthetic steps (Scheme 4.6). Finally, the hydroxyl nucleophilically (S<sub>N</sub>2) may attack the epoxide to form the furan ring of trienylfuranol A **2**. While the epoxidation positions in **142** (C-10 and C-11) are distinct from that of in **2** (C-8 and C-9). Certainly, it's an output from two different enzymes catalysis (Scheme 4.6 – 4.7).



**Scheme 4.7** The proposed biosynthetic pathway of trienylfuranol A **2**. X = Flavin species, or Fe-heme.

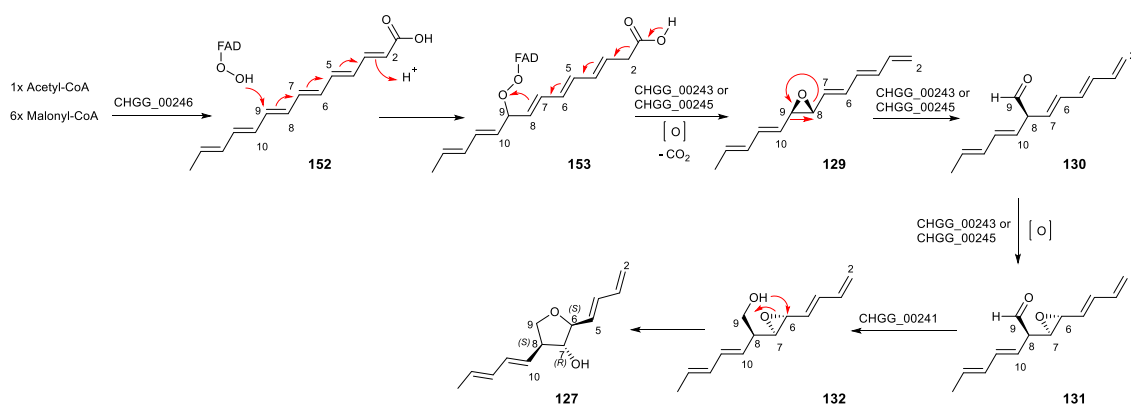
Additionally, because of the similarities between the  $^{13}\text{C}$  labelling patterns of depudecin **126**, aureonitol **127** and trienylfuranol A **2**, we also proposed depudecin **126** and aureonitol **127** biosynthetic pathways based on Scheme 4.6 and Scheme 4.7.

For depudecin **126**, the pathway (Scheme 4.8) could start with a hexaketide **149**. Then epoxidation/decarboxylation could yield the intermediate polyketide polyene **151**, the FMO (DEP2 or DEP4) is assumed to catalyze the reaction. Afterwards, the FMO (DEP2 or DEP4) may further catalyze the intermediate **151** to form the final depudecin **126**, and the shift of double bonds, as well as alkene epoxidation are observed during this process.



**Scheme 4.8** The proposed biosynthesis of depudecin **126**.

Nakazawa and co-workers<sup>91</sup> proposed a biosynthetic pathway (Scheme 4.3) for the  $\text{C}_{13}$  polyketide aureonitol **127**. But the mechanism of how the polyene precursor **129** is formed was not shown in their pathway. Based on the hypothetical biosynthesis of **142**, depudecin **126** and trienylfuranol A **2**, we proposed that the unique epoxidation/decarboxylation process could also be involved during aureonitol **127** biosynthesis, to form the polyketide polyene intermediate **129** as shown in Scheme 4.9. However, no evidence for the identification of such a catalyst yet exists and more work will be required to verify these ideas.

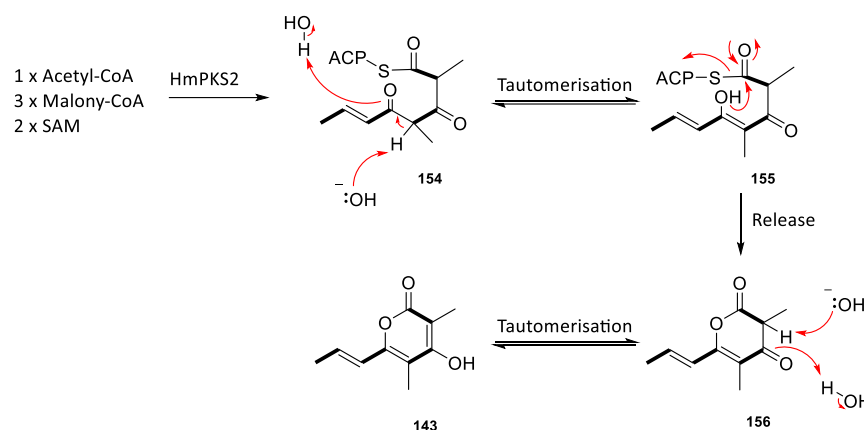


**Scheme 4.9** The proposed biosynthesis of aureonitol **127**.

#### 4.4.4 Biosynthetic Study of The hrPKS BGC ‘Cluster B’

Through the heterologous expression of PKS *hmPKS2* (Figure 4.27 – 4.28) in *A. oryzae* NSAR1, we isolated a new compound, 2-pyrone **143**, from the positive transformant with a production titre of 50 mg·L<sup>-1</sup>.

The biosynthetic pathway (Scheme 4.10) of **143** is likely to start with the tetraketide **154** which is assembled by PKS2 (Table 4.2, Figure 4.27), following a possible tautomerisation that affords **155**. And then it probably undergoes a lactone formation to be released from the PKS, which yields the pyrone **156**. As shown in Table 4.2, PKS2 has no TE domain for polyketide chain release. The final step could be the tautomerisation of **156** to the pyrone **143**.



**Scheme 4.10** The proposed pathway to **143**.

The methylated pyrone scaffold of **143** reveals that PKS2 (Table 4.2) must have a functional C-MeT domain and a defective ER domain. The ER domain in PKS2 was predicted to be inactive during the previous PKS architecture analysis (Table 4.2, Figure 7.86), which fits with the structure of **143**. However, the C-MeT domain was also predicted to be non-functional during the previous sequence alignments (Figure 7.87), because the C-MeT cofactor motif (‘GGGTG’) of PKS2 is not as conserved as the sequence (‘GAGTG’) in the functional C-MeT of TENS (PKS of tenellin),<sup>118</sup> SQTKS (PKS of squalestatin)<sup>119</sup> and CitS (PKS of citrinin).<sup>120</sup> But the fact is, that minor sequence mutation (from ‘GAGTG’ to ‘GGGTG’) of the cofactor motif does not appear to deactivate the C-MeT domain of the PKS in this case.

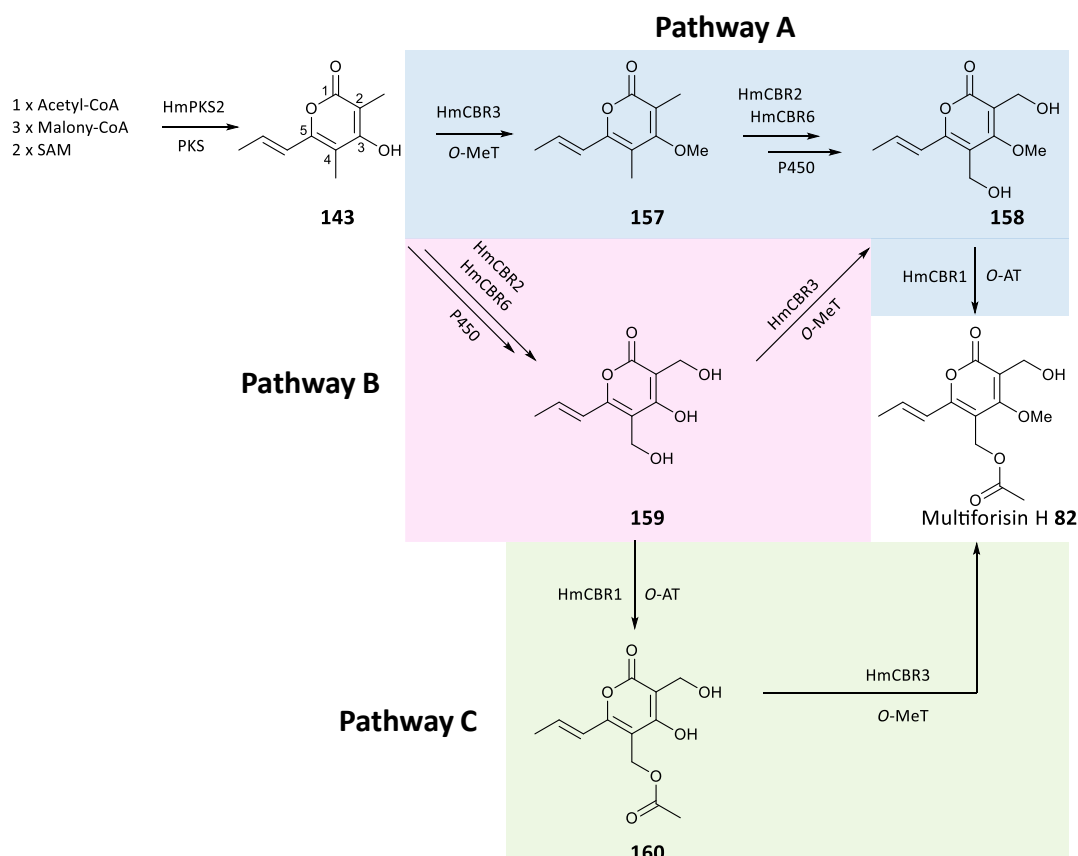
Interestingly, **143** is quite similar to the known compound multiforisin H **82**.<sup>57,121,122</sup> And multiforisin H **82** was also found by us from the *H. spongiphila* (Section 2.3.2), thus it is reasonable to make a biogenetic relationship between **143** and **82**.

A biosynthetic pathway for multiforisin H **82** is proposed in Scheme 4.11. Three chemical reactions are probably needed to convert pyrone **143** to multiforisin H **82**, they are oxidation of the C-2 and C-4 methyl groups, methylation of the C-3 hydroxyl, and acetylation at O-4,



respectively. In total, three pathways were assumed based on the various reaction orders. In ‘pathway A’, the methylation processes could firstly give **157**, then the followed oxidation may yield **158**, and afterwards acetylation could obtain **82**. However, distinct from ‘pathway A’, the order of oxidation and methylation in ‘pathway B’ are inverted. Therefore **159** could be formed firstly through the oxidation of **143**, then methylation and following acetylation may afford **82**. In ‘pathway C’, **159** obtained from the oxidation is likely to serve as the acetylation precursor for **160**. Finally, the hydroxyl methylation of **160** could yield **82** (Scheme 4.11).

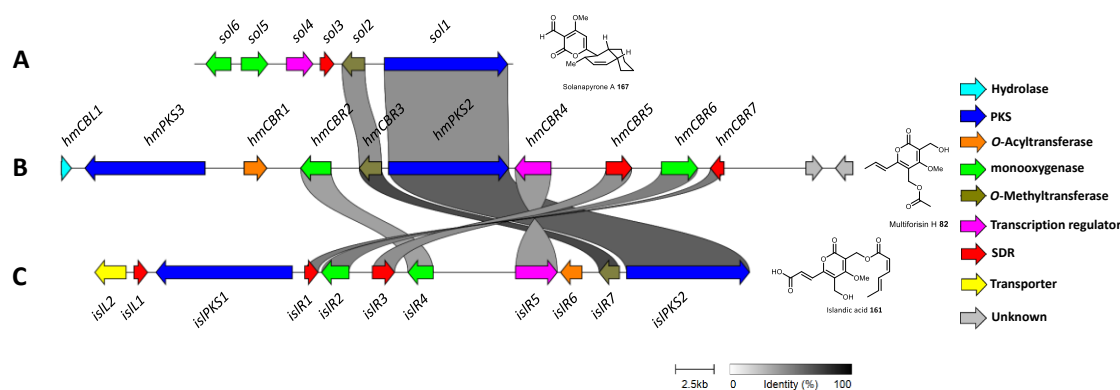
The required enzymes which could catalyse the reactions are all encoded in ‘Cluster B’ (Table 4.5). Therefore, ‘Cluster B’ is probably responsible for multiforisin H **82** biosynthesis. Even though multiforisin compounds have been discovered over 20 years,<sup>57</sup> the biosynthetic pathway and the related gene cluster was still unknown before our finding.



**Scheme 4.11** The proposed biosynthetic pathway to multiforisin H **82**. Blue colour: Pathway A; Pink colour: Pathway B; Green colour: Pathway C.

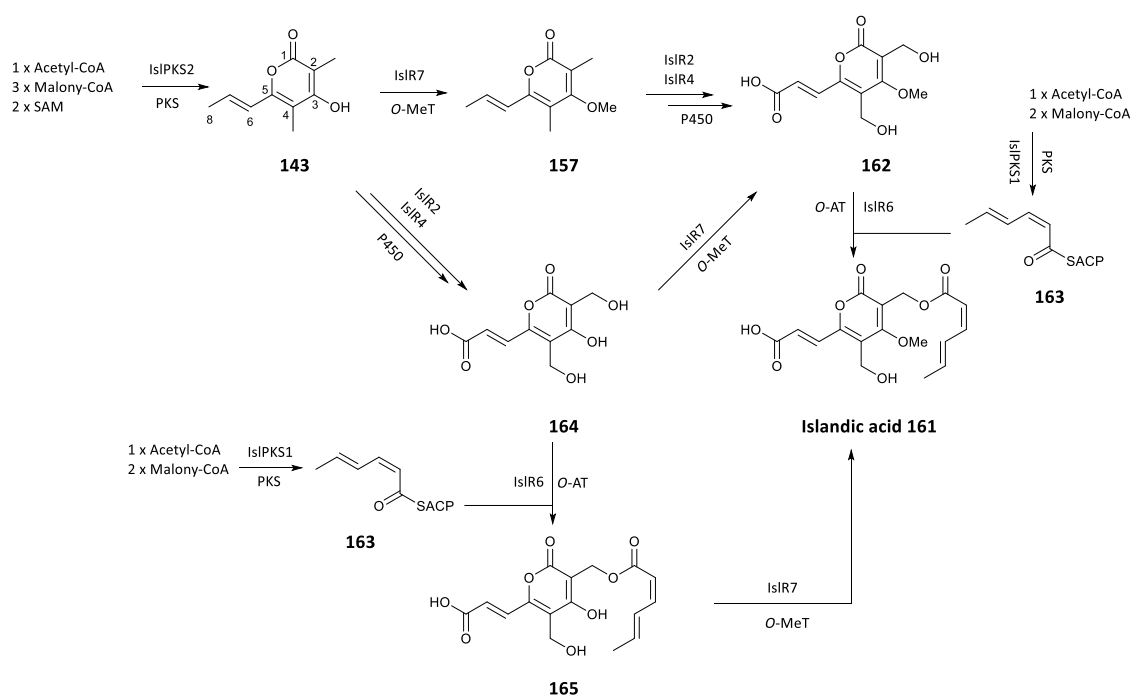
The known compound islandic acid **161** has a similar core scaffold as multiforisin H **82**.<sup>122–124</sup> Islandic acid was the first member of this class of compounds to have been reported.<sup>124</sup> It was first isolated from the fungus *Talaromyces islandicus* in 1982. But no information about islandic acid **161** biosynthesis was known before our work. A putative PKS BGC for islandic acid **161** was found by performing similarity searches of ‘Cluster B’ in the genome of *T. islandicus*.<sup>123,124</sup>

Several proteins encoded by the putative multiforin H **82** BGC ('Cluster B'; Figure 4. 30B) and putative islandic acid **161** BGC (Figure 4. 30C) share homologies through comparisons, such as PKS, P450, SDR and *O*-methyltransferase.



**Figure 4.30** Clinker & clustermap comparison of hrPKS 'Cluster B' (*hmCB*) with other pyrones BGC. **A**, the solanapyrones BGC<sup>82</sup> from *Alternaria solani*; **B**, the hrPKS Cluste B from *H. monticulosa*; **C**, the putative islandic acid **161** BGC from *Talaromyces islandicus*.

Several similarities were observed between the chemical structures of multiforin H **82** and islandic acid **161**, as well as between their putative BGC (Figure 4.30). Therefore, there should also exist several similarities between their biosynthetic pathways. For example, similar oxidation, methylation and acetylation in Scheme 4.11 are probably also involved during islandic acid **161** biosynthesis (Scheme 4.12). In a word, the proposed islandic acid pathway in Scheme 4.12 may closely resemble the proposed biosynthesis of multiforin H **82** (Scheme 4.11).

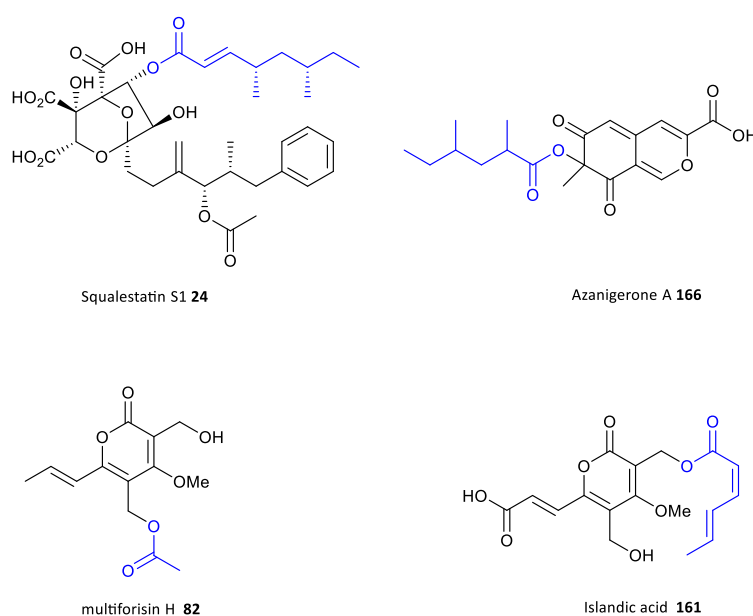


**Scheme 4.12** The proposed biosynthetic pathway for islandic acid **161**.

But it's worth noting that the acetylation position (C-2) of islandic acid **161** is different from that (C-4) of multiforisin H **82** (Scheme 4.11 – 4.12). In addition, the second PKS (IslPKS1) may be responsible for the assembly of triketide polyene **163** with a *Z*-configured double bond, which is similar to strobilurin PKS.<sup>92</sup> Moreover, one more oxidation than multiforisin H **82** at position C-8 was observed during islandic acid **161** biosynthesis.

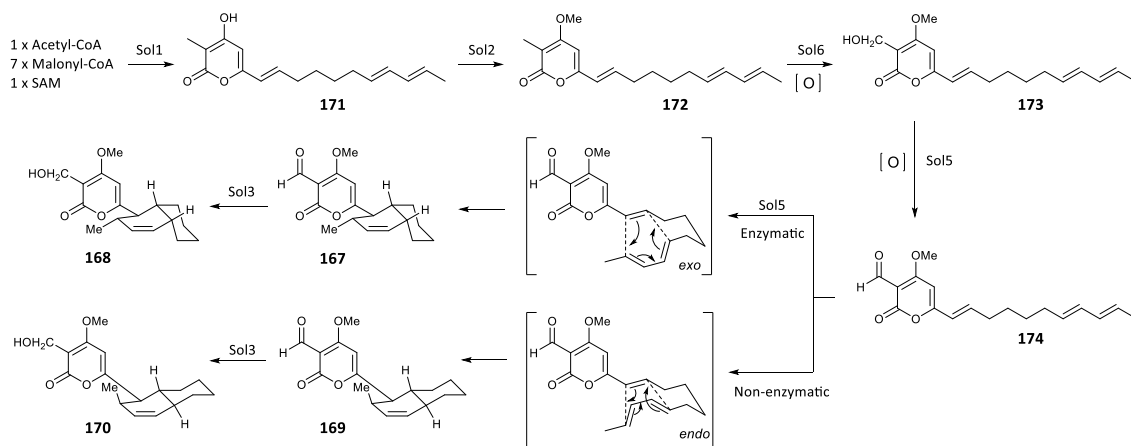
There is a second PKS (*hmPKS3*; Table 4.2, Figure 4.30) contained in the putative multiforisin H **82** BGC ('Cluster B'; Table 4.2, Figure 4.30), but no homology is found between the IslPKS1 and HmPKS3 in Figure 4.30. NCBI BLASTp<sup>61</sup> analysis of HmPKS3 showed that the top hits are SQTKS<sup>119</sup> (squalestatin tetraketide synthase; with similarity / identity: 52.4% / 34.5%) and AzaB<sup>125</sup> (azanigerones hrPKS; with similarity / identity: 52.7% / 35.1%).

SQTKS is known to assemble a dimethyl tetraketide chain (blue colour shown in Figure 4.31), and there exists an *O*-acyltransferase Mfm4 in the squalestatin S1 **24** BGC which is responsible for the addition of the tetraketide acyl group onto the squalestatin core. In addition, AzaB was reported to construct a dimethylated triketide (highlighted by blue colour; Figure 4.31). And the *O*-acyltransferase AzaD in azanigerone A **166** BGC could catalyze the transfer of the acyl-CoA substrate onto the azaphilone backbone. Based on these known examples, we proposed that HmPKS3 (Figure 4.30) is likely to produce a similar short-chain acyl group. In addition, HmCBR3 (*O*-acyltransferase) in 'Cluster B' could load the acyl substrate to the pyrone core scaffold, which may yield metabolites resembling islandic acid **161** and multiforisin H **82**. However, our results suggest that PKS2 may be inactive as no new compound was observed when it was expressed.



**Figure 4.31** Structures of squalestatin S1 **24**, azanigerone A **166** (the configurations of the chiral centers are yet to be determined), multiforisin H **82** and islandic acid **161**.

We also compared the solanapyrones **167** – **170** BGC<sup>81,82</sup> with the putative BGC of multiformisin **H 82** and islandic acid **161** (Figure 4.30). Results showed certain homologies between the encoded proteins such as PKS and *O*-methyltransferase. In addition, the solanapyrones **167** – **170** biosynthetic pathway (Scheme 4.13) and the proposed multiformisin **H 82** and islandic acid **161** pathways (Scheme 4.11 – 4.12) share similar hydroxyl methylation and methyl oxidation tailoring steps. However, the Diels-Alder reaction involved during solanapyrones **167** – **170** biosynthesis makes the structure more complicated.



**Scheme 4.13** The proposed biosynthetic pathway for solanapyrones by Kasahara *et al.*, 2010.<sup>82</sup>

## 4.5 Conclusion and Prospect

This project aimed to investigate the biosynthesis of trienylfuranol A **2**. Isotopic labelling experiments by using [1-<sup>13</sup>C], [2-<sup>13</sup>C] and [1,2-<sup>13</sup>C<sub>2</sub>] acetate revealed trienylfuranol A **2** is derived from a polyene hrPKS (Figure 4.10). Genome mining, PKS domain alignments and transcriptomic analysis led to the identification of 11 potential polyene hrPKS from *Hypomontagnella* genome (Table 4.2). The top two polyene hrPKS BGC (‘Cluster A’ and ‘Cluster B’) were analyzed and annotated in detail.

Heterologous expression of the *hmPKS1* from ‘Cluster A’ yielded a new polyene compound **142** with no hydroxyl at C-11 (Table 4.7), but the C-11 hydroxyl is probably required for trienylfuranol A **2** formation. Moreover, no FMO or P450 monooxygenases are encoded by ‘Cluster A’ (Table 4.3), but the oxidation is indispensable for trienylfuranol A **2** production based on the isotopic labelling experimental results (Figure 4.10). Therefore, ‘Cluster A’ may be not the correct gene cluster for trienylfuranol A **2** biosynthesis.

However, the proposed pathway for the observed polyene **142** may give a clue to the actual and still unknown process involved during depudecin **126**, aureonitol **127** and trienylfuranol A **2** biosynthesis (Scheme 4.6 – 4.9). Particularly, the proposed unique epoxidation/decarboxylation

catalysed by FMO or P450 monooxygenase (*via* iron peroxide) species could occur during the biosynthesis of these four polyene compounds.

For ‘Cluster B’ (Table 4.5), we isolated a new pyrone compound **143** through the heterologous expression of *hmPKS2* in *A. oryzae*. Based on the structural similarity, we propose ‘Cluster B’ is the putative BGC for multiformisin H **82**, a pyrone metabolite found from *H. spongiphila* WT. Heterologous expression to reconstitute the pathway and enzyme *in vitro* studies to fully delineate the multiformisin H **82** biosynthesis will be our next goal. Also, it will shed light on the biosynthesis investigation of a large family of bioactive pyrones related to islandic acid **161**.

## 5 Overall Conclusion and Outlook

The project focused on a detailed investigation of the biosynthesis of sporothriolide **1**, trienylfuranol A **2** and the sporochartines **3a** – **3d**. Firstly, isotope labelling experiments using different types of labelled acetate ([1-<sup>13</sup>C], [2-<sup>13</sup>C] and [1,2-<sup>13</sup>C<sub>2</sub>]) support the origin of sporothriolide as a polyketide or fatty acid and reveal the oxidative tailoring involved during the biosynthesis.

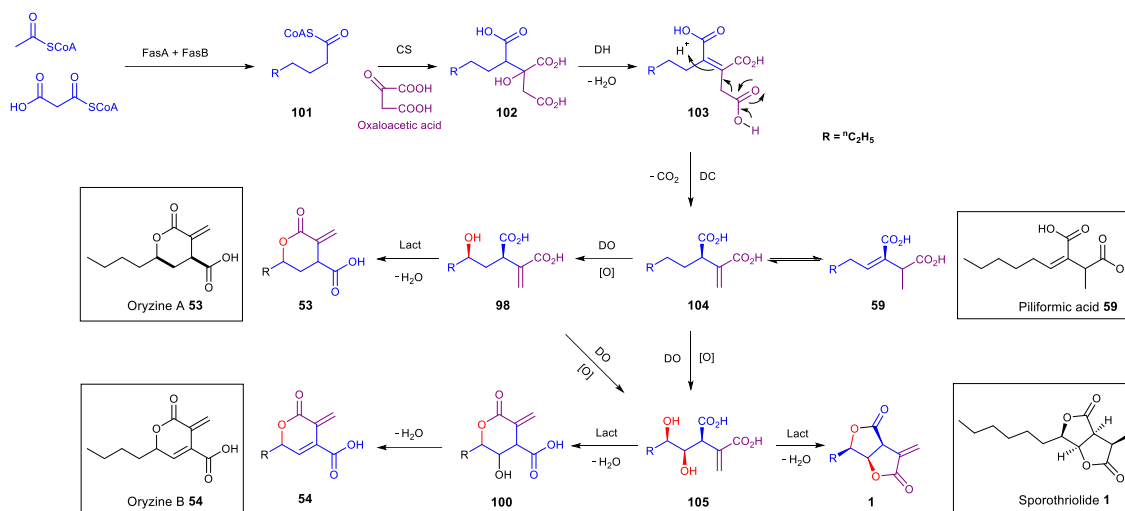
Genome sequencing of three producers (*H. monticulosa* MUCL 54604, *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471), transcriptome sequencing of RNA from sporothriolide producing and non-producing conditions, and subsequent gene expression analysis identified the *spo* alkyl citrate BGC as likely to be responsible for sporothriolide biosynthesis. Knockout of *spoE/spofasA* in *H. spongiphila* removed its ability to produce sporothriolide. By employing the strategies of heterologous expression and *in vitro* assay, we revealed the function of enzymes during sporothriolide formation. In summary, we delineated the biosynthetic pathway for sporothriolide for the first time.

The pathway (Scheme 5.1) begins with the assembly of decanoyl CoA **101** by a dedicated FAS. This is then condensed with oxaloacetate by SpoE (CS) which affords the octylcitric acid **102**, followed by dehydrated by SpoL to give octylaconitic acid **103**. This is decarboxylated by SpoK to yield octylitaconic acid **104**. The next enzyme, SpoG (non-heme iron-dependent dioxygenase) performs mono- and double-hydroxylation of **104**. The double hydroxylated intermediate **105** then serves as the substrate for the heterodimeric functional lactonases SpoH and SpoJ to achieve the formation of the  $\gamma$ -lactone sporothriolide **1**. Also, a self-resistance gene SpoI was found to efficiently hydrolyse the antifungal sporothriolide **1**.

Based on the evidence of sporothriolide pathway, the assumed piliformic acid **59** and oryzines **53** – **54** biosynthesis are well supported. Piliformic acid **59** (Scheme 5.1) may arise from the isomerization of alkylitaconic acid **104** (hexylitaconic acid), therefore no oxygenase is required which fits with the lack of a *spoG* homolog in its BGC.

For oryzines (Scheme 5.1), the intermediate hexylitaconic acid **104** can be mono-hydroxylated to **98** by OryG (SpoG homolog), also double-hydroxylated to **105**. Then lactonases OryH/OryL in oryzines BGC could catalyse the ring closure of **98** and **105** to form the  $\delta$ -lactone oryzines **53** – **54**, respectively. The final dehydration of **100** yields oryzine B **54**.

Understanding the selectivity of the lactonases controlling the formation of  $\gamma$ -lactones (*e.g.* sporothriolide **1**) and  $\delta$ -lactones (*e.g.* oryzines **53** – **54**) would provide opportunities for engineering and producing new compounds in the future.



**Scheme 5.1** The proposed pathway for sporothriolide **1**, piliformic acid **59** and oryzines **53 – 54**.

The polyketide polyene trienylfuranol **2** has a high titre in *H. spongiphila* CLL 205 and *H. submonticulosa* DAOMC 242471. The <sup>13</sup>C labelling experiments indicated that **2** is derived from a PKS assembled hexaketide, and with decarboxylation and oxidative modifications. Bioinformatic analysis identified several polyene hrPKS gene clusters, the top hit is ‘Cluster A’ (Table 4.3 – 4.4).

Heterologous expression of the *hmPKS1* in *A. oryzae* yields a new compound **142** which lacks the terminal hydroxyl required for trienylfuranol **2** biosynthesis. And ‘Cluster A’ does not encode oxygenase. Therefore, ‘Cluster A’ is probably not the correct BGC for trienylfuranol **2**. However, the structure showed unexpected oxidative tailoring by *A. oryzae*, which hints at a unique epoxidation/decarboxylation mechanism during **142** formation. In addition, we proposed the key epoxidation/decarboxylation catalysed by the FMO or P450 also exists during the trienylfuranol **2**, depudecin **126** and aureonitol **127** biosynthesis.

Sporochartines are derived from the DA cycloaddition of trienylfuranol **2** and sporothriolide, our acetate labelling experiments provide solid supports for this hypothesis. The sporochartines were obtained through our designed *in vitro* experiments, including mimicking the reaction environments of the fermentation and extraction process. Also, there are no predicted DAases in *spo* gene cluster and we therefore concluded that sporochartines are formed non-enzymatically.

Overall these results show how an approach combining modern sequencing and heterologous expression methods, combined with traditional isotopic labelling and natural products chemistry can elucidate and engineer new biosynthetic pathway in fungi. In this particular case, the pathway could not be predicted bioinformatically because of their close

relationship to primary metabolic pathways. The results also hint at how secondary metabolic pathways can evolve from primary metabolism, and how structural diversity can be programmed by relatively simple genetic changes.



## 6 Experimental

### 6.1 Biology

#### 6.1.1 DNA and RNA Extraction and Sequencing

*Hypomontagnella monticulosa* MUCL 54604, *H. spongiphila* UP-CLL-205 (MUCL 57903) and *H. submonticulosa* DAOMC 242471 were grown in 250 ml Erlenmeyer flasks containing 50 ml YMG media (Table 6.5) for 5 to 10 days (depending on growth speed) at 150 rpm and 25 °C in a shaking incubator. Afterwards, mycelia were harvested by vacuum filtration using a Büchner funnel with filter paper (MN 640 w, Macherey-Nagel, Germany). The biomass was then frozen with liquid nitrogen and ground to a fine powder in a mortar. The DNA extraction and purification were performed with the GenElute® Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, USA) according to manufacturer's instructions.

MinION sequencing library with genomic DNA from the different fungal strains was prepared using the Nanopore Rapid DNA Sequencing kit (SQK-RAD04, Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions. Sequencing was performed on an Oxford Nanopore MinION Mk1b sequencer using a R9.5 flow cell, which was prepared according to the manufacturer's instructions. MinKNOW (v1.13.1, Oxford Nanopore Technologies) was used to control the run using the 48h sequencing run protocol; base calling was performed offline using albacore (v2.3.1, <https://github.com/Albacore/albacore>).

Whole-genome-shotgun PCR-free libraries were constructed from 5 µg of gDNA with the Nextera XT DNA Sample Preparation Kit (Illumina, USA) according to the manufacturer's protocol. The libraries were quality controlled by analysis on an Agilent 2000 Bioanalyzer with Agilent High Sensitivity DNA Kit (Agilent Technologies, USA) for fragment sizes of 500 – 1000 bp. Sequencing was performed on the MiSeq platform (Illumina; 2x300 bp paired-end sequencing, v3 chemistry). Adapters and low-quality reads were removed by an in-house software pipeline prior to polishing as recently described.<sup>126</sup>

The assembly was performed using canu v1.6 and v1.7,<sup>127</sup> resulting in a few contigs representing the corresponding genome. These contigs were then polished with Illumina short read data using Pilon,<sup>128</sup> run for eight iterative cycles. BWA-MEM<sup>129</sup> was used for read mapping in the first four iterations and Bowtie2 v2.3.2<sup>130</sup> in the second set of four iterations. The respective sequences are stored on the ENA (European Nucleotide Archive) portal of the EMBL-EBI (<https://www.ebi.ac.uk/>) under the bioproject numbers PRJEB36647 (*H. monticulosa*), PRJEB37480 (*H. spongiphila*), and PRJEB36653 (*H. submonticulosa*).

Gene prediction was performed by applying Augustus version 3.2<sup>58</sup> and GeneMark-ES 4.3.6<sup>59</sup> using default settings. For Augustus, species parameter sets were established based on GeneMark-ES fungal version predictions. Predicted genes were functionally annotated using a modified version of the genome annotation platform GenDB 2.0<sup>131</sup> for eukaryotic genomes as previously described.<sup>132</sup> For automatic annotation within the platform, similarity searches against different databases including COG<sup>133</sup> KEGG<sup>134</sup> and SWISS-PROT<sup>135</sup> were performed.

*Hypomontagnella monticulosa* MUCL 54604 was grown in two 250 ml flasks each containing 50 ml of a different medium (DPY and PDB, Table 6.5) for 3 days at 25 °C and 150 rpm. Small quantities of mycelia (>100 µl) were removed with a sterile inoculating loop and RNA was extracted from the samples using the Quick-RNA Fungal/Bacterial Miniprep Kit (Zymo Research, Germany). Samples were treated with DNase I (Zymo Research) according to the manufacturer's recommendations. The High Capacity RNA-to-cDNA™ kit (Applied Biosystems by Thermo Fisher Scientific, USA) was used to obtain cDNA. The gDNA contamination of the extracted RNA was checked by PCR amplification of "Polymerase II subunit" (house-keeping gene) gene, the primers binding to the introns was used and cDNA as a template. PCR was conducted by using OneTaq® 2X Master Mix (New England BioLabs, USA), the manufacturer's protocol was followed and the annealing temperature was 60 °C. In total, 2 µg of RNA per sample was used for library preparation with the TruSeq mRNA Sample Preparation Kit (stranded, Illumina). Sequencing of the prepared cDNA libraries was carried out on the Illumina HiSeq 1500 platform (2 x 75 bp) using the 'Rapid Mode'. Data analysis and base calling were accomplished with in-house software.<sup>126</sup> The sequencing raw data for all libraries have been stored on the EBI ArrayExpress server, accession E-MTAB-8948.

### 6.1.2 Strains and Transformation

Strain	Genotype	Phylum	Origin
<i>Hypomontagnella monticulosa</i> MUCL 54604	Wildtype	Ascomycetes	Ref. 37, 52
<i>Hypomontagnella spongiphila</i> CLL 205	Wildtype	Ascomycetes	Ref. 37, 52
<i>Hypomontagnella submonticulosa</i> DAOMC 242471	Wildtype	Ascomycetes	Ref. 37, 52
<i>Saccharomyces cerevisiae</i> CEN.PK	MATa/α ura3-52/ura3-52 trp1-289/trp1-289 leu2-3_112/leu2-3_112 his3 Δ1/his3 Δ1 MAL2-8C /MAL2-8C SUC2/SUC2	Ascomycetes	Lazarus group Bristol
<i>Aspergillus oryzae</i> NSAR1	ΔargB, sC <sup>-</sup> , adeA <sup>-</sup> , niaD <sup>-</sup>	Ascomycetes	Lazarus group Bristol
<i>Escherichia coli</i> OneShot TOP10	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15Δ lacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG	Proteobacteria	Thermo Fisher Scientific
<i>Escherichia coli</i> OneShot ccdB survival 2T1 <sup>R</sup>	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15Δ lacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2	Proteobacteria	Thermo Fisher Scientific
<i>E. coli</i> BL21 (DE3)	F <sup>o</sup> ompT hsdSB (r <sup>-</sup> B m <sup>-</sup> B) gal dcm (DE3)	Proteobacteria	Thermo Fisher Scientific

**Table 6.1** Strains and origin.

### 6.1.2.1 *E. coli* Transformation

To 50  $\mu\text{L}$  of competent *E. coli* cells (Top10 or ccdB Survival ccdB Survival<sup>TM</sup> 2 T1R, Thermo Fisher Scientific, USA), either 1  $\mu\text{l}$  of purified plasmid or 10  $\mu\text{l}$  of a ligation mixture was added and incubated on ice for up to 20 min. After a 1 min heat shock at 42 °C, the cells were placed on ice for 3 min and 500  $\mu\text{L}$  of LB medium (Table 6.5) was added. Cells were incubated at 37 °C, 350 rpm for 1 h and then distributed on LB agar (Table 6.4) supplemented with antibiotics for 12 h at 37 °C.

### 6.1.2.2 *S. cerevisiae* Transformation (Yeast Homologous Recombination)

A fresh (3 – 5 days old ) single colony of *S. cerevisiae* was used to inoculate a 5 mL YPAD pre-culture and incubated at 30 °C, 200 rpm. After 18 h 20 mL of YPAD medium (Table 6.5) was added and incubated for 4 h at 30 °C, 200 rpm. Cells were collected by centrifugation (3000 x g, 5 min), washed with 25 mL ddH<sub>2</sub>O and centrifuged (3000 x g, 5 min). The supernatant was discarded and the cells were suspended in ddH<sub>2</sub>O before being transferred to a 1.5 mL microfuge tube. The mixture was centrifuged (21000 rpm, 15 s) and cells were resuspended in 400  $\mu\text{L}$  0.1 M lithium acetate. 50  $\mu\text{L}$  aliquots were prepared, centrifuged (21000 rpm, 15 s) and the supernatant discarded. 240  $\mu\text{L}$  of a 50% PEG 3350 solution, 36  $\mu\text{L}$  1 M lithium acetate, 50  $\mu\text{L}$  carrier DNA (denatured salmon sperm DNA, 2 mg / mL in TE buffer or ddH<sub>2</sub>O), and up to 34  $\mu\text{L}$  DNA were added to the cell pellet. The concentration of each linear DNA fragment was approximately 0.5 – 1  $\mu\text{g}$ . Cells were suspended in the transformation mixture and incubated at 42 °C for 40 min. Cells were collected by centrifugation (21000 rpm, 15 s), the supernatant was discarded and the pellet was suspended in 500  $\mu\text{L}$  ddH<sub>2</sub>O. 250  $\mu\text{L}$  were spread on SM-URA plates (Table 6.4) and incubated at 30 °C for 4 – 5 d.

### 6.1.2.3 *A. oryzae* Transformation

1 mL spore suspension (1/10) from a fresh *A. oryzae* NSAR1 DPY plate was used to inoculate 50 mL (250 mL flask) of GN liquid medium (Table 6.5). The culture was incubated for 24 h at 28 °C, 110 rpm. Cells were collected by filtration over sterile miracloth, washed with 0.8 M NaCl (50 – 100 mL) and suspended in 10 mL filter sterilised AO protoplast solution (10 mg/mL, Table 6.3) by inversion. The suspension was incubated at 30 °C and 50 rpm for 3.5 h. Protoplasts were released by repeated pipetting with a cut tip and gravity filtration through sterile miracloth removed remaining mycelia. Protoplasts were collected by centrifugation (3000 x g, 5 min) and the supernatant was discarded. The pellet was then directly suspended in the required amount of fungal transformation solution I (100  $\mu\text{l}$  per transformation, Table 6.3). After addition of the transformation vectors ( $\geq 1\mu\text{g}$ , in 10  $\mu\text{L}$  ddH<sub>2</sub>O) the mixture was incubated on ice for 5 min. Afterwards, 1 mL of fungal transformation solution II (Table 6.3) was added dropwise and the

mixture was incubated at 25 °C for 20 min. 14 mL molten CZD+S 0.8% agar (Table 6.4, 50 °C) was added and the mixture was distributed over two plates containing CZD+S 1.5% agar (Table 6.4). Plates were incubated at 28 °C for 4 – 5 days until colonies became visible. These were transferred to new plates containing CZD 1.5% agar (without sorbitol, Table 6.4). For further selection well growing colonies were transferred onto a new CZD plate. Viable transformants were placed on DPY agar and incubated for 5 – 7 days before being used for subsequent experiments.

#### **6.1.2.4 *Hypomontagnella spongiphila* Transformation**

200 µL mycelia from a cryo stock was used to inoculate 100 mL (500 mL flask) of DPY liquid culture and incubated for 24 – 48 h at 28 °C and 120 rpm. Cells were collected by filtration over sterile miracloth, washed with 0.7 M NaCl (50 – 100 mL) and resuspended in 11 mL filter-sterilised HYP (Table 6.3) protoplast solution (10 mg/mL) by inversion. The suspension was incubated at 28 °C and 110 rpm for 3 h. Protoplasts were released by repeated pipetting with a cut tip and gravity filtration through sterile miracloth removed remaining mycelia. Protoplasts were collected by centrifugation (4000 x g, 5 min) and directly resuspended in the required amount of HYP transformation solution I (100 µl per transformation, Table 6.3). Transformation vectors ( $\geq 1\mu\text{g}$ , in 10 µL ddH<sub>2</sub>O) were added to the solution and incubated on ice for 50 min. 1.25 mL of HYP transformation solution II was added dropwise and the mixture was incubated at 25 °C for 30 min. Subsequently, 5 mL fungal transformation solution I was added to the mixture and a 1 mL aliquot was distributed on a DPY/S (1.2M sorbitol, Table 6.4) agar plate containing 150 µg/mL hygromycin B. The plates were then air dried and thereafter incubated at 28 °C for 12 h. Each plate was overlaid with 5 mL DPY/S (Table 6.4) soft agar (0.8 %) containing 75 µg/mL hygromycin B. Plates were incubated at 28 °C for 8-10 days until colonies were observed. Colonies were transferred to new DPY plates (no sorbitol) containing 150 µg/mL hygromycin B. Well growing colonies were then again transferred to new DPY plates (no sorbitol) containing 150 µg/mL hygromycin B. Viable transformants were placed on DPY agar and incubated for 5-7 days before being used for subsequent experiments.

## 6.1.3 Primer Sets and Cloning

Primer no.	Sequence (5' to 3')	Primer no.	Sequence (5' to 3')
P1	GGAACCTGGGTACAGCTG	P62	TAATACGACTCACTATAGGG
P2	CATACCGCGCTCCGATCTAC	P63	CTAGTTATTGCTCAGCGGT
P3	CTTGGTCATTAGAGGAAGTAA	P64	CGCGGATCCATGAGCGTACCAACGGAACC
P4	TCCTCCGCTTATTGATATGC	P65	CCGCTCGAGCTAAGCCTGCGCTAACAGC
P5	TGTAAACACAGCGCCAGT	P66	CGCGGATCCATGGCCGACCAACAGAAGGAA
P6	CAGGAAACAGCTATGACC	P67	CCGGAATCTTAGTGAGAGCAACACCCAA
P7	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTTAAATACGGGCCGTGGGGT	P68	GCCAACCTTTGTACAAAAAGCAGGCTCCGCGGAATAAGACTCGGAGATAA
P8	TCTCCACTCGACTCGAGGATGCAAGCTTCAACAGATGACGCACTCCGG	P69	TGCAACCTTTGTACAAAAAGCTGGGTCCGCGTAAATGGATGGTTGATTA
P9	ACGTATTCAGTGTGAAAGATCCACTAGAAAGAACGCTTATACGGATA	P70	GCCAACCTTTGTACAAAAAGCAGGCTCCGCGCTGCTTACCAAGCTACTA
P10	TGCAACCTTTGTACAAAAAGCTGGTCCGTTATTCTTTCGGTTACGCT	P71	TGCAACCTTTGTACAAAAAGCTGGGTCCGTTGAAGCTGTACTTCTATTCT
P11	AGACGTGCGTCCGAGGTGCGTCACTGTTGAAGCTTGCATGCCTGAGGT	P72	GCCAACCTTTGTACAAAAAGCAGGCTCCGCGTATCTCCACCACCGAGAAA
P12	GCGGTGCCATATCCGTATAGACGTTCTTTCTAGTGGATCTTTCGACAC	P73	TGCAACCTTTGTACAAAAAGCTGGGTCCGCTACTTGTCAATACCAACCGCG
P13	TCAGACATTGTGTGAGCCG	P74	GGTTGACGAAAAGCTGGGT
P14	AGGAATCGGTCAATACACTA	P75	TCGCTTACCATGCTTTCGCT
P15	ACCCATTGCTAGACGTTCC	P76	GGCACCCGAAACAAATATAGT
P16	ATTCGAGAGAGCCTCTCC	P77	GAAGACCTTTCCGACCCGGAC
P17	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTGAGGAGCTCAGGAGGAAG	P78	AACCCCTGAGATACAGTGTGG
P18	CTAATGCTCTTAGGGCACTGCTCGTCAAGTGTTCCTCCACTCAACCCGGA	P79	CTTCCCGTGTCTAGCGATTGG
P19	TGGCGCAATACCCGCTTGTAGAGTGGGAAACACTTCGACGACAGTCCCTA	P80	GACACCCGCAACCGACCTGA
P20	TGCAACCTTTGTACAAAAAGCTGGTCCGTTAACTACAAGATTCTCTA	P81	GGGCATCGAAAACCTCCGGC
P21	TTTCTTCAACACAAGATCCCAAAGTCAAATGACTGCCACCAACAGCA	P82	TAAACGCCCAAGTTCAGC
P22	CGATGAGCTCATTITGAGCGTGATTTACCGCGGAGAGATTTTGGATC	P83	GTCCATCCGAGAGTATGCC
P23	TAAAGCAGTGCATCCAAATCTCTCCGGCGGTAAATCAGCTCCAAAT	P84	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTCCGCTAGTTGACTTTGAAA
P24	ACGCAATATCCCATATCATCAATCATGACCTTAAATACGGGCGGTGGGT	P85	ACGTATTCAGTGTGCAAGGATCCACTAGAAGCAACAGCGTTGACTTCTA
P25	CTGCACTGACCAATCCGAGCTGCTCAAATGCTTCTGCTACAGTCCC	P86	CCGAGCACTGCTCCGAGGCAAGGAAATGCGGCTACTACTCTCGTCC
P26	GGTGGCTGGTAGAGCTCATATAATCACTACTTCTTTCGGTTACGCT	P87	TGCAACCTTTGTACAAAAAGCTGGGTCCGCTGGTACGGCGCTCGATGGC
P27	GGTGGCTGGTAGAGCTCATATAATCACTAAATACGGGCCGTGGGGT	P88	GGTGACCTCTAGAAGTCAACGCTGTTGCTTCTAGTGATCTTTCGACAC
P28	TTACTTATGCGTTATGAACATGTTCCCTTAAATACGGGCGGTGGGT	P89	GGACGAGATGGGACGAGATGATAGTCCGCTATTCTTTCGCTCCGAC
P29	AACGACTACCCCGCTTGTAGCAGACATCACCATGATGCTCCACAGCAGAGA	P90	GCTTTCAGCTTCGATGATGG
P30	CTGCGGTACCAAAATAGCTCCGAGGTTATCGGATGGATGGCCCACTTCT	P91	CGTCAGACATTTTGGAG
P31	TCCCTGGTGCAGAAATGGGCGCATCCATCCGATAACCTCGGAGCTATTTTG	P92	Tian
P32	ACGCAATATCCATATCATCAATCATGACCTTATTTTGAATAGGCTA	P93	Tian
P33	AACGACTACCCCGCTTGTAGCAGACATCACCATGCGTTCGCTATCGCCAT	P94	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTGCCTCACTTCAAAGAAAG
P34	ATCCAGAGCTCATTCTGTGTCAGGACATAAACAGGTGCCTCGG3GTGCGCG	P95	CGGATTTGACATCTCCAGGGAGACCGGCACTTCCAATAACAGCGATA
P35	CCTCGATAGCGCCGACCCGAGGCACTGTTATGCTCCTGACACGAATG	P96	AACTCGACCTATCGCTGTTATTTGGAATGGGCTCCGCTCCCTGGAGAT
P36	CGGTGTTGACGCCCATGTTGCGGTGTACAAGGATCTGTGAACCAATACC	P97	ATCGAAAGTCTGTTATGGAACATTTCCCTAAGAACCATGGCAGCTG
P37	CAAGTGGACGGGTATGTTACAGATCCCTGTACAGCCGCAACATGGGC	P98	GGTGTATTAACGCTGCCATGTTTCTTAGGGATGGAATGTTCCATAACAT
P38	ACGCAATATCCATATCATCAATCATGACCTCATTCTGTAATAGATTGGG	P99	TGCAACCTTTGTACAAAAAGCTGGGTCCGCTACTCAATCTGGATGCCTT
P39	GTCGACTGACCAATCCGAGCTCGTCAAATGTTCTCGACGACCGAT	P100	TTTTCTTCAACACAAGATCCCAAAGTCAAATGGTTGCCAACAAGGTG
P40	GGTGGCTGGTAGAGCTCATATAATCACTACTCCACAATAGGACCC	P101	GGTTGGCTGGTAGAGCTCATATAATCACTTAAATCTTCACTTCAACA
P41	TTTTCTTCAACACAAGATCCCAAAGTCAAATGAGCGCTACCAACGGAAAC	P102	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTGCGCGCTACTCAAGGCC
P42	TTACTTATGCGTTATGAACATGTTCCCTTAAAGCTGCGCGCTAACAG	P103	CTTAGGAAATGGCCGCTAGTGTGTGATTTCTCTGTGCTCTCTCTT
P43	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTGGGAAACAGGCGCCAA	P104	GTAGATGGGCAAGGAGGAGACACAGAGAAATCACACATAGCGGCCA
P44	TGCCAACCTTTGTACAAAAAGCTGGGTCCGCTATCCCAACATTTCTTG	P105	GCATCGGTGCTATTGTTGCTGGTCAATAGCTCCCAAGGAAACAGCGGT
P45	GGTGGCTGGTAGAGCTCATATAATCACTACTTCTGTAATAGATTGGG	P106	GGATCCGAATACCGCTGTTTCTGCGGAGGCTATGACAGCAATACAATA
P46	ACGCAATATCCATATCATCAATCATGACCTAAGCCTGCGCGCTAACAG	P107	TCGTGGCAAGGAATTAATCTTTCGCACTTAAAGTCCATTTGGCTTG
P47	GGTGGCTGGTAGAGCTCATATAATCACTAAGCCTGCGCGCTAACAG	P108	AATATCCCAACAAGCAATGGCACGTTAAGGTCGCAACAAGTTAATCC
P48	TGCTTGAGGATAGCAACCG	P109	GGCTGAGGACTATCCCTTAGTAGTTCGTCATATAGGACCATGTTGCA
P49	GGGATGACAGCAGTAACGA	P110	ATGGGTACGGTGCAACATGGTCCCTATATGACGAACTACTAAGGGATGAG
P50	ATCCACCACTATTATCCCACTTATAATA	P111	CAAAGAAAGTAGAGATGATGCGGGGATGACATGACTTGCATCTCCTCAG
P51	GAGACGAAACAGACTTTTTATCGCTAAAA	P112	ATTGTTCCAACTGAGGATGATCAAGTCAATGTATCATCCCCGATCATCT
P52	GAACTGTCAGCGTCTCACC	P113	TGCGGAGGATACGTTGCTGATGCGGTGGCATTAAGAACCAGATCCGTTG
P53	TCGTATCTCTGTATCGCG	P114	GGGTTATCAACACGATCTGGTTCTAAATGCCACGCTCCGACACGCT
P54	CTTTCTTTCTCTTCTTTCCCATCTTC	P115	TGCAACCTTTGTACAAAAAGCTGGGTCCGCTATGACGACGCTCTTTGC
P55	TGACCTCTTAAACCCCAAGT	P116	GCCAACCTTTGTACAAAAAGCAGGCTCCGCTGCGCGCTCGAGACGAACA
P56	ACTTCATCGCAGCTTGACTA	P117	CTCGCATCTCCATCTGCGAGAAATGCGCGCGCATCGCCATGGTTCCCGG
P57	TCTTTCATTTCTTGCAGAC	P118	CTCAACCCCGGGAACCTGCGATGCGCGCGCATTTCTCGCAGATG
P58	CTTCTAAATATCGTTGTAACGTGTTCTGA	P119	GATGACGAGCAAGCTGTGTCGACCGTGAAGCTGCGCCTTGAAGTTGAA
P59	CGAAGTATATGGGAGACTATAGCTACTAG	P120	AAGTACATATCAACCTTCAAGGCGGAGTCTCCGCTGACACAGCTT
P60	CTTCCGTCCTCAAAGTTAGT	P121	TGCAACCTTTGTACAAAAAGCTGGGTCCGCTCACGAATCTCAGCTGGAG
P61	ACCATCTTCGATAATGTG		

Table 6.2 All oligonucleotides used in this work.

## 6.1.4 Components of Buffers, Solutions, Liquid Medium and Agar

Buffer	Conc.	Components
1x TAE buffer	40 mM 20 mM 1 mM	Tris-HCl Acetic acid EDTA
HYP protoplasting solution	0.7 M 10 mg/mL	NaCl lysing enzyme from <i>Trichoderma harzianum</i> (Sigma-Aldrich)
HYP transformation solution I	1.2 M 50 mM 10 mM	Sorbitol CaCl <sub>2</sub> Tris-HCl, pH 7.5
HYP transformation solution II	60% 50 mM 10 mM	PEG 6000 (w/v) CaCl <sub>2</sub> Tris-HCl, pH 7.5
<i>A. oryzae</i> NSAR1 protoplasting solution	0.8 M 10 mg/mL	NaCl lysing enzyme from <i>Trichoderma harzianum</i> (Sigma-Aldrich)
<i>A. oryzae</i> NSAR1 transformation solution I	10 mM 0.8 M 50 mM	CaCl <sub>2</sub> NaCl Tris-HCl, pH 7.5
<i>A. oryzae</i> NSAR1 transformation solution II	60% 10 mM 0.8 M 50 mM	PEG 3350 (w/v) CaCl <sub>2</sub> NaCl Tris-HCl, pH 7.5
Protein loading buffer	50 mM 150 mM 20 mM 10 %	Tris-HCl, pH 8.0 NaCl Imidazole Glycerol (v/v)
Protein elution buffer	50 mM 150 mM 500 mM 10 %	Tris-HCl, pH 8.0 NaCl Imidazole Glycerol (v/v)
Protein storage buffer	50 mM 20 %	Tris-HCl, pH 7.5 Glycerol (v/v)
Tris buffer (SpoG <i>in vitro</i> assay use)	50 mM	Tris-HCl, pH 7.5
PBS buffer (SpoI <i>in vitro</i> assay use)	0.80 % 0.02% 0.27% 0.027%	NaCl (w/v) KCl (w/v) Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O (w/v) KH <sub>2</sub> PO <sub>4</sub> (w/v)

Table 6.3 Components of buffers and solutions used in this work.

<b>Agar</b>	<b>Conc. [% (w/v)]</b>	<b>Ingredient</b>
LB agar	0.50	Yeast extract
	1.00	Tryptone
	0.50	NaCl
	1.50	Agar
YPAD agar	1.00	Yeast extract
	2.00	Tryptone
	2.00	D(+)-glucose monohydrate
	0.03	Adenine
	1.50	Agar
SM-URA agar	0.17	Yeast nitrogen base
	0.50	Ammonium sulfate
	2.00	D(+)-glucose monohydrate
	0.077	Complete supplement mixture minus uracil
	1.50	Agar
DPY agar	2.00	Dextrin from potato starch
	1.00	Polypeptone
	0.50	Yeast extract
	0.50	KH <sub>2</sub> PO <sub>4</sub>
	0.05	MgSO <sub>4</sub>
	2.50	Agar
PD agar	2.40	Potato dextrose broth
	1.50	Agar
DPY/S agar	2.00	Dextrin from potato starch
	1.00	Polypeptone
	0.50	Yeast extract
	0.50	KH <sub>2</sub> PO <sub>4</sub>
	0.05	MgSO <sub>4</sub>
	21.86	Sorbitol (1.2 M)
	1.50	Agar
DPY/S soft agar	2.00	Dextrin from potato starch
	1.00	Polypeptone
	0.50	Yeast extract
	0.50	KH <sub>2</sub> PO <sub>4</sub>
	0.05	MgSO <sub>4</sub>
	21.86	Sorbitol (1.2 M)
	0.80	Agar
CZD/S agar	3.50	Czapek Dox broth
<b>Used in transform pTAYAGSarg vector</b>	18.22	Sorbitol (1.0 M)
	0.10	Ammonium sulphate
	0.05	Adenine
	0.15	L-methionine
	1.50	Agar
CZD/S soft agar	3.50	Czapek Dox broth
<b>Used in transform pTAYAGSarg vector</b>	18.22	Sorbitol (1.0 M)
	0.10	Ammonium sulphate
	0.05	Adenine
	0.15	L-methionine
	0.80	Agar
CZD/S1 agar (CZD/S agar w/o adenine)	3.50	Czapek Dox broth
<b>Used in co-transform pTAYAGSarg and pTAYAGSade vectors</b>	18.22	Sorbitol (1.0 M)
	0.10	Ammonium sulphate
	0.15	L-methionine
	1.50	Agar
CZD/S1 soft agar (CZD/S soft agar w/o adenine)	3.50	Czapek Dox broth
<b>Used in co-transform pTAYAGSarg and pTAYAGSade vectors</b>	18.22	Sorbitol (1.0 M)
	0.10	Ammonium sulphate
	0.15	L-methionine
	0.80	Agar

Table 6.4 Agar used in this work.

Media	Conc. [% (w/v)]	Ingredient
LB	0.50	Yeast extract
	1.00	Tryptone
	0.50	NaCl
YPAD	1.00	Yeast extract
	2.00	Tryptone
	2.00	D(+)-glucose monohydrate
	0.03	Adenine
GN	2.00	D(+)-glucose Monohydrate
	3.00	Nutrient broth Nr.2 from Oxoid (Fisher Scientific)
DPY	2.00	Dextrin from potato starch
	1.00	Polypeptone
	0.50	Yeast extract
	0.50	KH <sub>2</sub> PO <sub>4</sub>
	0.05	MgSO <sub>4</sub>
MMK2	4.00	D(-)-Mannitol
	0.50	Yeast extract
	0.43	Murashige & Skoog salt
PDB	2.40	Potato dextrose broth
YMG	0.40	D(+)-glucose Monohydrate
	0.40	Yeast extract
	1.00	Malt extract

Table 6.5 Liquid medium used in this work.

## 6.2 Chemistry

### 6.2.1 Fermentation and Extraction of Compounds

#### 6.2.1.1 Small Scale

For product analysis of *Hypomontagnella* strains (wild type and transformant), fungi were grown in 100 mL of DPY or PDB medium in 500 mL flasks for 6 to 7 days at 28 °C and 130 rpm. *Aspergillus oryzae* transformants were grown in 100 mL of DPY medium in 500 mL flasks for 5 to 7 days at 28 °C and 110 rpm. Culture broths were separated by Büchner filtration into supernatant and biomass. Cells were disrupted with a hand blender and stirred in 100 ml acetone for 1 h. The organic phase was filtered by vacuum filtration and the solvent was removed under reduced pressure in a rotary evaporator until a water phase remained. 100 ml H<sub>2</sub>O was added to the aqueous phase and extracted twice with an equal amount of ethyl acetate. Combined organic layers were dried over MgSO<sub>4</sub> and solvent was removed under vacuum. Extracts were dissolved



in methanol and adjusted to a concentration of 10 mg/ml, filtered over glass wool and analysed by LCMS.

The supernatant was extracted twice with an equal amount of ethyl acetate. Combined organic layers were dried over MgSO<sub>4</sub> and solvent was removed under vacuum. Extracts were dissolved in methanol to a concentration of 10 mg/ml, filtered over glass wool and analysed by LCMS.

#### 6.2.1.2 Large Scale

For compound isolation fungi were grown in up to 1 L total volume of DPY or PDB medium (up to 10 flasks, each containing 100 mL medium) and grown for 6 to 7 days at 28 °C and 130 rpm in case of the *Hypomontagnella* strains and for 5 to 7 days at 28 °C and 110 rpm in case of *A. oryzae* transformants. Culture broths were separated by Büchner filtration into supernatant and biomass. The supernatant was extracted twice with an equal amount of ethyl acetate. Combined organic layers were dried over MgSO<sub>4</sub> and solvent was removed under vacuum. Extracts were dissolved in methanol to a concentration of 20 – 30 mg/ml, filtered over glass wool and processed by preparative LCMS.

#### 6.2.2 Analytical LCMS

A Waters 2545 binary gradient module with a Waters 515 HPLC pump coupled to a Waters 2767 autosampler, a Waters 2998 DAD, a Waters 2420 evaporative light scattering detector (ELSD) and a Waters single quadrupole mass detector 2 (SQ detector 2) were used for analytical LCMS. The DAD measured wavelengths from 210 to 600 nm and the mass detector was adjusted depending on the sample to measure in a range of 100 – 1000 *m/z*. A Phenomenex Kinetex column (2.6 µm, C<sub>18</sub>, 100 Å, 4.6 x 100 mm) equipped with a Phenomenex Security Guard precolumn (Luna, C<sub>5</sub>, 300 Å) served as stationary phase. The mobile phase was composed of HPLC-grade water mixed with 0.05 % formic acid (solvent A) and HPLC-grade acetonitrile mixed with 0.045 % formic acid (solvent B). A solvent gradient was run at a flow rate of 1 ml/min over 15 min starting at 10 % B and ramping up to 90 % B (Table 6.6).

#### 6.2.3 Preparative LCMS

Samples were dissolved in methanol or acetonitrile/water (9:1) mixture, adjusted to a concentration of 50 mg/mL and filtered over glass wool. For compound purification the same LCMS setup as above was used, but the stationary phase was replaced by a Phenomenex Kinetex Axia column (5 µm, C<sub>18</sub>, 100 Å, 21.2 x 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna, C<sub>5</sub>, 300 Å). The gradient was selected depending on the samples (Table 7.4) and run with a flow rate of 20 mL/min. Between 50 and 600 µL of the crude extracts were injected for each run. Fractions were collected with the Waters Sample Manager 2767 by either mass

directed or time-dependent trigger. Combined fractions were firstly evaporated under vacuum to remove the organic solvents, then the remaining aqueous phases were dried in a Freeze Dryers Rotational-Vacuum-Concentrators ALPHA 1-4 LDplus (Martin Christ Gefriertrocknungsanlagen GmbH, Germany), weighted, dissolved and analysed by HPLC and NMR.

	Time / min	Flow / mL/min	%A (Water)	%B (Acetonitrile)
<b>10-60-90%, 15 min Prep.</b>	Initial	20	90	10
	9	20	40	60
	13	20	40	60
	14	20	10	90
	15	20	90	10
<b>10-30-80%, 15 min Prep.</b>	Initial	20	90	10
	2	20	70	30
	13	20	20	80
	14	20	90	10
	15	20	90	10
<b>10-90%, 15 min Prep.</b>	Initial	20	90	10
	10	20	10	90
	12	20	10	90
	13	20	90	10
	15	20	90	10
<b>10-90%, 15 min Ana.</b>	Initial	1	90	10
	10	1	10	90
	12	1	10	90
	13	1	90	10
	15	1	90	10

**Table 6.6** LCMS gradient details.

## 6.2.4 HRMS

High-resolution mass spectra were acquired on a Waters Acquity ultra-performance liquid chromatography (UPLC) system coupled to a quadrupole time-of-flight mass spectrometer (Q-TOF).

## 6.2.5 Nuclear Magnetic Resonance (NMR) Analysis

NMR data were recorded using either a Bruker Ascend 400, a Bruker Ultrashield 500 or a Bruker Ascend 600 instrument each equipped with a cryo-cooled probe at 400/ 500/ 600 MHz ( $^1\text{H}$ ) and 100, 125 and 150 MHz ( $^{13}\text{C}$ ). Chemical shifts are shown in parts per million (ppm) in comparison to the TMS (tetramethylsilane) standard. The coupling constants  $J$  are given in Hz. The software MestReNova 10.0 was used for the analysis of the data. For known compounds only 1D NMR spectra were recorded and compared to literature to confirm the structure. For new compounds complete structural elucidation was carried out by conducting 2D experiments in addition including Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Coherence (HSQC),

and Heteronuclear MultipleBond Correlation (HMBC) as well as Nuclear Overhauser Effect spectroscopy (NOESY).

## References

- 1 T. Boruta, *Bioengineered*, 2018, **9**, 12–16.
- 2 N. P. Keller, G. Turner and J. W. Bennett, *Nat. Rev. Microbiol.*, 2005, **3**, 937.
- 3 P. Spiteller, *Nat. Prod. Rep.*, 2015, **32**, 971–993.
- 4 A. A. Brakhage, *Nat. Rev. Microbiol.*, 2013, **11**, 21–32.
- 5 A. M. Aiken, B. Allegranzi, J. A. Scott, S. Mehtar, D. Pittet and H. Grundmann, *Lancet Infect. Dis.*, 2014, **14**, 550–551.
- 6 B. Hamad, *Nat. Rev. Drug Discov.*, 2010, **9**, 675–676.
- 7 R. E. Lewis, *Mayo Clin. Proc.*, 2011, **86**, 805–817.
- 8 E. S. Istvan and J. Deisenhofer, *Science*, 2001, **292**, 1160–1164.
- 9 M. Manzoni and M. Rollini, *Appl. Microbiol. Biotechnol.*, 2002, **58**, 555–564.
- 10 G. Weber, K. Schörgendorfer, E. Schneider-Scherzer and E. Leitner, *Curr. Genet.*, 1994, **26**, 120–125.
- 11 D. A. Herbst, C. A. Townsend and T. Maier, *Nat. Prod. Rep.*, 2018, **35**, 1046–1069.
- 12 S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolásek and N. Ban, *Science*, 2007, **316**, 254–261.
- 13 S. W. White, J. Zheng, Y. M. Zhang and C. O. Rock, *Annu. Rev. Biochem.*, 2005, **74**, 791–831.
- 14 K. E. Lebe and R. J. Cox, *Chem. Sci.*, 2019, **10**, 1227–1231.
- 15 A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
- 16 R. J. Cox, E. Skellam and K. Williams, *Physiol. Genet.*, 2018, **15**, 385–412.
- 17 C. D. Campbell and J. C. Vederas, *Biopolymers*, 2010, **93**, 755–763.
- 18 J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas and C. Richard Hutchinson, *Science*, 1999, **284**, 1368–1372.
- 19 J. E. Baldwin, D. H. R. Barton and J. K. Sutherland, *J. Chem. Soc.*, 1965, 1787–1798.
- 20 J. E. Baldwin, D. H. R. Barton, J. L. Bloomer, L. M. Jackman, L. Rodriguez-Hahn and J. K. Sutherland, *Experientia*, 1962, **18**, 345–352.
- 21 K. Williams, A. J. Szwalbe, N. P. Mulholland, J. L. Vincent, A. M. Bailey, C. L. Willis, T. J. Simpson and R. J. Cox, *Angew. Chem. Int. Ed.*, 2016, **55**, 6784–6788.
- 22 P. J. Sidebottom, R. M. Highcock, S. J. Lane, P. A. Procopiou and N. S. Watson, *J. Antibiot.*, 1992, **45**, 648–658.
- 23 Z. Wasil, E. Kuhnert, T. Simpson and R. Cox, *J. Fungi*, 2018, **4**, 96.
- 24 S. Palys, T. T. M. Pham and A. Tsang, *Front. Microbiol.*, 2020, **11**, 1378.
- 25 J. R. Anderson, R. L. Edwards and A. J. S. Whalley, 1985, **34**, 1481–1485.
- 26 L. M. Elias, D. Fortkamp, S. B. Sartori, M. C. Ferreira, L. H. Gomes, J. L. Azevedo, Q. V. Montoya, A. Rodrigues, A. G. Ferreira and S. P. Lira, *Brazilian J. Microbiol.*, 2018, **49**, 840–847.
- 27 N. C. J. E. Chesters and D. O’Hagan, *J. Chem. Soc., Perkin Trans. 1*, 1997, 827–834.

- 28 H. Culceth, J. Fuchser, S. J. Moss, J. Nieschalk and D. O'Hagan, *Tetrahedron Lett.*, 1998, **39**, 1949–1952.
- 29 R. Schoenheimer and D. Rittenberg, *Science*, 1935, **82**, 156–157.
- 30 D.-S. Tian, E. Kuhnert, J. Ouazzani, D. Wibberg, J. Kalinowski and R. J. Cox, *Chem. Sci.*, 2020, **11**, 12477–12484.
- 31 A. C. Schotte, L. Li, D. Wibberg, J. Kalinowski and R. J. Cox, *Angew. Chem. Int. Ed.*, 2020, **59**, 23870–23878.
- 32 T. J. Simpson, *Chem. Soc. Rev.*, 1987, **16**, 123–160.
- 33 C. A. Jones, P. J. Sidebottom, R. J. P. Cannell, D. Noble and B. A. M. Rudd, *J. Antibiot.*, 1992, **45**, 1492–1498.
- 34 L. K. Caesar, N. L. Kelleher and N. P. Keller, *Fungal Genet. Biol.*, 2020, **144**, 103477.
- 35 J. E. Galagan, S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read, D. Jaffe, W. FitzHugh, L.-J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C. B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M. A. Nelson, M. Werner-Washburne, C. P. Selitrennikoff, J. A. Kinsey, E. L. Braun, A. Zelter, U. Schulte, G. O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R. L. Metzzenberg, D. D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R. J. Pratt, S. A. Osmani, C. P. C. DeSouza, L. Glass, M. J. Orbach, J. A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D. O. Natvig, L. A. Alex, G. Mannhaupt, D. J. Ebbole, M. Freitag, I. Paulsen, M. S. Sachs, E. S. Lander, C. Nusbaum and B. Birren, *Nature*, 2003, **422**, 859.
- 36 H. Nordberg, M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev and I. Dubchak, *Nucleic Acids Res.*, 2014, **42**, D26–D31.
- 37 D. Wibberg, M. Stadler, C. Lambert, B. Bunk, C. Spröer, C. Rückert, J. Kalinowski, R. J. Cox and E. Kuhnert, *Fungal Divers.*, 2021, **106**, 7–28.
- 38 C. Wang, V. Hantke, R. J. Cox and E. Skellam, *Org. Lett.*, 2019, **21**, 4163–4167.
- 39 K. Becker, S. Pfütze, E. Kuhnert, R. J. Cox, M. Stadler and F. Surup, *Chem. Eur. J.*, 2020, 1438–1450.
- 40 U. Kück and B. Hoff, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 51–62.
- 41 C. Fairhead, B. Llorente, F. Denis, M. Soler and B. Dujon, *Yeast*, 1996, **12**, 1439–1457.
- 42 M. L. Nielsen, L. Albertsen, G. Lettier, J. B. Nielsen and U. H. Mortensen, *Fungal Genet. Biol.*, 2006, **43**, 54–64.
- 43 J. S. Jeong, T. K. Mitchell and R. A. Dean, *FEMS Microbiol. Lett.*, 2007, **273**, 157–165.
- 44 F. Alberti, G. D. Foster and A. M. Bailey, *Appl. Microbiol. Biotechnol.*, 2017, **101**, 493–500.
- 45 J. Antosch, F. Schaefers and T. A. M. Gulder, *Angew. Chem. Int. Ed.*, 2014, **53**, 3011–3014.
- 46 T. B. Pedersen, M. R. Nielsen, S. B. Kristensen, E. M. L. Spedtsberg, W. Yasmine, R. Matthiesen, S. E. K. Kaniki, T. Sørensen, C. Petersen, J. Muff, T. E. Sondergaard, K. L.

- Nielsen, R. Wimmer and J. L. Sørensen, *Int. J. Mol. Sci.*, 2020, **21**, 7601.
- 47 R. Schor, C. Schotte, D. Wibberg, J. Kalinowski and R. J. Cox, *Nat. Commun.*, 2018, **9**, 1963.
- 48 L. Kahlert, E. F. Bassiony, R. J. Cox and E. J. Skellam, *Angew. Chem. Int. Ed.*, 2020, **59**, 5816–5822.
- 49 C. M. Lazarus, K. Williams and A. M. Bailey, *Nat. Prod. Rep.*, 2014, **31**, 1339–1347.
- 50 K. Krohn, K. Ludewig, H.-J. Aust, S. Draeger and B. Schulz, *J. Antibiot.*, 1994, **47**, 113–118.
- 51 F. Surup, E. Kuhnert, E. Lehmann, S. Heitkämper, K. D. Hyde, J. Fournier and M. Stadler, *Mycology*, 2014, **5**, 110–119.
- 52 C. Lambert, L. Wendt, A. I. Hladki, M. Stadler and E. B. Sir, *Mycol. Prog.*, 2019, **18**, 187–201.
- 53 C. Leman-Loubière, G. Le Goff, P. Retailleau, C. Debitus and J. Ouazzani, *J. Nat. Prod.*, 2017, **80**, 2850–2854.
- 54 L. Cao, W. Yan, C. Gu, Z. Wang, S. Zhao, S. Kang, B. Khan, H. Zhu, J. Li and Y. Ye, *J. Agric. Food Chem.*, 2019, **67**, 2811–2817.
- 55 L. L. Cao, Y. Y. Zhang, Y. J. Liu, T. T. Yang, J. L. Zhang, Z. G. Zhang, L. Shen, J. Y. Liu and Y. H. Ye, *Pestic. Biochem. Physiol.*, 2016, **129**, 7–13.
- 56 K. M. N. Burgess, A. Ibrahim, D. Sørensen and M. W. Sumarah, *J. Antibiot.*, 2017, **70**, 721–725.
- 57 H. Fujimoto, M. Sumino, J. Nagano, H. Natori, E. Okuyama and M. Yamazaki, *Chem. Pharm. Bull.*, 1999, **47**, 71–76.
- 58 M. Stanke, M. Diekhans, R. Baertsch and D. Haussler, *Bioinformatics*, 2008, **24**, 637–644.
- 59 V. Ter-Hovhannisyan, A. Lomsadze, Y. O. Chernoff and M. Borodovsky, *Genome Res.*, 2008, **18**, 1979–1990.
- 60 K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2019, **47**, W81–W87.
- 61 M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis and T. L. Madden, *Nucleic Acids Res.*, 2008, **36**, W5–W9.
- 62 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protoc.*, 2015, **10**, 845.
- 63 F. Chen, P. Lukat, A. A. Iqbal, K. Saile, V. Kaefer, J. van den Heuvel, W. Blankenfeldt, K. Büssow and F. Pessler, *Proc. Natl. Acad. Sci.*, 2019, **116**, 20644–20654.
- 64 C. L. M. Gilchrist and Y.-H. Chooi, *Bioinformatics*, , DOI:10.1093/bioinformatics/btab007.
- 65 S. Anders and W. Huber, *Genome Biol.*, 2010, **11**, R106.
- 66 M. H. Medema, E. Takano and R. Breitling, *Mol. Biol. Evol.*, 2013, **30**, 1218–1223.
- 67 K. A. K. Pahirulzaman, K. Williams and C. M. Lazarus, *Methods Enzymol.*, 2012, **517**, 241–260.
- 68 Y. J. Zhou, N. A. Buijs, V. Siewers and J. Nielsen, *Front. Bioeng. Biotechnol.*, 2014, **2**, 1–6.
- 69 Z. Zhang, P. Gibson, S. B. Clark, G. Tian, P. L. Zanonato and L. Rao, *J. Solution Chem.*, 2007, **36**, 1187–1200.

- 70 L. Liu, M. C. Tang and Y. Tang, *J. Am. Chem. Soc.*, 2020, **141**, 19538–19541.
- 71 J. Qi, D. Wan, H. Ma, Y. Liu, R. Gong, X. Qu, Y. Sun, Z. Deng and W. Chen, *Cell Chem. Biol.*, 2016, **23**, 935–944.
- 72 J. Gajewski, F. Buelens, S. Serdjukow, M. Janßen, N. Cortina, H. Grubmüller and M. Grininger, *Nat. Chem. Biol.*, 2017, **13**, 363–365.
- 73 N. P. Keller, *Nat. Chem. Biol.*, 2015, **11**, 671–677.
- 74 W. Weber, M. Semar, T. Anke, M. Bross and W. Steglich, *Planta Med*, 1992, **58**, 56–59.
- 75 C. Leman-Loubière, G. Le Goff, C. Debitus and J. Ouazzani, *Front. Mar. Sci.*, 2017, **4**, 399.
- 76 V. Hantke, E. J. Skellam and R. J. Cox, *Chem. Commun.*, 2020, **56**, 2925–2928.
- 77 H. J. Kim, M. W. Ruszczycky and H. wen Liu, *Curr. Opin. Chem. Biol.*, 2012, **16**, 124–131.
- 78 C. A. Townsend, *ChemBioChem*, 2011, **12**, 2267–2269.
- 79 B. S. Jeon, S. A. Wang, M. W. Ruszczycky and H. W. Liu, *Chem. Rev.*, 2017, **117**, 5367–5388.
- 80 K. Klas, S. Tsukamoto, D. H. Sherman and R. M. Williams, *J. Org. Chem.*, 2015, **80**, 11672–11685.
- 81 H. Oikawa, K. Katayama, Y. Suzuki and A. Ichihara, *J. Chem. Soc. Chem. Commun.*, 1995, 1321–1322.
- 82 K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka and I. Fujii, *ChemBioChem*, 2010, **11**, 1245–1252.
- 83 A. Sib and T. A. M. Gulder, *Angew. Chem. Int. Ed.*, 2017, **56**, 12888–12891.
- 84 M. Matsumoto, S. Matsutani, K. Sugita, H. Yoshida, F. Hayashi, Y. Terui, H. Nakai, N. Uotani, Y. Kawamura, K. Matsumoto, T. Yoshida and S. Studies, **45**, 879–885.
- 85 M. TANAKA, T. FUJIMORI and K. NABETA, *Biosci. Biotechnol. Biochem.*, 2005, **64**, 244–247.
- 86 W. D. Wight, K.-H. Kim, C. B. Lawrence and J. D. Walton, *Mol. Plant-Microbe Interact.*, 2009, **22**, 1258–1267.
- 87 J. Schuemann and C. Hertweck, *Physiol. Genet.*, 2009, 331–351.
- 88 A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *J. Am. Chem. Soc.*, 2011, **133**, 10990–10998.
- 89 W. R. Abraham and H. A. Arfmann, *Phytochemistry*, 1992, **31**, 2405–2408.
- 90 H. Seto, M. Saito, J. Uzawa and H. Yonehara, *Heterocycles*, 1979, **13**, 247–253.
- 91 T. Nakazawa, K. Ishiuchi, M. Sato, Y. Tsunematsu, S. Sugimoto, Y. Gotanda, H. Noguchi, K. Hotta and K. Watanabe, *J. Am. Chem. Soc.*, 2013, **135**, 13446–13455.
- 92 R. Nofiani, K. de Mattos-Shiple, K. E. Lebe, L. C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson and R. J. Cox, *Nat. Commun.*, 2018, **9**, 3940.
- 93 P. Brandt, M. García-Altares, M. Nett, C. Hertweck and D. Hoffmeister, *Angew. Chem. Int. Ed.*, 2017, **56**, 5937–5941.
- 94 D. Schwenk, M. Nett, H. M. Dahse, U. Horn, R. A. Blanchette and D. Hoffmeister, *J. Nat. Prod.*, 2014, **77**, 2658–2663.

- 95 P. Brandt, M. García-Altare, M. Nett, C. Hertweck and D. Hoffmeister, *Angew. Chem. Int. Ed.*, 2017, **56**, 5937–5941.
- 96 G. P. Horsman, S. G. Van Lanen and B. Shen, *Methods Enzymol.*, 2009, **459**, 97–112.
- 97 A. L. Smith and K. C. Nicolaou, *J. Med. Chem.*, 1996, **39**, 2103–2117.
- 98 K. S. Lam, J. A. Veitch, J. Golik, B. Krishnan, S. E. Klohr, K. J. Volk, S. Forenza and T. W. Doyle, *J. Am. Chem. Soc.*, 1993, **115**, 12340–12345.
- 99 J. Ahlert, E. Shepard, N. Lomovskaya, E. Zazopoulos, A. Staffa, B. O. Bachmann, K. Huang, L. Fonstein, A. Czisny, R. E. Whitwam, C. M. Farnet and J. S. Thorson, *Science*, 2002, **297**, 1173–1176.
- 100 W. Liu, S. D. Christenson, S. Standage and B. Shen, *Science*, 2002, **297**, 1170–1173.
- 101 Q. Gao and J. S. Thorson, *FEMS Microbiol. Lett.*, 2008, **282**, 105–114.
- 102 K. Belecki and C. A. Townsend, *J. Am. Chem. Soc.*, 2013, **135**, 14339–14348.
- 103 K. Belecki and C. A. Townsend, *Angew. Chem. Int. Ed.*, 2012, **51**, 11316–11319.
- 104 H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen and B. Shen, *J. Am. Chem. Soc.*, 2008, **130**, 6336–6337.
- 105 S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. A. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajananathan, P. D. Thomas, C. H. Wu, C. Yeats and S.-Y. Yong, *Nucleic Acids Res.*, 2012, **40**, 4725.
- 106 I. Letunic, S. Khedkar and P. Bork, *Nucleic Acids Res.*, 2021, **49**, D458–D460.
- 107 F. Sievers and D. G. Higgins, *Protein Sci.*, 2018, **27**, 135–145.
- 108 O. Piech and R. J. Cox, *RSC Adv.*, 2020, **10**, 18469–18476.
- 109 M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, 2016, **11**, 3319–3327.
- 110 M. C. Tang, Y. Zou, K. Watanabe, C. T. Walsh and Y. Tang, *Chem. Rev.*, 2017, **117**, 5226–5333.
- 111 R. Fujii, T. Ugai, H. Ichinose, M. Hatakeyama, T. Kosaki, K. Gomi, I. Fujii, A. Minami and H. Oikawa, *Biosci. Biotechnol. Biochem.*, 2016, **80**, 426–431.
- 112 Y. Fujimoto, H. Tsunoda, J. Uzawa, and T. Tatsuno, *J. Chem. Soc., Chem. Commun.*, 1982, 83–84.
- 113 H. Zhou, K. Qiao, Z. Gao, M. J. Meehan, J. W. H. Li, X. Zhao, P. C. Dorrestein, J. C. Vederas and Y. Tang, *J. Am. Chem. Soc.*, 2010, **132**, 4530–4531.
- 114 A. O. Zabala, Y. H. Chooi, M. S. Choi, H. C. Lin and Y. Tang, *ACS Chem. Biol.*, 2014, **9**, 1576–1586.
- 115 Y. Xu, P. Espinosa-Artiles, V. Schubert, Y. ming Xu, W. Zhang, M. Lin, A. A. Leslie Gunatilaka,



- R. Süssmuth and I. Molnár, *Appl. Environ. Microbiol.*, 2013, **79**, 2038–2047.
- 116 C. D. Reeves, Z. Hu, R. Reid and J. T. Kealey, *Appl. Environ. Microbiol.*, 2008, **74**, 5121–5129.
- 117 S. Wang, Y. Xu, E. A. Maine, E. M. K. Wijeratne, P. Espinosa-Artiles, A. A. L. Gunatilaka and I. Molnár, *Chem. Biol.*, 2008, **15**, 1328–1338.
- 118 K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2007, **8**, 289–297.
- 119 B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *Chem. Commun.*, 2016, **52**, 6777–6780.
- 120 Y. He and R. J. Cox, *Chem. Sci.*, 2016, **7**, 2119–2127.
- 121 H. Fujimoto, Y. Satoh, M. Nakayama, T. Takayama and M. Yamazaki, *Chem. Pharm. Bull.*, 1995, **43**, 547–552.
- 122 H. Fujimoto, *J. Nat. Med.*, 2018, **72**, 20–31.
- 123 T. Schafhauser, D. Wibberg, C. Rückert, A. Winkler, L. Flor, K. H. van Pée, D. P. Fewer, K. Sivonen, L. Jahn, J. Ludwig-Müller, T. Caradec, P. Jacques, M. M. E. Huijbers, W. J. H. van Berkel, T. Weber, W. Wohlleben and J. Kalinowski, *J. Biotechnol.*, 2015, **211**, 101–102.
- 124 Y. Fujimoto, H. Tsunoda, J. Uzawa and T. Tatsuno, *J. Chem. Soc. Chem. Commun.*, 1982, 83–84.
- 125 A. O. Zabala, W. Xu, Y. H. Chooi and Y. Tang, *Chem. Biol.*, 2012, **19**, 1049–1059.
- 126 D. Wibberg, L. Andersson, G. Tzelepis, O. Rupp, J. Blom, L. Jelonek, A. Pühler, J. Fogelqvist, M. Varrelmann, A. Schlüter and C. Dixelius, *BMC Genomics*, 2016, **17**, 245.
- 127 S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman and A. M. Phillippy, *Genome Res.*, 2017, **27**, 722–736.
- 128 B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young and A. M. Earl, *PLoS One*, 2014, **9**, e112963.
- 129 H. Li, *arXiv Prepr.*, 2013, arXiv:1303.3997.
- 130 B. Langmead and S. L. Salzberg, *Nat. Methods*, 2012, **9**, 357–359.
- 131 F. Meyer, A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich and A. Pühler, *Nucleic Acids Res.*, 2003, **31**, 2187–2195.
- 132 O. Rupp, J. Becker, K. Brinkrolf, C. Timmermann, N. Borth, A. Pühler, T. Noll and A. Goesmann, *PLoS One*, 2014, **9**, e85568.
- 133 R. L. Tatusov, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin and D. A. Natale, *BMC Bioinformatics*, 2003, **4**, 41.
- 134 M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno and M. Hattori, *Nucleic Acids Res.*, 2004, **32**, D277–D280.
- 135 B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout and M. Schneider, *Nucleic Acids*

- Res.*, 2003, **31**, 365–370.
- 136 I. Fujii, N. Yoshida, S. Shimomaki, H. Oikawa and Y. Ebizuka, *Chem. Biol.*, 2005, **12**, 1301–1309.
- 137 D. M. Roberts, C. Bartel, A. Scott, D. Ivison, T. J. Simpson and R. J. Cox, *Chem. Sci.*, 2017, **8**, 1116–1126.
- 138 Y. T. Kim, Y. R. Lee, J. Jin, K. H. Han, H. Kim, J. C. Kim, T. Lee, S. H. Yun and Y. W. Lee, *Mol. Microbiol.*, 2005, **58**, 1102–1113.
- 139 J. Huffman, R. Gerber and L. Du, *Biopolymers*, 2010, **93**, 764–776.
- 140 B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *Chem. Commun.*, 2016, **52**, 6777–6780.
- 141 L. Liu, Y. Han, J. Xiao, L. Li, L. Guo, X. Jiang, L. Kong and Y. Che, *J. Nat. Prod.*, 2016, **79**, 2616–2623.

## 7 Appendix

Protein Name	Protein sequence
SpoC	MSNDSRSTSVASNKDVEDVEAATPPPEVVQEEKSKVEEEKDPNLVWVWDGPDGPNQNFNSNGYKVFITLIWVYG NLTTTIASSIWSSGAGAIIVFDKSTIVVTLGVSLFLLGYAVGPPVWGPVSEFRGRKYPMLAGMFLFTIFCIPAAVGH NMETLLIARFFQGAFGSAPLSLAGGGIVDIWSPAHRGVAIAACIGTIFGSPILAPIMGNFIAASYLWGRWWTQWISCI MGGSCSILVFLPETLAPKILQGAAALRKSANPDAKTAFDGRKLAGPMDIVRIYLMRPFELLATEPILVLITIQS FIYGILYLVFVSYPIAFREIRHWALGVLSALPFLGLMVGVLGAGAVIWHTKTKFMATIKANGGKIIEQRLPMMIIG GCLLPVGLFIFAWTSHPETHWSGMVIGSIPTGMGMVMVVFQCFNYLVDVYAPIANSIAIGGNTFIRSFAGFPFLF APYMYHNLGVWDWATSTLGFISIAMIPILFYKFGHRIRSWSKNSVNTN
SpoD	MYASAQTSRRDAQPSQDPDSASATRADGMRKRPRSGNDEPPTAACDQCRLRKVRCRRQPECSNCRKA GVECNSSNTLKRNVNHTKQLRDFSVVLKHLNDVDHALGTLTELTRQIAARPCHTVNPHAVCAPNNEIAPLPTPGS LDFMLPSAGDEDRVNDPLALNGPLFETIEFDQGGERLYGYPAPLVLIKSLLRQATGALLESKQGESHENGD SYIA RALQDMSARATLRQKLDFFPNLPCRESVAVSDANPVTTTPRLMNVNLFVDGYLHNINTRTPIFDAGLRRAIDAH YGDEQPQESRAWALIINNIVLLELGLIEIQAARASHNSNRGMNDLILPSFLRNCDDRAIGNLEAFMGPSLVNVQALM TLTLAAREFYNNATAEKVCHAAACQVGRAIGLHRSRGARYPNEKGTLSQEPEQERERLFRVLYTMDKQRFVMTGQP CDLHMFDSDHRIGPDRNHEQAEPISDAFDDHMMTIWEEIYNLNSLRAASSGGETRMRQIRLVTSSEVKSQKHA GLMSPSCANGAADVPLQIELLYGYRVSQVLIIRCERGNESQEKMRRELARSSLRILEVCKTPLTTPRLLALLMSFR NYPMVAFVELIAFHLSLFRGECPTAQADVSLRAICDQLHLQHDNLTHIFYARLKLGLVWALETLEALGELIIRS SPQPRGMAGFSPQQRDSRRSTESSRNPTNAPSPMAPDISTACGLHPSRGNQSLSSISSRNGEEDFVQSGLAELT NFGFFTPGTDRLMDLASRPLSAACQFNTSSNSQSQSDLNSGPLTSSSNWGDFFQGVSA
SpoFasA	MTATNSNGRVSEDRRAEQQIAYNLLIELLYQFAFPVQWIDTQRELLTSECNVQRIIEIGPAKVLASMAKKSARL VGEQDLARSIEREFLNITDPE DARKIY EYDENSSAETISAKDGPAPPVAPVNVVAAPIQAPVAAPIAVVSAPIAA ASTVDKDFPTDVLISLVAQKLRRAFDEVLSSEIQNLGGKSTLQNELIGDLAAEFGLDPDGSESTAMDALGKLAS GFSGLKGSKSLVERFLSSKMPGGFGQTEMVAYLGSRWGLGNSQTAVCFCITIEPAARLSDVGVQVHEFLDSA VARYAKHAGVSLPTQSAGGASQGASGAVMQVDKAGLDALKNEQNSVLKRLQLEVLQAHLGVDITPNTASGVEGG DKLQEQDLDRFYAELDEEFLSGVQGMFDPKARSYSSWWNVVREDAARLLRHQDGTASPQRQLALNTRWTAE LEEMLRCAKAGPAKEAAESLLKLPSTQGASPVFRFTEPAMAPHTSVDEEQIHYTEKARQDSSGSRRTTYD VSSTRRDGPKSFSVHCLHRRGGSWYQDGLTNTYLDALFAGNTSGISYAGKTLVGTAGTGSIGIEVVRGLLAGGA RVIVTTSRTAANAGAIMSPLYKEFGARGSELILPFNAASKDVENLVAHIYDSSKIGADVDVFIPFAAIEPEGREID GIDARSEVAHRAMLTNVLRLMGLCKIQQEKEKRSYVGRPTTVVLPSPNHGDFGGDGLYSESKIGLETFLNRYSSERW SGYLSIIGAVIGWTRGTGLMSANNIVAEGIENLGMVMTFTAGEMAFNILALLYPSIIRKSDMEPIYADLSGGLMGMFQ LKEEIMAIRTNITGKRREQAIAERQHEEVLKGSKAASAQSKKSSPQKRSNIRQGFPRLSSHQEMTAGLES LGMVDSLRTVVVGFSELGPWSSRTRWQMESQGLAQDGLTEMAWMMGLVKHHDGLVDGKPYVGVWLDVE SKKPVQEDEFSAARYGEHIMNHSGIRIVEPEALDGFDPAKKELLHEVVDLDDLPADFDTSEALQSFKLRHGDKVTIFQ KGSADTWTVIVKRGATFIIPKSSTGHQTVAAQIPKGWNAATYGIPEIISQVDPITLVYLVCCVCEAMFSGADIEPFE LYKYIHVSELENCIGSGAGGLKSMRDMYRHRYPDEPVQGDILQETFLNSMAAWTNMMLFGATGPICKPTGTGTCATS VESLDNACEGIRSRRVKVALVGGTDDIQEEVAHEFSNMKATMVAEKELAKGYPQMSRPTATS RAGFVESAGC GVQIVMSAELAIQMGPIYAVVAYTQMAGDSVGRSVPAPGQVLTARETPAASRSPDLLDRYRRSRLEQIEAIE GWRLSQLASTSSHVGTHEAAHSQMIESAASRRKSDAQWMMWNGDIRQLDPSIAPMRAALAVWGLSIDDIGIASF HGTSTKANDKNESVINQMMTHLGRSGNPLLVICQKFLTGHPKGAAGAWMLNGCMQVLESGLVPGNRNADD VDGALRAFPHLLYPSSELQVANIKAFMLTSFGFGQKGGIVIGVTPRALFAALAAPKFEAYREQVERRRRRADRAFQL AMMTNSVFKAKDQSAWIEAGRAAGAVFLDPTARI
SpoE	MPSATVPKANGTNGTNGVHVKPGVDLHVIDSRTGEYHAVKIHNNAINATDLKAIKAPKDEHPEYQNEQGI RVYDPGFNTVSESKVTYIDLEGTIQRYGSIHDVIGKKFVDLSYLLIWGEWPSAEEAQKYQERLNNVPLIDESV FNVIRSFPKNGSIVGMMVAGLSALQSCDMAAVPAHAAKNLYLGQPQNVDEQIIRVMSSLSMITAAAYCHHTGRT FTPPRKDFSYIENFLMTGHVDESTGLPNPRVYNALERLWAVVADHEMTCTAALLQTASALPDVISSLISAFSAGY GPLHGGAIYVAYKNIEIETVDDVPAKLARVKAGKERLYGYGHRVYRVTDPFRFYISELGEELTEEVNKPDLQVAF LDRAAAQDEYFISRKLRPNADLFAAFYKALGFPPNFILPISMISRTQGFMAHWKEAMEGGPRIWRPQGIYTGK NRKE
SpoF	MVSTTNNAEERSLLPASNSLNTFVLPENEKPRNWSRYSRWLCVGVISLYGLMSPVMAAAIIPALPAISDLSISD EKTALGALVSIYLLWSVTPVFLGPLSEVYGRVGLLQIGHGLFMVFNLSVFAQTGPQLLVRFLAGGVGSGPLSIGAG IIGDLWAPPEERGVSIYTLGPLLGAIPGIAAAYISANFSWRWIFGFSSYILITLILGLFVLQETLLPVITERKRAAF FPQQGLVSDYGTVPVPEIPETEHKDFKAVRQSLMRPFILLWTQPIIQLAIFTGYQFGLNHLTITTFQSLWRD VYQ QDMLGASWNYFIYAVGVFVGSQATGMLNDRIYKRLDKKNSRNPRLRYMMLPASLLVPLGLLYGWSAERHMH WLIPDVGVCIYATGLIMSYQCTQAYIIDCYTSHAASSMSALMIVRSITGFTFPIFAPVLFVSVWYGLGSTWLAGC MGLGIPIMLKVYGPALRARSTYAVGE
SpoG	MSATNGTVQPLELSGALSAYEAIDMTPCIGTEFPTDLAEALRAPNSDEIIRDLAITICRRGVVAFRSQTNMTNELQ KELTHRLGELSGKPAGHRLSKHPLHIRKDDPEMGLDAGRQQALHGGDTTDKRQKASVEVHSDGSYEVCPDFT SLRMTDIPRTGGDTLFAAGYELDRLESEPYQRFESLTATHEVPALRKAETMEGIYTPGPRAPANTDMLFKQSH MVRTHPVTGWKTMFAGGLHCRRVNGVTEWESQELLEKILRLVADNHDLQVRIRWNTPCDMVIWDRVCLHCP TQDHYGLGKRMGYRTMSVAEKPFNLSSPSRLEANAVVGEKAGKKEVSEVPVAAPVKIPAVAAAPAVSAQA
SpoH	MRSLSPISDVTPSWGSSLVHLAGLLLTQATTAQVAKKCGNLDFFPTVCLDRYGTLLPGKFSRGAALDVSASSVD DTSFNSIKSASFLVFDQDKGLDVLGTAPVVEEMFDLDSAAPEAPVYVPTNELWIGGLQTGVTSQTVVDSQNP KPVKRTLNPPIYAANGMRYRDGRVWVSAAGGNDTLAGGPYHPGIYSFDPKTDGSRVEANNYYGWFINANDLD LDPSGRVWFTDPLYSRNMGNTEAPLLQAAVRYDPVSGQIQVMDDTLEFPNGIAFSFDGKILYLTNTAAGVGN DPGTPWQNAAGPKYVSTNKRTLYAFDVGTDGLLRNRRPLYTAMDYVADCVKVASNGYLVTAAGHGVDILDP VHLMRIQLSFLAVSIEFAGPKRDSLWIIHGKAARATINLTGQGASTASSRLRRHARTHAKRHVAPQSITE

SpoJ	MADQQKEVPEKTLPVVHRFITTHNAEGKPTFETGIKEEIDFERSPLGGDMFLAYSVEFPVALGHSDLNQYKAHL EKKPESFMIPGGFLSRDYDHPGCLPLWHRTITLDFGIVVEGQIQLELESSEKRIKKGDVAIQRGTNHAWSNPSKTE FARVYVAMDAKPPVNVNGQELGESLGVVSH
SpoJ	MCSRPRIRFLPYLLISSQLSALSFLGNDGLQIPLTVDADEVDEVAQEQSKLAAPCLGYEFPFHVICIHRYSGLIHGDF ERKVRNVLGDTETYPSTHAPGESTFTHISDADFLIWDTVGGQILGNSPISDFMFEVAPVSHEAPVYSPPTNELYFS RLQQGFPLQIVVDLNDQPPVLTEKLASPIYAAAGGRFYKGLIYFSTIGGNESLGGYFRPGIYTLDPKTKGESKTLNIN YWGWFNAADDLDVDDQGLWFDNNDYGRPVHVNTYAPQMGVATYRNISSGLVAIVEDTLKEPNGVAFSPD RNTLYLSDTGAGSSIIDGRVNPAPSIHVNTGPRLIYAYDVSKSRKGLSNKRPIYRAIDYAPDGVKISREGYIVTATGH GVDILSADGEPLVRVQNTFTVINIGWAGKESDELWVVGKGGVARVRWALRGPIVE
SpoK	MGKHGANVVFDETNDENSGVTAALCNWIAAGLKEDIPTVLERAKHLIDGCIACGLVGAHVWRWSEKAADAVLDY EPEGQCSVIGYEEKLGPLAAAVLNGTFIATELDDYHSVAPLHSASVVLPAALAAQVKNKTRKSAQNGNGHSAN GSTRTVSGLDFLIAAVVGFETGPRSGSAMHGADLLLRGWHSGPVFGCPAAAAASSKLLGLSADDTESAIGIACQQA GGLMAAQYEGMIKRVQHAFARNGFLGALLSRNGYVGIKKYERNYGGFLNMFSGQNGKTPPYDVRKVTGLG EVWQTTNIRVKLHACVGGCHGQIEAIEKLQKAHPERFAIGNLGHISIKVGLSGPIFAHDGWEPQERPLTETGAQ MNAAYIGAIQLVDGQVLIAEFANHKMDRDIVWDLVYKTKCHHDTQFDKPNHGGCAHIVVEFDGFTVEETIQM PRGFDPPITDEEIRTKYRKLALSADQQRMKEIEELIGIDKLDLDDISEIFEVLAQPTRNVLG
SpoL	MHVQPDDNSQRPYDEVINLIVDYAFDYEVKSEAAWARSKMALIDSGVAIESLVKSEKESLIKPLLPATNVG GFRLPGTSYSLDLQGAFFNMGAMIRYLDHNDAFPGEAWGHPSDNLGAILVADVLRDALARGKPEEAITMKQV LIGLIKTYEQGVFQIKNAFNRVGLDHTLVKIASTAMVSWMMGLSREQARAASVSHAWVDGHPLRFRQAPNAGP RKGWAAGDACMRAVHLATLARSGQPGRITPLTATRWGFYQVLNKDQEFQLPRPFGTWAVENVIFKVLTAEGHG LTTVEATMAASKELQARGLDPLKDIKNIHVRTQEAAMIINKKGPLHNPADRDHCLRYMMGVILLKNGVEVEAED YQDDSPFATDPRVEALRSIIMEEDVQFTQDYHNPTIRSCGSSIEIFLKDGTINVRQDFPLGHVVRDAETIPLVRKK AIHNLGLKFSQDEVTRIMETLEQPDFDTLPASKFIDLFQK
SpoFasB	MESSGGSTSSFDEVNPTPGIADTPGIFTPGIPVTFITLNYEEVEVEFSLAPSDATHLDEHRRFTLTLARSEGDNEEKK PMSAAALTFKFLEHLLRRSVSPGTLARFFYAVQSDLMEQKDIHDFISELPGASTRKSALRTYMTLSSKLSCLPSPGP SALLTAARRERSSILVAFGGQSSNPACVDDLAELYSLYRPLVEPLVSSLGAALLSLSRHPDTKAFFLGREIDLSAWLA DPSTRPAKNFIAGAAVSPFIIGLGLLHYAIICKMLGKTPAELGQLLSGITGHSQGVVAAAQVAKSHSWESFFVEAR WAVELLFWMGYESQMAAPQSPISAMVNDVSVESGVGVP SHMLLVGRMRRQLEAIVAASNKHLKPNERNLYLSL INSARNYVIAGPPRSRLRGLSLREICARDGLDQSRVPYSKRKPVILFQFLPVNAPFHSPLYNGAERISARISGSWPE VTTISSLHVPVFTYENGADMTKSYKADVDVDTQLLIDAVTTRVVDWPKTLQVGREKRLSHIITLGAAGRFSSMIHENV DGYGVRVIDGARIDPVDSTIMGAKAEIFAQFLSRSTMSPSTWQDQFKPRLVQSEGTNIETRLNRLRAPPVITAG MPTPTVPWDFVSAVINAGYHIELAGGGYHNAEAMTTAIEKVAASIPTRGRTCNVYVDPKAIQYQIPLIRQLIRKG VPIEGLTVGAGVSPDVAAEYIQTGKIKHISFKPGSIAAIEVIEIAKRHPTFPVILQWTTGGRGGGHHSCEDFHEPILLE TYSEIRRCQNLVYLVGSGFGDAGMFPYLTGSWSLQFGKAMPDCGILLGSRMMIATDAHTSPGAKKLLKAPG VDDAEWEKSYLKADAAGGVLTVTSEMGQPIHKLATRGVRLWKDMDDTIFSLPKPERKAALLKRKDEIIRRLNADY AKPWFGQDAAGQAVDVEDMTYADVLSRLVQLMYVKHQRWVWVQSYRELISEFAIQSLERLGSDFEPSWLNPS ESFVDQVKEACPDVTEQLLHPEDVRFQICCKRGRKPVNFVVALDDDFEHWFKKDSLWQSELDLAVFDQDPER VCILQSPVSVRYATRDDQSSKEILDEIHRDLVVLMMHVAEKPNGHVATRTNVTSRSRPMSENIMVDSMGDRIVFRP VPGEDLPSQEKWMECLDPYASSAILGLIREESLFEAASKRCRPNPFCRIFGPRHGYSLVLCRDYHEALLRDDSTGQTI VRVEARSAKDLRVEFTHRDSVPSGAATLVFQWEYDEHTRQLIDTTENRDKVIQDFYAHLWLQNGTNRTRGLTD RFFADSFELTQQLQALHSVVAHAFPSASPVGQTAVLPLESAVIAAWDVLMRPLISLDLGDILRLVHQSIGVEYVP GVSPMQIGESVTTSSIRSIETIEPSGKSVAVEARLIREGLHVATVTSEFFIKGKFSYQNTFRHKEELPIELKISSIDEAV LRDRSWLKLDDPSTPLVGKTVFVHTRSQTWNQTSANLEILGTVEQKLWNSDKRRLGSAVAFDASETHGNPVEF LQRKGTVDKVKPLKNPGWEGNSEVIVAPPHTHLYAQVSGDCNPIHASPVAELAEPLGPIMHGMYTAAVCRK VVEDLAVPGEPERMRRFNASVGMVVRPGDKLTVGLSHVAMKNGRMILEVIARQEESEEVLRGEAEVEQPSTAY LFTGQGSQSIGMGTALYESSPIAKALYDEMDKHLRDLFGWSILKIRESPELTVHFRGREGQRILENYLNMKTEIIG EDGIRRPAPIIPLSRDSTSYTFSEARGLLHATQFAQPAIILLEKATLEHMRANGLIKEGAVFAGHSLGEYGALSSMA GFVDFKMLSIGFYRGLLMQFAIPRDADGQTYAMMAANPGRVKGHFDDLSALRALVRHIAQEESEELLEIVNFNIE GDQYVCAGHVRLHCLTEILNAAAARKVHPESITEFVTASEPKTTTLGAIHSAIQSKTLPMSQLQRGKATIPLNG IDVPFHSARLRSGVPTFRKFFHERVKAEDIRPERLVGSIIPNVVVGKPFSEIKSFIQEVSKVTESPVLENLVC

Table 7.1 Protein sequence of *spo* (GenBank MT889334) gene cluster.



	<p>CACGGACAAGTTGCCCAACTACGACCTCTCGACCTCTTCACTACTAGCAACCGCCGCT          ACTCACTTCTCCAGCGGGTATGGTGTCTCAGTCTCGCAACCTGCAAGGCTTTCTCT          TCCGAGGAGATTGGAAGTCCGCGAAGAAGGCGGATCTAGCAATGAGAGCAACCTT          ATTTGGTATGGCTTCTACTGCGAGGGTCCCAATGGCAAGAAATGGGTCTGCTCATGG          CCTACTACCTACTTCTGAAGACCATCAGCAGCTGACATGGCTGGATGACCTTGTAT          GATGCTCTAAGTGAACATGAGGATGTTCTGCTGAGCCGAGGCCACTCGCGGGTGA          GCGAGGCGAGTCTCTAGCCCTCTGACCGCGTCAAAATGGTCTGTCAGGCTTCTCC          GCCTCTGGGAATCAAGCTCGGCTCAGTGTGCTCAGTCTGGTGGATCGGCGCCG          CTACCGCCGGATACCTACTGCTACCGAAGCCTGCTCTGCTACTACCGGTGAAGA          CGGTGGCGACATCAACATCAAGGCTGATGATGGCTGCTGGTCTGGTCTGAGCGCT          ATCTCTTACCTGAGACTACGATGACATCAAAATGGCTGCAACCTCCCGTCTGGT          TACTTGTCTGGTACCGGAGTCACTGAAGAAGCTCCAGGAGACCTCAATGACAAGGCT          ATCTTCCGCGCATCGTGAACCCGCTGTAAGGCTACCACTCAACACATGAAGCCTGC          TGCCGAGACATATGATCACTGAGGAGTGTGGCTCAACCTGGGGCAAGAACCGC          CTGCTCGGAGAAGCTAAGATGTTCTGCTGCTGACCAACCGCTCTCCCTGAGGATGC          CGTCTGGAGGAGTACTGAGGAGCTCACTGCTCAACCTGTGCTTCAACAGCGCCG          TCCAGATTTGCCAACCCCGAGTGGCTGACATCAACATGATGAGTGGCGCCCTCAC          TCCGCTCTCGGCGCATCAAGCAGATCAAGGCTGATTTAAATCGAGAACTAAAGTA          CCTTCCACCCCTTCTGGTGGTCTGACTGCTACCTCACTGCTCAAGCTTCTGGTGA          CTGGTGGCTGGTACCGCTGACTTGTAGGGCTACTCCTTGGAGAGATGAGCCCTA          GTGGCAAGATTCTGCTCGGCTGGCACTTCTGCTGACTGCTTCCCTCAAGTGGGAC          AAGACCAAGAAATCTGGCTGAGTCTGTAAGCCGAGAGACCTTAAACCAAGTACG          CTGCGCACGATTTCTGGTGAATCTTGGTGTCTCCCTCGCGAGCTACTGGAGG          AACATCTCGATCTGGTACATGCTGCTCAAGGACACAGCTTGTGGTGGAGGAC          TGTCTTCCGCGGTTACTTCTCATGCTCAAGGCTCAACAGATCAACGAGCTCT          CGGAGAACCCTGTTCCCTGATGCTAGTCTCCGAGTGTATCCATCAACAGGCACTT          GTGACTCTGACGATGACAATGGAGTGGTCTTGAACATGCGCCCTCGCTCAGCG          CCTGCGGAAGGCCAAGCAAGTGGTGGAGTTCAAGCTGCTTCCATCAACGAGACGG          TTTCAAGAAAGCAACATGCTGGGACGCTTGGTCAACATGCGCAAGGAGAGCGCCG          ATCAAGCCAGTCCGAACTACACAGACCGCTTGTGAAGCCTGGAACATGCGCCCTAAA          GCGTGGCGCTCGACTATGCGCCGCTTCCAGGATGGAACAACATCACTTGTGATGA          AGACTATGCGCGCACTGCACTACTGTTGTTAAGCAGAAGGTCAGACAGATGTTGGCGA          GTCCCGCTATGACTCAACCCGCGCTCTGACTCTGCTGCACTGATGTTGCTGCCAT          TTGGGGTGGTCCGCAATGCAATGAAGTTCGGTCTGTTCTGTCGAGCTGACGAGATC          GCCCTCTGGTCCGACCGAAGCCAGCTCGAAGACAGTGGCGTCAAGCTTCTTCTGGAT          CGACCCCGTGGTCTGCTCTTTAAGCTTCAACAGCTGATCGCAGATGCGCAGATGGCCAGC          TCTGATGAGATCAATGCAATGCTGTTGCTGCTCAAGCCGCTCTCCGCGCTTATCCG          AGGAGCCGCAAGCCTGACCATATGCCGTAACCAATCAAGCTGACGTTGGCCCTTCTC          AGCGGAGCAGAAAGCTGCTGCTGCGGACTTCTGCTGAGTGTGATTAAGAAAGCGT          GTCTGAAGGCTTATGCTGATGGCGCTATGCGGAGTCTGCTGCGCAAGATTCTGCTC          TAAGCCTACCGCTACTACGATGACCCGAGGAGTATGAGATGCTCCGCGAACTTCTC          CGGATCAAGAACGTAAGGCGCTCAAAATGATGTTGGTGGGATCTGAAGGCGAGGCT          GTTGTCTGAAGTCTTCTGAGCTGTTGCTCCGATGTTGCTGAGCCTGCACTCTCGCAAG          ATCCAGAACCTGGTCTGAGGGGCTGAGCCATCTTGAAGTCTGCGCCCTCAGTG          AAGCTGACCTGAAGACCTCGGATCTCCGGTGGTACCATCGAGGCTCTGCTGGCAAG          CATGTGACGGTGGTCTCCGCGGCTCCGAGCCGAGACTAAGGCGCTGCGGCAACGAG          TCCAACCTATCAACCGCAACAGCTCAGCCTGCTTCCGGTGGGAGGAGGCTCAAG          GGTGAGGCTTGAAGGTTGAGGCTGGAATTCGCACTCGACCCCTGCTTCCCTCAAGC          TTGTATGCTGCTGCACTGAGGTTCACTTCTGCGAGTATCTCCGAAAGGAGTCTTGG          ACATGAGGCTTCTGCTGCTGCTAACAACCTTCTGGGTTACTCGCGCGGCTTGTGA          CCAGCGAGAGCTCTGAGTACGCTGATGACAGTGGTCTTCCGCTGCTCCGATCCGAGCAG          GCTTCTGAAGCCACCTTACTGACTTGAAGAGGAGGCGCCCTCGACCAAGTCTGCGA          TTTAACTCACTGACCAAGAAAGGCAATGGCGAGGAACTGAGACTGAGTACTTCT          GTCTCAAGGGTCAAGTCCACATCAGGCTGATTTCCGATGAGCGCATCAACGACTCTA          TGCTCAGATGAGGAGGATTTACTTGGCCATTTGACCCCGCAACGCTTGAAGCA          AGGCGCGCACCCAGGAGTCTTCAATCCGACACCGGATCGAGAGACTGACTTGAAGC          TGAGAAATCGAAGTCAAGTGGTGGCCAGCGCTCAACAAGGAGGACATCGCGTTACT          TCCGTTACCAATTGCAACCGATTGCAACCAAAATTCGCGGATCTGAGCCCGGTTGG          CTCAGCGTCAAGGACTTCCGCGGAGCAGCGCTGCTGACTTCAAGCGCAACGATTTG          CCACCTTCCAGCGGCTCAACAGAGTCTGTTCAAAAGCTTAAAGCTGATGAGTCACTC          ACCATGCTAGTCTCCCTCTCTCAGCGCGCCCTCTACGGTATCGAAGACTAGCCCAT          CTCAGCAGGCGAGTCTGCTCTGCTCTCCCTGCTGCTGATCTCTCCGCTGCAATCAT          AAAGTCTGCAAGGCGCTGGTCTGCTCCATCGTCCGCGCGCGGCTGCGCTGAGGCG          GAAATCTAGCAAGCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT          GACTAAGTACCAAGCGGATGTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          AAGCTGGCTCACTGCAAGTCTTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TCTGCAAGCTCTGCAATGTTCCCGCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG          ATCAACTATTCAAGAAAGCCAAAGTCTGACTGACTACTTAAAGCTATGCTGCTGCTGCT          CGACAGGAACTATTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          CATTGCTTCTCTCGAGATCTCCAGAGCGCTAAGACCGCTACTTGAACGAGGCTTCCA          GCGGCACTACTCTGCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TGCCCAAACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TCGAAAACCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          GGAAGATGCTGGAAGGCTGCAAGGAAATCTCCGCGCCGCGGCTCAAGGGTGTATT          AACGCTGCAATGTTCTGAGGTAATCCATCCAGATCAGTCTGCTTATTATGCTAACA          CTAAGTAGGATGGAATGCTTCAATAACTGACTTCTGATGCTGGTCACTCAACTAACC          CCAAGGTTACGGAAGCGCAACCTCCACGAAGTAAAGGATCTGACTTGAATTTCTC          GTACTGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TAATGCTACATGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TGTCTACTATGATTTGGGATTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          CTCAGCGCAAGGATGATGTTGTTGAGAGTACGAGTCTGCTGCTGCTGCTGCTGCTGCTG          CCATCTGAGCAGCAACCTGACGCAAGGACTGCAACCTGTTGCTGCTGCTGCTGCTGCTG          CCTGCTCTGCGCAAGGCAAGTCCGAACTGAGGAGACGCTGATGCTGCTGCTGCTGCTG          GTTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TGGCGTGTGACTGCTGCTTCCGAGTCAAGGCTGCGCCACTGCTGAGGCTATG          AGCTGCTCAGGAGCACTTATTGGCAAGTACGCTGCTTCTGCTGCTGCTGCTGCTGCTG          TCGAAGGCAAGGCGGATCGGTTGCGATTTGATTTGACAGCATGATTTGCTGCTGAT          TGACAACTGATCTCAAGAGCTGGGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG          TCACTCACTCAGAAAGTCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          G</p>	<p>EVAVKFGTSTPVASNVMLDLEGFPLASIEKEFLDMQALLAANNLWVTRGG          LVTSEPYAMTSLRLRSEQASLNATLDFESEAPLSQFVLDLSTLAKKANGLEE          ETEYFVSKQVHTRLSIDERINDVYAHDEEYVSPDPKRLVKARTHEVVFESD          NGSETLEAEIEIQQVATVGLNKEIDIAVISEFATDFSHIAGIVTRVSSVQFAPS          DRVVSFANKFATFORVNSQLVKNLNADEFTTMSLPISGAALYGLNIAHLQFP          GESVLLVPGVGLSAAIKVLAGATPIVAARGAEALIKLGLSDQIQRWSTAAAL          TKYQPVVSTGSDVSDVAREAHRLPASFVSCGSEKASSNDVQIWPATRGASY          LSFVDINLFKPKILTQLNRIVTRFQGTSLPVAKNIAELNETIASFSDLDQSAKT          ILNEASSGTTLVPSRPHARLDNATYFLVGLGGLGRSLTWSMLKRGARNFALRS          GTDSKQJAILENLRAAGANVQFRGDASKEDVEKAVKEPADRPIKVINAAAM          VLRDGMFHMNTDFSWTSTNPKVHGSANHEVMKLDLDFVMTSSVSGTLTGP          GOSNYAAGNAYMDALRHRQTKRAAIVPMLIGVVAENASIEFALKRKM          YGVDEYVLDSEFAILEQTNPDAKDFHVLGDLPSLRKAKSEAFEGVDVFWAV          DPFSSLLHMETGSDVEGGADSVSRVKAATAAEANDVAQHEQIKSLRLLM          LDVEEFGGRSVASFSDMIGALRNLWIFKELGLDIAFQLGQLPSLTKRFSELVVA          NQGIQIE</p>
<p>HmCAR1</p>	<p>ATGGGGCCGGCCGAGATGGATGGCGATGCTATGACGCGGGGGCGTCCGAAGGCGT          AGCTAGAGCAAGCCGGTCTAGACGAGCTCGATGCACTGATTTCCGTTGGTCCAAAG          GAGTTCGGATTTGGCGCTGATCCCGCTTCACTCTTCAAGATACGGGTTGGTCTGCTG          GACTTGGATCCGAAAGTAAACGAAAGCTGTTGGGCGATCGACAACACGTAATGCCCT          TAAATCTGATCATCTGGTCAATACAAGTAAACCCCAAAAGGACATGAAGAAATAAGT          GGATGTAAGTTCGGTGTAGCTGCTGACTCCCAACGAGCCCTTCTGATGCGGTTAT          CTTAGGAATAAAGTAGTATAAAGTCCGCGAGACTTCTCAGATATATGATAAGACCGCC          ATCTCTTACTAGCGTAACTGGCAACAGAGTCTACTGACTTAAAGGAAACGAGTTAA          CTGGGTAGCTGATTTCTTCTTGGGCACTGGCATTGAAATTTCAACAATCCGCTAAT          ACAGTGTCTTCTAGCAAGTATGCTGCTGCAATGTTACTCTGGGGTGGCACTTACG          CTGCTGCGCGTGTACCAACTACGCGGGTGTAGTACTGCTGTTCTTCTGCGGGTATT          AGAAGCAGCTGTTGTCCAGCATGGGTAATTTACATCCAGTGGTATAGAAAGAGGAA          CAGGCAATTCGAGTGGCATATGTTTCTTACTCGGGGCGAGCAAAATGTTGGAGGCTT          TTTGCTACGCGTGTAGCGACGCTGGGAAAGCCAAACCTACTTTAAGGGTGG          CAGGCTACTTCTTATTTGGGTTACTACGCTGGTGTAGGAAATGCTTCTGCTGTTAATA          ATGCGGACTCTCCGCTGTTGCGGATTCTCCATAAGGAGCAGAAAGGACTTCACTTGA          GCGTATCAGACAATAAGCAAGGTTGATGATCAGACATCAAAATGGGCTGAGTTAGG          GAGGCGCTGACTGACACGATGACTGCTTACGTTTCTGGGCTTCTGCTGCTAACAATCC          GAACCTATCGCAACATCTTCCGAAATTTTGGTGAAGGAGTGGGCTATCGAATGAG          GAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG          GTATGCGGCGATGAAGCAAGGCAAGACTAATTTGATGCTGCTGCTGCTGCTGCTGCTG          ATAGTAGGCGCACTTCTGATGCGCAACACTGAGCGGGTCTGCTGCTGATGGATACTAC          AAGCGTGGTATCCCAATGCTGGACGACAACTTCTGCTGCTGAGATCCAACTGCTGCT          GGATCAACGAAAGGAGCTGACTGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG</p>	<p>MGAGPEMDGDAMPAGEVAVARDKPLDEPAMQSIIVRQSSIGADAALQLL          QDGTGLRRLDPEVNRRLRRIDKHMPLIICIVTGITESTLGNFENWVWASIFF          QLAFFETIRILQPLFLAKYASVNTLWGATLACLAACCTNYAGLLACRFVGLVLEAA          VPAWVLFSTQWYKEEQAFVGVIVFSTCGAGMGFFGAFVAVTHVFKDQPN          TURGWQVIFLFLGLLVVMGIAFWFIMPDPVAVGFLHKEQKHLHRIHNEQSL          GSQTFKWQVREALDWTNLYAFVFAANIPIASITFNLVKGMYGYSNEEGL          LLVPLGAWVEVALIGTYGAMKTRQLYFCIAGHPAIVGAILMATTVERPALIYVY          SGPIPIVWTLGLQSSNAGSTTKVTVACIGIATYTVNIISSPQTQFARDAPRYLPK          ISICIVLTLDLVMMRWLDKRNKREREKALNTYSPEDNHEFLDLQNLKFK          RYEV</p>



Gene ( <i>H. monticulosa</i> )	Predicted nucleotide sequence	Predicted protein sequence
HmCBL1	<p>ATGGCGACCTCGAGGACGACCGGTTGCTGACCCAGATGCTTCCGCCAAGGCGCTCCGCGCAC                      ACAATTCGTTGACCGAGAGGCTAAGTTTCATCCAGCTCATGTAGGATCCCAACATCAACAAT                      TCCCATCTTAACTACATATGACTATCTTAACTATCCAGCTACCTACCTACCTACCTA                      CTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAAGCATACTAACAAC                      GATGCTCTGCAGATGCAATGGGACGGCGTCCAGGGTTCTGCAACAAGAAATCGCGACGGG                      GAAGGATGCTGCTCTCCGCTCTCTCGACACTGACGGTAGTCTTCTACGGACGAT                      GCAGCTCGACGAGAAGGACCCATCTACGGGGGACGAAACAGCTCGCGCGCTGCGAGGGG                      GGAAGACCTGACTATCTCGCCAAACAGCGCGCGCTGTTGACCCAGCGCTGCTGCTGGA                      CGTGCATGATTGCACTCGCCCTCGCGGACAGTGTAGACGCTTGGCAATGGGGTTGGAG                      GGAGCCGAATAAGAGGGAGGAAAGTGGCGAAGAGGAGAAAGAGGAGGGTTGGTGGAA                      TAAGTTTTTGGTGGCTTGA</p>	<p>MATLEAAPVADPDAFAGELRRHNSLEKAKFIHMQWDEWVPGFMQTESRSEG                      CSVSPFDCTVLLYGRMLDKDEPIYGGNSVAVRRGGKTLILANSAAWVPS                      ACCWVDVDDCHLALGGQCRRLGIGVREPNKRERKVRKREKRWVWVNFGRS</p>
HmPKS3	<p>ATGGCGCATCTCAAGGCTACGCCACGCCAAGAAGCGAGTCCGTTCTTCTAAAGACT                      ATACACATAAATAAATGGGTTTCTTATCCCGATCCCGCCGCTAGCTATCGTCCGCTG                      GCTGTAGGTTTCCGGTGATGCTACGAGCCGCTCAATCATGCGACTTTGCAAAACGGCC                      AAGTTGGCGGTCGCCGATACCCGAGGCTAGACTCAACAAGTAGTGGCAAGAGGAGAGC                      ACAGAGAGAGTAGTCTCACTCCACCAAGCTTGGCCATCTCATCGTAAATGCTCGTCT                      TAGAATCACACATAGCGGCATTCTCAAGGACACACAAAGTCTATTGATGTCCCTTCT                      AAATCTTCTGTTGACAAAACGGGTTAGTTTCCGTTCCCTAAATGTCACTACTACTACT                      AACTCCACAGGTCACCGATCTCAAGCGAGACTATTACTGCAAAAGTATATAAAGTACCG                      AGGATGGTATGCTGACTTATCTCATGTGTCTAATAGTAACTAAACATGACAGCGGAA                      ACCAATCGAGAACCTGGCGGATCGAATACCGCTGTTTCTGGGAGGCTATGACACGCAAT                      AATAGCACCGATGCGGATGCGCTACATCCACAGGAAATCCAGGACATCAGGCGGAGC                      CTAGTCTCAAACTTCACTCAAGGCGGCGATGATGCAATGACACTGGAAGCTTCTCAGA                      CACTAGCCCTACACAGGTTGCAAACTCTCCGTTAGTGAAGCGGATTTCTATTGTCG                      GGGCTGTACATTAAATCAGGAGTTCGACGACGCTCAGAAGGAGTACCGGGGGAAG                      CGGTTGCAAGTGTGCTCAAACTCAGCGCTGTTAAAGATAAGGATGAATCATGCTG                      TATCATTGTAACACTGTCTCAACAAAGCGGAAAAACCGGAACTCCAGCTCCGCGGAA                      GCACAACTCAAGCTTACGAGGCTCAGCGCGCGCGGCTGATGCGAGCACTGCTT                      ATGTTGAAGCAGTATGCGGGGAGTGAAGTAGCAACGCTCCGAGATGGAAGCACTGCACA                      GGACTTCGCGAAGCGCGGGTTGCGAAGGCTATCTTCTGGTTCGGTTAAGCAAAATA                      TCGCAACAGCGAGCGGCTGAGTGGCTTAGCAGTATCATCAAGCGGCTAGCGATGAGA                      ACGGGTGTGACCGAGCTGAACCCCAATCCCAAGCAAGCAATGCGGCTTAAGTAA                      TTCTTACACTGCGTACGTTTACGAACTCTGCTAATGAAGTCAAGTGGCGCAAAAGTAA                      TTCTTGGCCAGAGACGAACTCGAGCTTCTTAAACAAATTTGGCGGACGGATCTAAC                      GCCACGCTCATATTGAGTGGCTAACCGGCTGCTAGCGCTTAAGTGAATAATCCCTCC                      GAGAAAAGCGGACAGTCCCGGCAAGTCCGCTTCTTCTTATGTCGAGAACTTCAAC                      AACGCGGAGAGTGGCGAAGACCTTAGTGTCTCAGTCCGCGGATTTGGAAGTGGCA                      AGCACTGGATCTAGTGCCTCTTATACGCTTGCAGCGGACGATCGGCTTCCATGGACC                      GTCAGTAGAGCAAGCAAGTCCGCGGCTGCTACCGGCTAGGGGAGCGGTAAGGCA                      AGTTCACCAAAAGGCTGCAATCGCTTCTGTTAATGGCCAAAGTGGCAATGGTAT                      GCTATGGCGGAGGCTCAGCGGAGTCTGCTTATCAGACGCTGCTAGGATGGCGGCA                      AGTTTAAATGGTGAAGTGCACATGCTTCCATATGTAAGAATAAAGCTTCAACATAA                      CTTTGAGAAATACTACCGCGACAGCAACTACTAAGGAGTGTCTCAAGCGGTTTCT                      TCAAGTATCTGGCAGACTGCTCAGTGTGCGCTTCAAGTTCGTTGCTGCTGACTAGAT                      CTGGGGAATGCTCCCGCAGCTTCTTCAAGCACTGAGTGGCAAGTTCAGCTGCTTACG                      CGCTGCTTCTTCAAGGAGCTCAGGAGTGTCTTCTGATGAGACTGCTAGCG                      AAGCTTGAATCAAAAGCGCTCCGACCGGAGTATGCTGCGCGGCTGACGAGCGGAC                      CAAGTTGAACCACTCCGCAATGAGGAGGAGTCTGCGCTACTTCTGTTTAAACAGCC                      CTGAGCGTGAAGTACGAGGACTGGGACCTCAAGCAATCTGCTGCTGACTGGAGA                      AGGATGGGATTTGCGGAGGAGCTCAAGTCTTCCGCTACTCCCAACATGTTGGTA                      TATGGCTCAAGAAATGCTGCGCTTAACTGCACTTCTTCCGCGCTAGTGGCTGGGA                      AAGCGCTATACCGCTCAGCGGACAGCGGATATTTAGTCCGCGGATCTTCCGCGCTGA                      ATACTGGTCCAAATCTCAGCGCCGCTTCTTCAAGTCAAGGCTCGAAGCTATGCTGCTG                      ATACAGAAATTTCCGCGCTCAGGCTTCAAGTGTGATGATGTTGAGAGTCCGCGCCACAG                      CACTCTGCTGCGCGGCTACCGCAATCTCAAGACTCGCATGATCCCTACACATCGTCCGA                      AGAGTCAAGAAATGCGCTCAACATGACGCGCTTGGCGGCGGAGCTTCAACCGCGGTTA                      CCTGTGCTTAAAGGAGTCACTTCCGCTCGCGGATAAGCGGACCGGACAGCTTCTG                      CCAATCTCCCACTACCGCTGGAGCGACGCTGACGACTGAGTCCAAAGCCCAAGACCA                      TCCGTCAGAGGAGTTGCGACGCTGAGTACTTGGACCCCACTACGACGAGCAGTGGTT                      GGTACAGAGTGGAAACTCGCTCCGCTTCCGATATGCTGATGCTGACTAAGGTA                      GACTGAAATGCTTCAACCGCGCGGATGATGCGCAATGGCGTGAAGCGCTGAGACTCC                      TGGCTGACCGAGCGGAAAGTCAACCGGATACAGCTTAGGAGAGTAGAGATCTCAACG                      CCTGGTAAATCCGGAATCGCGAGTAGGTTGAGACCACTGCGGACTACCCCTGCAAGCA                      AAAGGAGCTGATGCAAGGTTGATGACTCAATATCTTCAATGAACCGGATGGAGAC                      TGGTTAGCAATGTAAGGGAATGGTCTCCGAGCAGTAAGCGAAGCAGCAGCAGTACTG                      CCTCAAGCGCGGACTCAATGAGGAGGCTTCTTCAAGGGGCACTAAAGCGCGCGTATTT                      CCGTATCTCGCTACAGAGGATCTCGGAAATGGAAATGAGTATGGCGCGCTTCCAGAA                      CTTGATGGAAAGCGAGCAGCAACAACTGCTGAGCAGTATGCTTCAAGTCCCTGAG                      CCAAGATCAAGAACAGCTGAGTATGCTTCCCAACGAGCCTGATTCATCTCAGG                      TACTATAGCGGCTTACCGTACGCTAAGCGAGTACGACACTTCCCTAAGCTGTTCCGGA                      ATCTGATATCTCAGCGATCTTGAAGGATCAGTGGCCAAATGAAAGCTTCCGCAATCG                      GACCAAGCAGAGAAAGAGGCTTACTTCAAGGCTGACGCTGTTAAACACGACGCAAAAG                      AGGTTCTCAGATCGATGGCTTCTGCAAGTCAATCCCAACTCCGAGGAGATCTG                      AAGAGAGCAAGCACTCTGCTCAAGACCACTGGGAGCGGAGCTACCGAGTAC                      CGGCGTGGTCAAGGAGTCAAGGATTTATCTCGACGCGAGATGCGGAATCGAAGA                      AGATGGTCCGAGCTCATACTTATGATCATGACGCGTGCAGAACTGAAAGCGCAAAACC                      TGAGACTTGGTGTATCAGAGGAAATGCAAGTGGATGAAACTGCTGCGCCCAAGCC                      AAGCGGAGAACTTGGCACTCTGATGACACTTGGGAAAGCGCCACCGGCTCAAGCAG                      CTGATACGACCACTCAATCTTGGCGTCTGCGCGACTGTTCCGCGTGGTTCTCA                      ACTCGTGGTATGCTGCGGAGGAGTCTTCCCGGAAATTTAAAGAACCGCAATCTTCT                      AGTCAATTTCCGCGAGCTACCAAGCTGAGAGACCGCACTATAAGCAACTTCAAGGCTG                      CGGAACTCATGCGGTGCAAGTCCCGGCTAATGCTGCTGAACTGAGGCTGACCGGAG                      GTGTTTCCAGGCTTCTTCAAGCTTCCGCGGAGGCAAGTGAATCGGAACTTCTT                      GGCACTACACACAGCACTCAAGATGATGCTTACAGGCGCTCAAGCTTGGCTGCGC                      CATGGGAGATATGTACAGTTCCAGAGCTGACATGTTGCAAGCTGGCTAAGCAGTCTG                      TCAAGGGGGCAATACGACTAATGCTGACTTGGCGCTTAACTACTAGCGTCAAG                      GAATGCACTACATACAGGACTACTCAAGCTTATGGAAGCTTATCTTATTAGGCCCA                      CAAGCAATAGCTGACATGAGCTCTGCTGCAAGCTCAGCGAGTGGGCTCAATGACGA                      GCCGATAGCTGAGCCGATCTTCCCTCAAGCTGGGATGATACCTTAGGAAACTGCG                      TCACTGCTGATGATGATGATGAGGACTGCGAGCAGCAGAGTCCAGGTTACCGATCA                      TTTTACCGGCTACAGTACAGTACTGACTGAGGCTGTTGATGCTGACACCGCTGTT                      CCTGCAAGCAATGGCTGAGACTCTTCTGCAATAGAGGGCAAGCGCTGCTCCGCT                      TGTGGAGCATAGAGAAGCGCCAGCGGAGCAGGATCTGATATTACTGCGGAGTCT                      CGGACTGATGACAGTGGCGAAAAGGATTCGAACTCTACGCGCTTACTTCCGAGC                      ACGCGAGTGTCTATGTTGAGCAGTGGCGGCTTATGACGACGCGGCTTCTTCCCA                      AGATCAGGAGCTGCTGCAAGCTAAGAGTGAAGACCCCAAGAGATACGCTCACTGCA                      CTTGAGTATGCAAGAACCGTGGTGCAGCACAACATTTATCATATTACGTTTAAAGC                      ATGTTTGACTTGGCCCAACCGCTGGCAGTATGGGAATACTCCGTAAGGGGCTGCT                      GCTACTGCTCCACGCTACGCGGACTTGGAGCAAGTCCGCTGGTCAAGCGGACCGGCT                      GATCCGCGCCAGGCTCAACGCTTCTCAATGGAGGCTTCTGTTGAAACCTTGGC                      TACAACGATCCCAACTTCTGCTTGGTGAACATGAGGAGTACGCTGATATCTGCGG                      GTATGCTGAGATCAAGCAAGCTTGGTCTCAACCGCGAGATACCCGTTGATGAT                      CGAGGACCGGCTGGCAGCATGATCTCCGATGTTGTTGCGACTGGGCGGAGTACTAAA                      CAGAGCGGCTGAAGTGGCGACCGGCTACAGGACTTCAAGGCTGCTGCTCAATGTA                      AGCGGTCGCGTGGAGCAGTCCGCAAGTCCCGCGCAATGCTCTGAGACAGCGCGG                      GCTCCCAACCGCAATTTACTGCTACTACTGCTGCTCATGTCGCAAGCTCAAGGAG                      AGAATCCATTTGATCCAAATGGCGAAGGATTTGGCAAGCAGACTACCTTCCCA                      TATATTGAGCAAGATTTGTTCTGCTGCTGAGGAGGCGAATTCGCTAGTAGAG                      AGTACGCTTACCCCAAGGCTTCTAGGACAGGAGTGAATTCGCGGAGGATATCA                      TGGCTCAGACCGGAGTGGTGTGACGCTGCACTCACTGCTGCTGCTGCGGCTGCTTAA                      GGCAACTGGGAATGATCGACTGCTGCGCGCTTGGATATCGTGAAGGACCAATCGAAC                      ACTCGAAGCGCTGCACTGACCTTCCGCGCTGCTGCACTACCAAGCTGGATATCC                      TACAGTATGAGTATGAGCGAAGTCTGACGAGGACTTCAAGGACTTACTGATCTG</p>	<p>MARISRPTPRKNGVFPKSDYHTNNNGVSVSRPVAIVMACRFAGDATSPS                      NLWDLCAAGVGRSRIPEAVDSQVQDEESTERNHTSHGFLKDNSTSDVAFSN                      LPVDKTVDPQARLLVSYQATEAGIPIENLADSNVFLGGYDQVQNSDAVL                      PSYPTGKSRSTGASLVNFFNLQGSMSIDTSSSSDLAALHQGQTLRLGADVIV                      GACTLLNQFDDSSGDRREGVAVLVIKSLDAALKDKDRHAIIRNTLNQSGKT                      GTSVSAEQIKLIEDYCRAGLDMADTAYVSMASGEVANAIEALDRTFKSR                      GSEEPFVSGVKNQINFRSLGADIIKAALAMQNLVAPSLNPISTWQVHVK                      PNKLPWPRDKLRASINFRGDSNAHVIDGAPNARRLSGNSLREKAAQSPD                      KSRVFLSARDSTAEVMAKNLSAHLRRLREGAPGSSSLAYLATRIRSRFPWVT                      MRASNVELATGLGPEVKAVHSTKEPRIFVFNQGGAGQWYAMGRELIEYVFR                      AIEDADKVLNGYATWLSYDELHRESSRVSQVLAQSVYLIHQVLRLESVIV                      PHAVSSHSVGAAYAGLSFKALGVYFRDLGLAKLESQASRPGMLAALG                      GPDQVEPYLANTEGGRAVIACVNSPESVTLAGDLAINEVRLHIDDAIEGDKN                      AYHSHMLDMAQYAGALTAILPRPSPAKALYASPTGDIESPILTEPWVQ                      NLPDPLVLSQALEAMCFDIIQATYSQASNDVMDLVRLEPSTKFRMLHSRILK                      TSLKRENAHVMTQALAGELNRPVSLKENVFPLDNDGQPTVFNPLPYW                      SHASTTESKATIKTRVLEKMRHLGRDHLAASSSGLVWRNSLRSDIAEHDKVD                      SNVPLVAGVAVAMAEVRLADPAEKSTRGKRDVEILNVLIPDSVSTHRL                      LTPCEKELDYGWVDFNFRSLGADIIKAALAMQNLVAPSLNPISTWQVHVK                      NEEAFFPRGTAKARRISVSLQDLRKMIEYGFAPQLNIGSQAAANSASSMFINRP                      MPKIKNLQVLYHPTDIIQATYSQASNDVMDLVRLEPSTKFRMLHSRILK                      KAFANRTKEKGLTSSVTLNNDNEGLQIDGFCQSPHPEEISEEQTLVCT                      HWEPDVRVYRSPVLEKMRHLGRDHLAASSSGLVWRNSLRSDIAEHDKVD                      ASYQKELHVKMKTVAQAKRNLAPLSTWENATSPRLKYLVDLQNTSVGAVRT                      VYRVSQLAGIVRGSVQVLAQFAGGNSGGLTGLSYTYDLDLHGAHQRLA                      PWDGMVQFKLIDGQAKSFKYSGEYVLEPGLRTRVYKLSLHSLKLDLIDG                      KLILEPTSNKLDMLQTLPSWVNDPDKLSPILSQWDDTDRTEFTGDFD                      IGDCEQEPFQSTLTLGLQSLVYVHTVAVPDKVWLVHLSAIRGQTFAPV                      VESINAQEPDRICFAEMLPGLDMSGKFTLRFRSRSRVLVSSGGVDA                      APSFKIQGLRLTRVLEKMRHLGRDHLAASSSGLVWRNSLRSDIAEHDKVD                      AGIDWEYSVKSGLVYRPIADLETSAVSDRVDPPQPFQLERSLWVPLAT                      NDPHNFCVNEEDLSTPIAGMVEPEKAFGLNDRIPVDEIETASHLSDVIVRL                      GPDTQSGLVKDRVYGLAKRNLVSRAPVSIKAVPEMFSFAAALPHTAHTA                      YSLLHVARLQAGESLHNAASDVQAAATLQYIGAKIFRLLYKLSLHSLKLDLIDG                      IDPTRLVSKANFARDIMAGTGGVVDALNSLGLKATWECIAGFRGLVDIR                      TNSNNSKRLDMPFRSARSYVDIQLCEFRSGLTQELTRICFTNSGRTHIPI                      RSVPISELAAGHKEETHFGKSIIVPTEDQVNVIPRSSLNSQYETMVAAGSG                      EVNHAITSWLEKRRNVVSHDAESNLAAVLOQESAGGSCNIRHNDACEKSL                      VKLLKELAGLPIRIVDNTLIVNATASEHVSAGTWNLHKLHPLDFMFLSSAG                      VTGHPSQATYADQAFDRALRHRIARGPLAVSLDPLDSTAIETAQANEMTSLD                      MDKVLRLVEAATHSLKHGPDACVIVGLQWPDQISSLDATIARADPFGTLQAVP                      RATSSATAPTEGSMVGTPTDLQQLKLSSEDSIKLATEVAARLAEILLNVDAEIG                      HRDASIMSHGVDSLSAVERINWLTGVAKAKVIEILRDLPLFSLVLSRSEAGKE                      AAA</p>





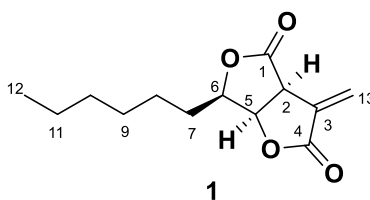


	<p>TATATTCTGCAAAATACAAAGCAACAGTCGCAAGGCAATACCGAAAACCTGACAGGTCATA          GTTGGATTGGACGAGGAAGATAACCATCTGCGCAATGGTCTCCATTACACCCAGATGCCCGC          GATGAGCTCAGCGCAATATTGACGGAGTCCAGATCTGACAGATCTTAGATTCTTAATACGAT          ACTCTGTGTCAGAAATAAAGTGTATGTTGAGTTGCTACGTATACTGCTGTGTCGTCGGCTA          TTCGTAATATTTTCCGGTATAGGATGACCAAGCAGTACACTGCGCAATGATGATGATC          TCGAATCTTCTCAACCCGGCGGAGCTCAAGAGCCCTGCTATATGTAATGGAGCTCAT          CCTCTAACAAAATTTGCAACTGCTAAGTGGCGTGGTTATCCGTAATGACTTGTCACTA          ACATAAATCTGAGTGGTCTCTGAGTACCAGCGGTTGGACTATAGGACGCCAACCG          GAGTACAGAGGTGACTTTGCGGTCTATCCTAAGGATTGTTACACGCTGTCTATTCTCT          CCTCGCCCTCTATACCTTGAACAATACCGCGGCTGTGCTGGCCGATATACGAAGCAGT          GTGATGATTTGGGAGACCTATCAGGAATTTAGCCGATTGACCGCCGGGCTCTTGGT          TCGTGTGACCTATCTTGTGATTAAGACCGGCTGTGTAAGGGAGGATCAACGACTTT          GGAAGCACTTGGCCGCTATCCGAGTACCAAAAGTGGCGCTCATAGTGGCTCGGGA          GTCGAGGCAATGGCTAGCAAAATAGCAAGGAATGCGACTAGGCTTTGTAACCTCTA          GCTCGGATGATCTCTCTAGACAGTTGACCGCATCCCTTTCTACCGGTGGTTGAGCC          CGCGAATTTGCCAGATGCGCTATATCCGGGCTGGTGTAGAGACTGAGTGGGAAGCGG          TCGAAGGTTTACCAAGCTACTCAAGCCGCTATCTGAATCTGGAGCAGTACAACCC          GCGCCGGAGTGCAGTACGATAGCCGTCGCGGAGGATCGATGAAAGTCTGTAGCGA          GTTTTGGCAACTTCCGTCAGCATTGCTGTCGAGCGGAATACAGTACGCAAGCGTCTT          ACAAACTGCTCTAACAAGCACTCTACATATAGCTATCTTGACTACTGTTGAAATTTCC          GGCAATGCTATTACGCGGCGAATGATACCAAGCTTACCGCATACAAGCGAGTCTCT          ATCAACAGAAAAGGCACTAGCAGTGGCGCTCTCCGCTCTCGATAACCTCAACCTTTA          CACGCGATCTTGGGGTGGCACTACTGATTCCCTGGCTGTATTCTGACTTCCGAGCGG          CAGTACTATAGGCTACTATGATGATCAGCTTTCCGGGAGCGGTAAGTACGAGTCTT          AATAAATCTAAAACAGATCCCTTCCGCTGGCATGCGCACTGACGCGCAGCTTGCATA          CAGGGGTACATGATGCTAAGCCGCTTCCGATGCTAGCAGAGTTAGCCACAGCGGAA          GGAGGGGCAATACGATAACAGCACTATTGGAATTTGAATAAGTGGAGGTAATAAGGAG          GCGGTGAGATAGCAAAATCCAGCAGGCTTACAACAGTCCAGTAAGCCAGGTGAGCGAGGT          ACTAAGCTGGTCTGCTGAGCCAGGATCAAAATCTGTTGAATGAAATTTCTCTAGTACCAT          GTCCAAAGATTACCCAACTGGGATACGATACCCGATTACGTCATA</p>	<p>CSEFLATPSAFALQPNATWDKRLHLKPLRQLLHIAFDLSCWNRPMMLRGNPDI          QSLPAYKRVLLTSKQKALAAVRLDNLVTLHMLGGCHTRFPGLVFSTFAEAVLL          AYLXMDPRFPDGDNDLVLKSTDRPRAGMASLRTDCIQGVHDLRLRMLAEVS          HTAEGGANTIRLLEYLNKVEVNEKEAVIANIQAYNSPVQSVSEVNWNLCEPTDP          NSLNEFLSMTSQDLHPNWDPIDFPH</p>
<p>HmCBR5</p>	<p>ATGAAGTTCTGCTGACGAAAGCGGCTGCTGCGCTCTCATCTCCAGACCTGTTGGCTGTAG          CGCTATTCTGCTGCTCCTGGACGCTCAATACCTCCGACCCACGGCTTCCCGTGGCGACT          GTCTTTACACAGTTCAGCAAGAACTGTTCCCGACGCTGTCAACGGTCTCCTCTTTGGAC          TCTCAGACCCCTCGTGTACCGCCATCGAGGATACAGCAGCAGCTATCCCGATGTCGA          GATCGTGTGACGCGCCGACCGAGAAGGATGTTCACTAATTGTATGTTTCACTACAGCTTCA          CATCTTACAGAGCTTAGCTAATAATAATTAGTCAAACTGCAATGACACAGCATTGAGT          TCTGGCTGTACAGCTGCGCACTTAGCTTACAGCTGACGACTTCAAGGGATGCAAAAT          CGACATGCGCGCTATTGGACTACCACTCAGCCGATGGAAGTCTGATGTTCCAGGCG          GGTACTATGACGCGCAAGTTATGAGTACTTGGGAGCGCGCTACTGCGGACCACTGGC          AGCTGACGCTGTGCTGTATGATGGAGCTGTTGGCGCGGTCAGGAGCGCAAGAAAGT          TTCTATGCGATGATCGAGATAACCTGCGCAACTGAACGTTGCTTGGCTGATGGCACTGGC          CCGAGTCAACTGACAGCCAGCCGCTTACTGTTGGGATGAAGGGCTGGCCCAACTCT          GGTATGCTACCACTGTTGATGAACTACTACCTGTTGAGGTCGACTGACTGCACTACCAAC          CTACACTGGAAGGTGACAAGCTGCACTGCTCAACGCTCTTAAACGCTTCAAGGCAAT          GGCAACCCCTGTTAACTAGGCTTAACTCGGCTCATCTCTCAACACAGCGTTTCTACC          ACTGAGCCAGTCTGTTGGACTTTGCTTCAAGGGAAGTCTGAAAGGTAACAAGGTC          CTGAAGCCTTCAACGACTTGTGCGGATGTAAGAGTTCGGCGATGTCCTCCACCCCAAA          TCTGATATGAGGGCACTGTTGTTGGCGCTTTCGCGCAAGACGCTAGCCACTAC          CTCACCGTCAACTGCTTACTACAACCTGACTGCTGAGCGCAAACTCAACCGCTTCAAGG          ACTGGATTAAGGAATACCCGAGCTGGGACCACTGCTCAGATTGTCACGAGGTTACTCTAC          TGAGGCGCTCGAAGTTCGCGCTGACGACTGCTGCTCCCTTCCGTCGCGGAGCTCATCTA          TGAAGTACTACTTCTGCAAGCACTTCAAGTCTAGCAAGCTAACATACTGAACAGGCT          CTTCGATGTCAGATCCCACTGAGAACCTCGGGAAATACTTCAACCTGTTCCCGGAGAT          GGGCCAGGAAGTCCAGACATGTTGAAGCAAGGCAAGCCACTCGCTCCCTGTTGCTTATG          CAACACGCAAGGCTGCGAGGTCGAAGATGTTGATGAGTCAAGGAGTGGCGGCGCAGGA          CCGTCTTCCGCTCAAGAAGAGTACGACCCCAAGACCGCTTCCGCTTCAACCCCACTG          TCTCGAAGTACCCACCGCTAA</p>	<p>MKFSQAQVLFSLQTLVAGSAPRAPGTQYFRPHFGTRRDLVSTVQVQELGP          QLSNGSLFGSPDRWYAAIERSTHAIPDVEVVOQATEKDVSTVYKXNDNSIEFL          AVNRHSRSYSVAAFKMQMDIAGLLDITQDPDGLVWFGGDTYVQDMVEYLVW          ERGVYATGSCVGMMPGLGGHGRQEGFYGMISDNLRLNVLVADRAVR          VNSTHADLLWGMKAGHNFVITFELNIPREVDVSWHYHTYKWDKLDITVFN          ALNLKHNGTTPVNMVNFSGFLNLSVSTTEASLWVFTGKGTAEANKLVKPF          NDIGAEYEEFDPVYQJSDMQGTGVLCAKNASHTSTNLLYTLFAEQIYN          RFSDWIKEYPELGTQAVIHQESIEAVDKFPDADAFPRRDLMLFDVQIPTEN          PRGINFTTVAREWAEVQTMWNEQPTRIPGAVYNYANGLEPKMWWYHEQW          RQDRLLALKKYDPQNRFRNYPVSEATTA</p>
<p>HmCBR6</p>	<p>ATGAAAAGCAACTATTGCTGTCGGCATAGCAACGCTGATCAGGAAGTGGCGCTACCTTTA          TAGTATACGAAGACGGTGTATCAGATTGCTTATCATAGTTACCCGAGATCTTTCGCGCACCA          AATTCGTCGAAGGTTTCTATTAGTGTGCTTTCTGTCGGCGAGTGTCTTTCGCGGCACTC          CGCGGTAAGGCTTAACTCTGACCGGATGTAAGCAAGGATCATGAGGTTAACTGAA          TGTACGACCGCAAGCCGCGGCGGTTTCTGCTACCAAGCTGACGATAATCCCGCTCCCA          ATGATCCGACACTAGGACAAAGCTAATGCTCAACAGTACGAGAATAGCTGTGGCGT          GATAACGTTGGCTATGACGGATATCCCGCAGGAAACAGATATGACAGATATCATATG          TAGACATCTAGTAACTGGCTAGCCAGTTGCTGCTTGAATTAATAAAGCTTCTGCTG          TCTGCTTTTGAAGTCTGATGATGTTGTTGCTAAGAAAGTACTGACTGAGCTCCGCT          TGGCAGTAAAGCTACGCAAGTCCCTTACCAAGGCAATGGCGTGTGCTACCAAGCCCAT          ATGCACTTCCAGTATGTTGCTGATGCTTATGCGCTATGATGCTTCCCTCCCAAGCCAT          CCGGGCATCCCATACAACCAAGGCAAGGCAAGGATGCTTCCGAGATGCCCGGATGTTG          CGCAAGTGAAGCTTACGAGAGCTGCGAGTGGTGTGCAAGCAAGTGAAGAAAGTGA          CTCGCTATTGCAAGGTTCTGCTGCTTTCGAAACCGTGGATTAATCCGCGATTTCCG          TGAGGCCGAGATCTGACACGACGAAAGGAAATCGCAAACTCTTCTCTATAACCGG          ATGGCTCTATGGAGACTTCCAGCTATCTACAAGCAGCGGAGGTTTCAAGGCCAAGAC          ACTTATCCAGGATCTATGACTCTAGCTTCTCAACACTGCTGCTGCTGCTGCTCAGCCCA          AGGGCTGAGGCTTATCAAGCTTTGAGACTAAGTAAAGTGGCAAGGGCGCGCATCA          GCTCAAGAGCGAATTTGAATGATCACTACGCTGATGTTAAGTTTCTTTAGCAATAAC          TGGGTAAGACTGCCATTGACTCAGTGAACCTTAAAGCAGTAAACCTTCTGAGATT          CAGACGCAAGCCGATGAGCGTGAACCTCCCTAAGCCCGCTGATGACTTCTCATGGC          GATCTATGAGCAGCCGAGTGGTGGAGGATGATCAATCCCGCGCGCAAGTTACGCT          GTGTGGTGGAAAAGCAAGCTGCTACAAGAAATTTGCAAGTAAAGGATGCTGCTCCG          GGAACAAGTGGCTATAGGATCGAAACTATCTGCGCGGCGGGTGAATCGGTTAGGAC          ACATGCTGATGCGCAAGGAGGCGGAGGAGCAAGGCAAGGCGGACTCAATTTTCAAGT          AGTTCTAGTGCATGAGGTGAGCTTCACTCCCGCTTGAAGCGCGGCAAGAAAGAAATG          ACTAATGATTTGCTGCTTATGATCTGCGGATATAATTTGGCGTCACTACACTGAGCGCG          CTATGATGCTGCTTAAATCTCAACGATCAAGGCGGATGACTAAGCTAAGGAGCA          GTTACAGAACTGCGCAAGGCTGCAAGGAAAGTGGCTCCGACTTGAAGGAGCTCGC          ATGGGCGAAAATCCCTATATGAGGCTATTATGAAAGATGCTAAGACTGAATGCAAGTCA          GTGACAGGGAGGCACTTGGCAACCCAAATCTCGGCAACCACTCCGCAAGGGGACACAA          GTTTCTGCTGCTCAACGGCCGGAATTTCTGCTGCGCTGATGCTTATTGACGACTATTG          CAGCGAGACTCGCTGCGCGCAAGATCGGCAAGTGGATGAGACGAGGACTTAAAGCT          CTTGACCCGAGCGCTGGCTGCTACAAGACGAGCAAGAAAGGGTGGAGACCGTAAAT          CGACGGCGCGGAGTCCGCACTGTTGCGATTGGGCTTCCGCTGCTGCTGGGCGAGG          GCTAGCGACATGAGAGTGAAGACTATAATCTCATGCTAGTGTGCACTGAGCTGCTACCG          ACGCCCAAGGCTATCCAGCTACTCGGCTCGAGGAAATGCGCGGCTGCCAGATGTTGCT          ATATAAGCCGAAAGAACTATA</p>	<p>MKSNYRHLGSHNADQEVLTFFVIERCFYFAYHSSRSPFRPNVSEVGLVYFRG          PSVLRPTSAVKASLDRIFRLSTNVDNPNANGSDTRDKANVSPINHEVPLPQAI          GLVSRVFGFISVAVLYGLYRLLPKPPIPIYQKATMLFGDAPDMVREKVP          GELRWCAKQVKNLSPICQVFLVPSKWLILADREARDLIRKRFKDFSSFLN          MAMPDFHIGVYGFAGKAMRQLQDMLTFLFNLLVWKKQAYKFIKDFK          MKLAKRPFVSDYFVSLDMLVSAFNSNWWWKTAIGPOLLEQSMNPSIPDAS          PDEPLTKPAPVDDFLMAYEAPVEKLNAPKRLTWVWWWKQAYKFIKDFK          DRVLREQVAIAIENYRGGVRESIEHMLRREAREAKQGRPLNFOSVLDIPEIFD          IGGHHTSGAMMVLKYLTVHQPALRKLAKLHLEPTELRKLFVTEELRWAK          IYMEAIIEMLRLNAVTVTREALDQTLGHIPKIPQVTFVLSYNGFLPSRMPIDD          SLRSTEAARKIRATWDETQDLVDFPDRVLETLVIAAGGDKAFPPFLKFLDVAFIG          LGPRACWGRRLAHMERMISMLVWHFLLTPQALSYSYSGELIARVPMQCYIRP          KKL</p>
<p>HmCBR7</p>	<p>ATGGCTTCAAGTAAACCACTTCTTATTATTGGTGCACCAACCGGCTGCGGAGGCGCTCG          CCGCTGTTTACCGCTTGGCAAGAGGCTCATCATCAGGACGAAGGCGAGGAGGCTGG          ATGCTTGGCGCTGAACTGAAAGGTTGTAACCCGCGAGGTAGGACTTTTATGAGGCGAAA          TCAAATCTATGCTTTGGCTAATGTCAGTTGTAAAGTTGACACTGCGGCACTTCCGCTCT          CCGGCTCAGTATCGCTACTCAAGGACTACCCTAAGTACAGACCGCTCTATGCTCAACCGG          TATCAGCAAGTCTACAACATTTTTGACAACCTCTCATCAACAGAAAGGTCGAAGCGAG          GTTGCATCAACTGACTGCCCTAACTCTGCGCAACCTCTTCCGCCCCCACTCTCAACTC          GCCAAGTCCGCAACCAAGACCACTTCTCATCAACCTCTCTGCTGCTACTTCCCTCAGC          TTTACCCGACATACTGCGCACGAAGGCGGCTGCTGAGGACTTCTGCAAGATCTTCCGACG          AACTGGCTTCCGCGGAGGCTGCTCAGAACATGAACGTTGTCAGATCTTCCCGCTACTG          AGACACTGGCTGACGCGCGCAGCTGACTACAGCTGCGCGGAGGTTGTAAGGACAA          AGCTTCCGCGCAGCGCTGAAGGAGTCTTCCGACGCGGTTCCCGCGCATCGAGGACTC          GTTCCGCGAGCTTCAAGAGGAGATGCCGCTGCGCTTCCGCGGCTGCGGAGCTGGGCTC          TGGAGGCTGCTTCCGAGAGGCTACGAGTCTATGCGGCTGACCACTTAA</p>	<p>MASRINTLIGATTGIGELARRFHAGLKVIITRRQRDLDAELKGVETRFQDI          GDIALPGLVSAIKDYKPLDVTYVWAGIQYQYKIFNNSITNEKVAEVAINTLTPN          LLANLFAHLLNLAQSKTHTTITSSLYIPFSPYFPCATKAGLQAFKFRQQLFAE          GEGAQNMNVVIEVPPVDTDLDAHRDVTIAAGGDKAFPPFLKFLDVAFIG          EDVPGDGSKEIAVFGELGVWTRGAFKVEYISGIMTI</p>

**Table 7.3** The predicted genes and corresponding nucleotides as well as protein sequences (re-annotate based on RNA transcriptome data) of the 'Cluster B' from *H. monticulosa* MUCL 54604.

Extract	Gradient (Preparation LCMS)	Targeted compound	Retention time* (min) 10-90%, 15 min Ana.	Amount (exp.) mg / L media
<i>H. monticulosa</i>	10-60-90%, 15 min Prep.	<b>1</b>	7.70 - 7.90	14
<i>H. spongiphila</i>	10-60-90%, 15 min Prep.	<b>3b</b>	8.85 - 8.95	1.2
<i>H. submonticulosa</i>	Flash Chromatogram (Petroleum ether/Ethyl acetate: 9/1 – 8/1)	<b>2</b>	5.30 - 5.40	23
<i>H. monticulosa</i>	10-60-90%, 15 min Prep.	<b>74</b>	7.90 - 7.95	12
<i>H. spongiphila</i>	10-60-90%, 15 min Prep.	<b>76</b>	7.40 - 7.50	9
<i>H. monticulosa</i>	10-60-90%, 15 min Prep.	<b>77</b>	6.20 - 6.40	4
<i>H. monticulosa</i>	10-60-90%, 15 min Prep.	<b>78</b>	6.00 - 6.20	4
<i>H. monticulosa</i>	10-60-90%, 15 min Prep.	<b>75</b>	5.70 - 5.80	5.5
<i>A. oryzae</i> ; EXP8	10-30-80%, 15 min Prep.	<b>92</b>	7.60 - 7.70	8
<i>A. oryzae</i> ; EXP5	10-30-80%, 15 min Prep.	<b>91</b>	6.40 - 6.60	3
<i>A. oryzae</i> ; EXP5	10-30-80%, 15 min Prep.	<b>90</b>	7.00 - 7.20	7.5
Originated from compound <b>88</b>	-	<b>93</b>	7.40 - 7.60	7
<i>A. oryzae</i> ; EXP4	10%-90%, 15 min Prep.	<b>83</b>	7.65 - 7.80	4
SpoG <i>in vitro</i> assay	-	<b>87</b>	5.40 - 5.60	/
<i>A. oryzae</i> ; EXP5	10-30-80%, 15 min Prep.	<b>89</b>	4.95 - 5.15	6
<i>A. oryzae</i> ; EXP5	10-30-80%, 15 min Prep.	<b>88</b>	5.90 - 6.30	7
Spol <i>in vitro</i> assay	10%-90%, 15 min Prep.	<b>94</b>	5.82 - 5.92	/
<i>A. oryzae</i> ; EXPt1	10-60-90%, 15 min Prep.	<b>142</b>	5.38 - 5.48	2
<i>A. oryzae</i> ; EXPt4	10%-90%, 15 min Prep.	<b>143</b>	5.10 - 5.30	50
<i>H. spongiphila</i>	10-90%, 15 min Prep.	<b>82</b>	4.50 - 4.60	5

**Table 7.4** Compounds LCMS purification details. Exp., indicates experimentally purified amounts. \*Retention times slightly over time.

Compound **1**<sup>51</sup>

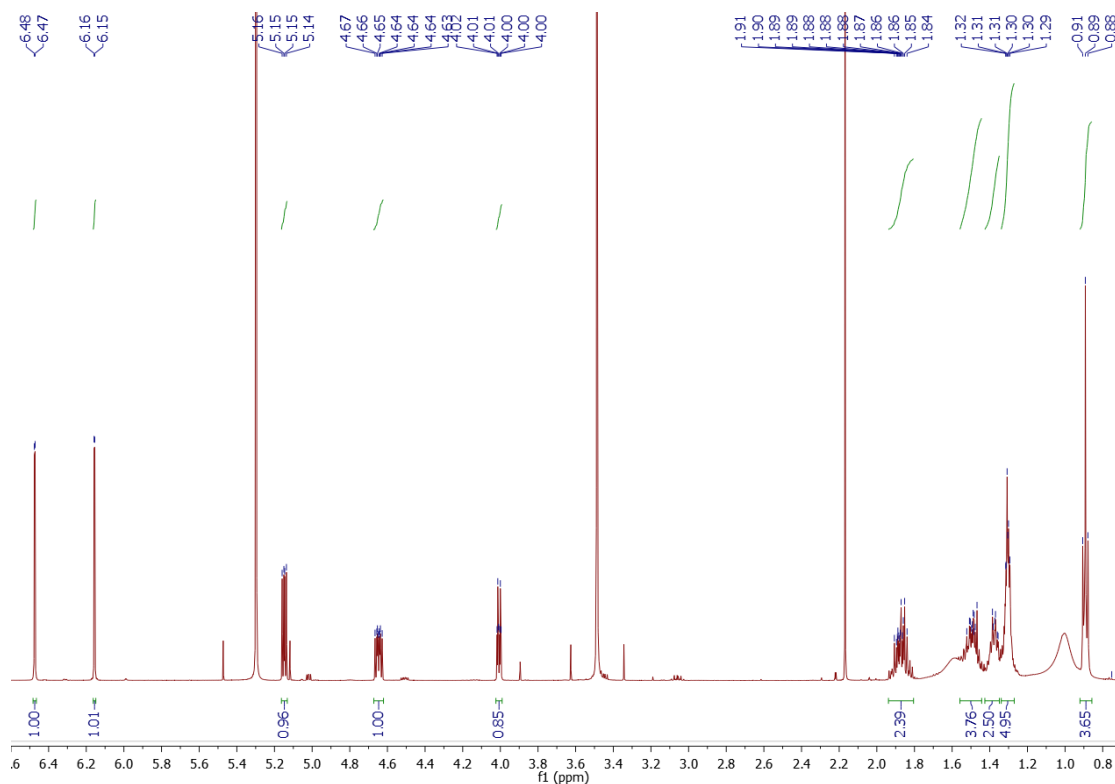
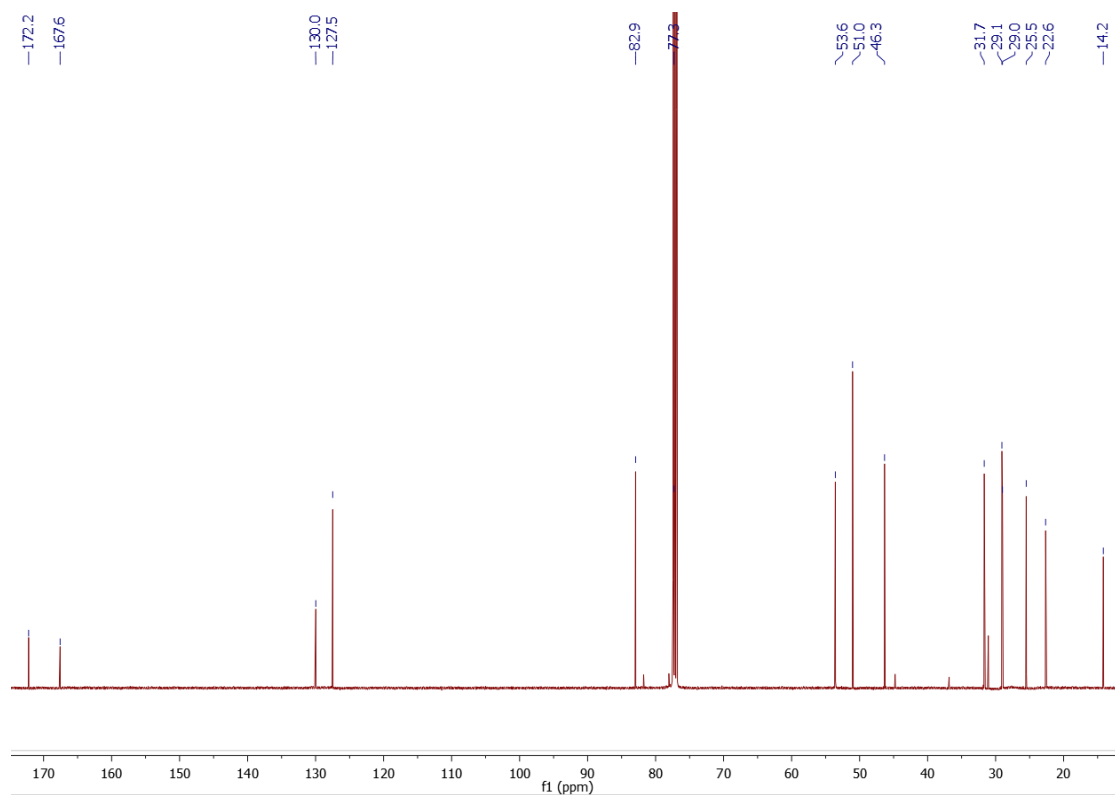
## Sporothriolide

Chemical Formula:  $C_{13}H_{18}O_4$ 

Colorless oil;  $[\alpha]_D^{26} -104$  ( $c = 0.5$ ,  $CH_3OH$ ); Literature,<sup>51</sup>  $[\alpha]_D^{25} -144$ , ( $c = 3.3$ ,  $CHCl_3$ ); UV ( $\lambda_{max}$ ): 206 nm.  $^{13}C$  NMR data ( $CDCl_3$ , 125 MHz):  $\delta_c$  172.2 (C-4), 167.6 (C-1), 130.0 (C-3), 127.5 (C-13), 82.9 (C-6), 77.3 (C-5), 46.3 (C-2), 31.7 (C-10), 29.1 (C-9), 29.0 (C-7), 25.5 (C-8), 22.6 (C-11), 14.2 (C-12);  $^1H$  NMR data ( $CDCl_3$ , 500 MHz):  $\delta_H$  6.47 (1H, d,  $J = 2.1$  Hz, H-13a), 6.16 (1H, d,  $J = 2.1$  Hz, H-13b), 5.15 (1H, dd,  $J = 6.8, 4.7$  Hz, H-5), 4.65 (1H, ddd,  $J = 7.9, 6.3, 4.7$  Hz, H-6), 4.01 (1H, dt,  $J = 6.8, 2.0$  Hz, H-2), 1.88 (2H, m, H-7), 1.49 (2H, m, H-8), 1.38 (2H, m, H-9), 1.30 (2H, m, H-10), 1.30 (2H, m, H-11), 0.89 (3H, t,  $J = 7.0$  Hz, H-12); NMR data (Table 2.13 and Table 7.5) are consistent with those previously reported;<sup>51</sup> ESI-MS  $m/z$  237  $[M - H]^-$ , 239  $[M + H]^+$ , 477  $[2M + H]^+$ ; HR-ESI-MS  $m/z$  237.1127  $[M - H]^-$  (calcd. for  $C_{13}H_{17}O_4$ , 237.1127).

pos.	$\delta_c$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_c$ / ppm literature <sup>51</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>51</sup>
<b>1</b>	167.6	-	167.5	-
<b>2</b>	46.3	4.01, 1H, dt (6.8, 2.0)	46.2	4.00, 1H, dt (6.7, 2.1)
<b>3</b>	130.0	-	129.9	-
<b>4</b>	172.2	-	172.1	-
<b>5</b>	77.3	5.15, 1H, dd (6.8, 4.7)	77.2	5.14, 1H, dd (6.7, 4.6)
<b>6</b>	82.9	4.65, 1H, ddd (7.9, 6.3, 4.7)	82.8	4.64, 1H, ddd (8.0, 6.3, 4.6)
<b>7</b>	29.0	1.88, 2H, m	28.9	1.86, 2H, m
<b>8</b>	25.5	1.49, 2H, m	25.4	1.50, 2H, m
<b>9</b>	29.1	1.38, 2H, m	29.0	1.37, 2H, m
<b>10</b>	31.7	1.30, 2H, m	31.6	1.30, 2H, m
<b>11</b>	22.6	1.30, 2H, m	22.5	1.31, 2H, m
<b>12</b>	14.2	0.89, 3H, t (7.0)	14.1	0.88, 3H, t (7.0)
<b>13</b>	127.5	6.47, 1H, d (2.1) 6.16, 1H, d (2.1)	127.4	6.46, 1H, d (2.1) 6.15, 1H, d (2.1)

**Table 7.5**  $^1H$  NMR (500 MHz) data and  $^{13}C$  NMR (125 MHz) data for **1** in  $CDCl_3$ . Literature<sup>51</sup> data were measured at 700 MHz in  $CDCl_3$ .

Figure 7.1  $^1\text{H}$  NMR of compound 1.Figure 7.2  $^{13}\text{C}$  NMR of compound 1.

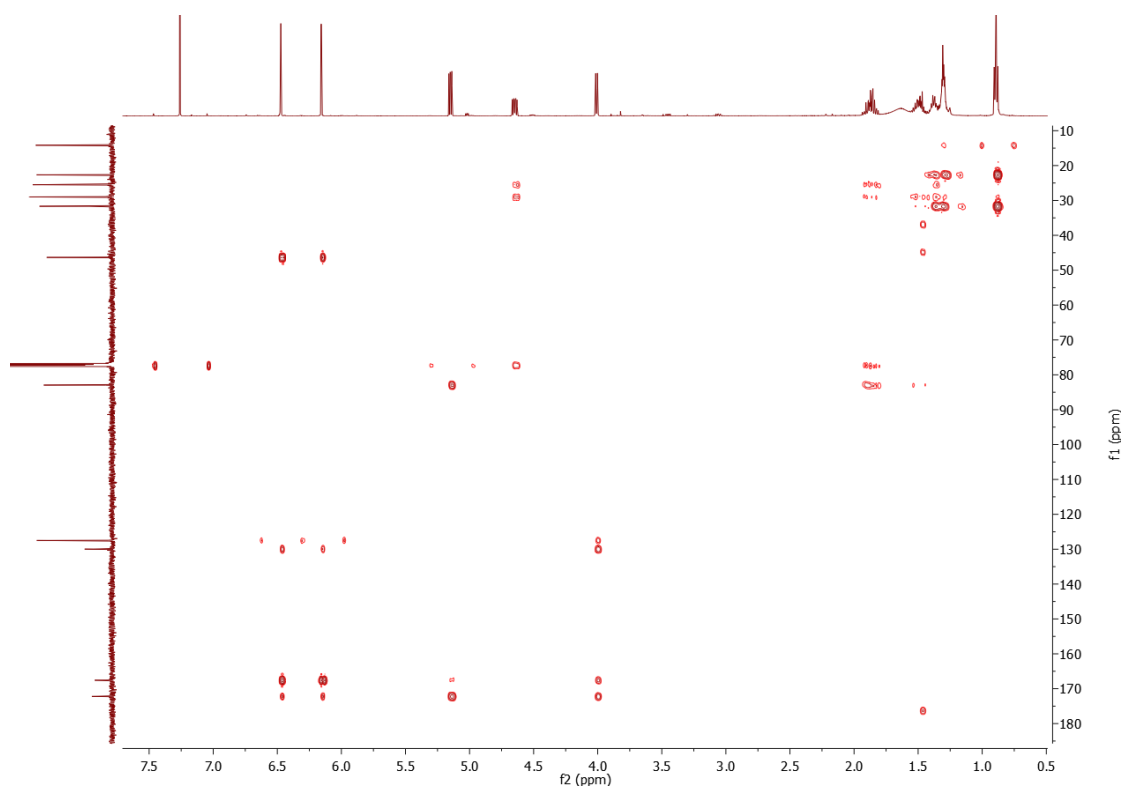


Figure 7.3 HMBC of compound 1.

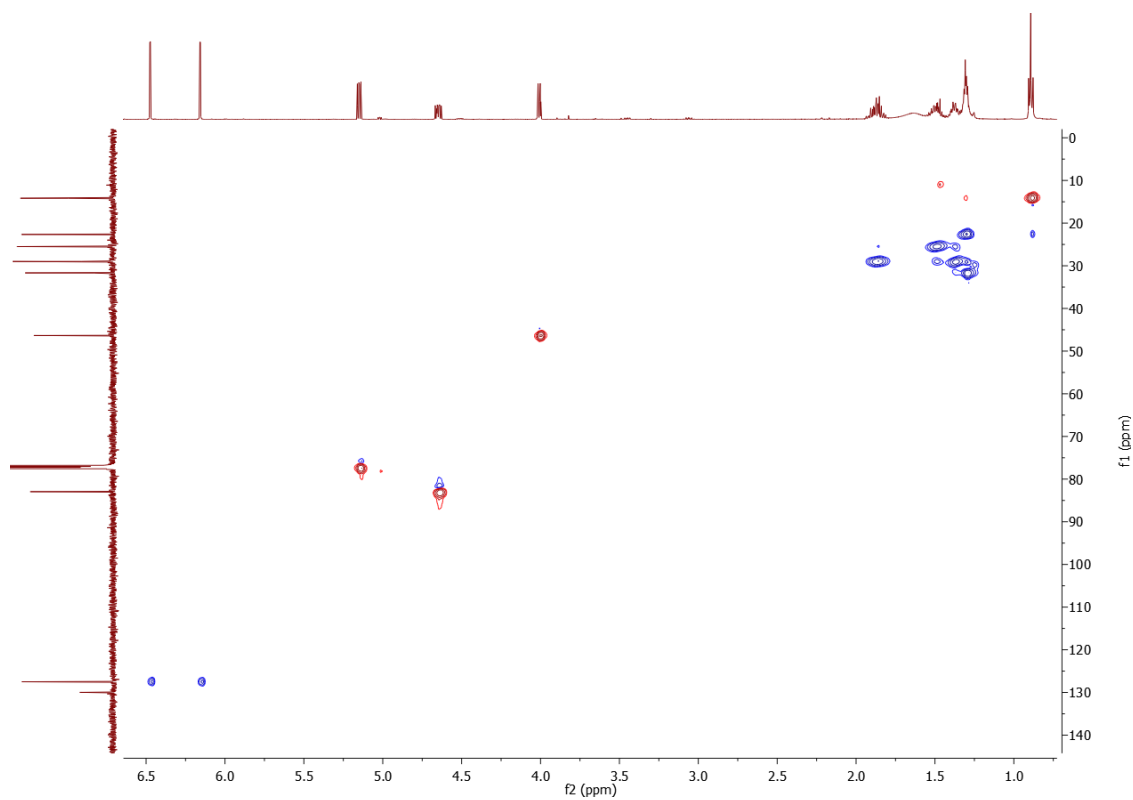


Figure 7.4 HSQC of compound 1.

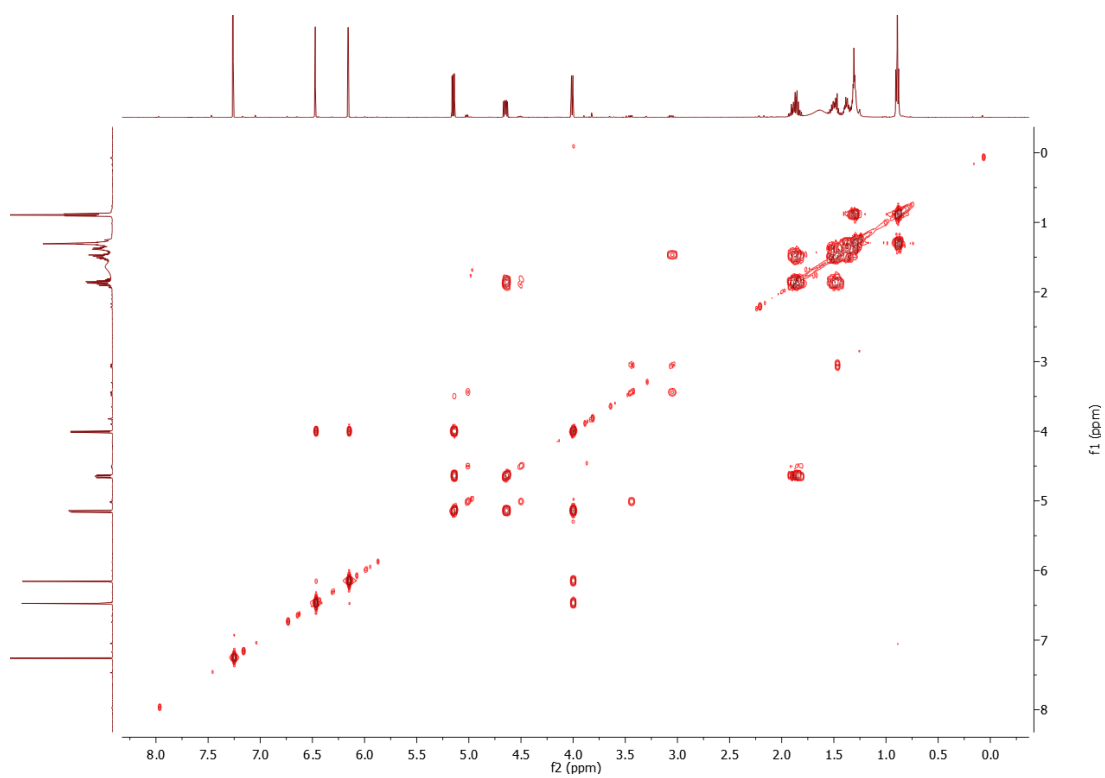
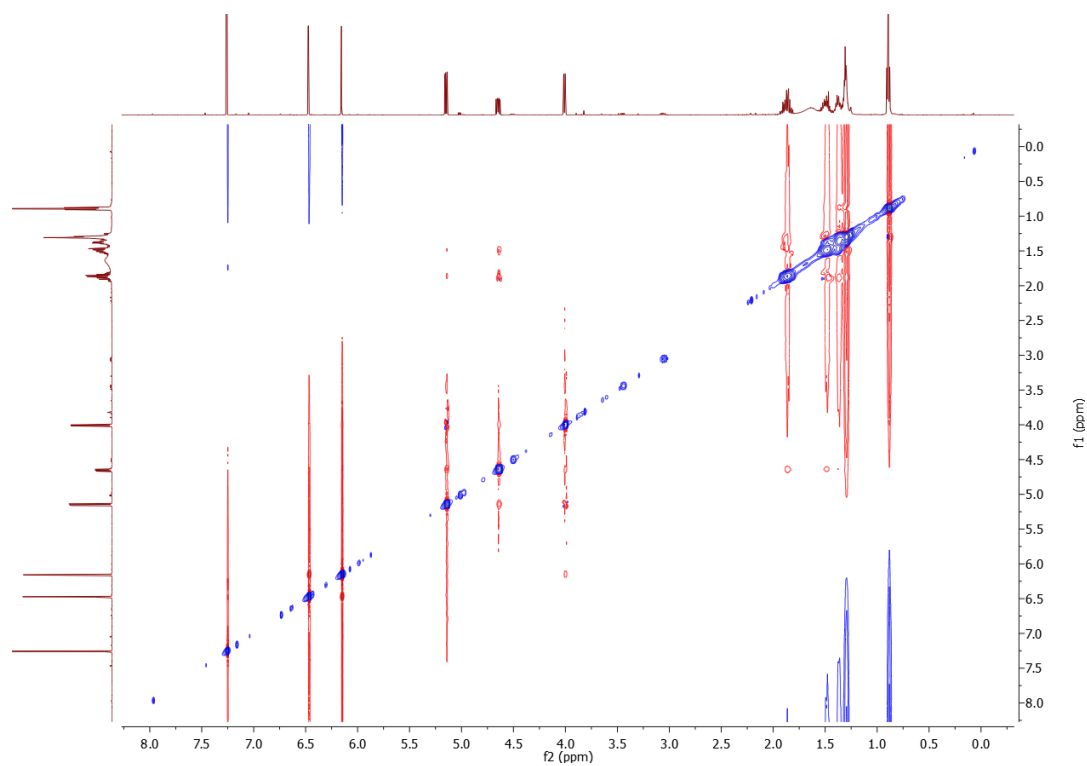
Figure 7.5  $^1\text{H}$ - $^1\text{H}$  COSY of compound 1.

Figure 7.6 NOESY of compound 1.



## Elemental Composition Report

Page 1

## Single Mass Analysis

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 50.0

Selected filters: None

Monoisotopic Mass, Even Electron Ions

31 formula(e) evaluated with 1 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-55 H: 0-100 O: 0-11

Tian

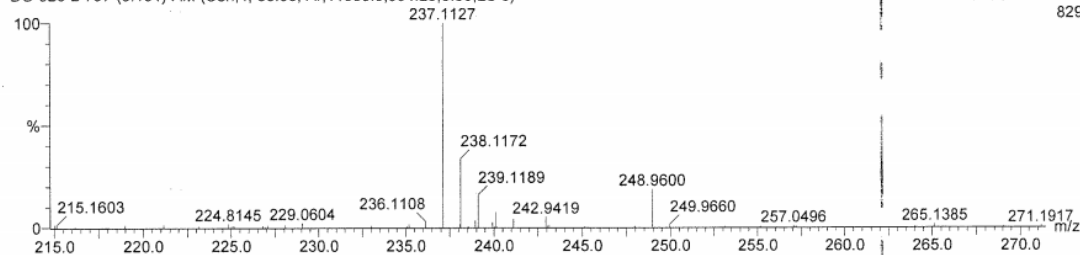
DO 026-2 797 (8.151) AM (Cen,4, 33.00, Ar,11000.0,554.26,0.55,LS 5)

Q-ToF Premier UPLC-MS

10-Dec-2018 10:42:16

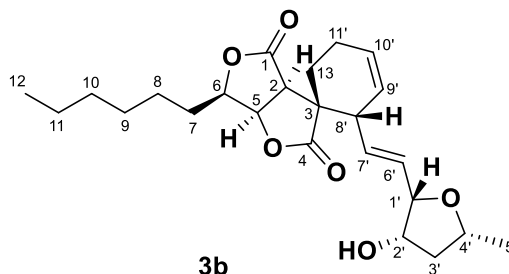
1: TOF MS ES-

829



Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
237.1127	237.1127	0.0	0.0	5.5	90.3	C13 H17 O4

Figure 7.7 HR-ESI-MS of compound 1.

Compound 3b<sup>53,75</sup>

## Sporochartine B

Chemical formula: C<sub>24</sub>H<sub>34</sub>O<sub>6</sub>

White powder;  $[\alpha]_D^{25} + 57$  ( $c = 0.8$ , CHCl<sub>3</sub>); Literature,<sup>53,75</sup>  $[\alpha]_D^{25} + 72$  ( $c = 1.0$ , CHCl<sub>3</sub>); UV ( $\lambda_{\max}$ ): 211 nm. <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 125 MHz):  $\delta_c$  178.7 (C-4), 173.0 (C-1), 130.8 (C-7'), 130.1 (C-6'), 129.9 (C-10'), 124.7 (C-9'), 82.8 (C-1'), 81.1 (C-6), 78.7 (C-5), 74.1 (C-4'), 73.8 (C-2'), 50.8 (C-3), 47.2 (C-2), 46.9 (C-8'), 42.4 (C-3'), 31.7 (C-10), 29.1 (C-9), 28.9 (C-7), 26.9 (C-13), 25.3 (C-8), 22.6 (C-11'), 22.8 (C-11), 22.4 (C-5'), 14.2 (C-12); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 500 MHz):  $\delta_H$  5.95 (1H, m, H-10'), 5.84 (1H, ddd,  $J = 1.6, 9.0, 15.5$  Hz, H-7'), 5.75 (1H, dd,  $J = 4.0, 15.5$  Hz, H-6'), 5.54 (1H, m, H-9'), 5.14 (1H, dd,  $J = 6.0, 4.1$  Hz, H-5), 4.40 (1H, m, H-6), 4.28 (1H, m, H-2'), 4.19 (1H, m, H-1'), 4.06 (1H, m, H-4'), 3.31 (1H, d,  $J = 5.9$  Hz, H-2), 3.23 (1H, m, H-8'),

2.25 (1H, m, H-11'), 2.03 (1H, m, H-13), 1.76 (1H, m, H-7), 1.59 (1H, m, H-3'), 1.45 (2H, m, H-8), 1.34 (2H, m, H-9), 1.33 (3H, d,  $J = 6.2$  Hz, H-5'), 1.29 (2H, m, H-11), 1.28 (2H, m, H-10), 0.88 (3H, m, H-12); NMR data (Table 3.1) are consistent with those previously reported;<sup>53,75</sup> ESI-MS  $m/z$  463 [M + HCOOH - H]<sup>-</sup>, 419 [M + H]<sup>+</sup>, 436 [M + H<sub>2</sub>O]<sup>+</sup>; HR-ESI-MS  $m/z$  417.2278 [M - H]<sup>-</sup> (calcd. for C<sub>24</sub>H<sub>33</sub>O<sub>6</sub>, 417.2277).

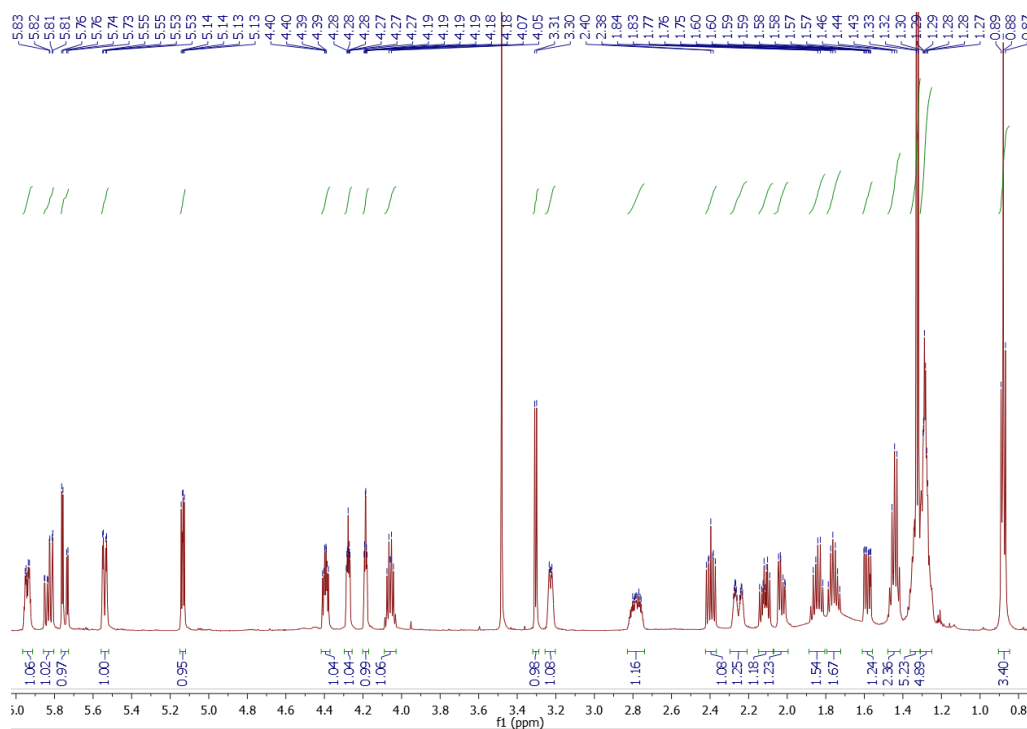


Figure 7.8 <sup>1</sup>H NMR of compound 3b.

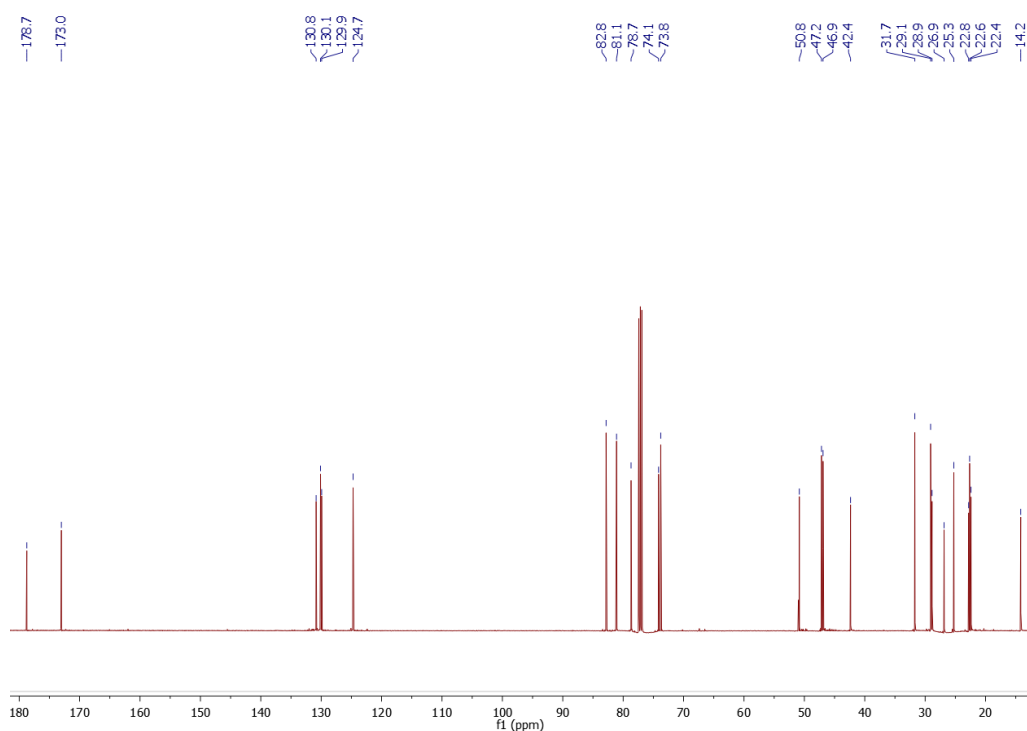


Figure 7.9 <sup>13</sup>C NMR of compound 3b.

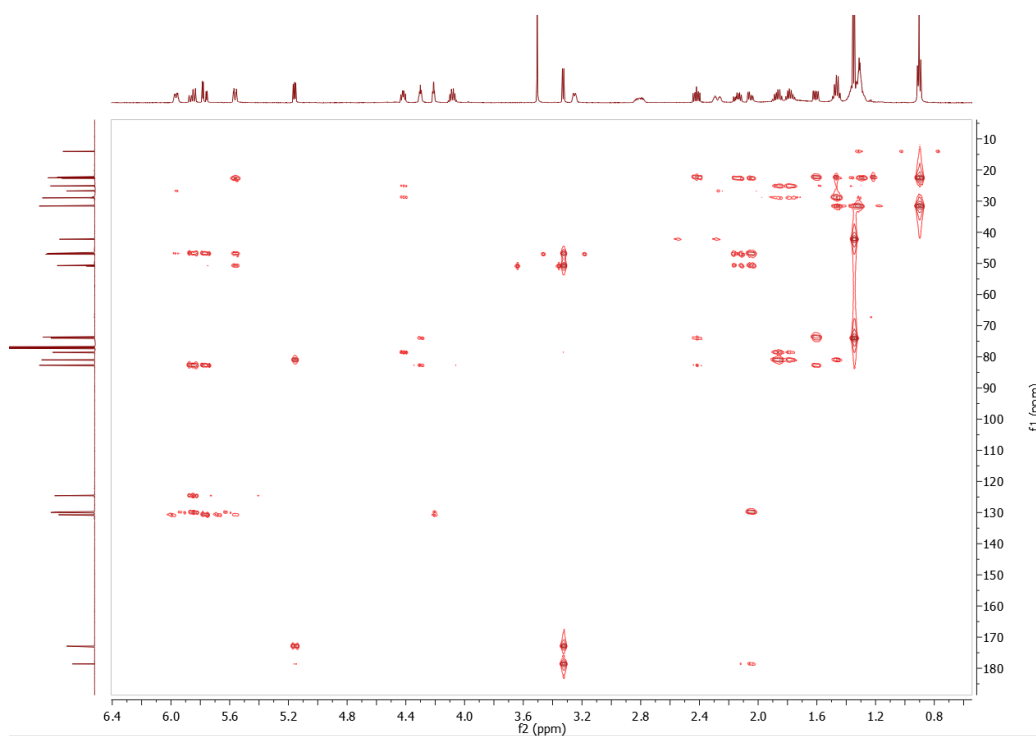


Figure 7.10 HMBC of compound 3b.

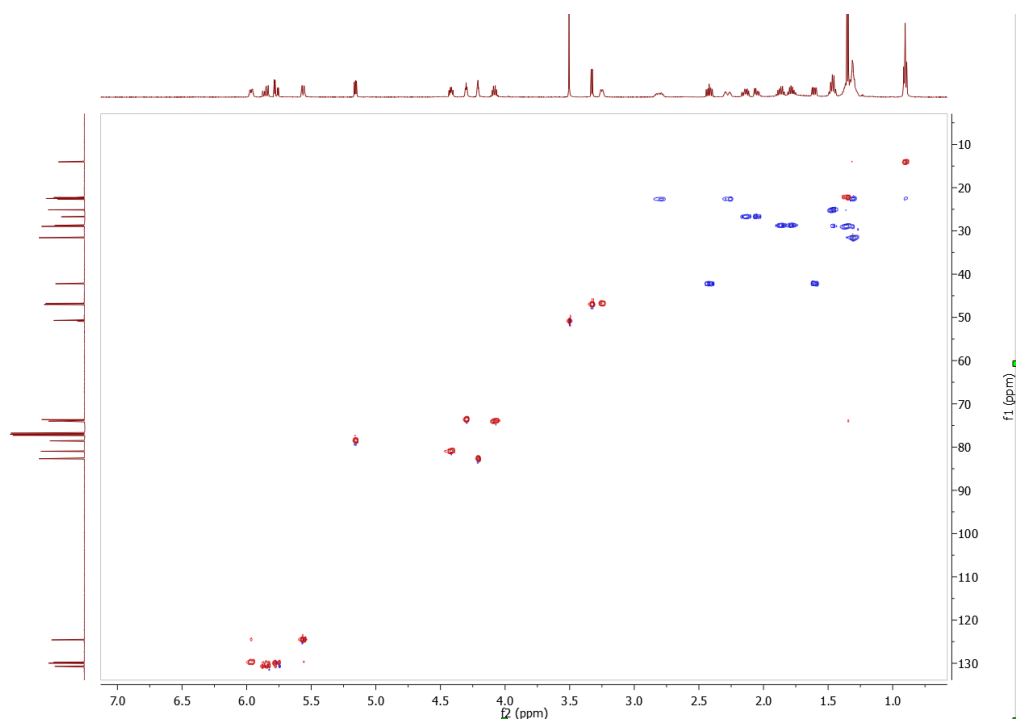
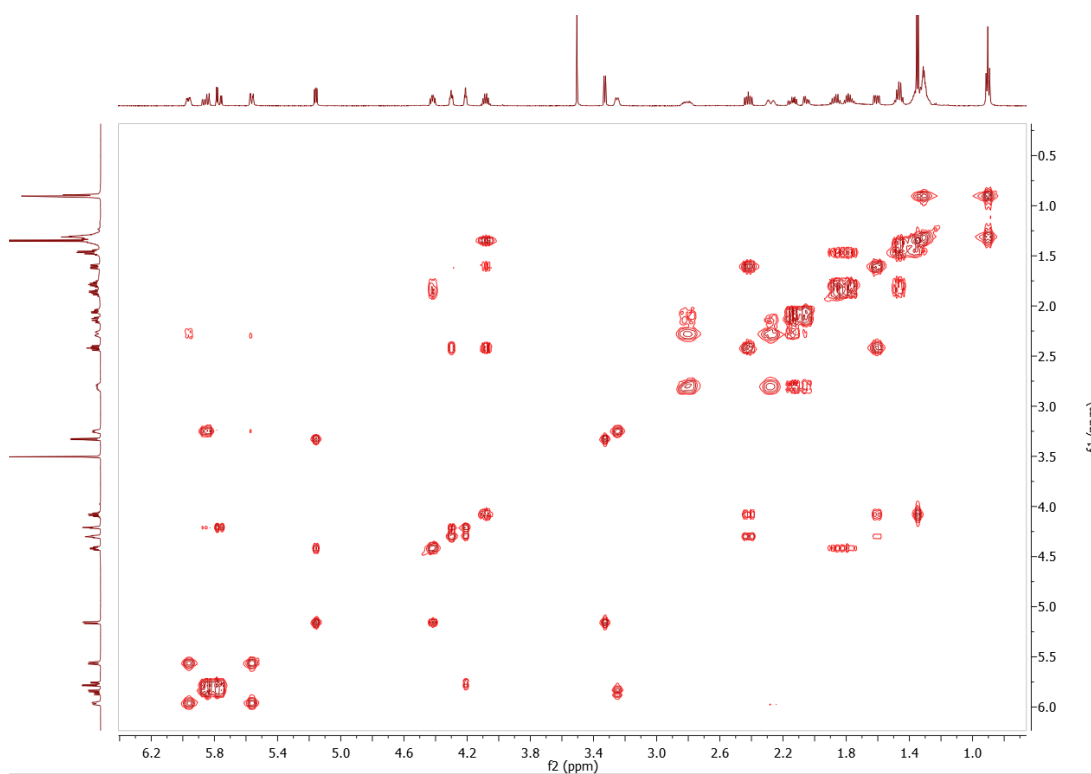
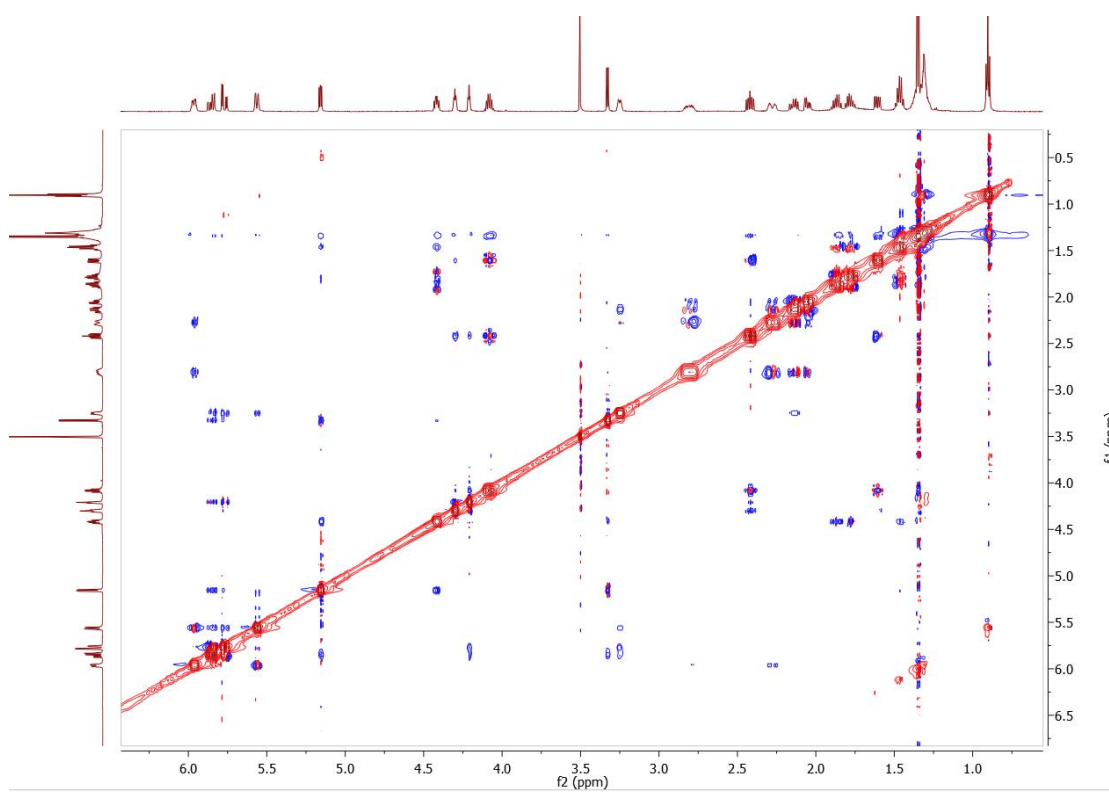


Figure 7.11 HSQC of compound 3b.

Figure 7.12  $^1\text{H}$ - $^1\text{H}$  COSY of compound **3b**.Figure 7.13 NOESY of compound **3b**.

## Elemental Composition Report

Page 1

## Single Mass Analysis

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 50.0

Selected filters: None

Monoisotopic Mass, Even Electron Ions

57 formula(e) evaluated with 3 results within limits (all results (up to 1000) for each mass)

Elements Used:

C: 0-55 H: 0-100 O: 0-11

Tian

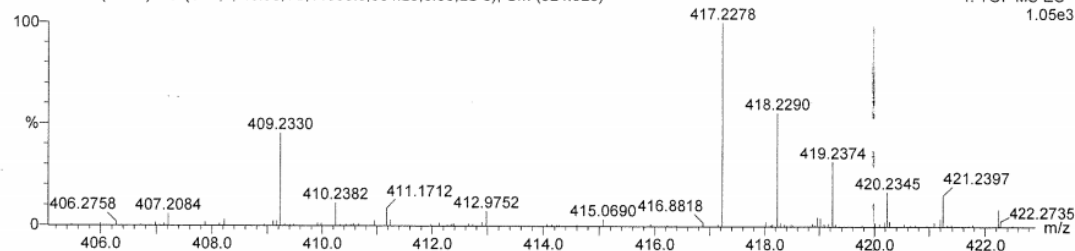
Q-ToF Premier UPLC-MS

06-Dec-201808:48:16

DO 025 924 (9.454) AM (Cen,4, 40.00, Ar,11000.0,554.26,0.55,LS 5); Cm (924:929)

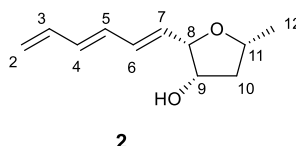
1: TOF MS ES-

1.05e3



Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
417.2278	417.2277	0.1	0.2	8.5	179.7	C24 H33 O6
	417.2336	-5.8	-13.9	-0.5	234.8	C17 H37 O11
	417.2218	6.0	14.4	17.5	135.7	C31 H29 O

Figure 7.14 HR-ESI-MS of compound 3b.

Compound 2<sup>56</sup>

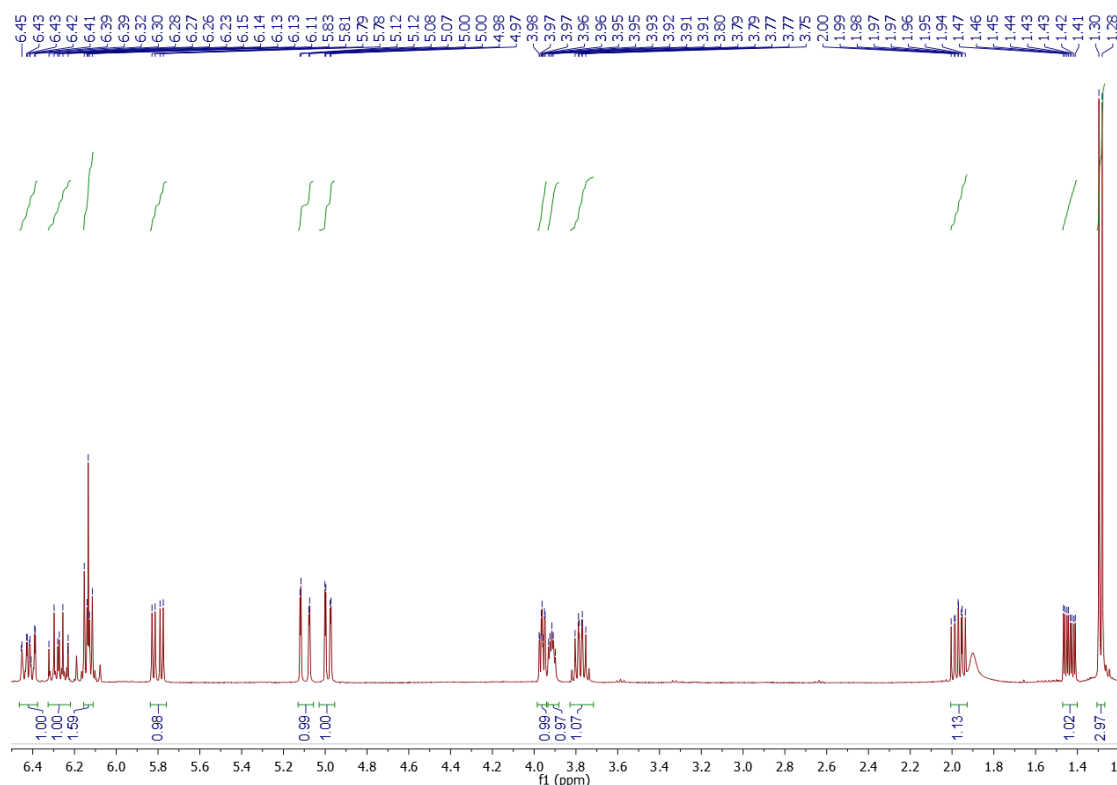
## Trienylfuranol A

Chemical formula: C<sub>11</sub>H<sub>16</sub>O<sub>2</sub>

White powder;  $[\alpha]_D^{25} + 3$  ( $c = 0.1$ , MeCN); Literature,<sup>56</sup>  $[\alpha]_D^{20} + 1.9$  ( $c = 1.10$ , MeCN); UV ( $\lambda_{\max}$ ): 252 nm, 264 nm, 275 nm. <sup>13</sup>C NMR data (C<sub>6</sub>D<sub>6</sub>, 100 MHz):  $\delta_c$  137.4 (C-3), 133.9 (C-4), 133.1 (C-6), 133.1 (C-5), 130.5 (C-7), 117.6 (C-2), 83.9 (C-8), 74.4 (C-9), 74.0 (C-11), 43.0 (C-10), 22.2 (C-12); <sup>1</sup>H NMR data (C<sub>6</sub>D<sub>6</sub>, 400 MHz):  $\delta_H$  6.42 (1H, m, H-6), 6.27 (1H, m, H-3), 6.16 (1H, m, H-5), 6.13 (1H, m, H-4), 5.80 (1H, dd,  $J = 15.4, 6.1$  Hz, H-7), 5.10 (1H, dd,  $J = 16.8, 1.7$  Hz, H-2b), 4.99 (1H, dd,  $J = 10.0, 1.7$  Hz, H-2a), 3.96 (1H, ddd,  $J = 5.9, 4.1, 1.3$  Hz, H-8), 3.91 (1H, m, H-9), 3.78 (1H, m, H-11), 1.97 (1H, ddd,  $J = 13.3, 7.7, 6.3$  Hz, H-10a), 1.44 (1H, ddd,  $J = 13.3, 6.8, 2.8$  Hz, H-10b), 1.29 (3H, d,  $J = 6.2$  Hz, H-12); NMR data (Table 7.6) are consistent with those previously reported;<sup>56</sup> ESI-MS  $m/z$  225 [M + HCOOH - H]<sup>-</sup>, 181 [M + H]<sup>+</sup>, 361 [2M + H]<sup>+</sup>; HR-ESI-MS  $m/z$  203.1049 [M + Na]<sup>+</sup> (calcd. for C<sub>11</sub>H<sub>16</sub>O<sub>2</sub>Na, 203.1048).

pos.	$\delta_C$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_C$ / ppm literature <sup>56</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>56</sup>
2	117.6	4.99, 1H, dd (10.0, 1.7) 5.10, 1H, dd (16.8, 1.7)	117.6	4.98, 1H, dd (10.1, 1.7) 5.09, 1H, dd (16.9, 1.7)
3	137.4	6.27, 1H, m	137.4	6.27, 1H, ddd (16.9, 10.1, 10.1)
4	133.9	6.13, 1H, m	133.9	6.12, 1H, dd (15.1, 10.3)
5	133.1	6.16, 1H, m	133.1	6.18, 1H, dd (15.4, 10.4)
6	133.1	6.42, 1H, m	133.1	6.43, 1H, ddd (15.4, 9.9, 1.5)
7	130.5	5.80, 1H, dd (15.4, 6.1)	130.3	5.75, 1H, dd (15.4, 6.0)
8	83.9	3.96, 1H, ddd (5.9, 4.1, 1.3)	83.8	3.95, 1H, ddd (5.7, 4.0, 1.4)
9	74.4	3.91, 1H, m	74.4	3.86, 1H, m
10	43.0	1.44, 1H, ddd (13.3, 6.8, 2.8) 1.97, 1H, ddd (13.3, 7.7, 6.3)	43.0	1.42, 1H, ddd (13.3, 6.6, 2.7) 1.93, 1H, ddd (13.3, 7.8, 6.3)
11	74.0	3.78, 1H, m	74.0	3.77, 1H, ddq (7.7, 6.3, 6.3)
12	22.2	1.29, 3H, d (6.2)	22.3	1.28, 3H, d (6.2)

**Table 7.6** <sup>1</sup>H NMR (400 MHz) data and <sup>13</sup>C NMR (100 MHz) data for **2** in C<sub>6</sub>D<sub>6</sub>. Literature<sup>56</sup> data was measured at 600 MHz in C<sub>6</sub>D<sub>6</sub>.



**Figure 7.15** <sup>1</sup>H NMR of compound **2**.

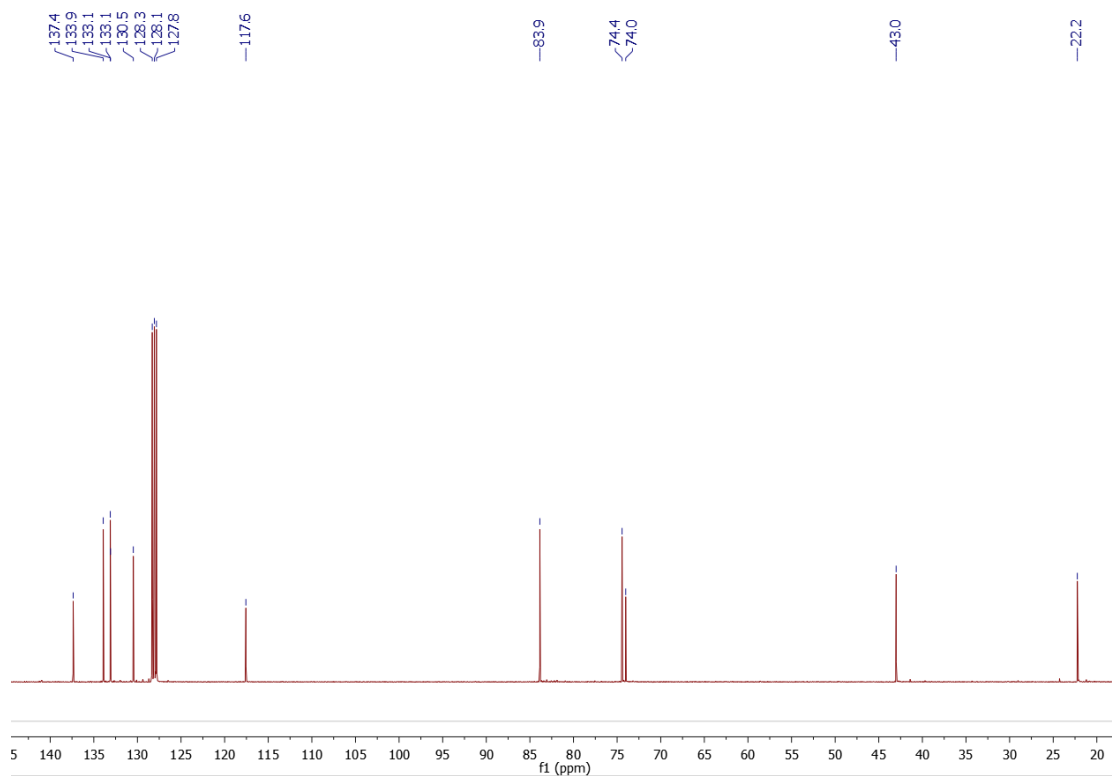
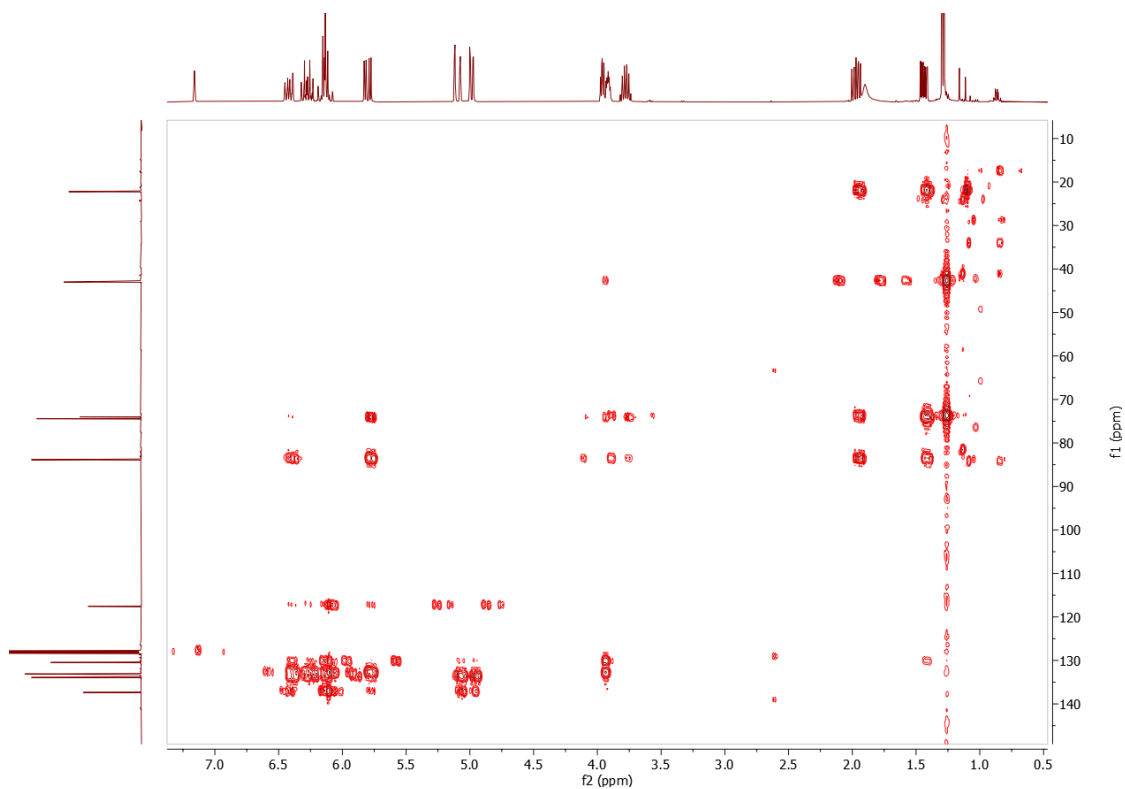
Figure 7.16 <sup>13</sup>C NMR of compound 2.

Figure 7.17 HMBC of compound 2.

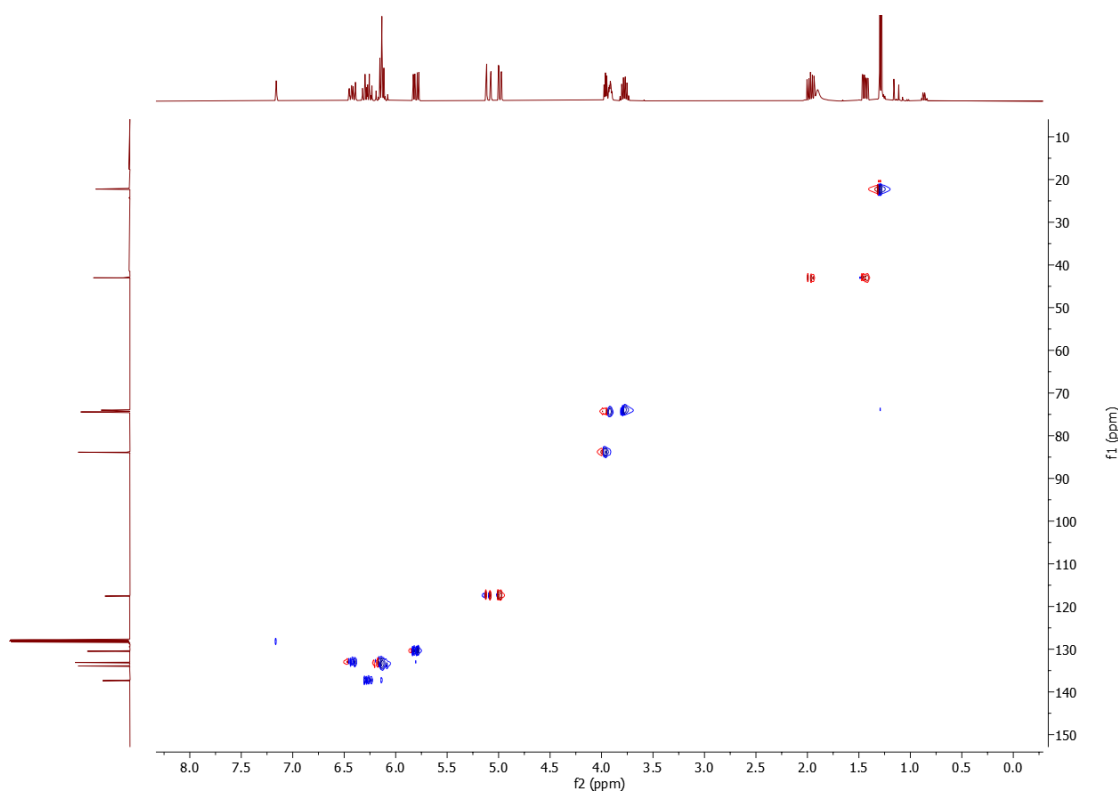
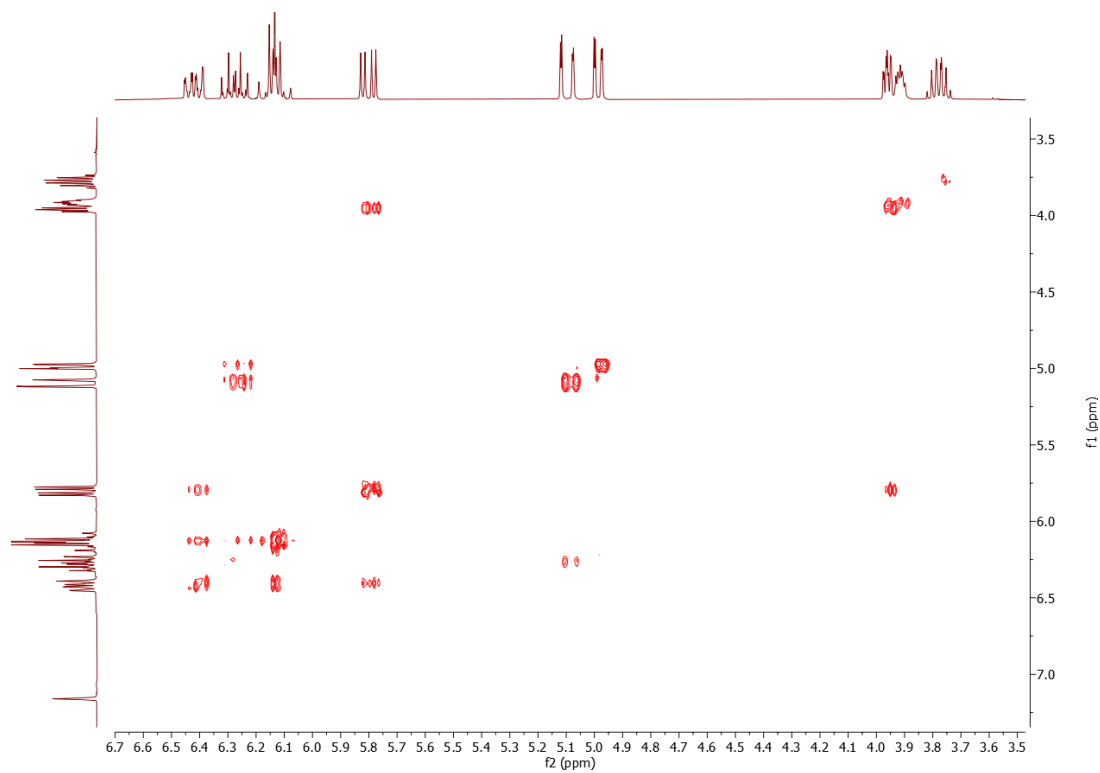


Figure 7.18 HSQC of compound 2.

Figure 7.19  $^1\text{H}$ - $^1\text{H}$  COSY of compound 2.



## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 10.0 PPM / DBE: min = -1.5, max = 50.0

Selected filters: None

Monoisotopic Mass, Odd and Even Electron Ions

298 formula(e) evaluated with 5 results within limits (up to 80 closest results for each mass)

Elements Used:

C: 0-30 H: 0-60 N: 0-9 O: 0-6 Na: 0-1

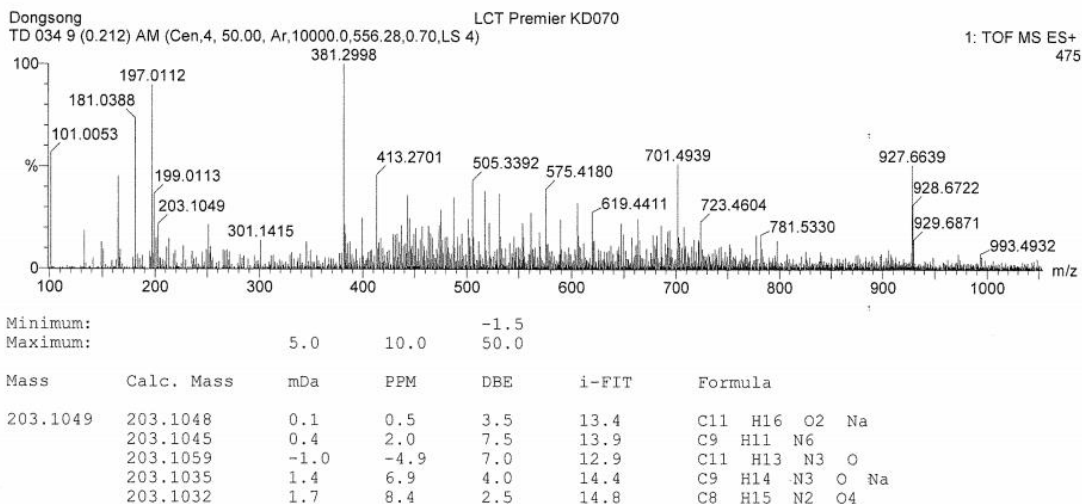
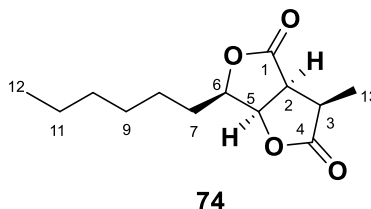


Figure 7.20 HR-ESI-MS of compound 2.

Compound 74<sup>51</sup>

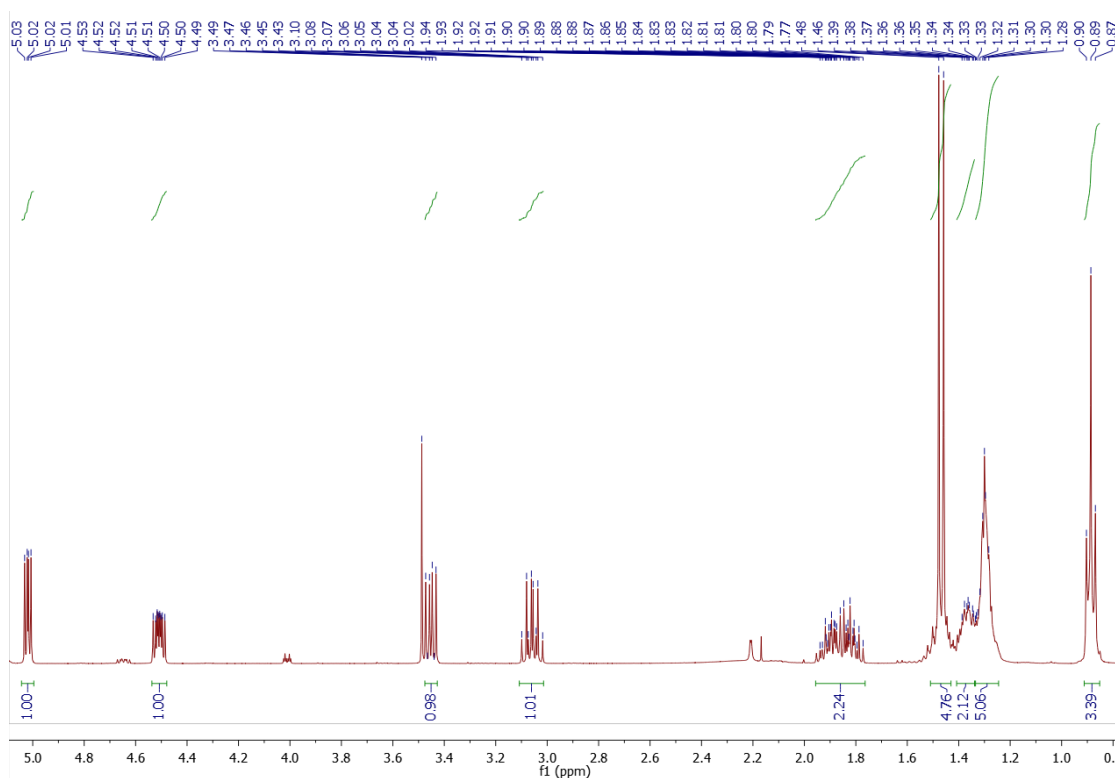
## Dihydrosporothriolide

Chemical formula: C<sub>13</sub>H<sub>20</sub>O<sub>4</sub>

Colorless oil;  $[\alpha]_D^{25} + 94$  ( $c = 0.1$ , CHCl<sub>3</sub>); Literature,<sup>51</sup>  $[\alpha]_D^{25} + 116$  ( $c = 0.1$ , CHCl<sub>3</sub>); UV ( $\lambda_{\max}$ ): 227 nm. <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 100 MHz):  $\delta_c$  176.3 (C-1), 172.2 (C-4), 81.8 (C-6), 78.1 (C-5), 44.8 (C-2), 36.8 (C-3), 31.7 (C-10), 29.1 (C-9), 28.9 (C-7), 25.5 (C-8), 22.6 (C-11), 14.2 (C-12), 11.1 (C-13); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 400 MHz):  $\delta_H$  5.02 (1H, dd,  $J = 6.0, 4.0$  Hz, H-5), 4.51 (1H, ddd,  $J = 8.3, 6.1, 4.0$  Hz, H-6), 3.45 (1H, dd,  $J = 10.1, 6.1$  Hz, H-2), 3.06 (1H, dq,  $J = 10.2, 7.6$  Hz, H-3), 1.91 (1H, m, H-7a), 1.81 (1H, m, H-7b), 1.50 (2H, m, H-8), 1.47 (3H, d,  $J = 7.5$  Hz, H-13), 1.37 (2H, m, H-9), 1.31 (2H, m, H-11), 1.30 (2H, m, H-10), 0.89 (3H, t,  $J = 7.0$  Hz, H-12); NMR data (Table 7.7) are consistent with those previously reported;<sup>51</sup> ESI-MS  $m/z$  239 [M - H]<sup>-</sup>, 479 [2M - H]<sup>-</sup>, 241 [M + H]<sup>+</sup>.

pos.	$\delta_c$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_c$ / ppm literature <sup>51</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>51</sup>
1	176.3	-	176.2	-
2	44.8	3.45, 1H, dd (10.1, 6.1)	44.7	3.44, 1H, dd (10.1, 6.0)
3	36.8	3.06, 1H, dq (10.2, 7.6)	36.8	3.05, 1H, dq (10.1, 7.5)
4	172.2	-	172.1	-
5	78.1	5.02, 1H, dd (6.0, 4.0)	78.1	5.01, 1H, dd (6.0, 3.9)
6	81.8	4.51, 1H, ddd (8.3, 6.1, 4.0)	81.7	4.50, 1H, ddd (8.0, 6.2, 3.9)
7	28.9	1.91, 1H, m 1.81, 1H, m	28.9	1.92, 1H, m 1.81, 1H, m
8	25.5	1.50, 2H, m	25.3	1.50, 2H, m
9	29.1	1.37, 2H, m	29.0	1.37, 2H, m
10	31.7	1.30, 2H, m	31.6	1.30, 2H, m
11	22.6	1.31, 2H, m	22.6	1.31, 2H, m
12	14.2	0.89, 3H, t (7.0)	14.2	0.88, 3H, t (7.0)
13	11.1	1.47, 3H, d (7.5)	11.1	1.47, 3H, d (7.5)

**Table 7.7**  $^1\text{H}$  NMR (400 MHz) data and  $^{13}\text{C}$  NMR (100 MHz) data for **74** in  $\text{CDCl}_3$ . Literature<sup>51</sup> data was measured at 700 MHz in  $\text{CDCl}_3$ .



**Figure 7.21**  $^1\text{H}$  NMR of compound **74**.

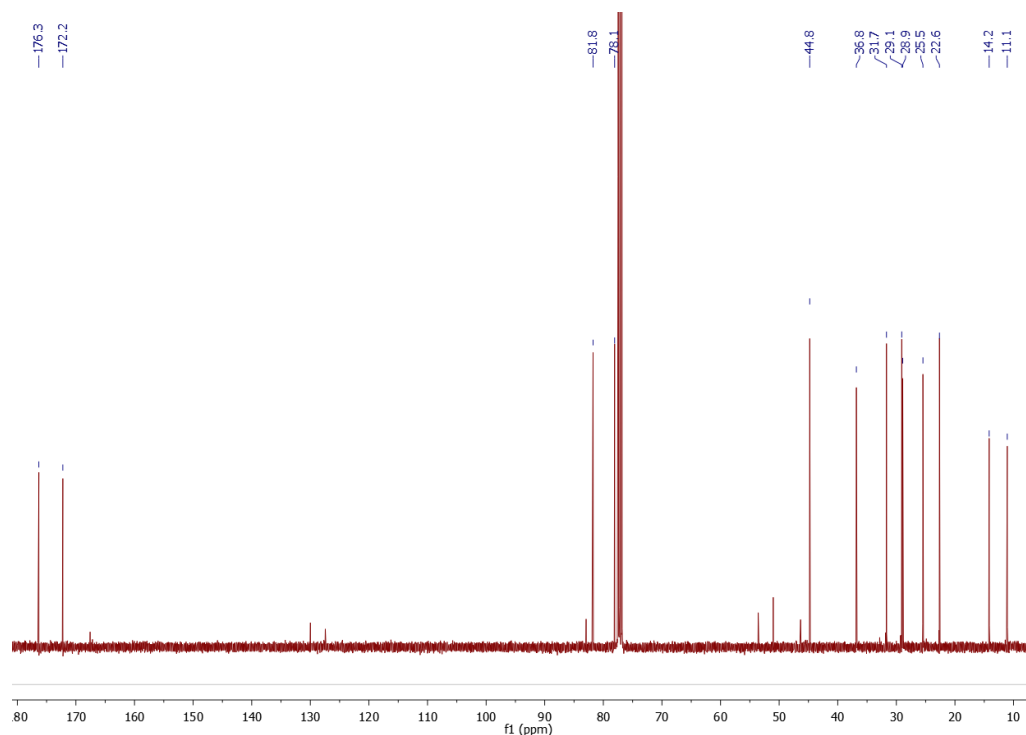
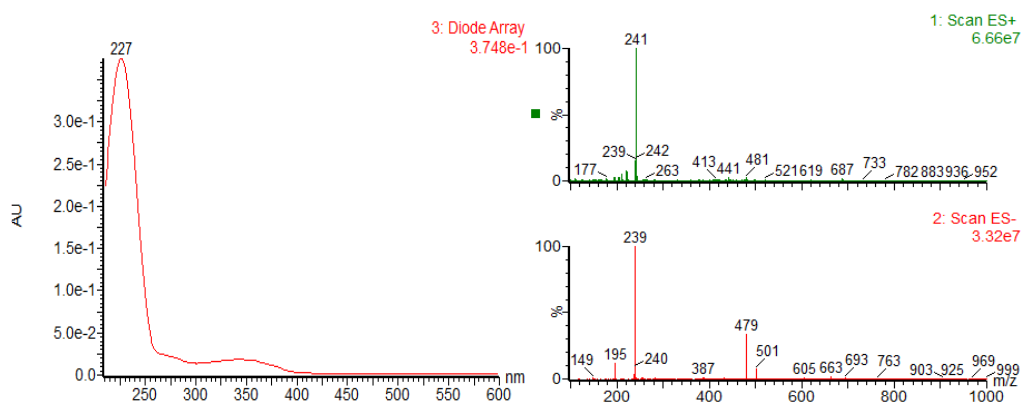
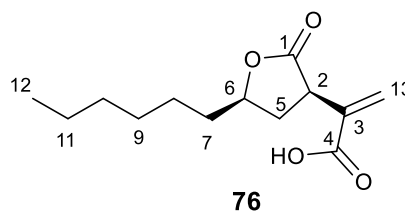
Figure 7.22  $^{13}\text{C}$  NMR of compound 74.

Figure 7.23 UV and mass spectra of compound 74.

Compound 76<sup>53</sup>

## Deoxysporothric acid

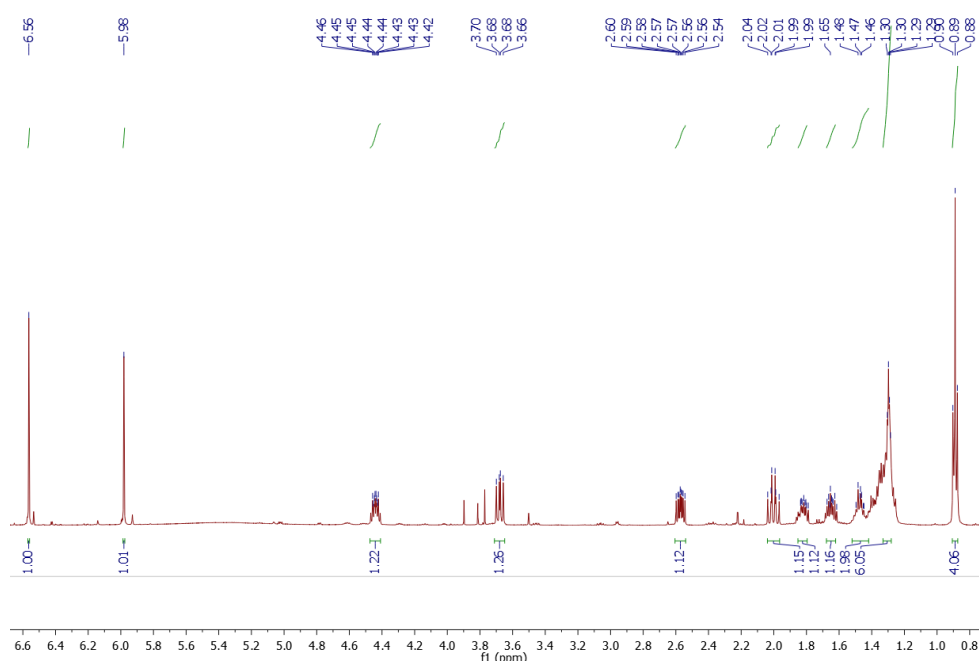
Chemical formula:  $\text{C}_{13}\text{H}_{20}\text{O}_4$ 

White powder;  $[\alpha]_{\text{D}}^{25} + 9$  ( $c = 0.1$ ,  $\text{CHCl}_3$ ); Literature,<sup>53</sup>  $[\alpha]_{\text{D}}^{25} + 18$  ( $c = 0.37$ ,  $\text{CHCl}_3$ ); UV ( $\lambda_{\text{max}}$ ): 211 nm.  $^{13}\text{C}$  NMR data ( $\text{CDCl}_3$ , 100 MHz):  $\delta_{\text{c}}$  176.0 (C-1), 169.8 (C-4), 135.5 (C-3), 131.8 (C-

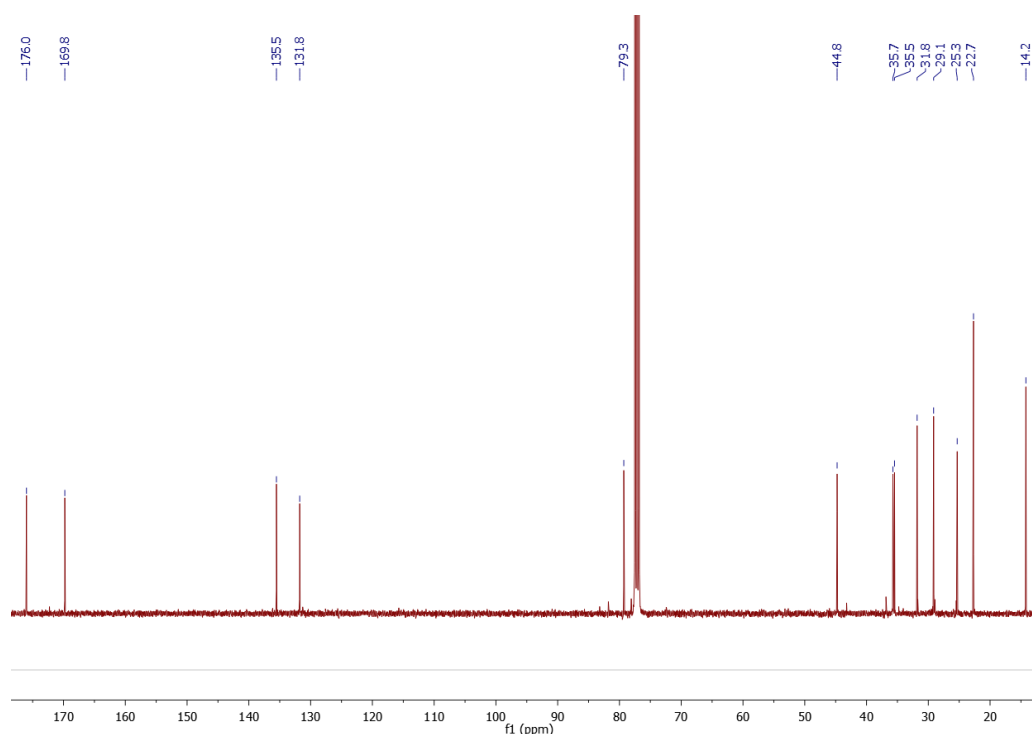
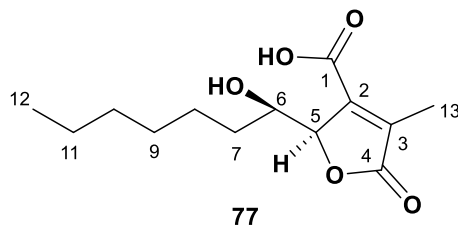
13), 79.3 (C-6), 44.8 (C-2), 35.7 (C-5), 35.5 (C-7), 31.8 (C-10), 29.1 (C-9), 25.3 (C-8), 22.7 (C-11), 14.2 (C-12);  $^1\text{H}$  NMR data ( $\text{CDCl}_3$ , 400 MHz):  $\delta_{\text{H}}$  6.56 (1H, s, H-13a), 5.98 (1H, s, H-13b), 4.44 (1H, m, H-6), 3.68 (1H, dd,  $J = 12.2, 8.9$  Hz, H-2), 2.57 (1H, ddd,  $J = 11.5, 9.3, 5.7$  Hz, H-5a), 2.00 (1H, q,  $J = 10.3$  Hz, H-5b), 1.82 (1H, m, H-7a), 1.65 (1H, m, H-7b), 1.47 (2H, m, H-8), 1.33 (2H, m, H-9), 1.30 (2H, m, H-11), 1.28 (2H, m, H-10), 0.89 (3H, t,  $J = 6.8$  Hz, H-12); NMR data (Table 7.8) are consistent with those previously reported;<sup>53</sup> ESI-MS  $m/z$  239  $[\text{M} - \text{H}]^-$ , 479  $[2\text{M} - \text{H}]^-$ , 241  $[\text{M} + \text{H}]^+$ , 481  $[2\text{M} + \text{H}]^+$ .

pos.	$\delta_{\text{C}}$ / ppm	$\delta_{\text{H}}$ / ppm ( $J$ / Hz)	$\delta_{\text{C}}$ / ppm literature <sup>53</sup>	$\delta_{\text{H}}$ / ppm ( $J$ / Hz) literature <sup>53</sup>
1	176.0	-	176.1	-
2	44.8	3.68, 1H, dd (12.2, 8.9)	44.8	3.68, 1H, dd (12.0, 8.9)
3	135.5	-	135.8	-
4	169.8	-	169.8	-
5	35.7	2.57, 1H, ddd (11.5, 9.3, 5.7) 2.00, 1H, q (10.3)	35.7	2.56, 1H, m 2.01, 1H, q (10.5)
6	79.3	4.44, 1H, m	79.3	4.43, 1H, m
7	35.5	1.82, 1H, m 1.65, 1H, m	35.5	1.82, 1H, m 1.65, 1H, m
8	25.3	1.47, 2H, m	25.3	1.48, 2H, m
9	29.1	1.33, 2H, m	29.1	1.33, 2H, m
10	31.8	1.28, 2H, m	31.8	1.28, 2H, m
11	22.7	1.30, 2H, m	22.7	1.30, 2H, m
12	14.2	0.89, 3H, t (6.8)	14.2	0.89, 3H, t (6.9)
13	131.8	6.56, 1H, s 5.98, 1H, s	131.4	6.54, 1H, s 5.95, 1H, s

**Table 7.8**  $^1\text{H}$  NMR (400 MHz) data and  $^{13}\text{C}$  NMR (100 MHz) data for **76** in  $\text{CDCl}_3$ . Literature<sup>53</sup> data was measured at 500 MHz in  $\text{CDCl}_3$ .



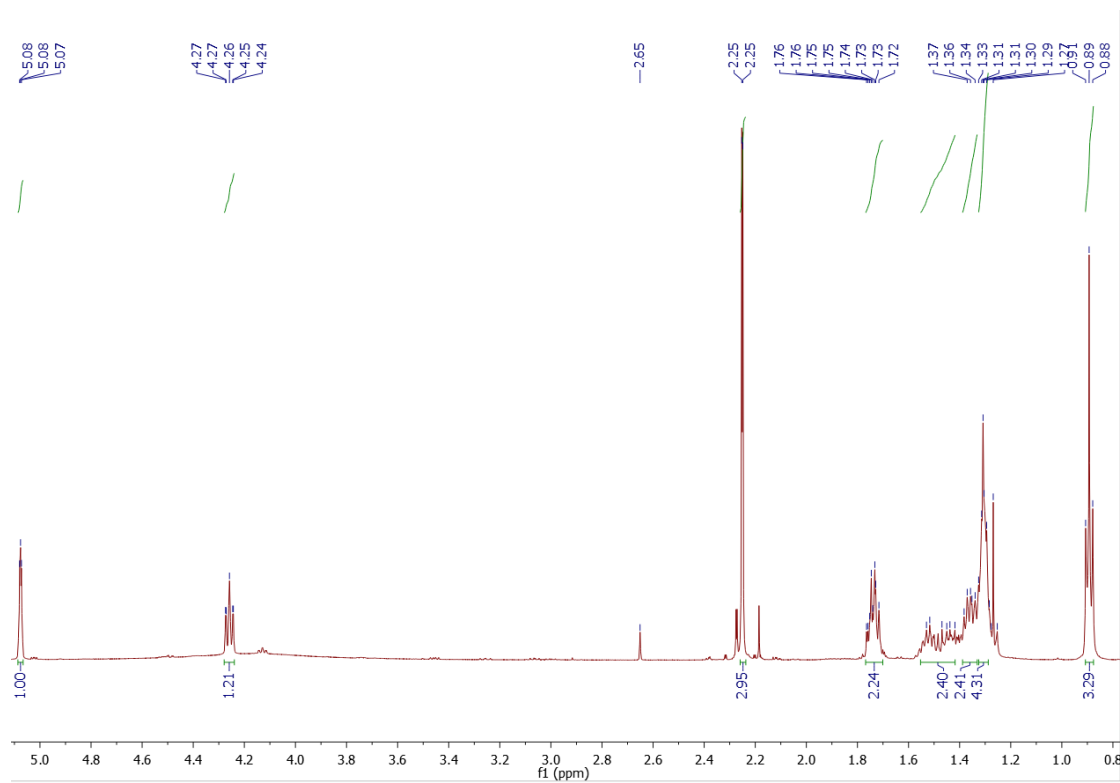
**Figure 7.24**  $^1\text{H}$  NMR of compound **76**.

Figure 7.25  $^{13}\text{C}$  NMR of compound 76.**Compound 77<sup>51</sup>****Isosporothric acid****Chemical formula:  $\text{C}_{13}\text{H}_{20}\text{O}_5$** 

Colorless oil;  $[\alpha]_{\text{D}}^{25} - 7$  ( $c = 0.1$ ,  $\text{CHCl}_3$ ); Literature,<sup>51</sup>  $[\alpha]_{\text{D}}^{25} - 18$  ( $c = 0.5$ ,  $\text{CHCl}_3$ ); UV ( $\lambda_{\text{max}}$ ): 236 nm.  $^{13}\text{C}$  NMR data ( $\text{CDCl}_3$ , 125 MHz):  $\delta_{\text{c}}$  173.2 (C-4), 165.0 (C-1), 144.8 (C-2), 140.4 (C-3), 83.1 (C-5), 70.2 (C-6), 34.5 (C-7), 31.8 (C-10), 29.2 (C-9), 25.9 (C-8), 22.7 (C-11), 14.2 (C-12), 11.3 (C-13);  $^1\text{H}$  NMR data ( $\text{CDCl}_3$ , 500 MHz):  $\delta_{\text{H}}$  5.08 (1H, m, H-5), 4.26 (1H, m, H-6), 2.24 (3H, d,  $J = 2.0$  Hz, H-13), 1.74 (2H, m, H-7), 1.49 (2H, m, H-8), 1.36 (2H, m, H-9), 1.30 (2H, m, H-11), 1.29 (2H, m, H-10), 0.89 (3H, t,  $J = 6.9$  Hz, H-12); NMR data (Table 7.9) are consistent with those previously reported;<sup>51</sup> ESI-MS  $m/z$  255  $[\text{M} - \text{H}]^-$ , 511  $[2\text{M} - \text{H}]^-$ , 257  $[\text{M} + \text{H}]^+$ , 513  $[2\text{M} + \text{H}]^+$ .

pos.	$\delta_C$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_C$ / ppm literature <sup>51</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>51</sup>
1	165.0	-	164.8	-
2	144.8	-	144.7	-
3	140.4	-	140.2	-
4	173.2	-	173.1	-
5	83.1	5.08, 1H, m	83.1	5.06, 1H, qd (2.1, 1.5)
6	70.2	4.26, 1H, m	70.1	4.25, 1H, dt (7.0, 1.5)
7	34.5	1.74, 2H, m	34.4	1.73, 2H, m
8	25.9	1.49, 2H, m	25.8	1.50, 2H, m
9	29.2	1.36, 2H, m	29.1	1.34, 2H, m
10	31.8	1.29, 2H, m	31.7	1.28, 2H, m
11	22.7	1.30, 2H, m	22.6	1.29, 2H, m
12	14.2	0.89, 3H, t (6.9)	14.1	0.88, 3H, t (7.0)
13	11.3	2.24, 3H, d (2.0)	11.2	2.24, 3H, d (2.1)

**Table 7.9** <sup>1</sup>H NMR (500 MHz) data and <sup>13</sup>C NMR (125 MHz) data for **77** in CDCl<sub>3</sub>. Literature<sup>51</sup> data was measured at 700 MHz in CDCl<sub>3</sub>.



**Figure 7.26** <sup>1</sup>H NMR of compound **77**.

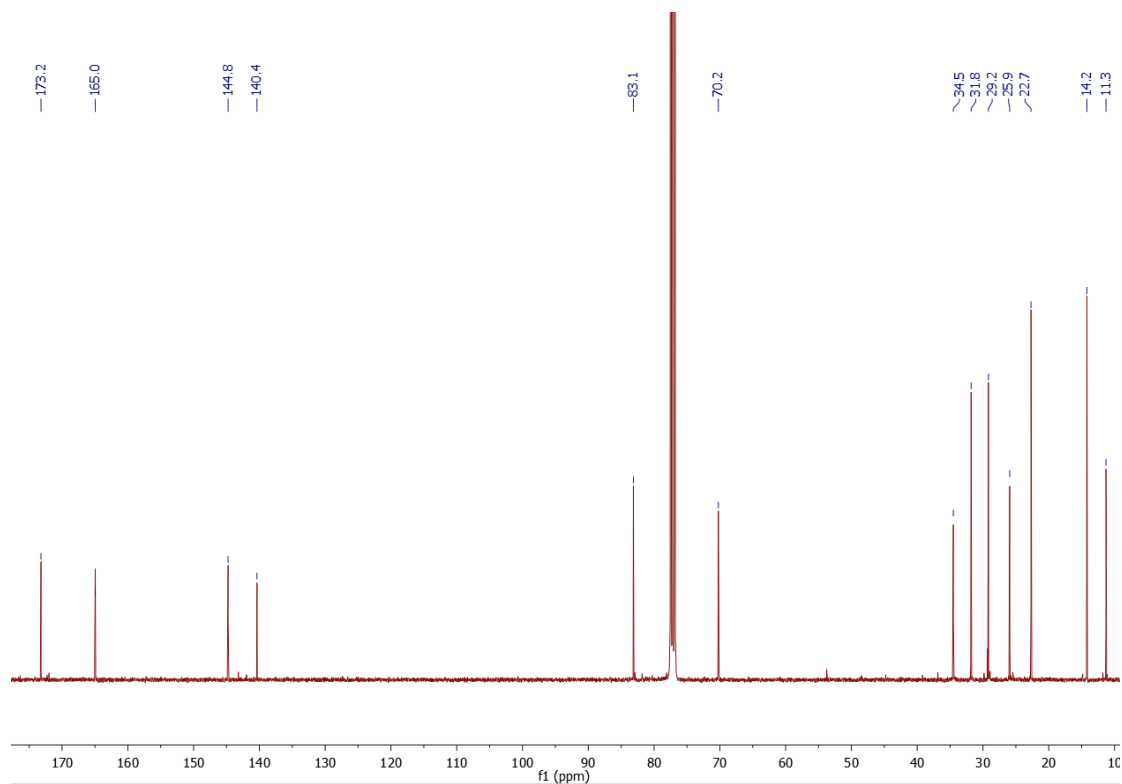
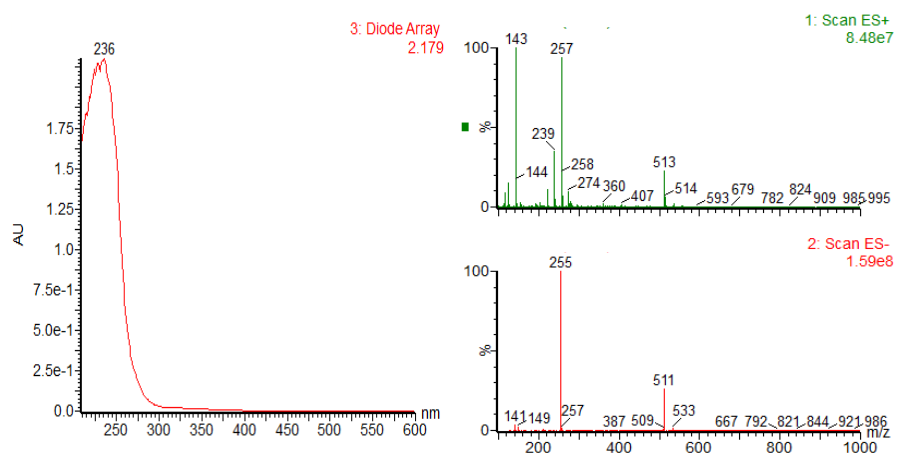
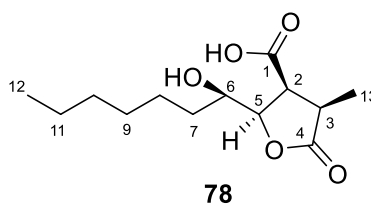
Figure 7.27  $^{13}\text{C}$  NMR of compound 77.

Figure 7.28 UV and mass spectra of compound 77.

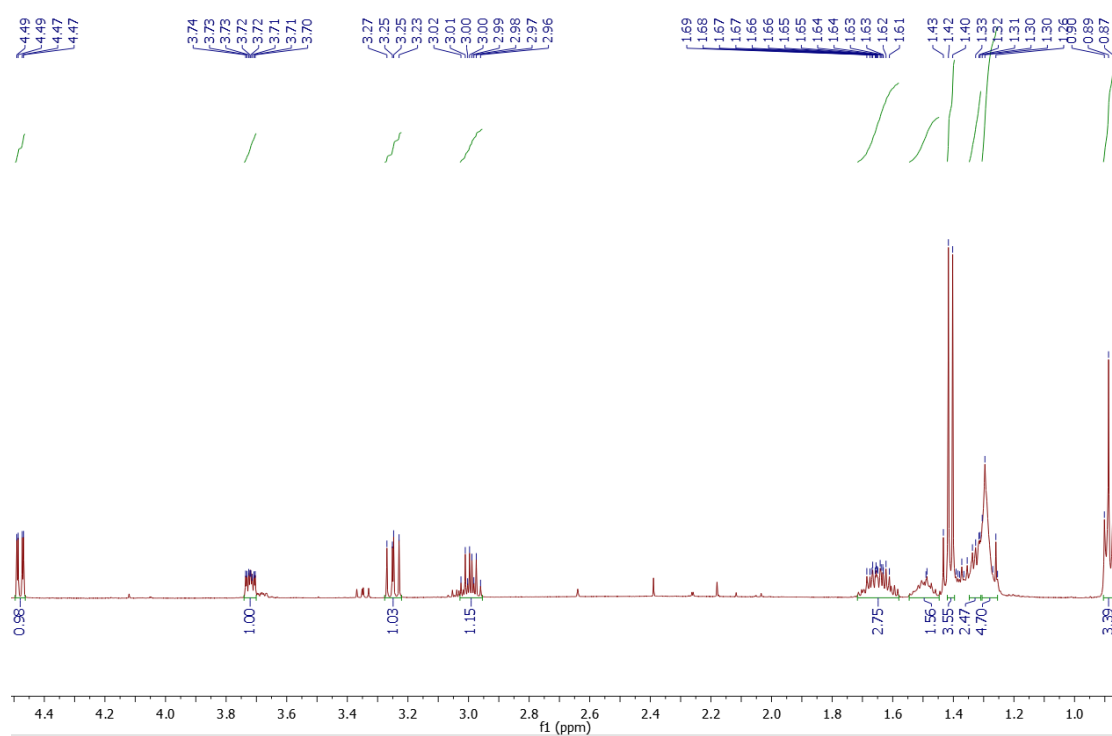
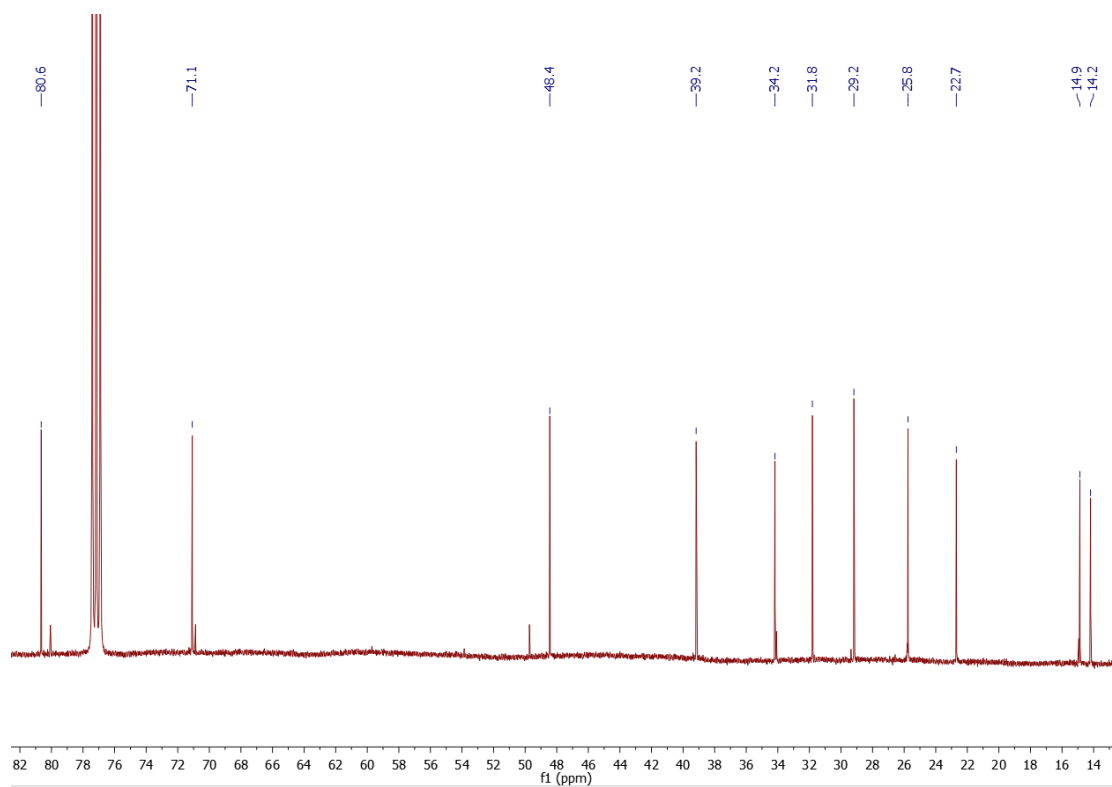
Compound **78**<sup>51</sup>**Dihydroisoporphric acid****Chemical formula: C<sub>13</sub>H<sub>22</sub>O<sub>5</sub>**

Colorless oil;  $[\alpha]_{\text{D}}^{25} + 12$  ( $c = 0.1$ , CHCl<sub>3</sub>); Literature,<sup>51</sup>  $[\alpha]_{\text{D}}^{25} + 16$  ( $c = 0.04$ , CHCl<sub>3</sub>); UV ( $\lambda_{\text{max}}$ ): 329 nm, 213 nm. <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 125 MHz):  $\delta_{\text{C}}$  176.8 (C-4), 174.1 (C-1), 80.6 (C-5), 71.1 (C-6), 48.4 (C-2), 39.2 (C-3), 34.2 (C-7), 31.8 (C-10), 29.2 (C-9), 25.8 (C-8), 22.7 (C-11), 14.9 (C-13), 14.2 (C-12); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 500 MHz):  $\delta_{\text{H}}$  4.48 (1H, dd,  $J = 9.0, 2.1$  Hz, H-5), 3.72 (1H, m, H-6), 3.25 (1H, dd,  $J = 11.0, 9.0$  Hz, H-2), 2.99 (1H, dq,  $J = 11.0, 7.0$  Hz, H-3), 1.65 (2H, m, H-7), 1.64 (1H, m, H-8b), 1.50 (1H, m, H-8a), 1.41 (3H, d,  $J = 7.1$  Hz, H-13), 1.32 (2H, m, H-9), 1.29 (2H, m, H-11), 1.28 (2H, m, H-10), 0.89 (3H, t,  $J = 6.9$  Hz, H-12); NMR data (Table 7.10) are consistent with those previously reported;<sup>51</sup> ESI-MS  $m/z$  257 [M – H]<sup>–</sup>, 515 [2M – H]<sup>–</sup>, 259 [M + H]<sup>+</sup>, 281 [M + Na]<sup>+</sup>, 241 [M – H<sub>2</sub>O + H]<sup>+</sup>, 517 [2M + H]<sup>+</sup>.

pos.	$\delta_{\text{C}} / \text{ppm}$	$\delta_{\text{H}} / \text{ppm} (J / \text{Hz})$	$\delta_{\text{C}} / \text{ppm}$ literature <sup>51</sup>	$\delta_{\text{H}} / \text{ppm} (J / \text{Hz})$ literature <sup>51</sup>
<b>1</b>	174.1	-	171.9	-
<b>2</b>	48.4	3.25, 1H, dd (11.0, 9.0)	48.2	3.23, 1H, dd (11.0, 9.0)
<b>3</b>	39.2	2.99, 1H, dq (11.0, 7.0)	39.2	2.99, 1H, dq (11.0, 7.1)
<b>4</b>	176.8	-	176.7	-
<b>5</b>	80.6	4.48, 1H, dd (9.0, 2.1)	80.6	4.47, 1H, dd (9.0, 2.2)
<b>6</b>	71.1	3.72, 1H, m	71.0	3.71, 1H, m
<b>7</b>	34.2	1.65, 2H, m	34.0	1.82, 2H, m
<b>8</b>	25.8	1.50, 1H, m 1.64, 1H, m	25.7	1.46, 2H, m
<b>9</b>	29.2	1.32, 2H, m	29.0	1.29, 2H, m
<b>10</b>	31.8	1.28, 2H, m	31.8	1.25, 2H, m
<b>11</b>	22.7	1.29, 2H, m	22.6	1.26, 2H, m
<b>12</b>	14.2	0.89, 3H, t (6.9)	14.2	0.88, 3H, t (7.0)
<b>13</b>	14.9	1.41, 3H, d (7.1)	14.8	-

**Table 7.10** <sup>1</sup>H NMR (500 MHz) data and <sup>13</sup>C NMR (125 MHz) data for **78** in CDCl<sub>3</sub>. Literature<sup>51</sup> data was measured at 700 MHz in CDCl<sub>3</sub>.



Figure 7.29  $^1\text{H}$  NMR of compound 78.Figure 7.30  $^{13}\text{C}$  NMR of compound 78.

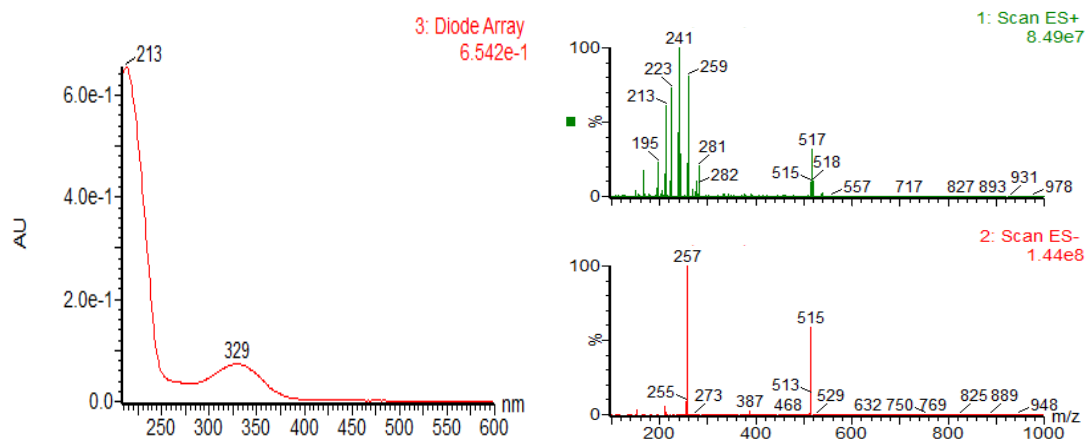
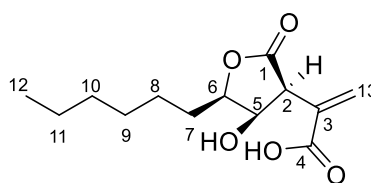


Figure 7.31 UV and mass spectra of compound 78.

### Compound 75<sup>51</sup>



75

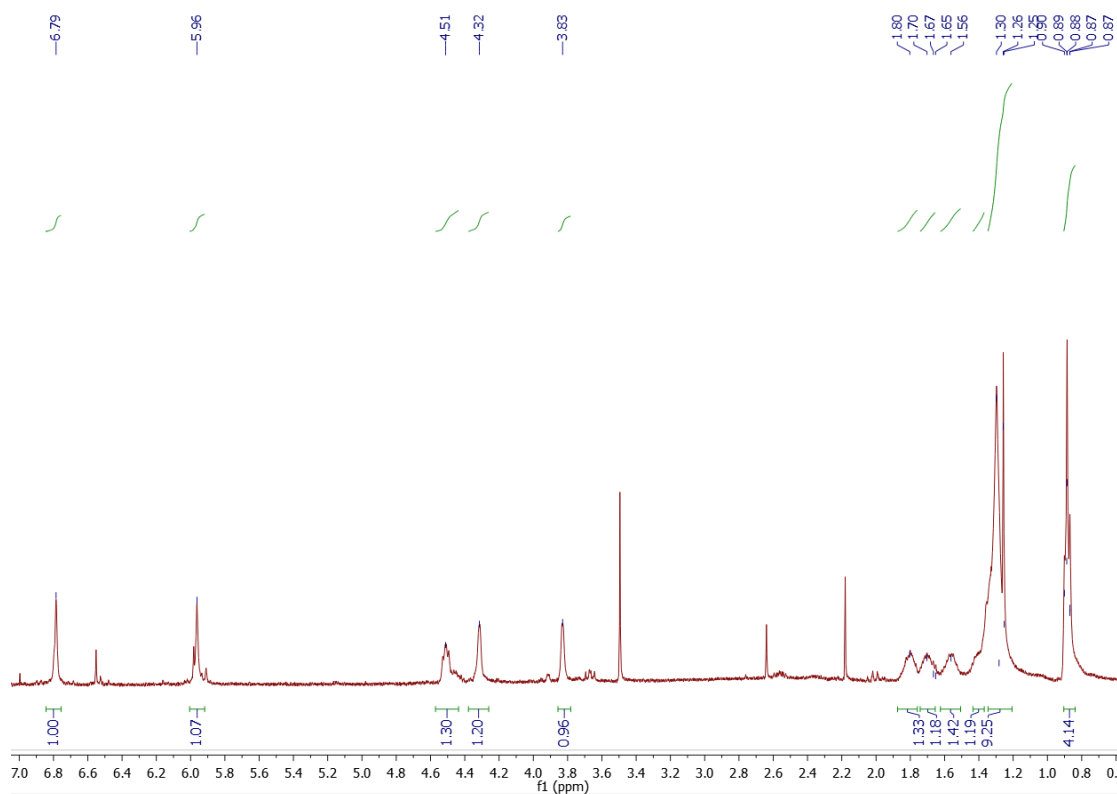
### Sporothric acid

Chemical formula:  $C_{13}H_{20}O_5$

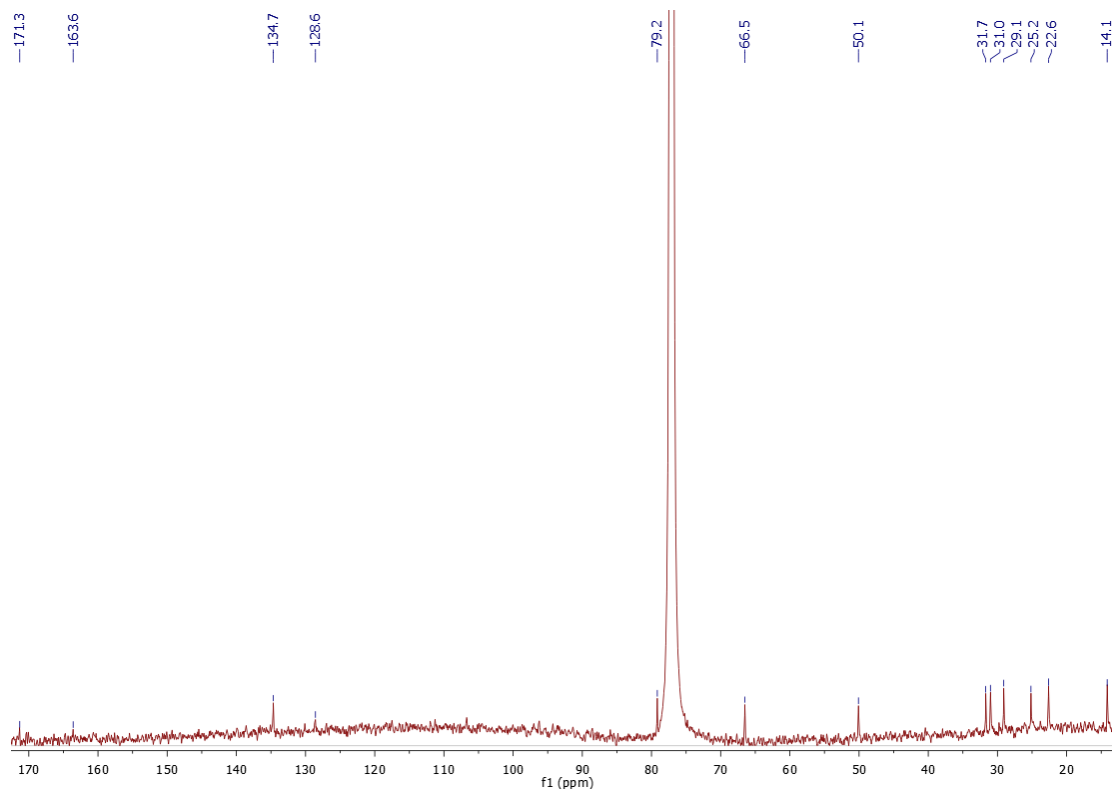
Colorless oil;  $[\alpha]_D^{25} + 5$  ( $c = 0.1$ ,  $CHCl_3$ ); Literature,<sup>51</sup>  $[\alpha]_D^{25} + 14$  ( $c = 0.1$ ,  $CHCl_3$ ); UV ( $\lambda_{max}$ ): 208 nm.  $^{13}C$  NMR data ( $CDCl_3$ , 100 MHz):  $\delta_c$  171.3 (C-4), 163.6 (C-1), 134.7 (C-13), 128.6 (C-3), 79.2 (C-6), 66.5 (C-5), 50.1 (C-2), 31.7 (C-10), 31.0 (C-7), 29.1 (C-9), 25.2 (C-8), 22.6 (C-11), 14.1 (C-12);  $^1H$  NMR data ( $CDCl_3$ , 400 MHz):  $\delta_H$  6.79 (1H, s, H-13a), 5.96 (1H, s, H-13b), 4.51 (1H, m, H-6), 4.32 (1H, m, H-5), 3.83 (1H, m, H-2), 1.80 (1H, m, H-7a), 1.70 (1H, m, H-7a), 1.56 (1H, m, H-8a), 1.40 (1H, m, H-8b), 1.27-1.34 (6H, m, H-9/H-10/H-11), 0.88 (3H, t,  $J = 6.5$  Hz, H-12); NMR data (Table 7.11) are consistent with those previously reported;<sup>51</sup> ESI-MS  $m/z$  255  $[M - H]^-$ , 511  $[2M - H]^-$ , 257  $[M + H]^+$ , 513  $[2M + H]^+$ ; HR-ESI-MS  $m/z$  255.1232  $[M - H]^-$  (calcd. for  $C_{13}H_{19}O_5$ , 255.1232).

pos.	$\delta_{\text{H}} / \text{ppm} (J / \text{Hz})$	$\delta_{\text{C}} / \text{ppm}$	$\delta_{\text{H}} / \text{ppm} (J / \text{Hz})$ literature <sup>51</sup>	$\delta_{\text{C}} / \text{ppm}$ literature <sup>51</sup>
1	-	163.6	-	163.5
2	3.83, 1H, m	50.1	3.82, 1H, dd (3.9, 1.4)	50.0
3	-	128.6	-	128.5
4	-	171.3	-	171.3
5	4.32, 1H, m	66.5	4.30, 1H, dd (3.9, 1.3)	66.5
6	4.51, 1H, m	79.2	4.51, 1H, ddd (8.6, 5.2, 1.4)	79.1
7	1.80, 1H, m 1.70, 1H, m	31.0	1.80, 1H, m 1.69, 1H, m	31.0
8	1.56, 1H, m 1.40, 1H, m	25.2	1.54, 1H, m 1.38, 1H, m	25.1
9	1.27-1.34, 2H, m	29.1	1.32, 2H, m	29.0
10	1.27-1.34, 2H, m	31.7	1.27, 2H, m	31.6
11	1.27-1.34, 2H, m	22.6	1.28, 2H, m	22.6
12	0.88, 3H, t (6.5)	14.1	0.88, 3H, t (7.0)	14.1
13	6.79, 1H, s 5.96, 1H, s	134.7	6.78, 1H, s 5.95, 1H, s	134.6

**Table 7.11**  $^1\text{H}$  NMR (400 MHz) and  $^{13}\text{C}$  NMR (100 MHz) data for **75** in  $\text{CDCl}_3$ . Literature<sup>51</sup> data was measured at 700 MHz in  $\text{CDCl}_3$ .



**Figure 7.32**  $^1\text{H}$  NMR of compound **75**.

Figure 7.33  $^{13}\text{C}$  NMR of compound 75.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

656 formula(e) evaluated with 9 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-100 H: 0-120 N: 0-15 O: 0-20 Na: 0-1

Tian

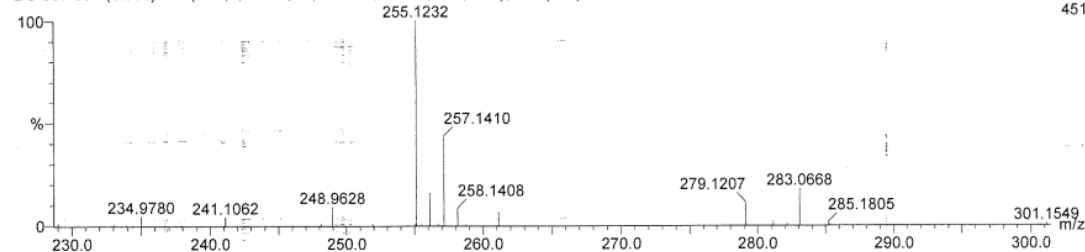
Q-ToF Premier UPLC-MS

13-Jan-2020 11:52:55

DO 057 651 (6.664) AM (Cen,4, 40.00, Ar,11000.0,554.26,0.55,LS 5); Cm (651)

1: TOF MS ES-

451

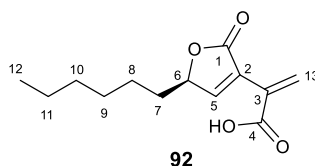


Minimum: -0.5  
Maximum: 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
255.1232	255.1232	0.0	0.0	4.5	90.3	C13 H19 O5
	255.1222	1.0	3.9	6.5	94.1	C12 H16 N4 O Na
	255.1246	-1.4	-5.5	9.5	92.4	C14 H15 N4 O
	255.1208	2.4	9.4	1.5	93.3	C11 H20 O5 Na
	255.1206	2.6	10.2	5.5	95.7	C9 H15 N6 O3
	255.1267	-3.5	-13.7	3.5	108.6	C2 H12 N14 Na
	255.1192	4.0	15.7	0.5	96.5	C8 H19 N2 O7
	255.1278	-4.6	-18.0	1.5	104.1	C3 H15 N10 O4
	255.1182	5.0	19.6	2.5	100.5	C7 H16 N6 O3 Na

Figure 7.34 HR-ESI-MS of compound 75.

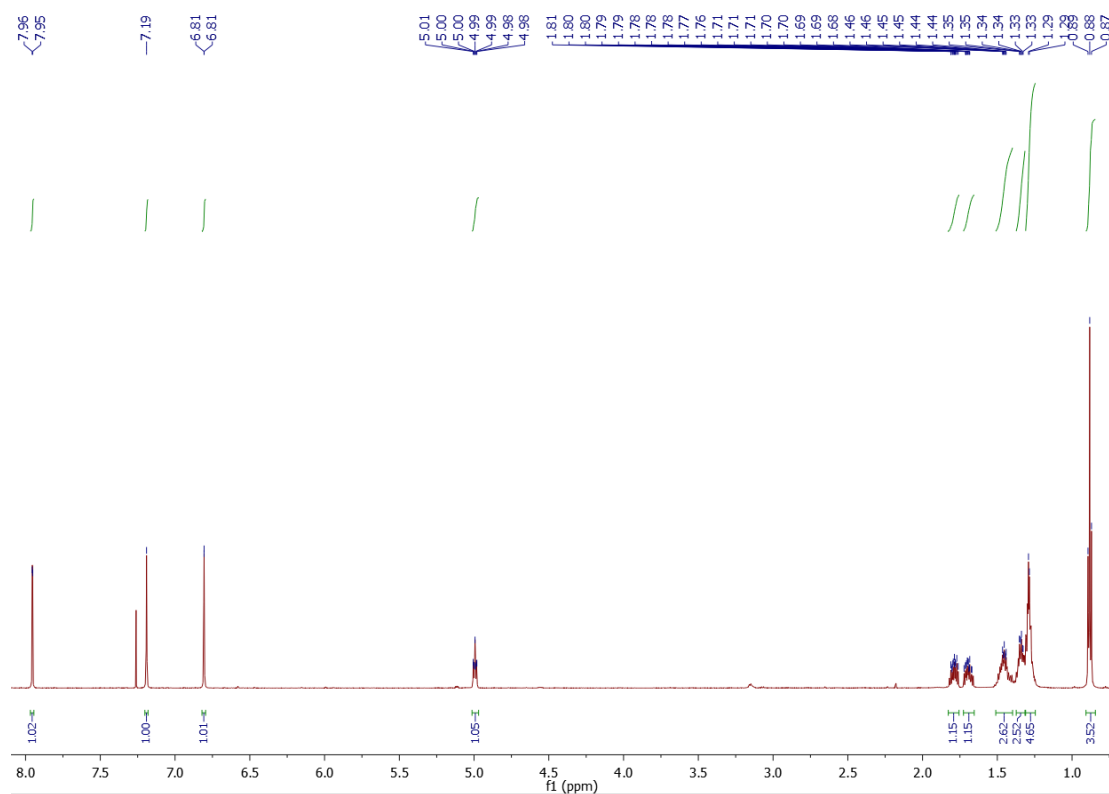
## Compound 92



## Dehydrodeoxysporothric acid

Chemical formula:  $C_{13}H_{18}O_4$ 

White powder;  $[\alpha]_D^{25} -25$  ( $c = 0.83$ , MeOH); UV ( $\lambda_{max}$ ): 229 nm.  $^{13}C$  NMR data ( $CDCl_3$ , 150 MHz):  $\delta_c$  171.7 (C-1), 170.4 (C-4), 153.4 (C-5), 133.7 (C-13), 128.6 (C-3), 124.9 (C-2), 80.8 (C-6), 33.4 (C-7), 31.7 (C-10), 29.1 (C-9), 25.1 (C-8), 22.6 (C-11), 14.1 (C-12);  $^1H$  NMR data ( $CDCl_3$ , 600 MHz):  $\delta_H$  7.95 (1H, d,  $J = 1.7$  Hz, H-5), 7.19 (1H, s, H-13a), 6.81 (1H, s, H-13b), 4.99 (1H, m, H-6), 1.79 (1H, m, H-7a), 1.69 (1H, m, H-7b), 1.46 (2H, m, H-8), 1.34 (2H, m, H-9), 1.29 (2H, m, H-11), 1.28 (2H, m, H-10), 0.88 (3H, t,  $J = 7.0$  Hz, H-12); NMR data see Table 2.13; ESI-MS  $m/z$  237  $[M - H]^-$ , 475  $[2M - H]^-$ , 239  $[M + H]^+$ , 261  $[M + Na]^+$ . HR-ESI-MS  $m/z$  237.1127  $[M - H]^-$  (calcd. for  $C_{13}H_{17}O_4$ , 237.1127).

Figure 7.35  $^1H$  NMR of compound 92.

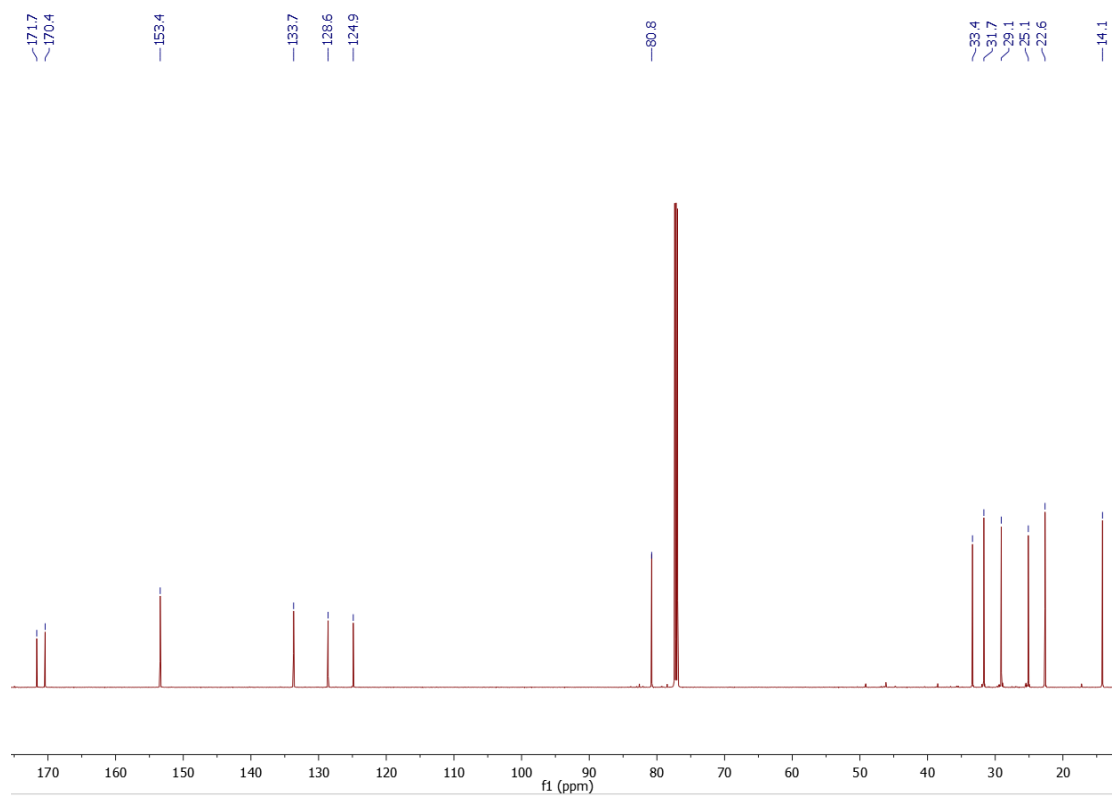
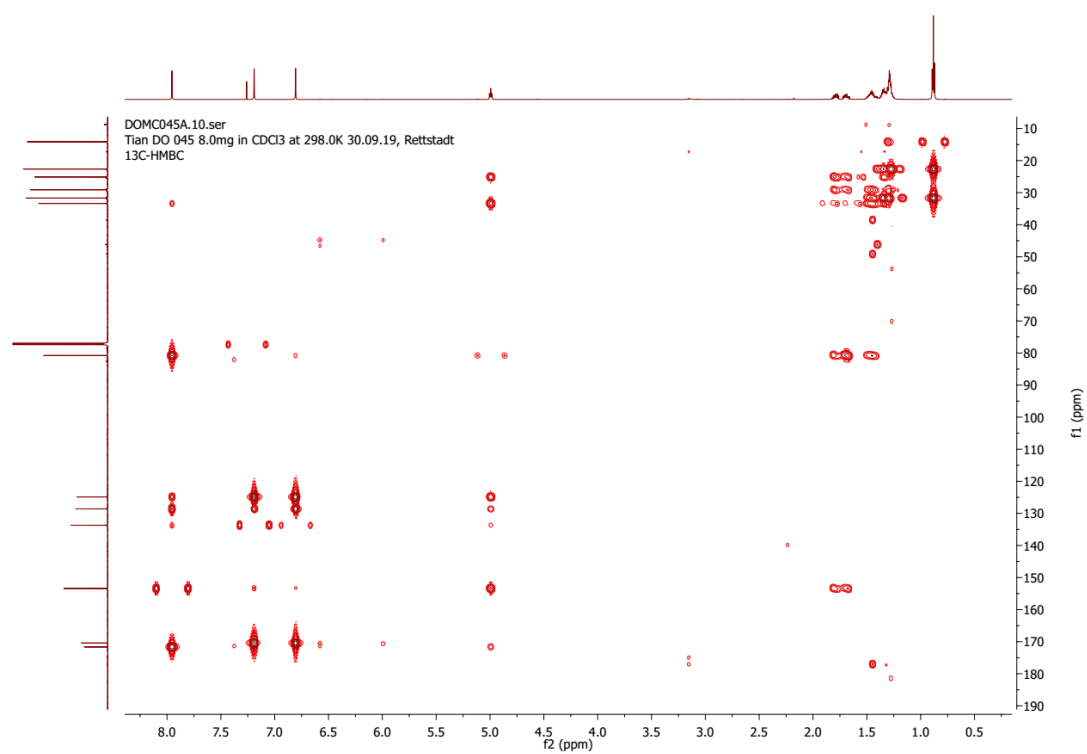
Figure 7.36  $^{13}\text{C}$  NMR of compound 92.

Figure 7.37 HMBC of compound 92.

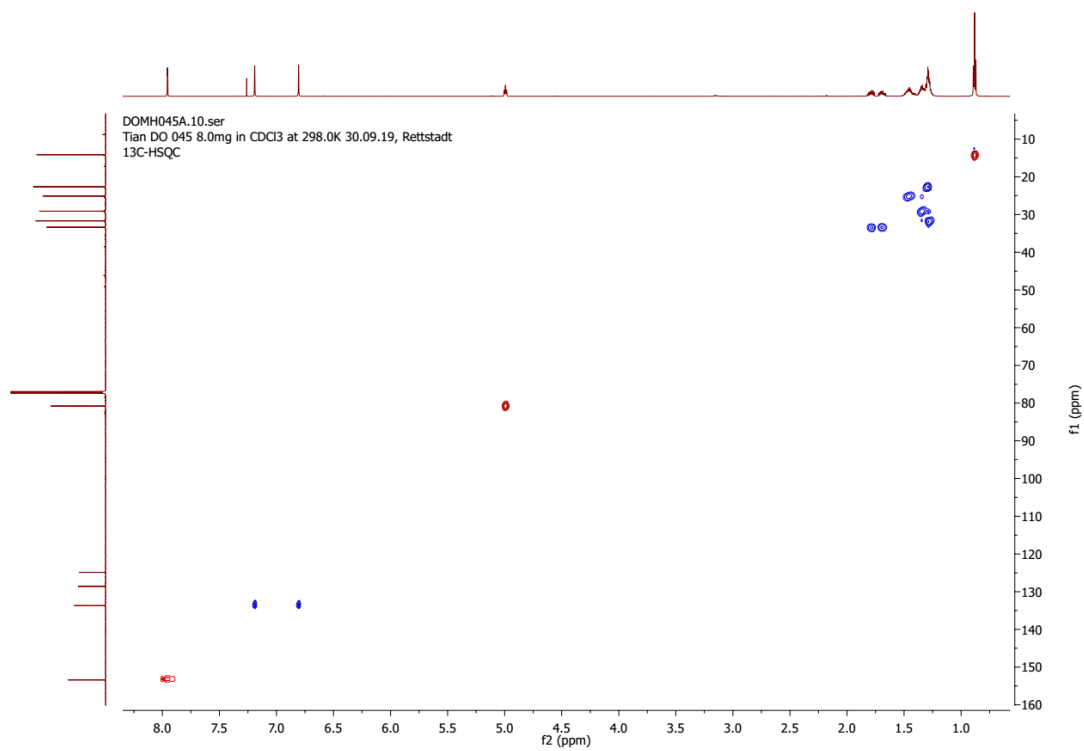
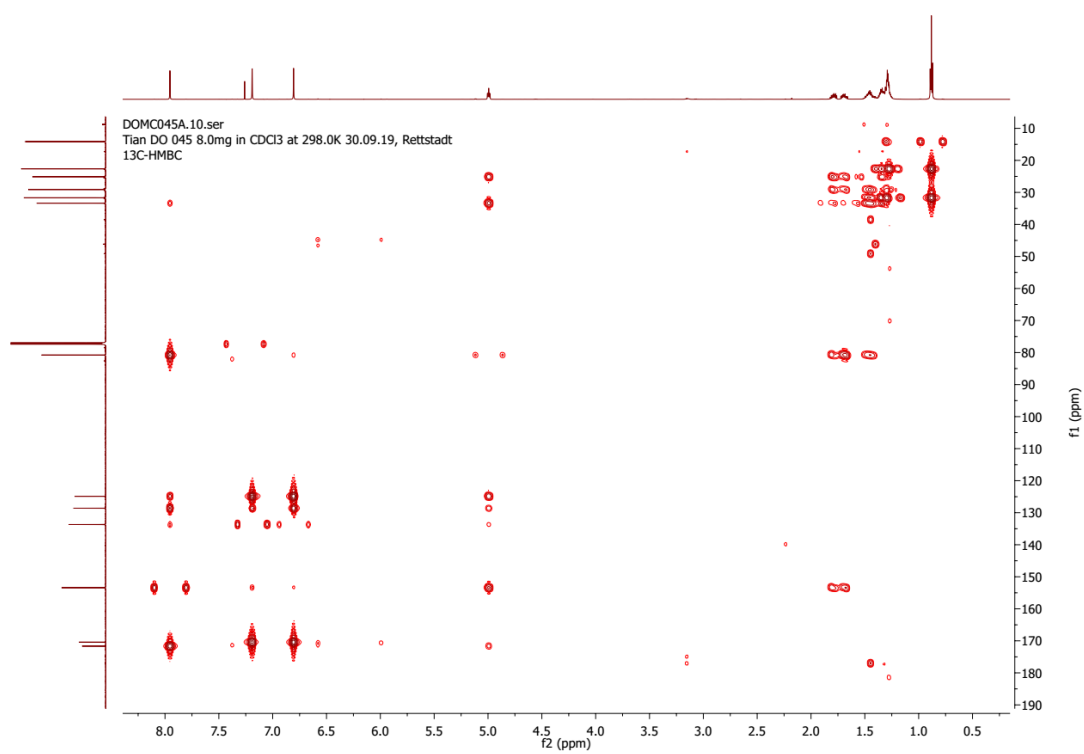


Figure 7.38 HSQC of compound 92.

Figure 7.39  $^1\text{H}$ - $^1\text{H}$  COSY of compound 92.

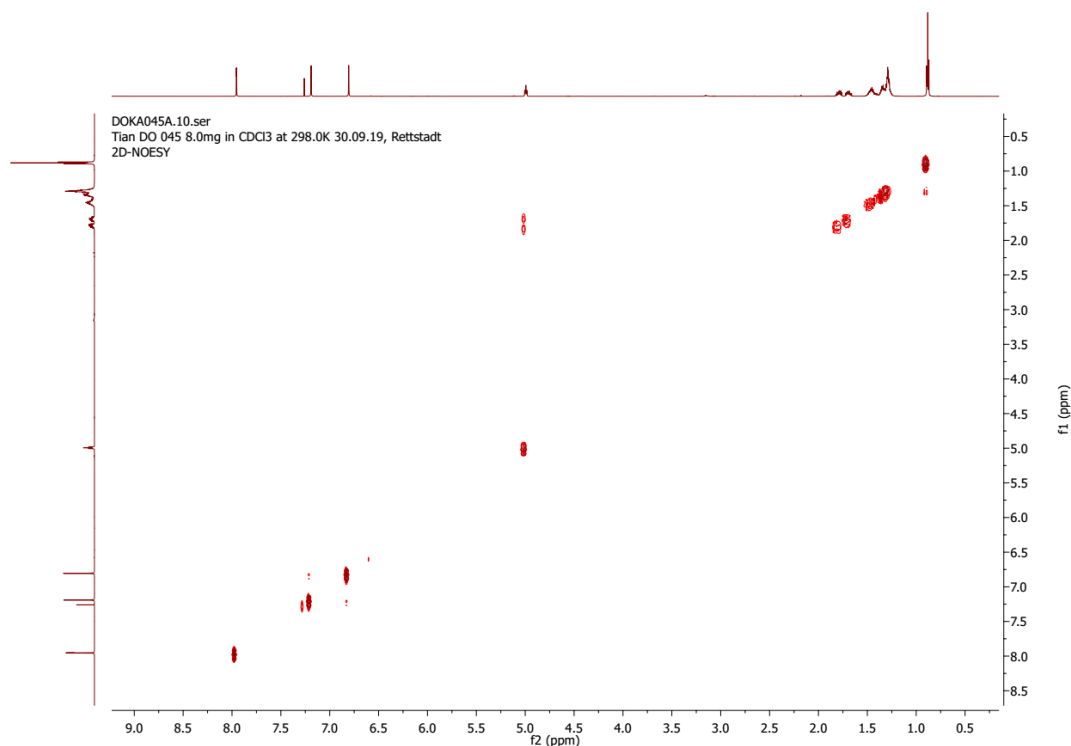


Figure 7.40 NOESY of compound 92.

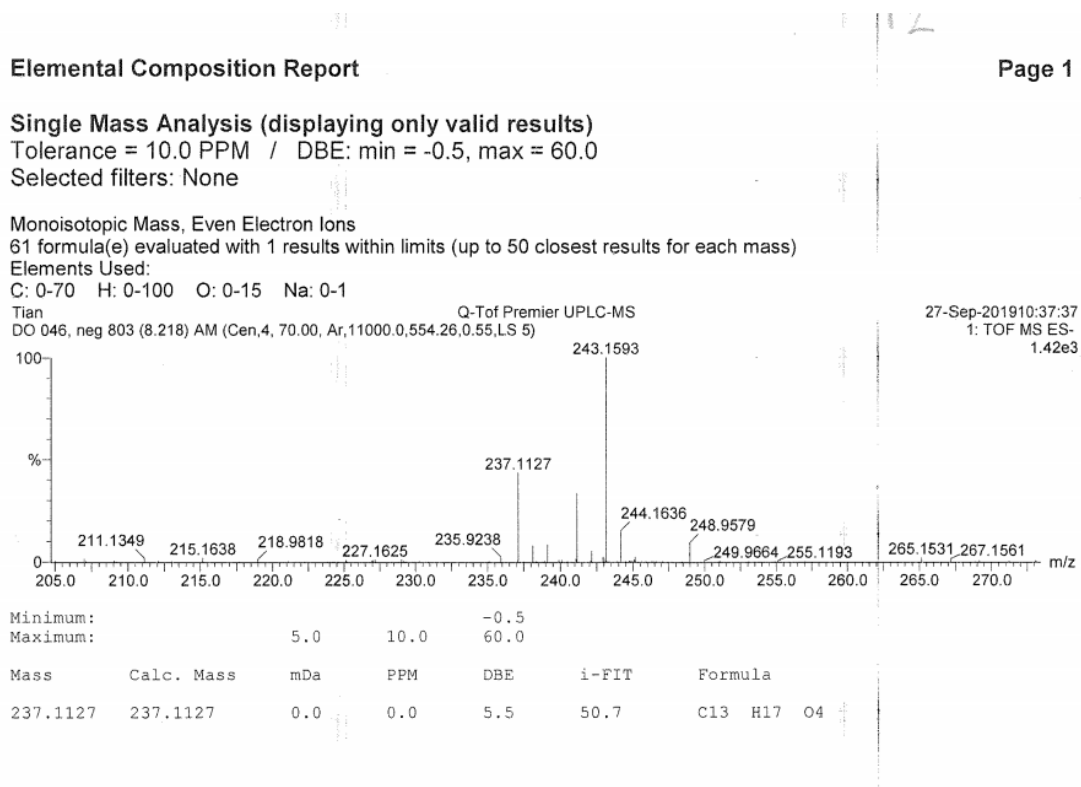
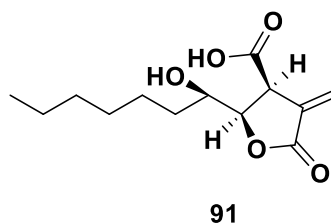


Figure 7.41 HR-ESI-MS of compound 92.

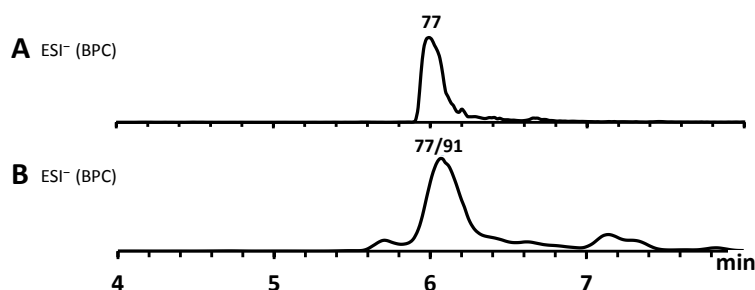


## Compound 91



**Chemical formula: C<sub>13</sub>H<sub>20</sub>O<sub>5</sub>**

UV ( $\lambda_{\text{max}}$ ): 203 nm. ESI-MS  $m/z$  237 [M – H<sub>2</sub>O – H]<sup>–</sup>, 255 [M – H]<sup>–</sup>, 511 [2M – H]<sup>–</sup>, 257 [M + H]<sup>+</sup>, 513 [2M + H]<sup>+</sup>. HR-ESI-MS  $m/z$  255.1234 [M – H]<sup>–</sup> (calcd. for C<sub>13</sub>H<sub>19</sub>O<sub>5</sub>, 255.1232).



**Figure 7.42** A, ESI spectrum of standard 77; B, ESI spectrum of standard 91 dissolved in CHCl<sub>3</sub> for 24 hrs.

## Elemental Composition Report

14

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

Monoisotopic Mass, Even Electron Ions

499 formula(e) evaluated with 6 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 N: 0-8 O: 0-14 Na: 0-1

Tian

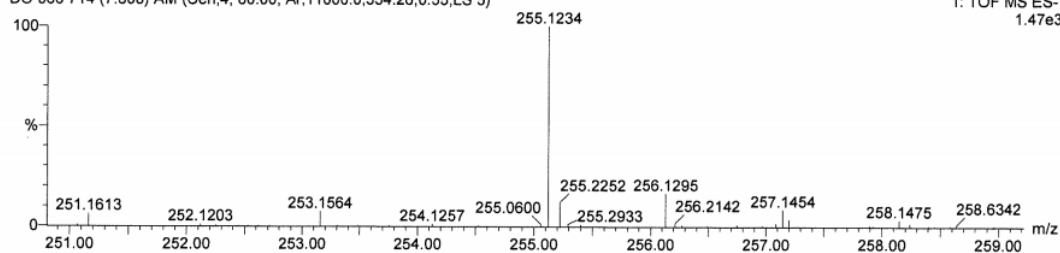
Q-ToF Premier UPLC-MS

06-May-2020 15:05:00

DO 066 714 (7.308) AM (Cen,4, 80.00, Ar,11000.0,554.26,0.55,LS 5)

1: TOF MS ES-

1.47e3

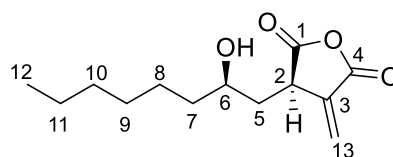


Minimum: -0.5  
Maximum: 5.0 20.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
255.1234	255.1232	0.2	0.8	4.5	35.9	C13 H19 O5
	255.1246	-1.2	-4.7	9.5	40.7	C14 H15 N4 O
	255.1222	1.2	4.7	6.5	45.4	C12 H16 N4 O Na
	255.1208	2.6	10.2	1.5	44.6	C11 H20 O5 Na
	255.1206	2.8	11.0	5.5	50.7	C9 H15 N6 O3
	255.1192	4.2	16.5	0.5	54.9	C8 H19 N2 O7

**Figure 7.43** HR-ESI-MS of compound 91.

## Compound 90



90

## Sporodride A

Chemical formula: C<sub>13</sub>H<sub>20</sub>O<sub>4</sub>

White powder;  $[\alpha]_D^{21} - 6$  ( $c = 0.32$ , CHCl<sub>3</sub>); UV ( $\lambda_{\max}$ ): 211 nm. <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 100 MHz):  $\delta_c$  175.9 (C-1), 164.6 (C-4), 132.5 (C-13), 130.9 (C-3), 77.4 (C-6), 42.3 (C-2), 35.7 (C-7), 31.8 (C-10), 30.7 (C-5), 29.2 (C-9), 24.8 (C-8), 22.7 (C-11), 14.2 (C-12); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 400 MHz):  $\delta_H$  6.59 (1H, s, H-13a), 5.87 (1H, s, H-13b), 4.55 (1H, m, H-6), 3.72 (1H, m, H-2), 2.31 (1H, dt,  $J = 2.6, 14.4$  Hz, H-5b), 1.90 (1H, ddd,  $J = 5.8, 11.1, 14.4$  Hz, H-5a), 1.71 (1H, m, H-7b), 1.60 (1H, m, H-7a), 1.50 (1H, m, H-8a), 1.39 (1H, m, H-8b), 1.29 (2H, m, H-11), 1.28 (2H, m, H-9), 1.27 (2H, m, H-10), 0.88 (3H, t,  $J = 6.8$  Hz, H-12); NMR data see Table 7.12; ESI-MS  $m/z$  239 [M – H]<sup>–</sup>, 479 [2M – H]<sup>–</sup>, 241 [M + H]<sup>+</sup>, 263 [M + Na]<sup>+</sup>. HR-ESI-MS  $m/z$  239.1282 [M – H]<sup>–</sup> (calcd. for C<sub>13</sub>H<sub>19</sub>O<sub>4</sub>, 239.1283).

pos.	$\delta_c$ / ppm	$\delta_H$ / ppm ( $J$ / Hz)	HMBC (H to C)	<sup>1</sup> H- <sup>1</sup> H COSY	NOESY
1	175.9	-	-	-	-
2	42.3	3.72, 1H, m	1, 3, 4, 5, 6, 13	5	6
3	130.9	-	-	-	-
4	164.6	-	-	-	-
5	30.7	1.90, 1H, ddd (5.8, 11.1, 14.4) 2.31, 1H, dt (2.6, 14.4)	1, 2, 6, 7	2, 6	6, 13
6	77.4	4.55, 1H, m	2, 7, 8	5, 7	2, 5, 13
7	35.7	1.60, 1H, m 1.71, 1H, m	5, 6, 8, 9	6, 8	-
8	24.8	1.50, 1H, m 1.39, 1H, m	7, 9, 10	7, 9	-
9	29.2	1.28, 2H, m	8, 10, 11	8, 10	-
10	31.8	1.27, 2H, m	9, 11	9, 11	-
11	22.7	1.29, 2H, m	10, 12	10, 12	-
12	14.2	0.88, 3H, t (6.8)	10, 11	11	-
13	132.5	6.59, 1H, s 5.87, 1H, s	2, 4	-	5, 6

Table 7.12 <sup>1</sup>H NMR (400 MHz) data and <sup>13</sup>C NMR (100 MHz) data for 90 in CDCl<sub>3</sub>.

Compound 90 was isolated as a white powder, with molecular formula C<sub>13</sub>H<sub>20</sub>O<sub>4</sub> (calc. [M – H]<sup>–</sup> HRMS 239.1283, measured 239.1282) which is same as deoxysporothric acid 76, also indicating

4 degrees of unsaturation. Analysis the 1D and 2D NMR data (Table 7.12) revealed that **90** contains the same number of methyl, methylene, methine, and quaternary carbons as **76**, in addition the proton and carbon chemical shifts are just slightly shift compared with **76**. This suggests **90** to be an isomer of **76**. However, the key evidence of the lactone scaffold with HMBC correlations from H-6 to C-1 and C-4 was not shown in **90** (Figure 7.44). This information suggest the only possibility of an anhydride scaffold for **90**. NOESY correlations of H-13 and H-5, as well H-13 and H-6 show the rotation of the five-membered anhydride ring. In addition the NOESY correlation of H-2 and H-6 establish the relative configuration of **90** (Figure 7.44). We designated **90** as sporodride A, to serve as the first sporothriolides anhydride structure.

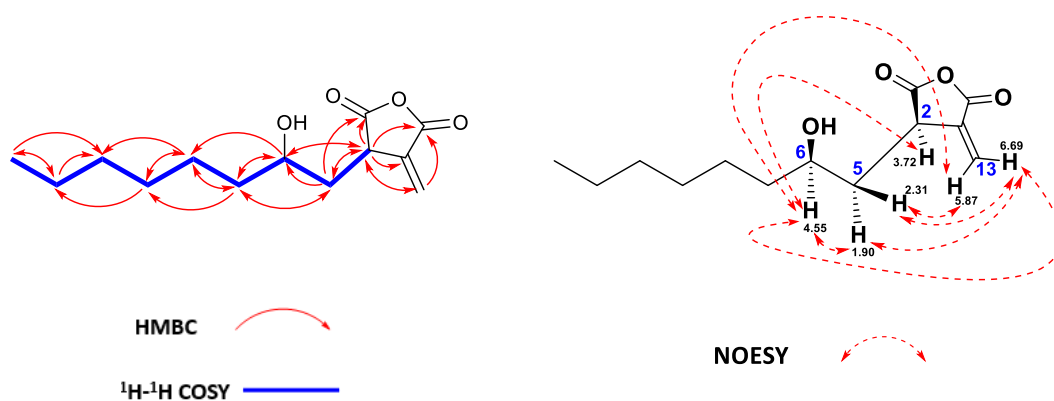


Figure 7.44 Key HMBC,  $^1\text{H}$ - $^1\text{H}$  COSY and key NOESY correlations of **90**.

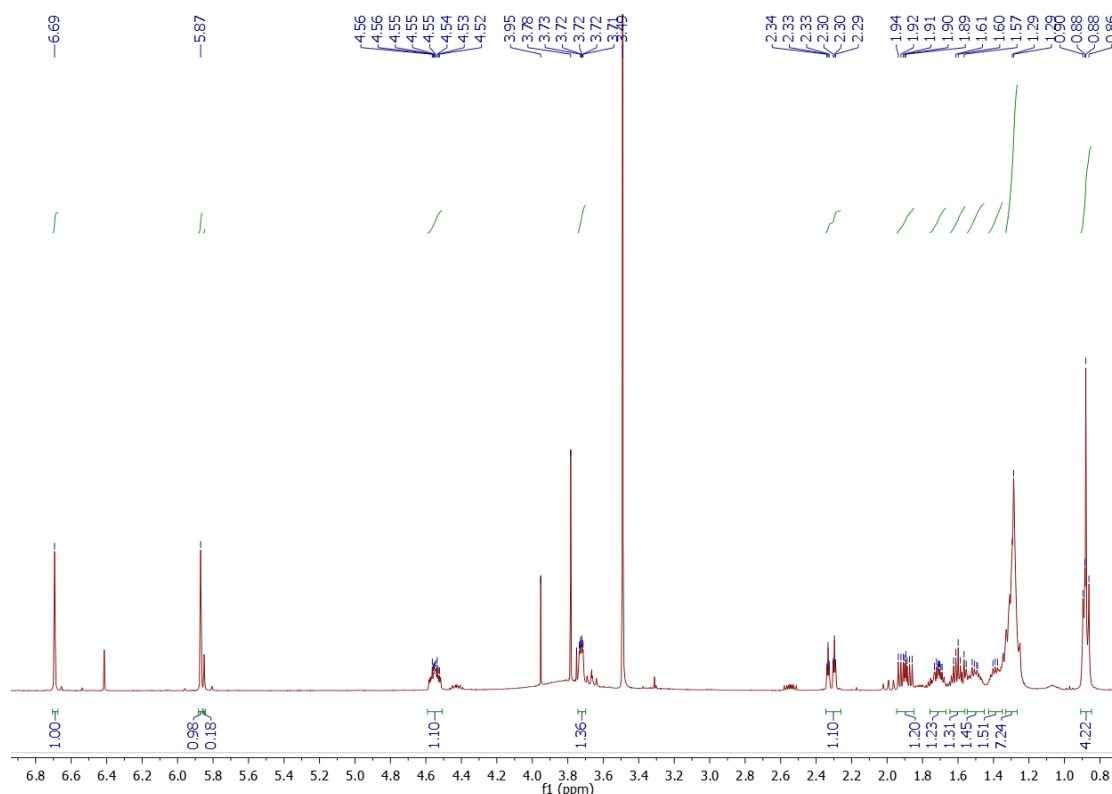


Figure 7.45  $^1\text{H}$  NMR of compound **90**.

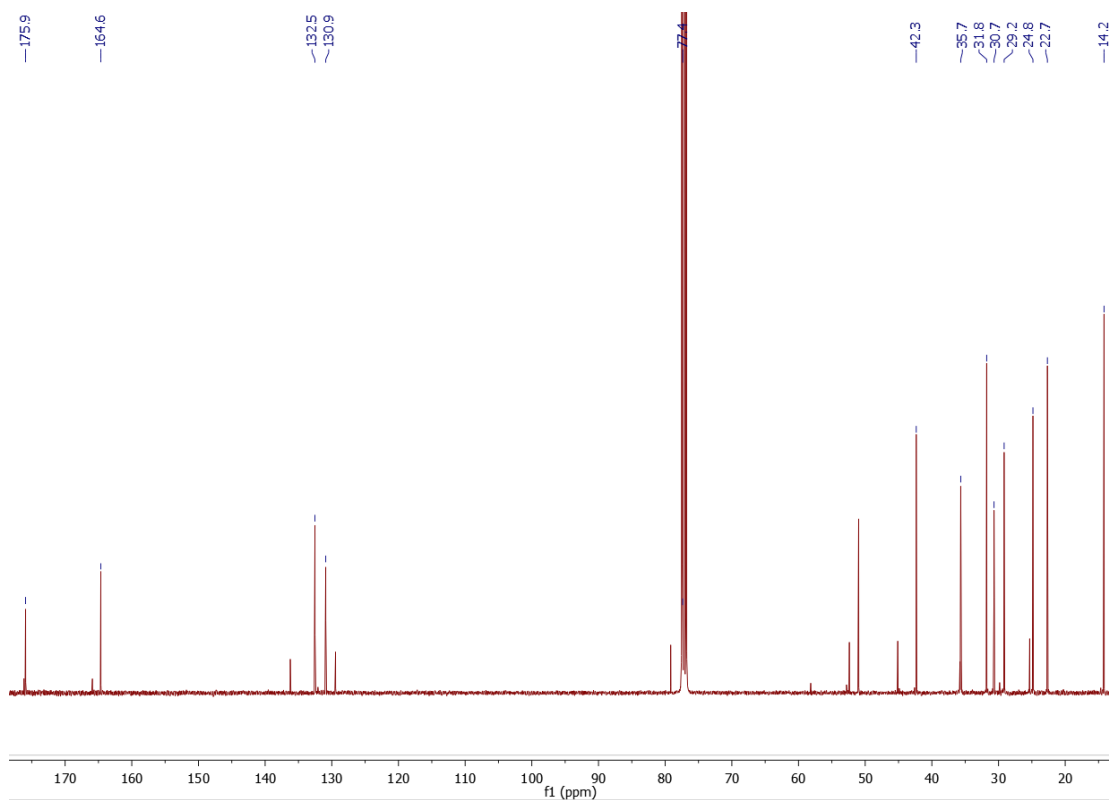
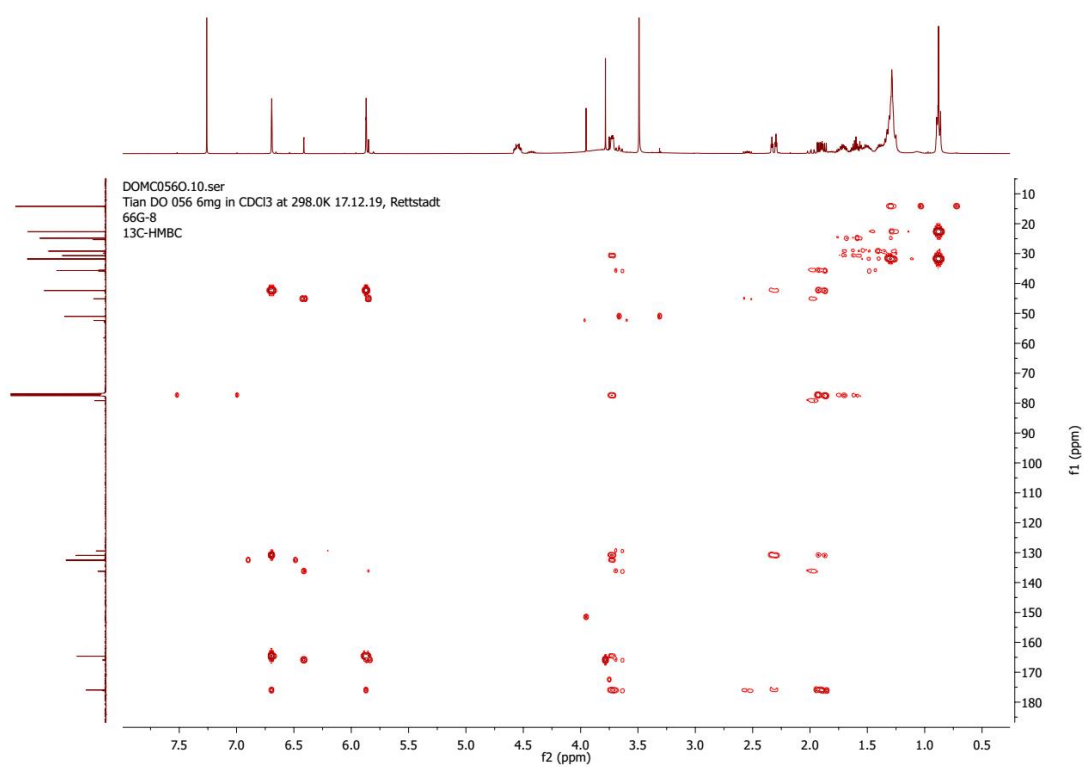
Figure 7.46  $^{13}\text{C}$  NMR of compound 90.

Figure 7.47 HMBC of compound 90.

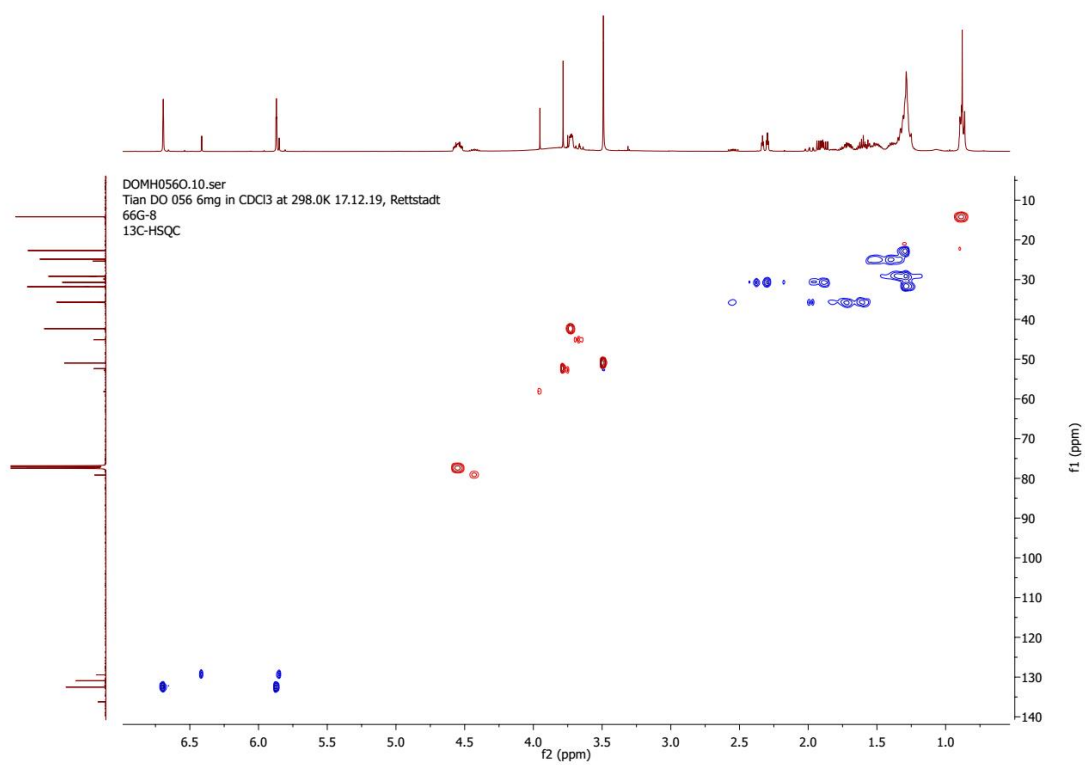
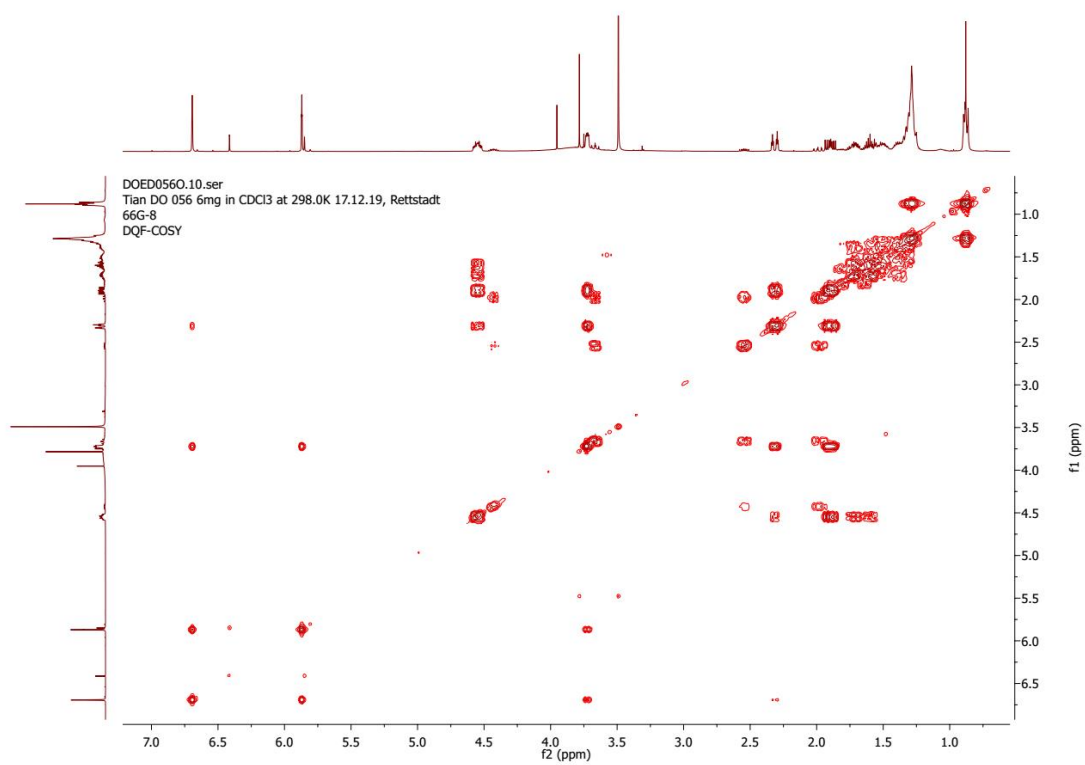


Figure 7.48 HSQC of compound 90.

Figure 7.49  $^1\text{H}$ - $^1\text{H}$  COSY of compound 90.

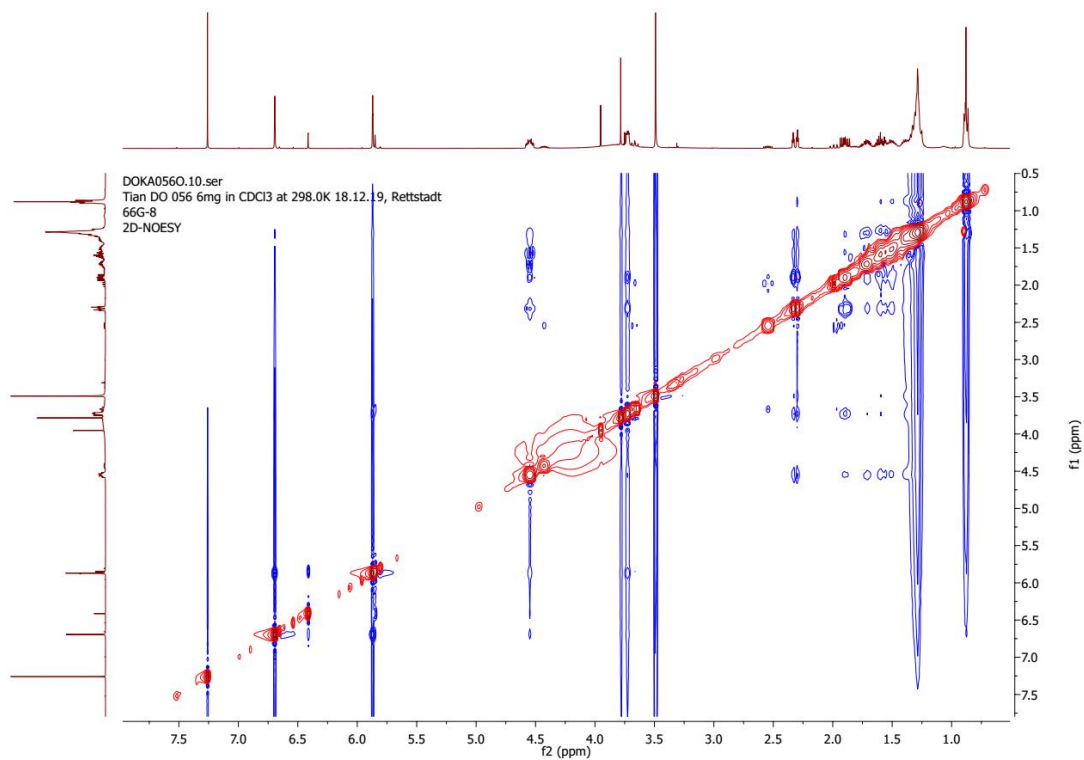


Figure 7.50 NOESY of compound 90.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

Monoisotopic Mass, Even Electron Ions

558 formula(e) evaluated with 7 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-100 H: 0-120 N: 0-15 O: 0-20 Na: 0-1

Tian

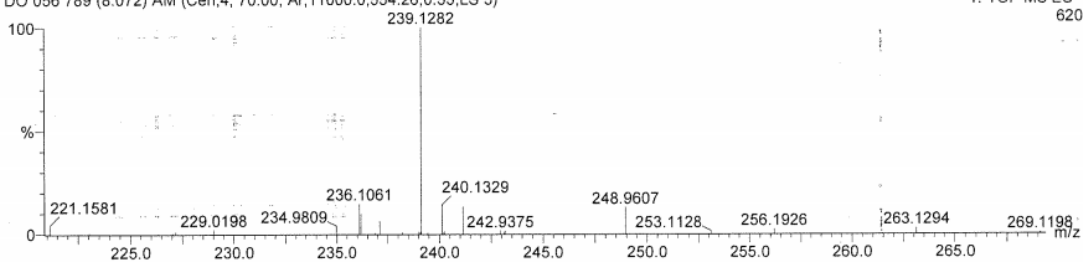
Q-ToF Premier UPLC-MS

DO 056 789 (8.072) AM (Cen,4, 70.00, Ar,11000.0,554.26,0.55,LS 5)

13-Jan-2020 11:37:01

1: TOF MS ES-

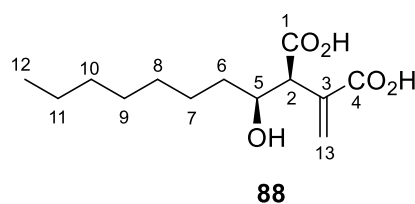
620



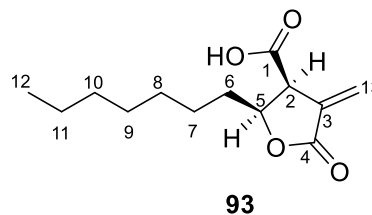
Minimum: -0.5  
Maximum: 5.0 20.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
239.1282	239.1283	-0.1	-0.4	4.5	31.0	C13 H19 O4
	239.1273	0.9	3.8	6.5	35.5	C12 H16 N4 Na
	239.1297	-1.5	-6.3	9.5	35.4	C14 H15 N4
	239.1259	2.3	9.6	1.5	33.0	C11 H20 O4 Na
	239.1256	2.6	10.9	5.5	35.6	C9 H15 N6 O2
	239.1243	3.9	16.3	0.5	35.3	C8 H19 N2 O6
	239.1329	-4.7	-19.7	1.5	45.8	C3 H15 N10 O3

Figure 7.51 HR-ESI-MS of compound 90.

Compounds **93** and **88**

Hydroxyalkylitaconic acid A  
Chemical formula: C<sub>13</sub>H<sub>22</sub>O<sub>5</sub>



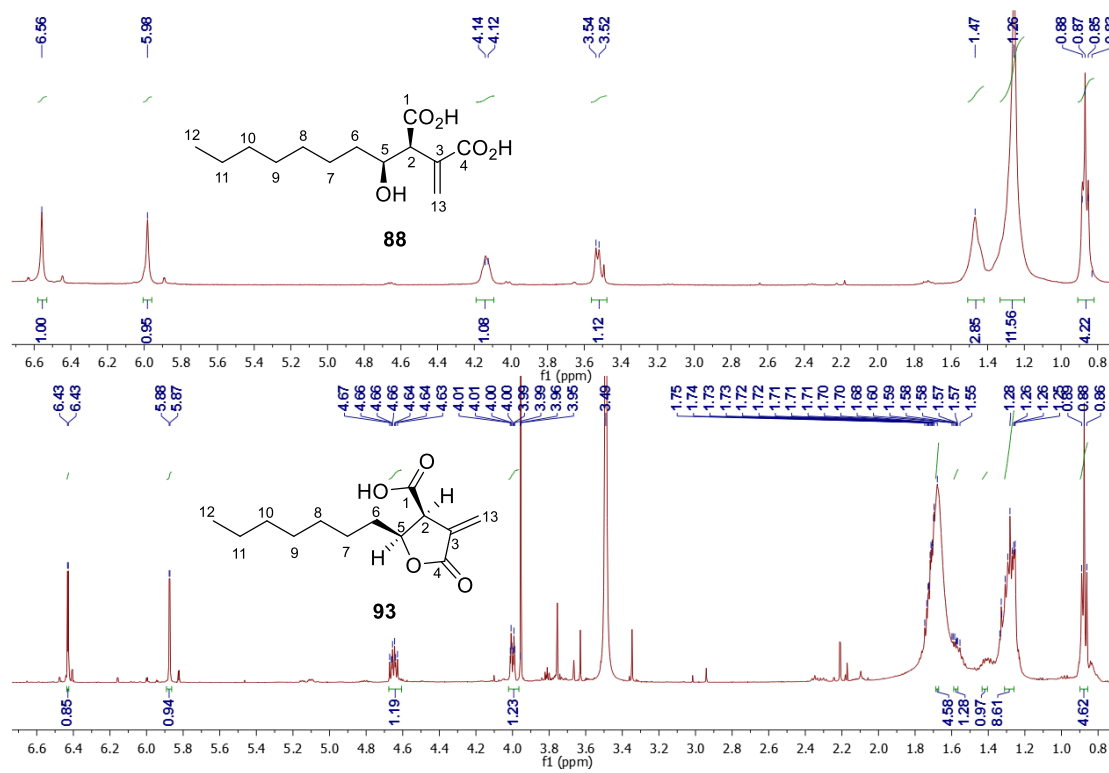
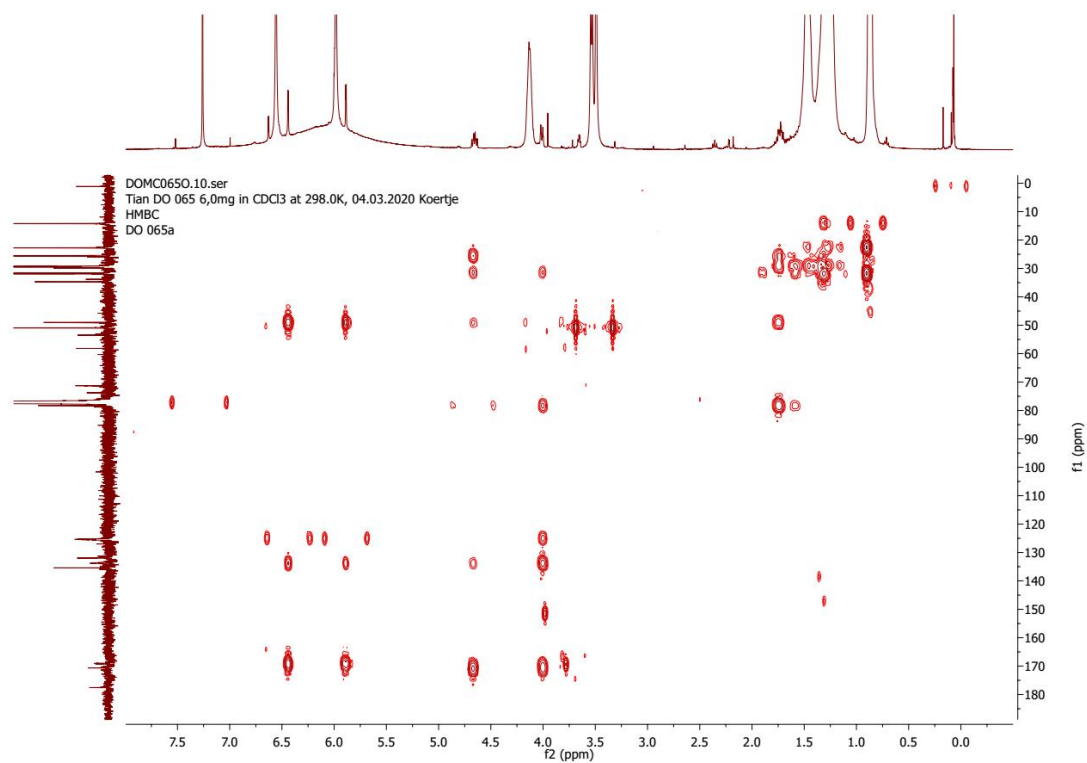
Epideoxyisoporothric acid  
Chemical formula: C<sub>13</sub>H<sub>20</sub>O<sub>4</sub>

Compound **88**

White powder;  $[\alpha]_D^{21} - 55$  ( $c = 0.12$ , CHCl<sub>3</sub>); UV ( $\lambda_{\max}$ ): 203 nm; <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 100 MHz):  $\delta_c$  177.6 (C-1), 169.1 (C-4), 132.0 (C-13), 135.4 (C-3), 34.7 (C-6), 53.5 (C-2), 25.5 (C-7), 31.9 (C-10), 73.8 (C-5), 29.5 (C-9), 29.4 (C-8), 22.8 (C-11), 14.2 (C-12); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 400 MHz):  $\delta_H$  6.56 (1H, s, H-13a), 5.98 (1H, s, H-13b), 1.43-1.51 (2H, m, H-6), 3.54 (1H, d,  $J = 6.8$  Hz, H-2), 4.13 (1H, m, H-5), 1.21-1.35 (10H, m, H-7/H-8/H-9/H-10/H-11), 0.87 (3H, t,  $J = 6.5$  Hz, H-12); NMR data see Table 2.11; ESI-MS  $m/z$  257 [M – H]<sup>–</sup>, 515 [2M – H]<sup>–</sup>, 259 [M + H]<sup>+</sup>, 517 [2M + H]<sup>+</sup>. HR-ESI-MS  $m/z$  257.1386 [M – H]<sup>–</sup> (calcd. for C<sub>13</sub>H<sub>21</sub>O<sub>5</sub>, 257.1389).

Compound **93**

White powder;  $[\alpha]_D^{21} - 91$  ( $c = 0.35$ , CHCl<sub>3</sub>); UV ( $\lambda_{\max}$ ): 211 nm; <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 125 MHz):  $\delta_c$  170.8 (C-1), 169.0 (C-4), 133.7 (C-3), 125.4 (C-13), 78.2 (C-5), 48.9 (C-2), 31.8 (C-10), 31.6 (C-6), 29.3 (C-9), 29.2 (C-8), 25.7 (C-7), 22.7 (C-11), 14.2 (C-12); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 500 MHz):  $\delta_H$  6.43 (1H, s, H-13a), 5.87 (1H, s, H-13b), 4.65 (1H, m, H-5), 4.00 (1H, dt,  $J = 2.2, 7.7$  Hz, H-2), 1.71 (2H, m, H-6), 1.55 (1H, m, H-7b), 1.41 (1H, m, H-7a), 1.32 (2H, m, H-9), 1.29 (2H, m, H-8), 1.28 (2H, m, H-11), 1.26 (2H, m, H-10), 0.88 (3H, t,  $J = 7.2$  Hz, H-12); NMR data see Table 2.11; ESI-MS  $m/z$  239 [M – H]<sup>–</sup>, 479 [2M – H]<sup>–</sup>, 241 [M + H]<sup>+</sup>, 481 [2M + H]<sup>+</sup>, 503 [2M + Na]<sup>+</sup>. HR-ESI-MS  $m/z$  239.1282 [M – H]<sup>–</sup> (calcd. for C<sub>13</sub>H<sub>19</sub>O<sub>4</sub>, 239.1283).

Figure 7.52  $^1\text{H}$  NMR of compound **93** and **88**.Figure 7.53 HMBC of compound **93**.



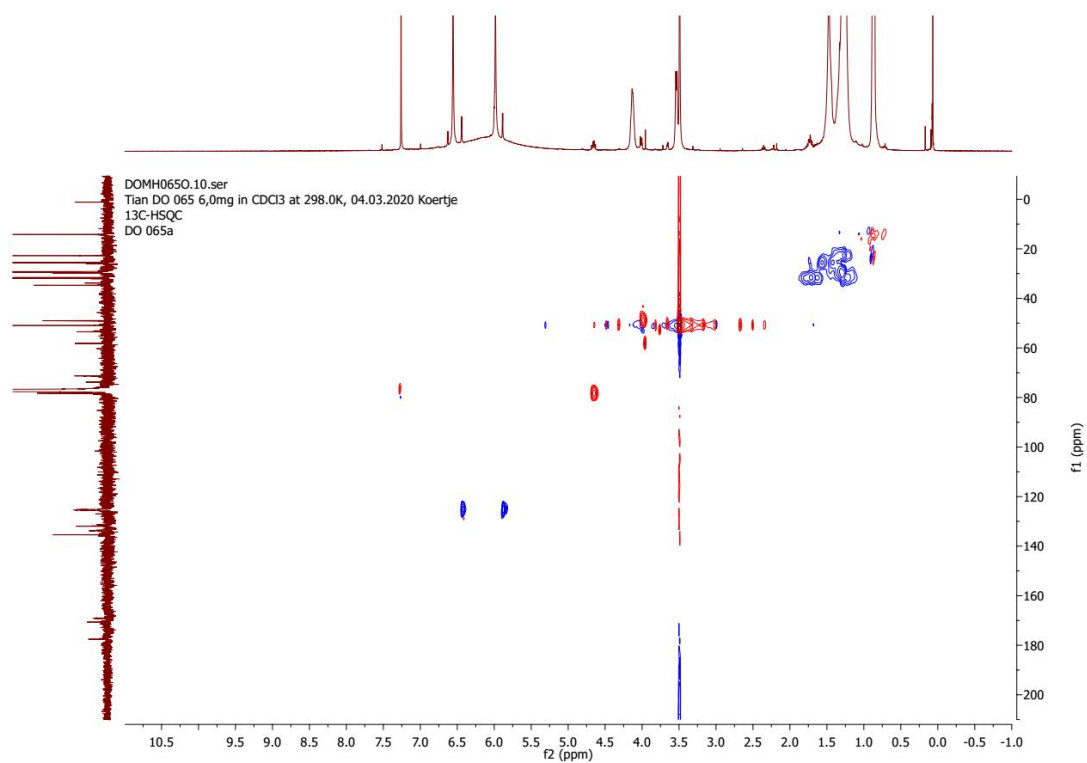
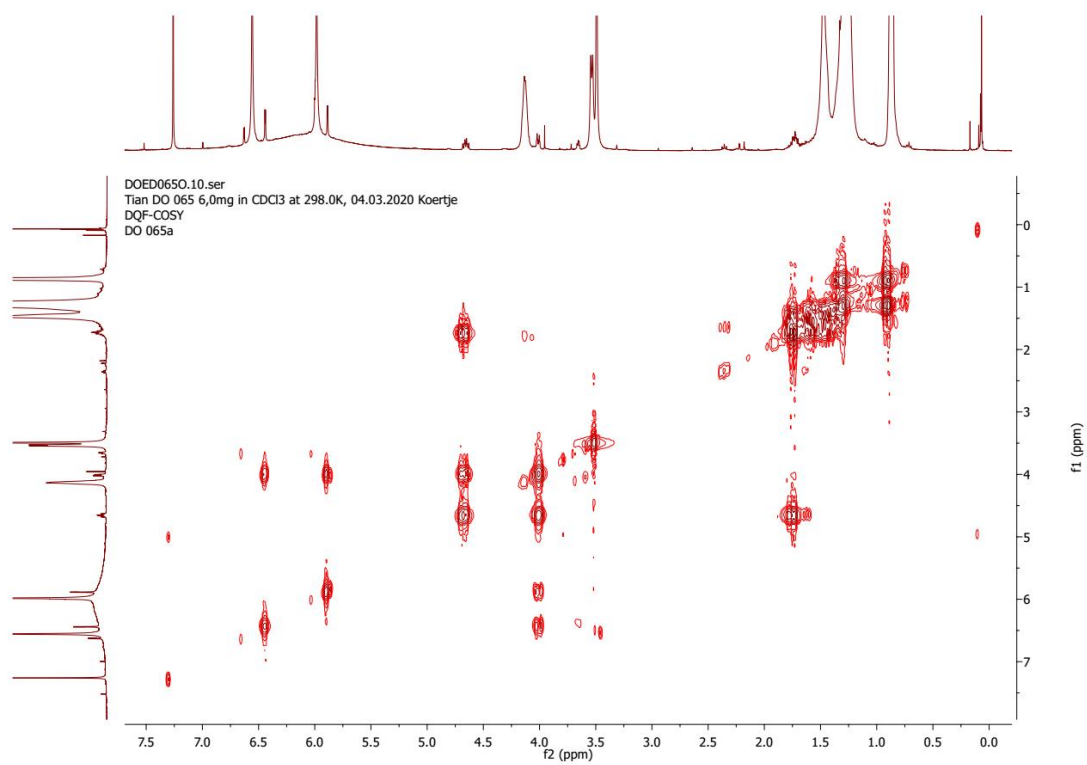


Figure 7.54 HSQC of compound 93.

Figure 7.55 <sup>1</sup>H-<sup>1</sup>H COSY of compound 93.

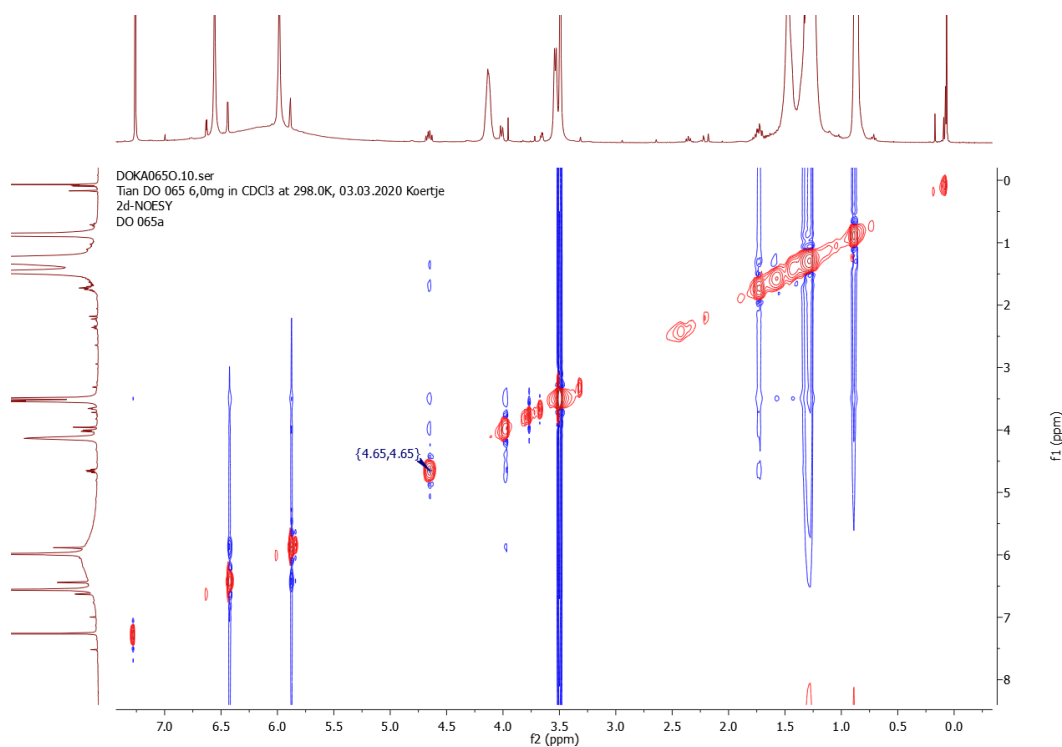


Figure 7.56 NOESY of compound 93.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

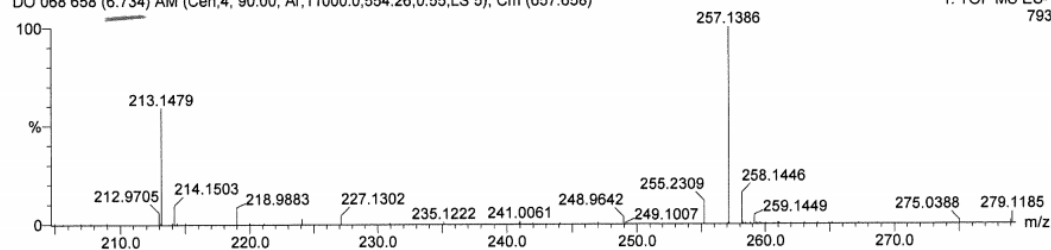
Selected filters: None

Monoisotopic Mass, Even Electron Ions

38 formula(e) evaluated with 1 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 O: 0-14

Tian Q-ToF Premier UPLC-MS  
DO 068 658 (6.734) AM (Cen,4, 90.00, Ar,11000.0,554.26,0.55,LS 5); Cm (657:658)06-May-2020 21:17  
1: TOF MS ES-  
793

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
257.1386	257.1389	-0.3	-1.2	3.5	4.5	C13 H21 O5

Figure 7.57 HR-ESI-MS of compound 88.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -1.5, max = 50.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

473 formula(e) evaluated with 6 results within limits (up to 80 closest results for each mass)

## Elements Used:

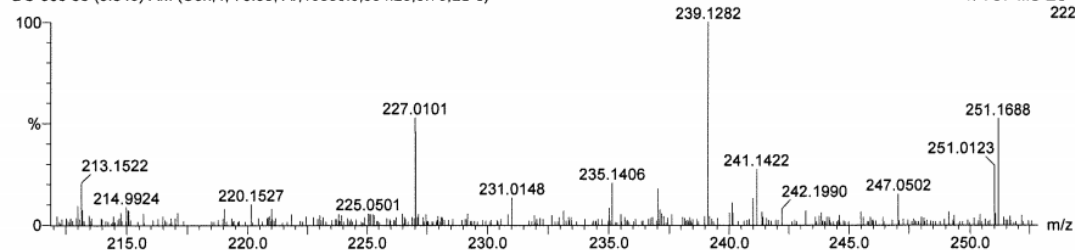
C: 0-70 H: 0-110 N: 0-9 O: 0-10 Na: 0-1

Tian

LCT Premier KD070

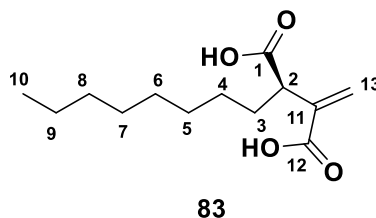
DO 065 38 (0.849) AM (Cen,4, 70.00, Ar,10000.0,554.26,0.70,LS 5)

1: TOF MS ES-



Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
239.1282	239.1283	-0.1	-0.4	4.5	28.1	C13 H19 O4
	239.1273	0.9	3.8	6.5	29.9	C12 H16 N4 Na
	239.1297	-1.5	-6.3	9.5	30.9	C14 H15 N4
	239.1259	2.3	9.6	1.5	27.8	C11 H20 O4 Na
	239.1256	2.6	10.9	5.5	28.8	C9 H15 N6 O2
	239.1243	3.9	16.3	0.5	27.6	C8 H19 N2 O6

Figure 7.58 HR-ESI-MS of compound 93.

Compound 83<sup>54, 141</sup>

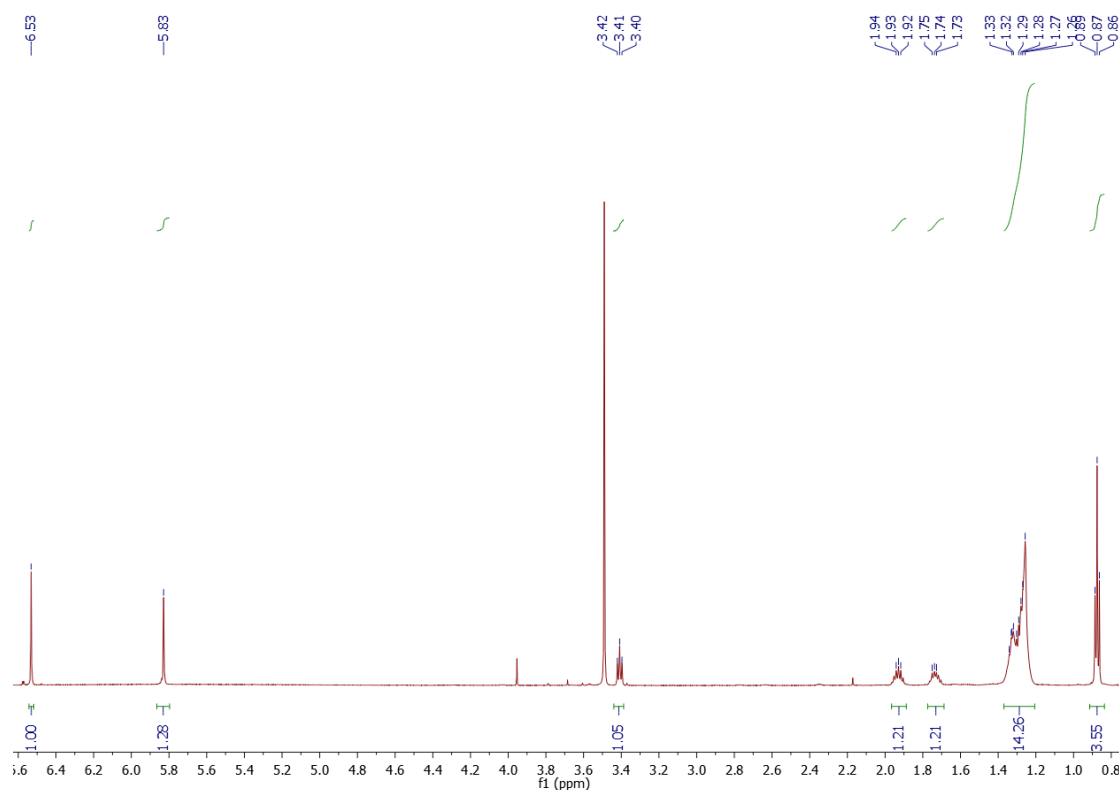
83

**2R-2-(2-acrylic)-decanoic acid****Chemical formula: C<sub>13</sub>H<sub>22</sub>O<sub>4</sub>**

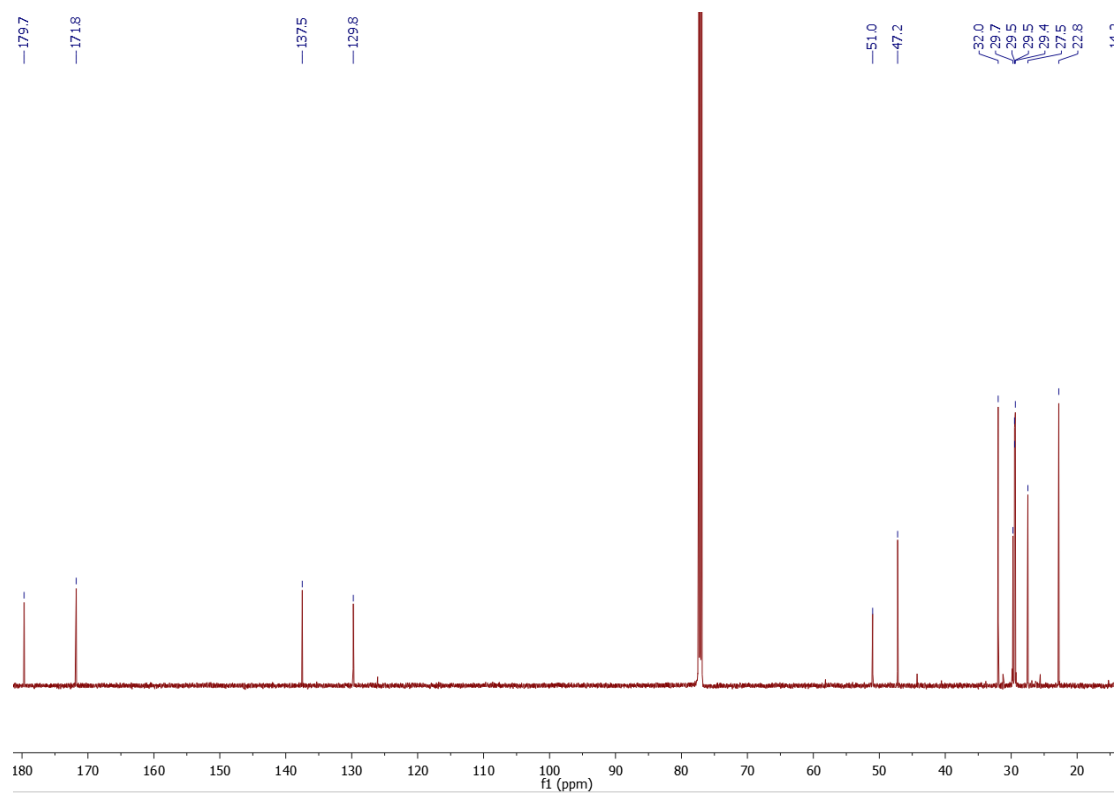
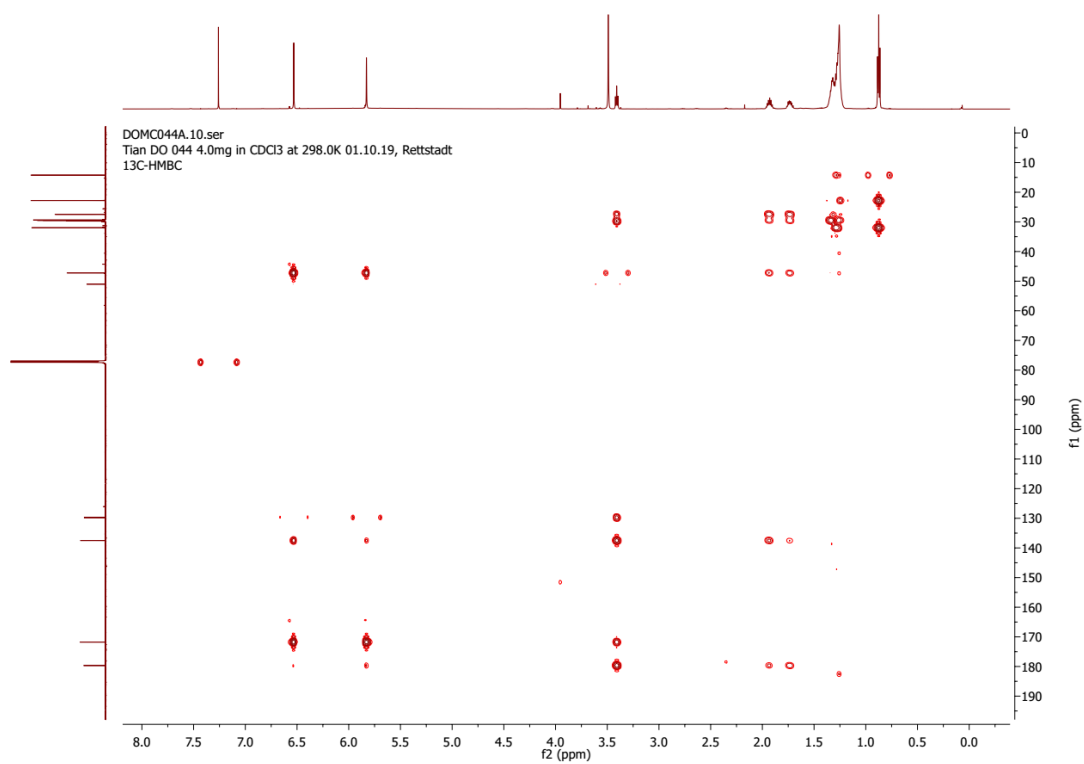
Yellow oil;  $[\alpha]_{\text{D}}^{25} - 9$  ( $c = 0.27$ , MeOH); Literature,<sup>54, 141</sup>  $[\alpha]_{\text{D}}^{25} - 22.1$  ( $c = 0.44$ , MeOH); UV ( $\lambda_{\text{max}}$ ): 211 nm. <sup>13</sup>C NMR data (CDCl<sub>3</sub>, 150 MHz):  $\delta_{\text{C}}$  179.7 (C-1), 171.8 (C-12), 137.5 (C-11), 129.8 (C-13), 47.2 (C-2), 32.0 (C-8), 29.7 (C-3), 29.5 (C-5), 29.5 (C-6), 29.4 (C-7), 27.5 (C-4), 22.8 (C-9), 14.2 (C-10); <sup>1</sup>H NMR data (CDCl<sub>3</sub>, 600 MHz):  $\delta_{\text{H}}$  6.53 (1H, s, H-13a), 5.83 (1H, s, H-13b), 3.41 (1H, t,  $J = 7.3$  Hz, H-2), 1.93 (1H, m, H-3b), 1.73 (1H, m, H-3a), 1.33 (2H, m, H-4), 1.28 (2H, m, H-9), 1.24-1.34 (6H, m, H-5/H-6/H-7), 1.25 (2H, m, H-8), 0.88 (3H, t,  $J = 6.9$  Hz, H-10); NMR data (Table 7.13) are consistent with those previously reported;<sup>54, 141</sup> ESI-MS  $m/z$  241 [M - H]<sup>-</sup>, 243 [M + H]<sup>+</sup>. HR-ESI-MS  $m/z$  241.1441 [M - H]<sup>-</sup> (calcd. for C<sub>13</sub>H<sub>21</sub>O<sub>4</sub>, 241.1440).

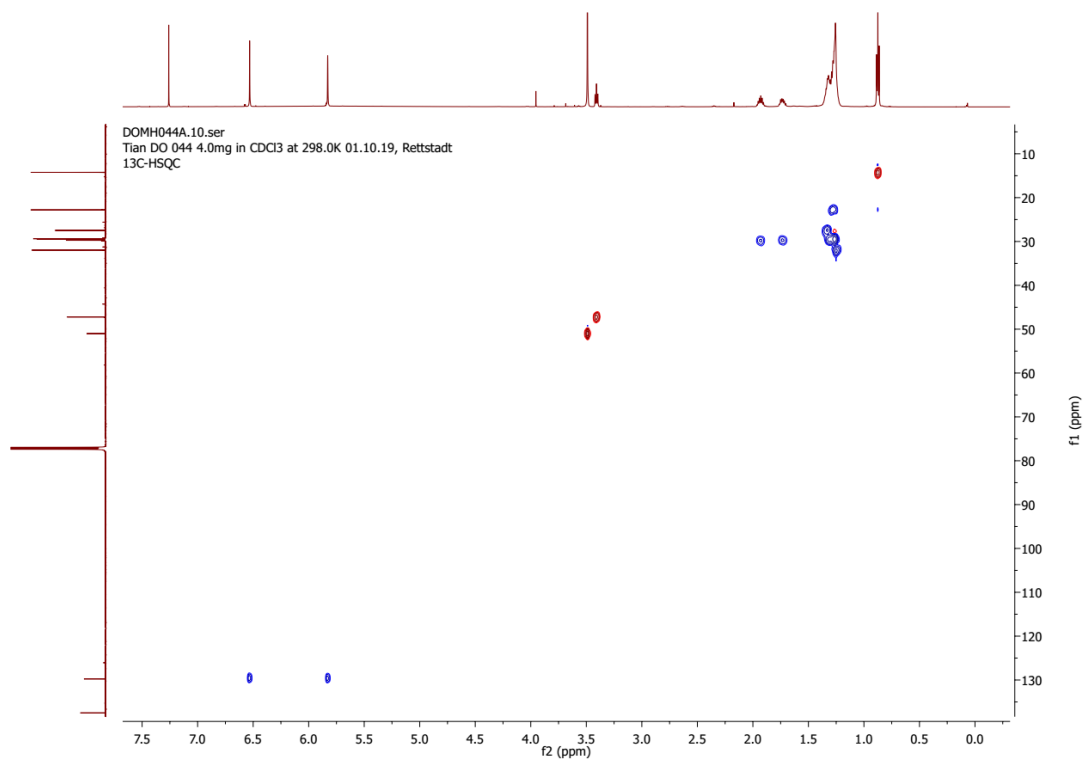
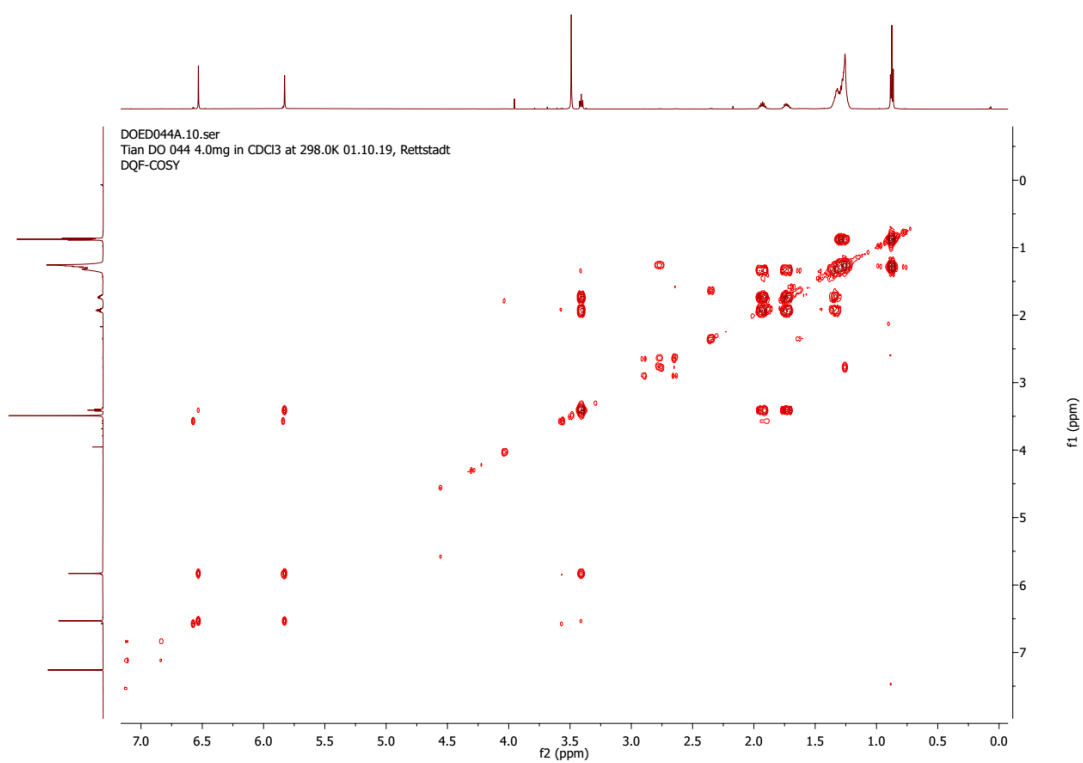
pos.	$\delta_c$ / ppm	$\delta_H$ / ppm (J / Hz)	$\delta_c$ / ppm literature <sup>54, 141</sup>	$\delta_H$ / ppm (J / Hz) literature <sup>54, 141</sup>
1	179.7	-	177.9	-
2	47.2	3.41, 1H, t (7.3)	48.0	3.39, 1H, t (7.4)
3	29.7	1.73, 1H, m 1.93, 1H, m	30.3	1.75, 1H, m 1.93, 1H, m
4	27.5	1.33, 2H, m	27.5	1.24-1.40, 2H, m
5	29.5	1.24-1.34, 2H, m	29.4	1.24-1.40, 2H, m
6	29.5	1.24-1.34, 2H, m	29.3	1.24-1.40, 2H, m
7	29.4	1.24-1.34, 2H, m	29.2	1.24-1.40, 2H, m
8	32.0	1.25, 2H, m	31.8	1.24-1.40, 2H, m
9	22.8	1.28, 2H, m	22.6	1.24-1.40, 2H, m
10	14.2	0.88, 3H, t (6.9)	14.1	0.86, 3H, t (6.7)
11	137.5	-	138.3	-
12	171.8	-	171.4	-
13	129.8	6.53, 1H, s 5.83, 1H, s	128.4	6.54, 1H, s 5.82, 1H, s

**Table 7.13**  $^1\text{H}$  NMR (600 MHz) data and  $^{13}\text{C}$  NMR (150 MHz) data for **83** in  $\text{CDCl}_3$ . Literature<sup>54, 141</sup> data was measured at 500 MHz in  $\text{CDCl}_3$ .



**Figure 7.59**  $^1\text{H}$  NMR of compound **83**.

Figure 7.60  $^{13}\text{C}$  NMR of compound **83**.Figure 7.61 HMBC of compound **83**.

Figure 7.62 HSQC of compound **83**.Figure 7.63  $^1\text{H}$ - $^1\text{H}$  COSY of compound **83**.

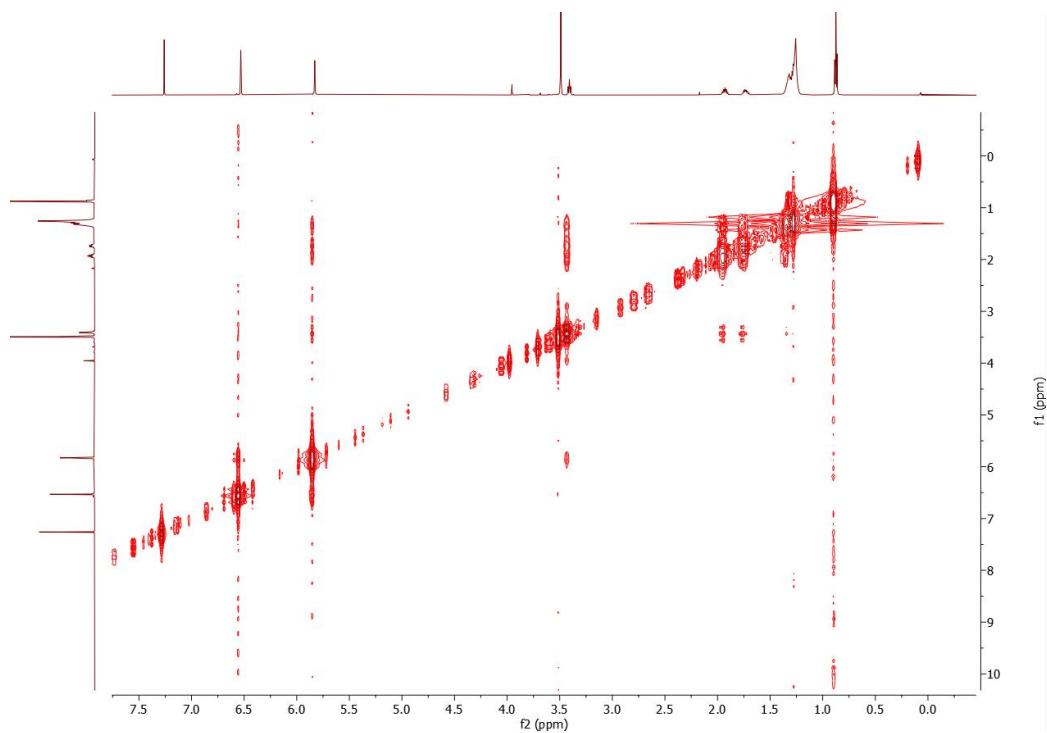


Figure 7.64 NOESY of compound 83.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 10.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

Monoisotopic Mass, Even Electron Ions

64 formula(e) evaluated with 1 results within limits (up to 50 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 O: 0-15 Na: 0-1

Tian

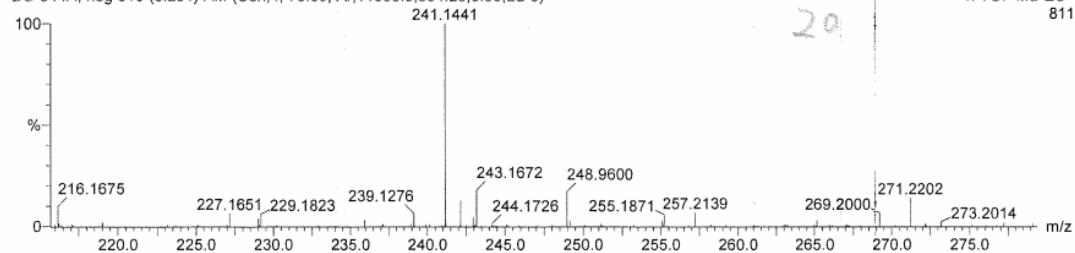
Q-ToF Premier UPLC-MS

DO 044A, neg 810 (8.291) AM (Cen,4, 70.00, Ar,11000.0,554.26,0.55,LS 5)

27-Sep-2019 10:21:27

1: TOF MS ES-

811

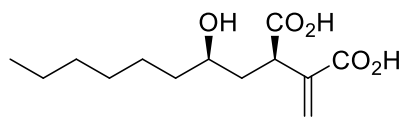


Minimum: -0.5  
 Maximum: 5.0 10.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
241.1441	241.1440	0.1	0.4	3.5	61.3	C13 H21 O4

Figure 7.65 HR-ESI-MS of compound 83.

## Compound 87



87

Chemical formula: C<sub>13</sub>H<sub>22</sub>O<sub>5</sub>

UV ( $\lambda_{\text{max}}$ ): 211 nm. ESI-MS  $m/z$  257 [M - H]<sup>-</sup>, 515 [2M - H]<sup>-</sup>, 241 [M - H<sub>2</sub>O + H]<sup>+</sup>, 259 [M + H]<sup>+</sup>, 281 [M + Na]<sup>+</sup>, 539 [2M + Na]<sup>+</sup>. HR-ESI-MS  $m/z$  257.1387 [M - H]<sup>-</sup> (calcd. for C<sub>13</sub>H<sub>21</sub>O<sub>5</sub>, 257.1389).

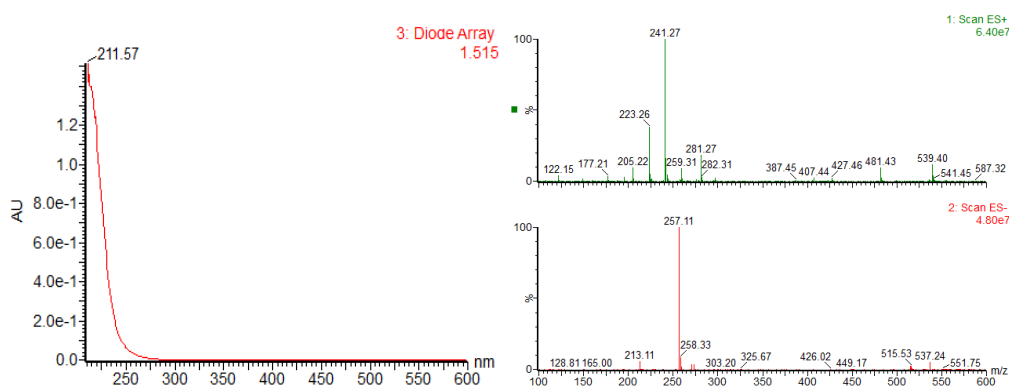


Figure 7.66 UV and mass spectra of compound 87.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

38 formula(e) evaluated with 1 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 O: 0-14

Tian

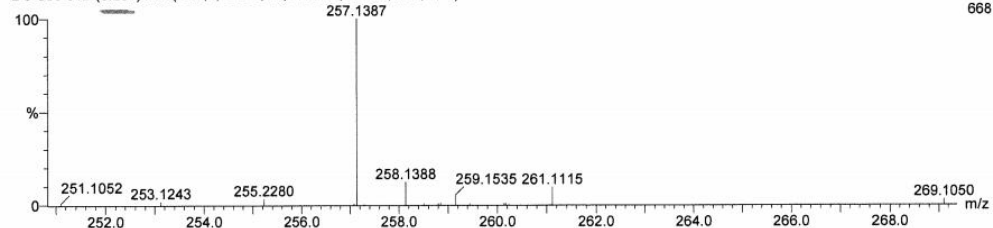
Q-ToF Premier UPLC-MS

06-May-2020 15:21:17

DO 068 612 (6.261) AM (Cen.4, 90.00, Ar, 11000.0, 554.26, 0.55, LS 5)

1: TOF MS ES-

668



Minimum:

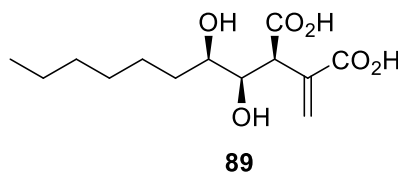
Maximum: 5.0 20.0 -0.5 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
257.1387	257.1389	-0.2	-0.8	3.5	8.3	C13 H21 O5

Figure 7.67 HR-ESI-MS of compound 87.



## Compound 89



**Chemical formula: C<sub>13</sub>H<sub>22</sub>O<sub>6</sub>**

UV ( $\lambda_{\text{max}}$ ): 200 nm. ESI-MS  $m/z$  273 [M – H]<sup>–</sup>, 547 [2M – H]<sup>–</sup>, 257 [M – H<sub>2</sub>O + H]<sup>+</sup>, 275 [M + H]<sup>+</sup>, 292 [M + H<sub>2</sub>O]<sup>+</sup>, 549 [2M + H]<sup>+</sup>. HR-ESI-MS  $m/z$  273.1339 [M – H]<sup>–</sup> (calcd. for C<sub>13</sub>H<sub>21</sub>O<sub>6</sub>, 273.1338).

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

75 formula(e) evaluated with 2 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 O: 0-14 Na: 0-1

Tian

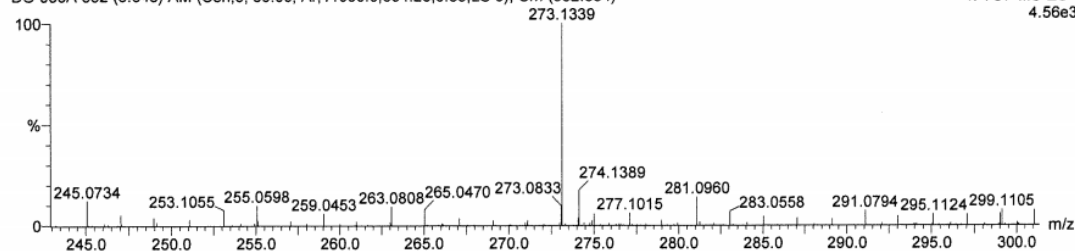
Q-ToF Premier UPLC-MS

08-May-202012:12:34

DO 068A 552 (5.643) AM (Cen,5, 60.00, Ar,11000.0,554.26,0.55,LS 5); Cm (552:554)

1: TOF MS ES-

4.56e3

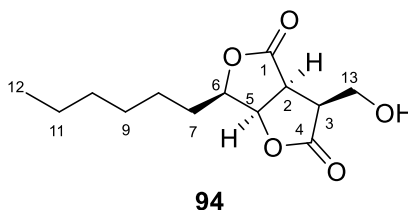


Minimum: -0.5  
Maximum: 5.0 20.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
273.1339	273.1338	0.1	0.4	3.5	2774773.5	C13 H21 O6
	273.1314	2.5	9.2	0.5	2774628.5	C11 H22 O6 Na

Figure 7.68 HR-ESI-MS of compound 89.

## Compound 94



**Chemical formula: C<sub>13</sub>H<sub>20</sub>O<sub>5</sub>**

UV ( $\lambda_{\max}$ ): 214 nm. ESI-MS  $m/z$  255  $[M - H]^-$ , 511  $[2M - H]^-$ , 239  $[M - H_2O + H]^+$ , 257  $[M + H]^+$ , 274  $[M + H_2O]^+$ , 535  $[2M + Na]^+$ . HR-ESI-MS  $m/z$  255.1217  $[M - H]^-$  (calcd. for  $C_{13}H_{19}O_5$ , 255.1232).

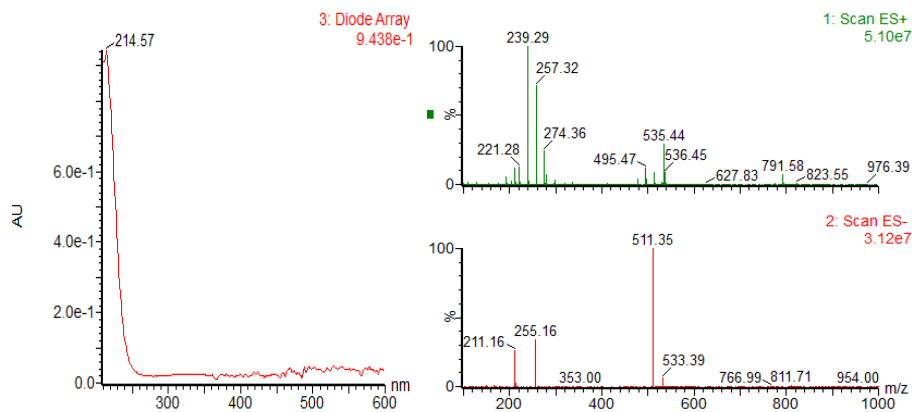


Figure 7.69 UV and mass spectra of compound 94.

24  
Elemental Composition Report

Page 1

Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0  
Selected filters: None

Monoisotopic Mass, Odd and Even Electron Ions

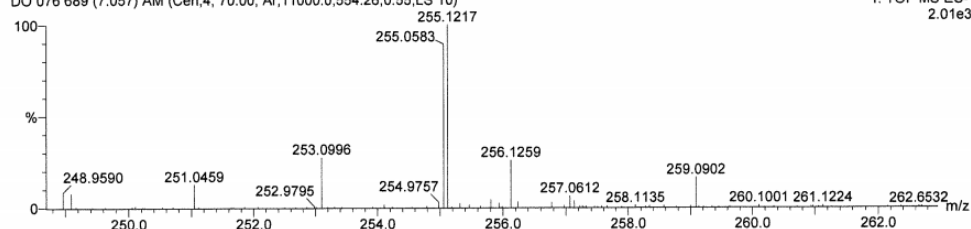
33 formula(e) evaluated with 2 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-80 H: 0-100 O: 0-10

Tian Q-ToF Premier UPLC-MS  
DO 076 689 (7.057) AM (Cen,4, 70.00, Ar,11000.0,554.26,0.55,LS 10)

09-Jul-202009:41:00  
1: TOF MS ES-  
2.01e3

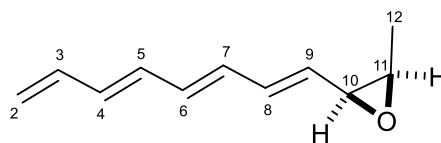


Minimum: -0.5  
Maximum: 5.0 20.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
255.1217	255.1232	-1.5	-5.9	4.5	47.9	C13 H19 O5
	255.1174	4.3	16.9	13.5	9.4	C20 H15

Figure 7.70 HR-ESI-MS of compound 94.

## Compound 142

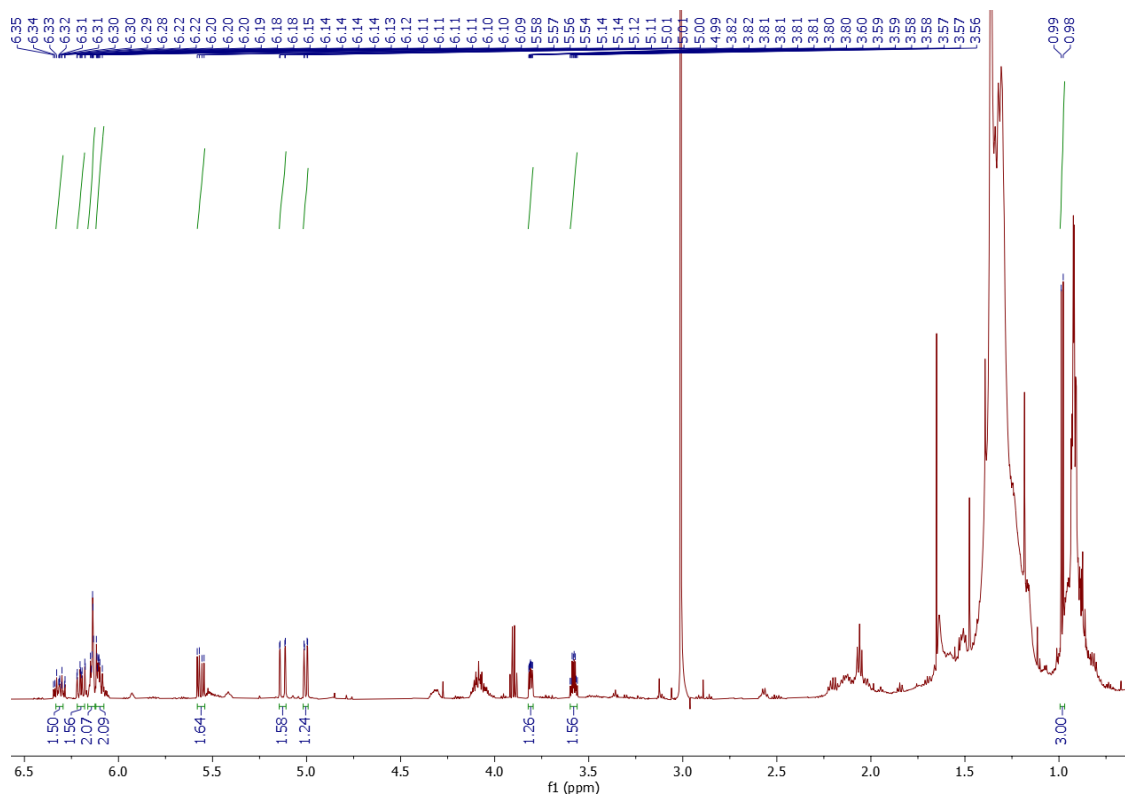


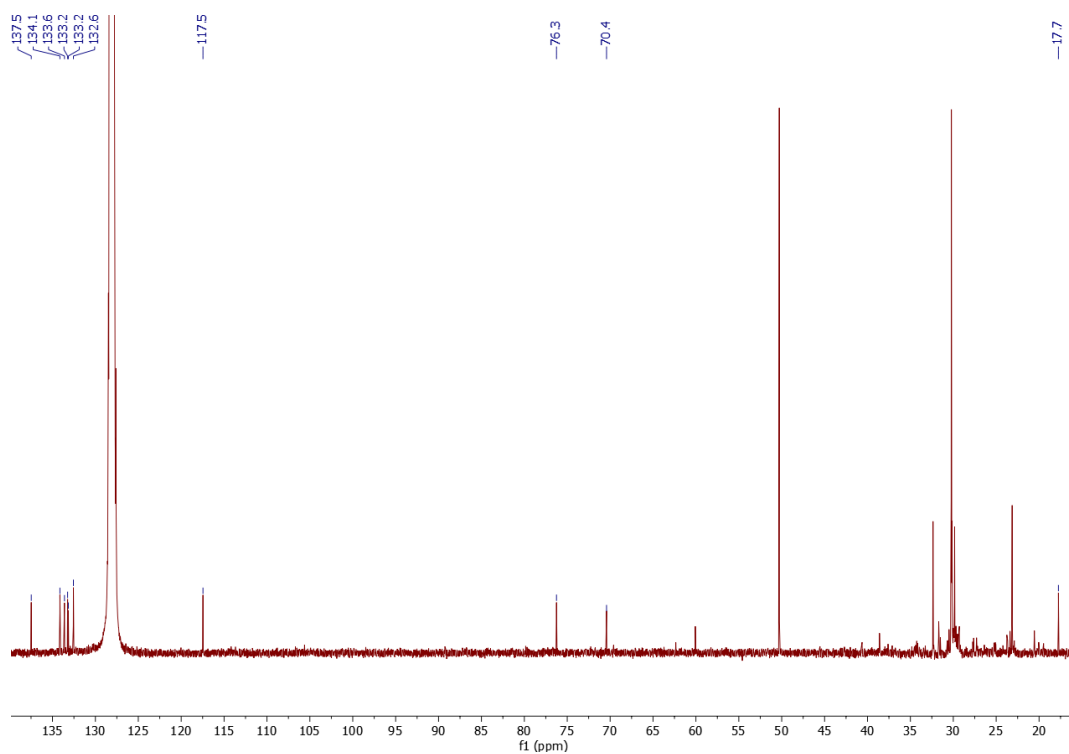
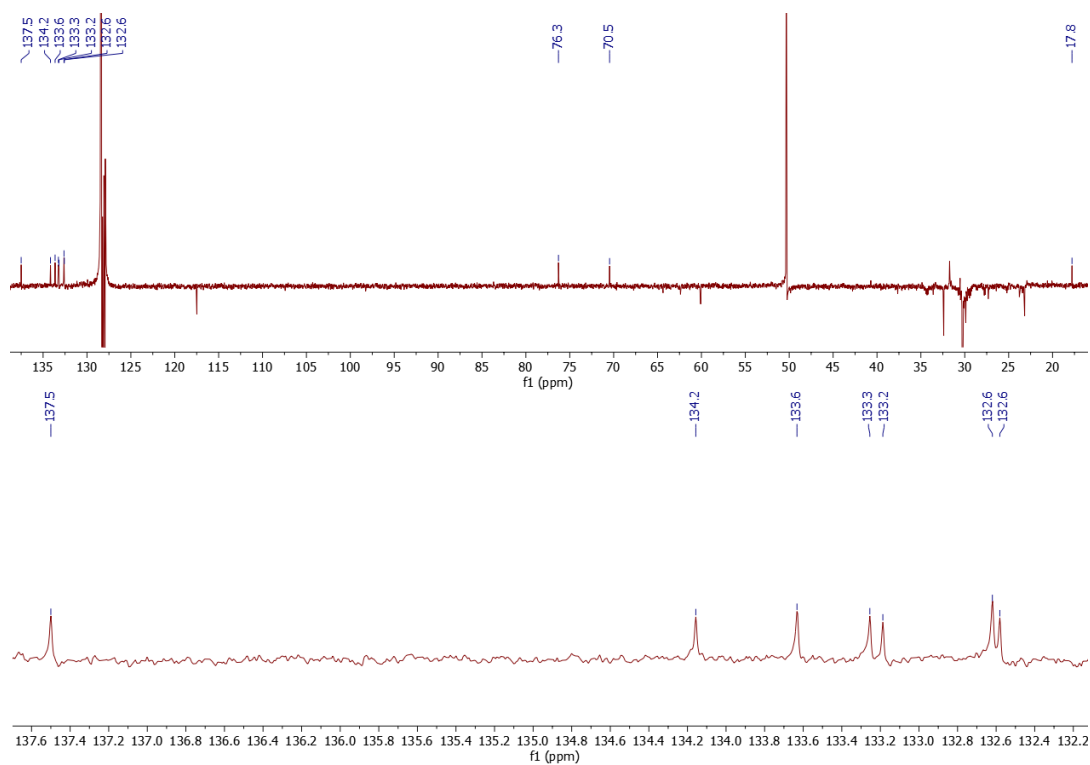
142

## Tetraenylmethoxriane A

Chemical formula:  $C_{11}H_{14}O$ 

Yellow oil; UV ( $\lambda_{\max}$ ): 284 nm, 296 nm, 309 nm.  $^{13}C$  NMR data ( $C_6D_6$ , 150 MHz):  $\delta_c$  137.5 (C-3), 134.1 (C-4), 133.6 (C-6), 133.2 (C-5), 133.2 (C-7), 132.6 (C-8), 132.6 (C-9), 117.5 (C-2), 76.3 (C-10), 70.4 (C-11), 17.7 (C-12);  $^1H$  NMR data ( $C_6D_6$ , 600 MHz):  $\delta_H$  6.31 (1H, m, H-3), 6.20 (1H, ddd,  $J = 15.4, 10.1, 1.3$  Hz, H-8), 6.14 (1H, m, H-4), 6.14 (1H, m, H-6), 6.11 (1H, m, H-5), 6.11 (1H, m, H-7), 5.56 (1H, dd,  $J = 15.4, 6.7$  Hz, H-9), 5.13 (1H, dd,  $J = 16.9, 1.5$  Hz, H-2a), 5.00 (1H, dd,  $J = 10.1, 1.5$  Hz, H-2b), 3.81 (1H, ddd,  $J = 6.7, 3.9, 1.2$  Hz, H-10), 3.58 (1H, qd,  $J = 6.4, 3.9$  Hz, H-11), 0.99 (3H, d,  $J = 6.4$  Hz, H-12); NMR data see Table 4.9; ESI-MS  $m/z$  161  $[M - H]^-$ , 163  $[M + H]^+$ ; HR-ESI-MS  $m/z$  163.1123  $[M + H]^+$  (calcd. for  $C_{11}H_{15}O$ , 163.1123).

Figure 7.71  $^1H$  NMR of compound 142.

Figure 7.72 <sup>13</sup>C NMR of compound 142.Figure 7.73 <sup>13</sup>C NMR (DEPT) of compound 142.

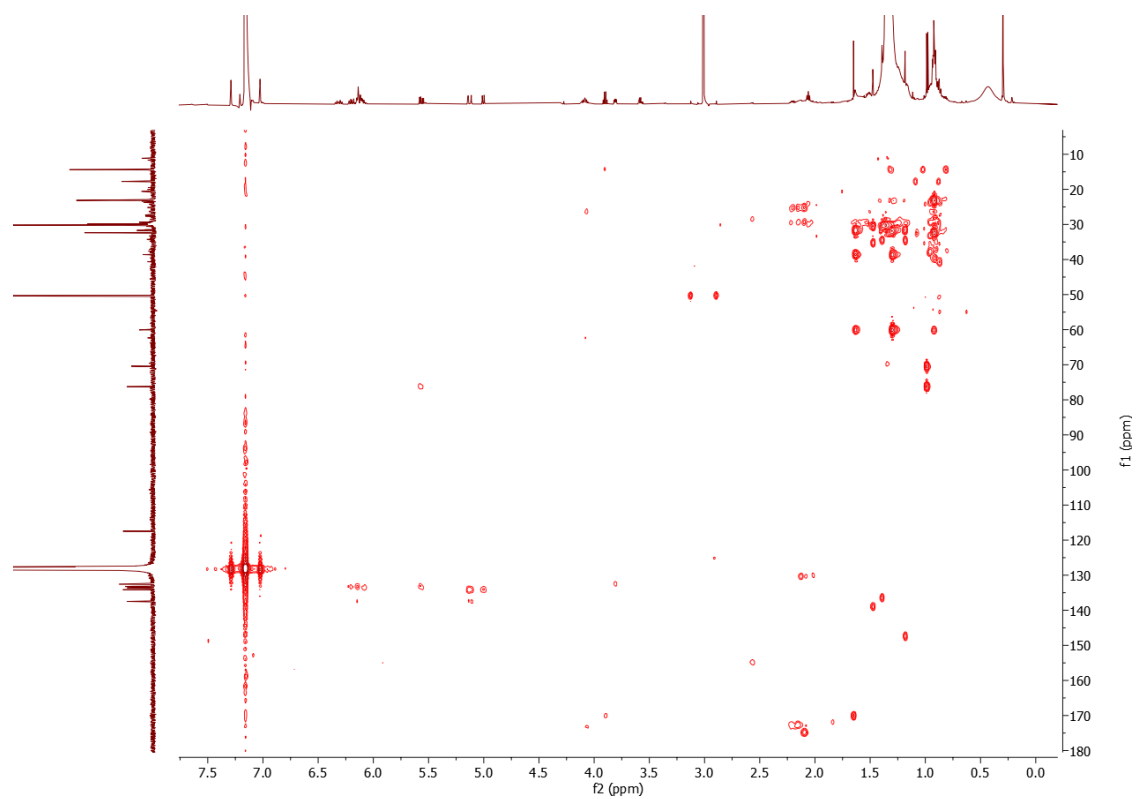


Figure 7.74 HMBC of compound 142.

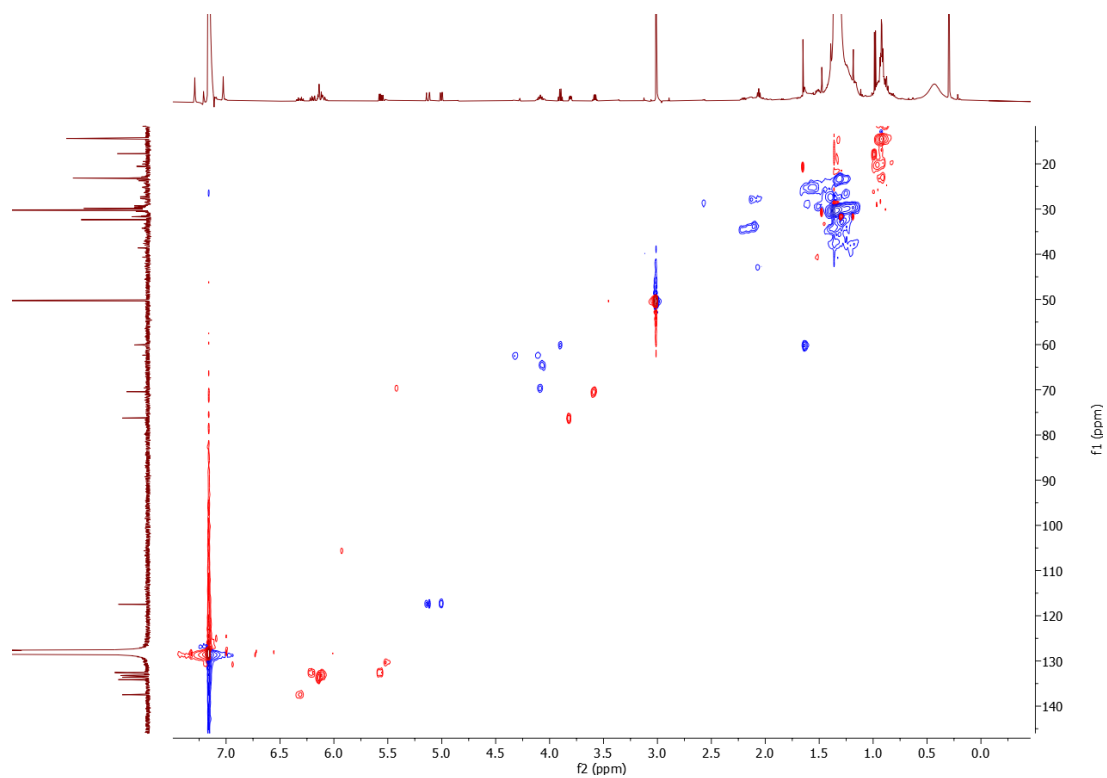


Figure 7.75 HSQC of compound 142.

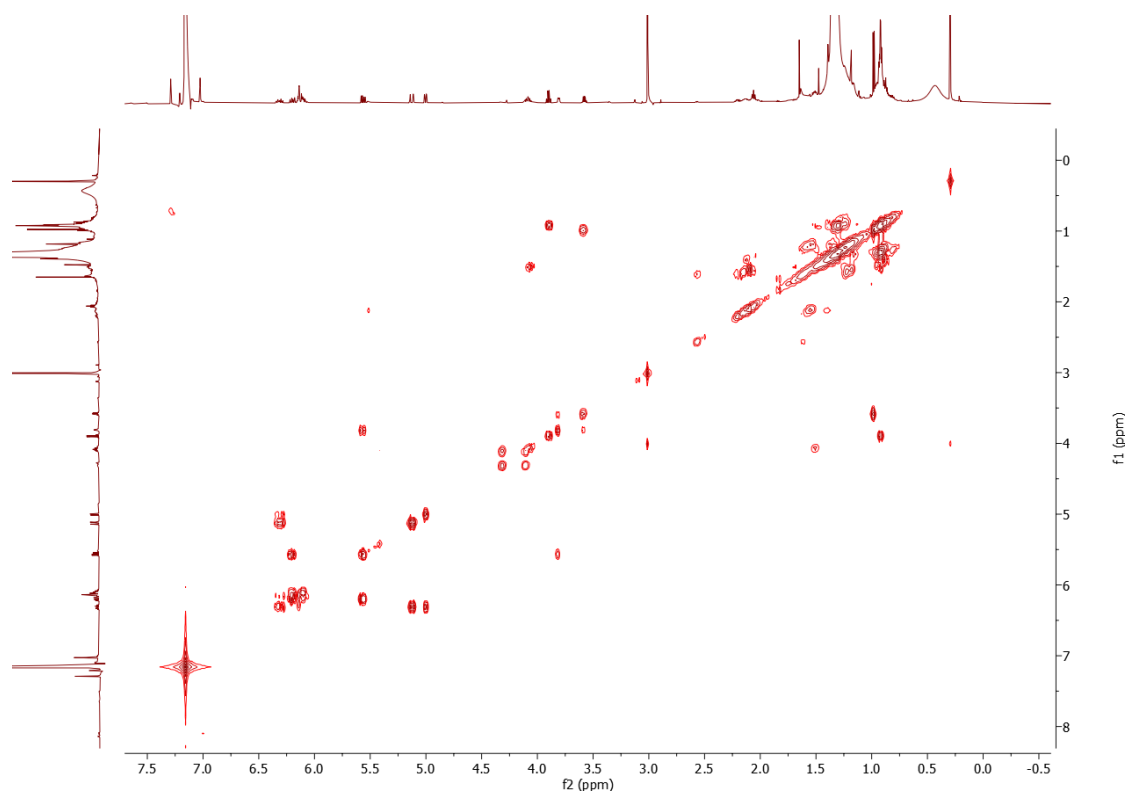


Figure 7.76  $^1\text{H}$ - $^1\text{H}$  COSY of compound 142.

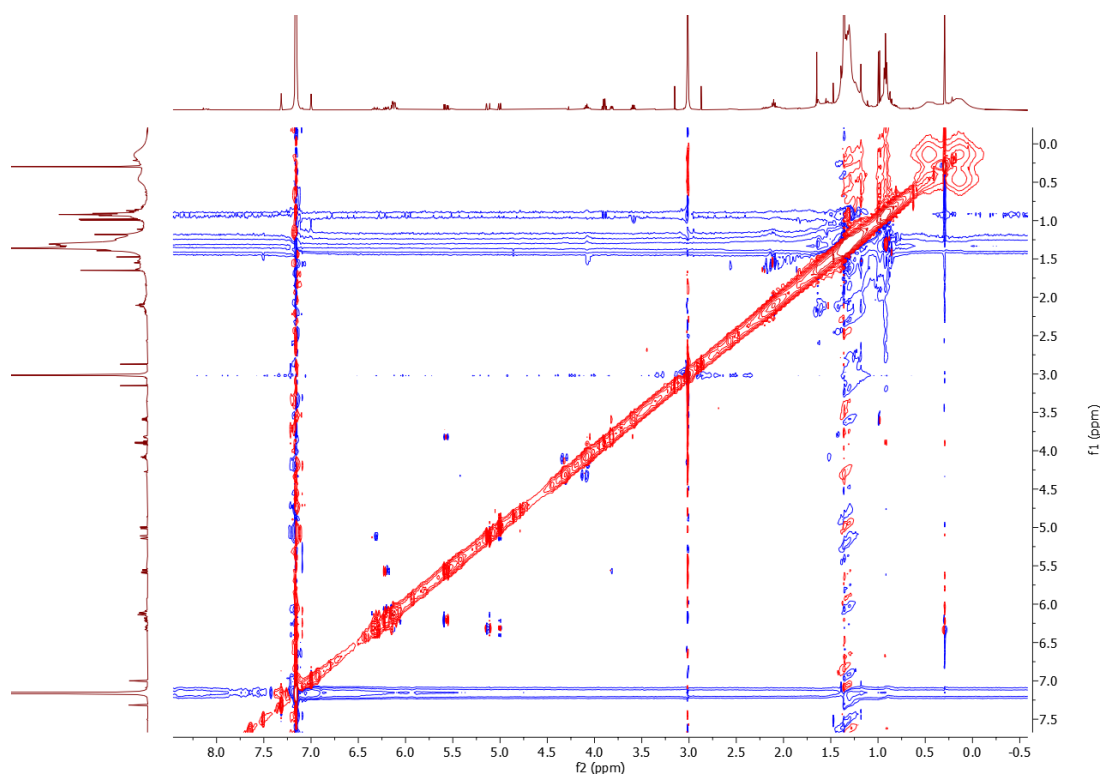


Figure 7.77 NOESY of compound 142.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

13 formula(e) evaluated with 1 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 0-70 H: 0-100 O: 0-4

Tian

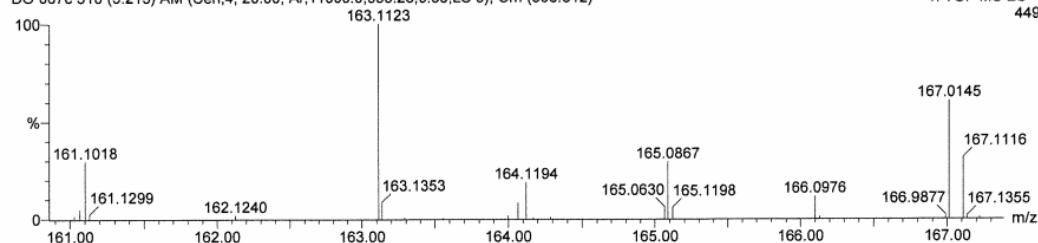
Q-ToF Premier UPLC-MS

11-Mar-2020 13:03:40

DO 067c 510 (5.215) AM (Cen,4, 20.00, Ar,11000.0,556.28,0.55,LS 5); Cm (506:512)

1: TOF MS ES+

449



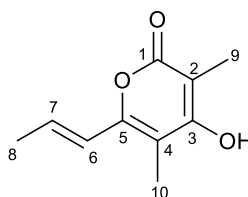
Minimum:

Maximum: 5.0 20.0 -0.5 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
163.1123	163.1123	0.0	0.0	4.5	13.2	C11 H15 O

Figure 7.78 HR-ESI-MS of compound 142.

## Compound 143

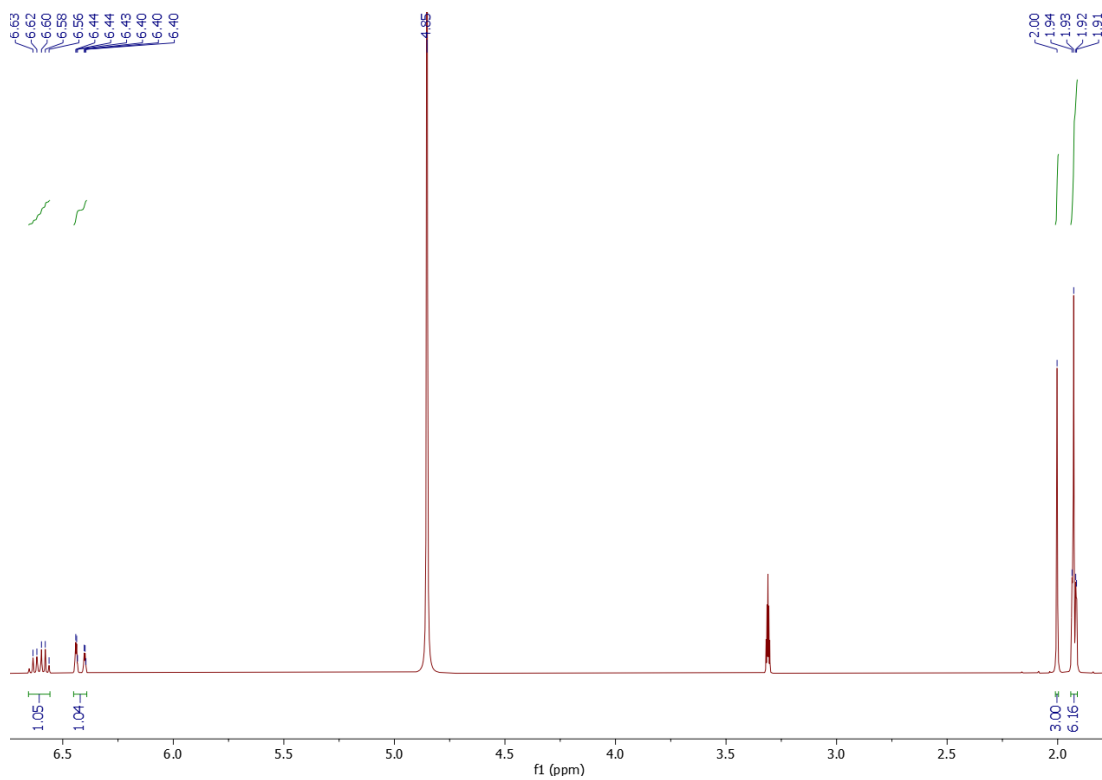
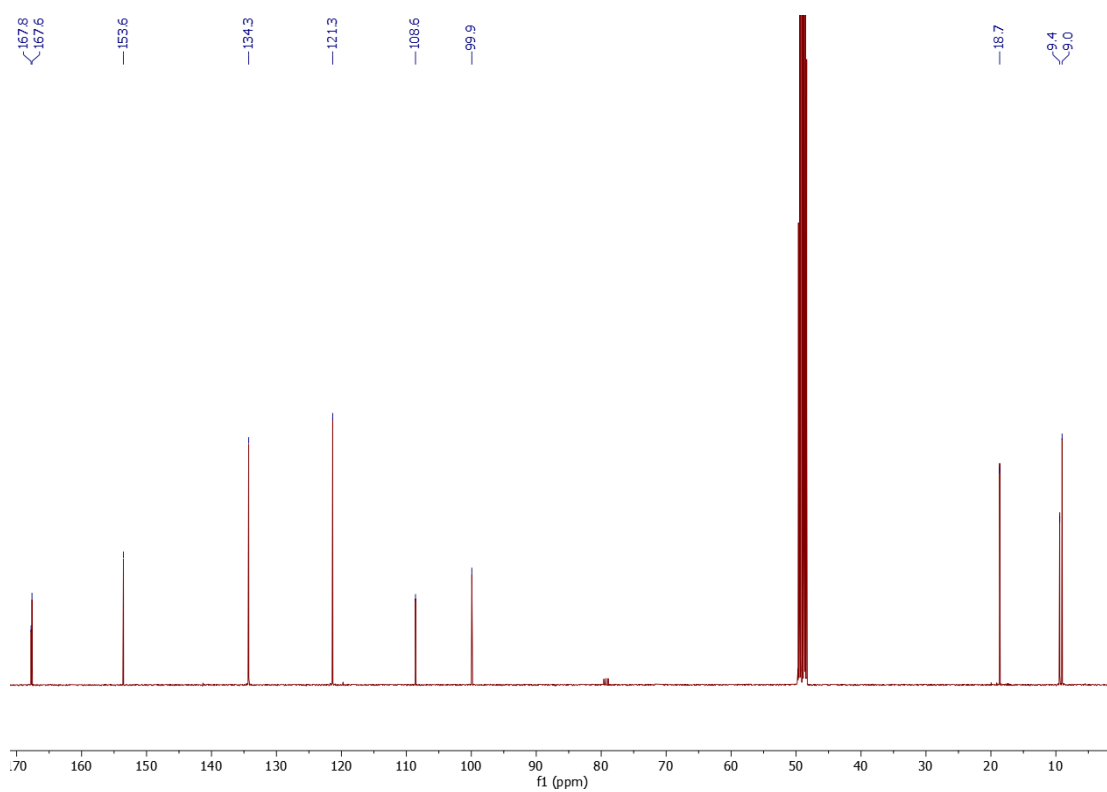


143

## Islandic acid A

Chemical formula: C<sub>10</sub>H<sub>12</sub>O<sub>3</sub>

White powder; UV ( $\lambda_{\max}$ ): 225 nm, 326 nm. <sup>13</sup>C NMR data (CD<sub>3</sub>OD, 100 MHz):  $\delta_c$  167.8 (C-3), 167.6 (C-1), 153.6 (C-5), 134.3 (C-7), 121.3 (C-6), 108.6 (C-4), 99.9 (C-2), 18.7 (C-8), 9.4 (C-10), 9.0 (C-9); <sup>1</sup>H NMR data (CD<sub>3</sub>OD, 400 MHz):  $\delta_H$  6.61 (1H, dq,  $J$  = 15.4, 6.9 Hz, H-7), 6.42 (1H, dq,  $J$  = 15.4, 1.7 Hz, H-6), 2.00 (3H, s, H-10), 1.93 (3H, m, H-9), 1.92 (3H, m, H-8; NMR data see Table 4.11; ESI-MS  $m/z$  179 [M - H]<sup>-</sup>, 181 [M + H]<sup>+</sup>; HR-ESI-MS  $m/z$  181.0875 [M + H]<sup>+</sup> (calcd. for C<sub>10</sub>H<sub>13</sub>O<sub>2</sub>, 181.0865).

Figure 7.79  $^1\text{H}$  NMR of compound 143.Figure 7.80  $^{13}\text{C}$  NMR of compound 143.



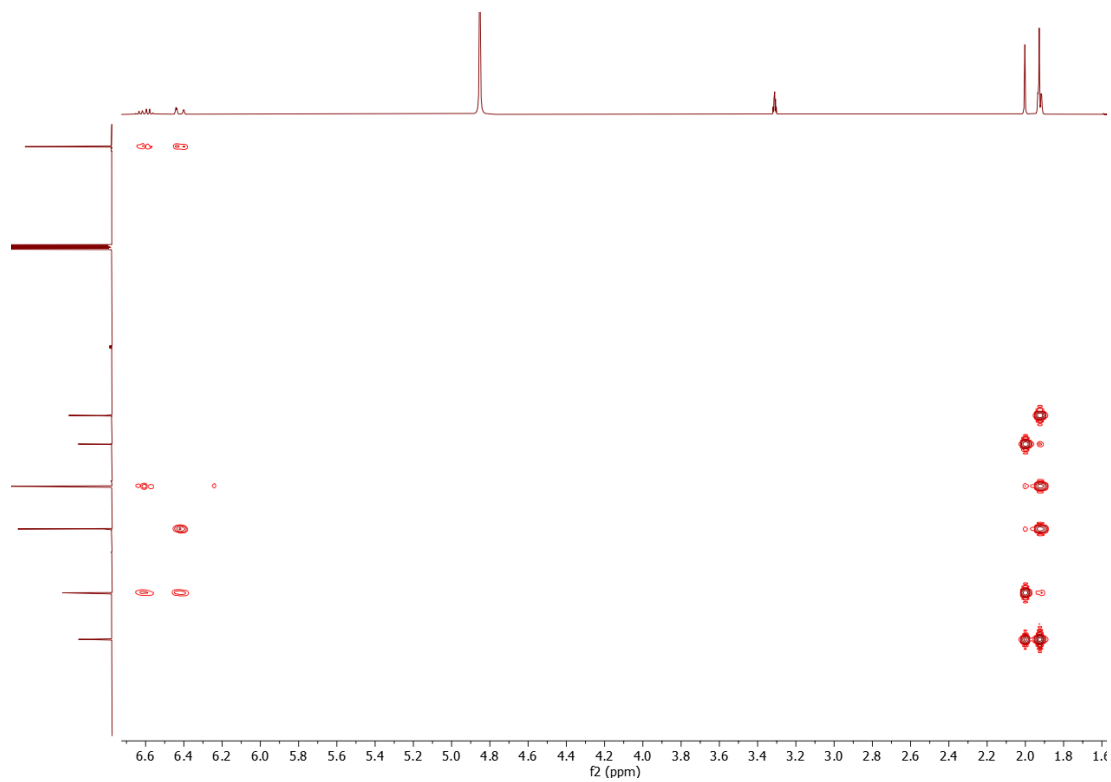


Figure 7.81 HMBC of compound 143.

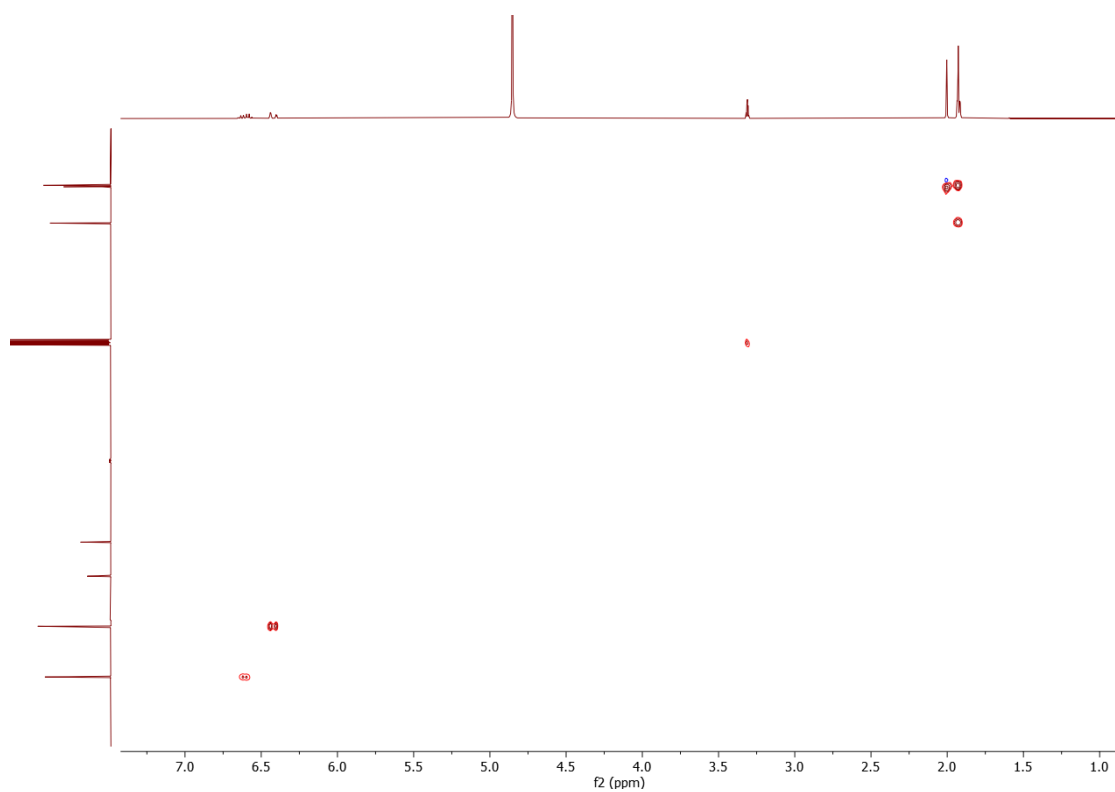


Figure 7.82 HSQC of compound 143.

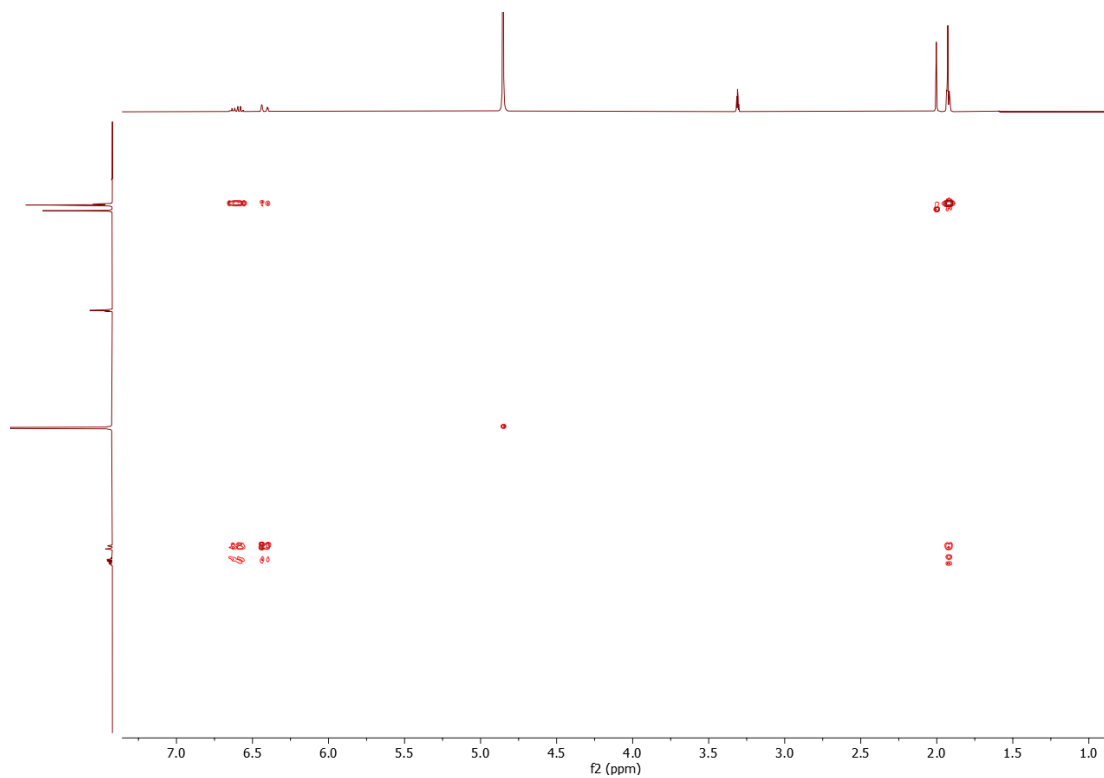


Figure 7.83 <sup>1</sup>H-<sup>1</sup>H COSY of compound 143.

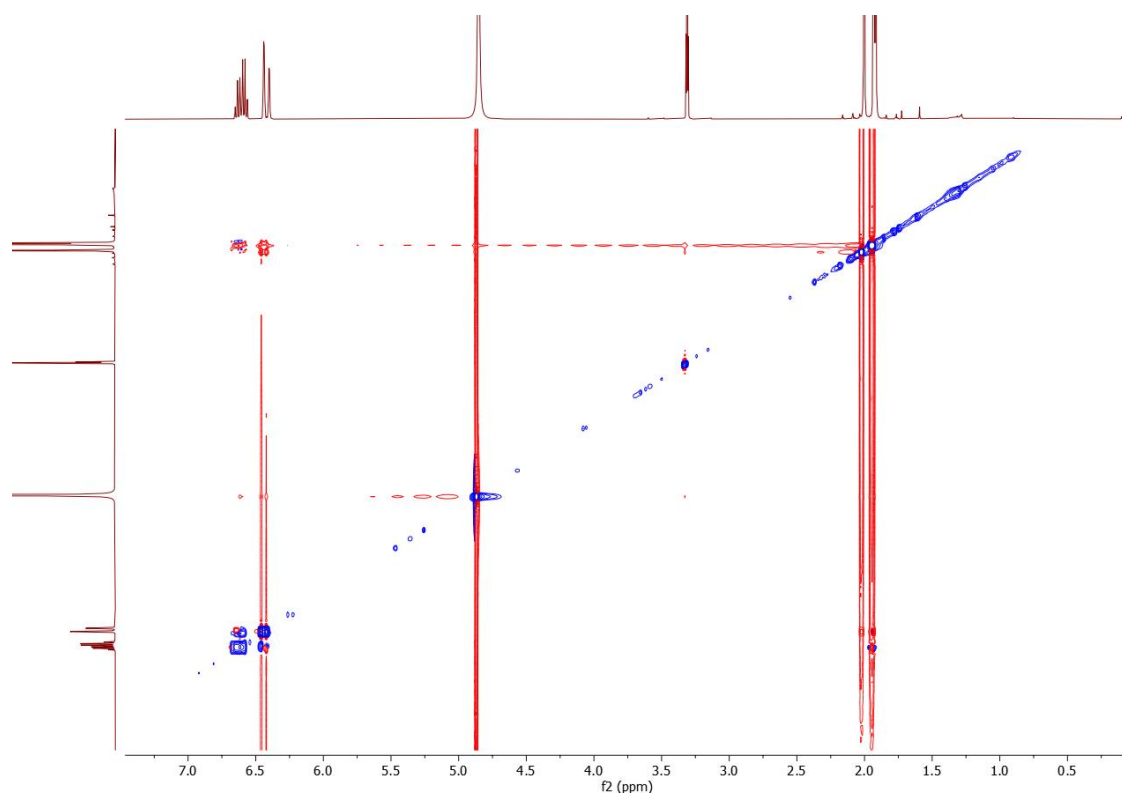


Figure 7.84 NOESY of compound 143.

## Elemental Composition Report

Page 1

## Single Mass Analysis (displaying only valid results)

Tolerance = 20.0 PPM / DBE: min = -0.5, max = 60.0

Selected filters: None

## Monoisotopic Mass, Even Electron Ions

262 formula(e) evaluated with 3 results within limits (up to 25 closest results for each mass)

Elements Used:

C: 1-50 H: 1-100 N: 0-4 O: 0-10 Na: 0-1 S: 0-1

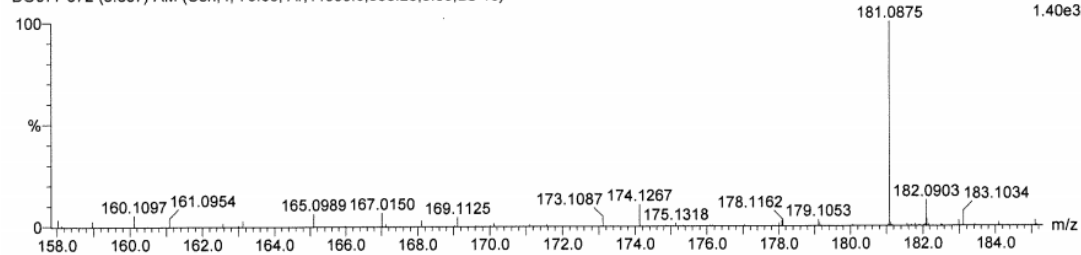
Tian Q-ToF Premier UPLC-MS

DO077 672 (6.867) AM (Cen,4, 70.00, Ar,11000.0,556.28,0.55,LS 15)

13-Aug-2020 10:04:47

1: TOF MS ES+

1.40e3



Minimum: -0.5  
 Maximum: 5.0 20.0 60.0

Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
181.0875	181.0865	1.0	5.5	4.5	34.8	C10 H13 O3
	181.0898	-2.3	-12.7	-0.5	11.2	C7 H17 O3 S
	181.0841	3.4	18.8	1.5	44.1	C8 H14 O3 Na

Figure 7.85 HR-ESI-MS of compound 143.



TENS	---RLYTEDEKGMHMANLFLANALEEITFKFPRCKILEIGAGTGATTWAA-LSAIG-----	51
SQTKS	-LYRYTDAIKWDRSYQQIDQLVKLHAKCPSAKIIEIGAGTGCTTRAV-LDALST-----	54
Citrinin	-----INAVWIIQQAEEFFLEQLVKRLPNTGEPRLRILEMGAGTGTTVKM-LPLLE-----R	49
PKS1	RTQSKPDVALLSQAQKLSVADFVLEFVKPKGLKVLSDGAYAESLVGK-IPALS-----	54
PKS2	--TPLYQTGLLMTGVYPOLYNVLDLSHANPNLRILEVGGTGGATRIA-MKAFR-----G	53
PKS3	--QYFAELPRLRDRTYKQLSKVAEFYAVTSPGANVLEIGAGTGGVVSQVILQAFGA-----R-G	56
PKS6	-----LAALYEFMSQSDTSLIQLSLSHSRPNLQVLEIGTGRSSVTSV-LKHLT-----R	49
PKS7	-----LAHQTPGMKILEVGGTGGTRNI-VSALCT-----RRGD	34
PKS8	-----YDELTSYRCMAQCESYLDLLAHQTPGLNILEVGGTASATRKF-IRALR-----AND	51
PKS11	-----TREETLEEMQDNKMAVCEKLLAL-KEGETMLDICCMTLAKFA-SVNYGAHVTVTLARNQAANGNDALRRAGV	73
PKS12	-HNRPHYHESSIINDTNHDLFLIAGQLSHRYPHMRVLEVGAGTGGSTGGV-LAALD-----	53
PKS14	-----LTEI-YNGTSFGYGFVRLLSNTRPNLRILEVGAGTGGTELI-LRDLV-----D	48
PKS15	SLNNMYRYGLSDERTPAIQCEYIRRLSHKR-PLRILEVAGTGSATSRI-LSSLG-----	53
PKS16	-ADIFYADLFQSLCADGRLLSFLDLTAHENPTLRILEIGAGTGGVTGHV-LSILQE-----R-EK	57
PKS17	-AQAFYNQNFSRHAEVGHVRVFLDLLSHEHPSLRILEVGAGTGSMSRLI-LETLOS-----F-EQ	57
PKS18	-----KDFLQLLGHRLPHMNVLEIGAGTGGTLAKI-LDFLQ-----S	36
TENS	----EAFDTYTYTDLVSGFFENAVERFSAFR--HRMVFRALDIEKDPASQSFDL--NSYDIIATNVLEATRNLGVTLG	122
SQTKS	-HGAARCAQYDFTDVSSGFFEAQQKFTAF--DVIRFQKLDIEKDIETQGFEC---GSYDLVIASQVLEATGKIEDTMA	128
Citrinin	---LGVPEVYTMTDLSSSLIAAARKRFKKYP---FMKFKVVNIESPPDPQLVH---SQHII LATNCVLEATRNL EISTR	118
PKS1	----LTATYDVPDEEY---EMSAELSGSK---NVKAVKFDIGADLEGOAVAA---KSFELVVSRIASP-----ATLA	113
PKS2	PNGIKRYREYTFDTDISAGFLGGARESLAEFK---DMNFSVFDAEVDPIEQGYE---PVYDLVIACQVLEATSSMKNTLT	126
PKS3	NGSGTLLGSYTYTDLQDDALHGAARLAPWG--DMVQFQKLDIGDLAKQSPFKG---GEYDLIVVPLALYSTTSVKNALH	131
PKS6	SDGQVLCRYTFTSTG---FISGKQKQKFPF---NMEYATLNI SKDLEEQEFEG---RQYDLIIASNAIVTRNKIQESLK	120
PKS7	PEGFLRCSRYDFTDVSAALDRAREEFHEYH---SQMTFGVLDIDRDFAEQGFGA---GSYDVLVAVSVLEISSDLVGVVK	109
PKS8	DTGSLRCSRYHFTDISASFLKAREEFSSFH---SQMTFSLDIERDLTEQGLEE---GAYDVIIVADGVLEVTHNIAHTLR	126
PKS11	PESQSNLLCMDYRDI PAKKFNKI-----SOLEM-----GEHVIRRLTTFFR	115
PKS12	----SRMSSYTFDTDISPSYFQKARKDFDKYS--DRMIFKTLDIEEDPSTQGYEL---GSYDLIIASNVLEATRRLSNTLR	124
PKS14	DAGFPYISYIYTFDVSGLVVKARERFSYAP---NMEYKVLDISKSPLEQGFEGGERGSYDLIIAANVLEATPFLKETLG	125
PKS15	PNVAGRLQKYTYTVDVSGVFFEAASEEFKDWK---SMMEFKVLDIEKDFLQGLEE---GSYDLVVAQVLEATSSIGTTLA	128
PKS16	QTGAPSAEYTYTDISPMFQEQATSRWSKLRDQGRMVFKTLDDQRTIESQGFVA---GSYDMI IAGSVLEATPYLEVTIR	134
PKS17	ETGQTRFSEYIYTDISAAYFEDARALFDEYK---DRMLFKTLDLLEVDPVQGGFQL---ESYDLIIAGGVLEVPSDLTSTLR	132
PKS18	ESSEELYQEYMYTDISSGFFVSAKERFHDHP---RIRYDVLDISKDPVEQGFDP---RHYDLIIASNVLEATPYLTETLK	110
TENS	NVRALLKPGGYLLNLEKGTPE--SLRATFNFGGLEGWLAEERQLSPLMSPDGWAQLQKASFSGVHDVHVD---	194
SQTKS	NVRRLLKPGGKLLVLETTTRD--EMDLQVLFGLLPGWWSSEERKMSPLSTSSWEKVLKKTGFNGLDVELRDCCD--	201
Citrinin	NIHRILRPDGFLLLEMTTEQV--PWVDFIFGLLEGWVLFEDGRRH--ALQPATHWKKILTS-----	175
PKS1	KIQNLVAEGGRAIIESAAPLS-----EADLKTSGFSGAVTIEAPAG-V	155
PKS2	NVRLLKPGGQLLVLETKNKF--MVPGVVGTFTGYWAGIPDGRVDAFPQSLEAWDKALQNVGFSGLDIVLDDFPPEL	201
PKS3	TIRSLKLDGKLLILEPTSNK--LDMQLLFGTSPWWWVND-EPDKLSPILSLQGWDDTLRETGTGVD-----	196
PKS6	NIKKLLHPGGRLLQLCPSS--KWINTVFLGLPSWVYGSADGRNLNEPYFDVNGWKSQLLAAGFDSIEGIVLDSDE	194
PKS7	RIRKALKPGGKLIQCSFTPS--GWTLGIFGLFPGWVFGVDDGRVLSPSISIDDDWDLKENGFSGVDIVRDLGRE-	184
PKS8	NIRKALKPGGKLIQCSFTPS--GWTLGIFGLFPGWVFGVDDGRVLSPSISIDDDWDLKENGFSGVDIVRDLGRE-	193
PKS11	QCYDMELEDDGAMYVQLSGLRKAQYEDFIWG-----LFLNKYIFPGADASTPLANY-	166
PKS12	QCRALLRPGGHLLMLGTRM--TTAFQLLFGVLPGWFLGLDDNRVWAPSTTISEWNVLKKAGFSGVDAVTPYCSE	199
PKS14	NIKSLKPDGMLVLTLELPTL--RTANYTFGHFAGWVLEADARPSNPLIPVERWDELKAGAGTGVDSFVSDDEL	200
PKS15	NCRKLLKPGGRLIVLETKNKI--ARRSVVFGVLSGWVLEDDRRQWGPTELTEEEDWSRLRAEGFSGVDCFRDRED-	202
PKS16	NVRKLLKPGGRLMLLVINPA-DIAIN-----	160
PKS17	KVRKLLKPKGHLVLELVTEPD-MACAKVGFPIETWWSAKEWRRYSPLVTEQRWDELRLRQTGFSGVDASFRDYESL	208
PKS18	HTKTLKPDGRLLLQLCPVVS--NWTNFI FGLFPGWVWGKDDGRI DEPYISPKEWDTRLRDAGFAGVDASALDADLYL	185

**Figure 7.87** Sequence alignment of the C-MeT domain from PKS genes in *H. submonticulosa* with other fungal C-MeT domains. Functional C-MeT: Tenellin (TENS),<sup>118</sup> SQTKS (Squalestatin Tetraketide Synthase),<sup>137,140</sup> and citrinin;<sup>120</sup> Green colours: 'GAGTG' conserved cofactor motif, as well as the His-Glu dyad.<sup>109</sup> The C-MeT domain in red colour highlighted PKS are predicted inactive.

## Curriculum Vitae

2017.10 – 2021.09	<b>PhD</b> , Organic Chemistry, Biomolekulares Wirkstoffzentrum (BMWZ), Gottfried Wilhelm Leibniz Universität Hannover, Germany
2014.09 – 2017.06	<b>Master</b> , Medicinal Chemistry, School of Pharmacy, Guizhou University, Guiyang, China
2008.09 – 2011.06	<b>College</b> , Chemical Pharmaceutical Technology, China Pharmaceutical University, Nanjing, China
2004.09 – 2007.06	<b>High school</b> , Lingbao No.1 Senior High School, Lingbao, China

## List of Publications

1. D. Tian, E. Kuhnert, J. Ouazzani, D. Wibberg, J. Kalinowski, and R. J. Cox. *Chem. Sci.*, **2020**, 11, 12477–12484.
2. H. Yang, D. Tian, Y. Zeng, L. Huang, W. Gu, X. Hao and C. Yuan. *Biochem. Syst. Ecol.*, **2020**, 88, 103981.
3. Z. Li, L. Zuo, Y. Shi, D. Tian, L. Liu, Y. Yang, L. Zhou, X. Zhang, J. Kang, X. Hao, C. Yuan, and Z. Sun. *J. Sep. Sci.*, **2020**, 43, 978–986.
4. X. Lin,<sup>#</sup> D. Tian,<sup>#</sup> Y. Fu, Y. Li, L. Huang, W. Gu, J. Song, Y. Li, Y. B. David, M. Wen, C. Yuan, and X. Hao. *Eur. J. Med. Chem.*, **2019**, 162, 765–780.
5. Y. Zeng, P. Yi, W. Gu, C. Xiao, L. Huang, D. Tian, H. Yan, D. Chen, C. Yuan and X. Hao. *Org. Biomol. Chem.*, **2018**, 16, 4195–4198.1
6. D. Tian, W. Gu, L. Huang, T. Huang, Y. Fan, Y. Fu, J. Zhang, C. Yuan, and X. Hao. *Phytochem. Lett.*, **2017**, 19, 132–135.
7. D. Tian, P. Yi, L. Xia, X. Xiao, Y. Fan, W. Gu, L. Huang, Y. B. David, Y. Di, C. Yuan, and X. Hao. *Org. Lett.*, **2016**, 18 (22), 5904–5907.