Engineering *Aspergillus oryzae* for the Production of Biosynthetic Proteins

Von der Naturwissenschaftlichen Fakultät

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Yunlong Sun, lixue shuoshi (Master of Science), Jiangxi Science and Technology Normal University

2024

Referent: Prof. Dr. Russell J. Cox

Korreferentin: Prof. Dr. rer. nat. Natalia Tschowri

Tag der Promotion: 19. 04. 2024

Abstract

Keywords: hrPKS, reprogramming, engineering, sesquiterpene, tetraketide, *Aspergillus oryzae*.

This investigation aims to strategically engineer the biosynthesis of fungal natural products through heterologous expression. It includes modifications to the heterologous host to boost productivity; leveraging the product diversity of an individual enzyme *via* mutasynthesis and combining biosynthetic genes from multiple pathways in a single heterologous host to creat new compounds.

To improve tetraketide **25** titer in *A. oryzae* NSAR1 through squalestatin tetraketide synthase (SQTKS) expression, we co-expressed two hydrolase-encoding genes. The 7-day fermentation in DPY medium involved daily titer measurements. Although the hydrolases had a modest impact, the titer reached around 3.2 mg/L, marking a fourfold increase from the previous experiment.

We conducted protein-level engineering of the *cis*-ER domain in SQTKS, introducing 10 mutation groups *via* yeast recombination into the complete SQTKS. These mutations were designed based on isolated ER mutants of SQTKS, analyzed through a combination of computational modeling and experimental assays with various mimic substrates by previous co-worker. After expressing them in *A. oryzae* NSAR1, we screened for correct transformants, performed fermentation and chemical extraction, and utilized LCMS for compound detection. However, no new compounds were generated in these expression experiments.

We performed pathway-level engineering by integrating genome-mined genes from multiple biosynthetic gene clusters (BGCs) into *A. oryzae* NSAR1. We co-expressed genes from the four BGCs of sporogen AO1 **27**, hypoxylan A **73**, eremoxylarin D **123** and PR-toxin **107**. These four BGCs share a common core carbon skeleton, while possessing distinct tailoring enzymes. Twenty oxygenated aristolochene congeners were synthesized, and their structures were characterized, featuring notable compounds such as the natural product hypoxylan A **73** and an epimer of guignaderemophilane C **37**. A novel fungal aromatase enzyme has been identified, which catalyses the production of phenols *via* oxidative demethylation.

We investigated fungal pathways for tetraketide multiforisins and islandic acid-related compounds. Heterologous expression experiments yielded high titers of these compounds and pathway intermediates, leading to the structure characterization of 14 isolated compounds, including multiforisin H **143h** and I **143i**. These results not only clarified the pathway but also laid the groundwork for the total biosynthesis of this metabolite class. Attempts to add genes from the islandic acid BGC into *A. oryzae* for synthesizing islandic acid **144** or Allantopyrone A **145** proved unsuccessful, as none of the transformants generated compounds beyond those already synthesized by the *A. oryzae* hosts.

Zusammenfassung

Schlagwörter: hrPKS, Reprogrammierung, Engineering, Sesquiterpen, Tetraketid, *Aspergillus oryzae*.

Diese Untersuchung zielt darauf ab, die Biosynthese von Pilznaturstoffen durch heterologe Expression strategisch zu optimieren. Dies umfasst Modifikationen am heterologen Wirt, um die Produktivität zu steigern; die Nutzung der Produktvielfalt eines einzelnen Enzyms durch Mutasynthese; sowie die Kombination von biosynthetischen Genen aus mehreren Wegen in einem einzigen heterologen Wirt, um neue Verbindungen zu schaffen.

Um den Titer von Tetraketid **25** in *A. oryzae* NSAR1 durch SQTKS-Expression zu verbessern, haben wir zwei Hydrolase-kodierende Gene koexprimiert. Die 7-tägige Fermentation in DPY-Medium beinhaltete tägliche Titermessungen. Obwohl die Hydrolasen einen bescheidenen Einfluss hatten, erreichte der Titer etwa 3,2 mg/L, was eine Vervierfachung im Vergleich zum vorherigen Experiment darstellt.

Wir führten eine Protein-Level-Engineering des *cis*-ER-Domäne in SQTKS durch, indem wir 10 Mutationsgruppen über Heferekombination in das gesamte SQTKS einschleusten. Diese Mutationen wurden auf der Grundlage isolierter ER-Mutanten von SQTKS entworfen, die durch eine Kombination aus computergestützter Modellierung und experimentellen Assays mit verschiedenen Imitat-Substraten von vorherigen Kollegen analysiert wurden. Nach ihrer Expression in *A. oryzae* NSAR1 haben wir nach korrekten Transformanten gescreent, Fermentation und chemische Extraktion durchgeführt und LCMS zur Verbindungserkennung genutzt. Jedoch wurden in diesen Expressionsversuchen keine neuen Verbindungen erzeugt.

Wir führten ein Engineering auf der Pathway-Ebene durch, indem wir genomgeförderte Gene aus mehreren biosynthetischen Genclustern (BGCs) in *A. oryzae* NSAR1 integrierten. Wir co-exprimierten Gene aus vier BGCs von sporogen AO-1 **27**, Hypoxylan A **73**, Eremoxylarin D **123**, PR-Toxin **107**. Diese vier BGCs teilen einen gemeinsamen Kernkohlenstoff-Skelett, und besitzen gleichzeitig unterschiedliche Anpassungsenzyme. Es wurden zwanzig oxygenierte Aristolochene-Kongenere synthetisiert und charakterisiert, wobei ihre Strukturen herausgearbeitet wurden. Hervorzuhebende Verbindungen sind dabei das Naturprodukt Hypoxylan A **73** und ein Epimer von Guignaderemophilane C **37**. Des Weiteren wurde eine neue pilzliche Aromatase entdeckt, die durch oxidative Demethylierung Phenole produziert.

Wir haben die pilzlichen Wege für Tetraketid-Multiforisine und mit Islandic-Säure verwandte Verbindungen untersucht. Heterologe Expressionsversuche ergaben hohe Titel dieser Verbindungen und Zwischenprodukte des Wegs. Dies führte zur Strukturcharakterisierung von 14 isolierten Verbindungen, einschließlich Multiforisin H **143h** und I **143i**. Diese Ergebnisse klärten nicht nur den Weg, sondern legten auch den Grundstein für die Gesamtbiosynthese dieser Metabolitenklasse. Versuche, Gene aus dem Islandic-Säure-BGC in *A. oryzae* einzufügen, um Islandic-Säure **144** oder Allantopyron A **145** zu synthetisieren, waren erfolglos, da keiner der Transformanten Verbindungen erzeugte, die über diejenigen hinausgingen, die bereits von den *A. oryzae*-Wirten synthetisiert wurden.

Acknowledgement

I express my sincere gratitude to Professor Russell J. Cox for his outstanding supervision over the past four years and his invaluable guidance throughout this captivating project. I am truly grateful for the opportunity to explore new ideas and the numerous inspiring discussions we've had over the years. Working under your supervision has been a thoroughly enjoyable experience, and I will always cherish the wonderful time I spent in this group.

I express my thanks to all collaboration partners, including Geraldine Le Goff, Guillaume Arcile, Jamal Ouazzani, Bart Verwaaijen, Daniel Wibberg, Jörn Kalinowski, and Marc Stadler. A special appreciation goes to Professor Thomas Ostenfeld Larsen for generously providing the strain for my pyrone project.

I thank Dongsong and Eric for their assistance with the pyrone project, and I extend my thanks to Jenny and Kevin for their help with the aristolochene project. I thank Ahmed for measuring my samples with NMR at the BMWZ. I thank Shenyu for great talks about solving NMR.

I am grateful to Lukas for the academic assistance he provided me in the first year of my PhD.

A heartfelt acknowledgment is owed to the team managing the service facilities at the OCI and the technical staff in the BMWZ. My sincere thanks to Katja, Tjorven, Serge, Birgit, and Doreen for consistently going above and beyond to help. I also appreciate the help of Toma and Evgeniya.

I would like to express my thanks to all current and former members of the Cox team for fostering such a positive and collaborative work environment. The current members Katharina, Ayodele, Maurice, Yuechen, Henrike, Jing, Ameer, Christina, Anna, Nick and former members Daowan, Jin, Chongqing, Sen, Haili, Mary, Slawik have all played vital roles in creating this excellent atmosphere.

I am grateful to the China Scholarship Council for covering my living expenses in Germany.

Lastly, I convey my gratitude and love to my family for their unwavering support throughout my long long studies.

Abbreviations and Units

ACP	acyl carrier protein	MOS	3-methylorcinaldehyde synthase
BGC	biosynthetic gene cluster	MSAS	6-methylsalicylic acid synthase
BLAST	basic local alignment search tool	NSAS	norsolorinic acid synthase
AT	acetyltransferase	NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
bp	base pair	nrPKS	non-reducing PKS
cDNA	complementary DNA	nOe	nuclear Oberhauser effect
C-MeT	*C*-methyltransferase	NMR	nuclear magnetic resonance
CoA	coenzyme A	NOESY	nuclear overhauser effect spectroscopy
COSY	correlation spectroscopy	nrPKS	non-reducing PKS
CON	condensation domain	NRPS	non-ribosomal peptide synthetase
carb	carbenicillin	PCR	polymerase chain reaction
DAD	diode array detector	PEG	polyethylene glycol
DH	dehydratase	PKS	polyketide synthase
ER	enoyl reductase	PPant	phosphopantetheinyl
ESI	electronspray ionization	ppm	parts per million
EIC	extracted ion chromatogram	P_{amyB}	taka-amylase A gene promoter
ELSD	evaporative light scattering detector	P_{gpdA}	glyceraldehyde-3-phosphate dehydrogenase gene promoter
FAS	fatty acid synthase	P_{eno}	enolase gene promoter
FMO	FAD-dependent monooxygenase	P_{adh}	alcohol dehydrogenase gene promoter
FAD	flavin adenine dinucleotide	prPKS	partially reducing PKS
FPP	farnesyl diphosphate/ pyrophosphate	P450	cytochrome P450 monooxygenase
gDNA	genomic DNA	PPTase	Phosphopantetheinyl transferase
HSQC	heteronuclear single quantum coherence	PT	product template
Hz	Hertz	RT-PCR	reverse transcription PCR
HRMS	high resolution mass spectrometry	R	reductive release domain
^{1}H-NMR	proton NMR	rpm	rounds per minute
hrPKS	highly reducing polyketide synthase	RNA	ribonucleic acid
HPLC	high performance liquid chromatography	SDR	short chain dehydrogenase/reductase
HMBC	heteronuclear multiple bond correlation	SQHKS	squalestatin hexaketide synthase
HeLa cells	cervical cancer cells	S	serine
iPKS	iterative polyketide synthases	SQTKS	squalestatin tetraketide synthase
IC50	half maximal inhibitory concentration	SAT	starter unit acyl carrier protein
kana	kanamycin	SAM	S-adenosyl methionine
kDa	kilo Dalton	TC	terpene cyclase
kb	kilo base pairs	TIC	total ion current
KR	ketoacyl reductase	TE	thiolesterase
KS	ketoacyl synthase	TH	thioester hydrolase
LPS	lipopolysaccharide	UV	ultra violet
MS	mass spectrometry	UPLC	ultra-performance liquid chromatography
m/z	mass to charge ratio	Vero cells	verda reno (kidney epithelial cells)
mRNA	messenger RNA	WT	wild type

Contents

1. Introduction

1.1 Natural Products from Fungi

The Fungi Kingdom encompasses an incredibly diverse array of organisms.[1] Fungi can be found in virtually every ecological habitat, and it has been estimated that there are approximately 1.5 million fungal species on Earth in 1991.[2, 3] According to literature published in 2017, a more accurate estimate for the total number of fungal species on Earth falls within the range of 2.2 to 3.8 million.[4,5] With just 120,000 officially named species, the best-case scenario sees only 8% formally classified, dropping to a worst-case scenario of as low as 3% with scientific nomenclature.[4–6]

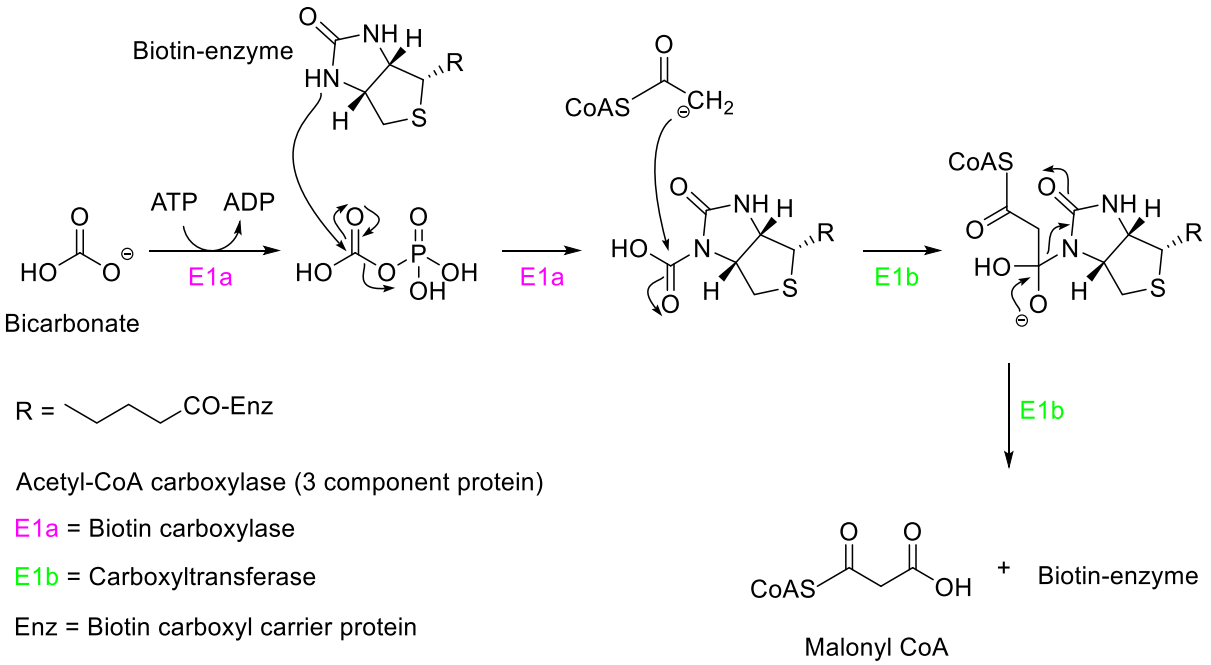
Nevertheless, only a small fraction of this fungal diversity has been scientifically explored.[2, 3] The isolation and characterization of natural products from fungal fruiting bodies, such as quinoid pigments, were first documented in 1877 by Stahlschmidt and Thörner.[7] Since then, organic chemists have recognized fungi as a valuable source of bioactive compounds, dedicating efforts to investigate these organisms for compound isolation, structure elucidation, and synthesis.[8–12] Fungal-derived natural products undergo a multitude of enzyme-catalyzed reactions, yielding a vast array of chemical structures characterized by the presence of chiral centers, fused rings, and a wide spectrum of functional groups, showcasing their remarkable diversity.[13]

The identification of genes for natural products has experienced substantial growth in recent decades, thanks to advancements in technologies such as whole-genome sequencing [14] and computational tools like antiSMASH,[15] FGENESH,[16] Clinker,[17] Cblaster,[18] and AlphaFold.[19] In addition, gene editing tools like CRISPR have played a pivotal role for fungal genome editing.[20] Furthermore, heterologous expression has become a common approach for elucidating gene functions and establishing connections to the chemical steps involved in biosynthesis of natural products.[21] Natural products synthesized through biosynthesis offer cost-effective and adaptable alternatives to chemically synthesized compounds for medical use.[22]

1.1.1 Fatty Acid Biosynthesis

Fatty acid biosynthesis is an intricate process of primary metabolism involving multiple enzymatic steps and regulatory mechanisms, unfolding in three key phases: the citrate shuttle, acetyl-CoA carboxylase activity, and the fatty acid synthase complex. The citrate shuttle initiates with the conversion of citrate into acetyl-CoA and oxaloacetate, a transformation catalyzed by ATP-citrate lyase.[23] This step serves as a supply of acetyl-CoA for the synthesis of fatty acids, polyketides, mevalonate, and other biosynthetic pathways.[24] Acetyl-CoA is converted into malonyl-CoA by acetyl-CoA carboxylase, a key, rate-limiting step in fatty acid biosynthesis.[25] The final phase entails the actual synthesis of fatty acids from malonyl-CoA and acetyl-CoA, a process carried out by the fatty acid synthase.

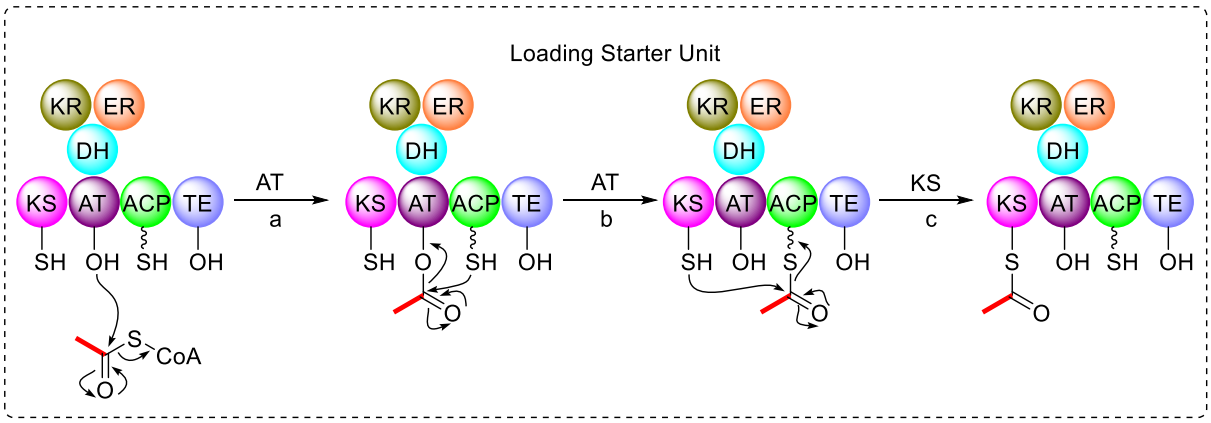
The conversion of acetyl-CoA into malonyl-CoA is a carboxylation process catalyzed by the three-component enzyme, acetyl-CoA carboxylase. This enzyme comprises three distinct components: biotin carboxylase (E1a), carboxyltransferase (E1b), and a biotin-enzyme complex in which biotin is bound to the biotin carboxyl carrier protein. Initially, bicarbonate is activated through phosphorylation by ATP. Subsequently, the phosphorylated bicarbonate carboxylates biotin-enzyme, resulting in the formation of carboxybiotin-enzyme under the catalysis of E1a. Finally, carboxybiotin carboxylates the enolate of acetyl CoA, leading to the production of malonyl CoA, a reaction facilitated by carboxyltransferase E1b.[26,27]



Scheme 1.1.1 Conversion of Acetyl-CoA into malonyl-CoA.[26,28]

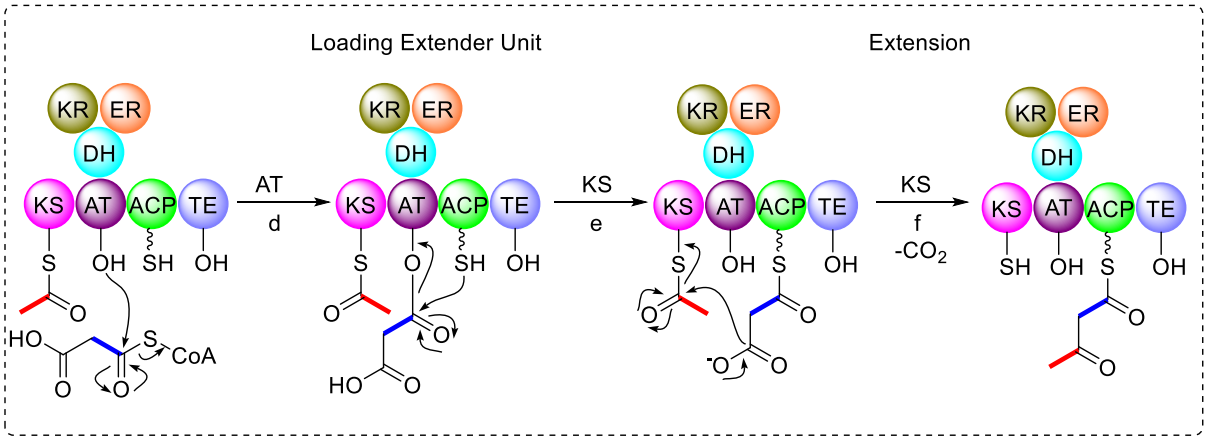
The third step involves the conversion of acetyl-CoA and malonyl-CoA into palmitate, a process that occurs in the presence of NADPH catalyzed by FAS. The main chain formation requires elongation through the incorporation of extender units *via* Claisen-like condensation and β-processing, which encompasses multiple enzymatic reactions involving reduction, dehydration, and subsequent reduction steps. These reactions collectively lead to the synthesis of saturated straight-chain fatty acids.

In the initiation step, the active site serine of malonyl-acetyl transferase (AT) binds with an acetyl-CoA starter unit (step **a**, Scheme 1.1.2) through the formation of an ester bond. Subsequently, the *holo*-acyl carrier protein (*holo*ACP) engages with the acetyl-CoA, and AT catalyses the transfer of the acetyl-CoA to ACP, resulting in the formation of a thioester (step **b**, Scheme 1.1.2) located at the end of the phosphopantetheine. Following this, this acetyl-CoA is transfered to the ketosynthase domain (KS), establishing another thioester (step **c**, Scheme 1.1.2).



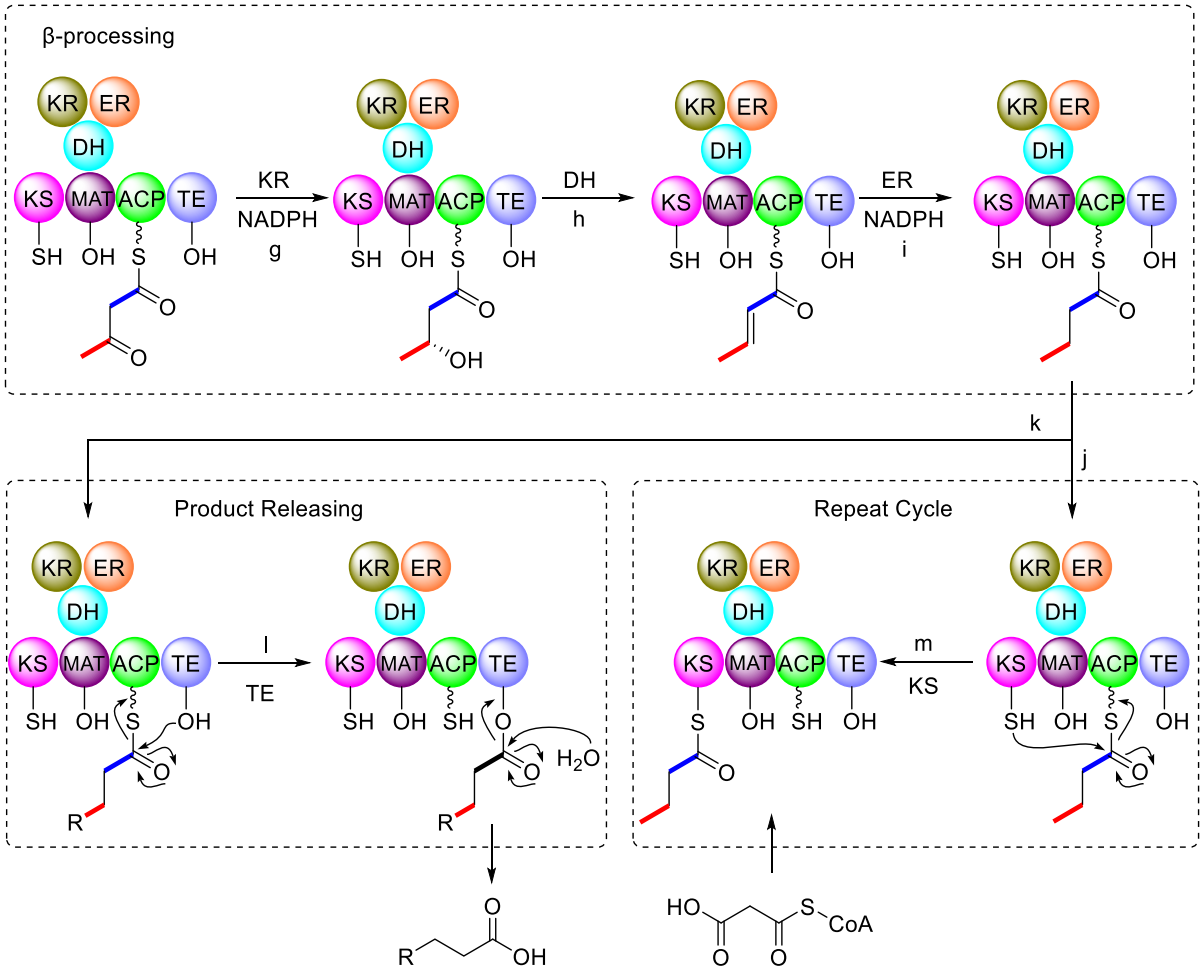
Scheme 1.1.2 Biosynthesis of loading starter unit by FAS.

The extender unit, malonyl-CoA, is loaded onto AT, and then it is transferred to *holo*ACP (steps **d** and **e**, Scheme 1.1.3). The chain extension process begins at step **f** when malonyl-CoA initiates a nucleophilic attack on the carbonyl of acetyl-CoA, simultaneously liberating CO_2. This reaction yields acetoacetyl-CoA bound *via* phosphopantetheine to *holo*-ACP (step **f**, Scheme 1.1.3). These steps can be iterated, ultimately culminating in the formation of a poly-β-keto ester if the domains responsible for β-processing are not active.



Scheme 1.1.3 Biosynthesis of loading extender unit and the chain extension by FAS.

After the formation of acetoacetyl-ACP (step **f**), it undergoes a series of β-modification steps, involving the enzymatic domains of ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER). The KR domain carries out a reduction of the β-keto group in acetoacetyl-ACP, resulting in the formation of β-hydroxyacetyl-ACP, a step that requires NADPH as a co-factor (step **g**, Scheme 1.1.4). Subsequently, the DH domain catalyses a dehydration by eliminating water, leading to the formation of an *E*-alkene and the generation of *α, β*-unsaturated acyl-ACP (step **h**, Scheme 1.1.4). Lastly, the ER domain reduces the double bond in this α, β-unsaturated acyl-ACP to yield saturated fatty acyl-ACP in the presence of NADPH as a cofactor (step **i**, Scheme 1.1.4).



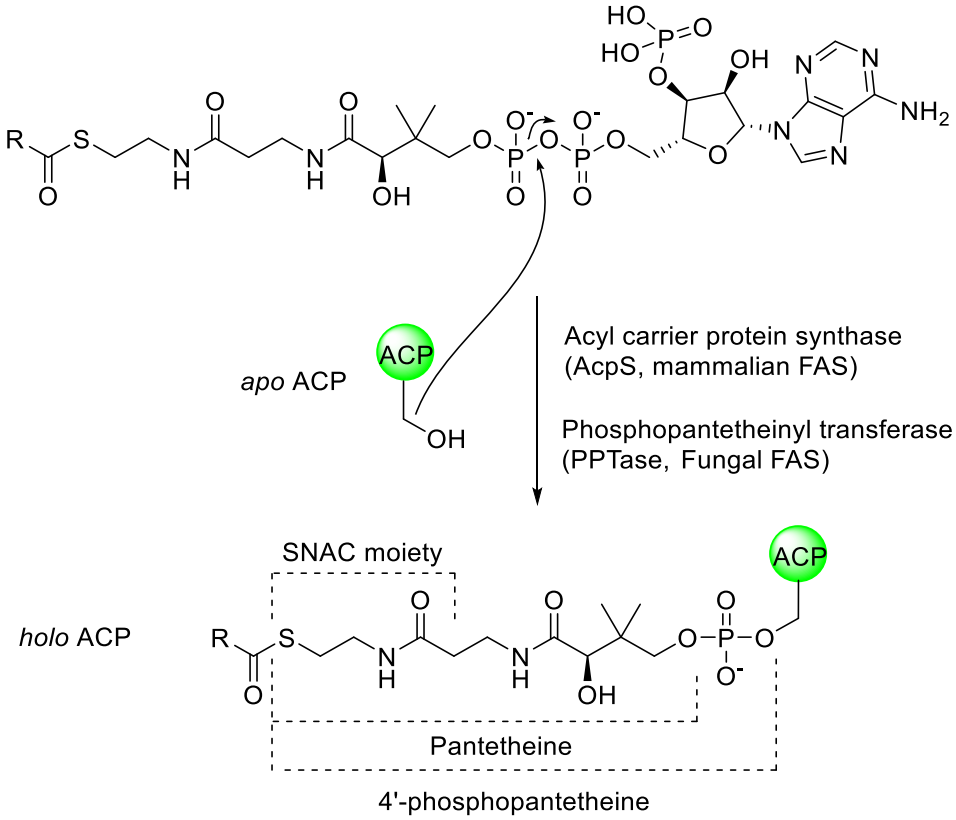
Scheme 1.1.4 Biosynthesis of β-processin, product releasing and repeating cycle by FAS.

When the growing fatty acyl chain reaches a specific length, it is liberated from the FAS by the thiolesterase (TE) domain. The saturated fatty acyl chain can be directly transferred to TE domain and subsequently released as the final product (step **k** and **l,** Scheme 1.1.4). Alternatively, it continues with another round of chain elongation and β-processing (step **j** and **m,** Scheme 1.1.4). In this case, the alkyl chain is transferred to the KS domain, while another malonyl-CoA molecule is introduced to support further elongation (step **j** and **m,** Scheme 1.1.4).

Fatty acid biosynthesis requires the essential involvement of the ACP in every step. It all begins with the AT, which systematically loads the starter and extender units onto the ACP. From there, the ACP takes charge, shuttling these intermediates through the complete catalytic cycle, which includes chain elongation and β-processing.[29]

The discovery of ACP, a small protein characterized by its thermal stability, represents a significant milestone in understanding fatty acid synthesis.[30] ACP proteins, which are found in both type I and type II systems, typically have a compact size of around 9 kDa and exhibit a conserved canonical structure comprised of four helices.[31]

A conserved serine residue located at the N-terminus of helix II undergoes post-translational modification catalyzed by a phosphopantetheinyl transferase (PPTase), which installs a 4'-phosphopantetheine arm (PPant) derived from coenzyme A (Scheme 1.1.5).[32] This modification converts the ACP from its *apo* form into its functional *holo*-ACP state.[33,34] The covalent binding of acyl substrates occurs at the free thiol group on the PPant arm, while the effective delivery of these substrates to the various different active sites of FAS is aided by the length and flexibility of this mobile arm (Scheme 1.1.5).[29, 31, 35, 36]



Scheme 1.1.5 Process of introduction of phosphopantetheine chain to *apo*ACP.

1.1.2 Fungal Polyketide Biosynthesis

Fungal polyketides exhibit a wide range of chemical structures, commonly categorized into two groups: aromatic and aliphatic. Type I iterative polyketide synthases (iPKS) are responsible for their synthesis, and they are further classified as non-reducing (nr), partially reducing (pr), or highly reducing (hr) PKS based on the composition of catalytic domains and oxidation levels of the polyketide intermediates.[37] Fungi are also known to possess Type III PKS.[38–40]

NrPKS usually synthesise polyketides characterized by fully oxidised carbon chains without any reductive steps during chain assembly.[41] In contrast, prPKS are responsible for the synthesis of polyketides in which particular carbonyl groups undergo selective, limited reductions, often involving just one reduction step.[41] HrPKSs produce polyketides distinguished by extensive reduction, where the

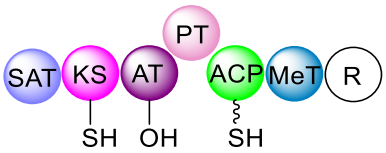
majority of carbonyl groups are transformed into hydroxyl groups, followed by additional reductions in subsequent stages.[37] The degree of reduction is variable and tightly regulated through the PKS programming.[37]

1.1.2.1 Non- and Partially- Reducing PKS

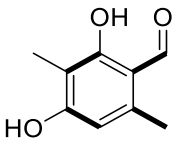
Non-reducing PKS typically exhibit a structural composition characterized by an N-terminal starter unit ACP transacylase domain (SAT), followed by KS, AT, product template (PT), and ACP domains. Located at the C-terminal region is the TE-CLC domain, which could affect the chain length.[42] Within the case of norsolorinic acid synthase (NSAS) participating in aflatoxin biosynthesis,[43] the TE-CLC domain serves as a catalyst for a subsequent (Claisen) cyclization reaction (Figure 1.1.1).[44] The SAT domain of NSAS discriminates and selectively acquires a C_6 fatty acid from a FAS, subsequently catalyzing the transfer of this unit onto the ACP (Figure 1.1.1). This process is followed by seven iterative extensions using malonyl-derived ketides, resulting in the synthesis of precursor of norsolorinic acid **1**.[45] The PT domain acts necessarily in binding the linear intermediate acyl-CoA and catalyzing the sequential aldol cyclizations and dehydrations, resulting in the formation of the bicyclic ACP-thiolester intermediate, which subsequently undergoes the conclusive Claisen cyclization (Figure 1.1.1).[46]

Figure 1.1.1 Domian architecture of NSAS and the product.

In another example, 3-methylorcinaldehyde synthase (MOS), encoded by *ASpks1* from *Acremonium strictum*, was expressed in *Aspergillus oryzae*, resulting in the production of 3-methylorcinaldehyde **2** (Figure 1.1.2).[47] The MOS consists of a set of catalytic domains, including SAT, KS, AT, PT, ACP, and *C*-MeT, with a domain corresponding to an NADPH-dependent thiolester reductase (R) situated at the C-terminus (Figure 1.1.2).[47] Reductive release is a well-documented mechanism that is commonly observed in the generation of aromatic aldehydes.[48, 49] In instances where a lot of nrPKS undergo methylation, a *C*-MeT domain is located between the ACP and R domains.[50]



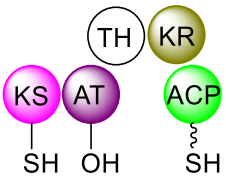
Non-reducing PKS MOS



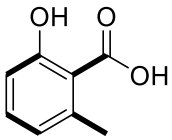
3-methylorcinaldehyde **2**

Figure 1.1.2 Domian architecture of MOS and the product.

In comparison to the nrPKS and hrPKS categories, prPKSs are relatively infrequent. A famous example is ATX, which is homologous to MSAS (6-methylsalicylic acid synthase) from *Aspergillus terreus* and also synthesize 6-methylsalicylic acid **3** (Figure 1.1.3).[51] ATX has been investigated in detail. It comprises functional domains including KS, AT, thiolester hydrolase (TH), KR, and ACP (Figure 1.1.3).[52] ATX catalyses the iterative elongation and modification of the polyketide chain, employing one acetyl-CoA and three malonyl-CoA molecules through three decarboxylative Claisen thiolester condensation reactions. It further involves one TH-mediated hydrolysis event after the third elongation cycle, resulting in the formation of 6-methylsalicylic acid **3** (Figure 1.1.3).[53, 54]



Partially-reducing PKS ATX



6-methylsalicylic acid **3**

Scheme 1.1.3 Domian architecture of ATX and the product.

1.1.2.2 Highly-Reducing PKS

HrPKS are commonly found in fungi and share strong structural and functional similarities with vertebrate FAS. These kinds of hrPKS such as AAL-Toxin PKS (ALT1)[55] and squalestatin tetraketide synthase (SQTKS),[56] contain essential catalytic domains like KS, AT, DH, *C*-MeT, and KR. The final domain is consistently the ACP.[56] Some hrPKS lack functional enoyl reductase domains (ER^0) and require a *trans*-acting ER to be encoded within the BGC for collaboration, as seen in cases like lovastatin nonaketide synthase (LNKS),[57] and tenellin pentaketide synthase (TENS).[58] There is normally no off-loading *cis*-domain in hrPKS.[59]

One classic example lies in the BGC responsible for the synthesis of lovastatin **15**, a well-known and extensively studied case, where two distinct types of hrPKS are present.[60] LNKS, encoded by *lovB*, possesses domains including KS, AT, DH, *C*-MeT, KR and ACP. LNKS ends with a condensation (CON) domain at C-terminus, a feature commonly found in nonribosomal peptide synthetases (Figure 1.1.4).[57] It possesses an ER domain that is inactive because it lacks critical active site residues, necessitating a *trans*-acting ER, which is encoded by *lovC* (Figure 1.1.4).[57]

In contrast, lovastatin diketide synthase (LDKS), encoded by *lovF*, has all the functional domains like KS, AT, DH, C-MeT, ER, KR and ACP (Figure 1.1.4). Neither LNKS nor LDKS possess C-terminal thiolesterase domains or any other common off-loading domains (Figure 1.1.4).[57]

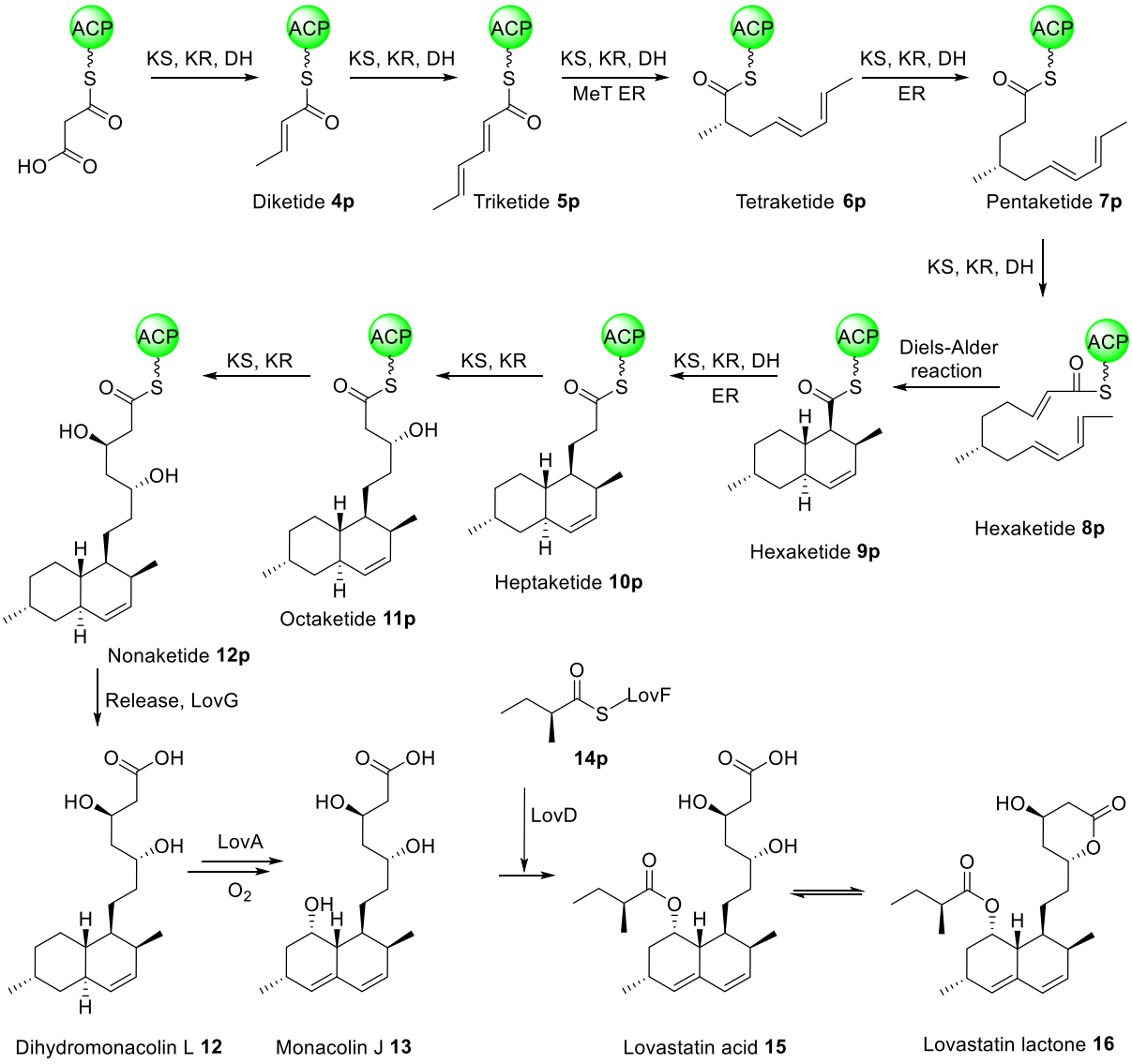
Figure 1.1.4 Domian architectures of LNKS and LDKS.

The biosynthesis of lovastatin **15** has been thoroughly investigated.[60] Initiation involves the attachment of the acetyl starter unit and the malonyl CoA extender unit to LovB. Each Claisen condensation is catalyzed by the KS domain, followed by subsequent ketoreduction and dehydration, forming diketide **4p**. Up to the point where the heptaketide **10p** is formed, the KS, KR, and DH domains are all active through the early six elongation steps (Scheme 1.1.6).

The *trans*-acting ER becomes involved in the programming starting at triketide **5p**. During this stage, the *C*-MeT domain adds a methyl group, resulting in the formation of tetraketide **6p**. The *trans*-acting ER continues its role, while the *C*-MeT domain disengages from this step, leading to the creation of pentaketide **7p.** In the following step, the *trans*-acting ER is not involved, and this leads to the formation of hexaketide **8p.** A Diels-Alder cycloaddition takes place at this stage,[61, 62] resulting in the synthesis of the fused rings that constitute the decalin system **9p** (Scheme 1.1.6).

A chain extention and a full reduction is then followed by collaboration of KS, KR, DH and ER to get heptaketide **10p**. In the next two steps, DH remains inactive and *trans*-acting ER is not engaged, allowing the chain to extend to octaketide **11p** and, in the final round, to nonaketide **12p**. Nonaketide **12p** is then released from the ACP, forming dihydromonacolin L **12** (Scheme 1.1.6).

Dihydromonacolin L **12** undergoes oxidation catalyzed by a cytochrome P450 oxygenase LovA, leading to monacolin J **13**.[63] Concurrently, the diketide **14p** produced by LovF (LDKS) is covalently linked to monacolin J **13** *via* the catalysis of LovD, a transferase enzyme.[64] This whole PKS programming and tailoring modification results in the formation of lovastatin in two forms, **15** and **16** (Scheme 1.1.6).



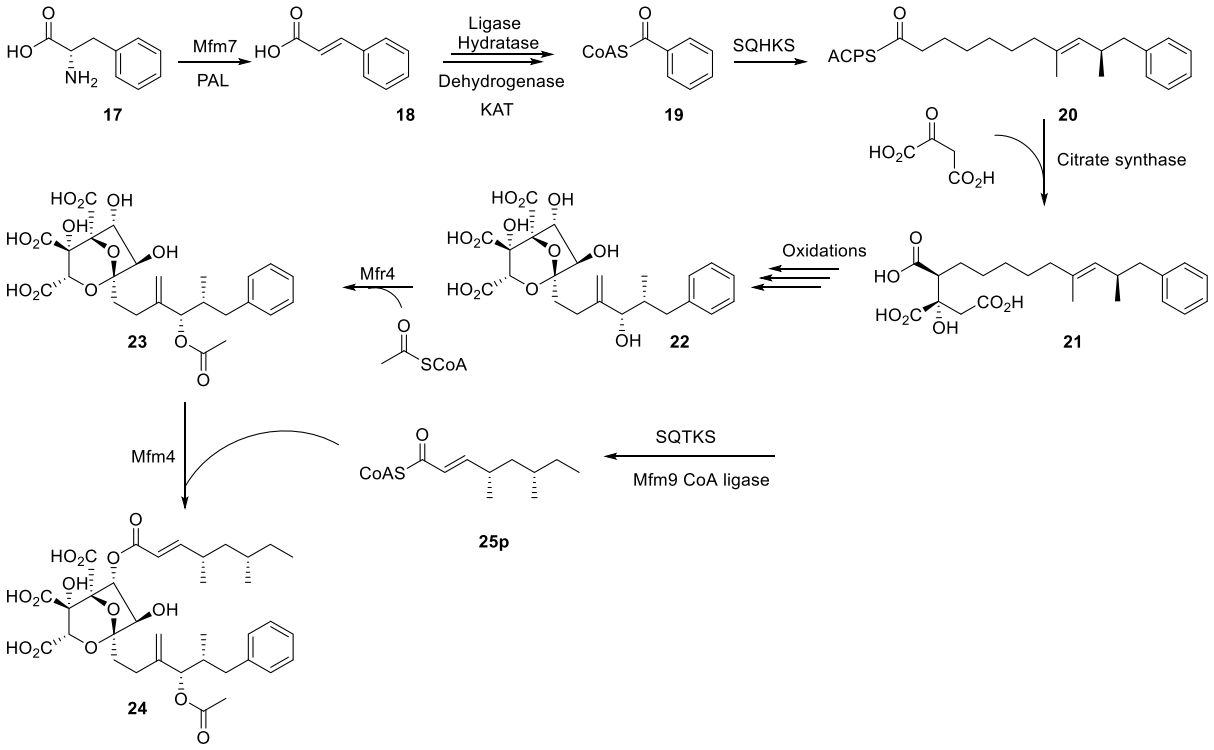
Scheme 1.1.6 Biosynthetic pathway of lovastatin.[65]

1.1.2.3 Squalestatin Biosynthesis

Squalestatin S1 **40** (SQS1, or zaragozic acid A) is a fungal metabolite obtained from *Phoma sp*. strain C2932,[66] an uncharacterized strain MF5453,[67] and *Setosphaeria khartoumensis*.[68] SQS1 displays a wide range of antifungal characteristics and has been shown to lower serum cholesterol levels by as much as 75% in marmosets.[69]

The proposed biosynthetic pathway of SQS1 starts with phenylalanine **17**. The first step in the conversion of phenylalanine **17** to **18** is catalyzed by phenylalanine ammonia lyase (PAL), which is encoded by *mfm7*. To catalyze the transition from **18** to benzyl CoA **19**, it has been reported that this process requires enoyl-CoA ligase, enoyl-CoA hydratase, dehydrogenase, and 3-ketoacyl-CoA thiolase (KAT), a series of enzymatic reactions reminiscent of biosynthetic processes in plants (Scheme 1.1.7).[70, 71]

Squalestatin hexaketide synthase (SQHKS) generates a carbon skeleton **20**, which is condensed with oxaloacetate through the action of citrate synthase (CS). This process results in the formation of an early alkyl citrate intermediate **21**. Then **21** undergoes oxidative transformations to yield **22**, which is then coupled with an acetyl group to produce **23**. Following this, tetraketide **25p**, synthesized by the squalestatin tetraketide synthase (SQTKS) [72], is incorporated into the structure, ultimately yielding squalestatin S1 **24** (Scheme 1.1.7).



Scheme 1.1.7 Proposed biosynthetic pathway for squalestatin S1 **24**.[73]

1.1.3 Eremophilane-type Sesquiterpenes

Eremophilanes, belonging to the sesquiterpene family, were first identified in 1932 by Simonsen et al. from the plant *Eremophila mitchelli*, with eremophilone as the inaugural term.[74] Typically featuring a decalin core structure connected to one-carbon substituents at positions C-4 and C-5, these sesquiterpenes exhibit structural diversity, incorporating either an isopropenyl or isopropyl group at C-7, resulting in various forms such as alcohol, acid, ester, and furan. Known for their wide range of biological activities, including anti-inflammatory, anti-neoplastic, and antibacterial effects, eremophilane-type derivatives have also been extensively discovered in fungi.[76–78] Despite the numerous structurally varied compounds identified over the decades, there remains limited exploration into the mining of tailoring enzymes and biosynthesis for eremophilane-type sesquiterpenes.

The metabolite phomenone **26** exhibits phytotoxicity and *in vitro* testing demonstrates its ability to impede polysome activity.[79] Sporogen AO-1 **27** was isolated from *A. oryzae* and showed the ability to induce the development of phialospores.[80] Moreover, Sporogen AO-1 **27** acts as an inhibitor of human

inducible nitric oxide synthase expression[81] and exhibits antimalarial properties against *Plasmodium falciparum*, along with cytotoxic effects against Vero cells.[82]

Xylarenones A **28** and B **29** were isolated from the endophytic fungus *Xylaria sp* NCY2.[83] Compound **28** was likewise obtained from an endophytic fungus that was isolated from *Alibertia macrophylla*.[84] Both **28** and **29** exhibit a hydroxyl group at C-1 with stereo configurations that are opposite to those of C-3 in **26** and **27**. This observation is of particular interest in mining enzymes that can selectively oxidize specific positions on the same backbone. Compound **28** demonstrated a significant inhibitory effect on pepsin activity, with an IC_{50} value of 2.43 μM.[84] Both **28** and **29** exhibited antibacterial properties against Gram-positive bacteria, with **29** showing slightly stronger activity than **28**. Both of them exhibited a moderate level of antitumor activity against HeLa cells.[83]

Phomadecalin D **30** was obtained from *Phoma sp* NRRL 25697, a fungal strain that inhabits the stromata of a *Hypoxylon* species.[85] Compound **30** was also reported to be isolated from the endophytic fungus *Microdiplodia sp* KS 75-1.[86] The formation of a double bond at carbon C-1 and C-2 in this structure is likely attributed to a dehydrogenase reaction or the elimination of alcohol from an intermediate, such as **28** or **29**. Phomadecalin D **30** exhibited moderate activity against *Bacillus subtilis* ATCC 6051 and *Staphylococcus aureus* ATCC 29213.[85]

Dacrymenone **31** was obtained through the fermentation of *Dacrymyces* sp 32-98.[87] Distinguished by a hydroxyl group at C-2, an acetoxy group at C-3, and a ketone at C-8, the structure of **31** differs from the previously mentioned compounds. In addition to its inhibitory effects on *Aspergillus ochraceus* and *Cladosporium cladosporioides*, **31** demonstrates the ability to inhibit bacterial growth within the concentration range of 25-100 μg/ml (86-343 μM).[87]

Periconianone A **32** and B **33** were obtained from an endophytic fungus *Periconia sp*.[88] **32** possesses a rare 6/6/6 tricyclic carbon skeleton, consists of two carbonyl groups, two olefinic double bonds, and two hydroxyl groups.[88] It was hypothesized to be formed from an atypical late-stage aldol cyclization reaction of the highly oxidized bicyclic compound **33**, but the biosynthesis remains unexplored.[88] Both **32** and **33** exhibited inhibitory effects on NO generation in lipopolysaccharide (LPS)-stimulated murine microglia BV2 cells, with IC_{50} values of 0.15 and 0.38 μM, respectively.[88]

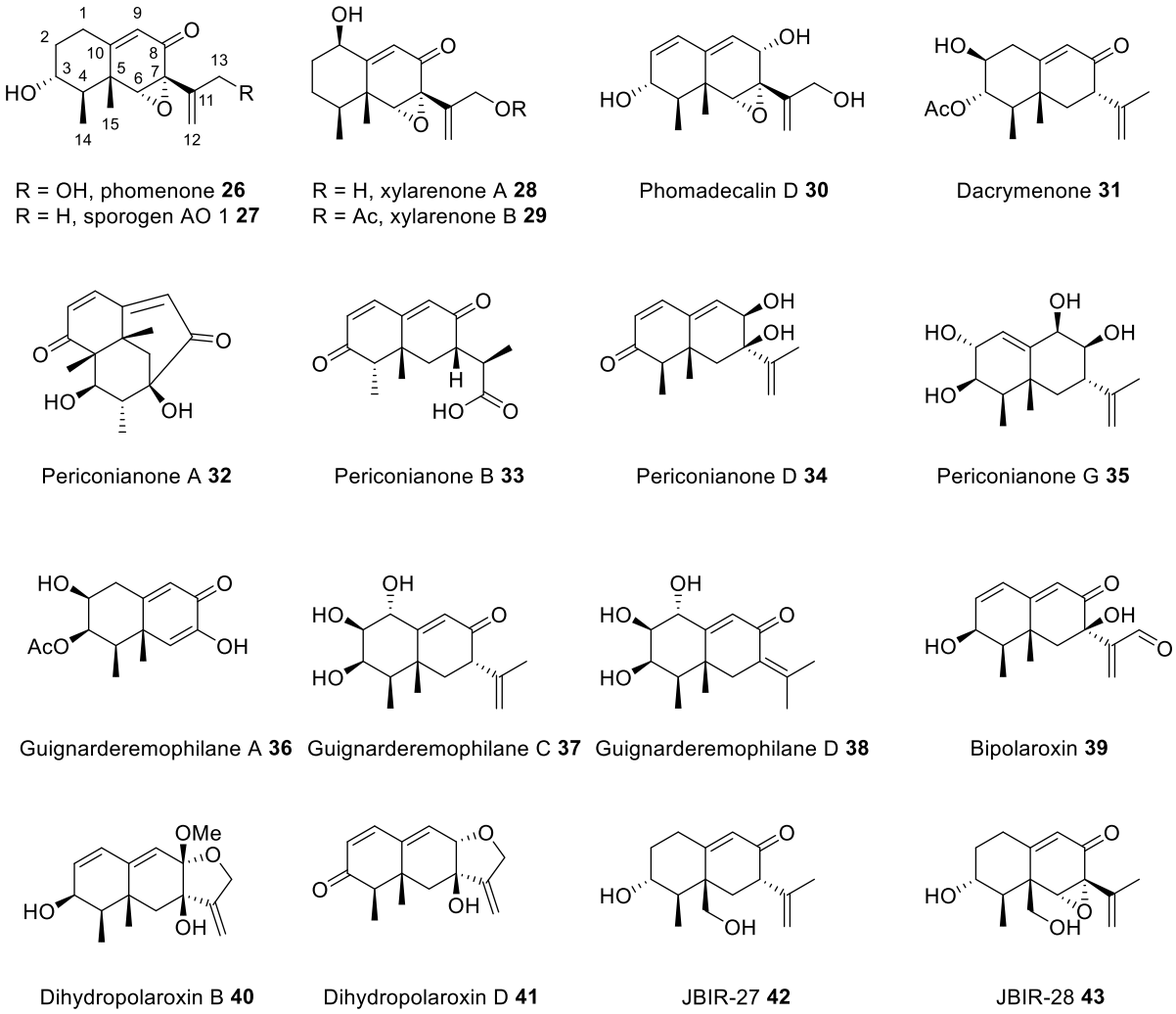


Figure 1.1.5 Structures of eremophilane-type derivatives

Periconianone D **34** and G **35** were isolated from the endophytic fungus *Periconia sp*. F-31.[89] Compound **34** possess two *cis*-configured hydroxyl groups at C-7 and C-8. Compound **35** exhibits high oxidation at various positions on its skeleton, specifically at C-2, C-3, C-8 and C-9. The hydroxyl groups located at C-2 and C-3 exhibit a *trans*-configuration, whereas the hydroxyl groups at carbon C-8 and C-9 exhibit a *cis*-configuration. This observation implies that there may be a distinct involvement of specific oxidations with unique enzymes in the biosynthesis of these compounds. Both compounds displayed inhibitory effects on the production of NO induced by lipopolysaccharide in BV-2 cells.[89]

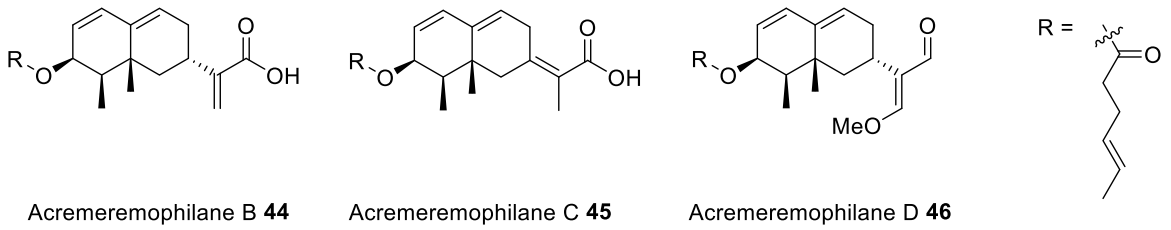
Guignarderemophilanes A **36,** C **37** and D **38** were isolated from the plant-devired fungus *Guignardia mangiferae*.[90] Compound **36** possesses an extra olefinic double bond located at C-6 and C-7. The isopropenyl group at carbon-7 was substituted by a hydroxyl group. Oxidations were observed at positions C-1, C-2, and C-3 in the structure of **37** and **38**. It is likely that an isomerization occurred within the C-7, C-11, and C-12 bonds, resulting in the spontaneous formation of **37** from **38.**[91] Both **37** and **38** possess anti-inflammatory properties in neural environments. They were also found to inhibit the production of NO induced by lipopolysaccharide (LPS) in BV2 cells, with IC_{50} values of 6.4 and 4.2

μM, respectively.[90] The first-ever total syntheses of **37** and **38** have been achieved, starting from hydroxyl carvone and involving 15 and 14 steps, respectively.[92] Identifying the specific enzymes responsible for modifying the skeletons would possibly enable a relatively straightforward synthesis of **37** and **38** using synthetic biological methods, reducing number of chemical synthesis steps.

Bipolaroxin **39** was isolated from the pathogenic fungus *Bipolaris cynodontis* and characterized by Sugawara et al in 1985.[93] Interestingly, **39** and **34** both exhibit oxidation at C-7, and compound **39** displays an aldehyde at C-13, none of which were present in the previously listed compounds. **39** exhibited high levels of phytotoxicity and selectivity in the assay system, as evidenced by the response of diverse crop and weed species. At a concentration of 3.8 mM, it elicited symptoms across 37 distinct plant species at the application level.[93,94]

Dihydrobipolaroxin B **40** and D **41** were isolated from the marine-derived fungus *Aspergillus sp*. SCSIOW2.[95] Compound **40** possesses an acetal moiety at C-8, which consists of two distinct R groups: one being a methoxy group, and the other forming a tetrahydrofuran ring with C-13. Similar to **40**, but without the methoxy group, **41** has a tetrahydrofuran ring and a C-8 acetal unit. They both showed some NO-inhibitory activity, although it was very low and neither was cytotoxic.[95]

JBIR-27 **42** and JBIR-28 **43**, along with sporogen-AO1 **27** and phomenone **26** were extracted from the culture medium of *Penicillium sp* SS080624SCf1.[96] Both **42** and **43** have a rare hydroxyl group at C-15 within the skeletal structure. Compounds **26**, **27**, and **43** demonstrated cytotoxic properties against HeLa cells, as evidenced by their respective IC_{50} values of 8.3, 19, and 92 μM. In contrast, the administration of **42** at a concentration of 80 μM did not result in any cytotoxic effects.[96]



Acremeremophilane B **44** Acremeremophilane C **45** Acremeremophilane D **46**

Figure 1.1.6 Examples of eremophilane-type derivatives containing a sidechain at C-3

Acremeremophilanes B **44**, C **45**, and D **46** were produced by *Acremonium sp*, a microorganism isolated from deep-sea sediments.[97] They all have the same backbone moiety, and also include an additional moiety at the C-3 that belongs to a 4'-hexenoic acid. Both **44** and **45** possess a carboxylic acid moiety, whereas **46** lacks the methine group at C-7, which is substituted by an olefinic isomer between C7-C11. Compound **46** features an olefinic methine moiety located at C-7, accompanied by a formyl substituent at C-13 and a methoxy substituent at C-12. They all demonstrated inhibitory properties against the production NO, with IC_{50} values ranging from 8 to 45 mM.[97]

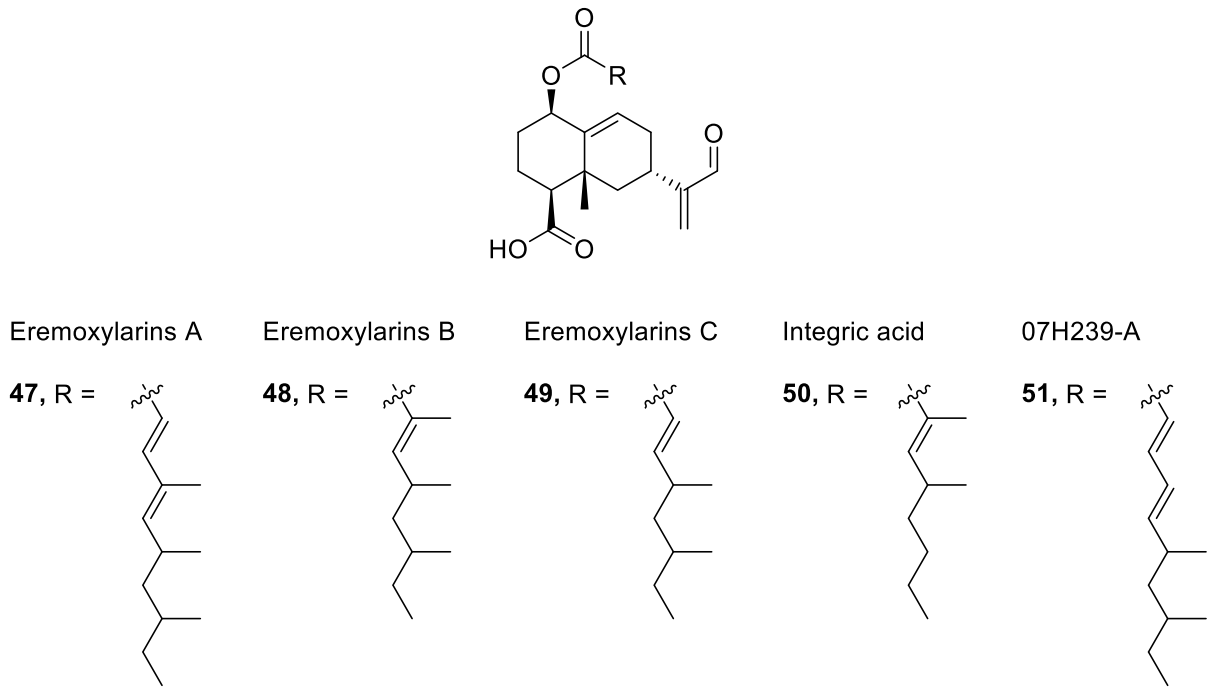


Figure 1.1.7 Examples of eremophilane-type derivatives containing a sidechain at C-1

Eremoxylarins A **47**, B **48**, and C **49**, integric acid **50**, and 07H239-A **51** are compounds isolated from Xylaria species.[98–102] They share a common decalin core, but their ester side chains at C-1 vary. This core features a carboxylic acid group at C-14 and a formyl group at C-13. Compounds **47** and **51** have nine-carbon chains with varying methyl group distributions, while **48**, **49**, and **50** possess chains with the same carbon count but different methyl group placements along the chain. These compounds display diverse biological activities. Compounds **47** and **48** possess antibacterial properties against *Staphylococcus aureus* and *Pseudomonas aeruginosa* and can restore growth activity in yeast mutants by engaging Ca^{2+} signal transduction.[98, 99] Compounds **49** and **51** exhibit cytotoxicity against cancer cell lines.[99,100] Compound **50** suppresses HIV-1 integrase, the enzyme responsible for viral integration into the host genome.[101,102]

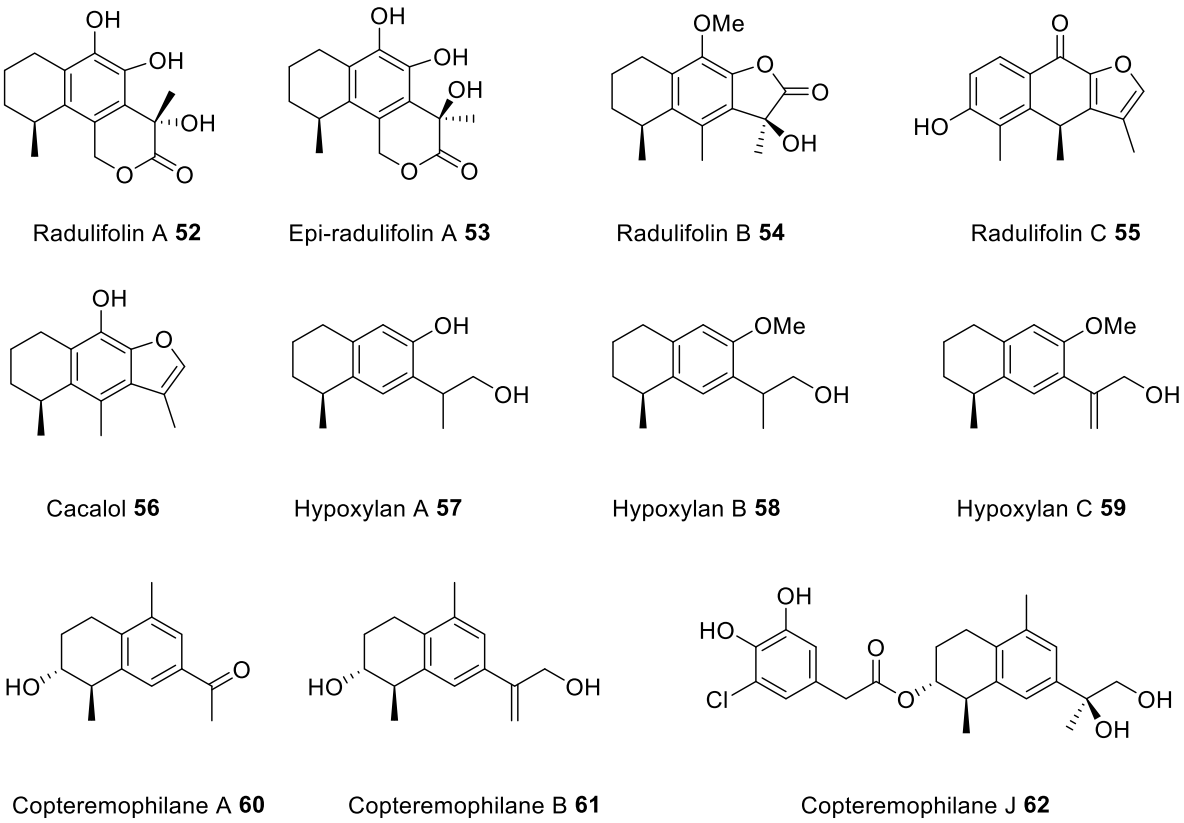


Figure 1.1.8 Examples of eremophilane-type derivatives containing a benzene ring.

A class of rearranged eremophilanes was found in the plant family *Asteraceae*. Compounds **52, 53, 54**, and **55** were isolated from the roots of *Psacalium radulifolium*.[103] These natural products were not previously identified in ascomycetes until Stadler et al. discovered compounds **57**, **58**, and **59** in a culture of *Hypoxylon rickii*, a xylariaceous ascomycete from Martinique.[104] These compounds were named Hypoxylan A **57**, B **58**, and C **59**.[104] They were believed to be sesquiterpenes with a unique 14-noreudesmane core structure, sharing structural similarities with cacalol **56**.[105] However, the absence of isotopic labeling experiments leaves these claims lacking definitive evidence.

Copteremophilanes A **60**, B **61**, and J **62**, were discovered in a marine sponge-associated *Penicillium copticola* fungus.[106] The formation of Hypoxylans **57** - **59** likely involves cyclization of (+)-aristolochene, followed by oxidative steps, including hydroxyl group insertion, right aromatization, and methyl group loss. However, in the formation of **60**, **61**, and the right part of **62**, methyl migration from C-10 to C-9, instead of from C-10 to C-5 during (+)-aristolochene cyclization, is expected. The lack of prior biosynthetic research on these compounds makes their biosynthesis an intriguing area for exploration.

1.2 Investigation of Natural Product Biosynthesis

1.2.1 Isotopic Labelling

Our understanding of natural product biosynthesis has significantly expanded through a rich history of isotopic labeling experiments, which unveil the origins of natural products and identify specific functional components.[107] Since the early days, exemplified by Frederick Soddy's Nobel Prize-winning discovery of *isotopes* in 1921, to the contemporary methods involving mass spectrometry for assessing incorporated labels and comprehensive NMR spectroscopy for confirming their precise locations, these experiments have greatly contributed to the advancement of our knowledge.[107,108]

Isotopic labeling experiments play a crucial role in investigating polyketide biosynthesis. In cases where polyketides exclusively originate from acetyl- and malonyl-CoA, the polyketide skeleton can be labeled by employing ^{13}C-labeled acetate.[107] If the labeled acetate bond remains intact throughout the biosynthetic pathway, the incorporation of double-labeled acetate [1, 2- $^{13}C_2$] results in the observation of ^{13}C-^{13}C coupling in the ^{13}C NMR spectra.[109] To elucidate the origins of hydrogen and oxygen, incorporation can also be achieved using atmospheric ^{18}O-labeled oxygen or doubly labeled [^{18}O, 1-^{13}C]-acetate.[109]

A famous case of isotopic labelings employed in biosynthesis investigation is squalestatin S1 **24** (Figure 1.2.1).[110] Feeding experiments using ^{13}C-labeled acetate revealed a hexaketide main chain and a tetraketide side chain.[110] The incorporation of ^{18}O ([1-^{13}C, $^{18}O_2$] acetate or $^{18}O_2$ as a precursor) into squalestatin S1 indicated that oxygen atoms (in green) at C-1, C-3, C-5, C-6, C-7, and C-12 are introduced by molecular oxygen rather than acetate, suggesting the involvement of several oxidation steps in the biosynthetic pathway. In contrast, the C-24 and C-34 carbonyl oxygen atoms are derived from acetate.[110]

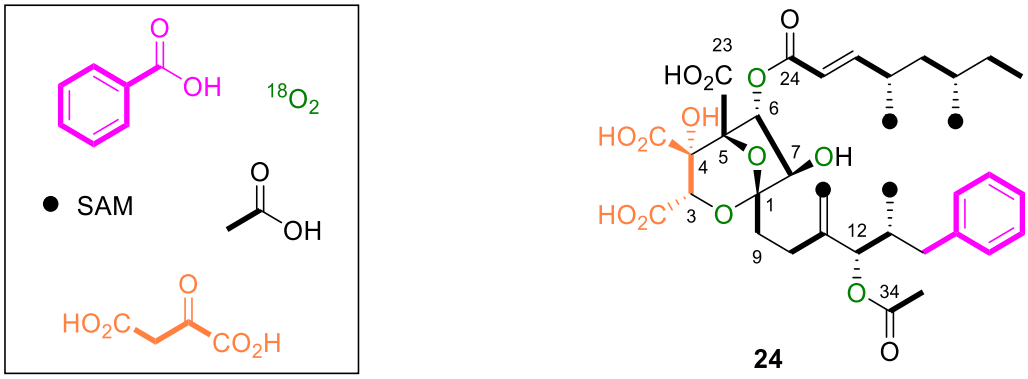
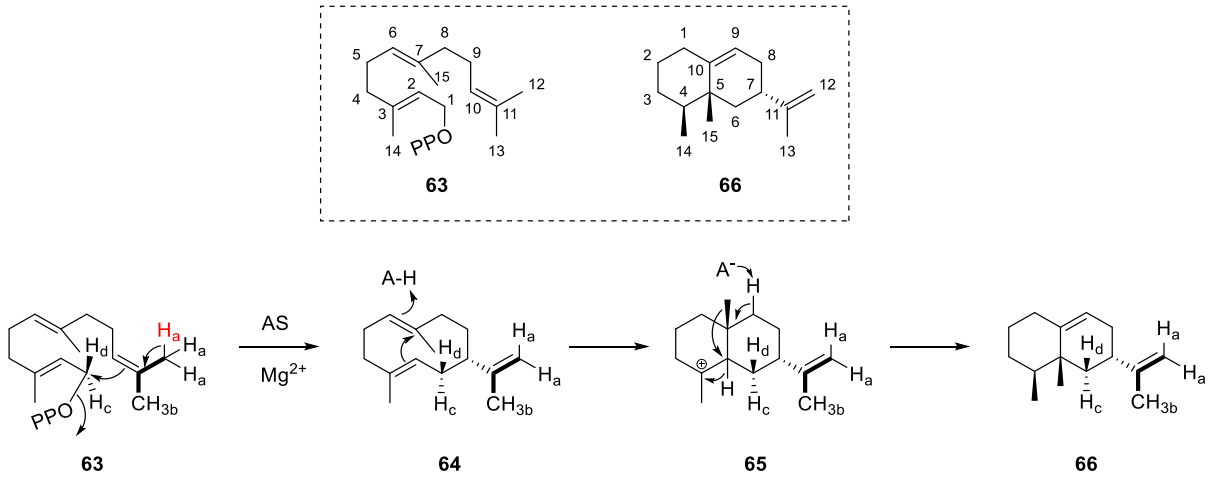


Figure 1.2.1 The labelling patterns of squalestatin S1 **24.**

Isotopic labeling has been employed in the biosynthetic investigation of terpenes as well.[111] A classic example is aritolochene **66**, and the cyclization process of aritolochene has been thoroughly examined

using isotopic labeling methods.[112] The conversion of the acyclic precursor farnesyl diphosphate (FPP) into aristolochene is catalyzed by aristolochene synthase (AS).[113]

Isotopic labeling experiments supporting a cyclization mechanism indicate FPP **63** undergoes ionization with an electrophilic attack on the π bond of C10-C11, resulting in the inversion of configuration at C-1 and the formation of the neutral hydrocarbon intermediate *S*-(-)-germacrene A **64**.[114,115] Protonation at C-6 is succeeded by the electrophilic attack of the resulting C-7 carbocation on the C2-C3 π bond, giving rise to the formation of the eudesmane cation **65**.[114,115] Subsequent steps involving methyl migration, hydride transfer, and stereoselective deprotonation culminate in the synthesis of (+)-aristolochene **66**.[114,115]



Scheme 1.2.1 Isotopic labelling for C-11, C-12, C-13 and protons of them. The C-1 of FPP **63** is the C-6 of aristolochene **66**

FPP **63** was selectively labeled at positions C-11, C-12, and C-13, resulting in the generation of [11, 12-$^{13}C_2$] and [11, 13-$^{13}C_2$]. Following this, the aristolochene synthase of *A. terreus* was employed in the incubation of FPP **63**, leading to the production of deuterated aristolochene **66**. The ^{13}C NMR spectrum revealed the expected pair of intensified and coupled doublets.[116] This finding implies that the cyclization of FPP to aristolochene did not take place at positions C-11, C-12, and C-13.[116]

Regardless of whether FPP **63** was labeled as 1*R*-[1-^{2}H]-FPP (H_d = D, H_a = H_b = H_c = H) through deuterium or labeled as 1*S*-[1-^{2}H]-FPP (H_c = D, H_a = H_b = H_d = H), both resulting products exhibited peaks exclusively composed of deuterium. These deuterium peaks corresponded to those detected in aristolochene at H-6_{eq} and H-6_{ax}, respectively.[116] The findings strongly suggest that the cyclization of FPP **63** to aristolochene **66** involves the inversion of configuration at the C-1 position of FPP, as indicated by ^{2}H NMR spectroscopy.[116]

It was determined that deprotonation occurs at the *cis*-(C-12) methyl group of FPP **63** through the incubation of [12, 12, 12-$^{2}H_3$]-FPP (H_a = D; H_b = H_c = H_d = H) with aristolochene synthase. This

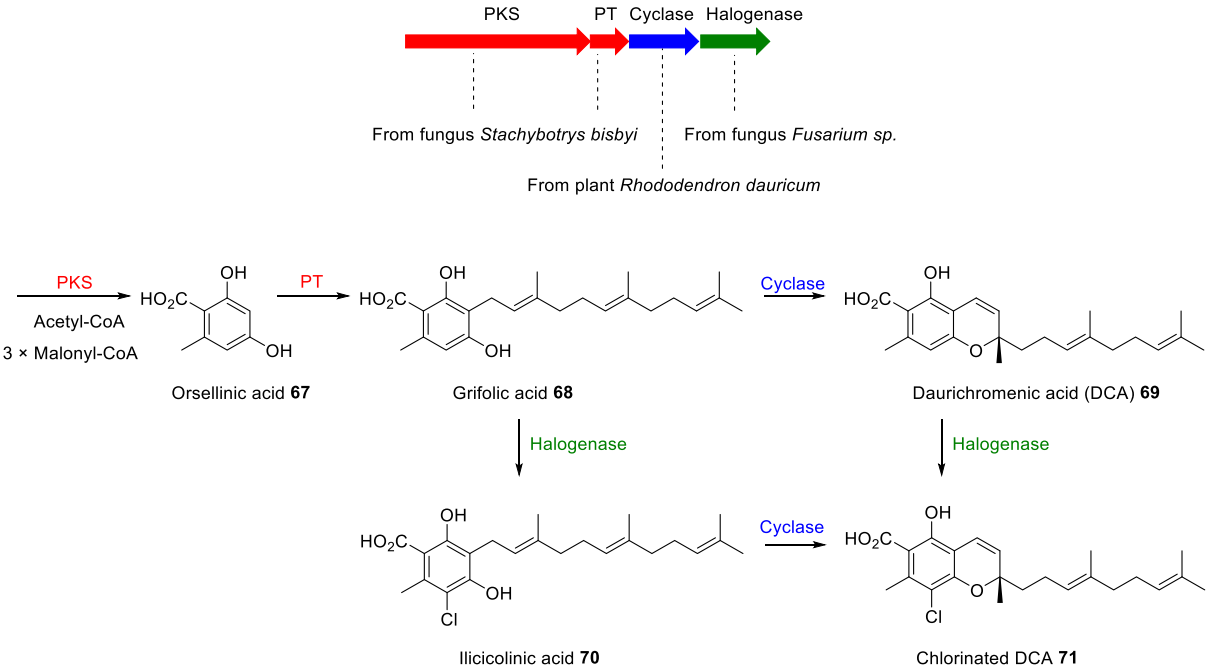
conclusion finds support in the ^{2}H-NMR observation of deuterium at the C-12 methylene in the resulting product.[117] Further validation is provided by the transformation of [13, 13, 13-$^{2}H_3$]-FPP (H_b = D, H_a = H_c = H_d = H) into [13, 13, 13-$^{2}H_3$]-aristolochene.[117]

1.2.2 Heterologous Expression in Filamentous Fungi

1.2.2.1 One Host Many Compounds

Heterologous biosynthesis enables access to compounds encoded by biosynthetic pathways identified through genome mining, with numerous reported examples of total biosynthesis of natural products and the discovery of novel compounds.[118] Achieving the reprogramming of natural biosynthetic pathways, involving mixing and matching genes from established biosynthetic clusters to produce unnatural designer analogs, presents a significant advantage in constructing pharmaceutical screening libraries and can be easily executed in a single heterologous host.[118–120]

For example, through heterologous expression, daurichromenic acid **69** was reconstructed by incorporating orsellic acid synthases (PKS) and a prenyltransferase (PT) from the fungus *Stachybotrys bisbyi*, along with a daurichromenic acid synthase (Cyclase) from the plant *Rhododendron dauricum*, into *Aspergillus oryzae* NSAR1 (Scheme 1.2.2).[121] Additional introduction of a halogenase from the fungus *Fusarium sp.* resulted in the biosynthesis of ilicicolinic acid **70** and (+) - 5 - chloro daurichromenic acid **71** (Scheme 1.2.2).[121]

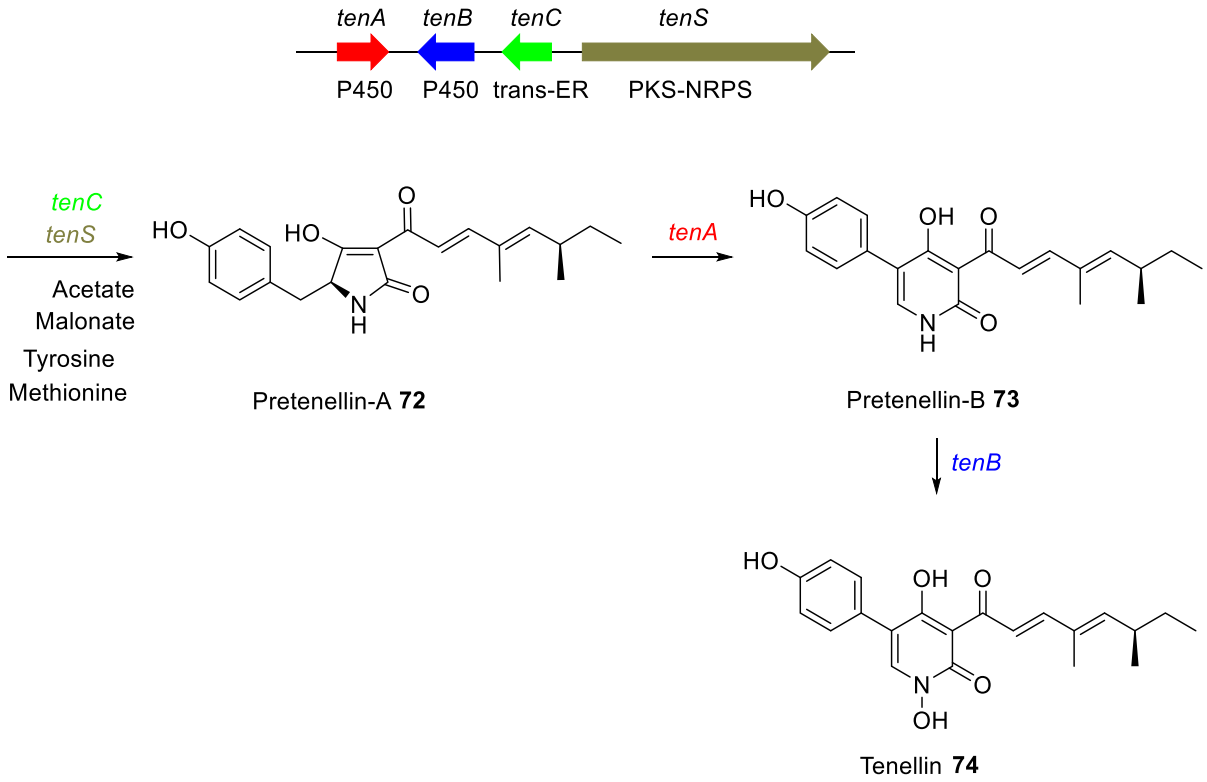


Scheme 1.2.2 Engineered biosynthesis of daurichromenic acid **69** and its derivatives.[121]

1.2.2.2 High Productivity

Achieving higher titers, reflecting enhanced productivity of pharmaceutically important compounds, is a sought-after goal often pursued through strain improvement and metabolic engineering.[122] To attain sufficient titers for full structural characterization or commercial use, the host-produced compounds need to be optimized, necessitating the incorporation of strong promoters in heterologous expression of biosynthetic gene clusters.[123–128]

For example, the tenellin BGC represents the pioneering case of heterologous expression encompassing an entire fungal secondary metabolism gene cluster (Scheme 1.2.3).[129] Using *A. oryzae* to express the tenellin BGC, with the robust P_{amyB}, resulted in the production of tenellin **74** at a titer of 243.0 mg/L, a fivefold increase over the wild-type organism (Scheme 1.2.3).[129] This methodology addresses typical challenges encountered in unraveling fungal secondary metabolite pathways, including difficulties posed by fungal strains that exhibit resistance to molecular techniques such as gene knockout.[129]



Scheme 1.2.3 Biosynthesis of tenellin **74.**[129]

1.2.2.3 BGC Gene Assembly

Another critical aspect to consider is the challenge posed by the transformation of multiple genes and massive size of certain core biosynthetic enzymes during the heterologous expression of BGCs. In some cases, individual gene from the BGC were cloned separately and then sequentially transferred step by step into the heterologous host, resulting in their random integration into the genome (Figure 1.2.2A).[120] For example, the assembly and expression of genes (*tenS, tenC, tenA, and tenB*) from the tenellin BGC

in *A. oryzae* enables the generation of intermediates and facilitates the elucidation of the gene expression order in the pathway.[129]

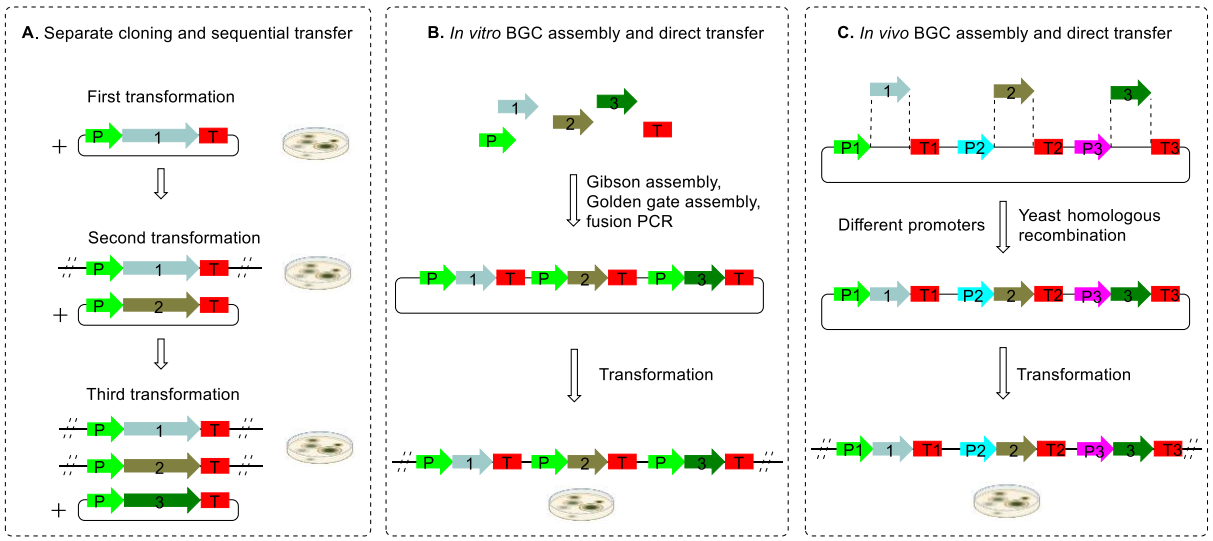


Figure 1.2.2 Commonly employed strategies for the assembly and transfer of BGCs in fungal heterologous expression; P represents promoters, T represents terminators.

An alternative method involves constructing multiple genes or all genes within a single cassette or several cassettes and subsequently transferring them collectively to the host. This approach requires *in vitro* assembly techniques like Gibson assembly, Golden Gate assembly, fusion-PCR, or *in vivo* assembly methods such as yeast homologous recombination (Figure 1.2.2 A and B).[130] Yeast transformation-associated recombination has been identified as an efficient and robust method for the assembly of multiple genes within large intact fungal BGCs.[130] This method entails cloning large BGCs through the combination of multiple DNA fragments, featuring homologous arms between each other, along with linearized vectors incorporating different promoters.[130] Following, these DNA fragments undergo transformation into *S. cerevisiae*, conducting assembly through homologous recombination. This transformation method is simple, chemical-mediated, and cost-effective.[130] For example, Li et al. assessed this method for assembling 14 fungal BGCs ranging in size from 7 kb to 52 kb, showcasing high efficiency with an average positive rate exceeding 80%.[131]

1.3 Overall Aims

This thesis will focus on engineering the biosynthesis of fungal natural products from three different levels: host-level engineering, which involves modifying the heterologous host to enhance the productivity; enzyme-level engineering, which exploits the product diversity of individual enzymes by mutasynthesis; and pathway-level engineering, which combines biosynthetic genes from multiple pathways in a single heterologous host.

First, we will apply host-level engineering to enhance the productivity of the heterologous host for the expression of squalestatin tetraketide synthase. This involves co-expressing the wild-type SQTKS with predicted hydrolysis enzymes to increase tetraketide production in *A. oryzae* NSAR1.

Second, engineering individual enzymes is a more rational approach to create new natural product derivatives. Mutasynthesis is the simplest method to engineer natural product analogues. In this study, we will introduce mutations into the *cis*-ER domain of the complete SQTKS, and then expressed them in the optimised *Aspergillus oryzae* by heterologous expression. The mutations were derived from the isolated ER mutants of SQTKS, which were analysed using a combination of computational modelling and experimental assays with various mimic substrates by previous co-workers.

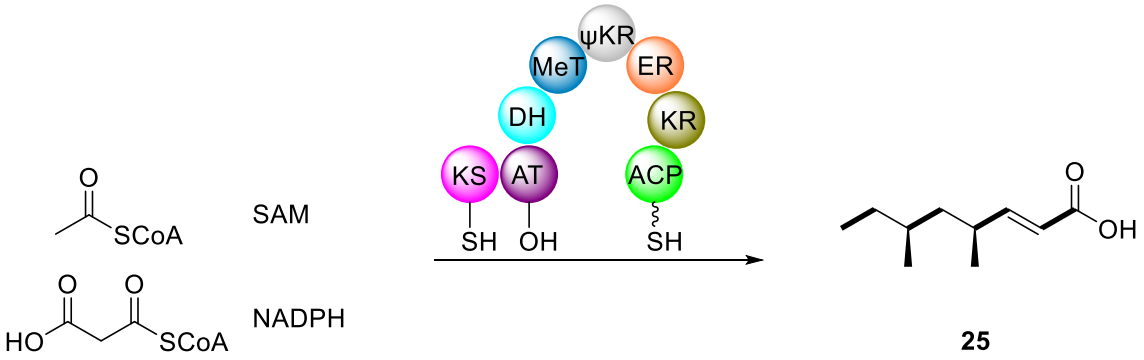
Third, we will conduct pathway-level engineering, which entails combining genome-mined genes from multiple BGCs in a single heterologous host *via* a single process. This can either redirect the original pathway, which is mostly exclusively linear, to produce related natural products, or extend the original pathway to render it more web-like, which could feasibly generate more congeners. This part is the main contribution of this thesis, and it comprises two projects. First, we will mix and match tailoring enzymes from genome-mined and related biosynthetic pathways to rapidly access differently oxidised aristolochenes. Second, we will elucidate the pathway to the multiforisins and engineer the intermediates by adding genes from other BGCs from distinct organisms to redirect the pathway towards islandic acid **144** and allantopyrone A **145** intermediates.

2. Reprogramming of SQTKS

2.1 Introduction

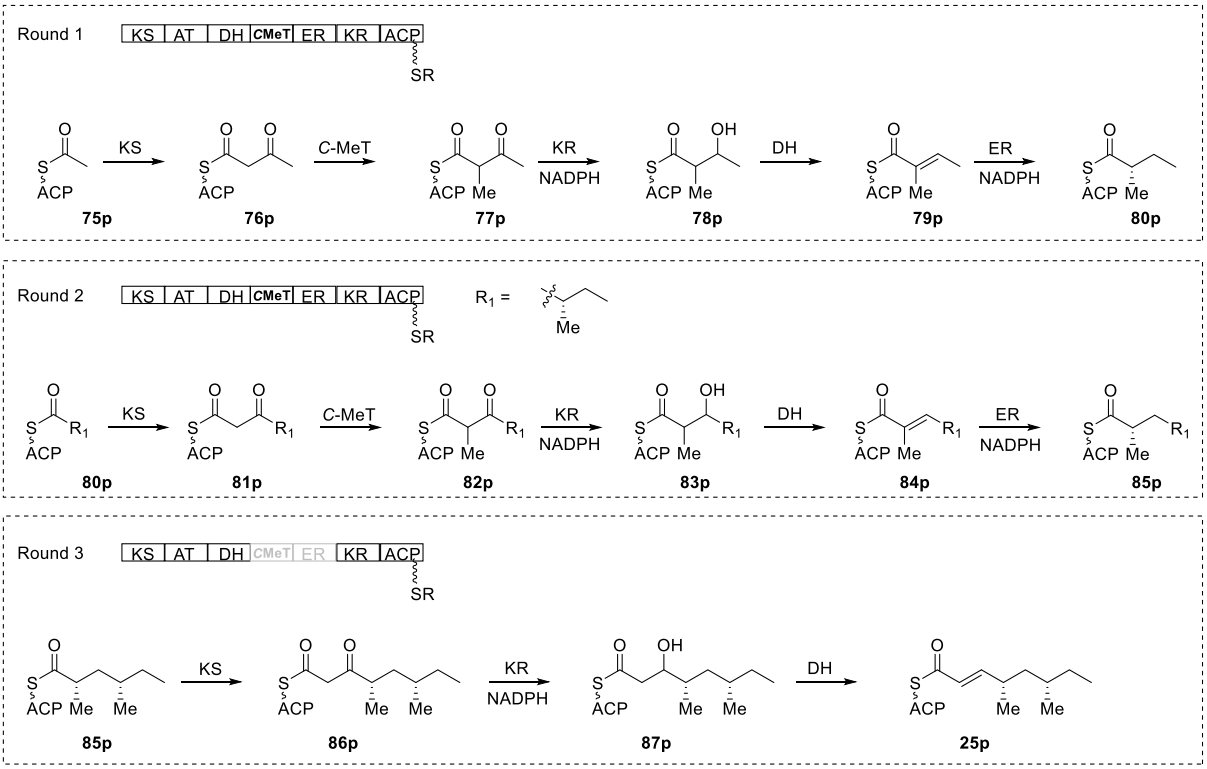
2.1.1 Programming of Squalestatin Tetraketide Synthase

Squalestatin tetraketide synthase (SQTKS), is type I fungal polyketide synthase, recognized for its iterative and highly reducing operational mode (hrPKS). This enzyme consists of seven distinct catalytic domains (Scheme 2.1.1**)**, each playing a specialized role during the synthesis of a tetraketide chain **25**, which is chain B of squalestatin S1**24**.[132–134] These domains include ketosynthase (KS), acyltransferase (AT), dehydratase (DH), *C*-methyltransferase (*C*-MeT), enoyl reductase (ER), ketoreductase (KR), and acyl carrier protein (ACP). The KR domain is further devided into two distinct segments: the catalytic KR, responsible for catalyzing chemical transformations, and the structural ψKR, which fulfills the role of providing structural support for the overall functionality of the KR.[135]



Scheme 2.1.1 Doamins of SQTKS and structure of squalestatin tetraketide **25.**

SQTKS catalyses a multi-step programming, comprising three rounds of chain extension and modification (Scheme 2.1.2). In the first round, all functional domains are active. First, the AT domain transfers an acetate starter unit onto the ACP domain, resulting in the formation of **75p**, and also loads the malonyl extender unit onto the ACP domain. Once these units are loaded, the KS domain catalyzes the condensation of the starter and extender units on the ACP, resulting in the formation of a β-keto thiolester bond with the ACP **76p**. Following this step, the *C*-MeT domain introduces a methylation on the α-carbon, producing **77p**. Next, the KR domain intervenes to convert the β-ketone group into a β-OH, forming **78p**. Afterward, the DH domain carries out the removal of water, leading to the creation of an α, β-unsaturated thiolester **79p**. This intermediate is then further reduced by the ER domain, culminating in the production of a diketide thiolester **80p**. In the second round, all domains are active again, with the diketide **80p** serving as the initial unit to start this phase. The malonyl extender unit is loaded and coupled to form a triketide-ACP by KS domain to form **81p**. Then **81p** is modified by the *C*-MeT domain, resulting in **82p** through methylation. Next, the KR domain reduces **82p** to create **83p**. It is then subjected to dehydration by the DH domain, leading to the formation of **84p**. Finally, the ER domain concludes the process by reducing it further to yield the fully saturated triketide product **85p**.



Scheme 2.1.2 Biosynthesis of squalestatin tetraketide **25**

In the final round, the *C*-MeT and ER are not active. The triketide-ACP **85p** undergoes condensation with a third malonyl extender unit, giving the formation of a tetraketide-ACP **86p**. In this step, the *C*-MeT domain is bypassed, while both the KR and DH domains remain active. Their actions yield an α, β-unsaturated thiolester **25p**. The ER domain remains unengaged and does not participate in the reaction. This sequence of events results in the final product **25**, which maintains its α, β-unsaturation. It is unclear which process leads to the hydrolysis and release of compound **25** as there is no TE domain. It is assumed that hydrolysis is either spontaneous or perhaps catalyzed by a second enzyme.[136]

Prior studies involving the heterologous expression of SQTKS in *A. oryzae* have revealed that the fully synthesized polyketide is released from the polyketide synthase, resulting in the formation and identification of the 4*S*, 6*S*, *E*-α, β-unsaturated acid squalestatin tetraketide **25**.[137] The gene encoding SQTKS was cloned from the producer strain *Phoma sp*. C2932 and then was introduced into *A. oryzae* strain M-2-3 using a plasmid under the control of P_{amyB}, incorporating a selection marker *argB*.[137] The yield of compound **25** achieved 0.84 mg/L when cultured in the CDS medium. Attempts were made to express the full-length gene encoding SQTKS in both *E. coli* and *S. cerevisiae*. While the expected protein was successfully expressed in *E. coli*, purification proved to be challenging.[138] Consequently, the crystallization and structural investigation of SQTKS has remained elusive.

2.1.2 Engineering the Programming of ER Domain of SQTKS

Previous coworkers have established that the isolated ER domain of SQTKS displays a wide substrate selectivity, encompassing various di- and triketides. However, when the chain length and methylation pattern of substrates increase, the enzyme's activity decreases.[139] Interestingly, tetraketide product **25** is an effective inhibitor of the isolated ER showing that it can enter the active site, but it can not react.[139] This result lends support to the hypothesis that within the complete SQTKS system, the programming for chain construction appears to halt due to the ER domain's inability to reduce tetraketide **25**.

Hence, Piech and coworkers hypothesized that altering the selectivity of the isolated ER held the potential to reprogram SQTKS. Their first step in this endeavor involved mutating the isolated ER to accept substrate **25** and extending its capacity to handle longer substrates, marking the initial step towards reprogramming the entire SQTKS.[140]

Structural models of the isolated ER were constructed and modified using tools such as Swiss-Model [141], AutoDock [142], and YASARA [143]. The resulting model exhibits three primary structural features (Figure 2.1.1). First, a prominent globular domain is established between L1908 and I2001, playing a pivotal role in interacting with the acyl-pantetheine substrate. The second salient feature encompasses the central sequence spanning from V2002 to V2144, forming the cofactor-binding domain and featuring a Rossmann-fold motif. Finally, the C-terminal sequence from D2145 to P2208 acts as a connecting link between the cofactor and substrate binding domains and contributes to the formation of a cap-like structure situated above the active site.

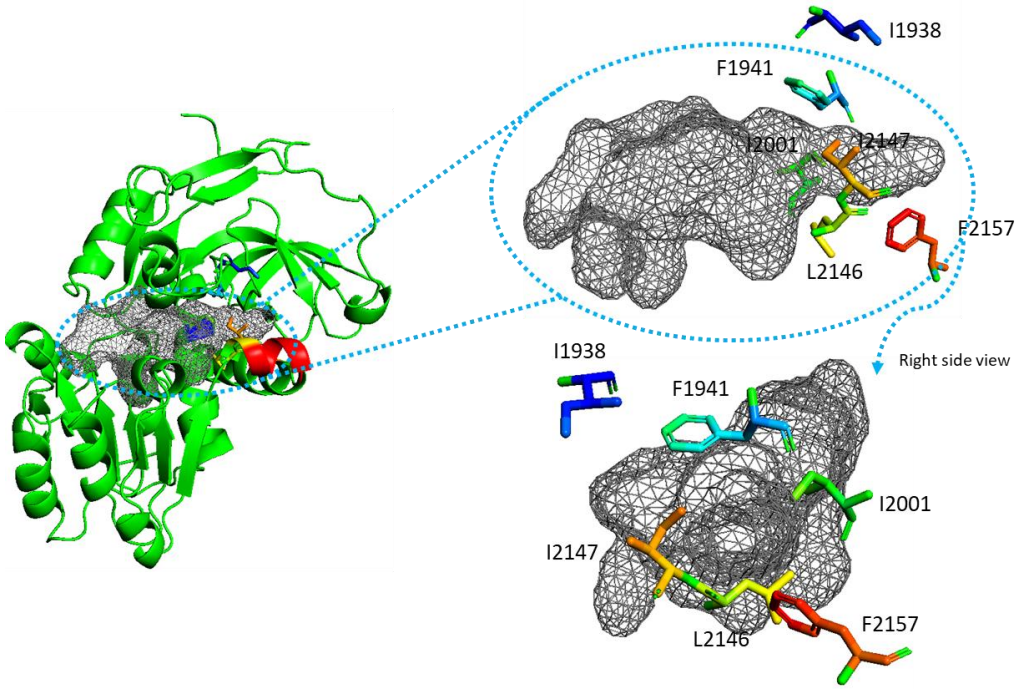


Figure 2.1.1 A model for the ER domain, the mutations were conducted in ER domain including F1941, I2147, F2157, etc around the grey pocket.

Mutagenesis was selectively conducted on specific amino acids, and *in vitro* kinetic analyses were performed on both the wild-type and mutant forms of ER using synthetic substrates.[140] The mutations encompassed F1941A, F2157A, F1941A/F2157A, I2147A/F2157V, and F1941A/I2147A/F2157V (Figure 2.1.1). Mimic compounds, including diketide (**88p**), triketides (**89p**, **90p, 91p**), tetraketides (**92p, 93p, 25p**), pentaketides (**94p, 95p**), heptaketide (**96p**), and cinnamoyl pantetheine (**97p**), were synthesized and assessed in conjunction with both the WT and mutant ER proteins. The WT protein demonstrated kinetic parameters closely mirroring those previously reported.[140,144] The mutant ER proteins exhibited a variety of significant changes in their behavior.[140]

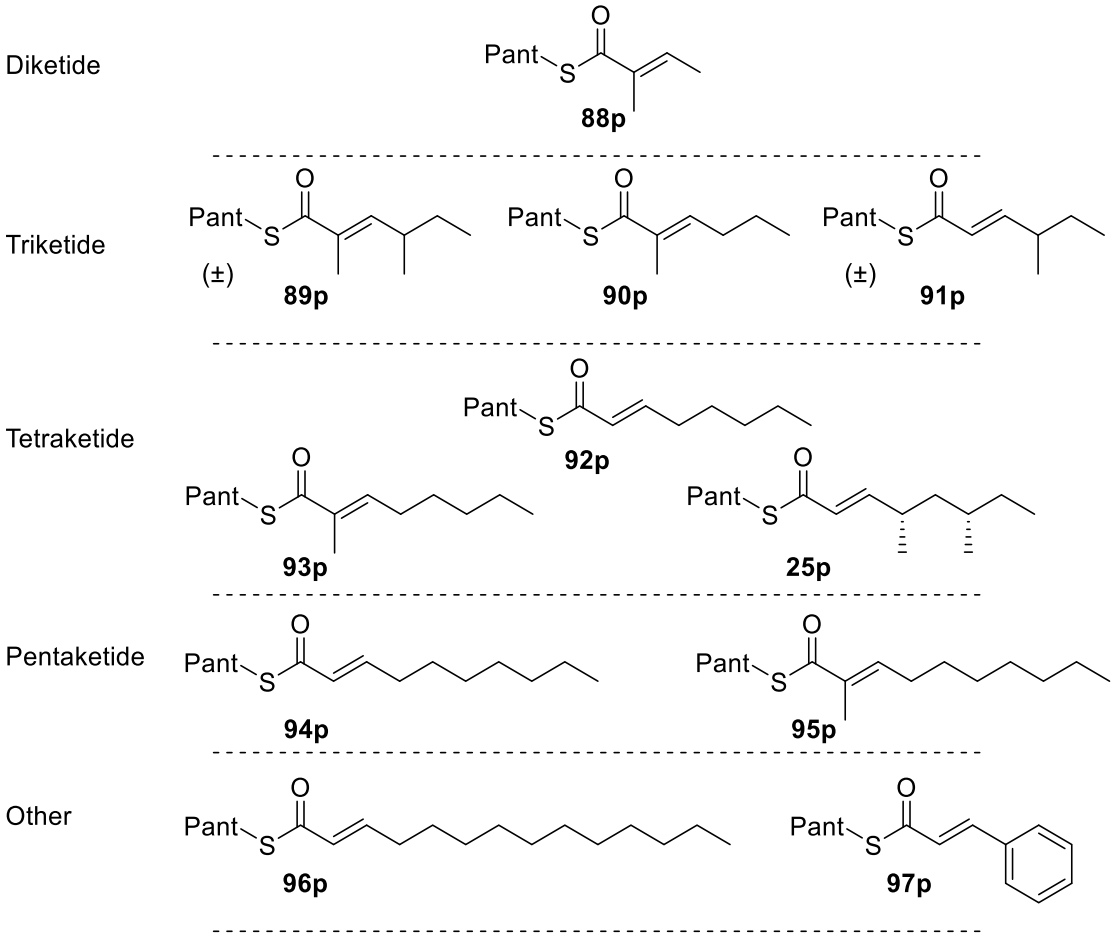


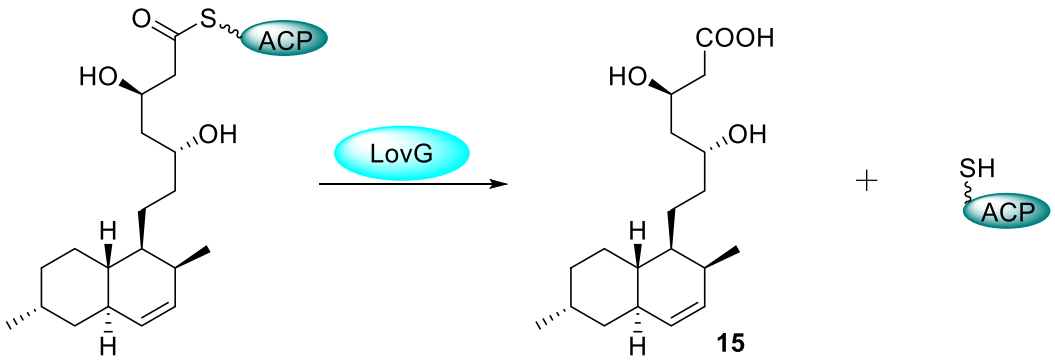
Figure 2.1.2 Structures of the substrates for the *in vitro* assay with WT and mutant ER proteins.

The F1941A mutant typically exhibited activity levels that were either comparable to or lower than those of WT across all substrates. It displayed some activity with 4*S*, 6*S*-tetraketide **25p**, a substrate that the WT enzyme cannot process. This finding suggests that the docking experiment with the 4*S*, 6*S*-tetraketide structure **25p** resulted in the attainment of a productive s-*trans* conformation within the active site. The F2157A mutation accommodated shorter substrates, and it significantly enhanced its activity with pentaketides **94p** and **95p**. However, it was unable to catalyze the reaction with 4*S*, 6*S*-tetraketide **25p**, in alignment with the result of *in silico* experiments.

The double mutation F1941A/F2157A combines the effects of both individual mutations of F1941A and F2157A. The double mutation I1247A/F2157V was predicted to decrease the active site volume, which is particularly noticeable with larger substrates but not with 4*S*, 6*S*-tetraketide **25p**. The triple mutant F1941A/I2147A/F2157V appears to synergize the effects of other mutations. Heptaketide **96p** and cinnamoyl pantetheine **97p** were also tested but displayed inactivity in the presence of all ER variants.

2.1.3 Polyketide Off-Loading by Trans-Acting Hydrolases

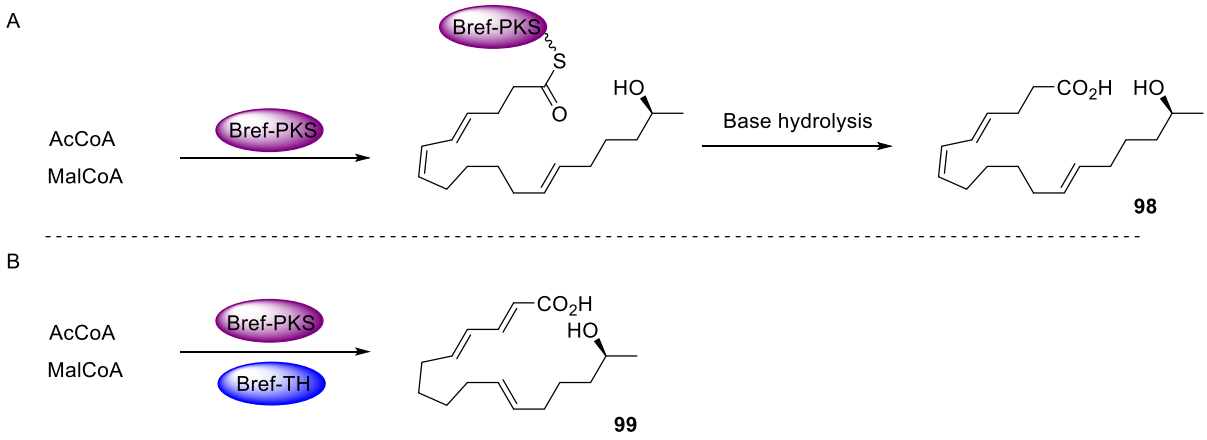
Many hr-PKSs lack *cis* offloading domains and instead utilize a variety of *trans*-acting hydrolytic enzymes, often from the α-β hydrolase family. These partner enzymes are pivotal in stopping chain elongation and influencing the final length of product. Certain instances have demonstrated the ability of *trans*-acting hydrolases to facilitate the release of polyketides in hr-PKS biosynthesis. A well-known example is LovG, a member of the serine hydrolase esterase-lipase family, which was determined to plays an important role in releasing dihydromonacolin L acid **12** from LNKS.[145] The function of LovG was investigated by employing genetic disruption, and the lovG mutants revealed a significant reduction in lovastatin acid **15** levels, exceeding 95% when compared to the wild-type strain.[145] The LovG protein was expressed and purified. When incubated with purified LovB and LovC, along with malonyl-CoA and the required cofactors, the production of **15** was observed, all without the need for base hydrolysis or a heterologous TE. Moreover, when lovG was coexpressed with lovB and lovC in yeast, it resulted in the elevated production of **15** at a titer of approximately 35 mg/L. These findings serve as strong evidence confirming LovG's natural role as the partner responsible for releasing **15** during the biosynthesis of lovastatin in *A. terreus*.



Scheme 2.1.3 Function of LovG

Another notable case is Bref-TH, a member of α, β-hydrolase_6 family known for its α-β hydrolase fold and functions.[146] Through reconstitution experiments, it was revealed that, when the purified Bref-TH enzyme was present, the predominant product of Bref-PKS was an acyclic polyketide **99**, matching the length of Brefeldin A. This product exhibited the anticipated β-reduction patterns necessary for subsequent conversion into Brefeldin A. In the absence of Bref-TH, Bref-PKS generated longer

polyketide products **98** after a chemical hydrolysis, underscoring the vital role of Bref-TH in regulating the chain length of the hrPKS.



Scheme 2.1.4 A, Function of Bref-PKS and the product after chemical hydrolysis; **B**, function of Bref-TH and the product.

2.2 Project Aims

The isolated ER mutants of SQTKS were investigated using a combination of *in silico* modeling and *in vitro* assays with a diverse array of mimic substrates. As a result, our forthcoming work will focus on incorporating these mutations into the complete SQTKS, followed by their expression in *Aspergillus oryzae* through heterologous expression.

Our initial plan involves co-expressing the wild-type SQTKS with the predicted hydrolysis enzymes to enhance trtraketide **25** production in *A. oryzae* NSAR1. Subsequently, we will introduce the SQTKS ER domain mutants into *A. oryzae* NSAR1 to assess the potential generation of new reprogrammed products with high yields under the optimized fermentation conditions.

2.3 Results

2.3.1 Bioinformatic Analyses

The gene cluster responsible for SQS1 was previously identified *via* genome sequencing of two SQS1-producing ascomycetes *Phoma sp*. C2932 and an unidentified fungus MF5453.[147] The SQS1 BGC in both MF5453 and C2932 exhibit a remarkably similar gene composition, displaying a substantial identity from 74% to 93% among proteins with corresponding predicted functions. Two potential hydrolases, Mfm8 and Mfm10, are present, and these enzymes are implicated in processes such as hexaketide chain release from the hexaketide synthase, or hydrolysis of the tetraketide chain release from its respective synthase. Either of these hydrolases has the potential to be involved in the PKS chain unloading process. In this project, aimed at releasing tetraketide, we will investigate the co-expression of *mfm8* and *mfm10* to enhance product release.

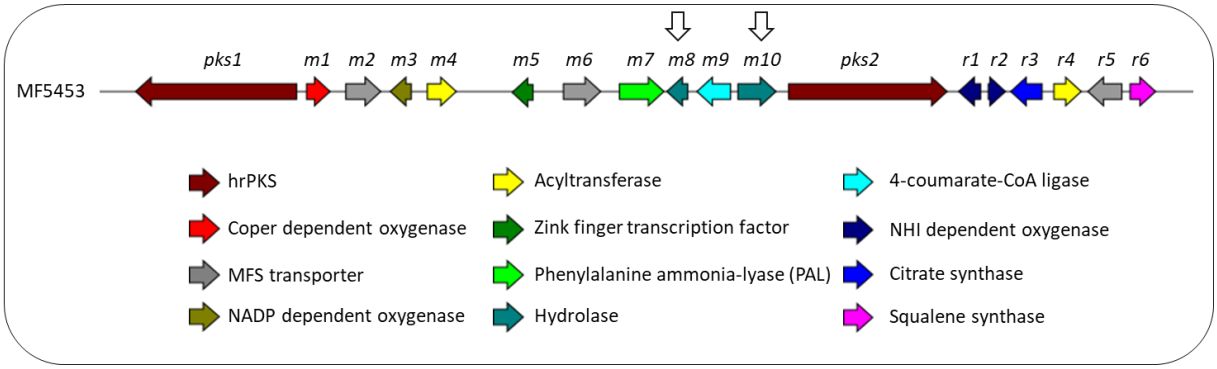


Figure 2.3.1 Gene cluster of SQS1 from unidentified strain *Phoma sp*. MF5453; Mfm8 and Mfm10 were pointed by arrows.

First, we conducted an analysis of the conserved domains within Mfm8. The protein sequence was subjected to a BLAST search, which yielded a specific hit for an acetyl esterase/lipase. The BLAST results indicated that Mfm8 is likely a carboxylesterase, belonging to the alpha/beta hydrolase family of proteins. It possesses the potential to catalyze the hydrolysis of substrates with varying chemical compositions through the utilization of a nucleophile-His-acid catalytic triad.

Table 2.3.1 The list of domain hits of Mfm8

Name	Accession	E-value	Description
Abhydrolase_3	pfam07859	2.45e-13	Alpha/beta hydrolase fold; this catalytic domain is found in a very wide range of enzymes.
Aes	COG0657	5.45e-13	Acetyl esterase/lipase; lipid transport and metabolism.
Esterase_lipase	cd00312	7.15e-07	Esterases and lipases (includes fungal lipases, cholinesterases, etc.) These enzymes act on carboxylic esters. The catalytic apparatus involves three residues: a serine, a glutamate or aspartate and a histidine.

Mfm8, Clz11, and LovG protein sequences were compared. The Mfm8 protein consists of 320 amino acids and shares an 80% sequence identity with Clz11, with 179 identical sites. The pairwise positive value is as high as 84%. However, it exhibits substantially lower sequence identity to LovG, with only

a 20% match. Importantly, the active site residues, namely S122, D201, and H229 from LovG, are conserved in Mfm8. In contrast, Clz11 is shorter than the other two proteins and lacks the active residue H229. Nevertheless, reported experimental evidence confirmed Clz11 as an active hydrolase enzyme during the biosynthesis of Zaragozic Acid A **24**.[148]

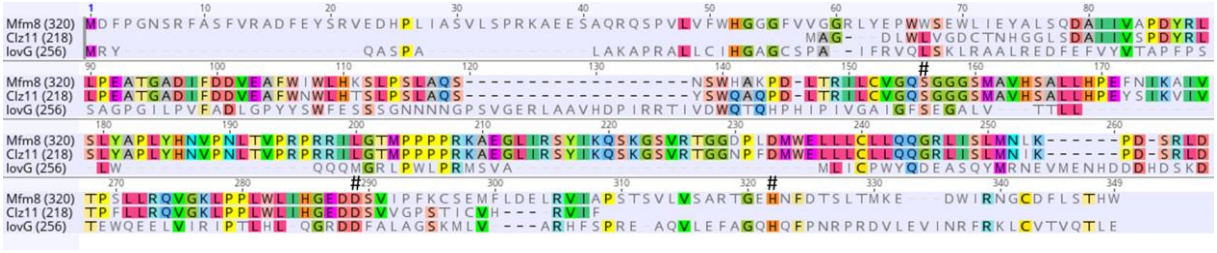


Figure 2.3.2 The protein sequence alignment of Mfm8, Clz11, and LovG. The '#' symbol represents the active site from LovG.

Mfm10 underwent a detailed analysis, including a BLAST search of its protein sequence, which resulted in specific hits such as beta-lactamase and AmpC.[149] The BLAST results suggest that Mfm10 is likely a protein containing a serine hydrolase domain, which may indicate hydrolase activity akin to that found in *Staphylococcus aureus* teichoic acid D-alanine hydrolase [150] or transferase activity similar to *Aspergillus terreus* monacolin J acid methylbutanoyl transferase.[148]

Table 2.3.2 The list of domain hits of Mfm10.

Name	Accession	E-value	Description
Beta-lactamase	pfam00144	4.16e-34	This family appears to be distantly related to pfam00905 and pfam00768 D-alanyl-D-alanine carboxypeptidase.
AmpC	COG1680	4.02e-28	CubicO group peptidase, beta-lactamase class C family.
PRK10662	PRK10662	1.47e-08	Beta-lactam binding protein AmpH; Provisional.

Mfm10, Clz13, and LovG were aligned together. Mfm10 is composed of 595 amino acids, and its sequence shares an 83% identity with Clz13, featuring 507 identical sites. The pairwise positive value is notably high, reaching 88%. However, the sequence identity with LovG is considerably lower, at only 17%. Mfm10 conserves the active site residues S122 and D201 found in LovG. However, H229 is not conserved in Mfm10 and Clz13. It's interesting to note that this site in Clz11 is also not conserved with LovG, even though Clz11 has been experimentally confirmed as an active hydrolase enzyme during the biosynthesis of zaragozic acid A **24**.[148]

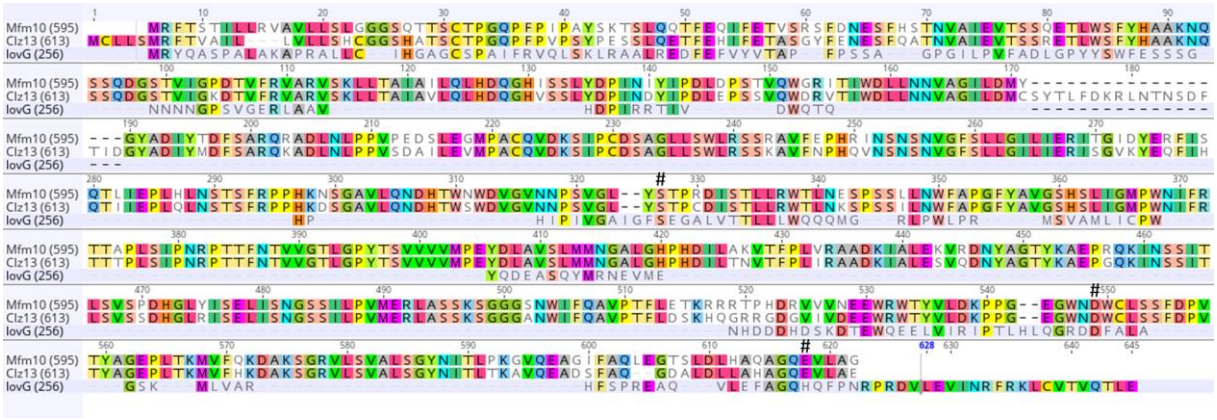


Figure 2.3.3 The protein sequence alignment of Mfm10, Clz13, and LovG; the '#' symbol represents the active site from LovG.

2.3.2 Plasmid Construction and Transformation of *A. oryzae*

In all subsequent experiments, we employed the vector pTYGs-*argB*, which is equipped with a *2μ* origin of replication and a *URA3*, enabling replication and facilitating auxotrophy selection in *Saccharomyces cerevisiae*. Of importance in the replication and selection of *E. coli* are the *colE1* and *AmpR* resistance genes. Each vector incorporates four promoters P_{amyB}, P_{adh}, P_{gpdA}, and P_{eno} and related terminators. P/T_{amyB} possesses attR sites that facilitate LR recombination through cooperation between the entry vector (PE-YA_attL) and the destination vector (pTYGs_attR).

We constructed three expression plasmids through yeast and LR recombination. In the first step, the gene encoding SQTKS was divided into three fragments, each of which contained overlapping regions. These intron-free fragments were cloned using a template from a plasmid created by former colleagues.[137] The resulting fragments were individually inserted into the entry vector PE-YA to create *sqtks*-PE-YA through yeast recombination. Next, the empty vector pTYGs_*argB* served as the destination vector, accepting *sqtks*-PE-YA to form the expression clone *sqtks*-pTYGs_*argB* via LR recombination (Table 2.3.3). In this construct, the expression of *sqtks* was under the control of the P_{amyB}. The plasmids generated were sequenced to confirm their integrity.

For the second plasmid, we introduced *mfm10* under the control of P_{adh} into the pTYGs_*argB*, yielding *mfm10*-pTYGs_*argB*, using yeast recombination. Then, *sqtks*-PE-YA was combined with *mfm10*-pTYGs_*argB*, resulting in *sqtks-mfm10*-pTYGs_*argB*. For the third plasmid, we extended the system by adding *mfm8* under the control of P_{gpdA} to *sqtks-mfm10*-pTYGs_*argB*. This led to *sqtks-mfm10-mfm8*-pTYGs_*argB* (Table 2.3.3, figure 2.3.4).

Table 2.3.3 Plasmids constructed in this study.

Construct ID	Plasmids	Features
SQ01	*sqtks*-PE-YA	An entry vector to transfer *sqtks*
SQ02	*sqtks* - pTYGs_*argB*	P_{amyB} induces *sqtks;* selection marker is *argB*
SQ03	*sqtks-mfm10*-pTYGs_*argB*	P_{amyB} induces *sqtks*, P_{adh} induces *mfm10*; selection marker is *argB*
SQ04	*sqtks-mfm10-mfm8*-pTYGs_*argB*	P_{amyB} induces Hrtc, P_{adh} induces *mfm10*, P_{gpdA} induces *mfm8*; selection marker is *argB*

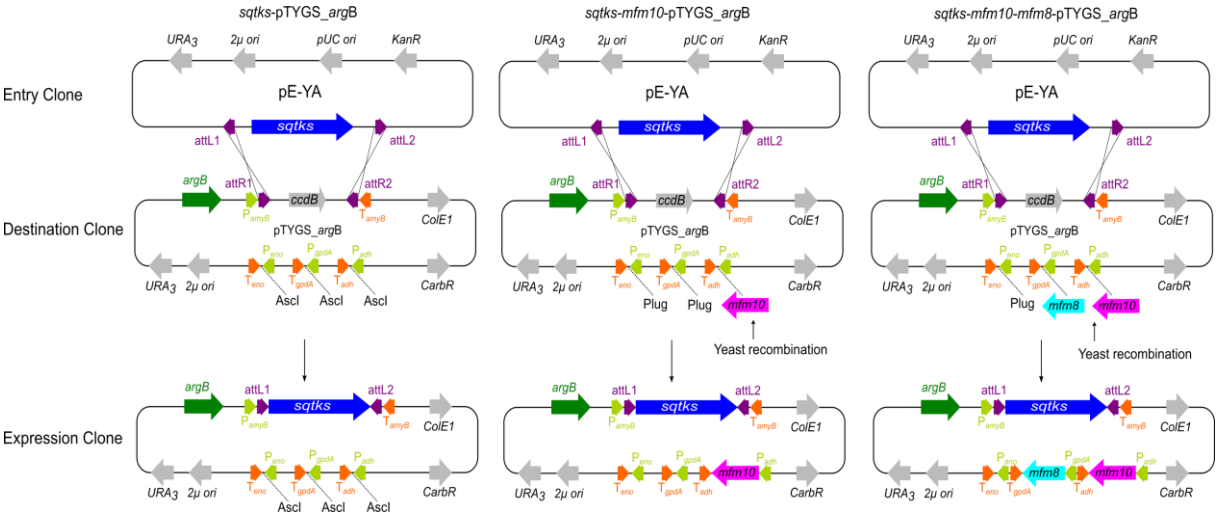


Figure 2.3.4 Overview of plasmids generated through yeast recombination and gateway cloning; 'Plug' refers to a 500-bp overlap that covers a 250-bp promoter and a 250-bp terminator, enabling the reconstitution of the restricted promoter and terminator.

Here, the quadruply auxotrophic *A. oryzae* NSAR1 strain (*ΔargB, sC*$^-$*, adeA*$^-$*, niaD*$^-$) served as the heterologous expression host. Fresh mycelium was enzymatically lysed, and the resulting protoplasts were examined under a microscope to assess their quality. Three groups, SQ02, SQ03, and SQ04, were individually transformed into protoplasts using the $CaCl_2$/PEG-mediated method. The selection of transformants was carried out on CZD/S agar lacking arginine for two rounds. A small mycelium sample was scraped from the DPY agar for genomic DNA extraction.

PCR was performed to verify the insertion of target genes into the genomic DNA of transformants obtained from three separate experiments. Primers situated at the junction of the promoter and the beginning of the terminator were employed. P1 corresponds to the terminal region of the P_{amyB}, while P2 marks the initial segment of the T_{amyB}. Likewise, P3 is positioned at the end of the P_{adh}, and P4 is located at the commencement of the T_{adh}. P5 designates the end of the P_{gpdA}, and P6 corresponds to the start of the T_{gpdA} (Figure 2.3.5A). For the transformants containing SQ02, eight of them were confirmed using two pairs of primers (P1/Sqtks-R1 and Sqtks-F9/P2, Figure 2.3.5B). In the case of transformants containing SQ03, seven of them were confirmed using a pair of primers (P3/P4, Figure 2.3.5C). Lastly, for the transformants containing SQ04, seven of them were confirmed using two pairs of primers (P3/P4 and P5/P6, Figure 2.3.5D).

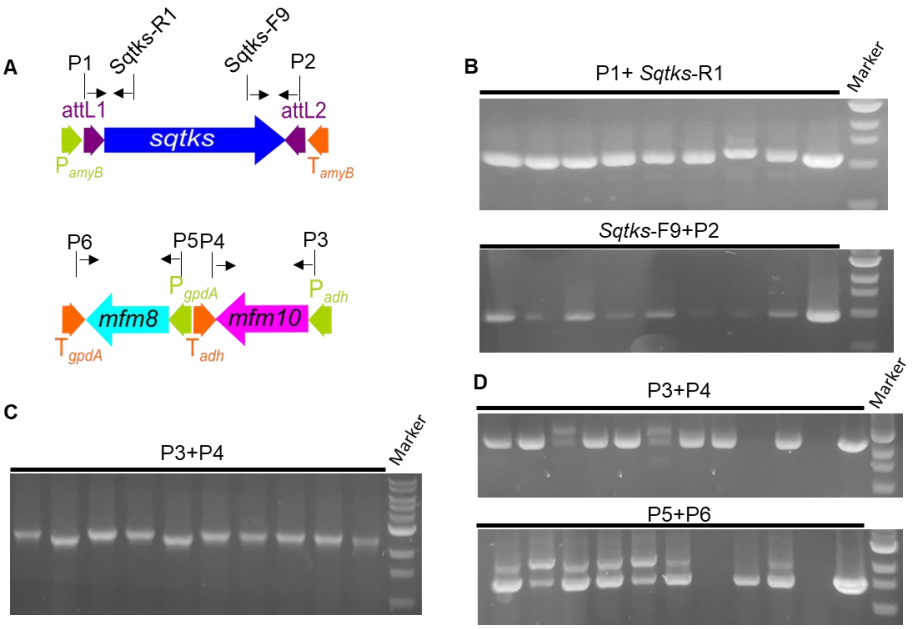


Figure 2.3.5 PCR analysis of candidate transformants. **A** shows the gene cassette that includes the primer positions. **B** displays the bands of transformants harboring SQ02. **C** shows the bands of transformants harboring SQ03. **D** shows the bands of transformants harboring SQ04.

2.3.3 Expression of SQTKS, SQTKS + Mfm10, SQTKS + Mfm10 + Mfm8 Respectively

The transformants and an untransformed NSAR1 strain were cultured in DPY medium for 7 days. After extracting chemicals from each transformant, we conducted LCMS analysis on the resulting crude extracts of transformants for SQTKS, SQTKS + Mfm10, and SQTKS + Mfm10 + Mfm8, in comparison to an untransformed strain. In the ultraviolet absorption chromatograms obtained using a diode array detector (DAD), we observed a distinctive peak at 7.5 minutes in all three experimental groups, which was absent in the untransformed strain (Figure 2.3.6A and B).

Further analysis was carried out on the total ion chromatograms (TIC) using electrospray ionization (ESI) in positive ion mode (ES^+). We not only identified peak **25**, but also discovered another novel peak **25a** with weak UV-absorption, which was also absent in the untransformed strain. We proceeded to perform a detailed analysis of compound **25** and **25a**, employing mass spectrometry, UV-absorption spectrometry, high-resolution mass spectrometry (HRMS), and nuclear magnetic resonance (NMR) techniques.

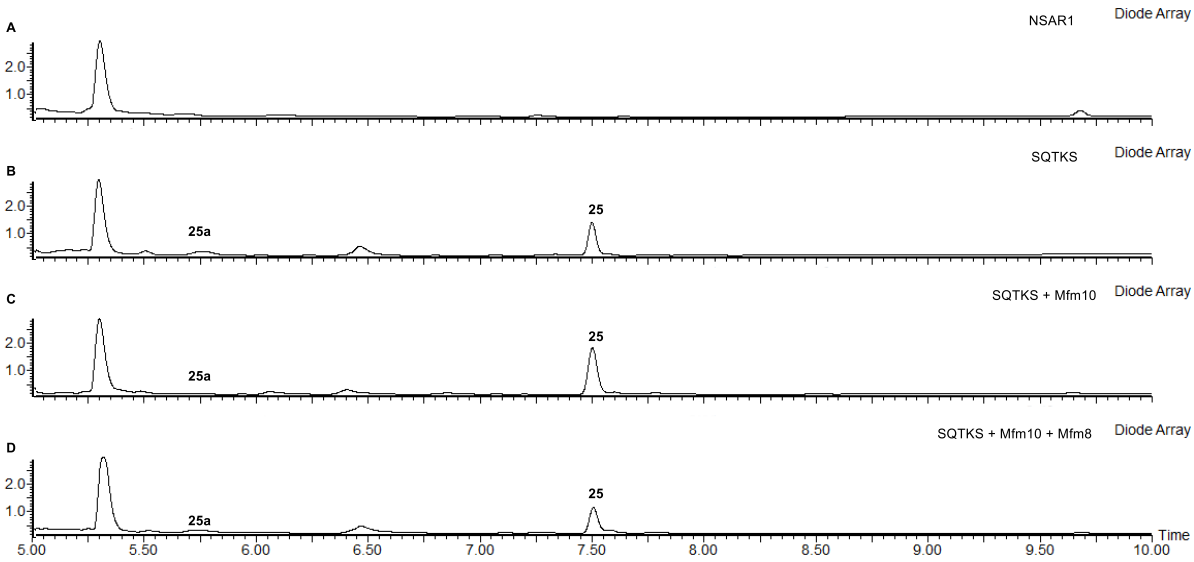


Figure 2.3.6 A, DAD of untransformed strain NSAR1; **B**, DAD of SQTKS transformant; **C**, DAD of SQTKS + Mfm10 transformant; **D**, DAD of SQTKS + Mfm10 + MfM8 transformant.

2.3.3.1 Identification of Tetraketide 25

Our goal is to express SQTKS in *A. oryzae* NSAR1 and optimize the titre of tetraketide **25**. To identify compound **25**, we referred to the mass data associated with the reported structure of tetraketide **25**, which exhibited a mass of *m/z* 170.[137] We searched for a mass fragment pattern of 171 using extracted ion chromatograms (EIC) in positive ion mode (ES^+, Figure 2.3.7B). In ES^+ mode, we detected comound **25**, which had the same retention time as observed in the total ion chromatograms. Additionally, we observed compound **25a**, with a mass of 171 in ES^+ (Figure 2.3.7B). Furthermore, we conducted a search for a mass fragment pattern of 169 using EIC in negative ion mode (ES^-, Figure 2.3.7A). In ES^-, we identified compound **25** at the same retention time as that in ES^+.

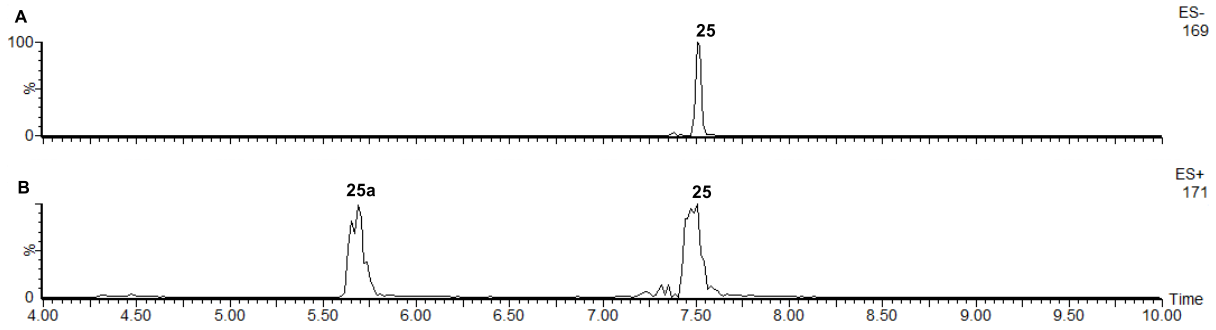


Figure 2.3.7 A, the EIC for searching 169 in ES^-; **B**, the EIC for searching 171 in ES^+

Compound **25** exhibited distinct mass fragmentation patterns in both positive ion mode (ES^+) and negative ion mode (ES^-, Figure 2.3.8). In ES^+, the mass spectra displayed peaks at *m/z* 153.3, 171.3 and 212.4. The *m/z* 153.3 peak suggests the loss of a water molecule, represented as $[M + H - H_2O]^+$. The *m/z* 171.3 peak is likely associated with the molecular ion $[M + H]^+$. The *m/z* 212.4 peak may correspond to $[M + H + C_2H_3N]^+$. In ES^-, compound **25** showed *m/z* 169.2, which is indicative of $[M - H]^-$. A larger

fragment at *m/z* 265.2 was attributed to the background signal from the untransformed strain, while a weak *m/z* 339.1 is consistent with $[2M - H]^-$. To support this identification, we observed UV absorption at 212 nm within the wavelength range of 200-600 nm (Figure 2.3.8).

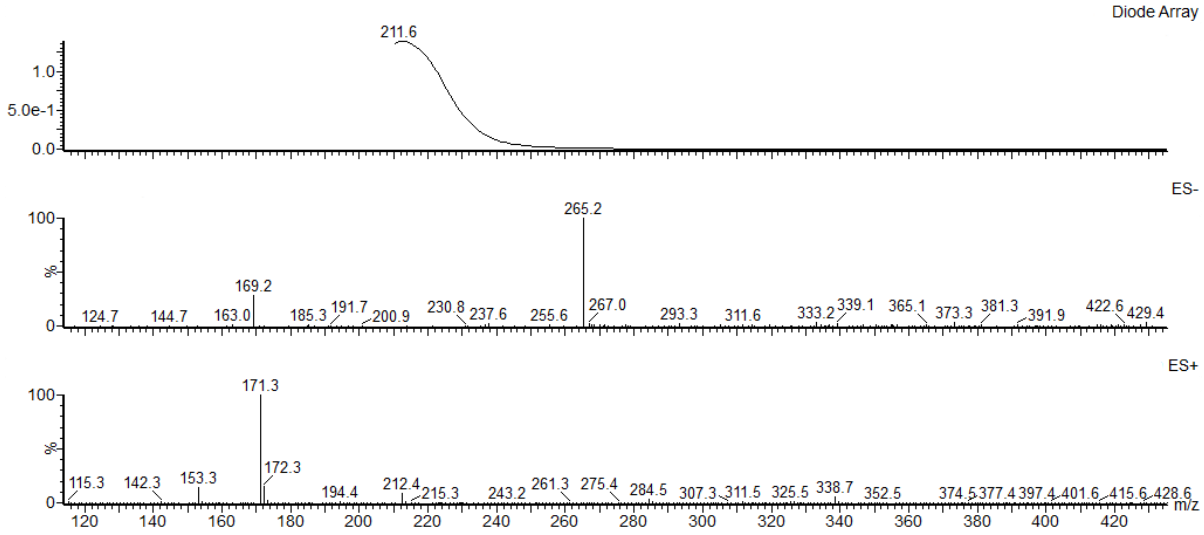


Figure 2.3.8 UV-absorption (top), mass fragmentation pattern in ES^- (middle) and ES^+ (bottom) of compound **25** from the transformant of SQTKS.

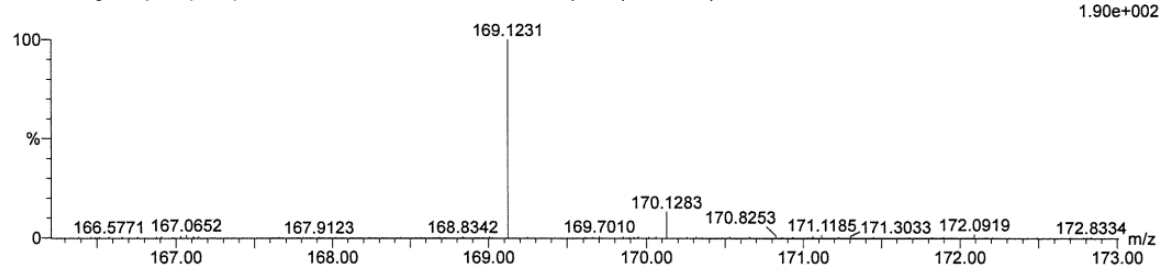
Furthermore, through HRMS analysis, we also confirmed that compound **25** is indeed tetraketide **25**, as the theoretical calculation for the $[M - H]^-$ ion resulted in an expected *m/z* value of 169.1231, which closely matched the observed *m/z* of 169.1229 (Figure 2.3.9).

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 20.0 PPM / DBE: min = -1.5, max = 50.0
Element prediction: Off
Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions
132 formula(e) evaluated with 2 results within limits (all results (up to 1000) for each mass)
Elements Used:
C: 0-80 H: 0-125 N: 0-5 O: 0-4 Na: 0-1

Sun QTof Premier HAB321
YS 002, neg 803 (8.211) AM (Cen,4, 60.00, Ht,10000.0,554.26,0.70,LS 10); Sm (SG, 1x5.00) 1: TOF MS ES-
1.90e+002



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
169.1231	169.1229	0.2	1.2	2.5	27.8	0.1	C10 H17 O2
	169.1204	2.7	16.0	-0.5	29.7	2.0	C8 H18 O2 Na

Figure 2.3.9 HRMS data for compound **25** from the transformant of SQTKS.

2.3.3.2 Charactrization of 25a

To purify compound **25a**, we conducted a 2-liter scale fermentation in DPY medium. Then we subjected the crude extracts to preparative LCMS, resulting in the isolation of **25a**. We conducted a thorough analysis of the mass information for **25a**, revealing distinctive mass fragmentation patterns in two electrospray ionization modes (Figure 2.3.10). In ES^+, the mass spectra exhibited peaks at *m/z* 153.3, 171.3, 189.3. The *m/z* 153.3 peak represented as $[M + H - 2H_2O]^+$. The *m/z* 171.3 peak likely corresponds to the molecular ion $[M + H - H_2O]^+$, and the weak *m/z* 189.3 peak suggests the presence of $[M + H]^+$. Furthermore, in the mode of ES^-, compound **25a** displayed an *m/z* 187.1 peak, representing $[M - H]^-$, and a weak *m/z* 375.3 peak, consistent with $[2M - H]^-$. However, it's hard to observe a prominent UV absorption signal for this compound in the chromatograms.

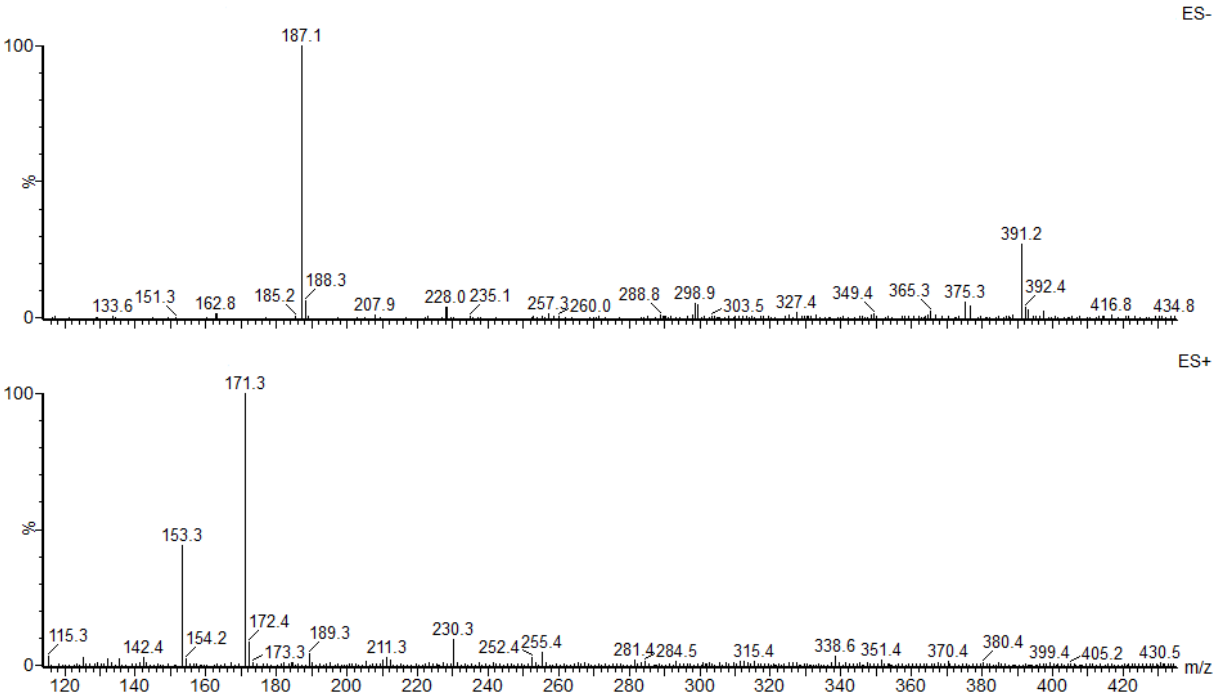


Figure 2.3.10 Mass fragmentation pattern in ES^+ (bottom) and ES^- (top) of compound **25a** from the transformant of SQTKS.

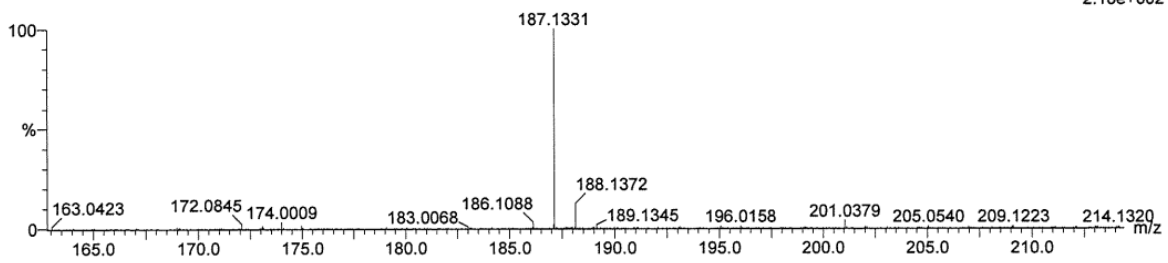
Through high-resolution mass spectrometry (HRMS) analysis, we determined the molecular formula of **25a** to be $C_{10}H_{20}O_3$. Calculations for $[M - H]^-$ resulted in a theoretical *m/z* value of 187.1334, which closely matched the observed mass of 187.1331 (Figure 2.3.11).

Elemental Composition Report **Page 1**

Single Mass Analysis
Tolerance = 20.0 PPM / DBE: min = -1.5, max = 50.0
Element prediction: Off
Number of isotope peaks used for i-FIT = 3

Monoisotopic Mass, Even Electron Ions
145 formula(e) evaluated with 2 results within limits (all results (up to 1000) for each mass)
Elements Used:
C: 0-80 H: 0-125 N: 0-5 O: 0-4 Na: 0-1

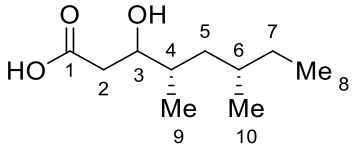
Sun QTof Premier HAB321
YS 003 769 (7.865) AM (Cen,4, 60.00, Ht,10000.0,554.26,0.70,LS 10); Sm (SG, 1x5.00) 1: TOF MS ES- 2.18e+002



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
187.1331	187.1334	-0.3	-1.6	1.5	26.1	0.2	C10 H19 O3
	187.1310	2.1	11.2	-1.5	27.7	1.8	C8 H20 O3 Na

Figure 2.3.11 HRMS data for compound **25a** from the transformant of SQTKS.

The structure of **25a** was determined using ^{1}H, ^{13}C, and 2D NMR. In the ^{1}H NMR data, three methyl signals were observed at δ_H 0.79 (3H, m), δ_H 10.8 (3H, m), and δ_H 0.84 (3H, m). Three methine signals were found at δ_H 3.72 (1H, ddd, *J* 9.3, 4.6, 3.6), δ_H 1.55 (1H, m), and δ_H 1.37 (1H, m), as well as three sets of methylene protons at δ_H 2.10 (1H, dd, *J* 14.9, 9.3) and 2.25 (1H, m), δ_H 0.88 (1H, m) and δ_H 1.28 (1H, m), and δ_H 1.03 (1H, m) and δ_H 1.32 (1H, m). The ^{13}C NMR and HSQC data for **25a** revealed three methyl signals (δ_C 15.5, δ_C 10.8, δ_C 20.0), three methine signals (δ_C 70.9, δ_C 35.7, δ_C 31.1), and three methylene signals at δ_C 38.4, δ_C 39.0, and δ_C 28.0. Furthermore, a carbonyl carbon was observed at δ_C 173.8 (Table 2.3.4).



25a

Chemical Formula: $C_{10}H_{20}O_3$
Exact Mass: 188.1412

Table 2.3.4 Summarized NMR signals (^{1}H-400 MHz, ^{13}C-100 MHz) for **25a** recorded in DMSO-d6.

Pos.	δ_C / ppm	δ_H / ppm (J/Hz)	^{1}H-^{1}H COSY	HMBC (H-C)
1	173.8			
2	38.4	2.10, 1H, dd (14.9, 9.3); 2.25, 1H, m	3 3	1, 3, 4 1, 3, 4
3	70.9	3.72, 1H, ddd (9.3, 4.6, 3.6)	2, 4	1, 5, 9
4	35.7	1.55, 1H, m	3, 5, 9	3, 5, 9
5	39.0	0.88, 1H, m; 1.28, 1H, m	4 4	6, 7, 9, 10 6, 7, 9, 10
6	31.1	1.37, 1H, m	7, 10	5, 7, 8, 10
7	28.0	1.03, 1H, m; 1.32, 1H, m	6, 8 6, 8	5, 6, 8, 10 6, 8, 10
8	15.5	0.79, 3H, m	7	7
9	10.8	0.82, 3H, m	4	3, 4, 5
10	20.0	0.84, 3H, m	6	5, 6, 7

The carbon atom C-1 (δ_C 173.8) was identified. The position of C-2 was ascertained through an HMBC correlation between H-2 (δ_H 2.10, 1H, dd, *J* 14.9, 9.3 Hz; δ_H 2.25, 1H, m) and C-1 (Table 2.3.4). H-3 was confirmed through a ^{1}H-^{1}H-COSY correlation between H-2 and H-3 (δ_H 3.72, 1H, ddd, *J* 9.3, 4.6, 3.6 Hz). The HSQC spectrum provided further confirmation of C-3 at δ_C 70.9, likely indicating a hydroxyl group at C-3. H-4 was verified through a ^{1}H-^{1}H-COSY correlation between H-3 and H-4, as well as between H-4 and H-5. A comparison with the ^{13}C-NMR data from existing literature led to the assumption that three methyl groups were still present at C-4, C-6, and C-7.[137] A ^{1}H-^{1}H-COSY correlation between H-4 and H-9 confirmed the location of the methyl group at C-9. Additionally, the H-6 exhibited a ^{1}H-^{1}H-COSY correlation with H-10, providing confirmation for the position of the methyl C-10. The protons of H-7 displayed ^{1}H-^{1}H-COSY correlations with H-8, thus confirming the presence of a methyl group at C-8. Therefore, we established the structure of **25a**, indicating that the double bond most likely underwent hydrolation, transforming into a hydroxyl group, based on the information provided by the structure of compound **25**. However, we were unable to determine the stereochemistry of the β-hydroxyl group by NMR. A more in-depth experiment is required for this purpose.

2.3.3.3 Time Course of transformant SQTKS

In the last section, we observed the reappearance of tetraketide **25** in the chromatograms generated from the LCMS analysis of three experimental groups after seven days of fermentation. To obtain the optimal fermentation duration for maximizing titer, we conducted a time course experiment. We cultivated the producer strain in DPY medium for a total of seven days, extracting the medium daily to collect crude chemicals. On each day, we subjected a 10 mL culture sample to two rounds of extraction, using an equivalent volume of ethyl acetate each time. The organic phase was evaporated until it reached complete desiccation. We then employed 500 µL of methanol to dissolve the resulting extract. The liquid component of the sample was collected and analyzed using analytical LCMS. This entire process was repeated three times with 20 µL of each sample.

The chromatograms were displayed through UV-absorption, allowing us to observe the "phenotypic characterization" based on the height of the peaks for compound **25**. On the first day, the peak appeared relatively small; however, by the third day, it significantly increased, reaching almost three times the height observed on the first day. On the fourth day, the peak's height was lower, indicating the beginning of a decline that continued from day 3 to day 6. By the seventh day, the peak's height showed a minor increase compared to day 6. This may be attributed to the onset of mycelium degradation and cell rupture, releasing chemicals from the cells (Figure 2.3.12).

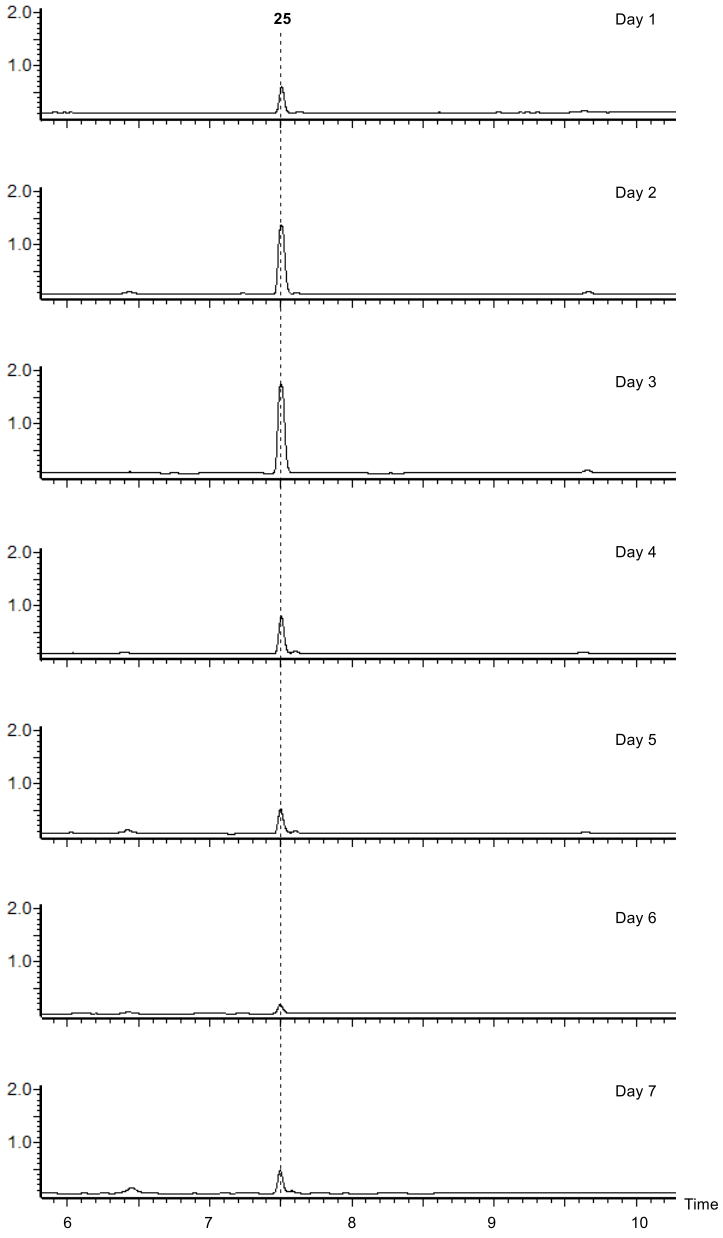


Figure 2.3.12 Chromatograms of extracts detected by DAD after the time course of the transformant harboring SQTKS from day 1 to day 7.

Although we can visually discern the variations in peak heights within the chromatograms, it's crucial to perform a quantitative determination of the titer. We determined the titer by employing a calibration

curve. To begin, we dissolved the purified compound **25** in methanol and then carried out dilutions to create a range of concentrations, from 0.02 mg/mL to 0.002 mg/mL, after which they were subjected to LCMS analysis utilizing the selected ion recording (SIR) detection method in ES^+ mode. A calibration curve was then constructed based on these concentrations and their respective peak areas (Figure 2.3.13A). We then utilized this calibration curve to quantitatively analyze the transformants (Figure 2.3.13B).

Analysis of the curve revealed that from day 1 to day 3, the titer experienced an increase, reaching its peak on day 3 at approximately 2.9 mg/L. However, consistent with the trends observed in the UV chromatograms, the titer began to decrease from day 3 to day 6. Although there was a slight uptick on day 7, the titer remained significantly lower than that observed on day 3 (Figure 2.3.13B).

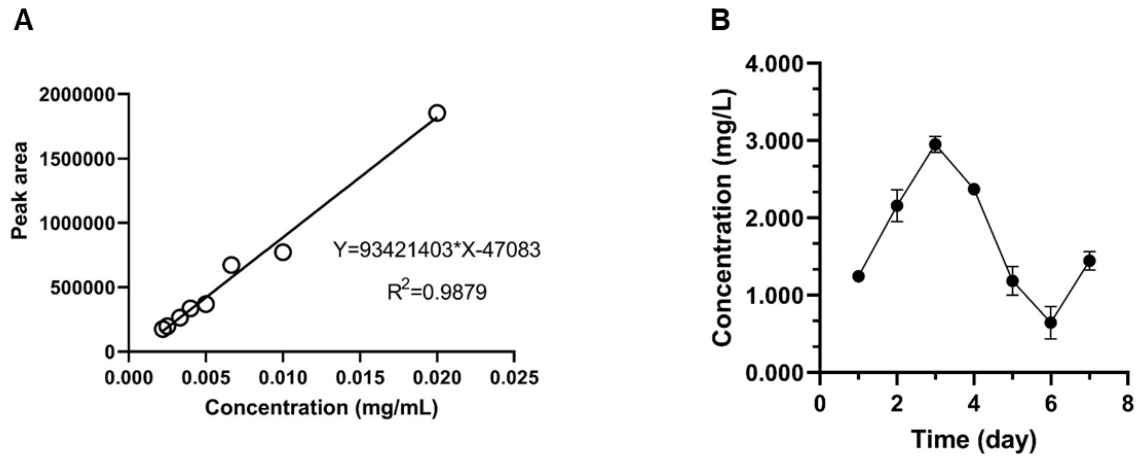


Figure 2.3.13 A, Calibration curve of **25**; **B**, titers of transformant harbouring SQTKS from day 1 to day 7.

2.3.3.4 Time Course of transformant SQTKS + Mfm10

We hypothesised that Mfm10 may possess hydrolase activity, potentially aiding in the release of tetraketide **25** from the SQTKS enzyme. In our experimental approach, we co-expressed Mfm10 alongside SQTKS under the regulation of the P_{adh}, following a method akin to the expression of SQTKS. The chromatogram revealed that the SQTKS + Mfm10 transformant could indeed generate tetraketide **25** after a seven-day fermentation period (Figure 2.3.6B). However, determining the yield solely by comparing peak heights presented a challenge. As a result, we conducted a time course experiment to accurately calculate the titers of tetraketide **25** within the SQTKS + Mfm10 transformant in different time of fermantation.

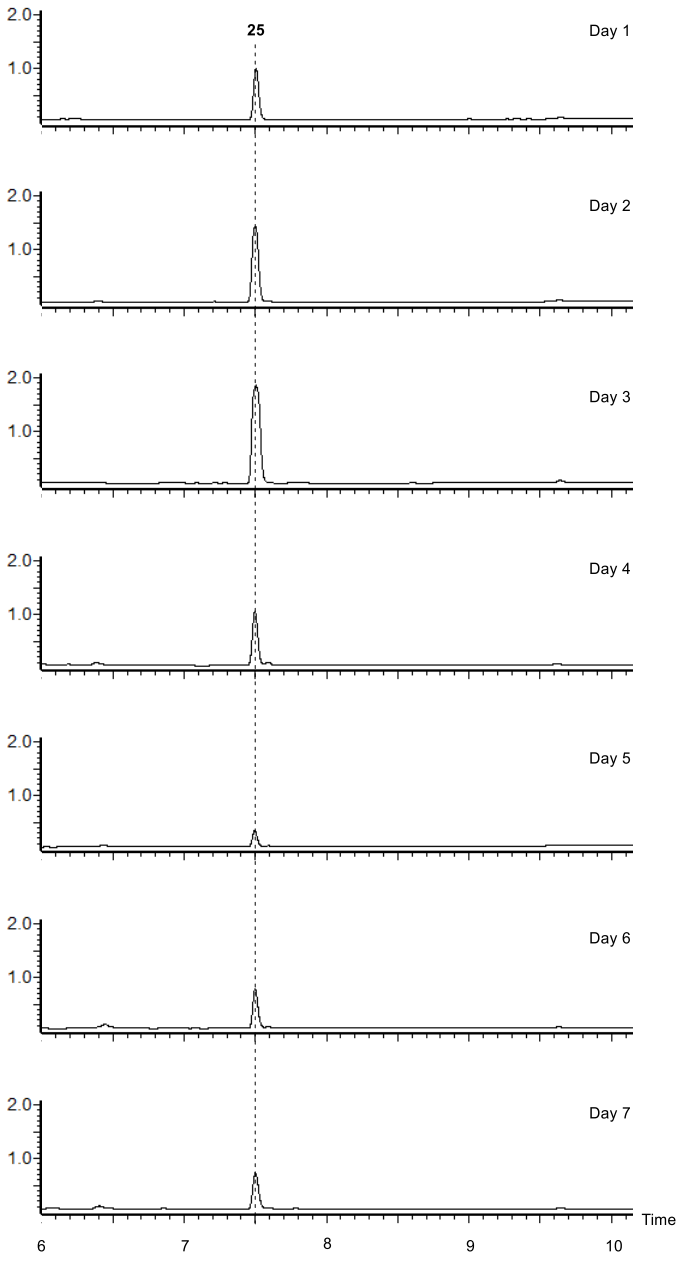


Figure 2.3.14 Chromatograms of extracts after the time course of the transformant SQTKS + Mfm10 from day 1 to day 7 detected by DAD.

We conducted an analysis of tetraketide **25** titer in the SQTKS + Mfm10 transformant, employing a formula derived from the calibration curve (Figure 2.3.13A). Subsequently, we organized the daily titers and presented them as a graphical curve (Figure 2.3.15). The curve reveals a consistent rise in titer from day 1 to day 3. However, after the third day, the titer exhibited a diminishing trend, reaching its lowest point on day 5. There was a marginal uptick on day 6, followed by a plateau from day 6 to day 7. The peak titer was observed on day 3, measuring approximately 3.2 mg/L. This value, while only slightly exceeding the 2.9 mg/L titer observed in the SQTKS transformant, but marks a nearly fourfold increase compared to the titer 0.8 mg/L produced by *A. oryzae* M-2-3. Despite these findings, it remains

premature to conclude that Mfm10 possesses a strong capacity for enhancing the off-loading of polyketides based on these results. We next tried another predicted hydrolase Mfm8.

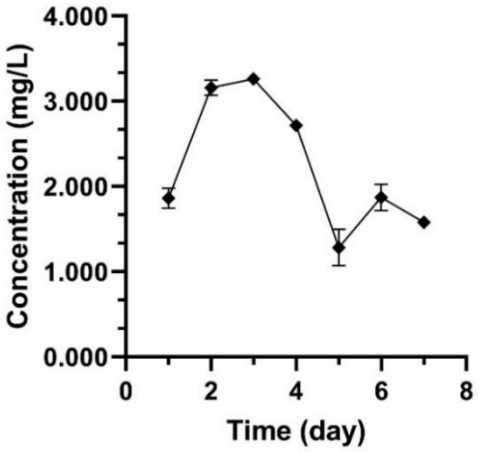


Figure 2.3.15 titers of transformant SQTKS + Mfm10 from day 1 to day 7.

2.3.3.5 Time Course of transformant SQTKS + Mfm10 + Mfm8

We have proposed the possibility of Mfm8 having hydrolase activity, which might aid in releasing tetraketide **25** from the SQTKS enzyme. To investigate this, we co-expressed Mfm8, SQTKS, and mfm10 using the same procedure as employed for SQTKS expression. Our chromatogram data clearly shows that in the transformant with SQTKS + Mfm10 + Mfm8, tetraketide **25** was successfully produced after seven days of fermentation (Figure 2.3.16). Additionally, we carried out a time-course experiment to accurately determine the titer of tetraketide **25** at various stages during the fermentation process in the SQTKS + Mfm10 + Mfm8 transformant.

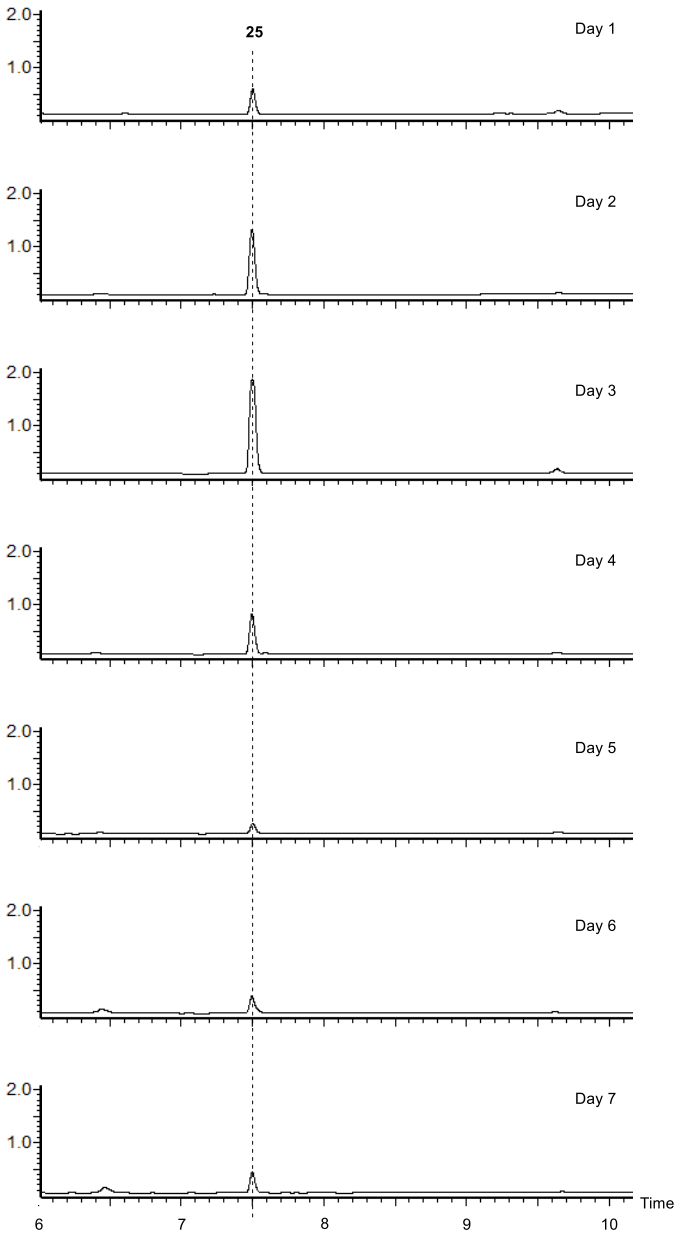


Figure 2.3.16 Chromatograms of extracts after the time course of the transformant SQTKS + Mfm10 + Mfm8 from day 1 to day 7 detected by DAD.

While we can observe changes in the chromatogram peak heights, it's vital to measure the titer quantitatively. This will allow us to make meaningful comparisons with the results of previous expression experiments. To determine the titer, we applied the same calibration curve as in the two prior experiments for the quantitative analysis of the SQTKS + Mfm10 + Mfm8 transformants. The examination of the curve showed that the titer increased from day 1 to day 3, reaching its highest point at approximately 2.9 mg/L on day 3. Nevertheless, in alignment with the patterns seen in the UV chromatograms, the titer started declining from day 3 to day 6. While there was a slight upturn on day 7, the titer consistently remained notably lower than what was observed on day 3 (Figure 2.3.17).

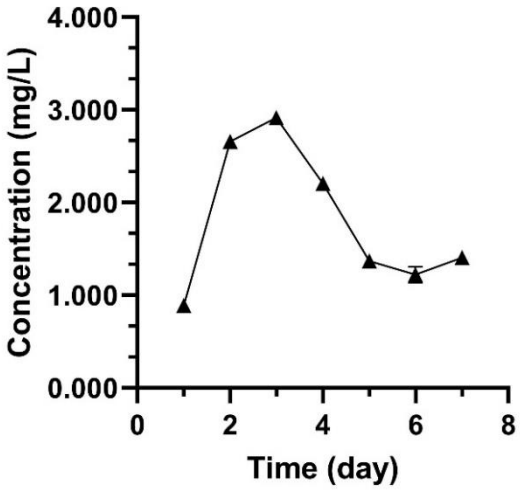


Figure 2.3.17 Titers of transformant SQTKS + Mfm10 from day 1 to day 7.

The graph combines three curves representing the time course of experiments conducted by three distinct groups. Analyzing the curve heights, it becomes evident that for each day, the titer of SQTKS + Mfm10 is slightly higher than the other two experiments (Figure 2.3.18 A). On day 3, all experiments reach their peak titers, which were then extracted to create a column graph (Figure 2.3.18 B). Specifically, the highest titer observed in the SQTKS transformant is approximately 2.9 mg/L, whereas for SQTKS + Mfm10, it stands at approximately 3.2 mg/L. The highest titer for SQTKS + Mfm10 + Mfm8 is also around 2.9 mg/L.

Hence, it would be premature to say that the hydrolases within Mfm10 or Mfm8 contribute significantly to the release of the product by SQTKS. Although there exists a slight distinction between the SQTKS + Mfm10 transformant and the other two experiments, this difference is likely attributed to the strain's growth capacity, which may, in turn, be influenced by subtle environmental variations. As a result, this does not provide sufficient evidence to support the hydrolytic role of Mfm10 and Mfm8 in the process.

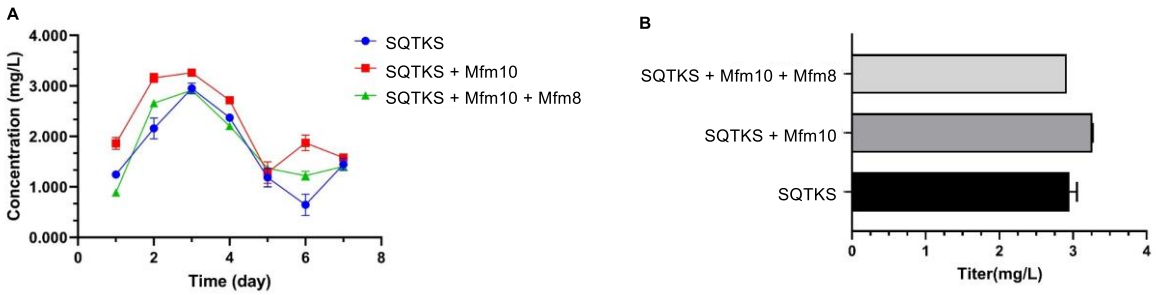


Figure 2.3.18 A, Overlaps of three curves from the time-course experiment for SQTKS, SQTKS + Mfm10, and SQTKS + Mfm10 + Mfm8; **B,** A column graph displaying the three highest titers at day 3.

Nevertheless, when we compared the obtained titers with the previously reported titer of 0.8 mg/L,[137] we found that our results are approximately four times higher. Consequently, we plan to proceed with the subsequent mutation experiments on SQTKS in the *A. oryzae* NSAR1 host.

2.3.4 Expression of Mutant SQTKS in *A. oryzae* NSAR1

Former co-workers successfully developed a structural model for the SQTKS ER, demonstrating its remarkable practicality even in the absence of structures of isolated HR-PKS *cis*-ER domains. The models were subjected to *in vitro* validation, affirming that the introduction of the F1941A mutation, or mutations containing F1941A, led to the specific conversion of 4*S*, 6*S*-tetraketide **25p** from an inhibitor when interacting with the WT protein to becoming a substrate for the mutant protein. Moreover, the F2157A mutation expanded the active site pocket and significantly improved substrate specificity, particularly for longer substrates.

These remarkable changes in activity and substrate range prompted us to consider introducing these mutations throughout the entire SQTKS (Figure 2.3.19). This would help us understand how these alterations impact the overall function of all active domains and what kinds of products can be generated within the system.

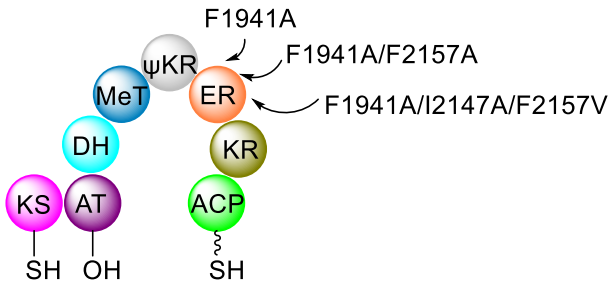


Figure 2.3.19 Domains of SQTKS and some of the mutations in ER.

To introduce mutant sites into the SQTKS, we initiated the process by modifying the primer sequences. We amplified the mutant DNA fragments within these mutant sites using the wild-type SQTKS DNA as a template. The SQTKS DNA was segmented into three fragments, with the mutant sites specifically located within the third fragment (Figure 2.3.20). Amplification was carried out using 43-base pair primers, featuring a 3 base-pair mutant site in the middle (Figure 2.3.20). Following this, the resulting three amplified fragments were combined using yeast recombination. In cases where there were two or three mutant sites, the third fragment was further subdivided into one or two additional fragments for yeast recombination. Afterward, we confirmed the presence of mutant SQTKS DNA on the plasmid pTYGs-*arg* through sequencing. This confirmed mutant DNA was subsequently introduced into *A. oryzae* NSAR1 following the same protocol as described in section 2.3.3.

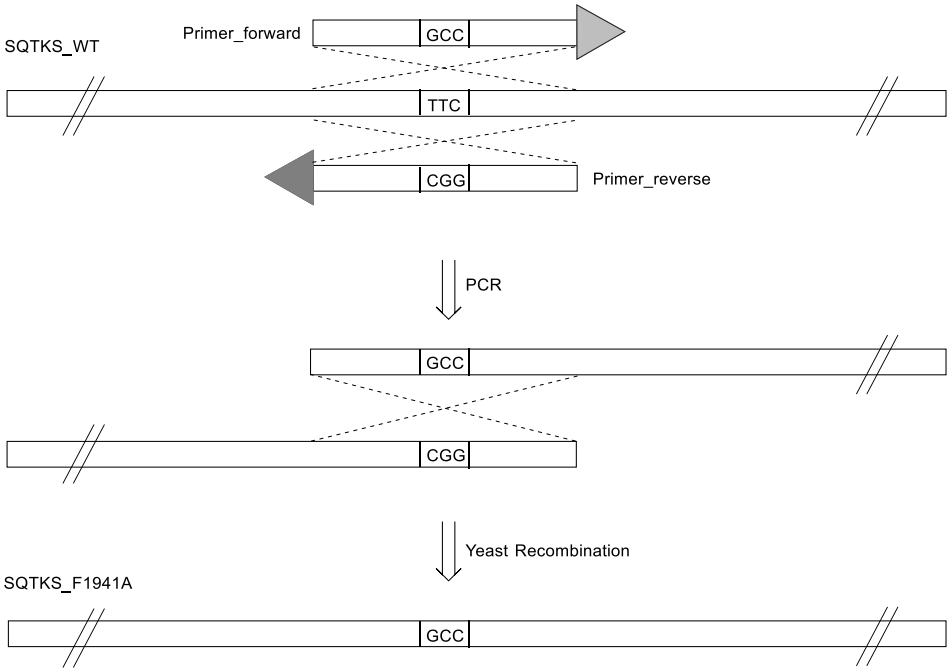


Figure 2.3.20 A brief protocol of inserting the mutant sites into the DNA of WT SQTKS

2.3.4.1 Expression of SQTKS_F1941A

The genomic DNA of the transformants was extracted and utilized as a template for PCR amplification, employing two pairs of primers: S-F/sqtks-R5 and sqtks-F5/S-R, designed to cover the entire DNA sequence of SQTKS (Figure 2.3.21A). We also performed an amplification using the genomic DNA as a template to target a specific fragment containing the mutant site GCC. Subsequently, the resulting PCR product was purified and subjected to sequencing (Figure 2.3.21B). The sequenced results confirmed the successful introduction of the mutation into the genome of *A. oryzae* NSAR1.

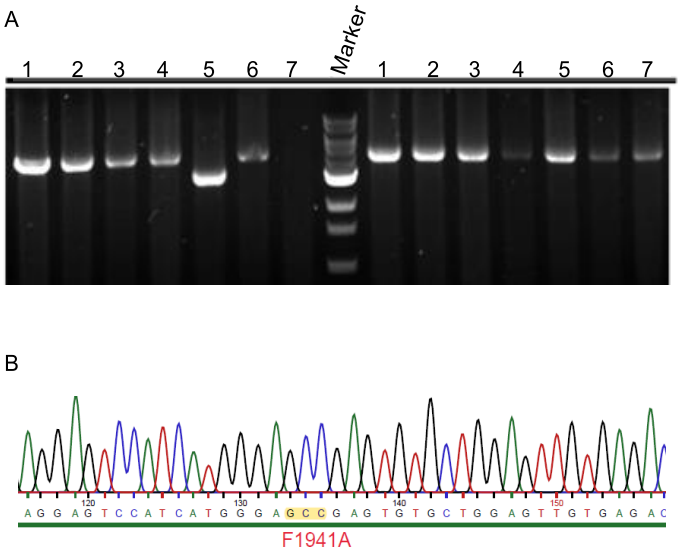


Figure 2.3.21 A, PCR characterization of transformants of SQTKS_F1941A; **B**, sequencing of the PCR product containing the mutation site F1941A.

After completing fermentation and preparing the crude extracts, we subjected the products to LCMS analysis. We compared the UV spectra of transformants containing the SQTKS_F1941A with the transformant containing the SQTKS_WT. We observed a peak at the same retention time as compound **25** by DAD, which originated from the SQTKS_WT transformants (Figure 2.3.22 A). The UV absorption pattern was identical to that of compound **25** (Figure 2.3.23 A). We also conducted TIC and analyzed the mass fragments from the identified peaks, encompassing both ES^+ and ES^-, revealing mass values of 171 for ES^+ and 169 for ES^-, along with several other identical mass fragments to those of compound **25** (Figure 2.3.23 A).

There is a peak labeled as "#" that we further examined via ESI (Figure 2.3.22 A**)**. We determined it had a distinct mass of 341.3 in ES^+ and 165.0 in ES^-, matching the values observed in both SQTKS WT transformants and the untransformed strain NSAR1 (Figure 2.3.23 B). As a result, we cannot classify this peak as a novel one. Our analysis of the UV spectra and ESI did not reveal any new products in the chromatograms. Thus, the introduction of the F1941A mutation into SQTKS did not result in a change in the enzymatic programming, potentially yielding new products. Nonetheless, tetraketide **25** was still produced by the SQTKS_F1941A transformants.

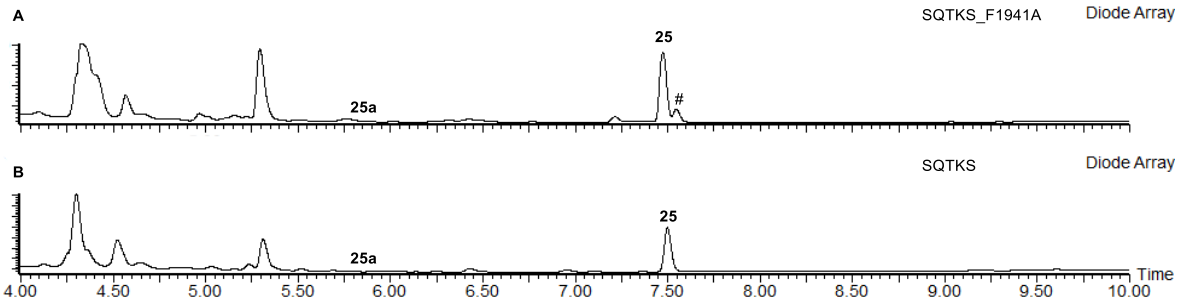


Figure 2.3.22 A, Chromatogram of transformant SQTKS_F1941A by DAD; **B**, chromatogram of transformant SQTKS_WT by DAD; # is a peak that from wild-type of *A. oryzae*.

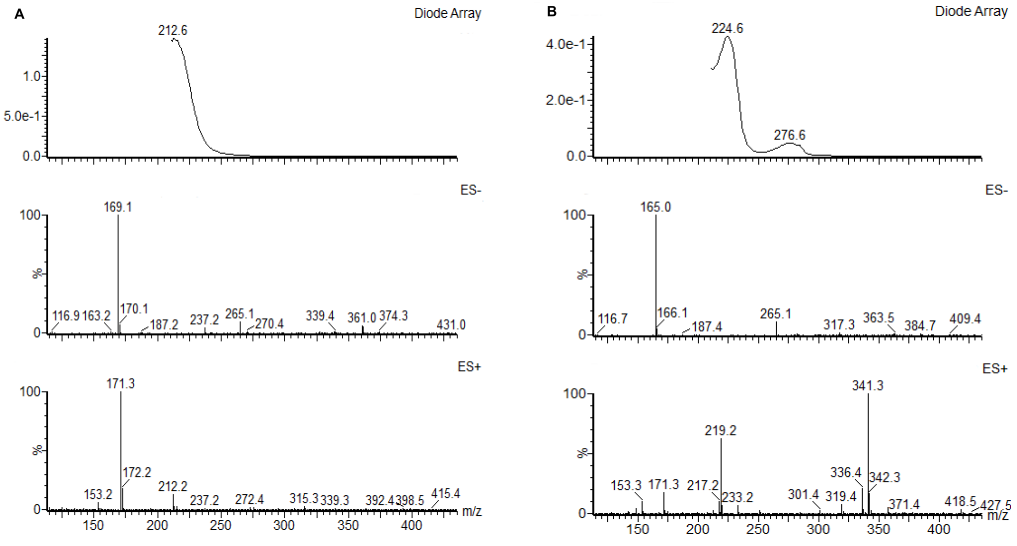


Figure 2.3.23 A, UV-absorption and mass spectra of the peak at 7.5 minutes from transformant SQTKS_F1941A; **B**, UV-absorption and mass spectra of the peak #.

2.3.4.2 Expression of SQTKS_ F1941A/F2157A

In vitro work showed that the double mutation F1941A/F2157A in the isolated ER can catalyse the reduction of the 4*S*, 6*S*-tetraketide **25p**, which is not a substrate of the wild-type ER enzyme. In this work, we introduced the F1941A/F2157A mutation into SQTKS_WT, resulting in the creation of SQTKS_F1941A/F2157A, and transformed it into *A. oryzae* NASR1 to select the SQTKS_F1941A/F2157A transformants. To validate the successful insertion of the entire DNA into the genome, we extracted genomic DNA and employed PCR (Figure 2.3.24A). Furthermore, we sequenced the mutant sites, GCC (Figure 2.3.24B).

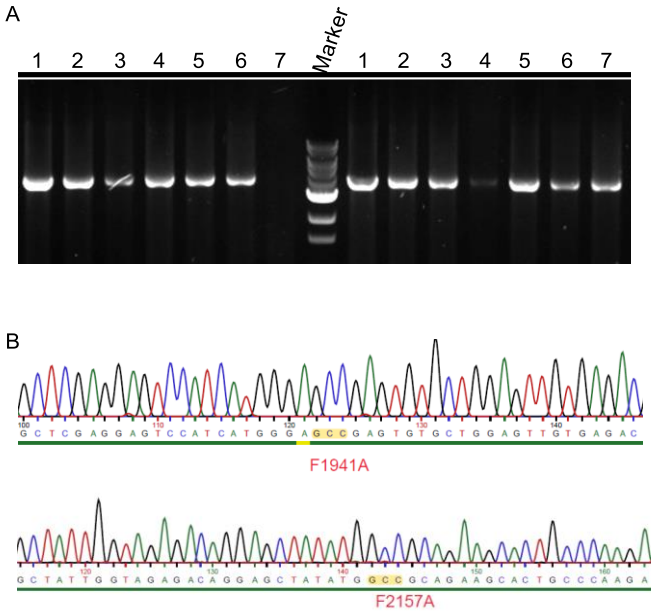


Figure 2.3.24 A, PCR characterization of transformants of SQTKS_ F1941A/F2157A; **B**, sequencing of the PCR product containing the mutation site F1941A and F2157A.

Following the same procedure mentioned in section 2.3.4.1, which involves fermentation and the preparation of crude extracts, we proceeded to analyze the products using LCMS. We then conducted a comparative examination of the UV spectra between the transformants containing SQTKS_F1941A/F2157A and the transformant SQTKS_WT. Notably, in the SQTKS_F1941A/F2157A transformants, we confirmed the presence of compounds **25** and **25a** through DAD and ESI analysis. No additional new products were detected in this analysis (Figure 2.3.25).

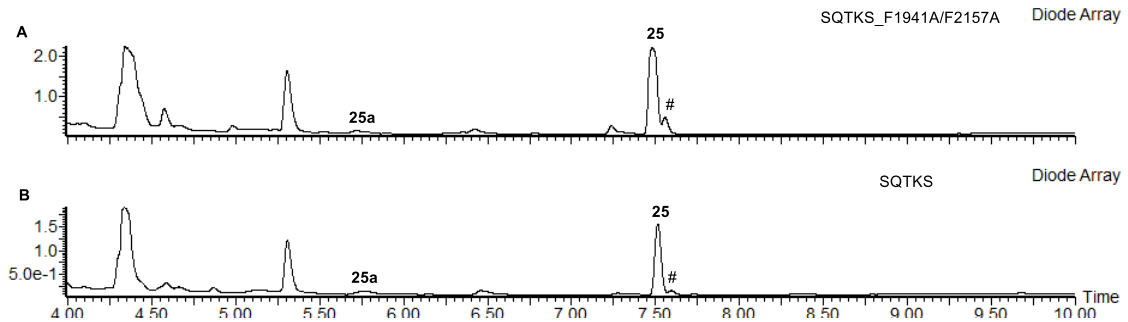


Figure 2.3.25 A, Chromatogram of transformant SQTKS_ F1941A/F2157A by DAD; **B**, chromatogram of transformant SQTKS_WT by DAD; # is a peak that from wild-type of *A. oryzae*.

2.3.4.3 Expression of SQTKS_ F1941A/I2147A/F2157V

The previous experiments also provided evidence that the triple mutant F1941A/I2147A/F2157V combines the effects of the individual mutations. It exhibits the ability to react with 4*S*, 6*S*-tetraketide **25p**, a function that the WT enzyme cannot perform. Here, we introduced the F1941A/I2147A/F2157V mutation into the SQTKS_WT sequence and transformed it into *A. oryzae* NASR1 using the pTYGs-*arg* plasmid to select for SQTKS_F1941A/I2147A/F2157V transformants. To confirm the successful integration of the entire gene into the genome, genomic DNA was extracted and subjected to PCR analysis (Figure 2.3.26A). Furthermore, we conducted sequencing to verify the presence of the mutant sites, GCC and GAC (Figure 2.3.26B).

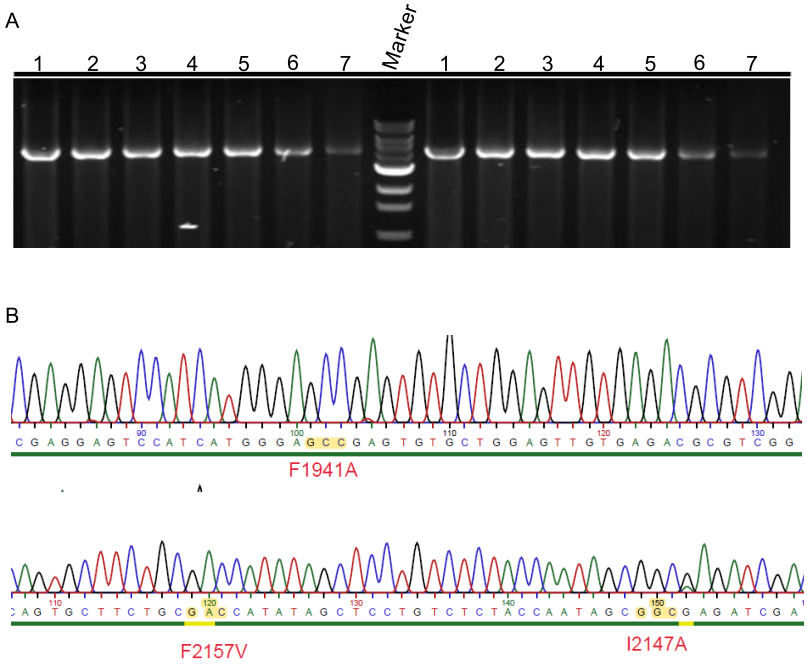


Figure 2.3.26 A, PCR characterization of transformants of SQTKS_F1941A; **B**, sequencing of the PCR product containing the mutation site F1941A.

After completing the fermentation and preparation of crude extracts, we proceeded to analyze the products using LCMS. We also conducted the examination of the UV spectra and ESI between the

transformants containing SQTKS_F1941A/I2147A/F2157V and the SQTKS_WT transformant. In the SQTKS_F1941A/I2147A/F2157V transformants, we were able to confirm the presence of compounds **25** and **25a** through DAD and ESI analysis. However, the amounts of **25** and **25a**, as observed by both DAD and ESI, were relatively low in all transformants.

This analysis did not reveal any additional new products, but the F1941A/I2147A/F2157V mutation appeared to significantly reduce the activity of SQTKS, leading to a notable decrease in the production of **25** and **25a** (Figure 2.3.27).

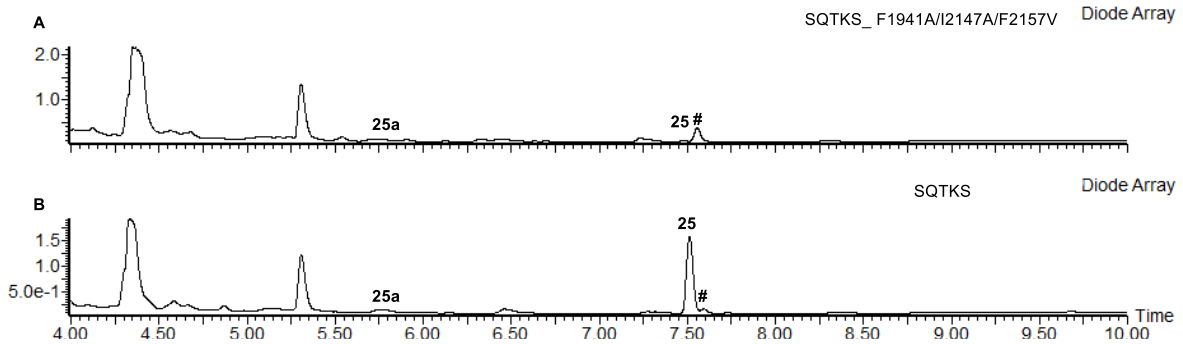


Figure 2.3.27 Chromatograms of WT (A) and F1941A/F2157A (B) by DAD.

Furthermore, we expressed other SQTKS mutants with the following mutations: I2147A/F2157V, I1938A, L2146A, L2146V, I2147A, and L2146A/I2147A. Several of these mutants had been previously evaluated through *in vitro* assays conducted within isolated ER enzyme (Table 2.3.5). We employed HPLC equipped with DAD and ESI to analyze the products, but no new products were detected in these experiments.

Table 2.3.5 All the mutagenesis residues and the products in this project.

Mutagenesis Residue/WT	Pocket Volume Å^3	Product
WT	1383	**25, 25a**
F1941A (A)	1405	**25, 25a**
F2157A (B)	1444	**25, 25a**
I2147A/F2157V (C)	1317	**25, 25a**
F1941A/F2157A (D)	1467	**25, 25a**
F1941A/I2147A/F2157V (E)	1392	**25, 25a**
I1938A (F)	1480	**25, 25a**
L2146A (G)	1367	**25, 25a**
L2146V (H)	1376	**25, 25a**
I2147A (I)	1307	**25, 25a**
L2146A/I2147A (J)	1269	**25, 25a**

2.4 Conclusion and Discussion

2.4.1 Optimizing the Expression Host

To enhance the production of tetraketide **25** through SQTKS expression in the *A. oryzae* NSAR1, we introduced the co-expression of two genes encoding predicted hydrolases. These hydrolases were expected to assist in the off-loading of polyketides from the PKS enzyme. Previously, SQTKS had been expressed in *A. oryzae* M-2-3 and fermented in a CDS medium, [137] yielding a titer of approximately 0.8 mg/L. In our current study, the fermentation of transformants was carried out in the DPY medium. We conducted time-course experiments over a period of 7 days and measured the titer each day. We observed that the titer reached its peak on day 3 and subsequently began to decrease. However, tetraketide **25** did not accumulate between day 3 and day 7. This could potentially be attributed to the host, *A. oryzae* NSAR1, diverting or degrading the product, as evidenced by the presence of **25a**. Therefore, it appears that host engineering may be necessary to reduce the shunt of the host.

The inclusion of Mfm10, either alone or in combination with Mfm8, did not result in a substantial improvement in titer when compared to the expression of SQTKS on its own. With the individual expression of SQTKS, we achieved a titer of approximately 2.9 mg/L, and when Mfm10 was added, the titer increased marginally to 3.2 mg/L. However, this increment did not represent a significant improvement and can not have a substantial impact on the subsequent experiments.

The involvement of *trans*-acting hydrolases in facilitating the off-loading of PKS products is relatively rare, especially when dealing with highly reducing PKS. Despite the fact that the Mfm8 protein shares an 80% sequence identity with Clz11,[148] a protein known to facilitate main-chain release in the zaragozic acid A pathway, our results did not provide evidence that Mfm8 assists in the release of tetraketide **25**. It's possible that Mfm8 might also contribute to main-chain release, but not to side-chain tetraketide **25** release. Additionally, the inclusion of Mfm10 did not lead to any significant improvement in titer, indicating that it did not enhance the production.

2.4.2 Engineering the Programming of Intact SQTKS

Some approaches have been investigated for the generation of novel polyketides by modifying hrPKS, including methods like domain swapping.[151] Another approach entails the modification of active sites to induce substrate specificity or broaden the range of accepted substrates. In this project, we introduced mutations into the intact SQTKS to observe the resulting products in *A. oryzae* NSAR1. However, these mutations failed to alter the programming of the SQTKS, as no new products were identified, and tetraketide **25** continued to be produced.

One potential explanation lies in the extraction process, as the products might be exported either into the surrounding medium or retained within the cells. It is also possible that the product exists in extremely low concentrations and may be lost during the ethyl acetate workup. Therefore, a more precise

detection method is required, along with the quantitative assessment of mutated SQTKS enzyme activities. Quantifying both the product and SQTKS enzyme activity proves challenging when employing heterologous expression in *A. oryzae*. This is primarily due to the variation in yields among different transformants, with the workup process during chemical extraction significantly affecting product quantities.

Ideally, if we could isolate the entire SQTKS protein and incubate it *in vitro*, product detection, enzyme activity measurement, and precise quantification would become more feasible. Regrettably, we have not yet succeeded in obtaining a soluble and active SQTKS enzyme, despite previous attempts by former colleagues who explored various methods, including *E. coli* and *S. cerevisiae*.

3. Discovery of Tailoring Enzymes for Modification of Aristolochene

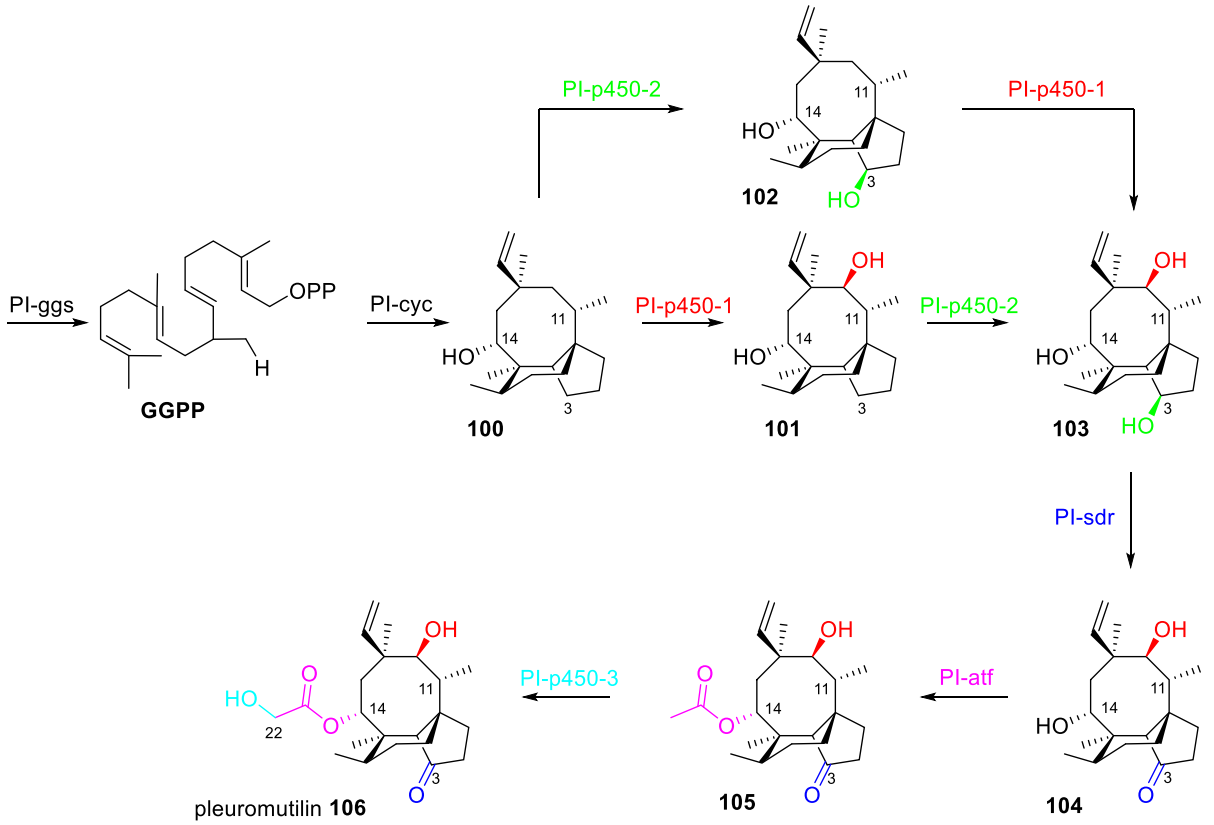
3.1 Introduction

3.1.1 Total Biosynthesis in *Aspergillus oryzae*

Total biosynthesis can be achieved either by introducing selected genes into a host organism or by directly manipulating an organism's DNA, resulting in the systematic synthesis of specific target compounds. The heterologous expression strategy, which can be employed in plants, fungi, and bacteria, streamlines the rapid assembly and expression of biosynthetic gene clusters (BGCs) responsible for the production of specialized metabolites. For example, the contemporary antibacterial agent pleuromutilin **106** can be efficiently synthesized through an expression single-process in the fungal host *Aspergillus oryzae*.[152,153]

The pathway leading to pleuromutilin **106** is initiated by geranylgeranyl diphosphate synthase (Pl-ggs), which synthesizes geranylgeranyl diphosphate (GGPP). Subsequently, the mutilin cyclase (Pl-cyc), a bifunctional diterpene synthase, orchestrates the cyclization of GGPP, resulting in the formation of the initial tricyclic intermediate, 3-deoxo-11-dehydroxymutilin **100**. Next, Pl-p450-1 introduces a hydroxy group at C-11, yielding **101**, and then Pl-p450-2 further modifies the molecule by adding another hydroxy group at C-3, resulting in **103**. Alternatively, Pl-p450-2 can initially add a hydroxy group at C-3 to form **102**, followed by subsequent reaction from Pl-p450-1 to produce **103** (Scheme 3.1.1).[153]

The hydroxy group at C-3 undergoes oxidation to form a keto group, a process facilitated by the short chain dehydrogenase reductase (SDR) Pl-sdr, leading to the production of mutilin **104**. Pl-atf, an acetyltransferase, then appends an acetate group to O-14, yielding 14-*O*-acetylmutilin **105**. Finally, Pl-p450-3 hydroxylates the C-22 acetate methyl group, culminating in the formation of pleuromutilin **106** (Scheme 3.1.1). These enzymatic processes offer a highly competitive alternative to total chemical synthesis, where the current best route to pleuromutilin **106** necessitates a minimum of 16 chemical steps, along with a multitude of reagents and solvents.[154]



Scheme 3.1.1 Proposed biosynthetic route of pleuromutilin **106**

3.1.2 Biosynthesis of PR-toxin

PR-toxin is a toxic fungal metabolite produced by specific strains of *Penicillium roqueforti* used in the maturation of French Roquefort cheese.[155] This compound was initially isolated and partially structurally elucidated by Wei et al.[156,157] It is a bicyclic sesquiterpene with various functional groups, including acetoxy (C-3), aldehyde (C-12), ketone (C-8), and a double bond (C-9 and C-10). Additionally, it features two stable epoxide rings (C-1 and C-2, C-7 and C-11) and belongs to the eremophilane terpenoid family. PR-toxin **107** originates from aristolochene **66**, a 15-carbon backbone, catalyzed by aristolochene synthase.[113] In 2015, the Dickschat group employed two-dimensional NMR techniques such as ^{1}H, ^{1}H-COSY, HSQC, HMBC, and NOESY to reevaluate and precisely correct the chemical shifts in the PR-toxin's structure.[91]

While PR-toxin **107** production has been reported in *Penicillium chrysogenum*, most of the biosynthetic research has primarily focused on *Penicillium roqueforti*.[158] *P. roqueforti* is known for generating several secondary metabolites, some of which have been identified and characterized. These metabolites include eremofortin A **114**, eremofortin B **113**, and eremofortin C **115**, all of which have been proposed as potential intermediates within the PR-toxin biosynthesis pathway.[159] Additionally, researchers have uncovered backbone compounds **66** and oxygenated eremophilanes **108-112** (Figure 3.1.1).[160]

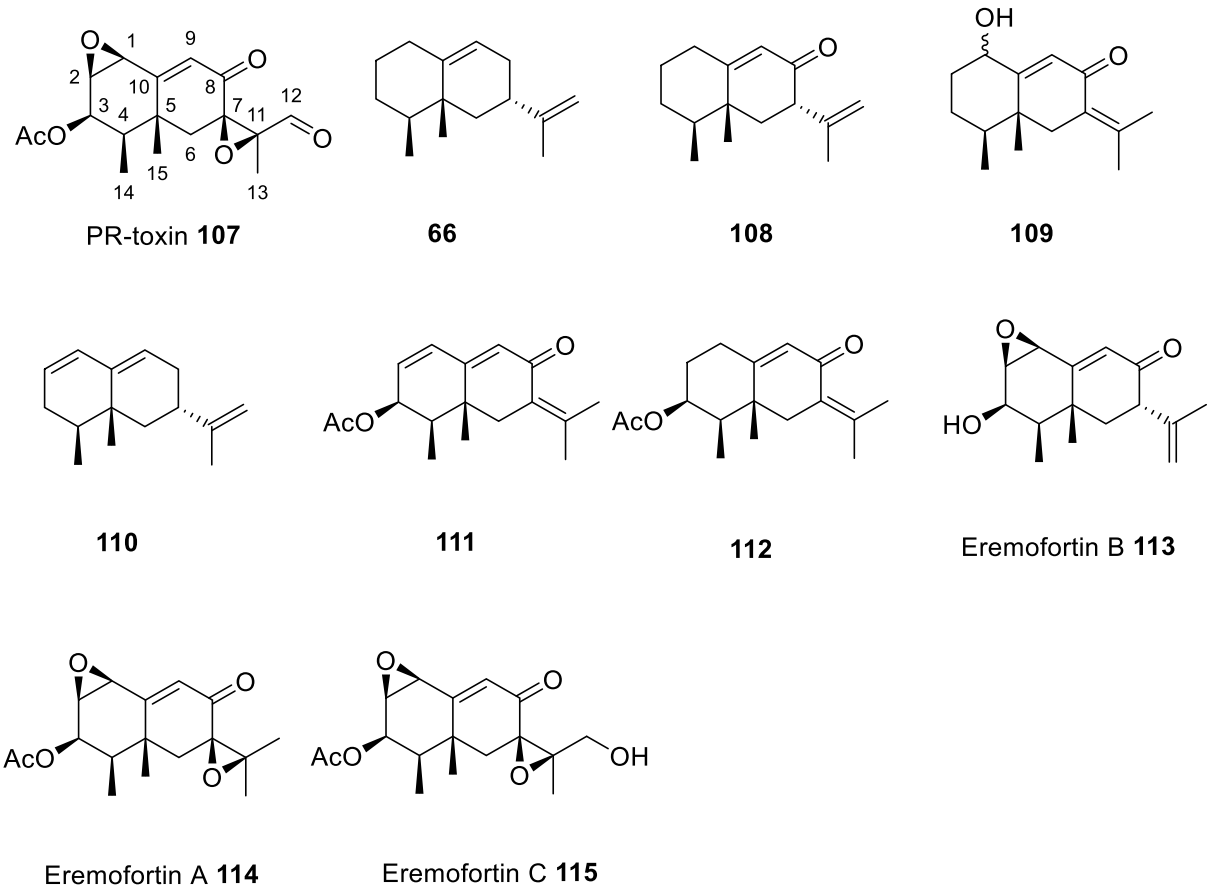


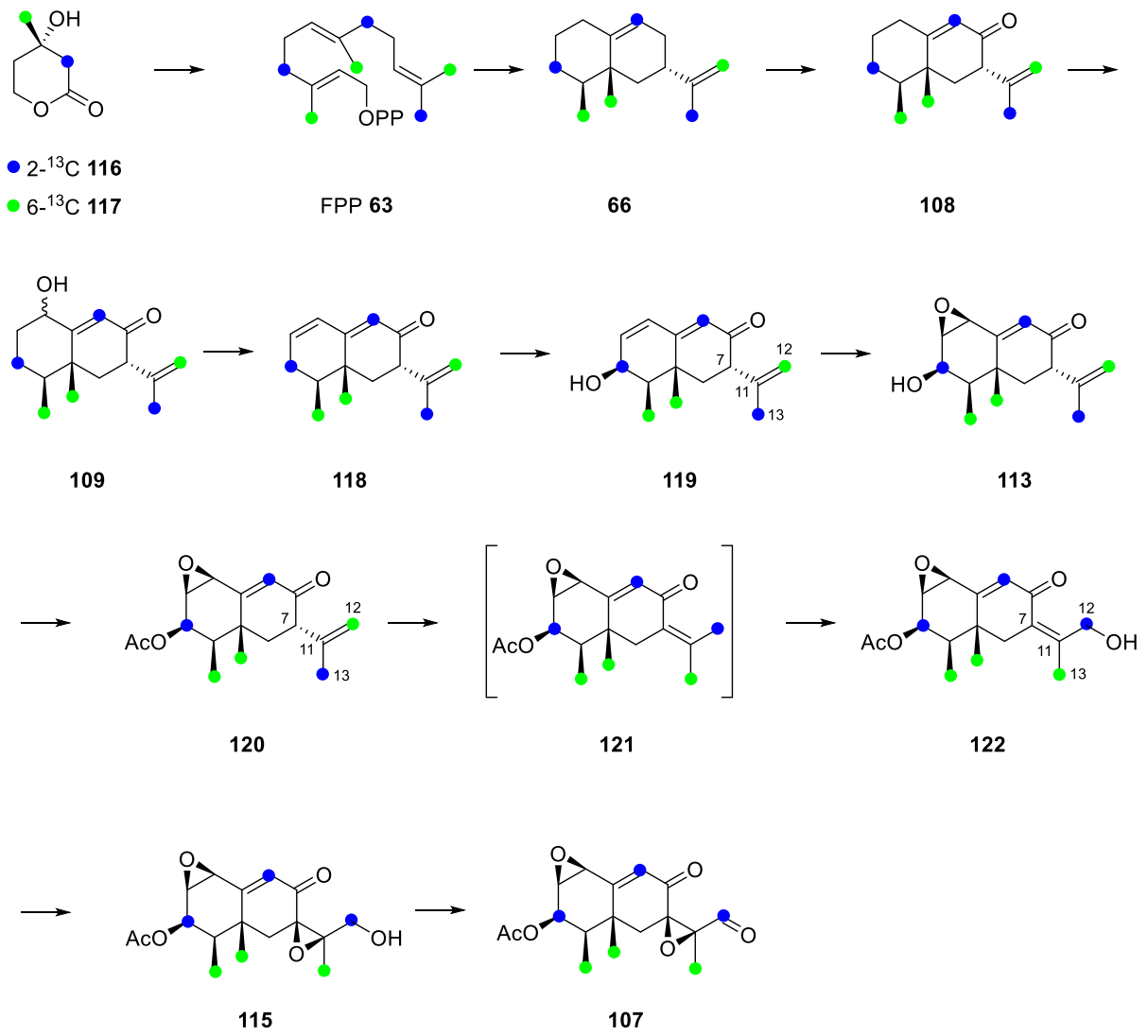
Figure 3.1.1 Structures of PR-toixn and related compounds

Although there have been recent advancements in acquiring partial information about the biosynthetic gene cluster responsible for PR-toxin **107**,[161] there is still a significant knowledge gap when it comes to understanding the exact roles of the genes and enzymes responsible for converting pathway intermediates. To address this, the Dickschat research group conducted feeding experiments using isotopically labeled precursors, offering valuable insights into the PR-toxin biosynthetic pathway.[91,162]

The backbone compound aristolochene **66** may undergo initial oxidation, resulting in the formation of **108**. It is conceivable that **108** could be transformed into **118** through the introduction of an olefin. During this step, **118** may originate from the oxidative process accompanied by the elimination of water from **109**, a phenomenon detectable in the secondary metabolites of *P. roqueforti*. Subsequently, **118** can be oxidized at C-3 to yield **119**. The epoxidation of **119** may lead to the production of eremofortin B **113**. An acetyl group is then added to the hydroxyl group at C-3 to generate **120**. A double-bond isomerization event may follow, allowing **120** to transform into **121**, which is subsequently oxidized to form **122** (Scheme 3.1.2).

Analysis of PR-toxin **107** through ^{13}C-NMR revealed that C-12 was enriched by (2-^{13}C)-mevalonolactone, while C-13 was enriched by (6-^{13}C)-mevalonolactone. These findings suggest that C-12 in **120** must undergo conversion to become C-13 in **121** and **122**, while C-13 in **120** becomes C-12

in **121** and **122**. This observation challenges the notion that **120** is directly transformed into **122** via a C7-C11-C12 allyl radical mechanism. Instead, **121** may be spontaneously generated through isomerization and rotation around the C7–C11 bond of **120**. Subsequent epoxidation gives rise to eremofortin C **115**, and further oxidation ultimately leads to the formation of PR-toxin **107** (Scheme 3.1.2).



Scheme 3.1.2 Proposed biosynthetic patyway of PR-toxin **107**. [91,162]

The PR-toxin biosynthetic pathway has recently been investigated in both *P. chrysogenum* and *P. roqueforti* following the functional characterization of select biosynthetic genes. The overall organization of these pathways in *P. chrysogenum* and *P. roqueforti* exhibits notable similarities. In the genome of *P. chrysogenum*, a gene cluster comprising 17 genes was identified. However, in the genome of *P. roqueforti*, the biosynthetic gene cluster (BGC) was found to be divided into two distinct regions: one encompassing *prl10 to prr1*, and the other containing *prr2 to prr8*. Among the encoded proteins are key enzymes, including aristolochene synthase, which plays a pivotal role in the initial cyclization of

farnesyl-diphosphate. Additionally, the pathway involves an oxidoreductase, an oxidase, two P450 monooxygenases, a transferase, and two dehydrogenase enzymes (Figure 3.1.2).

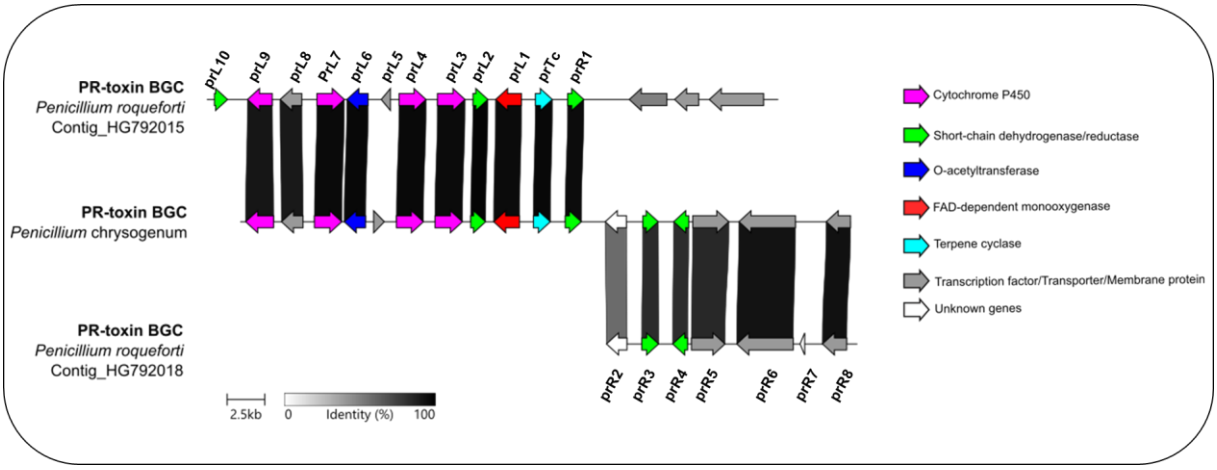


Figure 3.1.2 Proposed biosynthetic patyway of PR-toxin

3.2 Project Aims

Eremophilane sesquiterpenoids have attracted much attention because many of them possess a range of biological or therapeutic activities, including cytotoxic and antitumor antimicrobial, anti-inflammatory, antiviral and antiallergic effects.[77] Moreover, there is a limited amount of research that has been conducted about the discovery of the biosynthesis of eremophilane sesquiterpenes.

Numerous biosynthetic pathways exhibit a shared foundational carbon skeleton that later diverges as a result of the unique modifications introduced by tailoring enzymes. The evolution of these biosynthetic pathways is likely driven by the acquisition and loss of genes encoding these enzymes.[163] This phenomenon is exemplified by four compounds: sporogen AO-1 **27** (produced by *A. oryzae* and *A. flavus*),[80,81] hypoxylan A **73** (from *Hypoxylon rickii*),[104] eremoxylarin D **123** (from *Xylaria hypoxylon*),[164] and PR-toxin **107** (from *Penicillium roquefortii*),[160] all originating from the aristolochene skeleton but undergoing diverse tailoring reactions (Figure 3.2.1). The formation of aristolochene **66**, catalyzed by the terpene cyclase aristolochene synthase using farnesyl-diphosphate (FPP), involves a single step.[165] Genomes were assembled and automatically annotated, with manual annotation applied when necessary to resolve ambiguities, following a previously described informatic pipeline.[166,167] Notably, the biosynthetic gene cluster responsible for PR toxin biosynthesis in *P. roquefortii* (*prl1-prl9*) has been characterized.[161] This knowledge guided the search for analogous BGCs in the genomes of the other three organisms, shedding light on the intricate evolution of these biosynthetic pathways.

If the hypothesis that natural biosynthetic pathways gain or lose functions by the gain or loss of biosynthetic genes is correct, we reasoned that it should be possible to construct artificial BGCs that

combine skeleton and tailoring genes from different pathways for the rational synthesis of new specialised metabolites.

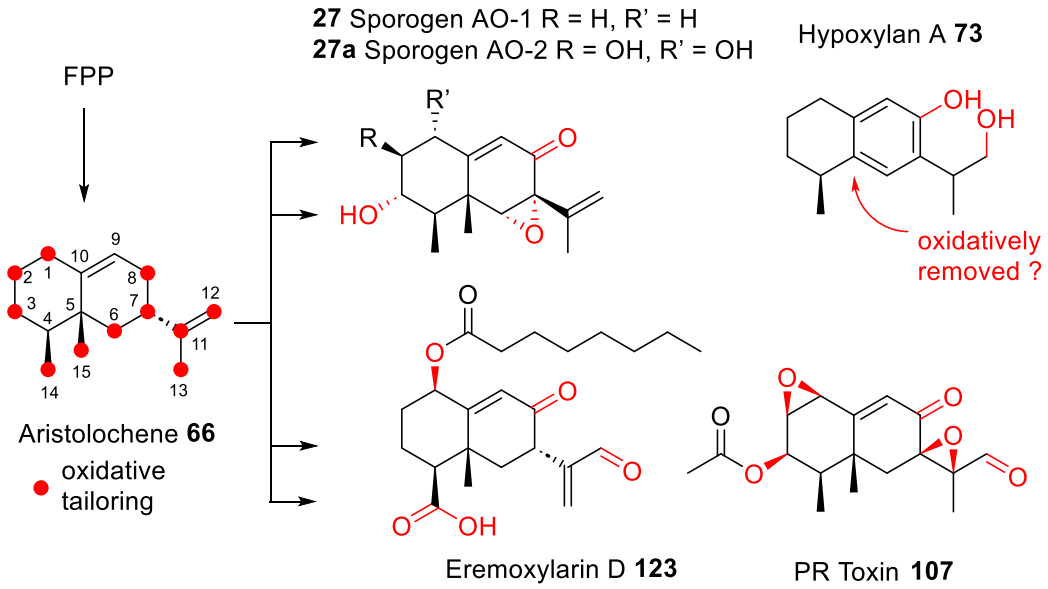


Figure 3.2.1 Terpenoids in fungi with aristolochene **66** as a backbone in common.

Our research therefore aims to increase the diversity of eremophilane sesquiterpenes by identifying key taioring genes, such as P450 oxygenases, capable of modifying the core aristolochene skeleton. To discover potential enzymes for this study, we are exploring both public and in-house databases containing fungal species information. We will employ various bioinformatic tools, including C-blaster and C-linker to investigate the possible interconnections between distinct gene clusters. Through gene manipulation techniques, we will identify a sufficient quantity of suitable tailoring genes and co-express them with the core genes. *A. oryzae* NSAR1, selected as our heterologous expression host, will undergo chemical-mediated transformation with the overall aim of synthesising new specialised metabolites.

3.3 Results

3.3.1 Bioinformatic Analyses

3.3.1.1 Hypoxylan-A BGC

Researchers affiliated with CeBiTec utilized a hybrid approach that combined Oxford Nanopore and Illumina technologies to sequence the genome of *Hypoxylon rickii* (hr).[168] Gene prediction was performed using Augustus version 3.2 and GeneMark-ES version 4.3.6, maintaining their default parameters.[169,170] Species-specific parameter sets for Augustus were established based on predictions generated by GeneMark-ES fungal version. Functional annotation of the predicted genes was carried out using an advanced iteration of the GenDB 2.0 genome annotation platform, specifically tailored for eukaryotic genomes. This comprehensive methodology has been previously detailed in literature.[171,172]

The *Hypoxylon rickii* protein database was manually constructed within the Geneious platform. To initially identify candidate gene clusters, we conducted manual BLASTP searches, employing the terpene cyclases aristolochene synthase from *Aspergillus terreus* (Attc) and *Penicillium roqueforti* (Prtc) as templates.[173,174].

When we used Attc as a template, Geneious identified five candidate cyclase proteins. Among them, 698_t displayed a 60.7% identity with Attc, while Hr2g6183.t1 exhibited a 67.1% identity with Attc (Table 3.3.1). Unusually, 698_t and Hr2g6183.t1 displayed 100% identity, indicating that they are identical genes with distinct annotations (Table 3.3.1).

Table 3.3.1 Blast results by Attc as a query against Hr genome.

% Identity	Attc	698_t	Hr2g6183.t1	8626_t	Hr2g4332.t1	11271_t
Attc		60.7	67.1	23.2	23.2	23.1
698_t	60.7		100	18.2	18.2	16.4
Hr2g6183.t1	67.1	100		18.1	18.1	16.4
8626_t	23.2	18.2	18.1		100	26.8
Hr2g4332.t1	23.2	18.2	18.1	100		26.8
11271_t	23.1	16.4	16.4	26.8	26.8	

In the case of the Prtc template, 698_t showed 51.0% identity with Prtc, while Hr2g6183.t1 demonstrated 56.6% identity with Prtc. This suggests that a candidate aristolochene synthase from *H. rickii* bears a closer resemblance to Attc than Prtc. Based on the percentage of identical sites, it can be concluded that there is only one aristolochene synthase (Hrtc) in the *H. rickii* genome (Table 3.3.2).

Table 3.3.2 Blast results by PrTc as a query against Hr genome.

% Identity	Prtc CDS	698_t	Hr2g6183.t1
Prtc CDS		51.0	56.6
698_t	51.0		100
Hr2g6183.t1	56.6	100	

The protein sequence of Hrtc was acquired and subsequently compared with Attc and Prtc, resulting in the generation of an alignment map (Figure 3.3.1). Hrtc consists of 339 amino acids, making it longer than Attc, which comprises 320 amino acids, and shorter than Prtc, which comprises 342 amino acids. Among these three protein sequences, they share 165 identical sites, resulting in a pairwise identity of 57%. Additionally, their pairwise positivity, calculated using BLSM62, stands at 71.4%. Notably, several active sites found in Prtc are conserved in Hrtc, including residues such as T89, Y92, D115, D116, N244, E252, and others.[165] Furthermore, the "aspartate-rich" segment DDLLE, which begins with Asp 90 in Attc, exhibits a high degree of conservation.[175–177]

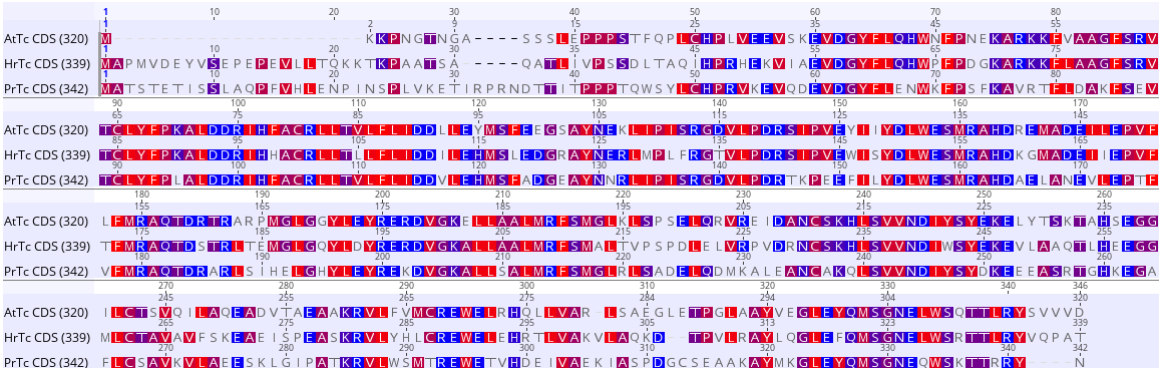


Figure 3.3.1 Alignment of aritolochene synthase protein sequences from *A. terreus* (Attc) *H. rickii* (Hrtc) and *P. roqueforti* (Prtc). Regions with high similarity across all sequences are labeled with colors.

Once we identified the candidate aristolochene synthase, designated as Hrtc (at locus tag Hr2g6183.t1), we conducted a thorough search for the location of the candidate BGC within the genome of *H. rickii* using the Geneious software. We precisely captured a 20-kilobase DNA segment on both sides of the located gene to encompass the entire genomic region of interest. Subsequently, this precise region underwent a meticulous editing and annotation process (Figure 3.3.2). To gain deeper insights into the gene function of the entire gene cluster, we subjected it to an additional BLAST analysis within public databases.

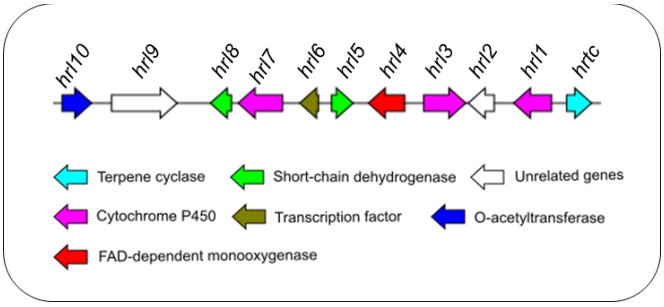


Figure 3.3.2 The proposed gene cluster for biosynthesis of hypoxylan A **73**.

Three genes (*hrl1*, *hrl3*, *hrl7*) were annotated to encode cytochrome P450 monooxygenases, each containing a conserved domain superfamily (Table 3.3.3). Two genes (*hrl5*, *hrl8*) were annotated as encoding short-chain dehydrogenase/reductases, featuring a Rossmann-fold nicotinamide cofactor

binding domain (Table 3.3.3). *Hrl4* was associated with the function of encoding an FAD-dependent monooxygenase, as indicated by the FAD/FMN-containing dehydrogenase hit. *Hrl10* was annotated as encoding an *O*-acetyltransferase. *Hrl6* was linked to a transcription factor function, encompassing a GAL4-like Zn_2Cys_6 binuclear cluster DNA-binding domain (Table 3.3.3). Notably, *Hrl2* and *hrl9* did not yield predicted functions within the available public database. Consequently, this gene cluster is presumed to be responsible for the biosynthesis of hypoxylan A **73**.

Table 3.3.3 Gene annotations of proposed BGC for hypoxylan A **73**.

Locus_tag	CDS Label	Predicted protein function	Domain hits
Hr2g6175	Hrl10	Trichothecene 8-O-acetyltransferase	Transferase
Hr2g6176	Hrl9	Unnamed protein product	/
Hr2g6177a	Hrl8	Short-chain dehydrogenase;	FabG; ADH_SDR_c_like
Hr2g6177b	Hrl7	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
Hr2g6177c	Hrl6	Transcription factor	$Zn(II)_2Cys_6$ transcription factor
Hr2g6178	Hrl5	Short-chain dehydrogenase	Rossmann-fold NAD(P)-binding domain
Hr2g6179	Hrl4	FAD-dependent monooxygenase	FAD_binding_4
Hr2g6180	Hrl3	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
Hr2g6181	Hrl2	SAT4 family membrane protein	No putative conserved domains
Hr2g6182	Hrl1	Cytochrome P450 monooxygenase	CYP503A1-like; CypX
Hr2g6183	Hrtc	Aristolochene synthase	Terpene_cyclase_nonplant_C1; Terpene_syn_C_2

3.3.1.2 Eremonxylarin D-like BGC

Cblaster was utilized to search for potential gene clusters within our in-house database, using the PR-toxin **107** BGC as a reference point (Figure 3.3.3). This exploration revealed the presence of two BGCs with the potential to synthesize eremonxylarin D **123** in *Xylaria hypoxylon*, a species known for its production of eremonxylarin D[178] (Figure 3.3.3). Subsequently, these candidate BGCs were precisely identified and extracted using the Geneious platform (Figure 3.3.4).

Xylaria species are known to have the capacity to produce a range of eremonxylarin-type natural products, including eremoxylarins D-J.[83,98–100,178,179] Eremoxylarin D **123** shares a structural backbone with PR-toxin **107** and Sporogen AO-1 **27**, but it also possesses an additional sidechain composed of a polyketide. Eremoxylarin D **123** is hypothesized to share similar early biosynthetic steps with PR toxin, as their backbones are both constructed through the action of a cyclic terpene aristolochene synthase, and some modifications, such as oxidation at C-8, are involved in tailoring.

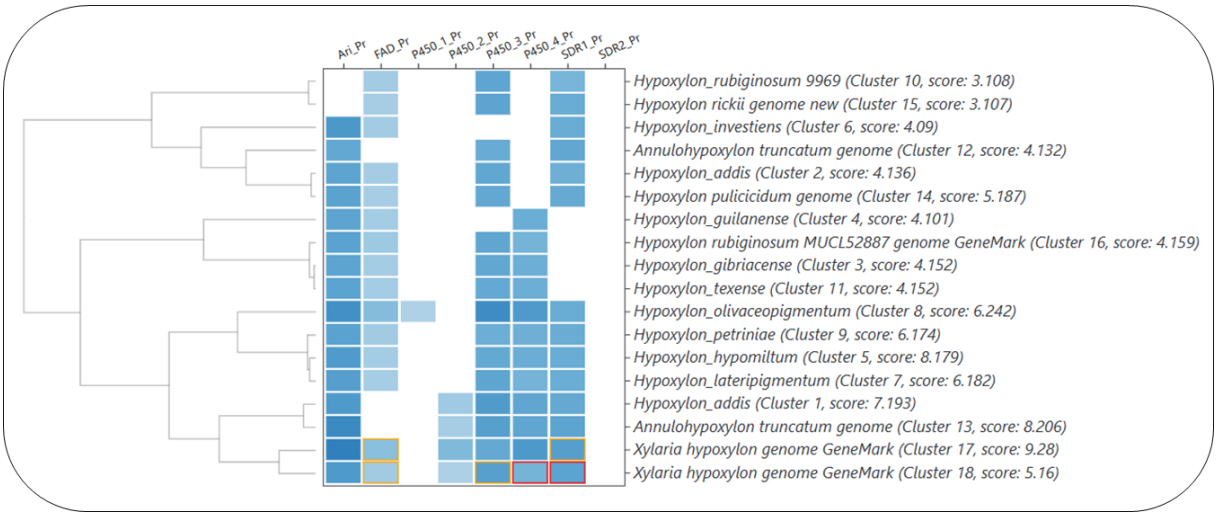


Figure 3.3.3 Gene cluster mining against in-house database by the reference PR-toxin BGC using Cblaster.

A gene cluster A was identified from *Xylaria hypoxylon* with an aristolochene synthase encoding gene. Remarkably, it exhibits a high degree of homology with the PR-toxin BGC. This particular BGC encompasses several key genes, including genes that encode: aristolochene synthase; four P450 oxygenases; two SDRs; one FMOs; one hrPKS; an *O*-AcT; and a putative transporter and a transporter (Figure 3.3.4).

Another gene cluster B (Figure 3.3.4) was also found. It has genes that encode: aristolochene synthase; six P450 oxygenases; two SDRs; two FMOs; an esterase; an *O*-AcT; and two putative transporters. Despite the presence of two candidate gene clusters harboring the aristolochene gene within the genome of *X. hypoxylon*, it's noteworthy that the hrPKS is specifically situated within gene cluster A (Figure 3.3.4). This observation serves as compelling evidence for directing our focus towards the gene cluster A associated with eremoxylarin D **123** biosynthesis.

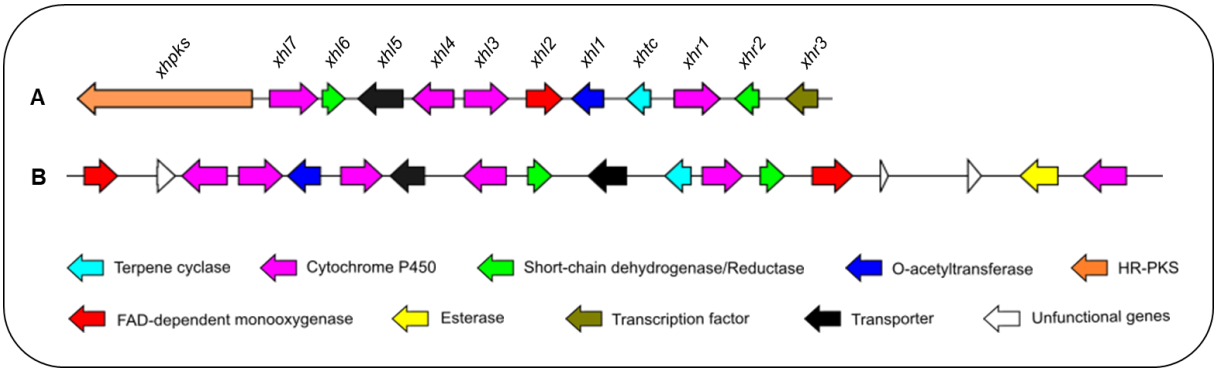


Figure 3.3.4 Gene clusters found from *Xylaria hypoxylon* genome by the query BGC of PR-toxin.

We utilized BLASTp against the NCBI database to annotate and predict the putative functions of all proteins encoded by the candidate BGC (Table 3.3.4). In addition to the expected aristolochene synthase, several other genes were identified, including four P450-encoding genes (*xhr1, xhl3, xhl4, xhl7*), two SDR-encoding genes (*xhr2, xhl6*), an FMO-encoding gene (*xhl2*), two transporter-encoding genes (*xhl5,*

xhl9), a transcription factor (*xhr3*), an *O*-AcT-encoding gene (*xhl1*), and a gene encoding an hrPKS (*xhPKS*). Each of these genes exhibited hits corresponding to their respective functional domains.

Table 3.3.4 Gene annotations of eremoxylarins D BGC

Locus_tag	CDS Label	Predicted protein function	Domain hits
7855_g	Xhl9	Cytosine/purine transport protein	CodB; SLC-NCS1sbd_CobB-like
7856_g	Xhl8	Highly reducing polyketide synthase	PKS_KS; PS-DH; PKS_KR; Enoyl_red; PP-binding; PKS_MT
7857_g	Xhl7	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
7858_g	Xhl6	Short-chain dehydrogenase	FabG; ADH_SDR_c_like
7859_g	Xhl5	MFS-type transporter, drug resistance transporter	MFS_Azr1_MDR_like; Efflux_EmrB
7860_g	Xhl4	Cytochrome P450 monooxygenase	CYP56-like
7861_g	Xhl3	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
7862_g	Xhl2	FAD-dependent monooxygenase	FAD_binding_4
7863_g	Xhl1	Eremophilane *O*-acetyltransferase	Transferase; PLN02663
7864_g	Xhtc	Aristolochene synthase	Terpene_cyclase_nonplant_C1;
7865_g	Xhr1	Cytochrome P450 monooxygenase	CYP7_CYP8-like; CypX
7866_g	Xhr2	Short-chain dehydrogenase	Retinol-DH_like_SDR_c_like; PRK06197
7867_g	Xhr3	C6 finger domain transcription factor	GAL4-like $Zn(II)_2Cys_6$; Fungal Zn(2)-Cys(6)

3.3.1.3 Sporogen AO-1 BGC

Sporogen AO-1 **27**, which was isolated from *A. oryzae* and possesses bioactive properties,[80,81] has yet to undergo thorough investigation regarding its biosynthesis. The backbone of sporogen AO1 **27** shares identities with aristolochene **66**, albeit with some modifications involving oxidation. In pursuit of understanding its biosynthesis, we manually conducted homology searches against the genome of *A. oryzae* RIB40, using the protein sequence of aristolochene synthase from PR-toxin BGC (Prtc) as a reference. This search led to the identification of a protein sequence (XP_023093357) bearing 54.2% identity to the protein sequence of Prtc (Table 3.3.5). Subsequently, a gene cluster containing the potential aristolochene synthase (Aotc) was extracted from the genome of *A. oryzae* NSAR1 (Figure 3.3.5).

Table 3.3.5 Blast result by Attc as a query against genome of *A. oryzae* NSAR1.

% identity	Prtc	XP_023093357	XP_023093164	XP_001825829	XP_023093930	XP_023091495
Prtc		54.2	29	24.0	21.5	24.1
XP_023093357	54.2		23.5	21.7	17.8	17.8
XP_023093164	29	23.5		24.7	22.1	21
XP_001825829	24.0	21.7	24.7		22.8	12.3
XP_023093930	21.5	17.8	22.1	22.8		13.9
XP_023091495	24.1	17.8	21	12.3	13.9	

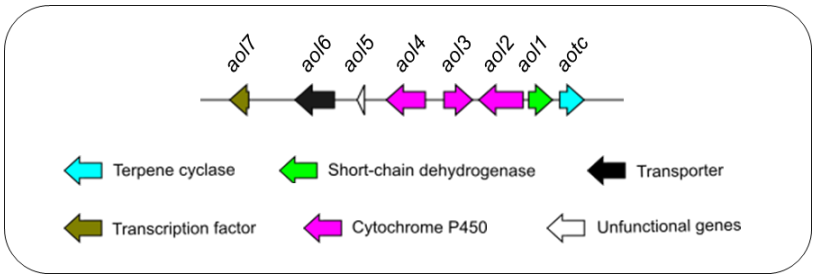


Figure 3.3.5 The proposed gene cluster for biosynthesis of sporogen AO-1 **27**.

We employed BLASTp against the NCBI database to predict and annotate the putative functions of all proteins encoded by the putative sporogen AO-1 BGC. In addition to the aristolochene synthase Aotc, six other genes were found encode proteins with predicted functions (Table 3.3.6). These include three genes (*aol2, aol3, aol4*) encoding P450 enzymes, one gene (*aol1*) encoding an SDR, a gene (*aol6*) encoding a drug resistance transporter protein, and a gene (*aol7*) encoding a transcription factor (Table 3.3.6).

Table 3.3.6 Gene annotations of sporogen AO1 BGC.

Locus tag	CDS Label	Predicted protein function	Domain hits
AO090011000096	Aol7	Transcription factor	Fungal_TF_MHR
AO090011000097	Aol6	Drug resistance transporter	MFS_Azr1_MDR_like; efflux_EmrB
AO090011000098	Aol5	Unnamed protein product	/
AO090011000099	Aol4	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
AO090011000100	Aol3	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
AO090011000101	Aol2	Cytochrome P450 monooxygenase	CYP60B-like; CypX; p450
AO090011000102	Aol1	Short-chain dehydrogenase	FabG; ADH_SDR_c_like; adh_short
AO090011000103	Aotc	Aristolochene synthase	Terpene_cyclase_nonplant_C1;

A global sequence alignment was then conducted among the hypoxylan A **73**, PR-toxin **107**, sporogen AO-1 **27**, and eremonxylarin D **123** BGCs using Clinker. This analysis revealed that the four distinct clusters exhibit a remarkably high degree of homology (Figure 3.3.6). The proteins encoded by the core genes (*aotc, hrtc, prtc, and xhtc*) were consistently conserved across all four clusters, displaying remarkable similarities of over 50% among them. These core proteins play a fundamental role in forming the central (+)-aristolochene skeleton.

Similarly, the genes responsible for P450 enzymes (*aol4, hrl3, prl7*, *xhl3*), were found to be preserved in all clusters. Conservation in all 4 clusters may indicate function in a common (and early) catalytic process such as the required oxygenation at C-8. Equally conserved were the genes encoding SDR enzymes (*aol1, hrl8, prl2*, *xhl6*), and we hypothesised that these enzymes may catalyse the conversion of C-8 alcohols to ketones. Collectively, these three shared genes contribute to the formation of skeleton **66**. Furthermore, three of the clusters shared genes coding for FAD-MO (*hrl4, prl1, xhl2*), which are likely responsible for modifying C-13 in the structures of hypoxylan A **73**, PR-toxin **107**, and eremonxylarin D **123**. The diversity of the final structures arises through the involvement of additional tailoring enzymes unique to each cluster (Figure 3.3.6).

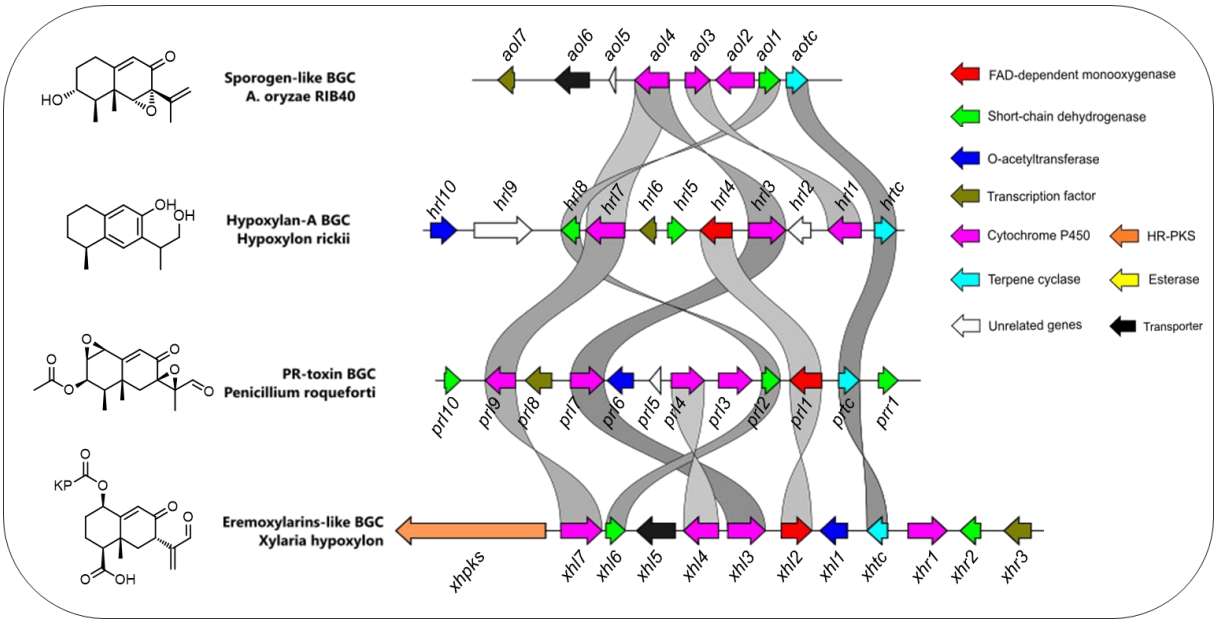
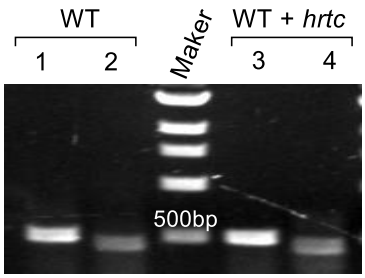


Figure 3.3.6 Alignment of four candidate gene clusters by Clinker.

3.3.2 RT-PCR for AoL4 and AoL1

Our established expression host, *A. oryzae*, traces its origins to a cultivated variant of *A. flavus*.[180] This specific variant is distinguished by the inactivity of several specialized metabolite BGCs. This attribute proves highly advantageous for heterologous expression experiments, primarily due to the absence of competition between the introduced biosynthetic pathways and the native counterparts for precursor molecules. Although *A. oryzae* appears to possess a putative BGC that could encode the synthesis of sporogen AO-1 **27**, this compound is not detectable in the fermentation extract of *A. oryzae* NSAR1 in our hands. The absence of **27** suggests potential damage to, or a lack of expression of, the aristolochene cyclase. However, it is plausible that the tailoring genes are still active. To delve deeper into this matter, we conducted reverse transcription polymerase chain reaction (RT-PCR) analyses, with a specific focus on *aol4* (encoding P450) and *aol1* (encoding SDR), which share similarities with genes identified in the three other gene clusters. This approach allowed us to assess the extent of their transcriptional expression.

Mycelia obtained from both *A. oryzae* NSAR1 and its transformants carrying *hrtc* were collected after a 5-day cultivation in DPY medium. These samples were promptly frozen in liquid nitrogen and then finely powdered. RNA extraction was carried out using the rapid-RNA fungal microprep kit (ZYMO). To convert the RNA into cDNA, we utilized the high-capacity RNA-to-cDNA kit. The resulting cDNA, derived from these procedures, served as templates for the amplification of *aol4* and *aol1*, producing fragments of approximately 450 to 500 base pairs in size (Figure 3.3.7).



1. The band of *aol4* in WT
2. The band of *aol1* in WT
3. The band of *aol4* in WT+*hrtc*
4. The band of *aol1* in WT+*hrtc*

Figure 3.3.7. RT-PCR for the expression level of *aol4* and *aol1*. Primers RT-PCR-*aol4*-F and RT-PCR-*aol4*-R were used for *aol4*; RT-PCR-*aol1*-F and RT-PCR-*aol1*-R were used for *aol1*.

The results of the RT-PCR analysis demonstrated the transcriptional activity of Aol4 and Aol1, regardless of whether it was in the untransformed strain or in the transformants carrying *hrtc* (Figure 3.3.7). *A. oryzae*, recognized for its clean host nature with limited interactions with heterologous pathways, has long been the preferred platform for pathway expression and engineering endeavors. However, the conspicuous activation of *aol4* and *aol1* could be beneficial since we hypothesise that all pathways require the chemical steps catlysed by these enzymes. This activation could potentially enhance the availability of precursors with the 8-oxo modification, required as starting materials for later enzymatic steps.

3.3.3 Plasmid Construction and Transformation of *A. oryzae*

3.3.3.1 Plasmid Constructions

All the plasmids in this study were constructed through yeast recombination and subsequently selected in *E. coli*. The DNA fragments derived from the hypoxylan A BGC were cloned from the cDNA of *H. rickii*. For the PR-toxin BGC, the DNA fragments, which included introns, were directly cloned from the gDNA of *P. roquefortii*. Likewise, the DNA fragments of the sporogen AO-like BGC were obtained through cloning from the gDNA of *A. oryzae* NSAR1. In the case of *xhr1*, exon DNA fragments were cloned from *Xylaria hypoxylon* gDNA and then recombined to create a coding DNA sequence through yeast recombination. All the constructed plasmids underwent thorough sequencing to ensure their correctness and accuracy.

The sequenced vectors were labelled with ID XX and used in different combinations for *A. oryzae* transformations (Table 3.3.7). The first plasmid, designated as XX01, was constructed by placing the aristolochene gene *hrtc* under the control of the P_{amyB} within the pTYGS_*arg* vector (Table 3.3.7). Next, a multi-gene plasmid, XX02, was assembled, incorporating genes *hrtc, hrl4, hrl5*, and *hrl8*. These genes were cloned and inserted into the pTYGS_*arg* vector under the control of different promoters: P_{amyB},

P_{adh}, P_{gpdA}, and P_{eno}, respectively (Table 3.3.7). Another multi-gene plasmid, XX03, was created, featuring genes *hrl3*, *hrl1*, and *hrl7*, which were cloned and inserted into the pTYGS_ade vector. These genes were regulated by P_{amyB}, P_{adh}, and P_{gpdA}, respectively (Table 3.3.7).

Table 3.3.7 Plasmids used in this study

Construct ID	Plasmids	Features
XX01	pTYGS_*arg-hrtc*	Hrtc under control of P_{amyB}, *argB* as the selection marker
XX02	pTYGS_*arg-hrtc-hrl4-hrl5-hrl8*	P_{amyB} induces Hrtc, P_{adh} induces Hrl4, P_{gpdA} induces Hrl5, P_{eno} induces Hrl8; selection marker is *arg*
XX03	pTYGS_*ade-hrl3-hrl1-hrl7*	P_{amyB} induces Hrl3, P_{adh} induces Hrl1, P_{gpdA} induces Hrl7; selection marker is *ade*
XX04	pTYGS_*arg-hrtc-hrl4 -hrl8*	P_{amyB} induces Hrtc, P_{adh} induces Hrl4, P_{eno} induces Hrl8; selection marker is *arg*
XX05	pTYGS_*arg-hrtc-hrl5-hrl8*	P_{amyB} induces Hrtc, P_{gpdA} induces Hrl5, P_{eno} induces Hrl8; selection marker is *arg*
XX06	pTYGS_*ade-hrl1-hrl7*	P_{adh} induces Hrl1, P_{gpdA} induces Hrl7; selection marker is *ade*
XX07	pTYGS_*ade-hrl3-hrl7*	P_{amyB} induces Hrl3, P_{gpdA} induces Hrl7; selection marker is *ade*
XX08	pTYGS_*arg-hrtc-hrl8*	P_{amyB} induces Hrtc, P_{eno} induces Hrl8; selection marker is *arg*
XX09	pTYGS_*ade-prl3-prl4*	P_{adh} induces Prl3, P_{gpdA} induces Prl4; selection marker is *ade*
XX10	pTYGS_*met-prl7-prl9*	P_{gpdA} induces Prl7, P_{eno} induces Prl9; selection marker is met
XX11	pTYGS_*ade-prl3*	P_{adh} induces Prl3; selection marker is *ade*
XX12	pTYGS_*ade-prl4*	P_{gpdA} induces Prl4; selection marker is *ade*
XX13	pTYGS_*met-prl7*	P_{gpdA} induces Prl7; selection marker is *met*
XX14	pTYGS_*ade-xhr1*	P_{gpdA} induces Xhr1; selection marker is *ade*

Plasmid XX04 is identical to XX02 but lacks the *hrl5* gene, while plasmid XX05 is similar to XX02 but lacks the *hrl4* gene (Table 3.3.7). Plasmid XX06, a variant of XX03, lacks the *hrl3* gene, and XX07, another variant of XX03, lacks the *hrl1* gene (Table 3.3.7). Plasmid XX08 is similar to XX05 but lacks the *hrl5* gene. XX09 contains *prl3* and *prl4* genes, each driven by the P_{adh} and P_{gpdA}, respectively. Additionally, *prl7* and *prl9* were introduced into pTYGS_met under the control of the P_{gpdA} and P_{eno}, respectively, forming plasmid XX10. *Prl3* was independently inserted into pTYGS_ade (XX11), while *prl4* found its place in pTYGS_ade as well (XX12). Plasmid XX13 carries *prl7*, inserted into pTYGS_met, and *xhr1* was cloned and placed within pTYGS_ade driven by P_{gpdA} (XX14, Table 3.3.7). For clarity, the features of all these plasmids are visually represented in corresponding cartoon maps (Figure 3.3.8).

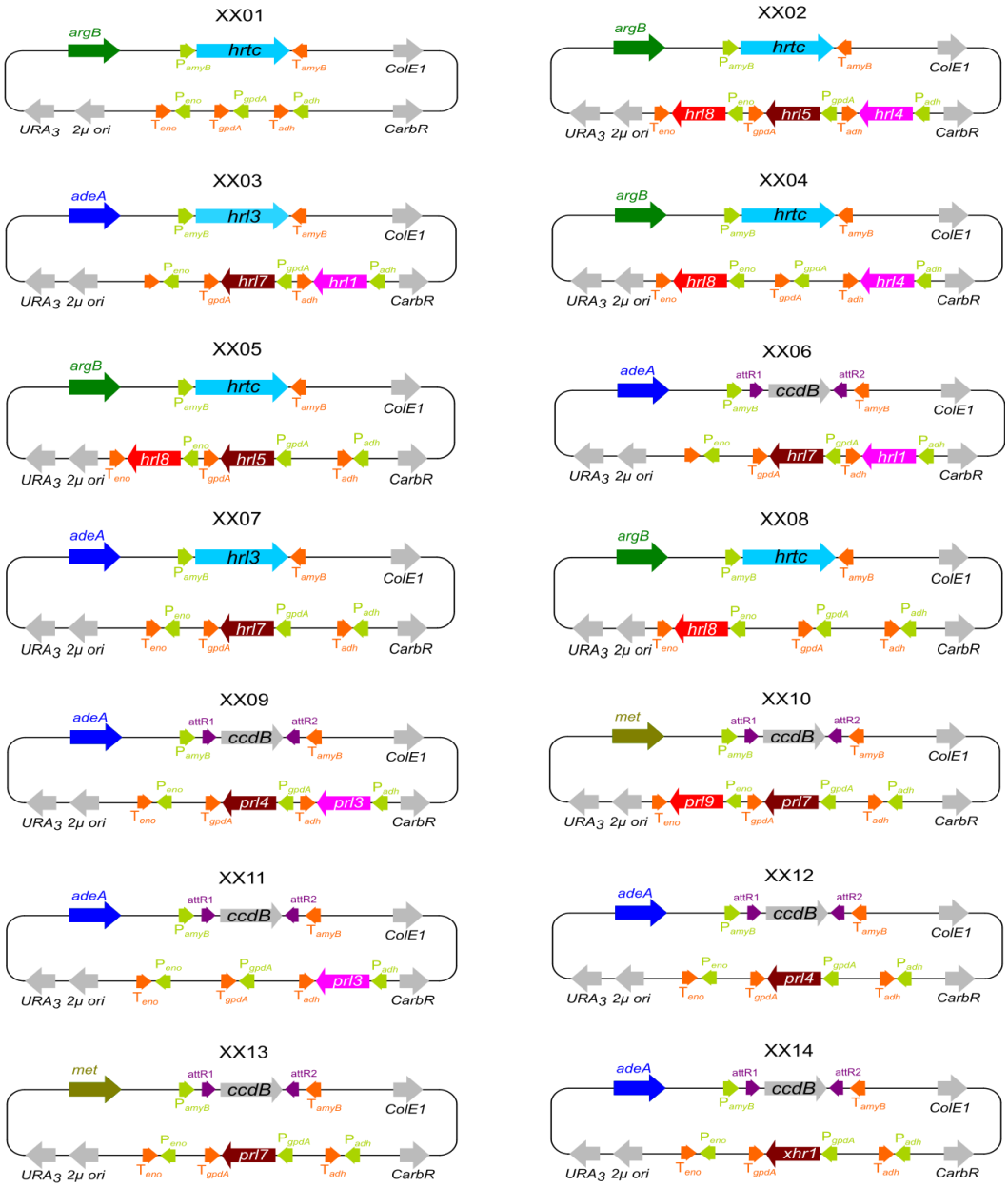


Figure 3.3.8 the maps of plasmids in this study

3.3.3.2 Gene Combinations and *A. oryzae* Transformation

Heterologous expression in *A. oryzae* was employed to elucidate the functions of genes originating from the four clusters and to generate additional compounds. In this study, the *A. oryzae* NSAR1 quadruple auxotrophic mutant (*niaD*$^{-}$, *sC*$^{-}$, △*argB*, *adeA*$^{-}$) was utilized as the host strain. The constructed vectors were introduced into *A. oryzae* NSAR1 following the methods outlined in Chapter 2, using various combinations (Table 3.3.8). After two rounds of screening, transformants were successfully obtained on the selection medium CZD/S. Subsequently, they were transferred to DPY medium to induce P_{amyB} during the fermentation. The culture broth of each transformant was then subjected to extraction and analysis using LCMS or GCMS techniques.

Table 3.3.8 Combinations of plasmids for each experimental group.

Exp	*hrtc*	*hrl1*	*hrl3*	*hrl4*	*hrl5*	*hrl7*	*hrl8*	*prl3*	*prl4*	*prl7*	*prl9*	*xhr1*	Plasmids
1	✓												XX01
2	✓	✓	✓	✓	✓	✓	✓						XX02+XX03
3	✓	✓	✓	✓		✓	✓						XX03+XX04
4	✓	✓	✓		✓	✓	✓						XX03+XX05
5	✓	✓		✓	✓	✓	✓						XX02+XX06
6	✓		✓	✓	✓	✓	✓						XX02+XX07
7	✓	✓	✓			✓	✓						XX03+XX08
8	✓	✓				✓	✓						XX06+XX08
9	✓							✓	✓	✓	✓		XX01+XX09+XX10
10	✓							✓					XX01+XX11
11	✓								✓				XX01+XX12
12	✓								✓	✓			XX01+XX12+XX13
13	✓											✓	XX01+XX14

3.3.4 Exp 1_Expression of *hrtc*

Hrtc was expected to function as an aristolochene synthase responsible for forming the skeleton structure. Plasmid XX01, containing the *hrtc* gene under the control of P_{amyB}, was introduced into *A. oryzae* NSAR1. Crude extracts of the mycelia from the transformants were analyzed using GCMS alongside a wild-type *A. oryzae* NSAR1 control (Figure 3.3.9A). During the analysis, a new peak emerged at 12.2 minutes. The mass spectrum of this compound revealed crucial information, including a molecular weight of 204 and main structural fragments at 55, 80, 91, 105, 121, 133, and 189 *m/z* (Figure 3.3.9B). This mass spectrum data was compared to the standard spectra of (+)-aristolochene **66** available in the literature (Figure 3.3.9C, D), confirming a close match.[165,181]

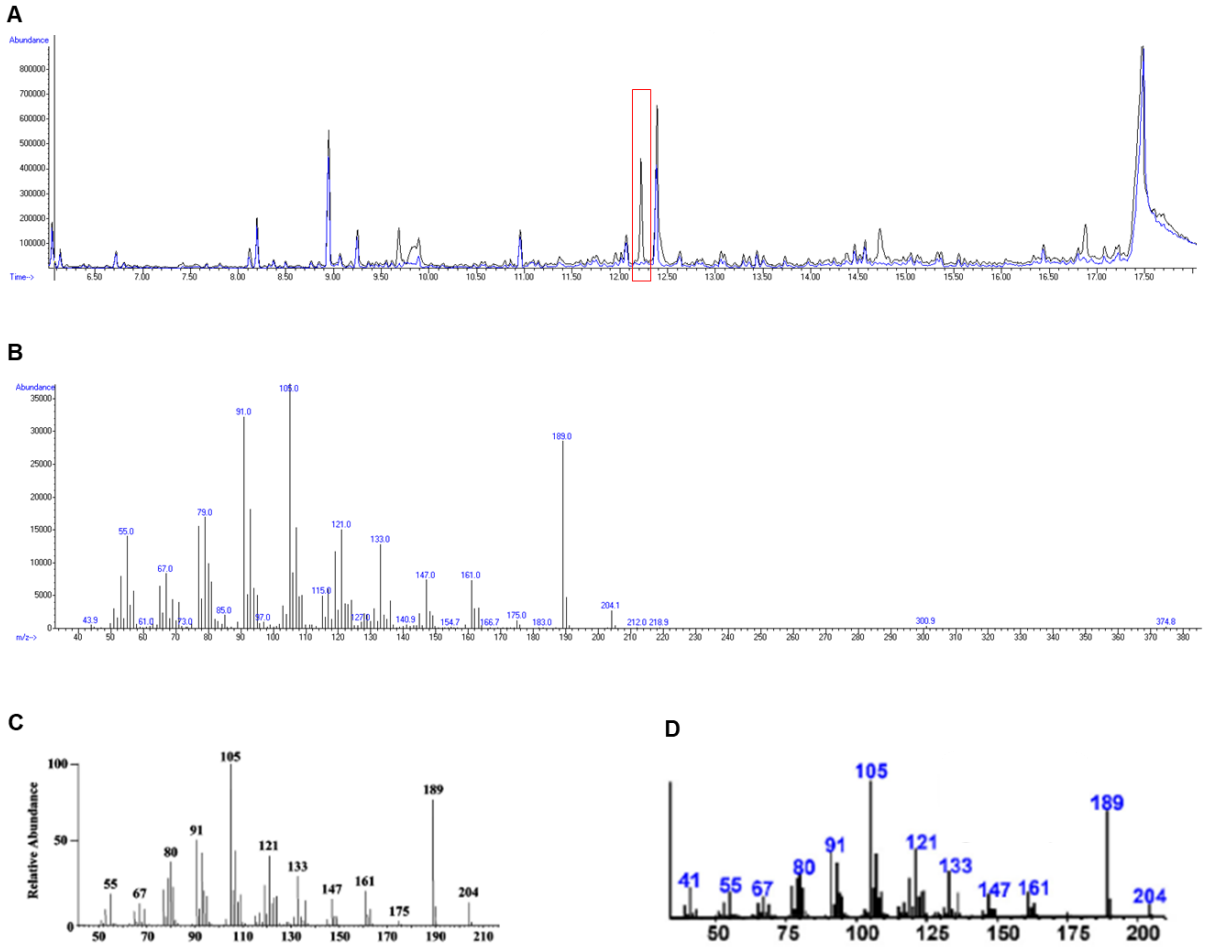


Figure 3.3.9. Detection of **66** by GC-MS; **A,** the chromatogram overlay of the control *A. oryzae* NSAR1 (in blue) and the transformant *A. oryzae* NSAR1+ *hrtc* (in black); the peak of **66** was labelled in a red box; **B,** mass information of **66; C** and **D,** standard spectrums of **66** from literature.

The crude extract, comprised of a mixture of cells and medium, underwent LCMS analysis, revealing the presence of two new peaks, designated as **124** and **125**, when compared to the control *A. oryzae* NSAR1 (Figure 3.3.10). To elucidate their chemical structures, a 1-liter scale-up fermentation of the **124** and **125** producing transformants was conducted. Subsequently, the resulting crude extract was employed for preparative LCMS, which facilitated the successful purification of **124** and **125**. The structures of **124** and **125** were conclusively determined as novel compounds through a comprehensive analysis involving 1D, 2D NMR, and HRMS techniques.

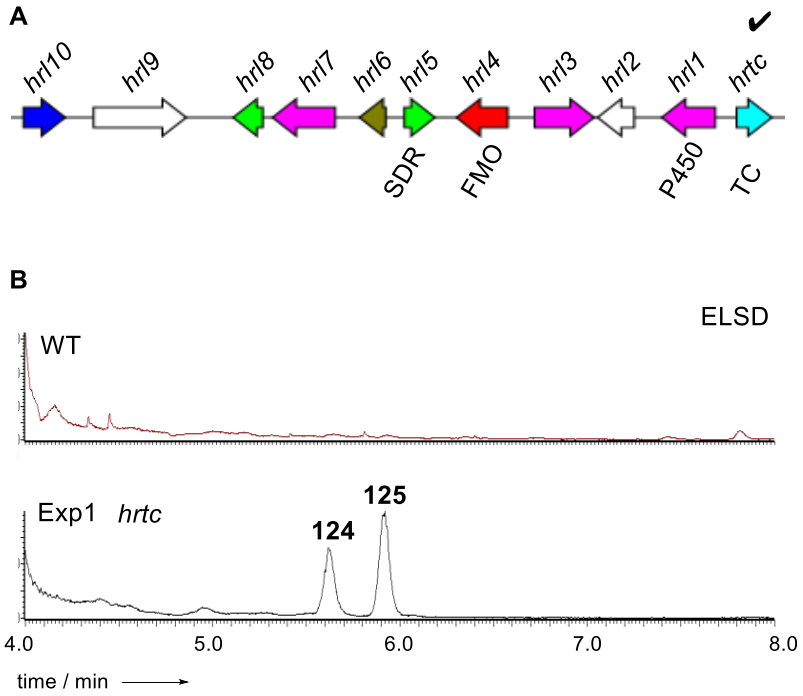


Figure 3.3.10 HPLC analysis of crude extracts from *A. oryzae* transformant; **A**, ELSD chromatogram of *A. oryzae* NSAR1; **B**, ELSD chromatogram of Exp1 (*hrtc*) transformant.

3.3.4.1 Characterization of 124

The structure of **124** was hypothesized to be derived from **66** with certain modifications. This hypothesis was formulated due to the detection of **66** in the cells of the *hrtc* transformant. **124** exhibited mass fragmentation patterns with *m/z* values of 219 and *m/z* 201 (ES^+), which are likely indicative of $[M + H - H_2O]^+$ and $[M + H - 2H_2O]^+$, respectively (Figure 3.3.11B). Additionally, it displayed a maximum UV absorption at 211 nm within the 200-600 nm range (Figure 3.3.11A).

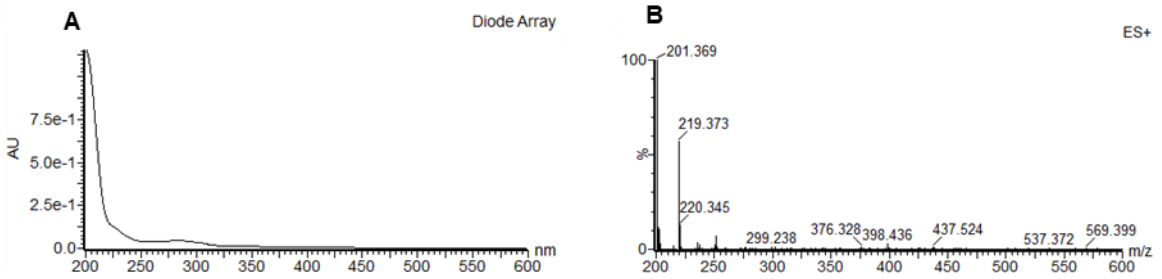


Figure 3.3.11 A, UV-absorption; **B**, fragmentation pattern in ES+ TIC of **124** by LR-LCMS.

The molecular formula of **124** was deduced as $C_{15}H_{24}O_2$ based on HR-GCMS analysis. The calculated *m/z* [M] value was 236.1776, and the observed mass closely matched at 236.1781 (Figure 3.3.12).

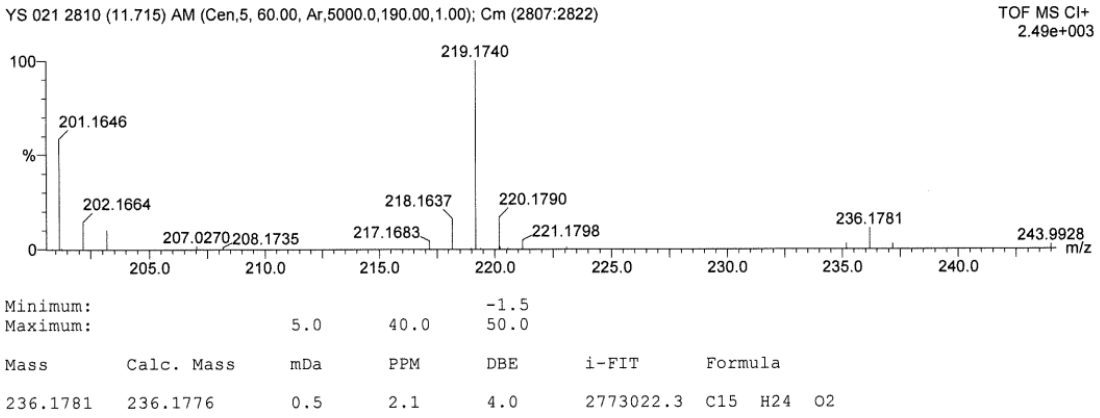


Figure 3.3.12 HRMS data for **124** *m/z* [M] calc. mass is 236.1776, 236.1781 was found by HR-GCMS.

The ^{1}H NMR data of **124** exhibited one olefinic proton at δ_H 5.42 (dd, J = 1.9, 1.9 Hz), two vinyl protons at δ_H 4.89 (m) and δ_H 4.78 (m), three methyl signals at δ_H 1.73 (dd, J = 1.1, 1.1 Hz), δ_H 0.87 (d, J = 6.7 Hz), and δ_H 1.02 (s), four methines at δ_H 4.07 (ddd, J = 9.5, 2.2, 2.2 Hz), δ_H 3.6 (dddd, J = 13.2, 11.4, 2.2, 2.2 Hz), δ_H 2.22 (ddd, J = 12.6, 9.5, 2.6 Hz) and δ_H 1.31(m), and three sets of methylene protons at δ_H 1.26- δ_H 2.37 (Table 3.3.9).

The ^{13}C NMR data of **124** showed three methyl signals (δ_C 19.5, δ_C 15.1, δ_C 17.6), three sp^3 methylene signals (δ_C 41.7, δ_C 40.1, δ_C 41.0), four sp^3 methine signals (including two oxygenated methines) at δ_C 70.9, δ_C 69.1, δ_C 41.4 and δ_C 47.7, one sp^3 quaternary carbon δ_C 38.1, four olefinic carbons (δ_C 125.1, δ_C 143.6, δ_C 146.5 and δ_C 112.7, Table 3.3.9). The DQF-COSY spectrum showed the sequence among C-1, C-2, C-3, and an oxymethine (δ_C 70.9) was confirmed to be the C-2 position (Table 3.3.9). The other oxymethine was supported to be at C-8, because the DQF-COSY spectrum suggested the ^{1}H-^{1}H correlations between H-9 and H-8 and the sequence C-6/C-7/C-8/C-9 (Table 3.3.9). The structure of **124** was confirmed, which has two hydroxy groups at C-2 and C-8 on the basis of structure of aristolochene **66**.

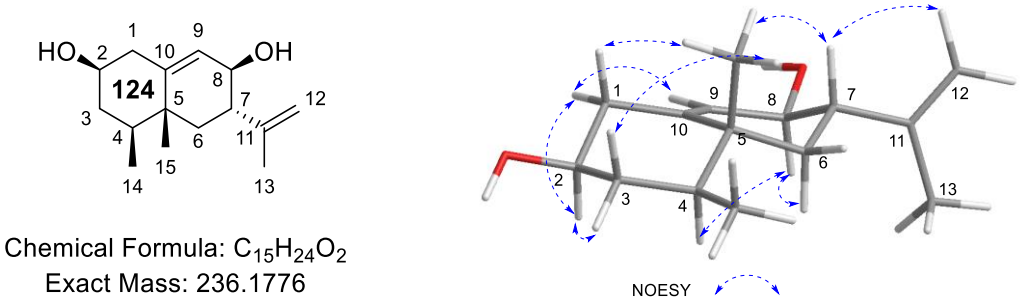


Figure 3.3.13 Structure and a stick model of **124.**

Table 3.3.9 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **124** recorded in $CDCl_3$.

Pos.	δ_C / ppm	δ_H / ppm (J/Hz)	^{1}H-^{1}H COSY	HMBC (H-C)
1	41.7	2.15, 1H, dddd (13.2, 11.4, 2.2, 2.2);	1, 2, 8, 9	2, 9, 10
		2.37, 1H, ddd (12.8, 5.0, 2.3)	1, 2, 3	2, 3, 5, 9, 10
2	70.9	3.6, 1H, dddd (11.3, 11.3, 4.8, 4.8)	1, 3	
3	40.1	1.39, 1H, ddd (12.5, 12.3, 10.8);	3, 2, 4	2, 4, 5, 14
		1.77, 1H, m	3, 2, 1	4
4	41.4	1.31, 1H, m	14	3, 5
5	38.1			
6	41.0	1.26, 1H, m;	6, 7	7, 8, 15
		1.67, 1H, dd (13.2, 2.6)	6, 7	7, 8, 10, 15
7	47.7	2.22, 1H, ddd (12.6, 9.5, 2.6)	6, 8	6, 8, 11, 12, 13
8	69.1	4.07, 1H, ddd (9.5, 2.2, 2.2)	7, 9, 1	9, 10, 11
9	125.1	5.42, 1H, dd (1.9, 1.9)	1, 8	1, 5, 7
10	143.6			
11	146.5			
12	112.7	4.89, 1H, m;	12, 13	7, 13
		4.91, 1H, m	12, 13	7, 13
13	19.5	1.73, 3H, dd (1.1, 1.1)	12	7, 11,12
14	15.1	0.87, 3H, d (6.7)	4	3, 5
15	17.6	1.02, 3H, s		4, 5, 10

The stereochemical configurations of H-2 and H-8 in **124** were elucidated based on the coupling constants among key protons surrounding these stereocenters. H-2 was proposed to occupy an axial position (H-2_{ax}), which was supported by proton coupling constants of H-2_{ax}/H-1_{ax} (^{3}J = 11.3 Hz) and H-2_{ax}/H-3_{ax} (^{3}J = 11.3 Hz). This was further corroborated by the proton coupling constants of H-2_{ax}/H-1_{eq} (^{3}J = 4.8 Hz) and H-2_{ax}/H-3_{eq} (^{3}J = 4.8 Hz). The NOESY correlations observed between H-1_{ax}/H-15 and H-2_{ax}/H-1_{eq} suggested that H-1_{ax} and H-15 were in axial positions on the same face of the (+)-aristolochene backbone, while H-2_{ax} and H-1_{eq} were on opposite faces (Figure 3.3.13). Additionally, the correlation between H-2_{ax} and H-3_{eq} further confirmed the axial position of H-2 (Figure 3.3.13). Consequently, the relative configuration at C-2 was determined to be 2*R*.

H-8 was also proposed to be axial (H-8_{ax}), supported by proton coupling constant of H-8_{ax}/ H-7_{ax} (^{3}J = 9.5 Hz). The coupling constant of H-6_{eq}/ H-7_{ax} (^{3}J = 2.6 Hz), H-7_{ax}/ H-6_{ax} (^{3}J = 12.6 Hz) indicated that the H-7 was also axial. The DQF-COSY spectrum also revealed a weak correlation of H-8_{ax}/H-1_{ax} (^{3}J = 2.2 Hz), and another correlation of H-8_{ax}/H-9 (^{3}J = 2.2 Hz). The NOESY correlations of H-8_{ax}/H-4_{ax} and H-8_{ax}/H-6_{ax} supplied the evidence for the proposition of axial position of H-8. Consequently, the relative configuration at C-2 was elucidated as 8*S*.

3.3.4.2 Characterization of 125

Another new peak, **125**, exhibited mass fragmentation patterns with *m/z* 235 (ES^{+}) and m/z 257 (ES^{+}), likely corresponding to $[M + H]^{+}$ and $[M + Na]^{+}$, respectively (Figure 3.3.14). The UV spectrum showed maximum UV absorption at 240 nm (200 - 600 nm) (Figure 3.3.14). Compound **125** was found

to possess a structure closely resembling that of compound **124**. The structure of **125** was also inferred based on **66**.

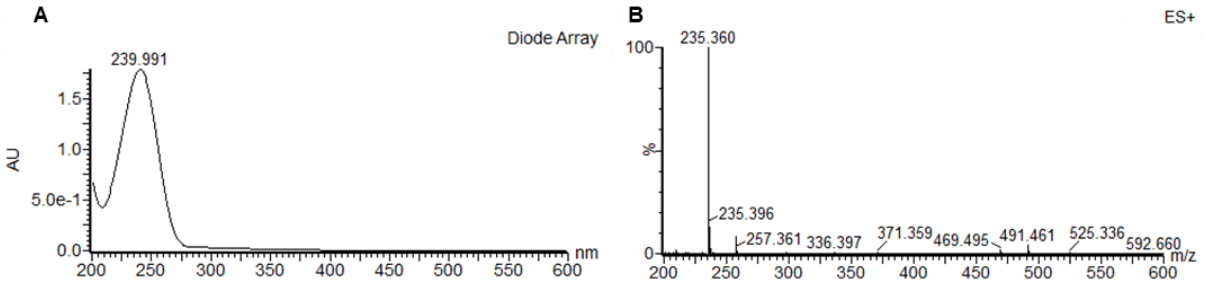
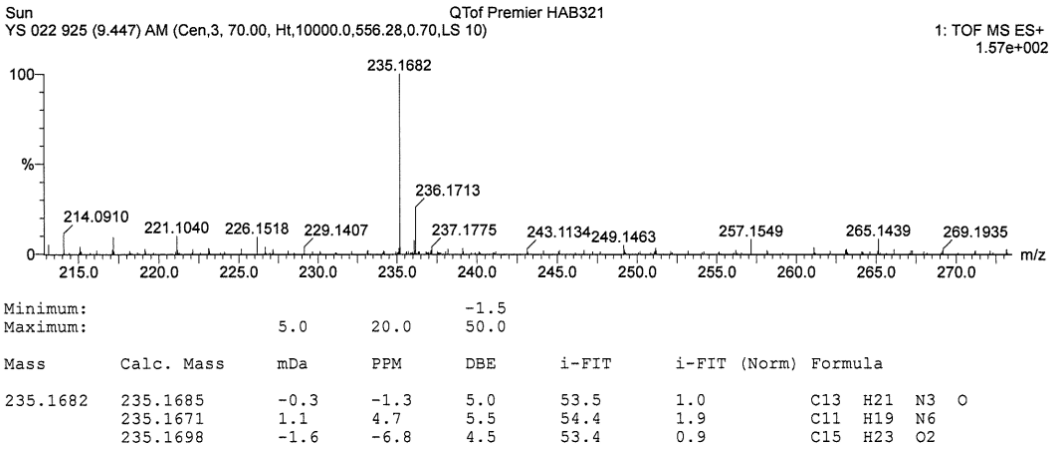


Figure 3.3.14 UV-absorption **(A)** and fragmentation pattern; **(B)** of **125** in ES$^+$ TIC by LR-LCMS

The molecular formula of **125** was determined to be $C_{15}H_{22}O_2$ through HRMS analysis. The calculated *m/z* [M + H] $^+$ mass was 235.1698, and the observed mass was 235.1682 (Figure 3.3.15).



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
235.1682	235.1685	-0.3	-1.3	5.0	53.5	1.0	C13 H21 N3 O
	235.1671	1.1	4.7	5.5	54.4	1.9	C11 H19 N6
	235.1698	-1.6	-6.8	4.5	53.4	0.9	C15 H23 O2

Figure 3.3.15 HRMS data for **125** *m/z* [M + H] $^+$ calc. mass is 235.1698, 235.1682 was found

The ^{1}H NMR data of **125** exhibited the resonances of one olefinic protons at δ_H 5.79 (d, J = 2.0 Hz), two vinyl protons at δ_H 4.82 (m) and δ_H 4.98 (m), three methyls signals at δ_H 1.74 (m), 0.96 (d, J = 6.2 Hz), and 1.16 (s), three methines at δ_H 3.75 (dddd, J = 11.2,11.2, 4.7, 4.7 Hz), 1.5 (m), 3.14 (dd, J = 14.5, 4.5 Hz), and three sets of methylene protons in the range of δ_H 1.5 - 2.56 ppm (Table 3.3.10).

The ^{13}C NMR data of **125** (Table 3.3.10) showed three methyls (δ_C 20.2, δ_C 15.0, δ_C 16.1), three sp^3 methylenes (δ_C 42.2, δ_C 39.7, δ_C 40.8), three sp^3 methines (including one oxygenated methines) at δ_C 41.4, δ_C 51.0, δ_C 69.6, one sp^3 quaternary carbon (δ_C 38.7), four olefinic carbons (δ_C 125.9, δ_C 166.6, δ_C 143.6 and δ_C 114.5). There was a carbonyl carbon (δ_C 198.9). The DQF-COSY spectrum showed that sequence among C-1, C-2, C-3, an oxymethine (δ_C 69.6) was similar to the C-2 position of **125**. The carbonyl carbon (δ_C 198.9) was supported to be the C-8, because the HMBC correlations suggested the observed correlations between H-7/C-8, H-6$_{ax}$ or H-6$_{eq}$/ C-8. The structure of **125** was confirmed, which has one hydroxy group at C-2 and a keto at C-8 on the basis of **66**. According to the elucidation of the structure, a known and identical structure was found.[182]

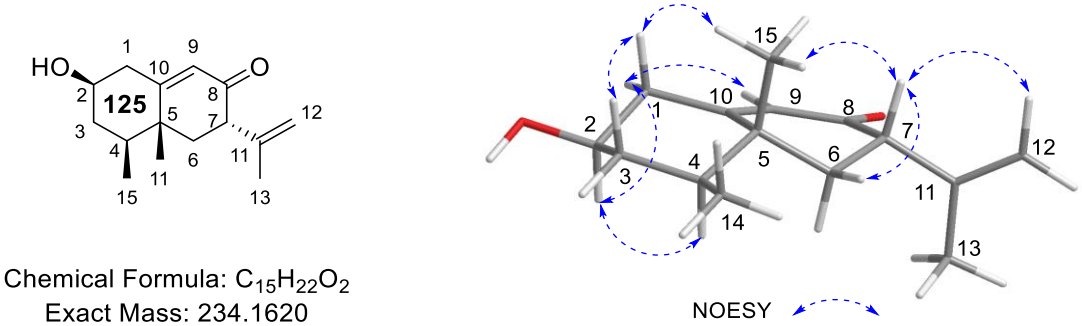


Figure 3.3.16 Structure and a stick model of **125.**

Table 3.3.10 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **125** recorded in $CDCl_3$, Literature [182] data was measured in $CDCl_3$.

Pos.	δ_C / ppm	δ_H / ppm (J/Hz)	^{1}H-^{1}H COSY	HMBC (H-C)	δ_C / ppm literature	δ_H / ppm (J/Hz) literature
1	42.2	2.33, 1H, ddd (13.6, 11.5, 2.0); 2.56, 1H, ddd (13.7, 4.9, 2.4)	1, 2, 9 1, 2, 3	2, 3, 9, 10 2, 3, 5, 9, 10	42.0	2.32, 1H, t (12.0); 2.57, 1H, dd (4.8, 12.0)
2	69.6	3.75, 1H, dddd (11.2,11.2, 4.7, 4.7)	1, 3		69.3	3.75, 1H, dddd (4.8, 6.0, 10.0, 12.0)
3	39.7	1.5, 1H, m; 1.85, 1H, m	3, 2, 4, 3, 2, 4, 1	4 4	39.5	1.56, 1H, dt (10.0, 12.0); 2.05, 1H, ddd (4.0, 6.0, 12.0)
4	41.4	1.5, 1H, m	14	6	40.6	1.50, 1H, ddq (4.0, 7.0, 12.0)
5	38.7				38.5	
6	40.8	1.8, 1H, m; 2.01, dd (13.1, 4.5)	6, 7 6, 7	5, 7, 8, 15 5, 7, 8, 10, 15	41.2	1.83, 1H, dd (12.0, 14.0); 2.00, 1H, dd (4.4, 12.0)
7	51.0	3.14, 1H, dd (14.5, 4.5)	6	6, 8, 11, 12, 13	50.8	3.15, 1H, dd (4.4, 14.0)
8	198.9				198.9	
9	125.9	5.79, 1H, d (2.0)	1	1, 5, 7	125.6	5.77, s
10	166.6				166.7	
11	143.6				143.4	
12	114.5	4.82, 1H, m; 4.98, 1H, m	12, 13 12, 13	7, 8, 11, 13 7, 8, 11, 13	114.2	4.80, 1H, s 4.97, 1H, s
13	20.2	1.74, 3H, m	12		20.0	1.72, 3H, s
14	15.0	0.96, 3H, d (6.2)	4	7, 11, 12	14.7	0.95, 3H, d (7.0)
15	16.1	1.16, 3H, s		4, 5, 6, 10	15.9	1.16, 3H, s

The stereochemical configurations of H-2 in compound **125** were established by analyzing the coupling constants. Based on this analysis, H-2 was proposed to be in the axial position (H-2_{ax}). This is supported by the following proton coupling constants: H-2_{ax} (dddd, J = 11.2, 11.2, 4.7, 4.7 Hz), H-1_{ax} (ddd, J = 13.6, 11.5, 2.0 Hz), and H-1_{eq} (ddd, J = 13.7, 4.9, 2.4 Hz, Table 3.3.10).

The presence of a J of 11.2 Hz is likely due to the correlation between the protons of H-2_{ax} and H-1_{ax}. Another J of 11.2 Hz probably arises from H-2_{ax} and H-3_{ax}, although it is not observable in the spectrum. Similarly, a J of 4.7 Hz is likely attributed to the interaction between the proton of H-2_{ax} and H-1_{eq}, while another J of 4.7 Hz may originate from H-2_{ax} and H-3_{eq} (Table 3.3.10).

The NOESY correlations between H-2_{ax} and H-1_{eq}, H-2_{ax} and H-4_{ax} provide further evidence that H-1_{ax} is also in an axial position, on the same face as H-1_{eq} and H-4_{ax} (Figure 3.3.16). Consequently, it can be concluded that the relative configuration at C-2 is 2*R*.

3.3.5 Exp 2_ Expression of hr BGC

The results obtained from the data acquired in Exp1 provide insights into the role of the skeletal gene *hrtc*. The activation of this gene achieved the synthesis of (+)-aristolochene through heterologous expression. Hypoxylan A **73**, a known product derived from the fungus *Hypoxylon rickii*, prompted efforts to identify the specific gene responsible for aromatase function and discover additional tailoring enzymes for modifying the skeleton within this gene cluster. To achieve this goal, a series of heterologous experiments covering the entire gene cluster were conducted. A systematic approach was devised, where genes were systematically eliminated one by one during the heterologous expression process.

Initially, all the potentially functional genes within this cluster were simultaneously integrated into plasmids pTYGS_*arg* and pTYGS_*ade*, and subsequently expressed in *A. oryzae* NSAR1. For example, the plasmids XX02+XX03 encompassing *hrtc, hrl1, hrl3, hrl4, hrl5, hrl7*, and *hrl8* were expressed, giving rise to Exp2. In Exp3, a heterologous expression trial, the gene *hrl5* was omitted, with the plasmids XX03+XX04 being employed. Similarly, Exp4 involved the removal of *hrl4*, utilizing plasmids XX03+XX05. For Exp5, wherein *hrl3* was excluded, plasmids XX02+XX06 were utilized. In Exp6, *hrl1* was omitted, and plasmids XX02+XX07 were employed.

In Exp7, two genes (*hrl4* and *hrl5*) were removed, with *hrtc, hrl3, hrl1, hrl7*, and *hrl8* co-expressed using plasmids XX03+XX08. Finally, in Exp8, three genes (*hrl3, hrl4, hrl5*) were excluded, and only *hrtc, hrl1, hrl7*, and *hrl8* were co-expressed *via* plasmids XX06+XX08. Detection of the crude cell and medium extracts was performed using LC-MS, with the crude extract of *A. oryzae* NSAR1 serving as the control. Comprehensive protocols for these experiments are elaborated upon in Chapter 2.

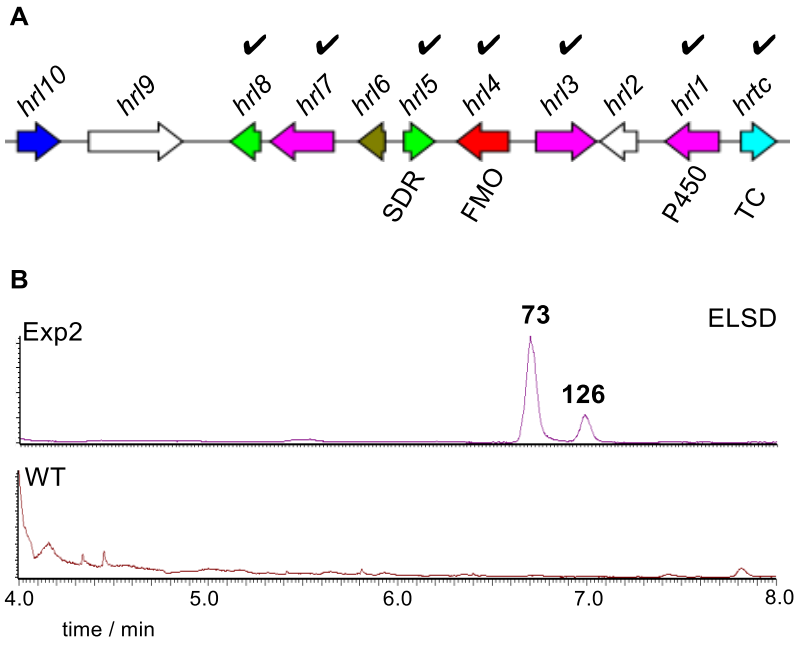


Figure 3.3.17 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp2 by HPLC analysis.

3.3.5.1 Characterization of 73

During Exp2, the expression of *hrtc, hrl1, hrl3, hrl4, hrl5, hrl7*, and *hrl8* led to the emergence of two new peaks (Figure 3.3.17). Peak **73** displayed mass fragmentation patterns at *m/z* 221 (ES^+) and *m/z* 219 (ES^-), thereby corroborating its proposed mass as 220 (Figure 3.3.18B). This identification was further supported by its UV absorption at 282 nm within the wavelength range of 200 - 600 nm (Figure 3.3.18A). The molecular formula of **73** was deduced as $C_{14}H_{20}O_2$ through HRMS analysis. Calculated at *m/z* $[M - H]^-$ 219.1385, the observed mass aligned close at 219.1392 (Figure 3.3.19).

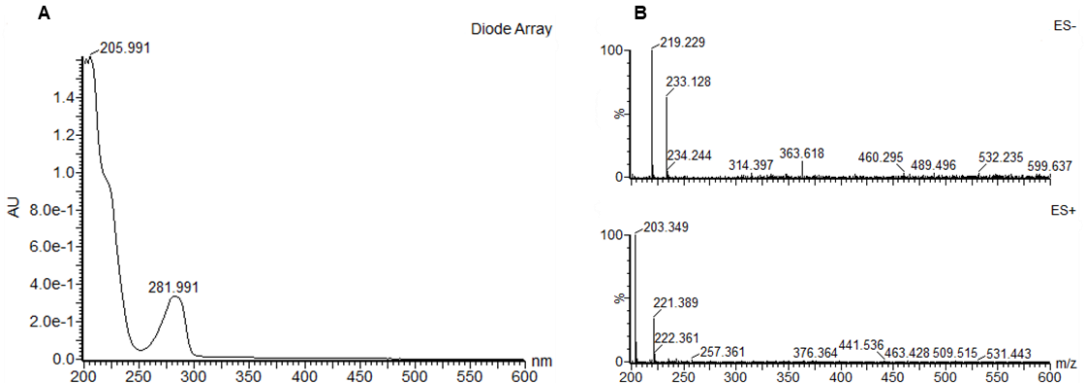


Figure 3.3.18 UV-absorption **(A)** and fragmentation pattern **(B)** of **73** in ES^- (top) ES^+ (bottom) TIC by LR-LCMS.

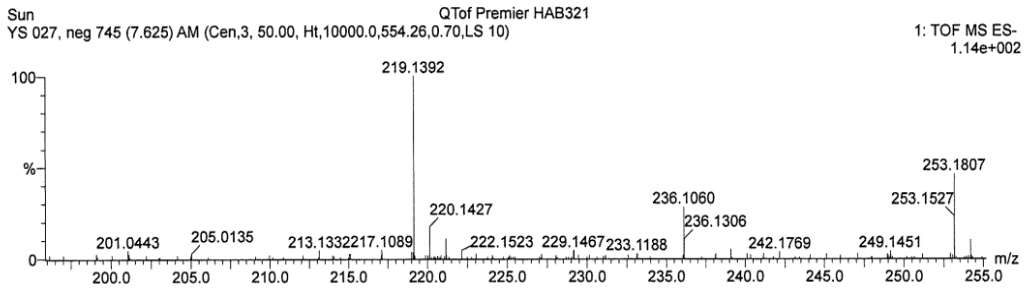
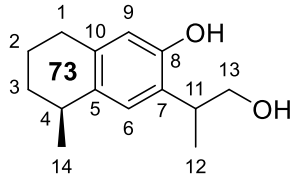


Figure 3.3.19 HRMS data for **73,** *m/z* $[M - H]^-$ calc. mass is 219.1385, 219.1392 was found.

The ^1H NMR spectroscopy data revealed a tetra-substituted aromatic ring, evident from the presence of two aromatic protons exhibiting low-field shifts at δ_H 6.6 (s) and δ_H 6.93 (s). Additionally, two methyl signals were observed at δ_H 1.32 (d, J = 7.3 Hz) and 1.26 (d, J = 7.0 Hz), along with two sp^3 hybridized methine protons resonating at δ_H 2.82 (m) and 3.19 (pd, J = 7.4, 7.4, 7.3, 7.3, 3.7 Hz). Furthermore, three sets of methylene protons were identified within the range of δ_H 1.5 - 2.68 ppm (Table 3.3.11). A distinctive hydroxymethyl group was discerned based on the interconnected proton signals of H-13 (δ_H 3.73, dd, J = 9.8, 7.6 Hz; δ_H 3.93, dd, J = 9.8, 3.8 Hz). The substitution pattern at C-8 was confirmed as

a hydroxyl group. Through COSY correlations, a sequential arrangement was established for C-1/C-2/C-3/C-4, constituting the left cyclohexene ring. This ring was interconnected with C-5 and C-10, a connection validated by HMBC correlations involving H-1/C-10 and H-4/C-5.

The ^{13}C NMR dataset (Table 3.3.11) revealed the presence of six carbon signals in the low-field region, aligning with the aromatic ring's carbon atoms. The carbon signal at δ_C 152.3 was identified as the oxygenated C-8. By employing HSQC and HMBC analyses (H-9 and C-1, H-6 and C-4 correlations), the assignment of C-6 (δ_C 127.4) and C-9 (δ_C 116.8) carbons was definitively confirmed. A distinctive carbon resonance at δ_C 64.5 indicated the presence of C-13, corroborated by correlations with C-7, C-11, and C-12. Additionally, two methyl groups were localized to C-4 (δ_C 32.0) and C-11 (δ_C 37.1). Comparison of the ^{13}C and ^{1}H NMR data with previously reported information on hypoxylan A **73** from the literature[183] yielded congruent results, further affirming the consistency in the observed chemical shifts and resonance patterns (Table 3.3.11).



Chemical Formula: $C_{14}H_{20}O_2$
Exact Mass: 220.1463

Table 3.3.11 Summarized NMR signals (^{1}H-400 MHz, ^{13}C-100 MHz) for **73** recorded in $CDCl_3$, compound from literature [183] was measured in $CDCl_3$.

Pos.	δ_C / ppm	δ_H / ppm (J/Hz)	^{1}H-^{1}H COSY	HMBC (H-C)	δ_C / ppm literature [183]	δ_H / ppm (J/Hz) literature[183]
1	29.8	2.68, 2H, m	2	2, 3, 5, 9, 10	29.7	2.67, 2H, m
2	20.6	1.67, 1H, m;	2, 1, 3	1, 3, 4, 10	20.5	1.67, 1H, m;
		1.85, 1H, m	2, 1, 3	1, 3, 4, 10		1.82, 1H, m
3	31.9	1.5, 1H, m;	3, 2, 4	1, 2, 4, 5	31.8	1.49, 1H, m;
		1.88, 1H, m	3, 2, 4	1, 2, 4, 5		1.87, 1H, m
4	32.0	2.82, 1H, m	3, 14	2, 3, 5, 10, 14	31.9	2.83, 1H, m
5	134.4				134.3	
6	127.6	6.93, 1H, s		4, 8, 10, 11	127.3	6.92, 1H, s
7	128.3				128.2	
8	152.4				152.4	
9	117	6.6, 1H, s		1, 5, 7, 8	116.9	6.59, 1H, s
10	136.7				136.6	
11	37.1	3.19, 1H, pd (7.4, 7.4, 7.3, 7.3, 3.7)	12, 13	7, 8, 12, 13	36.8	3.19, 1H, dqd (7.8, 7.3, 3.7)
12	16.0	1.32, 3H, d (7.3)	11	7, 11, 13	15.8	1.30, 3H, d (7.3)
13	69.6	3.73, 1H, dd (9.8, 7.6);	11, 13	7, 11, 12	69.7	3.72, 1H, dd (9.8,7.8);
		3.93, 1H, dd (9.8, 3.8)	11, 13	7, 11, 12		3.92, 1H, dd (9.8,3.7)
14	23.2	1.26, 3H, d (7.0)	4	3, 4, 5	23.1	1.24, 3H, d (7.9)

3.3.5.2 Characterization of 126

Another compound **126** exhibits a comparable NMR assignment profile to **73**. The shifts of the ^{13}C NMR from C-1 to C-10, along with the methyl groups at C-13 and C-14, are consistent with this observation

of the NMR of **73** (Table 3.3.12). The hydroxymethyl group's C-13 position in **126** undergoes substitution with a keto moiety of a carboxylic acid (δ_C 181.4), resulting in the formation of hypoxylan A acid **126**.

This deduction was further evidenced by the mass fragmentation patterns observed at *m/z* 235 (ES^+) and *m/z* 233 (ES^-), which lend support to the proposed mass value of 234 for **126** (Figure 3.3.20B). Corroborating the identification, UV absorption at 282 nm within the wavelength range of 200-600 nm is observed (Figure 3.3.20A). Employing HRMS analysis, the molecular formula of **126** is determined to be $C_{14}H_{18}O_3$. The calculated *m/z* for $[M + H]^+$ is 235.1334, closely aligned with the observed mass of 235.1343 (Figure 3.3.21). **126** represents a novel compound not hitherto reported among the products of *Hypoxylon rickii*. It also remains absent from existing literature. The carboxylic acid of **126** was probably formed by the shunt of *A. oryzae.*

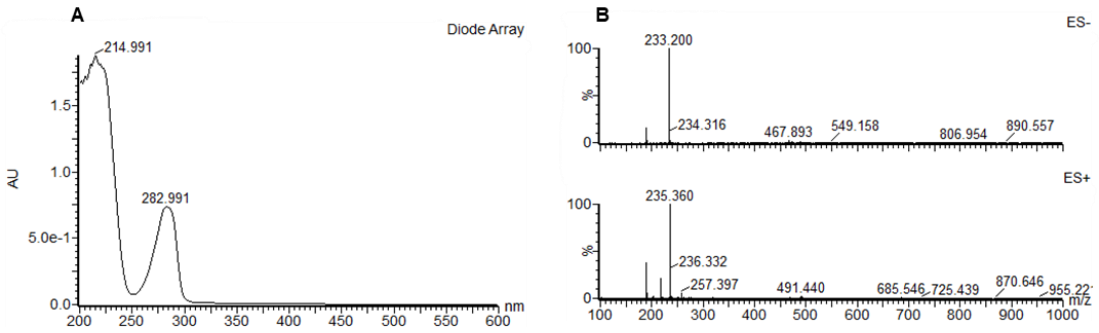


Figure 3.3.20 UV-absorption **(A)** and fragmentation pattern **(B)** of **126** in ES^+ TIC by LR-LCMS.

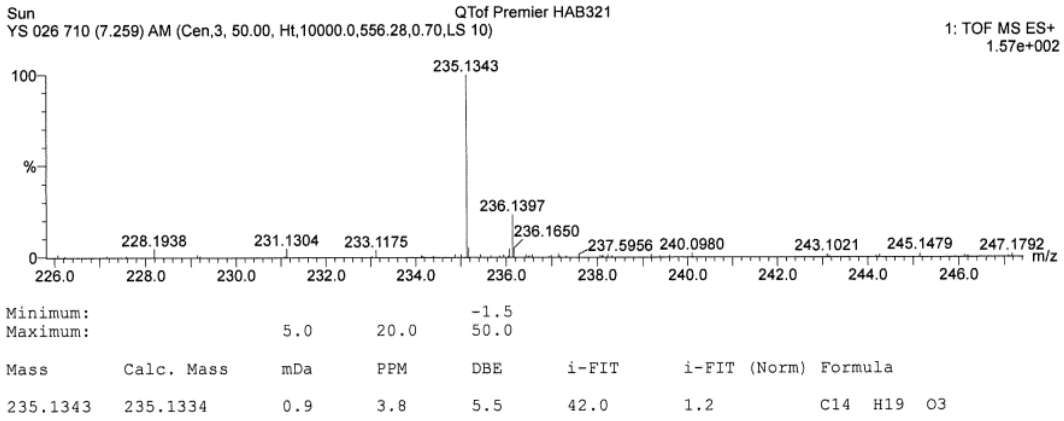
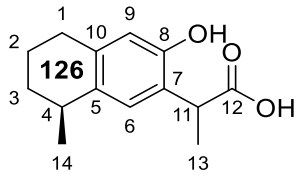


Figure 3.3.21 HRMS data for **126**, Calculated at *m/z* $[M + H]^+$ 235.1334, the observed mass aligned close at 235.1343.



Chemical Formula: $C_{14}H_{18}O_3$
Exact Mass: 234.1256

Table 3.3.12 Summarized NMR signals (^{1}H-400 MHz, ^{13}C-100 MHz) for **126** recorded in $CDCl_3$.

Pos.	δ_C / ppm	δ_H / ppm (J/Hz)	^{1}H-^{1}H COSY	HMBC (H-C)
1	29.8	2.67, 2H, m	1, 2	2, 3, 5, 9, 10
2	20.5	1.68, 1H, m;	2, 1, 3	1, 3, 10
		1.84, 1H, m	2, 1, 3	1, 3, 10
3	31.7	1.49, 1H, m;	3, 2, 4	1, 2, 4, 14
		1.86, 1H, m	3, 2, 4	1, 2, 4, 14
4	31.9	2.82, 1H, m	3, 14	2, 3, 5, 6, 14
5	135.0			
6	128.4	6.97, 1H, s		8, 10, 11
7	123.5			
8	151.7			
9	117.2	6.58, 1H, s		1, 8, 10
10	137.8			
11	41.4	3.88, 1H, ddd (7.2, 7.2, 7.2)	13	7, 8, 12, 13
12	181.4			
13	16.2	1.54, 3H, d (7.3)	11	7, 11, 12
14	23.1	1.24, 3H, d (7.0)	4	3, 4, 5

In Exp2, *hrtc, hrl1, hrl3, hrl4, hrl5, hrl7*, and *hrl8* were expressed in *A. oryzae* NASR1. This expression led to the production of hypoxylan-A **73**, along with the isolation and elucidation of a derivative, hypoxylan A acid **126**. This indicates the likely presence of an aromatase responsible for the phenolic formation or an enzyme performing oxidative demethylation among this gene cluster. Subsequently, we employed a strategic approach referred to as "knockout by expression." This approach entails the sequential removal of individual genes from the complete cluster in a stepwise manner.

3.3.6 Exp 3_ Expression of hr BGC w/o *hrl5*

The first gene omission involved *hrl5* encoding a predicted SDR. In the subsequent step (Exp3), *hrtc, hrl1, hrl3, hrl4, hrl7*, and *hrl8* were co-expressed. Following the completion of transformation and fermentation, the chromatograms revealed an absence of changes in the products; they remained as **73** and **126** (Figure 3.3.22). This result suggested that *hrl5* is not involved in the formation of hypoxylan A **73**.

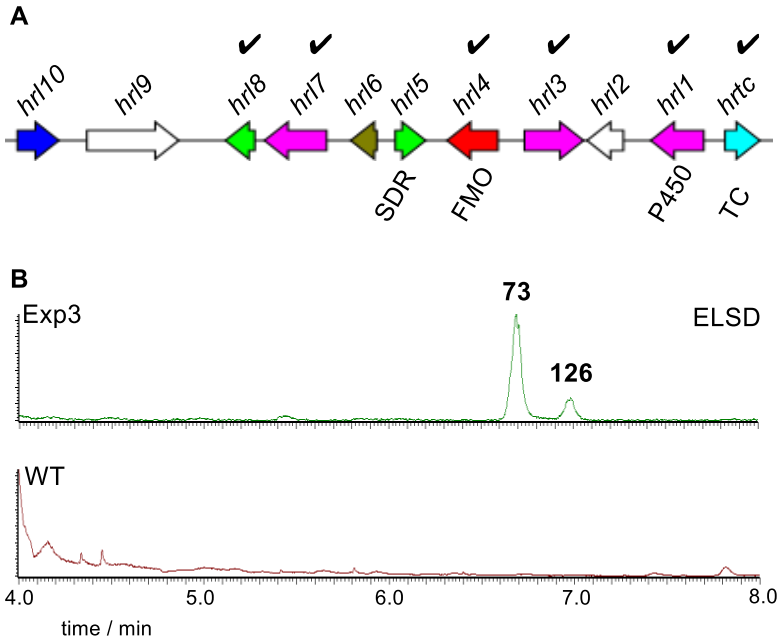


Figure 3.3.22 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp3 by HPLC analysis.

3.3.7 Exp 4_ Expression of hr BGC w/o *hrl4*

Following that, a subsequent experiment (Exp4) involving the omission of *hrl4* was undertaken (Figure 3.3.23). Upon completion of transformation and fermentation, the chromatograms exhibited three new peaks. Subsequent isolation yielded four distinct compounds (**127, 128**, **129**, and **130**), with **129** and **130** constituting a mixed entity, sharing identical MS data and retention time. The structural identification for all these compounds was accomplished through comprehensive 1D and 2D NMR analyses.

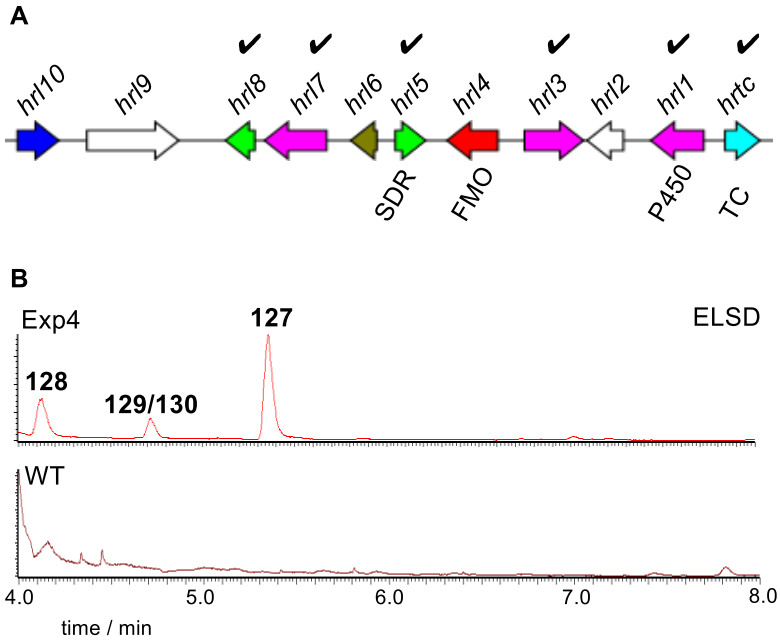


Figure 3.3.23 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp4 by HPLC analysis.

3.3.7.1 Characterization of 127

127 displayed some mass fragmentation patterns at *m/z* 237, 219, and 473 in the positive ion mode (ES^+), as well as *m/z* 235 in the negative ion mode (ES^-, Figure 3.3.24B). The presence of the *m/z* 219 (ES^+) fragment can be attributed to the loss of a water molecule $[M + H - H_2O]^+$, while the 473 peak likely corresponds to a dimeric species $[2M + H]^+$, lending further credence to its proposed mass of 236 (Figure 3.3.24B). Reinforcing this identification, UV absorption at 282 nm within the wavelength range of 200-600 nm was observed (Figure 3.3.24A).

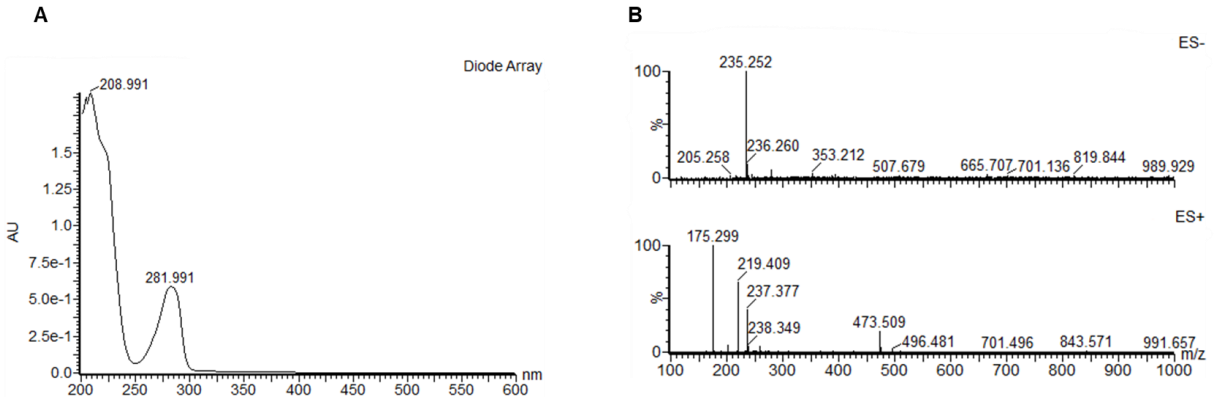
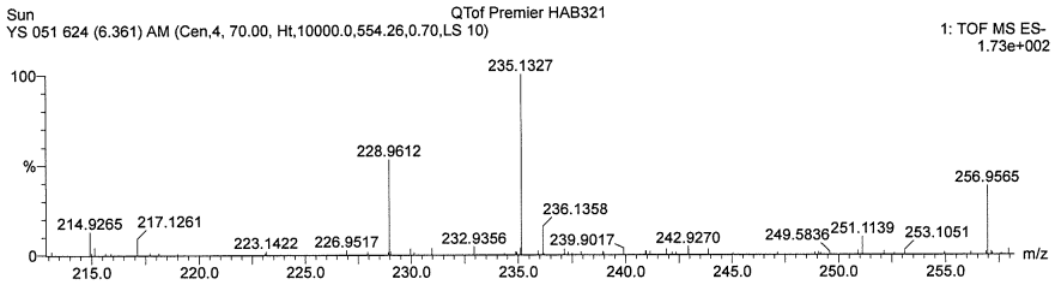


Figure 3.3.24 UV-absorption **(A)** and fragmentation pattern **(B)** of **127** in ES^+ TIC (bottom) and ES^- TIC (top) by LR-LCMS.

Employing HRMS analysis, the molecular formula of **127** was established as $C_{14}H_{20}O_3$. Calculations for $[M - H]^-$ yielded a theoretical *m/z* of 235.1334, closely aligned with the observed mass of 235.1327 (Figure 3.3.25).



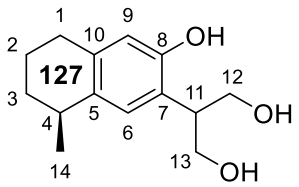

```
Minimum:                                  -1.5
Maximum:                    5.0     20.0   50.0

Mass        Calc. Mass      mDa     PPM    DBE     i-FIT     i-FIT (Norm)  Formula

235.1327    235.1328        -0.1    -0.4   2.0     14.9      4.2           C5  H17  N9  S
            235.1321        0.6     2.6    6.0     11.8      1.1           C12  H17  N3  O2
            235.1334        -0.7    -3.0   5.5     11.6      0.9           C14  H19  O3
```

Figure 3.3.25 HRMS data for **127**, Calculated at *m/z* $[M - H]^-$ 235.1334, the observed mass aligned close at 235.1327.

The structural elucidation of **127** revealed a resemblance to the structure of **73**. In the ^{1}H-NMR spectrum, key features persisted, including the tetra-substituted aromatic ring, as well as two protons with low-field shifts: H-9 (δ_H 6.45, s) and H-6 (δ_H 6.91, s, Table 3.3.13). But only a single methyl signal was detected at δ_H 1.23 (d, J = 7.0 Hz), whereas the other methyl had been substituted by a hydroxyl group,

leading to H-12 (δ_H 3.86, 2H, m). The hydroxymethyl group on C-13 remained unaltered (δ_H 3.86, 2H, m, Table 3.3.13). Verification of the substitution pattern at C-8 confirmed the persistence of a hydroxyl group.

Upon scrutiny of the ^{13}C NMR dataset (Table 3.3.13), six carbon signals in the low-field range corresponded to the carbon atoms within the aromatic ring. Further supported by 2D NMR analysis, the structure of **127** was determined to be similar to that of **73**, featuring an additional hydroxymethyl group at C-13. It is worth noting that the precise assignment of the two hydroxymethyl groups could not be deduced from the 2D NMR data. Consequently, the omission of the FAD-dependent dehydrogenase encoded by *hrl4* resulted in the formation of the diol **127**. This reveals that Hrl4 served as a role in the process of the formation of **73** and **126**.



Chemical Formula: $C_{14}H_{20}O_3$
Exact Mass: 236.1412

Table 3.3.13 Summarized NMR signals (^1H-600 MHz, ^{13}C-150 MHz) for **127** recorded in CD_3OD.

Compound 127				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^1H-^1H COSY**	**HMBC (H-C)**
1	30.8	2.62, 2H, m	1, 2	2, 3, 5, 9, 10
2	21.6	1.66, 1H, m;	2, 1, 3	1, 3, 10
		1.82, 1H, m	2, 1, 3	1, 3, 10
3	33.1	1.47, 1H, m;	3, 2, 4	1, 2, 4, 5, 14
		1.87, 1H, m	3, 2, 4	1, 2, 4, 5, 14
4	33.1	2.78, 1H, m	3, 14	2, 3, 5, 6, 10, 14
5	134.1			
6	129.7	6.91, 1H, s		1, 4, 8, 9, 10, 11
7	125.6			
8	154.1			
9	116.2	6.45, 1H, s		1, 5, 7, 8, 11
10	136.9			
11	46.4	3.22, 1H, dddd (6.4, 6.4, 6.4, 6.4)	12, 13	6, 7, 8, 12, 13
12	64.1	3.86, 2H, m;	12, 11	7, 11, 13
13	64.0	3.86, 2H, m	13, 11	7, 11, 12
14	23.5	1.23, 3H, d (7.0)	4	3, 4, 5

3.3.7.2 Characterization of 128

128 exhibited mass fragmentation patterns at *m/z* 217 in the positive ion mode (ES^+), as well as *m/z* 233 in the negative ion mode (ES^-). The presence of the 217 fragment is attributed to the loss of a water molecule $[M + H - H_2O]^+$, lending further support to its proposed mass of 234 (Figure 3.3.26B). Reinforcing this identification, UV absorption at 288 nm within the wavelength range of 200-600 nm was observed (Figure 3.3.26A).

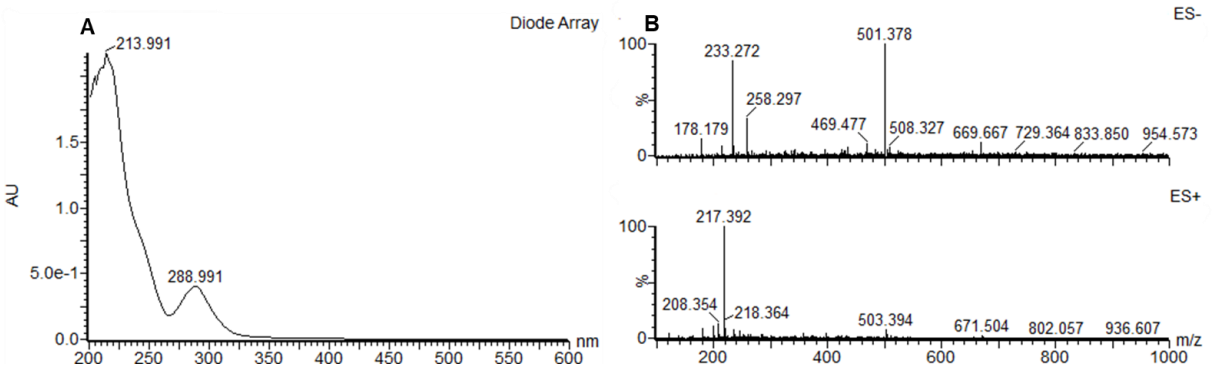


Figure 3.3.26 UV-absorption **(A)** and fragmentation pattern **(B)** of **128** in ES^+ TIC (bottom) and ES^- TIC (top) by LR-LCMS.

Through HRMS analysis, the molecular formula of **128** was deduced as $C_{14}H_{18}O_3$. Calculations for $[M - H]^-$ resulted in a theoretical *m/z* of 233.1178, closely aligned with the observed mass of 233.1171 (Figure 3.3.27).

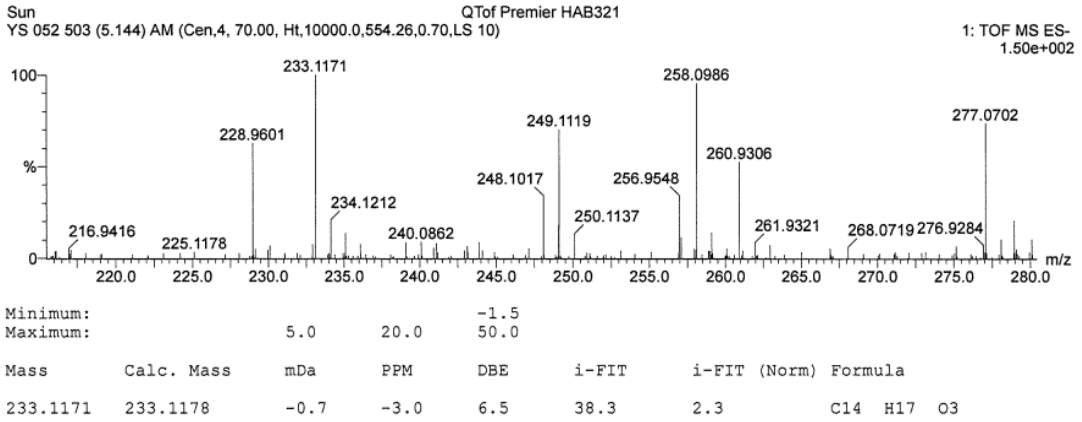
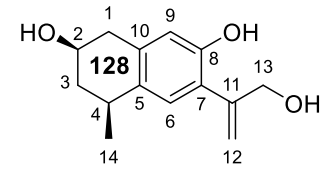


Figure 3.3.27 HRMS data for **128**, calculated at *m/z* $[M - H]^-$ 233.1178, the observed mass aligned close at 233.1171.

The structural analysis of **128** revealed a similarity to the configuration of **73** (Figure 3.3.28). The ^{1}H-proton NMR spectrum retained pivotal features, encompassing the tetra-substituted aromatic ring, as well as two protons exhibiting distinct low-field shifts: H-9 (δ_H 6.48, s) and H-6 (δ_H 7.02, s, Table 3.3.14). A single methyl signal was discernible at δ_H 1.3 (d, J = 6.8 Hz), whereas the other methyl had undergone conversion into a double bond, manifesting as H-12 (δ_H 5.15, 2H, m; δ_H 5.35, 2H, m). The hydroxymethyl group on C-13 remained consistent, registering at δ_H 4.35 (2H, m). Confirmation of the presence of a hydroxyl group at C-8 was obtained. Similar to **124** and **125**, an oxidative hydroxyl group was identified at C-2 (δ_H 3.9, 1H, m) according to the HMBC, ^{1}H-^{1}H COSY (Table 3.3.14).

Through detailed examination of the ^{13}C NMR dataset (Table 3.3.14), it was evident that six carbon signals in the low-field range corresponded to the carbon atoms within the aromatic ring (Table 3.3.14). Further substantiated by 2D NMR analysis, the structure of **128** was determined to be similar to **73**, with

maintaining olefinic carbons at C-11 and C-12 similar to aristolochene **66**. The hydroxyl group at C-13 remained consistent with its presence in **73**. The oxidative C-2 assignment was established at δ_C 68.3.



Chemical Formula: $C_{14}H_{18}O_3$
Exact Mass: 234.1256

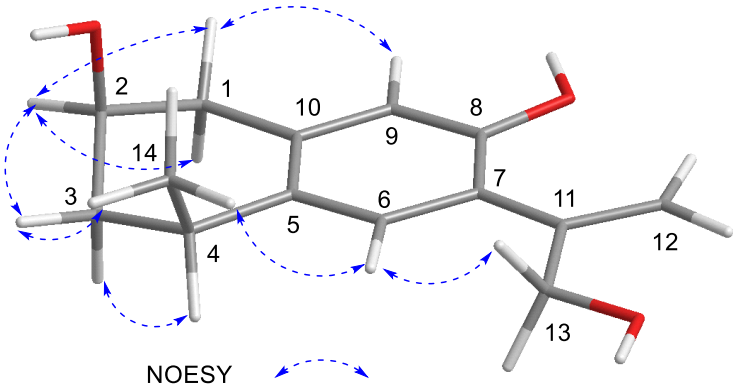


Figure 3.3.28 Structure and a stick model of **128** for the NOESY.

Table 3.3.14 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **128** recorded in CD_3OD.

Compound 128				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	40.4	2.59, 1H, dd (15.5, 10.3);	1, 2	2, 3, 5, 9, 10
		2.9, 1H, ddd (15.6, 5.3, 2.4)	1, 2, 3	2, 3, 5, 9, 10
2	68.3	3.9, 1H, m	1, 3	
3	43.5	1.33, 1H, m;	3, 2, 4	1, 2, 4, 5, 14
		2.14, 1H, dddd (11.9, 5.8, 3.5, 2.4)	3, 2, 4, 1	1, 2, 4, 5
4	33.3	2.86, 1H, m	3, 14	3, 5, 14
5	132.7			
6	129.2	7.02, 1H, s		4, 8, 10, 11
7	127.2			
8	153.5			
9	116.5	6.48, 1H, s		1, 5, 7, 8
10	136.6			
11	149.6			
12	114.4	5.15, 1H, m;	12, 13	7, 11, 13
		5.35, 1H, m	12, 13	7, 11, 13
13	65.9	4.35, 2H, m	12	7, 11, 12
14	22.4	1.3, 3H, d (6.8)	4	3, 4, 5

The stereochemical configuration of H-2 in **128** was determined by analyzing the coupling constants between key protons adjacent to the stereocenters. It was established that H-2 occupied the lower position (H-2_{ax}), with a stereochemistry of 2*R*. This conclusion was supported by the proton coupling constants observed between H-2_{ax} and H-1_{ax} (^3J = 10.3 Hz), as well as between H-2_{ax} and H-1_{eq} (^3J = 5.3 Hz), along with the proton coupling constants between H-2_{ax} and H-3_{eq} (^3J = 5.8 Hz).

3.3.7.3 Characterization of 129, 130

Compounds **129** and **130** underwent purification and appeared as individual peaks upon LR-LCMS analysis. Their UV absorption maxima (λ_{max}) were both observed at 290 nm. ESI-MS analysis yielded *m/z* values of $[M - H]^-$ = 233 and $[M + H - H_2O]^+$ = 217 for both compounds.

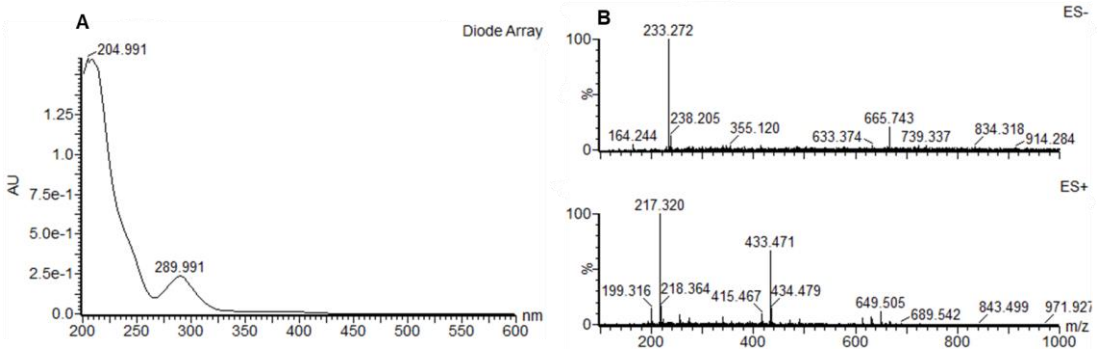
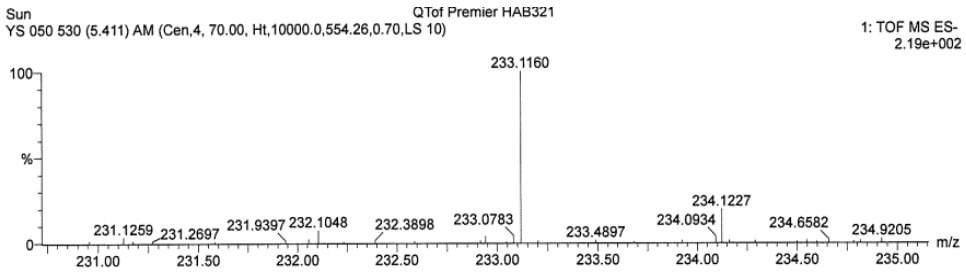


Figure 3.3.29 UV-absorption **(A)** and fragmentation pattern **(B)** of **129/130** in ES^+ (bottom) and ES^- (top) by LR-LCMS.

The HR-ESI-MS result showed *m/z* $[M - H]^-$ at 233.1160, the calculated $[M - H]^-$ was 234.1256 - 1.0078 = 233.1178. Through HRMS analysis, the molecular formula of **129, 130** was deduced as $C_{14}H_{18}O_3$.



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
233.1160	233.1164	-0.4	-1.7	7.0	48.4	1.1	C12 H15 N3 O2
	233.1151	0.9	3.9	7.5	48.6	1.4	C10 H13 N6 O
	233.1171	-1.1	-4.7	3.0	54.3	7.1	C5 H15 N9 S
	233.1146	1.4	6.0	1.5	55.5	8.3	C10 H21 N2 S2
	233.1144	1.6	6.9	-1.5	55.5	8.3	C2 H17 N8 O3 S
	233.1178	-1.8	-7.7	6.5	48.9	1.7	C14 H17 O3

Figure 3.3.30 HRMS data for **129** and **130**, calculated at *m/z* $[M - H]^-$ 233.1178, the observed mass aligned close at 233.1160.

NMR investigations indicated that compounds **129** and **130** formed a mixture. Their characterization suggested them to be a pair of epimers, as gleaned from comprehensive 1D and 2D NMR analyses (Figure 3.3.31). Compound **129** featured a hydroxyl group at C-1, defining it as the (1*R*)-epimer, while compound **130** also possessed a hydroxyl group at C-1, classifying it as the (1*S*)-epimer. Comparative analysis of ^{13}C-NMR of compound **128** with compounds **129** and **130** showed high similarity in the resonances of C-8, C-11, C-10, C-5, C-6, C-9, C-12, C-13, and C-14 (Table 3.3.15, Table 3.3.16). However, shifts in C-1, C-3, and C-4 resonances were observed, attributed to steric hindrance caused by the hydroxyl group.

In light of HMBC data, the proton signal of H-9 exhibited coherence to C-1 at chemical shifts of δ_C 69.1 and δ_C 69.2, affirming the presence of the hydroxyl group at C-1 in both epimers. Corroborating this, the ^{1}H, ^{1}H-COSY spectrum illustrated a correlation between the proton signal of H-9 and that of H-1 at

δ_H 4.57. The absence of a correlation between the proton signal of H-4 and H-3 at δ_H 4.57 indicated the lack of a hydroxyl group at C-3 in either of the mixtures (Table 3.3.15, Table 3.3.16).

Further distinction between these (1*R*) and (1*S*) epimers was achieved using NOESY (Figure 3.3.31). Starting from the distinct proton signals of H-4 (δ_H 2.84, δ_H 2.75), we mapped out the positions of the corresponding H-3 protons *via* NOESY correlations (Figure 3.3.31). Discrepancies in the NOESY profiles of the two compounds facilitated the differentiation of their signals, effectively classifying these epimers (Table 3.3.15, Table 3.3.16).

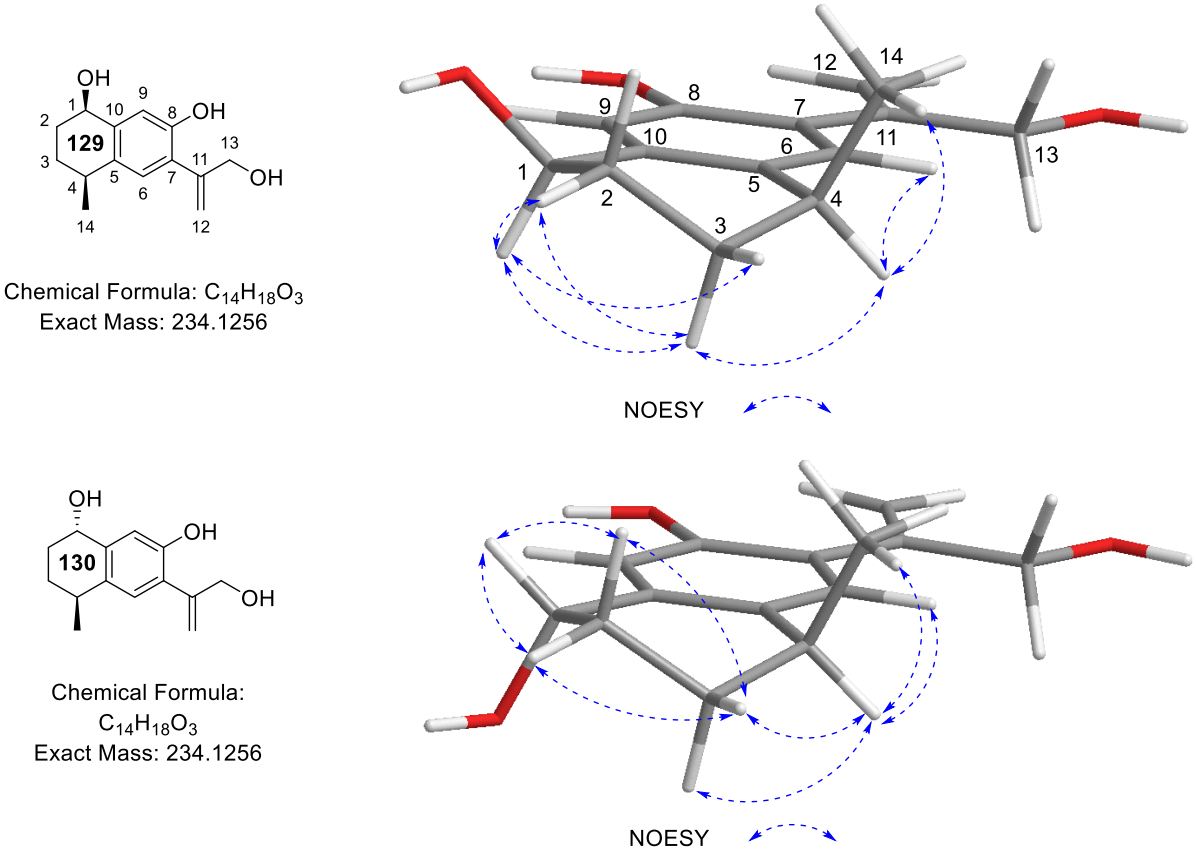


Figure 3.3.31 Structure and a stick model of **129, 130** for the NOESY.

Table 3.3.15 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **129** recorded in CD_3OD.

Compound 129				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	69.1	4.57, 1H, m	2, 9	3, 5, 10
2	31.2	1.71, 1H, m;	2, 3	1, 3, 4, 10
		2.07, 1H, m	1, 2, 3	1, 3, 4, 10
3	29.1	1.42, 1H, m;	2, 3, 4	1, 2, 4, 5
		2.11, 1H, m	3, 4	1, 2, 4, 5
4	33.1	2.84, 1H, m	3, 14	2, 3, 5, 14
5	134.0			
6	130.4	6.97, 1H, m		4, 8, 10, 11
7	128.5			
8	153.5			
9	115.9	6.85, 1H, d (0.8)	1	1, 5, 7, 8
10	140.6			
11	149.6			
12	114.5	5.16, 1H, m	12, 13	7, 11, 13
		5.37, 1H, m	12, 13	7, 11, 13
13	65.8	4.36, 2H, m	12	7, 11, 12
14	23.1	1.22, 3H, d (7.0)	4	3, 4, 5

Table 3.3.16 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **130** recorded in CD_3OD.

Compound 130				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	69.2	4.57, 1H, m	2, 9	3, 5, 10
2	30.8	1.88, 2H, m	1, 3	1, 3, 4, 10
3	28.7	1.68, 1H, m;	2, 4	1, 2, 4, 5, 14
		1.84, 1H, m	2, 4	1, 2, 4, 14
4	33.0	2.75, 1H, m	3, 14	2, 3, 5, 14
5	134.2			
6	130.3	6.98, 1H, m		4, 8, 10, 11,
7	128.5			
8	153.5			
9	115.8	6.86, 1H, d (0.9)	1	1, 5, 7, 8
10	140.5			
11	149.6			
12	114.5	5.16, 1H, m;	12, 13	7, 11, 13
		5.37, 1H, m	12, 13	7, 11, 13
13	65.8	4.36, 2H, m		7, 11, 12
14	23.0	1.28, 3H, d (7.0)	4	3, 4, 5

3.3.8 Exp 5_ Expression of hr BGC w/o *hrl3*

Following Exp4, the gene exclusion process focused on *hrl3*, which is predicted to encode a P450 enzyme. In the subsequent step (Exp5), a co-expression approach was employed, involving *hrtc, hrl1, hrl5, hrl4, hrl7*, and *hrl8*. Upon the successful transformation and fermentation, the ELSD chromatograms confirmed the production of **128** as the product (Figure 3.3.32). It is highly likely that *hrl3* encodes a P450 enzyme responsible for the oxidation at the C-13 position on the backbone.

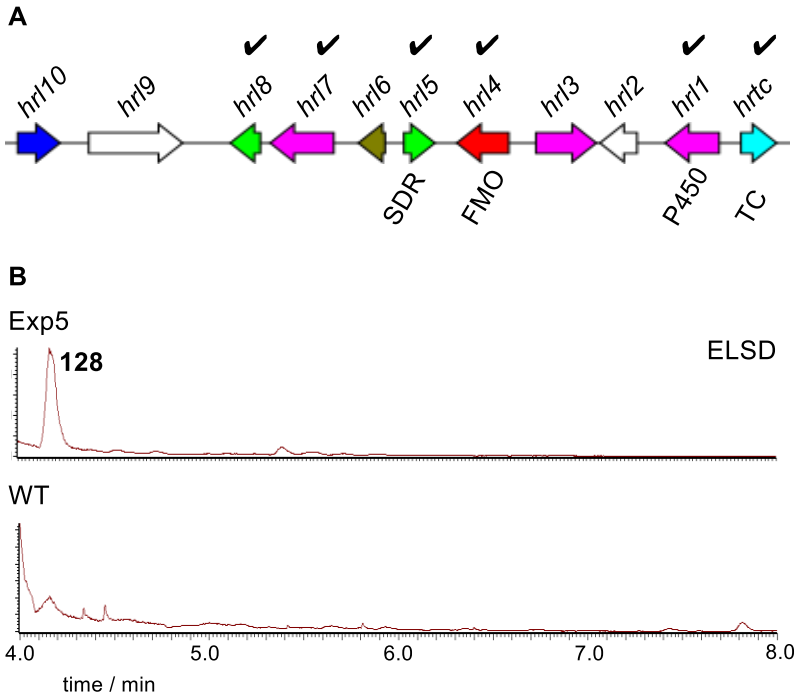


Figure 3.3.32 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp5 by HPLC analysis.

3.3.9 Exp 6_ Expression of hr BGC w/o *hrl1*

In Exp6, we intentionally omitted *hrl1*, which is co-expressing *hrtc* along with *hrl3*, *hrl4*, *hrl5, hrl7* and *hrl8* (Figure 3.3.33A). After the transformation and fermentation stages, the chromatograms exhibited the emergence of a novel peak. Subsequent isolation procedures allowed us to obtain compound **131** (Figure 3.3.33B). The structural characterization of this compound was effectively accomplished through an extensive analysis using both 1D and 2D NMR spectroscopy techniques.

The elucidation of structure of **131** revealed that the six carbon signals associated with the aromatic ring are conspicuously absent within the low-field region, when compared to compounds derived from Exp 2, 3, 4, and 5. This conspicuous difference distinctly implicated Hrl1 as the pivotal protein responsible for the aromatic ring's assembly and concurrently, the observed methyl group's loss.

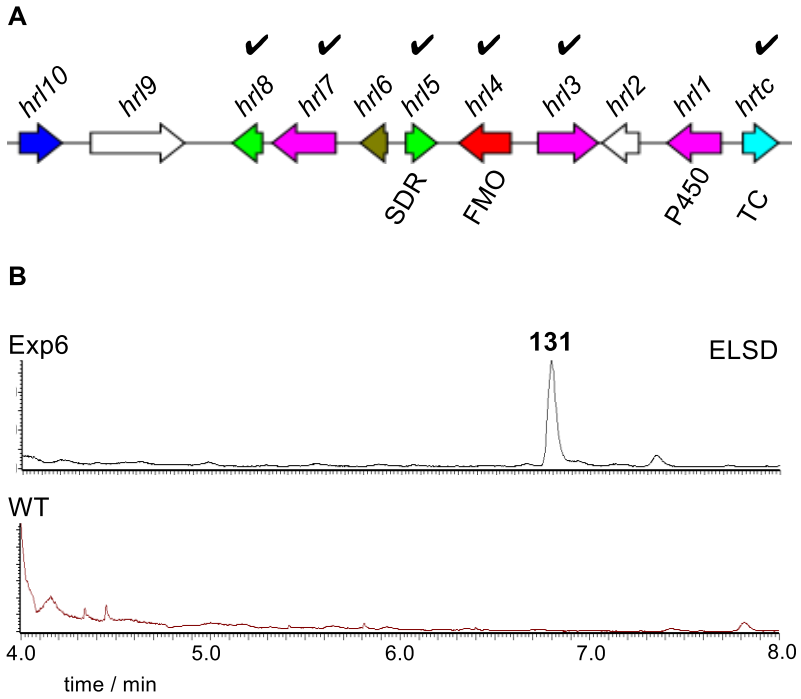


Figure 3.3.33 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp6 by HPLC analysis.

3.3.9.1 Characterization of 131

131 exhibited mass fragmentation patterns at *m/z* 249 and 231 in the positive ion mode (ES^+), as well as *m/z* 247 in the negative ion mode (ES^-, Figure 3.3.34B). The presence of the 231 fragment can be ascribed to the loss of a water molecule $[M + H - H_2O]^+$, while the 249 peak likely corresponds to an $[M + H]^+$ ion and 247 corresponds to $[M - H]^-$ ion. These observations robustly support its hypothesized mass of 248. Further affirming this characterization, a UV absorption peak at 242 nm within the wavelength range of 200-600 nm was discerned (Figure 3.3.34A).

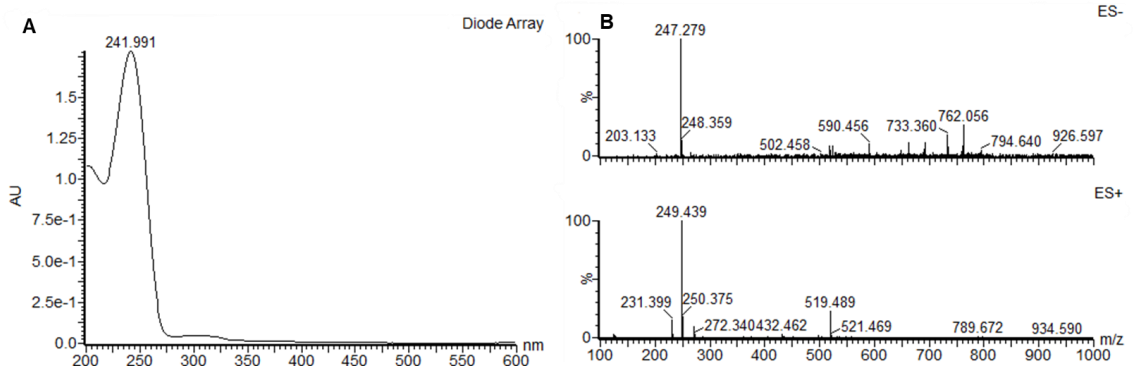
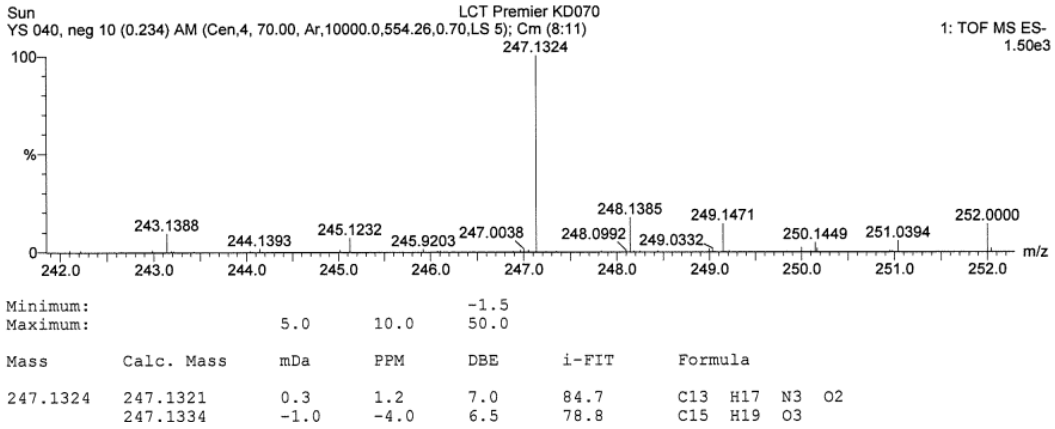


Figure 3.3.34 UV-absorption **(A)** and fragmentation pattern **(B)** of **131** in ES^+ (bottom) and ES^- (top) by LR-LCMS.

By leveraging HRMS analysis, the molecular formula of **131** was definitively identified as $C_{15}H_{19}O_3$. Calculations for $[M - H]^-$ yielded a theoretically expected *m/z* of 247.1334, which closely aligns with the observed mass of 247.1324 (Figure 3.3.35).

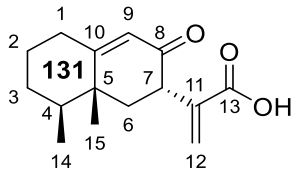


Minimum:				-1.5		
Maximum:		5.0	10.0	50.0		
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
247.1324	247.1321	0.3	1.2	7.0	84.7	C13 H17 N3 O2
	247.1334	-1.0	-4.0	6.5	78.8	C15 H19 O3

Figure 3.3.35 HRMS data for **131**; *m/z* [M – H]$^-$ calc. mass is 247.1334, 247.1324 was found.

The structural confirmation of **131** was accomplished *via* comprehensive utilization of both 1D and 2D NMR spectroscopy. The ^{1}H NMR spectral analysis revealed distinct resonances: a sole olefinic proton at δ_H 5.78 (d, J = 1.7 Hz), situated at position H-9; two vinyl protons at δ_H 5.71 (m) and δ_H 6.45 (m), corresponding to H-12; a pair of methyl signals at δ_H 0.92 (d, J = 5.7 Hz) and 1.19 (s), representing the methyl groups at C-14 and C-15, respectively (Table 3.3.17). Notably, sp^3 hybridized methine protons resonating at H-4 (δ_H 1.48, m) and H-7 (δ_H 3.56, dd, J = 12.0, 6.9 Hz) were also identified. Further characterization involved the identification of four sets of methylene protons with chemical shifts ranging from δ_H 1.46 to δ_H 2.02 (Table 3.3.17). The presence of a carbonyl group at C-8 was confirmed. A sequential arrangement was established for C-1/C-2/C-3/C-4, contributing to the configuration of the left cyclohexene ring by employing ^{1}H-^{1}H COSY correlations (Table 3.3.17).

The ^{13}C NMR spectroscopic analysis (Table 3.3.17) unveiled pertinent carbon resonances: two methyl groups, C-14 and C-15 (δ_C 15.3, δ_C 16.2); four sp^3 hybridized methylene carbons, C-1, C-2, C-3, C-6 (δ_C 33.2, δ_C 26.4, δ_C 30.5, δ_C 41.8); three sp^3 hybridized methine carbons, C-4, C-7 (δ_C 43.7, δ_C 45.90); and a single sp^3 hybridized quaternary carbon, C-5 (δ_C 39.9). Additionally, the spectra exhibited four olefinic carbons: C-9, C-10, C-11, C-12 (δ_C 123.7, δ_C 170.9, δ_C 139.0, δ_C 128.5, Table 3.3.17). The resonance of the carbonyl carbon was observed at δ_C 198.0, substantiating its assignment to C-8. This assignment found support in HMBC correlations, which revealed significant correlations between H-7 and C-8, as well as between H-6 and C-8 (Table 3.3.17). The confirmation of **131**'s structure was achieved, including the introduction of a keto group at C-8 and the incorporation of a carboxylic acid moiety at C-13, both on the basis of the (+)-aristolochene backbone. The coherent elucidation of this structural configuration of **131** as a novel compound, hitherto unreported in scientific literature.



Chemical Formula: $C_{15}H_{20}O_3$
Exact Mass: 248.1412

Table 3.3.17 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **131** recorded in CD_3OD.

Compound 131				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	33.2	2.32, 2H, m	1, 2, 9	2
2	26.4	1.46, 1H, m;	2, 1, 3	4
		1.87, 1H, m	2, 1, 3	4, 10
3	30.5	1.46, 1H, m;	3, 2, 4	2, 4, 5
		1.56, 1H, m	3, 2	1, 2, 4, 5
4	43.7	1.48, 1H, m	14	3, 5
5	39.9			
6	41.8	2.02, 2H, m	6, 7	5, 7, 8, 11, 13
7	45.9	3.56, 1H, dd (12.0, 6.9)	6	6, 8, 11, 12, 13
8	198.0			
9	123.7	5.78, 1H, d (1.7)	1	1, 5, 7
10	170.9			
11	139.0			
12	128.5	5.71, 1H, s;		7, 11, 13
		6.45, 1H, s		7, 11, 13
13	170.0			
14	15.3	0.92, 3H, d (5.7)	4	3, 4, 5
15	16.2	1.19, 3H, s		4, 5, 6, 10

3.3.10 Exp 7_ Expression of hr BGC w/o *hrl4* and *hrl5*

Exp7 involves the removal of *hrl4* and *hrl5*, leading to the co-expression of *hrtc*, *hrl1*, *hrl3, hrl7, hrl8* (Figure 3.3.36A). The ELSD analysis depicted the presence of **127**, **129**, and **130** as products, similar to the products of Exp4 (Figure 3.3.23B). Consequently, Exp7 substantiates the findings of Exp3, signifying that Hrl5 is not a requisite factor in hypoxylan-A **73** formation. Moreover, the results of Exp7 reinforce the significance of Hrl4 (FMO) in the generation of hypoxylan-A acid **126**, confirming the observations made in Exp4.

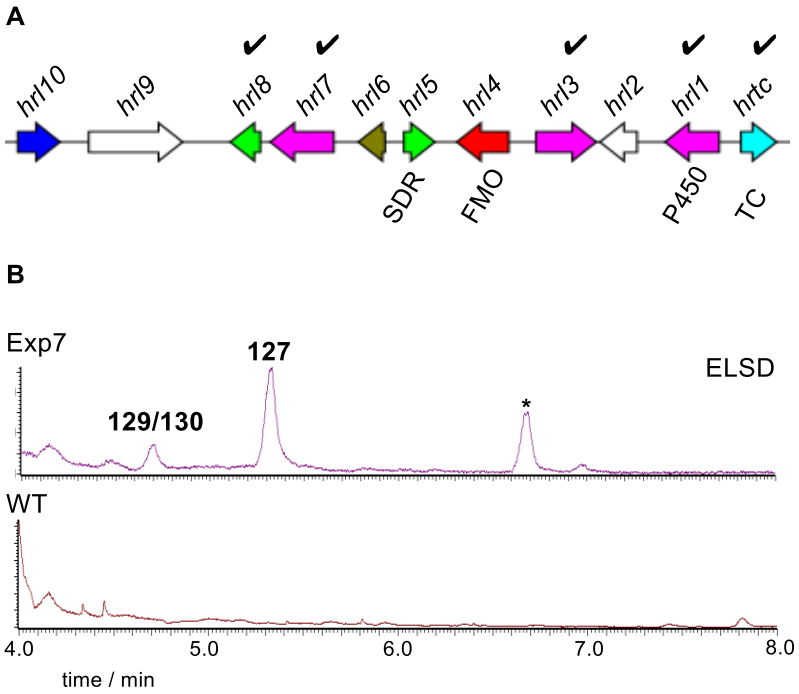


Figure 3.3.36 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp7 by HPLC analysis.

3.3.11 Exp 8_ Expression of *hrtc* and *hrl1*

Exp8, involving the co-expression of *hrtc, hrl1, hrl7*, and *hrl8*, also resulted in the production of **128** as the primary product (Figure 3.3.37), with phenol being present in the compound. Notably, Exp8 entailed the removal of *hrl3*, *hrl4*, and *hrl5* while expressing *hrl1*, *hrl7*, and *hrl8*. An earlier RT-PCR experiment confirmed the activity of Aol4 and Aol1, demonstrating their ability to fulfill the roles of Hrl7 and Hrl8, respectively. This indicates that Exp8 effectively only included the genes *hrtc* and *hrl1*. Therefore, when *hrl1* was co-expressed with *hrtc*, the synthesis of **128** was made possible. The presence of hydroxyl groups at C-2 and C-13 in **128** is likely formed by the *A. oryzae* shunt.

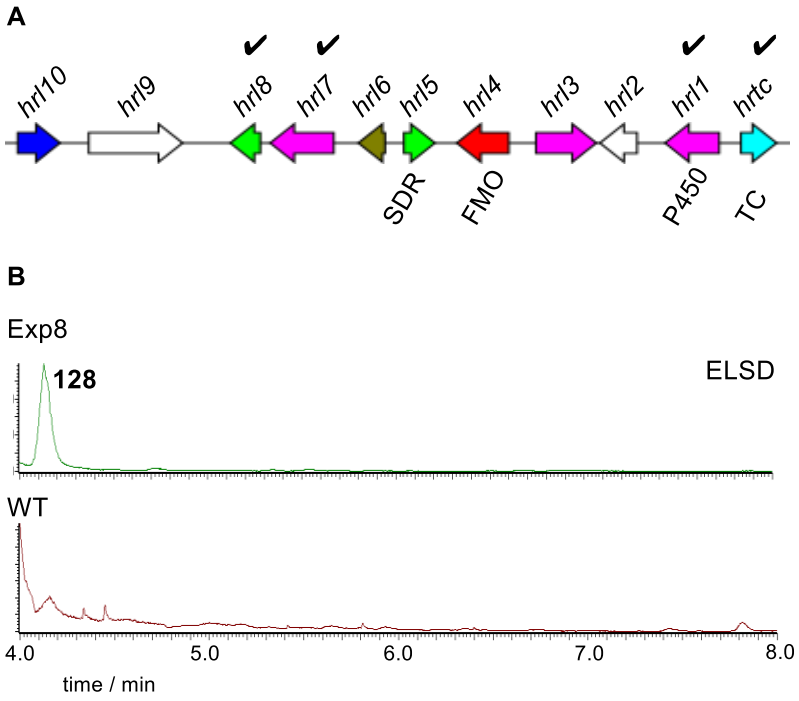


Figure 3.3.37 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp8 by HPLC analysis.

3.3.12 Exp 9_Co-expression of *hrtc* and pr genes

Analyzing the results from experiments 1 to 8, it becomes apparent that expression of *hrtc* in *A. oryzae* leads to the formation of (+)-aristolochene as well as compounds **124** and **125**. To be more specific, compound **124** exhibits a hydroxyl group situated at the C-8 position, whereas **125** has a carbonyl group at the same site. Our hypothesis postulates that the hydroxyl group was generated through the catalytic activity of the oxidative enzyme P450, Aol4 (Hrl7). This enzymatic step is then succeeded by a subsequent oxidation process by a SDR, Aol1 (Hrl8). Aol4 and Aol1 expressed from the native *A. oryzae* BGC are active in all experiments, in agreement with the earlier RT-PCR predictions (Section 3.3.2).

Subsequently, we embarked on a series of experiments utilizing a distinctive approach. We employed the host organism responsible for producing **124** and **125**. We then introduced tailoring genes sourced from the PR-toxin BGC into this host, enabling reactions with **124** or **125**.

Four genes (*prl3, prl4, prl7, prl9*) sourced from the PR-toxin BGC were selected for expression analysis. *Prl3* and *prl4* were individually integrated into the pTYGS_ade vector, under the control of P_{adh} and P_{gpdA} respectively. On the other hand, *prl7* and *prl9* were introduced in the pTYGS_met vector, driven by the P_{gpdA} and P_{eno} respectively. These two plasmids were then introduced into *A. oryze* NSAR1 to generate the transformants of Exp9. Following the completion of fermentation and crude extract preparation, the products underwent analysis *via* HPLC. The ELSD chromatograms divulged the emergence of new peaks when compared to those observed in the untransformed strain (Figure 3.3.38).

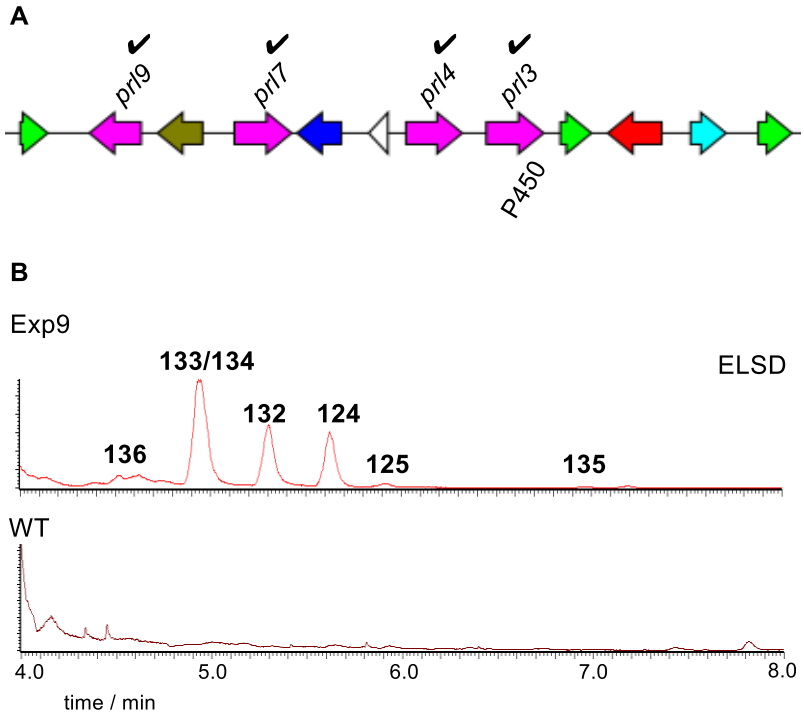


Figure 3.3.38 ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp9 and untransformed strain by HPLC analysis.

In Exp9, we observed five new peaks namely **132, 133, 134, 135** and **136** (Figure 3.3.38). These peaks emerged in conjunction with the previously identified products, **124** and **125**. The isolation of these compounds was accomplished employing preparative HPLC after a large-scale (2L) fermentation in DPY medium. The elucidation of structures of **132, 133, 134, 135** and **136** was undertaken through both 1D and 2D NMR spectroscopy, coupled with HRMS.

3.3.12.1 Characterization of 132

132 exhibited distinctive mass fragmentation patterns in the positive ion mode (ES^+) at *m/z* 251, 233, and 273, while the negative ion mode at *m/z* 249 (ES^-) demonstrated very weak signals (Figure 3.3.39B). The appearance of the *m/z* 233 can be attributed to the loss of a water molecule $[M + H - H_2O]^+$, with the *m/z* 251 and 273 peaks likely corresponding to the $[M + H]^+$ and $[M + Na]^+$ ions, respectively. These observations provide robust confirmation of the anticipated mass of *m/z* 251. Further substantiating this characterization, an absorption peak was observed in the UV spectrum at 237 nm within the wavelength range of 200-600 nm. This finding lends support to the presence of the aristolochene backbone in **132** (Figure 3.3.39A).

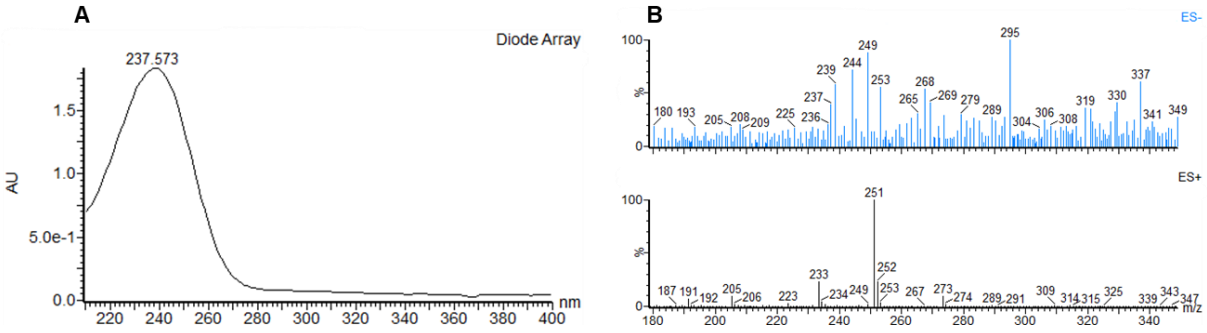
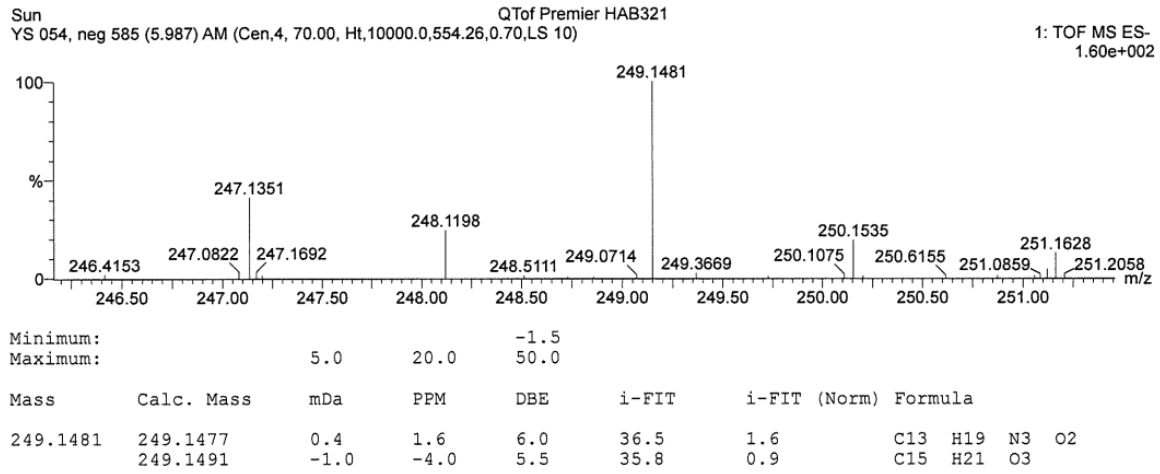


Figure 3.3.39 UV-absorption **(A)** and fragmentation pattern **(B)** of **132** in ES^- and ES^+ by LR-LCMS.

Through HRMS analysis, the molecular formula of **132** was conclusively determined as $C_{15}H_{22}O_3$. Theoretical calculations for the $[M - H]^-$ ion indicated an expected *m/z* value of 249.1491, closely aligning with the observed mass of 249.1481 (Figure 3.3.40).



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
249.1481	249.1477	0.4	1.6	6.0	36.5	1.6	C13 H19 N3 O2
	249.1491	-1.0	-4.0	5.5	35.8	0.9	C15 H21 O3

Figure 3.3.40 HRMS data for **132**; *m/z* $[M - H]^-$ calc. mass is 249.1491, 249.1481 was found.

The NMR analysis of **132** was undertaken based on the structure of either **124** or **125**. Given the UV absorption data, we opted to prioritize the utilization of **125** as a reference. **125** exhibited a UV absorption at 239 nm, whereas **124** displayed a UV absorption at 200 nm. As a result, we inferred that the keto group likely persisted within the structure of **132**.

The ^{1}H NMR spectrum unveiled a distinctive olefinic proton singlet at C-9 (δ_H 5.92, s), accompanied by a pair of vinyl protons at C-12 (δ_H 4.82, m; δ_H 4.98, m, Table 3.3.18). Furthermore, a trio of well-defined methyl signals materialized at C-13 (δ_H 1.70, m), C-14 (δ_H 0.96, d, J = 6.7 Hz), and C-15 (δ_H 1.32, s), each closely aligning with the characteristic pattern of **125**. However, a marked difference emerged in the case of **132**, where the NMR spectrum indicated the presence of two distinct sets of methylene protons within the chemical shift range of δ_H 1.62-1.96 ppm (Table 3.3.18). Adding to this distinction, **132** exhibited an additional methine proton in the low field region at δ_H 4.28 (d, J = 3.2 Hz).

The ^{13}C NMR dataset (Table 3.3.18) exhibited distinct carbon resonances, comprising three methyl groups (δ_C 20.2, δ_C 14.9, δ_C 18.1), two sp^3 methylene groups (δ_C 33.6, δ_C 43.0), four sp^3 methine groups

at δ_C 75.8, δ_C 71.0, δ_C 40.6, and δ_C 51.4, a single sp^3 quaternary carbon (δ_C 38.2), and four olefinic carbons (δ_C 128.9, δ_C 165.0, δ_C 143.3, δ_C 114.6). Notably, the carbonyl carbon signal (δ_C 199.6) was consistent with that observed in compound **125** (Table 3.3.18).

The DQF-COSY spectrum elucidated the sequential connectivity between proton resonances H-9, H-1, H-2, H-3 and H-4. This analysis established that the newly observed oxymethine signal (δ_H 4.28) was situated at the C-1 position. Furthermore, the other oxymethine signal was attributed to the C-2 position. The HMBC spectrum played a pivotal role in reinforcing the assignment of the oxymethine positions by correlations between the proton H-1 and carbon C-9, as well as between proton H-9 and carbon C-1. These spectroscopic findings confirmed the structure of **132**, characterizing the presence of an additional hydroxyl group at the C-1 position, distinguishing it from the structure of **125**.

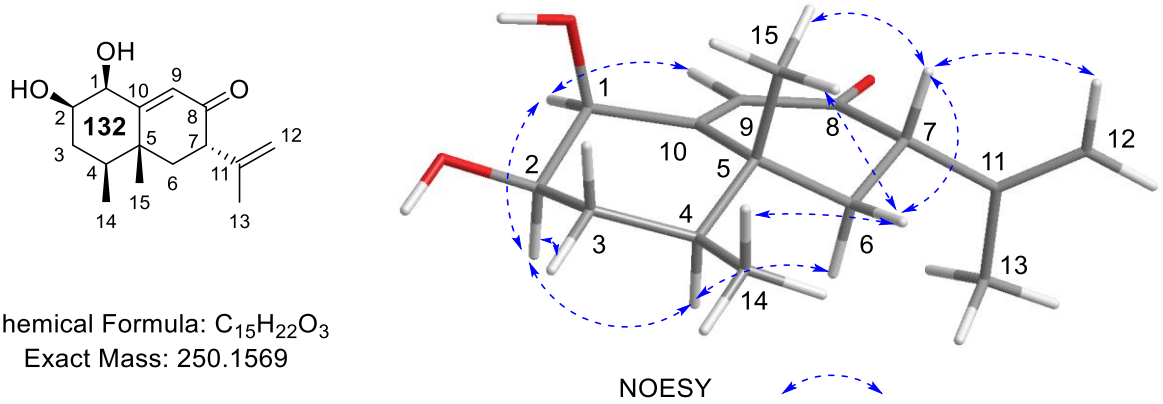


Figure 3.3.41 Structure and a stick model of **132** for the NOESY.

Table 3.3.18 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **132** recorded in $CDCl_3$.

Compound 132				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	75.8	4.28, 1H, d (3.2)	2, 3	2, 3, 5, 9, 10
2	71.0	3.73, 1H, ddd (11.7, 4.7, 3.5)	3, 1	1, 3, 4
3	33.6	1.62, 1H, dddd (12.9, 4.6, 3.2, 1.2);	3, 4, 2, 1	1, 2, 4, 5
		1.85, 1H, m	3, 4, 2	1, 2, 4, 5
4	40.6	1.48, 1H, m	3, 14	3, 4, 5
5	38.2			
6	43.0	1.85, 1H, m;	6, 7	5, 7, 8, 10, 15
		1.97, 1H, dd (13.0, 4.4)	6, 7	5, 7, 8, 10, 15
7	51.4	3.24, 1H, dd (14.4, 4.4)	6	5, 6, 8, 11, 12, 13
8	199.6			
9	128.9	5.92, 1H, s		1, 5, 7, 8, 10
10	165.0			
11	143.3			
12	114.6	4.82, 1H, m;	12, 13	7, 11, 13
		4.98, 1H, m	12, 13	7, 11, 13
13	20.2	1.7, 3H, m	12	7, 11, 12
14	14.9	0.96, 3H, d (6.9)	4	3, 4, 5
15	18.1	1.32, 3H, s		4, 5, 6

The determination of the stereochemical configurations of H-1 in **132** relied on the analysis of coupling constants between crucial protons neighboring the stereocenters. Specifically, the positioning of H-1 was proposed to be equatorial (H-1_{eq}), a proposal substantiated by proton coupling constants between H-1_{eq} and H-2_{ax} (^3J = 3.2 Hz). The stereochemical configurations of H-2 were further corroborated through analyses involving coupling constants. This included relationships such as H-2_{ax}/H-1_{eq} (^3J = 3.5 Hz), H-2_{ax}/H-3_{ax} (^3J = 11.7 Hz), and H-2_{ax}/H-3_{eq} (^3J = 4.7 Hz). These coupling constants provided decisive evidence regarding the stereochemical arrangement of H-2 in the molecule.

The NOESY correlations of H-1_{eq}/H-2_{ax} indicated that H-1_{eq} and H-2_{ax} were on the same face of the (+)-aristolochene backbone. In addition, correlation between H-2_{ax}/H-3_{eq}, H-2_{ax}/H-4_{ax} confirmed the axial position of H-2 (Figure 3.3.41). Therefore, the relative configuration of H-1 was determined to be at the bottom, and the hydroxyl group at the top. This relative arrangement was elucidated as 1*S* based on the comprehensive spectroscopic evidence.

3.3.12.2 Characterization of 133

Up to this point, we have successfully ascertained the structure of **132**, wherein a hydroxyl group is positioned at C-1. This hydroxyl group emerges as a result of a reaction involving substrates **125** and the enzyme expressed within this specific transformant. Our research endeavors now shift towards the ongoing task of elucidating the compositions of the remaining novel peaks in Exp 9.

133 and **134** exhibited identical retention times and experienced overlap under the HPLC conditions spanning from 10% to 90% acetonitrile over a 10-minute gradient as usual. In response, the LCMS methodology was adjusted, employing a gradient ranging from 10% to 50% acetonitrile over the same 10-minute duration. This modification resulted in the distinct separation of the peaks corresponding to **133** and **134**, leading to their appearance at disparate retention times (Figure 3.3.42).

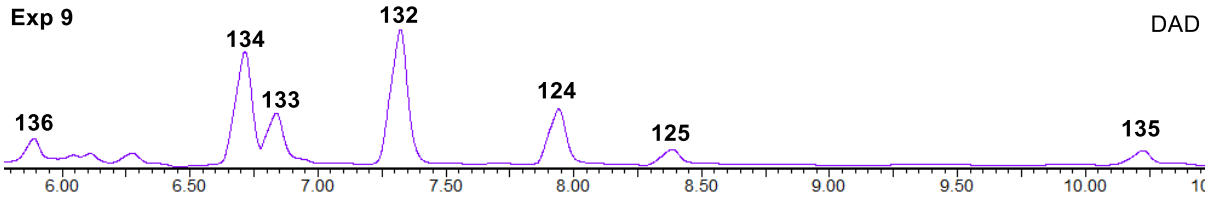


Figure 3.3.42 DAD chromatograms of Exp 9 by 10-50% acetonitrile over a 10-minute gradient.

133 exhibited distinctive patterns of mass fragmentation upon ionization in the positive ion mode (ES^+) at *m/z* 217, 235, 253, and 275. Conversely, the negative ion mode yielded weak signals that were not discernible (Figure 3.3.43). The presence of the *m/z* 235 can be attributed to the elimination of a water molecule $[M + H - H_2O]^+$, while the *m/z* 253, 275, and 217 peaks are attributed to the $[M + H]^+$, [M +

Na] $^{+}$, and [M + H - $2H_2O$] $^{+}$ ions, respectively. These observations provide part-confirmation of the anticipated mass of 253.

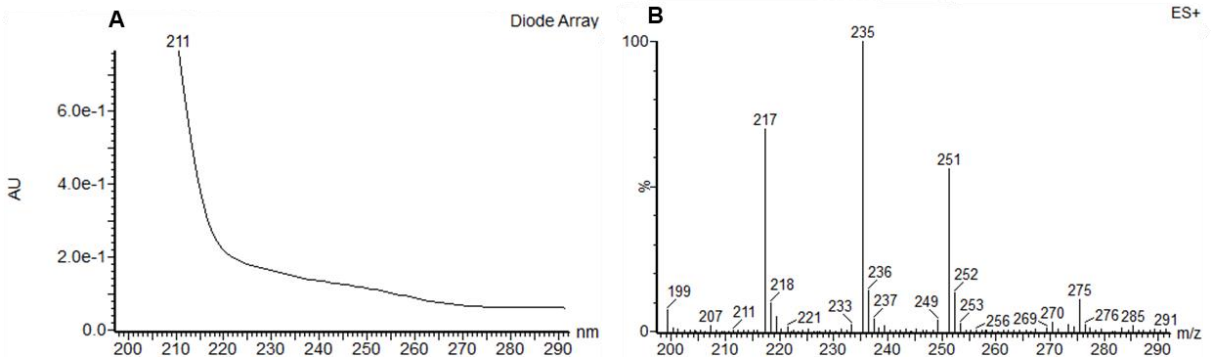


Figure 3.3.43 UV-absorption **(A)** and fragmentation pattern **(B)** of **133** in ES^{+} by LR-LCMS.

Further augmenting this characterization, an absorption peak was observed within the UV spectrum at 211 nm, operating within the wavelength range of 200-600 nm, which is fit with that of **124**. This finding lends credence to the presence of the aristolochene backbone within **133** (Figure 3.3.45). Employing HRMS analysis, the molecular formula of **133** was ascertained as $C_{15}H_{24}O_3$. Theoretical calculations concerning the [M + H] $^{+}$ ion yielded an anticipated *m/z* value of 253.1804, which closely aligns with the observed mass of 253.1798 (Figure 3.3.44).

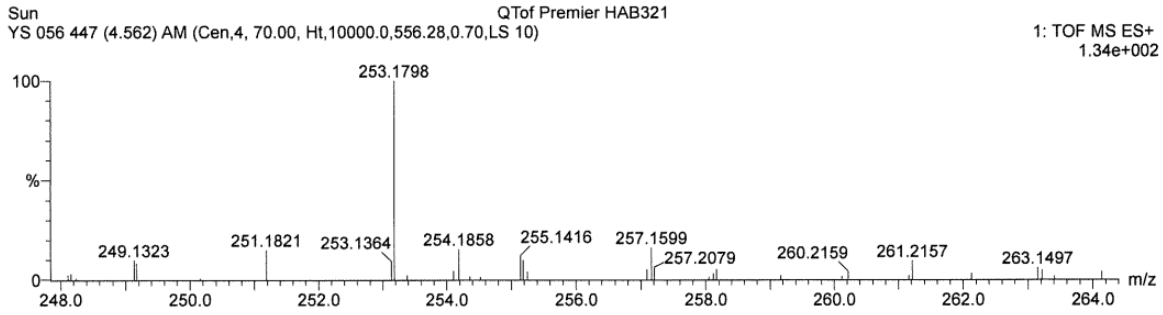


Figure 3.3.44 HRMS data for **133**; *m/z* [M + H] $^{+}$ calc. mass is 253.1804, 253.1798 was found.

The NMR analysis of **133** was carried out with a structural context derived from either **124** or **125**. Considering the UV absorption data, we chose to accord higher priority to employing **124** as a point of helper. **124** exhibited a UV absorption peak at 211 nm, which closely aligns with the UV profile of **133**. This stands in contrast to **125**, which manifested its peak at 239 nm. In light of these spectral distinctions, we postulated that **133** could be regarded as a derivative rooted in the structural framework of **124**. This derivative would involve the replacement of the keto group at C-8 with a hydroxy group. This supposition is substantiated by the consideration that both **124** and **125**, as well as the (+)-aristolochene synthesized by the host, possess the potential to serve as substrates in this reaction context.

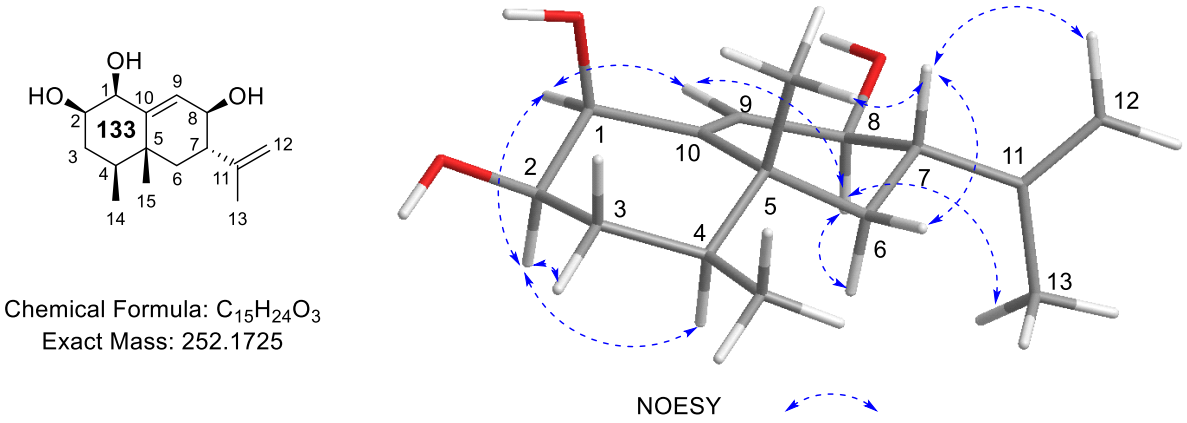


Figure 3.3.45 Structure and a stick model of **133** for the NOESY.

Table 3.3.19 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **133** recorded in MeOD.

Compound 133				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	77.7	4.08, 1H, dd (3.4, 1.2)	2, 3	2, 3, 5, 9
2	73.2	3.51, 1H, ddd (11.8, 4.5, 3.4)	1, 3	1, 3
3	34.7	1.46, 1H, dddd (12.8, 4.5, 3.2, 1.2);	3, 4, 2, 1	1, 2, 4, 5, 14
		1.81, 1H, ddd (12.7, 12.7, 12.7)	3, 4, 2	1, 2, 4, 5, 14
4	42.8	1.3, 1H, m	3, 14	3, 5, 14
5	38.6			
6	44.9	1.33, 1H, m;	6, 7	5, 7, 8, 10, 15
		1.59, 1H, dd (13.0, 2.6)	6, 7	5, 7, 8, 10, 15
7	47.9	2.34, 1H, ddd (12.7, 9.7, 2.6)	6, 8	6, 8, 11, 12, 13
8	70.6	4.04, 1H, dd (9.7, 1.9)	7, 9	7, 9, 10, 11, 13
9	132.0	5.58, 1H, d (1.9)	8	1, 5, 7, 10, 11
10	146.2			
11	148.3			
12	111.9	4.83, 2H, m	13	7, 11, 13
13	20.1	1.74, 3H, m	12	7, 11, 12
14	15.2	0.85, 3H, d (6.9)	4	3, 4, 5
15	20.5	1.18, 3H, s		4, 5, 6, 10

Comparison of the ^{1}H NMR spectra of **133** and **124** revealed a conspicuous difference: **133** displayed a lack of one methylene proton signal (Table 3.3.19). Additionally, a distinctive feature in the spectrum of **133** was the presence of an extra methine proton signal, resonating in the low-field region at δ_H 4.08 (dd, J = 3.2, 1.2 Hz), a characteristic not observed in the spectrum of **124** (Table 3.3.19). Remaining proton assignments in **133** were largely analogous to those in **124**, encompassing the olefinic proton at C-9 (δ_H 5.58, d, J =1.9), a pair of vinyl protons at C-12 (δ_H 4.83, 2H, m), and three clearly resolved methyl signals (δ_H 1.70, m; δ_H 0.96, d, J =6.7; δ_H 1.32, s, Table 3.3.19).

The ^{13}C NMR spectrum of **133** portrayed distinct carbon resonances, embracing three distinct methyl groups (δ_C 20.1, δ_C 15.2, δ_C 20.5), two sp^3 methylene groups (δ_C 34.7, δ_C 44.9), five sp^3 methine groups at δ_C 77.7, δ_C 73.2, δ_C 42.8, δ_C 47.9, and δ_C 70.6, one sp^3 quaternary carbon (δ_C 38.6), and four olefinic carbons (δ_C 132.0, δ_C 146.2, δ_C 148.3, δ_C 111.9). A difference between **133** and **124** lies in the chemical

shift of C-1, which now locates at δ_C 77.7. In comparison to **132**, the presence of a hydroxyl group at C-1 is still evident, while the carbonyl group at C-8 is absent in **133** (Table 3.3.19).

The DQF-COSY spectrum facilitated the elucidation of proton resonance interconnections between H-1 and H-2. Through this analysis, the oxymethine signals (δ_H 4.08, δ_H 4.04) were positioned at the C-1 and C-8 positions respectively. Subsequently, the HMBC spectrum played a pivotal role in reinforcing the assignment of these oxymethine positions by correlations between proton H-1 and carbon C-9. The presence of an additional hydroxyl group at the C-1 position was delineated, distinguishing **133** from the configuration of **124**.

The determination of the stereochemical configuration for H-1, H-2, and H-8 in **133** hinged upon the coupling constants among pivotal protons proximal to the stereocenters. The orientation of H-1 was inferred to be equatorial (H-1_{eq}), as indicated by proton coupling constants between H-1_{eq} and H-2_{ax} (^3J 3.4). The stereochemical arrangements of H-2 were further validated as H-2_{ax} through coupling constants involving H-2_{ax}/H-1_{eq} (^3J = 3.4 Hz), H-2_{ax}/H-3_{ax} (^3J = 11.8 Hz), and H-2_{ax}/H-3_{eq} (^3J = 4.5 Hz). As for H-8, they were ascertained as H-8_{ax} through coupling constants established between H-8_{ax} and H-7_{ax} (^3J = 9.7 Hz), as well as H-8_{ax} and H-9 (^3J = 1.9 Hz). These coupling constants collectively contributed to a comprehensive understanding of the stereochemistry within **133**.

The NOESY correlations observed between H-1_{eq} and H-2_{ax} in **133** also provided evidence that H-1_{eq} and H-2_{ax} were spatially situated on the same face. Furthermore, the correlations between H-2_{ax} and H-3_{eq}, as well as H-2_{ax} and H-4_{ax}, substantiated the axial orientation of H-2 (Figure 3.3.45). Consequently, the relative configuration of H-1 was deduced to be at the bottom, with the hydroxyl group situated at the top.

The structure of **132** and **133** all have one more oxidative hydroxyl at C-1 on the base of substrates **124**, **125** respectively. This result likely stems from the same reaction, executed by the same enzyme. **134**, which has very close retention time with **133** (Figure 3.3.42), was purified and the structure was also elucidated by 1D and 2D NMR and HRMS.

3.3.12.3 Characterization of 134

134 exhibited mass fragmentation patterns in the positive ion mode (ES^+) at *m/z* 233, 251, and 273, while the negative ion mode displayed notably weak signals that were not discernible (Figure 3.3.46). The emergence of the 233 fragment is attributed to the loss of a water molecule $[M + H - H_2O]^+$, whereas the peaks at *m/z* 251 and 273 are indicative of the $[M + H]^+$ and $[M + Na]^+$ ions, respectively. These observations collectively establish partial confirmation of the projected mass of 250 (Figure 3.3.46). An absorption peak at 244 nm within the wavelength range of 200-600 nm was evident in the UV spectrum. This observation further substantiates the presence of the aristolochene backbone and possibly the substrate is **125** because of the absorption peak at 244 nm (Figure 3.3.46).

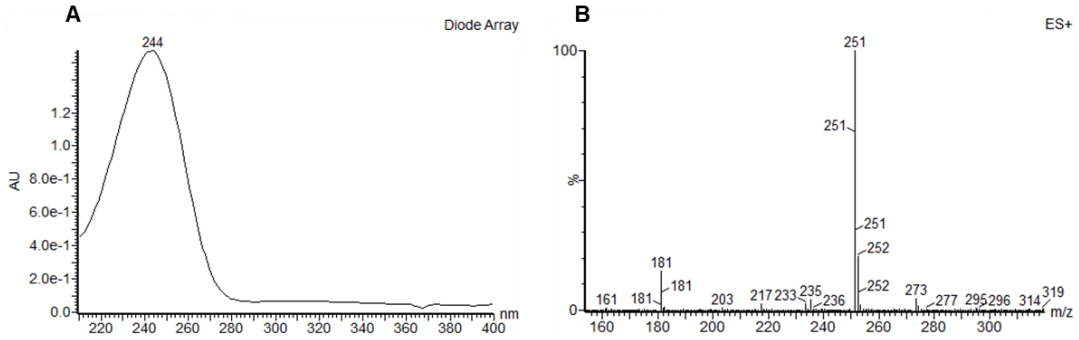


Figure 3.3.46 UV-absorption **(A)** and fragmentation pattern **(B)** of **134** in ES^+ by LR-LCMS.

By means of HRMS analysis, the molecular formula of **134** was established as $C_{15}H_{22}O_3$. Theoretical calculations for the $[M + H]^+$ ion yielded an anticipated *m/z* value of 251.1647, which closely harmonized with the observed mass of 251.1658 (Figure 3.3.47).

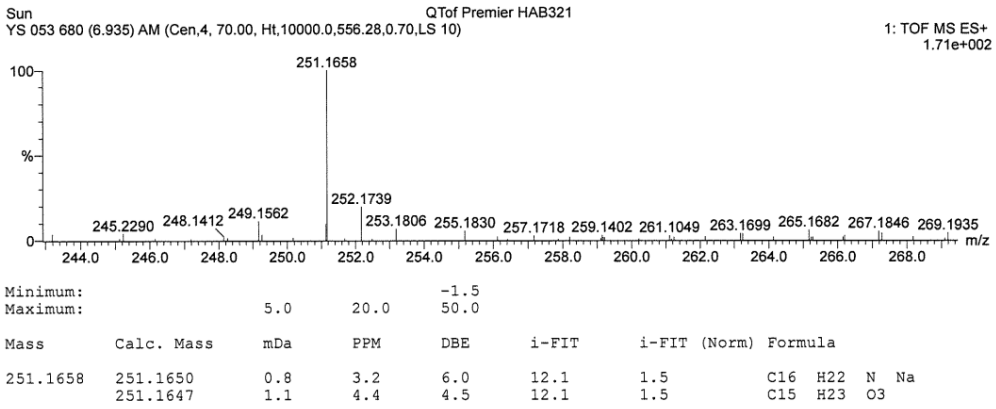


Figure 3.3.47 HRMS data for **134**; *m/z* $[M + H]^+$ calc. mass is 251.1647, 251.1658 was found.

The mass information and UV spectrum of **134** align closely with those of **132**. They share an identical proposed molecular formula. However, the key distinction between them is believed to lie in the arrangement resulting from the oxidative reaction. In the case of **132**, hydroxyl groups are situated at positions C-1 and C-2. While we initially presumed that both compounds feature a hydroxyl group at C-2, further analysis revealed a divergence: **134** possesses a hydroxyl group at C-3, or alternatively, a hydroxyl group at C-1 with an opposing stereochemical configuration.

The ^1H and ^{13}C NMR data of **134** were subjected to a comparative analysis with the NMR spectra of **132**, revealing highly analogous assignments between the two. Notably, the ^1H NMR spectrum of **134** exhibits an olefinic proton singlet at C-9 (δ_H 5.82, s), along with vinyl protons at C-12 (δ_H 4.81, m; δ_H 4.97, m), and methyl signals manifesting at C-13 (δ_H 1.72, m), C-14 (δ_H 1.16, d, J = 7.1), and C-15 (δ_H 1.36, s). These resonances closely mirror those observed in the ^1H NMR spectrum of **132**. Additionally, **134** possesses two oxygenated methines (δ_H 3.73, m; δ_H 3.86, m).

The ^{13}C NMR spectrum of **134** reveals the presence of three distinct methyl groups at δ_C 20.2, δ_C11.9, and δ_C 18.8, as well as two sp^3 methylene groups at δ_C 36.2 and δ_C 42.2. Further analysis indicates the

existence of four sp^3 methine groups, located at δ_C 44.5 and δ_C 50.6, among which two are oxygenated and resonate at δ_C 71.4 and δ_C 74.5. A solitary sp^3 quaternary carbon is observed at δ_C 38.6, while four olefinic carbons emerge at δ_C 125.7, δ_C 167.1, δ_C 143.5, and δ_C 114.6. Significantly, the signal corresponding to the carbonyl carbon (δ_C 199.1) aligns precisely with that found in compound **132**.

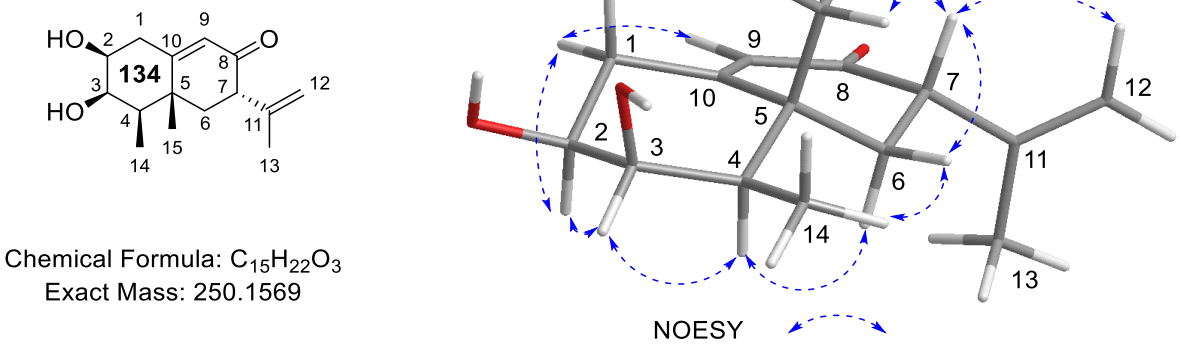


Figure 3.3.48 Structure and a stick model of **134** for the NOESY.

Table 3.3.20 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **134** recorded in MeOD.

Compound 134				
Pos.	***δC* / ppm**	***δH* / ppm (*J*/Hz)**	**1H-1H COSY**	**HMBC (H-C)**
1	36.2	2.34, 1H, dd (13.1, 4.7);	1, 2	2, 3, 9, 10
		2.79, 1H, ddd (13.1, 12.5, 2.1)	1, 2, 9	2, 3, 9, 10
2	71.4	3.73, 1H, ddd (12.2, 4.8, 3.1)	1, 3	
3	74.5	3.86, 1H, dd (3.3, 3.3)	2, 4	2, 4, 5, 14
4	44.5	1.47, 1H, dddd (7.1, 7.1, 7.1, 2.6)	14, 3	5, 6, 14, 15
5	38.6			
6	42.2	1.78, 1H, dd (13.7, 13.7);	6, 7	4, 5, 7, 8, 15
		1.98, 1H, dd (13.0, 4.5)	6, 7	5, 7, 8, 10, 15
7	50.6	3.15, 1H, dd (14.4, 4.5)	6	5, 6, 8, 11, 12, 13
8	199.1			
9	125.7	5.82, 1H, d (1.9)	1	1, 5, 7
10	167.1			
11	143.5			
12	114.6	4.81, 1H, m;	12, 13	7, 11, 13
		4.97, 1H, m	12, 13	7, 11, 13
13	20.2	1.72, 3H, m	12	7, 11, 12
14	11.8	1.16, 3H, d (7.1)	4	3, 4, 5
15	18.8	1.36, 3H, s		4, 5, 6, 10

The HMBC spectroscopic analysis played a pivotal role in pinpointing the precise locations of the oxygenated methines. Notably, H-9 demonstrated a strong correlation with C-1, and reciprocally, H-1 exhibited a correlation with C-9. The HSQC data further illuminated the proton multiplicity at C-1. Moreover, insights from the ^{1}H-^{1}H COSY experiment indicated that H-4 is correlated with H-3, which was corroborated by HSQC revealing the presence of a lone proton at C-3. This accumulation of evidence led us to hypothesize that one hydroxyl group is positioned at C-3 rather than C-1. By extending our analysis to the ^{1}H-^{1}H COSY correlations between H-1 and H-2, and H-2 and H-3, in conjunction with HSQC data, a coherent depiction emerged: the other hydroxyl group is located at C-2.

In conclusion, we can confidently affirm that the arrangement comprises two hydroxyl groups, precisely positioned at C-2 and C-3.

The stereochemical configuration of H-2 and H-3 in **134** was ascertained by analyzing coupling constants among pivotal protons in close proximity to the stereocenters. The orientation of H-2 was determined to be axial (H-2_{ax}), substantiated by proton coupling constants between H-2_{ax} and H-1_{ax} (^3J 12.2), and H-2_{ax} and H-1_{eq} (^3J 4.7). Likewise, the stereochemical arrangement of H-3 was confirmed as H-3_{eq} through coupling constants involving H-3_{eq} and H-2_{ax} (^3J 3.3), and H-3_{eq} and H-4_{ax} (^3J 3.3). These coupling constants collectively provided a comprehensive understanding of the stereochemistry within **134**.

NOESY correlations were observed between H-4_{ax} and H-3_{eq} in **134**. Subsequently, H-3_{eq} exhibited correlations with H-2_{ax}, indicating a spatial arrangement where H-4, H-3, and H-2 were positioned on the same face (Figure 3.3.48). Further support for this arrangement came from correlations between H-2_{ax} and H-1_{eq}, which substantiated the axial orientation of H-2 (Figure 3.3.48). Consequently, the relative configuration of H-2 was deduced to position it at the bottom as H-2_{ax}, with the hydroxyl group situated at the top. Similarly, H-3 was deduced to be at the bottom as H-3_{eq}, with the hydroxyl group located at the top. The absolute configuration at C-2 and C-3 are 2*S* and 3*R* respectively.

3.3.12.4 Characterization of 135

So far, the elucidation of the structures of the prominent products **132, 133**, and **134** has been accomplished. Two further minor products, **135** and **136** were isolated, purified, and subsequently subjected to NMR analysis. In the positive ion mode (ES^+), **135** demonstrated mass fragmentation patterns at *m/z* 235, 217, and 257. Conversely, the negative ion mode produced weak signals that were indistinguishable (Figure 3.3.49). The appearance of the *m/z* 217 can be ascribed to the loss of a water molecule $[M + H - H_2O]^+$, whereas the peaks observed at *m/z* 235 and 257 correspond to the $[M + H]^+$ and $[M + Na]^+$ ions, respectively. These findings collectively provide partial support of the anticipated mass of 234 (Figure 3.3.49). The UV spectra displayed a distinct absorption peak at a wavelength of 232 nm within the range of 200-600 nm.

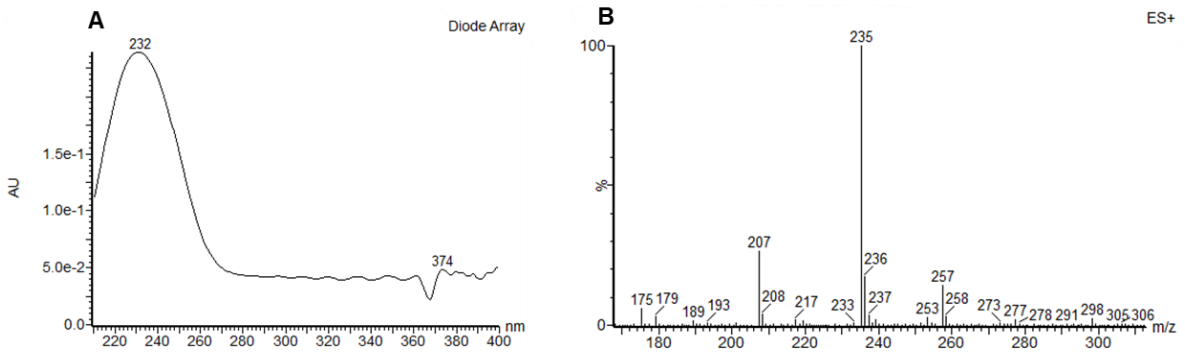
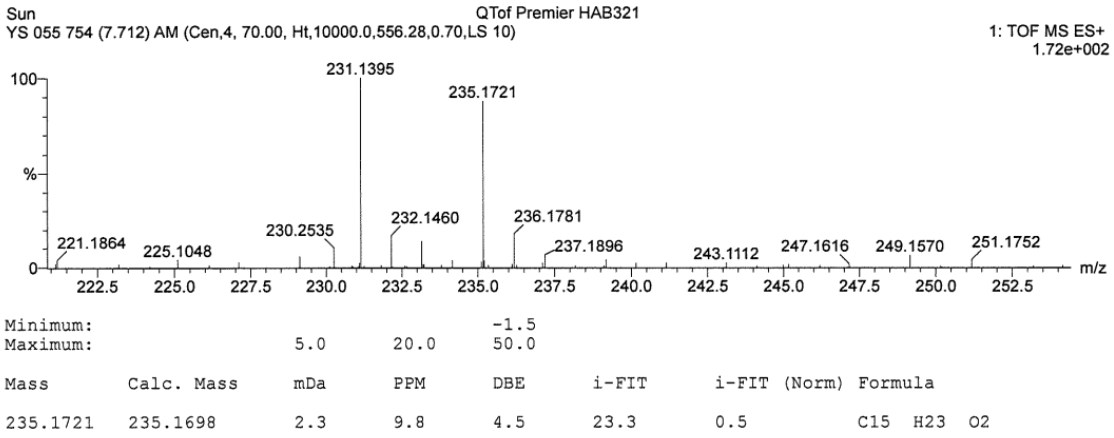


Figure 3.3.49 UV-absorption **(A)** and fragmentation pattern **(B)** of **135** in ES^+ by LR-LCMS.

Using HRMS analysis, we established that **135** has a molecular formula of $C_{15}H_{22}O_2$. Theoretical calculations for the $[M + H]^+$ ion yielded an anticipated *m/z* value of 235.1698, which closely matched the observed mass of 235.1721 (Figure 3.3.50)



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
235.1721	235.1698	2.3	9.8	4.5	23.3	0.5	C15 H23 O2

Figure 3.3.50 HRMS data for **135**; *m/z* $[M + H]^+$ calc. mass is 235.1698, 235.1721 was found.

The ^{1}H NMR spectrum of **135** revealed an olefinic proton singlet at C-9 (δ_H 5.84, s), accompanied by a pair of vinyl protons at C-12 (δ_H 4.82, m; δ_H 4.98, m). Additionally, three well-defined methyl signals appeared at C-13 (δ_H 1.72, dd, *J* 1.5, 0.8 Hz), C-14 (δ_H 0.95, d, *J* 6.8 Hz), and C-15 (δ_H 1.37, s), each closely resembling **132**'s characteristic pattern. However, **135**'s NMR spectrum indicated the presence of one more set of methylene protons (δ_H 1.69, m; δ_H 2.01, m, Table 3.3.21).

The ^{13}C NMR dataset of **135** reveals distinctive carbon resonances, encompassing three methyl groups (δ_C 20.1, 15.3, 18.4), three sp^3 methylene groups (δ_C 33.0, 24.9), three sp^3 methine groups at δ_C 73.2, 43.5, and 51.6, a lone sp^3 quaternary carbon (δ_C 39.1), and four olefinic carbons (δ_C 126.7, 167.5, 143.6, 114.5, Table 3.3.21). Notably, the carbonyl carbon signal (δ_C 200.0) aligns with that observed in compound **132**. Upon scrutinizing the ^{1}H and ^{13}C NMR spectra, we discerned the presence of a singular hydroxyl group within the backbone of **135**, setting it apart from **132**.

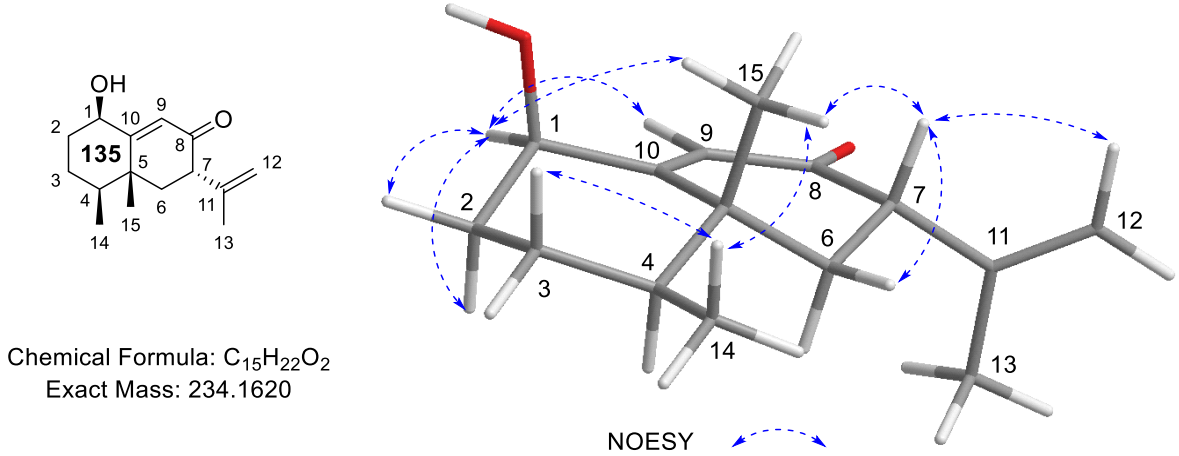


Figure 3.3.51 Structure and a stick model of **135** for the NOESY.

Table 3.3.21 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **135** recorded in $CDCl_3$.

Compound 135				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	73.2	4.33, 1H, dd (3.0, 3.0)	2	3, 5, 9
2	33.0	1.69, 1H, m;	2, 1, 3	3, 4
		2.01, 1H, m	2, 1, 3	1, 3, 4, 10
3	24.9	1.41, 1H, m;	3, 2, 4	2, 4, 5
		1.88, 1H, m	3, 2, 4	2, 4, 5
4	43.5	1.46, 1H, m	3, 14	14
5	39.1			
6	43.1	1.89, 1H, m;	6, 7	5, 7, 8, 10
		1.95, 1H, m	6, 7	5, 7, 8, 10
7	51.6	3.24, 1H, dd (14.3, 4.6)	6	6, 8, 11, 12, 13
8	200.0			
9	126.7	5.84, 1H, s		1, 5, 7
10	167.5			
11	143.6			
12	114.5	4.82, 1H, m;	12, 13	7, 13
		4.98, 1H, m	12, 13	7, 13
13	20.1	1.72, 3H, dd (1.5, 0.8)	12	7, 11, 12
14	15.3	0.95, 3H, d (6.8)	4	3, 4, 5
15	18.4	1.37, 3H, s		4, 5, 6, 10

The HMBC helped to locate the position of oxygenated methine. Specifically, H-9 displayed correlations with C-1, and conversely, H-1 exhibited correlations with C-9. The HSQC analysis revealed that C-1 is associated with one proton. Furthermore, the ^{1}H, ^{1}H-COSY data indicated correlations between H-1 and H-2, as well as between H-2 and H-3. Combining this information with the HSQC results, we conclusively determined the presence of a hydroxyl group at C-1.

The stereochemical configuration of H-1 in **135** was determined by analyzing the coupling constants among pivotal protons in close proximity to the stereocenter. Our analysis indicated that the orientation of H-1 is equatorial (H-1_{eq}), as supported by the proton coupling constants observed between H-1_{eq} and H-2_{ax} (^{3}J 3.0 Hz) and between H-1_{eq} and H-2_{eq} (^{3}J 3.0 Hz, Figure 3.3.51). These findings were further

corroborated by the NOESY correlations observed between H-1_{eq} and H-2_{eq}, as well as between H-1_{eq} and H-2_{ax} (Figure 3.3.51). Consequently, we can confidently deduce that the relative configuration of H-1 is equatorial, corresponding to 1*R*.

3.3.12.5 Characterization of 136

Another minor product, **136**, eluted at an earlier retention time (Figure 3.3.42). In the positive ion mode (ES^+), **136** exhibited mass fragmentation patterns with peaks detected at *m/z* 249, 267, and 533. In the negative ion mode, these peaks were observed at *m/z* 265 and 531. The presence of the *m/z* 249 can be attributed to the removal of a water molecule $[M + H - H_2O]^+$. The peaks noted at *m/z* 267 and 533 correspond to the $[M + H]^+$ and $[2M + H]^+$ ions, respectively. Likewise, the appearance of the *m/z* 265 peak likely represents the $[M - H]^-$ ion, and the *m/z* 531 peak represents $[2M - H]^-$. These findings collectively offer partial confirmation of the expected mass of 266 (Figure 3.3.52). The UV spectra displayed a significant absorption peak at a wavelength of 237 nm within the 200-600 nm range, offering support for the presence of the keto group at C-8 in accordance with the previously established product structures.

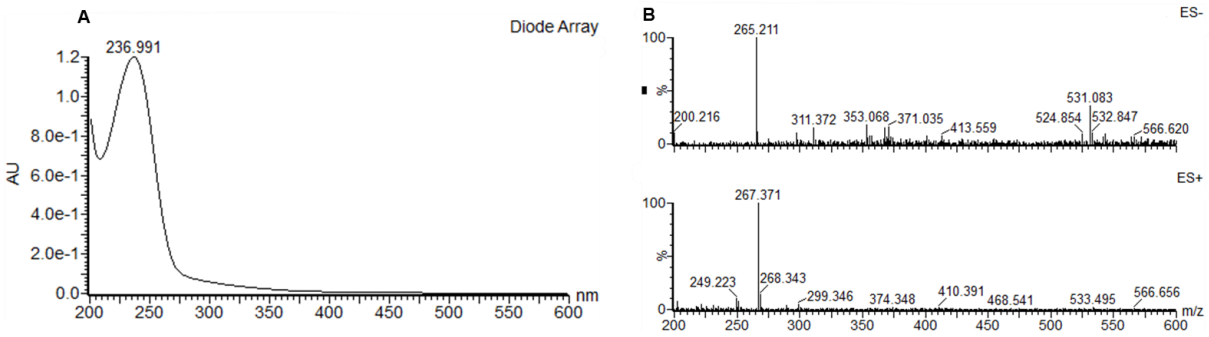
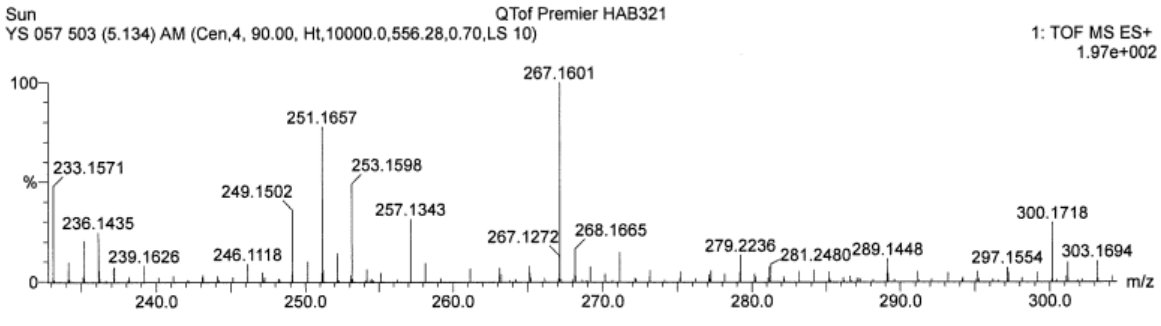


Figure 3.3.52 UV-absorption **(A)** and fragmentation pattern **(B)** of **136** in ES^- (top) and ES^+ (bottom) by LR-LCMS.

Through HRMS analysis, we determined that **136** possesses a molecular formula of $C_{15}H_{22}O_4$. Theoretical calculations for the $[M + H]^+$ ion predicted an expected *m/z* value of 267.1596, closely mirroring the observed mass of 267.1601 (Figure 3.3.53).



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
267.1601	267.1596	0.5	1.9	4.5	24.0	0.2	C15 H23 O4

Figure 3.3.53 HRMS data for **136**; *m/z* $[M + H]^+$ calc. mass is 267.1596, 267.1601 was found.

Based on the mass and UV data, we postulated that **136** contains three hydroxyl groups along its backbone, with the carbonyl group still positioned at C-8. Drawing inspiration from the structure of PR-toxin, we proposed that these three hydroxyl groups are situated at C-1, C-2, and C-3.

The ^{1}H-NMR analysis of **136** provided further insights. It revealed the presence of three protons residing on oxygenated methines at δ_H 4.32 (dd, *J* 3.4, 1.5 Hz), δ_H 3.6 (dd, *J* 3.4, 3.4 Hz), and δ_H 3.88 (ddd, *J* 3.0, 3.0, 1.5 Hz). Additionally, we identified one olefinic proton at δ_H 5.87 (s), two vinyl protons at δ_H 4.82 (m) and δ_H 4.93 (m), three distinct methyl signals at δ_H 1.69 (m), 1.19 (d, *J* 7.1 Hz), and 1.53 (s), as well as a pair of methylene protons at δ_H 1.9 (m) and δ_H 2.02 (dd, *J* 12.8, 4.5 Hz). Furthermore, we assigned two methines at δ_H 1.56 (m) and δ_H 3.35 (m), typically located at C-4 and C-5 (Table 3.3.22).

The HSQC analysis was instrumental in connecting specific carbon atoms with their corresponding protons. This allowed us to precisely identify three oxygenated methines at δ_C 79.2, 72.1, and 77.8, as well as olefinic carbons (δ_C 128.3, 168.8), vinyl carbons (δ_C 144.8, 114.8), methyl groups (δ_C 20.2, 12.1, 21.4), a methylene group (δ_C 45.6), two methine groups at (δ_C 45.9, 52.1), and a quaternary carbon (δ_C 38.9). Notably, the carbonyl carbon signal (δ_C 202.0) matched that observed in the previous compound. These findings provided conclusive evidence for the presence of three hydroxyl groups integrated into the backbone of **136**, thereby validating its structural composition (Table 3.3.22).

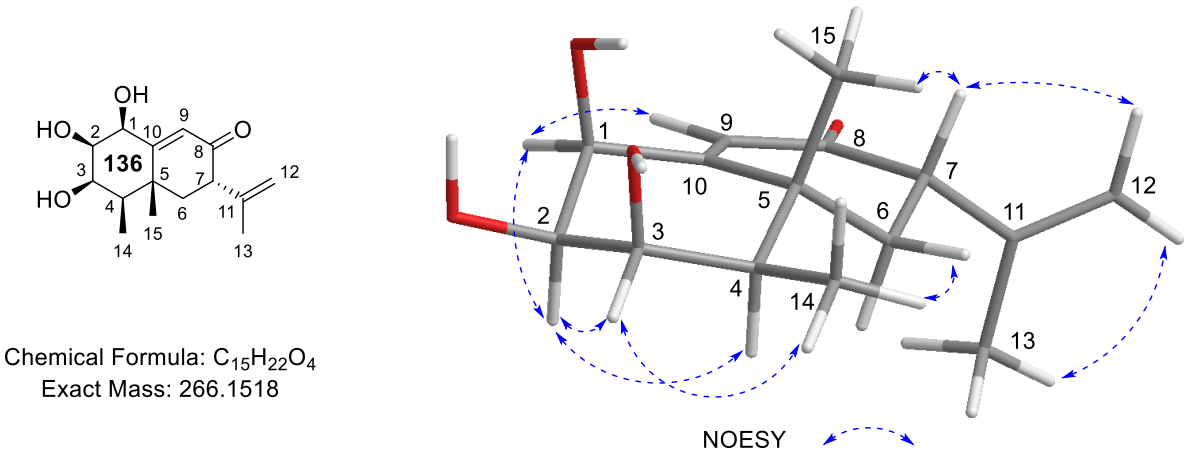


Figure 3.3.54 Structure and a stick model of **136** for the NOESY.

Table 3.3.22 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **136** recorded in MeOD.

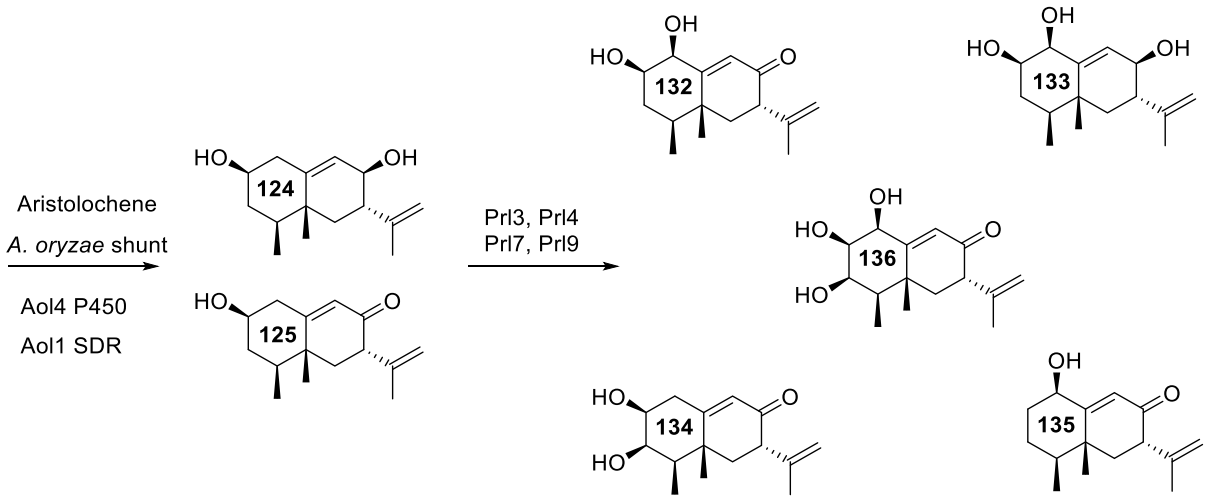
Compound 136				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	79.2	4.32, 1H, dd (3.4, 1.5)	2, 3	2, 3, 5, 9
2	72.1	3.6, 1H, dd (3.4, 3.4)	1, 3	1
3	77.8	3.88, 1H, ddd (3.0, 3.0, 1.5)	2, 4, 1	1, 2, 4, 5, 14
4	45.9	1.56, 1H, m	3, 14	5, 6, 14, 15
5	38.9			
6	45.6	1.9, 1H, m;	6, 7	4, 5, 7, 15
		2.02, 1H, dd (12.8, 4.5)	6, 7	5, 7, 10, 15
7	52.1	3.35, 1H, m	6	6, 8, 11, 12, 13
8	202.0			
9	128.3	5.87, 1H, s		1, 5, 7
10	168.8			
11	144.8			
12	114.8	4.82, 1H, m;	12, 13	7, 13
		4.93, 1H, m	12, 13	7, 13
13	20.2	1.69, 3H, m	12	7, 11, 12
14	12.1	1.19, 3H, d (7.1)	4	3, 4, 5
15	21.4	1.53, 3H, s		4, 5, 10

The stereochemical arrangement of H-1, H-2, and H-3 in **136** was determined by analyzing coupling constants among crucial protons situated near the stereocenters. We deduced that H-1 adopts an equatorial orientation (H-1_{eq}), H-2 is positioned axially (H-2_{ax}), and H-3 also takes on an equatorial position (H-3_{eq}). The coupling constants between H-1_{ax} and H-2_{ax} is ^{3}J 3.4 Hz, between H-2_{ax} and H-3_{eq} is ^{3}J 3.4 Hz, and between H-3_{eq} and H-4_{ax} is ^{3}J 3.0 Hz. The stereochemical configuration of H-4 was confidently established as axial, guided by the aritolochene skeleton. Notably, H-1_{eq} displayed a weaker coupling constant of ^{4}J 1.5 Hz with H-3_{eq}. These coupling constants collectively offer a comprehensive insight into the stereochemistry of **136**, where H-1 is 1*S*, H-2 is 2*R*, and H-3 is 3*R* (Figure 3.3.54).

Furthermore, the NOESY correlations played an important role in corroborating the stereochemical configuration of H-1, H-2, and H-3. These correlations were evident between H-1_{eq} and H-2_{ax}, while H-2_{ax} displayed correlations with H-3_{eq} and H-4_{ax}. These findings collectively indicated an arrangement where H-4, H-3, H-2, and H-1 were all positioned on the same face, specifically on the underside of the backbone (Figure 3.3.54).

3.3.13 Exp 10-12_Co-expression of *hrtc* and *prl3, prl4, prl7*

From the results of Exp 9, it became evident that when *prl3*, *prl4*, *prl7*, and *prl9* were co-expressed alongside the aristolochene synthase Hrtc, a total of five new compounds, namely **132, 133, 134, 135,** and **136**, were synthesized. The structures and absolute configurations of these compounds were elucidated (Scheme 3.3.1).



Scheme 3.3.1 Structures of all the products from Exp9.

Among these compounds, **132**, **135** and **136** likely require an enzyme to oxidize the C-1 position, possibly originating from the precursor **125**. On the other hand, **133** may depend on a precursor from **124**. Additionally, both **134** and **136** seem to require enzymatic oxidation at the C-3 position, with their precursors likely stemming from **125**. To investigate the individual functions of the enzymes Prl3, Prl4, Prl7, and Prl9, we designed three separate experiments named Exp 10, Exp 11, and Exp 12.

Prl3 was introduced into the *hrtc* transformant, which is capable of producing (+)-aristolochene, **124**, and **125**. This led to the creation of Exp10. After completing the fermentation and crude extract preparation, we subjected the products to HPLC analysis. The ELSD chromatogram was employed to identify peaks, which were then compared to those observed in the untransformed strain (Figure 3.3.55). Interestingly, we noticed two peaks that shared identical retention times, mass, and UV characteristics with **124** and **125**. This outcome suggests that Prl3 is not responsible for these reactions for forming those compounds in Exp9.

Following a process similar to Exp10, we introduced *prl4* into the *hrtc* transformant, and this led to the emergence of three peaks in the ELSD chromatogram when compared to the untransformed strain. To gain further insights, we examined the retention time, mass, and UV characteristics of these three peaks and compared them to the data from Exp9. It became evident that **132, 133,** and **135** aligned with these peaks.

Notably, all three compounds, **132, 133,** and **135**, share a common feature, they undergo oxidation at the C-1 position. However, this oxidation occurs at different stages in their respective reactions. Specifically, **132** is generated from the substrate **125** through its reaction with the enzyme Prl4, **133** forms from the substrate **124** in the presence of Prl4, and **135** arises from the C-8-oxidized aristolochene substrate when it reacts with Prl4. Based on the structures of these three compounds, we can conclude that Prl4 functions as a P450 enzyme capable of oxidizing the C-1 position of aristolochene.

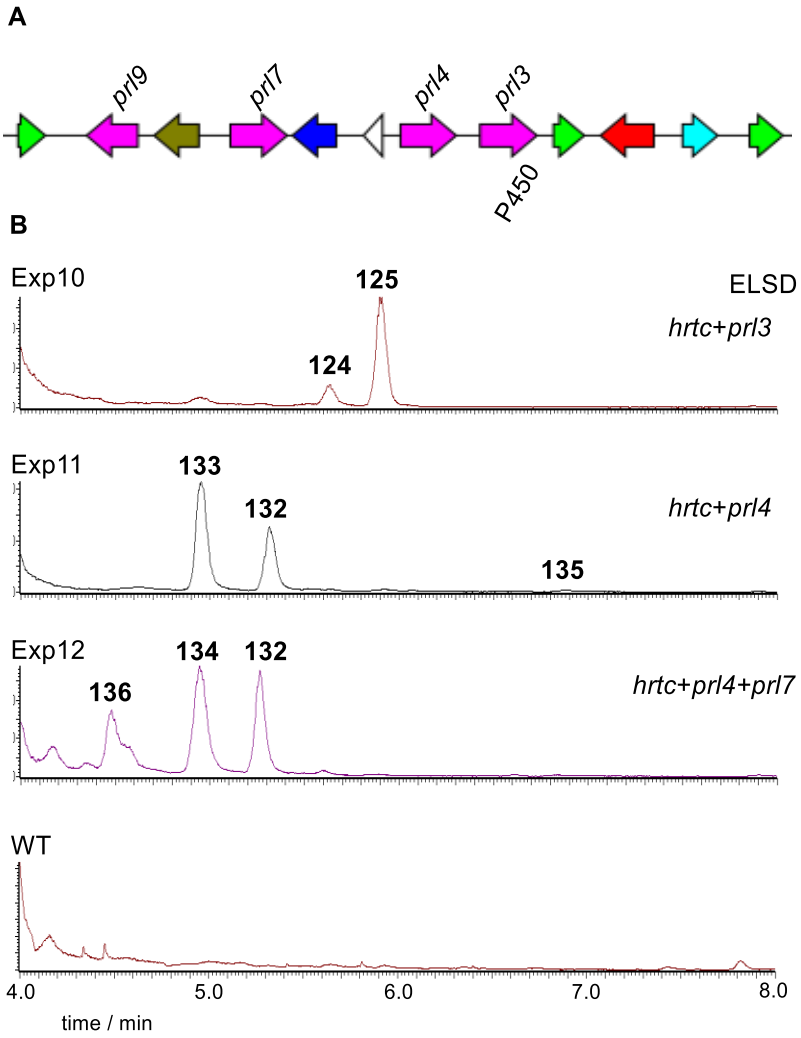


Figure 3.3.55 (**A**) BGC of PR-toxin (**B**) ELSD chromatograms of crude extracts from *A. oryzae* transformants of Exp10-Exp12 and untransformed strain by HPLC analysis.

Continuing our investigations, we conducted Exp12, in which *prl7* was introduced and co-expressed alongside *prl4* within the *hrtc*-transformant. In the ELSD chromatogram, we observed the emergence of three novel peaks when compared to the untransformed strain. To deepen our understanding, we scrutinized these peaks, considering their retention time, mass, and UV characteristics, and then compared them to the data gathered in Exp9. It became clear that **132**, **134**, and **136** were in alignment with these peaks.

Significantly, both **134** and **136** share a common skeleton, they undergo oxidation at C-3. More specifically, **134** is the result of substrate **125** reacting with Prl7, while **136** forms when substrate **132** engages with Prl7. Consequently, we concluded that Prl7 functions as a P450 enzyme with the capability to oxidize the C-3 position on the aristolochene backbone.

From experiments 9 through 12, we established that two active P450 enzymes, namely Prl4 and Prl7, can catalyze oxidation at C-1 and C-3, respectively, through a single process when utilizing an *A. oryzae*

host expressing aristolochene synthase. This approach is both rapid and convenient, allowing us to efficiently generate five new compounds using just three genes.

3.3.14 Exp 13_Co-expression of *hrtc* and *xhr1*

The genome of *Xylaria hypoxylon*, which produces eremoxylarin D **123**, was sequenced using a combination of Illumina and Oxford Nanopore technologies. A candidate gene cluster, believed to be responsible for the biosynthesis of eremoxylarin D **123** (Figure 3.3.56), was proposed. This cluster contains the aristolochene synthase gene and a hrPKS gene. Upon comparing the protein sequences of various BGCs, we observed that Xhl3 shares similarity with Prl7, Hrl3, and AoL4. Xhl4 shows similarity to Prl4 but lacks resemblance to genes found in the hypoxylan-A BGC and sporogen BGC. Additionally, Xhl7 exhibits similarity to Prl9, Hrl7, and Aol4. However, Xhr1 is dissimilar to the other three BGCs. Consequently, we cloned the gene *xhr1* and expressed it, as it showed distinct characteristics from the other identified BGCs.

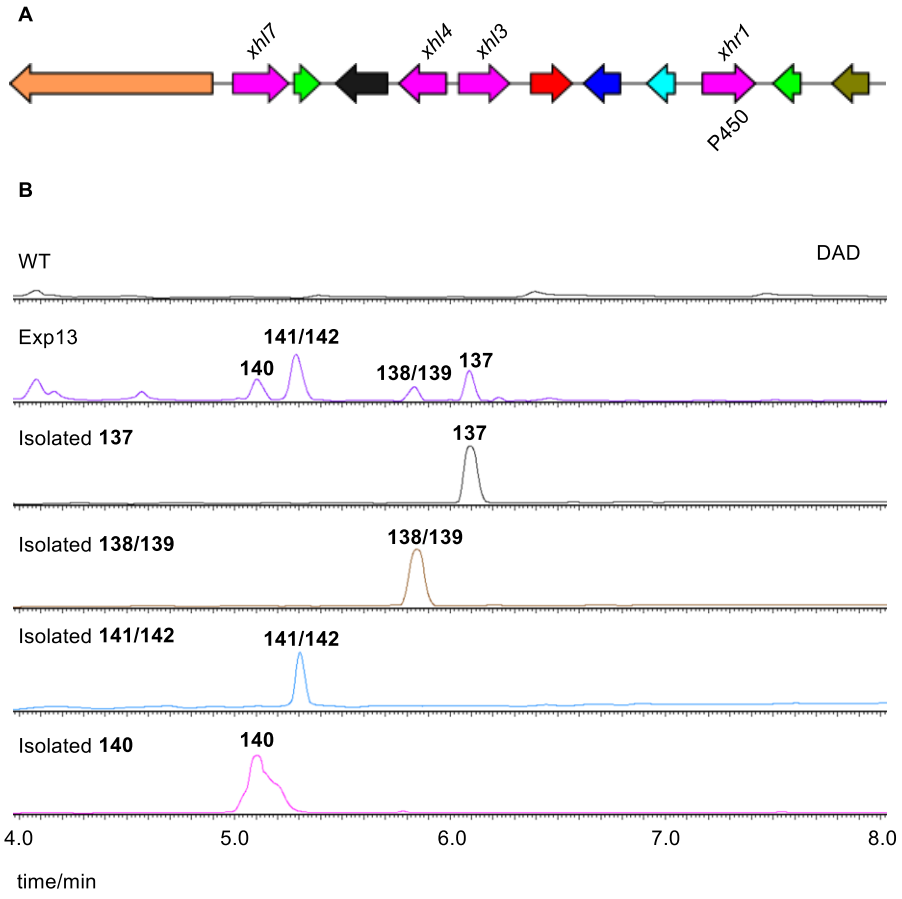


Figure 3.3.56 (**A**) BGC of eremoxylarin D **123** (**B**) DAD chromatograms of crude extracts from *A. oryzae* transformant of WT, Exp13 and isolated compounds by HPLC analysis.

The gene *xhr1* was incorporated into the *hrtc* transformant using the pTYGS-*ade* vector under the control of P_{gpdA}. This transformant possesses the capacity to synthesize (+)-aristolochene **66**, **124**, and **125**. Once the fermentation and crude extract preparation procedures were completed, we performed

HPLC analysis on the resulting products. To identify and quantify the compounds present, we employed an ELSD chromatogram, comparing the results to those obtained from the unmodified strain (Figure 3.3.56). Intriguingly, we observed the emergence of several new peaks in the transformed strain. To isolate these newly discovered compounds, we carried out preparative HPLC following a large-scale fermentation conducted in a 2-liter volume using DPY medium. For a comprehensive understanding of the structures of **137**-**142**, we utilized a combination of 1D and 2D NMR spectroscopy techniques in conjunction with HRMS.

3.3.14.1 Characterization of 137

137 displayed distinct mass fragmentation patterns upon ionization in the positive ion mode (ES^+), showing prominent peaks at *m/z* 221, 203, and 261. In contrast, the negative ion mode produced feeble signals that were indistinguishable (Figure 3.3.57). The appearance of the *m/z* 221 can be ascribed to the elimination of a water molecule $[M + H - H_2O]^+$, while the 261 and 203 peaks can be attributed to the formation of $[M + Na]^+$ and $[M + H - 2H_2O]^+$ ions, respectively. These findings offer partial confirmation of the expected mass of 238 (Figure 3.3.57).

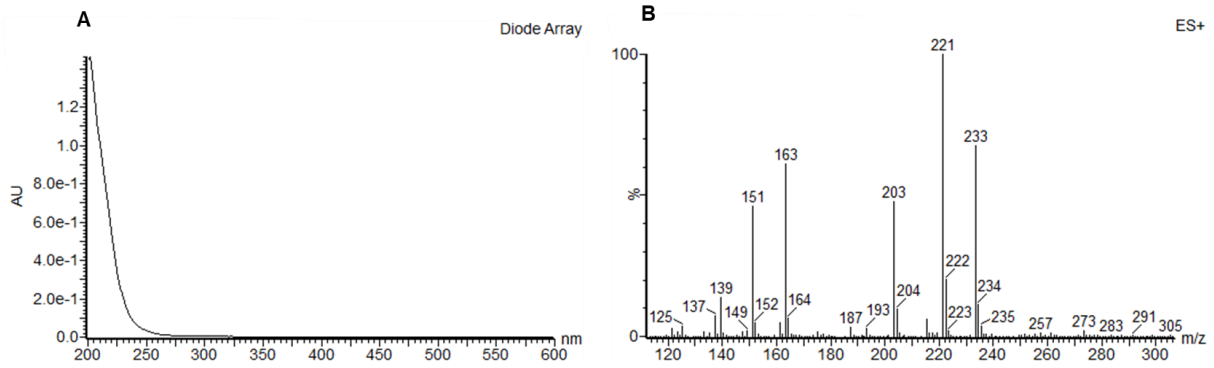
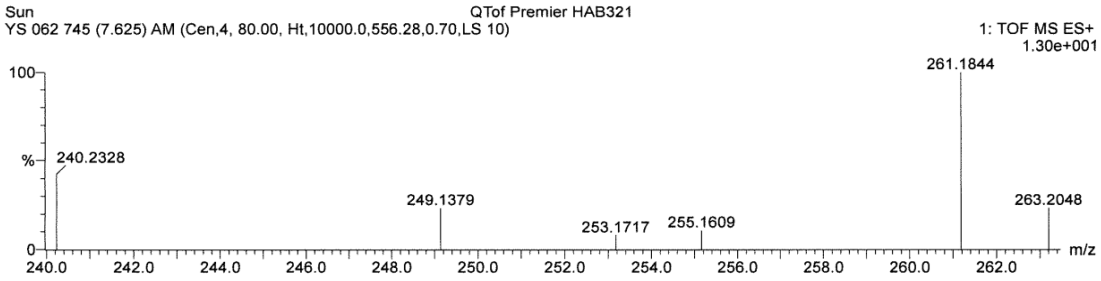


Figure 3.3.57 UV-absorption **(A)** and fragmentation pattern **(B)** of **137** in ES^+ by LR-LCMS.

Utilizing HRMS analysis, we confirmed that **137** possesses a molecular formula of $C_{15}H_{26}O_2$. Theoretical calculations for the $[M + Na]^+$ ion pointed to an expected *m/z* value of 261.1831, a value in close agreement with the observed mass of 261.1844, (Figure 3.3.58).



Minimum:				-1.5					
Maximum:		5.0	20.0	50.0					
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula		
261.1844	261.1855	-1.1	-4.2	5.5	10.1	0.6	C17	H25	O2
	261.1831	1.3	5.0	2.5	10.2	0.8	C15	H26	O2 Na

Figure 3.3.58 HRMS data for **137**; m/z *m/z* [M + Na] $^{+}$ calc. mass is 261.1855, 261.1831 was found.

The analysis ^{1}H-NMR spectrum of **137** disclosed the presence of methyl protons exhibiting doublet splitting patterns at both C-12 (δ_H 0.93, d, 6.9) and C-14 (δ_H 0.90, d, 6.7), along with the oxygenated C-13 (δ_H 3.51, dd *J* 10.5, 6.7 Hz; δ_H 3.64, dd *J* 10.5, 5.7 Hz). Furthermore, another proton was detected on the oxygenated methines at δ_H 4.21 (dd, *J* 3.0, 3.0 Hz) at C-1. Other proton signals closely resembled those of **135**, with the exception of the absence of the C-8 keto group. This includes four pairs of methylene protons at C-2, C-3, C-6, and C-8. Additionally, there was an olefinic proton at C-9 (δ_H 5.60, dd, *J* 5.3, 2.1 Hz), along with two a methyl signals at C-15 (δ_H 1.13, s), as well as two methines at C-4 (δ_H 1.31, m) and C-7 (δ_H 1.80, m, Table 3.3.23).

Through ^{13}C-NMR and HSQC analyses, we precisely identified two oxygenated carbons at δ_C 74.9 and 66.6, alongside olefinic carbons (δ_C 125.9, 145.3), methyl groups (δ_C 13.1, 15.8, 21.0), four methylene groups (δ_C 33.7, 25.6, 43.1, 28.4), two methine groups (δ_C 43.9, 31.3), and a quaternary carbon (δ_C 38.0). The signal corresponding to the carbonyl carbon was conspicuously absent. These findings unequivocally confirm the presence of two hydroxyl groups incorporated into **137**'s backbone, thereby validating its structural composition.

We inferred that H-1 adopts an equatorial orientation (H-1_{eq}), as evident from the coupling constants between H-1_{eq} and H-2_{ax} (^{3}J 3.0 Hz) and between H-1_{eq} and H-2_{eq} (^{3}J 3.0 Hz). The stereochemical configuration of H-1 was confidently established as equatorial, with additional guidance from the structure of X76 (Table 3.3.23). Furthermore, NOESY correlations clearly demonstrated interactions between H-1_{eq} and H-2_{ax}, as well as between H-1_{eq} and H-2_{eq} and H-9. These correlations indicate that H-1_{eq} possesses spatial angles of less than 60° with H-2_{ax} and H-2_{eq} (Figure 3.3.59).

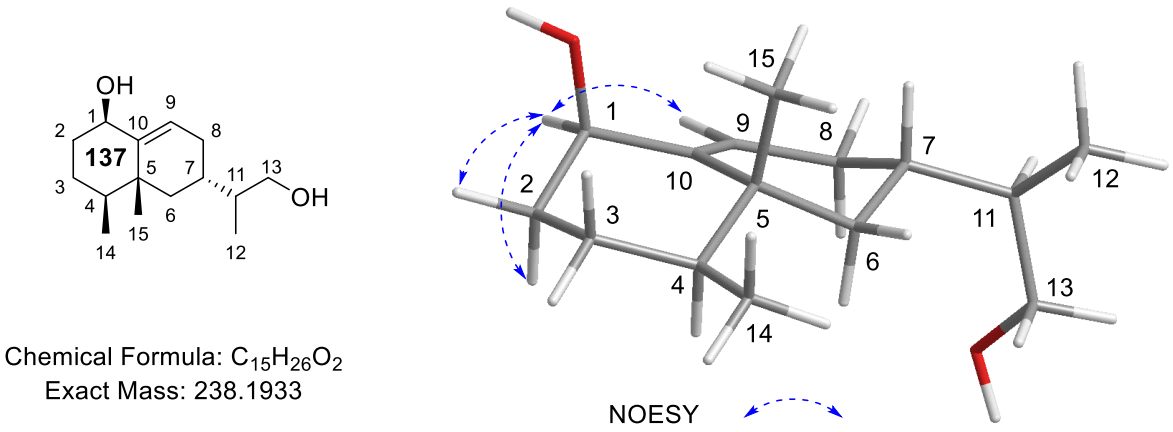


Figure 3.3.59 Structure and a stick model of **137** for the NOESY.

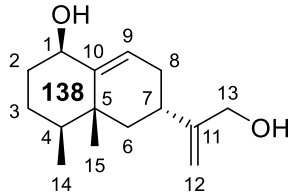
Table 3.3.23 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **137** recorded in $CDCl_3$.

Compound 137				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	74.9	4.21, 1H, dd (3.0, 3.0)	2	3, 9
2	33.7	1.56, 1H, m;	2, 3, 4	3, 4
		1.89, 1H, dddd (13.8, 2.8, 2.8, 2.7)	2, 3, 1	1, 2, 3
3	25.6	1.31, 1H, m;	2, 3	1, 2, 4, 5
		1.79, 1H, m	2, 3, 4	2, 4
4	43.9	1.31, 1H, m	3, 14	5, 15
5	38.0			
6	43.1	1.01, 1H, dd (12.2, 12.2);	6, 7	4, 5, 7, 8, 11, 15
		1.66, 1H, m	6	5, 7, 8, 10, 11, 15
7	31.3	1.8, 1H, m	6, 8, 11	6, 8, 11
8	28.4	1.76, 1H, m;	8, 7, 9	6, 7, 9, 10
		2.02, 1H, m	8, 7, 9	6, 7, 9, 10
9	125.9	5.6, 1H, dd (5.3, 2.1)	8	1, 5, 7, 8
10	145.3			
11	40.5	1.59, 1H, m	7, 12, 13	6, 7, 8, 12, 13
12	13.1	0.93, 3H, d (6.9)	11	7, 11, 13
13	66.6	3.51, 1H, dd (10.5, 6.7);	13, 11	7, 11, 12
		3.64, 1H, dd (10.5, 5.7)	13, 11	7, 11, 12
14	15.8	0.9, 3H, d (6.7)	4	3, 4, 5
15	21.0	1.13, 3H, s		4, 5, 6, 10

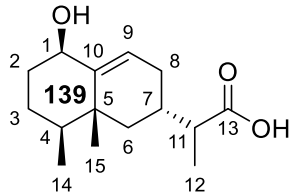
3.3.14.2 Characterization of 138 and 139

The NMR analysis of compounds **138** and **139** revealed a complex mixture, with **138** featuring a hydroxyl group at C-13. Starting with the H-13 protons (δ_H, 4.15, 2H, m) of **138** in the HMBC spectrum, we were able to deduce the corresponding ^{13}C chemical shifts for C-11 (δ_C 153.8), C-12 (δ_C 108.3), and C-7 (δ_C 33.5). Subsequently, the proton shifts of H-12 (δ_H 4.92, 1H, m; δ_H 5.08, 1H, m) and H-7 (δ_H 2.40, 1H, m) were determined using HSQC. Utilizing the information obtained from the ^{1}H, ^{1}H-COSY spectrum of H-7, we identified the proton shifts of H-8 (δ_H 1.90, 1H, m) and H-6 (δ_H 1.21, 1H, m; δ_H 1.77, 1H, m). This, in turn, led to the discovery of the ^{13}C shifts for C-8 (δ_C 32.1) and C-6 (δ_C 43.9) through HSQC. We further elucidated the proton shift of H-9 (δ_H 5.63) by referencing the proton shift of H-8 in the ^{1}H, ^{1}H-COSY spectrum. Lastly, we determined the proton shift of H-1 (δ_H 4.2, 1H, dd, *J*

3.0, 3.0 Hz) by analyzing the HMBC spectrum of H-9. Proton shifts for H-2 (δ_H 1.55, 1H, m; δ_H 1.90, 1H, m), H-3 (δ_H 1.29, 1H, m; δ_H 1.77, 1H, m), and H-4 (δ_H 1.29, 1H, m) were subsequently determined, along with their corresponding ^{13}C shifts as revealed by HSQC (δ_C-2 33.7, δ_C-3 25.6, δ_C-4 43.8).



Chemical Formula: $C_{15}H_{24}O_2$
Exact Mass: 236.1776



Chemical Formula: $C_{15}H_{24}O_3$
Exact Mass: 252.1725

Figure 3.3.60 Structures and of **138** and **139.**

139 exhibits a keto group at the C-13 position. Similar to **138**, **139** shares the same decalin backbone, which includes the hydroxyl group at C-1 (δ_H 4.21, 1H, dd, J 3.0, 3.0 Hz), as well as the two methyl groups, H-14 (δ_H 0.89, d, J 6.9) and H-15 (δ_H 1.15, s), and the olefinic proton at C-9 (δ_H 5.57, dd, J 5.3, 2.2 Hz). However, there are notable differences in the substituents at the C-11 position for **139**. Specifically, it features a methyl group, H-12 (δ_H 1.21, d, J 7.0), and a carbonyl group (δ_C 180.2).

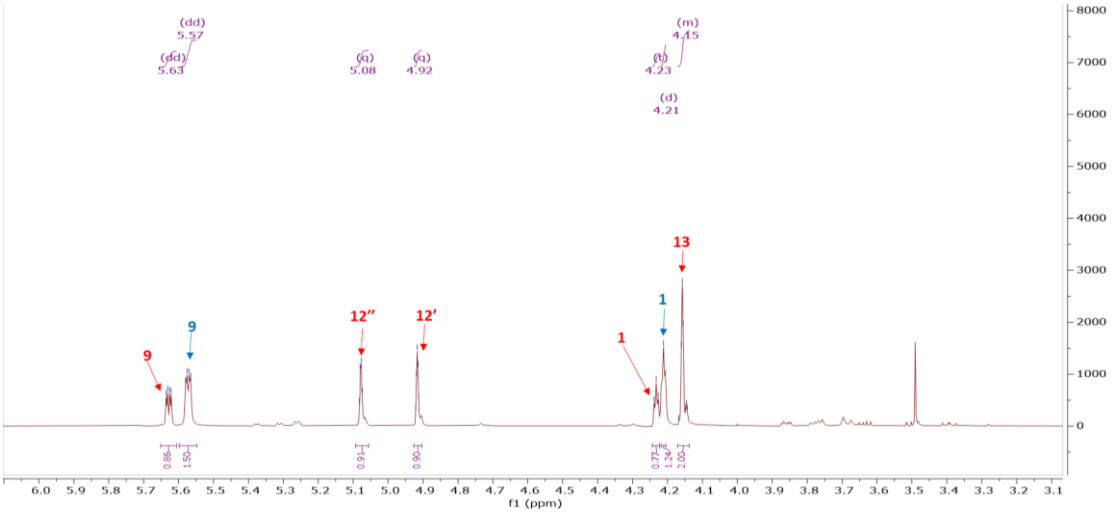
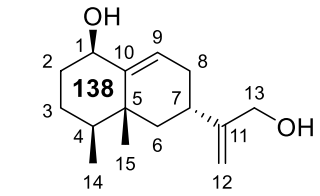


Figure 3.3.61 A part of ^{1}H-NMR spectrum (3.1 ppm – 6.0 ppm) of **138** and **139** mixtures recorded at 500 MHz in $CDCl_3$. The red numbers and arows represent the proton positions of **138**; the blue numbers and arrows represent the proton positions of **139**. The shift labels are on the top of the integral curves; the types of the peaks are in the brackets.

The stereochemical configuration of the hydroxyl group at C-1 was determined by analyzing their coupling constants. For compounds **138**, the configuration of H-1 was found to be H-1_{eq}, as evidenced by the coupling constant between H-1_{eq} and H-2_{ax} (^3J 3.0 Hz), and the coupling constant between H-1_{eq} and H-2_{eq}, also at J 3.0 Hz, which closely resembled that of **137**. Similarly, for compounds **139**, the configuration of H-1 was also determined to be H-1_{eq}. This conclusion was drawn based on the identical

coupling constants between H-1_{eq} and H-2_{ax} (J 3.0 Hz) and between H-1_{eq} and H-2_{eq} (J 3.0 Hz), mirroring the findings in **137**. Therefore, it was deduced that both of these compounds exhibit *R*-configuration at C-1, indicating that the hydrogen atom is situated on the lower side of the molecule. To validate this deduction, information obtained from NOESY experiments was employed (Figure 3.62, 3.63).



Chemical Formula: $C_{15}H_{24}O_2$
Exact Mass: 236.1776

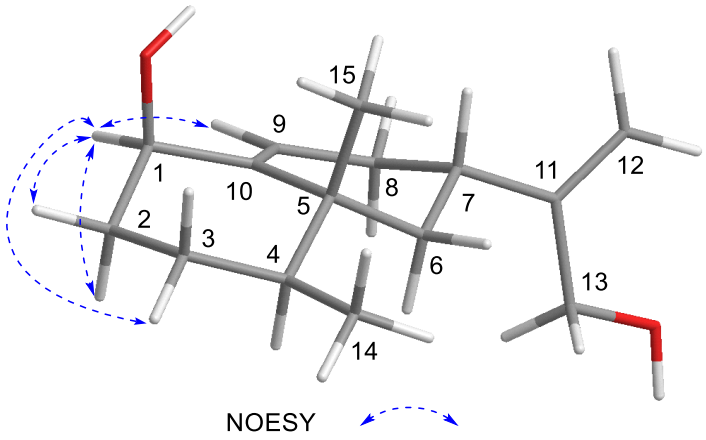


Figure 3.3.62 Structure and a stick model of **138** for the NOESY.

Table 3.3.24 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **138** recorded in $CDCl_3$.

Compound 138				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	74.9	4.23, 1H, dd (3.0, 3.0)	2	3, 5, 9
2	33.7	1.55, 1H, m;	2, 1, 3	3, 4
		1.90, 1H, m	2, 1, 3	3, 4, 10
3	25.6	1.29, 1H, m;	3, 2	1, 4, 5
		1.77, 1H, m	3, 2, 4	1, 4, 5
4	43.8	1.29, 1H, m	3, 14	6, 15
5	38.3			
6	43.9	1.21, 1H, m;	6, 7	4, 5, 15
		1.77, 1H, m	6, 7	4, 5, 15
7	33.5	2.40, 1H, m	6, 8	11, 12, 13
8	32.1	1.90, 1H, m;	8, 7, 9	6, 7, 9, 10
		2.19, 1H, m	8, 7, 9	6, 7, 9, 10
9	125.5	5.63, 1H, dd (5.2, 2.2)	8	1, 5, 7
10	145.2			
11	153.8			
12	108.3	4.92, 1H, m;	12, 13	7, 11, 13
		5.08, 1H, m	12, 13	7, 11, 13
13	65.4	4.15, 2H, m	12	7, 11, 12
14	15.8	0.89, 3H, d (6.9)	4	3, 4, 5
15	21.0	1.15, 3H, s		4, 5, 10

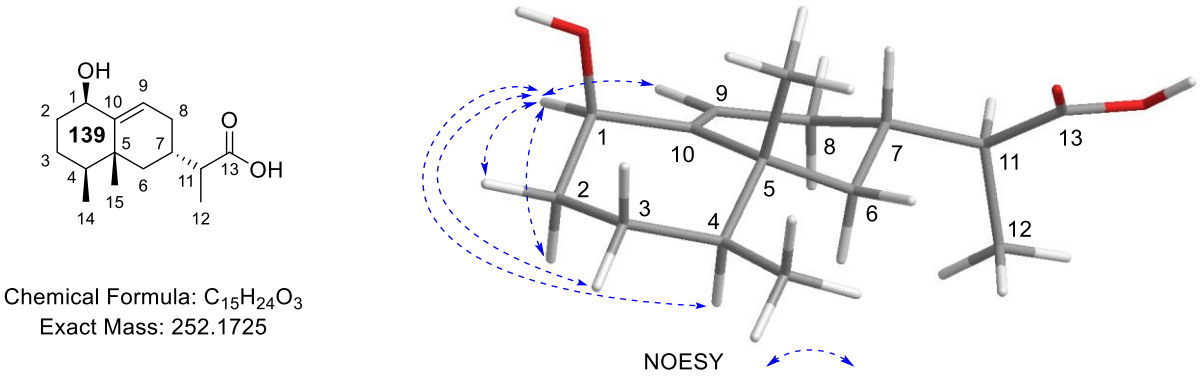


Figure 3.3.63 Structure and a stick model of **139** for the NOESY.

Table 3.3.25 Summarized NMR signals (^{1}H-500 MHz, ^{13}C-125 MHz) for **139** recorded in $CDCl_3$.

Compound 139				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	74.8	4.21, 1H, dd (3.0, 3.0)	2	3, 5, 9
2	33.8	1.55, 1H, m;	2, 1, 3	3, 4
		1.90, 1H, m	2, 1, 3	1, 3, 4, 10
3	25.5	1.29, 1H, m;	3, 2	1, 4, 5
		1.77, 1H, m	3, 2, 4	1, 4, 5
4	43.8	1.29, 1H, m	3, 14	2, 5, 6
5	38.0			
6	41.8	0.96, 1H, m;	6, 7	4, 5, 8, 10, 15
		1.77, 1H, m	6, 7	4, 5, 8, 10, 15
7	33.4	2.01, 1H, m	6, 8, 11	6, 11, 12
8	30.3	1.83, 1H, m;	8, 7, 9	7, 9, 10, 11
		2.11, 1H, m	8, 7, 9	6, 7, 9, 10
9	125.1	5.57, 1H, dd (5.3, 2.2)	8	1, 5, 7, 8
10	145.3			
11	44.3	2.32, 1H, dddd (6.9, 6.9, 6.9, 6.9)	7, 12	6, 7, 8, 12, 13
12	14.2	1.21, 3H, d (7.0)	11	7, 11, 13
13	180.2			
14	15.8	0.89, 3H, d (6.6)	4	3, 4, 5
15	21.0	1.15, 3H, s		4, 5, 10

The structural elucidation of **138** was supported by both LCMS and HRMS analyses. In LCMS, **138** exhibited distinctive mass fragmentation patterns when ionized in the positive ion mode (ES^+), resulting in peaks at *m/z* 219 and 201 (Figure 3.3.64B). Conversely, the negative ion mode yielded weak signals that were unidentifiable. The appearance of the *m/z* 219 fragment can be attributed to the elimination of a water molecule $[M + H - H_2O]^+$, while the 201 peak can be ascribed to the formation of $[M + H - 2H_2O]^+$ ions. These observations provided partial confirmation of the expected mass of 236. Furthermore, through HRMS analysis, we definitively established that **138** possesses a molecular formula of $C_{15}H_{24}O_2$. Theoretical calculations for the $[M + H - H_2O]^+$ ion indicated an expected *m/z* value of 219.1749, which closely matched the observed mass of 219.1754 (Figure 3.3.65).

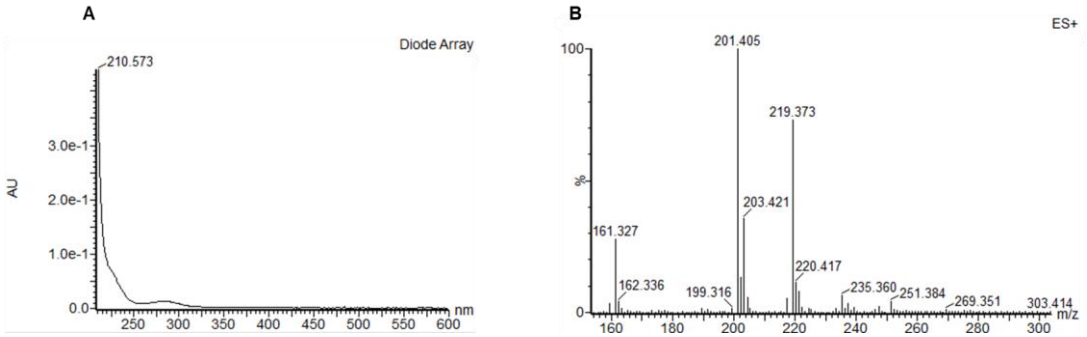
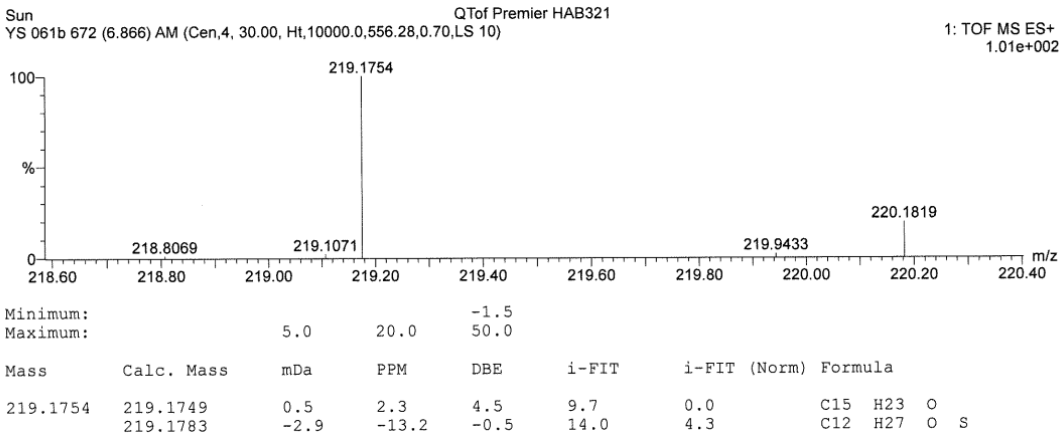


Figure 3.3.64 UV-absorption **(A)** and fragmentation pattern **(B)** of **138** in ES^+ by LR-LCMS.



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
219.1754	219.1749	0.5	2.3	4.5	9.7	0.0	C15 H23 O
	219.1783	-2.9	-13.2	-0.5	14.0	4.3	C12 H27 O S

Figure 3.3.65 HRMS data for **138**; *m/z* $[M + H - H_2O]^+$ calc. mass is 219.1749, 219.1754 was found.

The structural characterization of **139** received strong support from both LCMS and HRMS analyses. In LCMS, **139** displayed distinct mass fragmentation patterns upon positive ion mode (ES^+) ionization, resulting in peaks at *m/z* 235 and 275. Conversely, in the negative ion mode (ES^-), it yielded peaks at *m/z* 251 and 503 (Figure 3.3.66B). The presence of the *m/z* 235 fragment can be attributed to the removal of a water molecule $[M + H - H_2O]^+$, while the 275 peak is indicative of the formation of $[M + Na]^+$ ions. The signals at 251 and 503 correspond to $[M - H]^-$ and $[2M - H]^-$ ions, respectively. These findings provided partial confirmation of the expected mass of 252 (Figure 3.3.66B). Additionally, employing HRMS analysis, we conclusively established that **139** possesses a molecular formula of $C_{15}H_{24}O_3$. Theoretical calculations for the $[M - H]^-$ ion pointed to an expected *m/z* value of 251.1647, which closely matched the observed mass of 251.1644 (Figure 3.3.67).

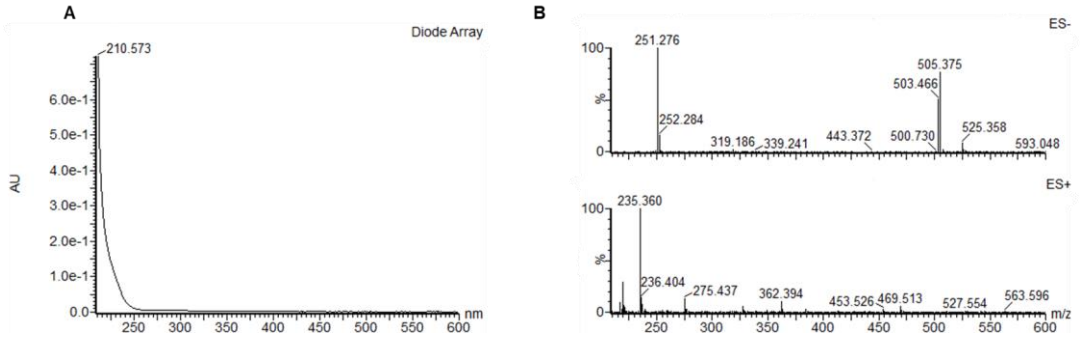
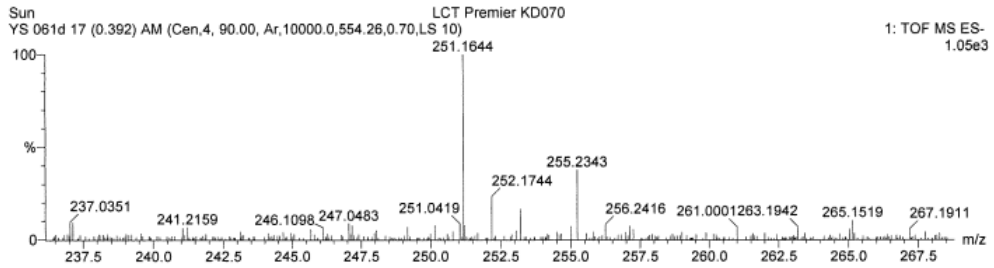


Figure 3.3.66 UV-absorption **(A)** and fragmentation pattern **(B)** of **139** in ES^+ (bottom) and ES^- (top) by LR-LCMS.



Minimum:				-1.5		
Maximum:		5.0	5.0	50.0		
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
251.1644	251.1647	-0.3	-1.2	4.5	77.0	C15 H23 O3

Figure 3.3.67 HRMS data for **139**; *m/z* $[M - H]^-$ calc. mass is 251.1647, 251.1644 was found.

3.3.14.3 Characterization of 140

140 exhibited distinct mass fragmentation patterns upon ionization in the positive ion mode (ES^+), with prominent peaks observed at *m/z* 219 and 201 (Figure 3.3.68B). The appearance of the *m/z* 219 fragment can be attributed to the removal of a water molecule $[M + H - H_2O]^+$, while the presence of the 201 peaks can be ascribed to the formation of $[M + H - 2H_2O]^+$ ions. Through HR-GCMS analysis, we definitively determined that **140** possesses a molecular formula of $C_{15}H_{24}O_2$. Theoretical calculations for the [M] ion indicated an expected *m/z* value of 236.1776, which closely matched the observed mass of 236.1778 (Figure 3.3.69).

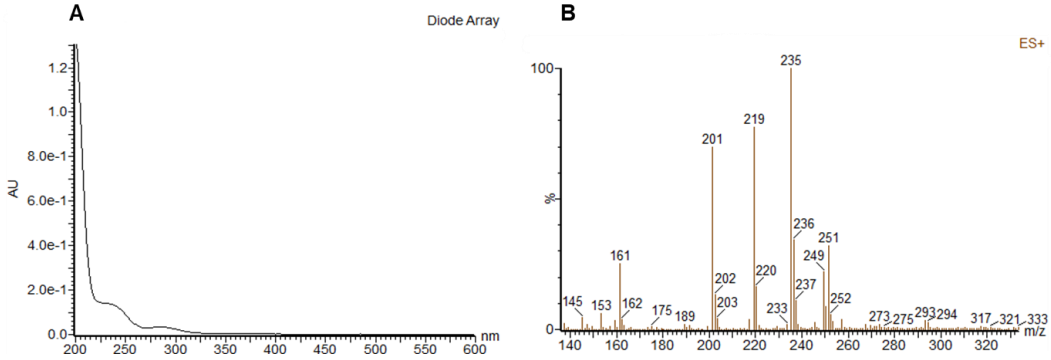
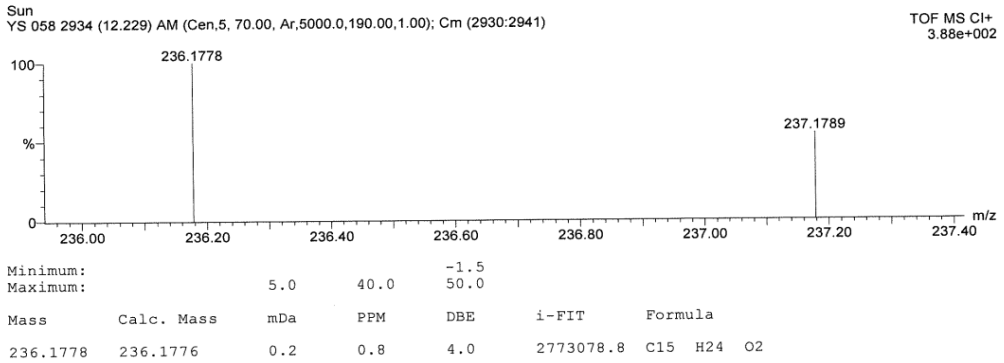


Figure 3.3.68 UV-absorption **(A)** and fragmentation pattern **(B)** of **140** in ES^+ by LR-LCMS.



Minimum:				-1.5		
Maximum:		5.0	40.0	50.0		
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
236.1778	236.1776	0.2	0.8	4.0	2773078.8	C15 H24 O2

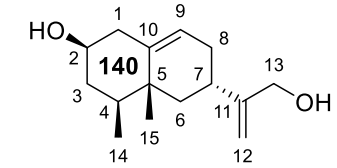
Figure 3.3.69 HRMS data for **140**; *m/z* [M - H] $^{+}$ calc. mass is 236.1776, 236.1778 was found by HRGC-MS.

The ^{1}H NMR spectrum of **140** revealed some distinct signals, including an olefinic proton at H-9 (δ_H 5.40, m), two vinyl protons at H-12 (δ_H 4.91, m; δ_H 5.07, m), two methyl signals at H-14 (δ_H 0.88, d, J 6.6 Hz) and H-15 (δ_H 0.98, s), as well as an oxidated methane proton at H-2 (δ_H 3.57, dddd, J 11.0, 11.0, 4.8, 4.8 Hz). Additionally, the spectrum displayed four sets of methylene protons at H-1 (δ_H 2.11, m; δ_H 2.34, m), H-3 (δ_H 1.37, m; δ_H 1.73, m), H-6 (δ_H 1.16, m; δ_H 1.82, ddd, J 12.7, 2.3, 2.3 Hz), and H-8 (δ_H 1.90, dddd, J 17.2, 11.7, 3.7, 2.3 Hz; δ_H 2.11, m, Table 3.3.26).

The ^{13}C NMR data (Table 3.3.26) revealed some important carbon signals, including two methyl groups at δ_C 15.5 and 18.2, four sp3 methylene groups at δ_C 42.1, 40.3, 43.3, and 32.1, an oxymethine signal at δ_C 71.1, two methane signals at δ_C 41.4 and 33.9, one sp^3 quaternary carbon at δ_C 38.1, and four olefinic carbons at δ_C 121.6, 141.5, 153.8, and 108.2. Additionally, the oxidized C-13 carbon was identified at δ_C 65.4.

The stereochemical configuration of H-2 was established by analyzing the coupling constants between key protons located around the stereocenters and by examining NOESY correlations. Specifically, H-2 was identified as being in the axial position (H-2$_{ax}$), supported by proton coupling constants H-2$_{ax}$/H-1$_{ax}$ (^3J 11.0), H-2$_{ax}$/H-3$_{ax}$ (^3J 11.0), H-2$_{ax}$/H-1$_{eq}$ (^3J 4.8), and H-2$_{ax}$/H-3$_{eq}$ (^3J 4.8).

Furthermore, NOESY correlations between H-2$_{ax}$ and H-4$_{ax}$ indicated that both H-2$_{ax}$ and H-4$_{ax}$ occupied the same face of the (+)-aristolochene backbone (Figure 3.3.70). Additionally, the correlation observed between H-2$_{ax}$ and H-3$_{eq}$ provided further confirmation of the axial position of H-2 (Figure 3.3.70). As a result, the configuration at C-2 was determined to be 2*R*.



Chemical Formula: $C_{15}H_{24}O_2$
Exact Mass: 236.1776

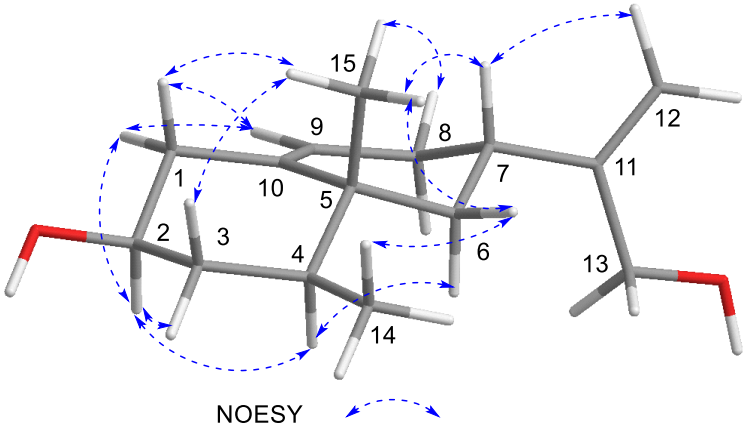


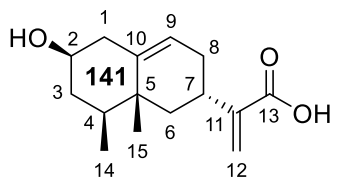
Figure 3.3.70 Structure and a stick model of **140** for the NOESY.

Table 3.3.26 Summarized NMR signals (^{1}H-600 MHz, ^{13}C-150 MHz) for **140** recorded in $CDCl_3$.

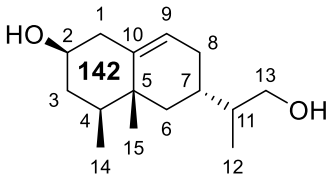
Compound 140				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	42.1	2.11, 1H, m;	1, 2	2, 3, 5, 9, 10
		2.34, 1H, m	1, 2, 3	2, 3, 5, 9, 10
2	71.1	3.57, 1H, dddd (11.0, 11.0, 4.8, 4.8)	1, 3	
3	40.3	1.37, 1H, m	3, 2, 4	2, 4
		1.73, 1H, dddd (11.7, 4.4, 2.4, 2.4)	3, 2, 4, 1	1, 2, 4, 5
4	41.4	1.33, 1H, m	3, 14	
5	38			
6	43.3	1.16, 1H, m;	6, 7	4, 5, 7, 8, 15
		1.82, 1H, ddd (12.7, 2.3, 2.3)	6, 7, 8	5, 7, 8, 10, 15
7	33.9	2.32, 1H, m	6, 8, 12	6, 8, 11, 12
8	32.1	1.9, 1H, dddd (17.2, 11.7, 3.7, 2.3);	8, 7, 9, 6	7, 9, 10
		2.11, 1H, m	8, 7, 9	9, 10
9	121.6	5.4, 1H, m	8	7
10	141.5			
11	153.8			
12	108.2	4.91, 1H, m;	12, 13	7, 11, 13
		5.07, 1H, m	12, 13	7, 11, 13
13	65.4	4.15, 2H, dd (1.3, 1.3)	12	7, 11, 12
14	15.5	0.88, 3H, d (6.6)	4	3, 4, 5
15	18.2	0.98, 3H, s		4, 5, 6, 10

3.3.14.4 Characterization of 141, 142

141 and **142** form a mixture in which **141** is the major component. The determination of **141**'s structure was aided by the presence of stronger signals observed in both 1D and 2D NMR spectra. Elucidation of the ^{1}H-NMR unveiled a shared decalin backbone between **141** and **142**, marked by oxymethine signals occupying the same positions and displaying signal overlap at H-2 (δ_H 3.45, 1H, m).



Chemical Formula: $C_{15}H_{22}O_3$
Exact Mass: 250.1569



Chemical Formula: $C_{15}H_{26}O_2$
Exact Mass: 238.1933

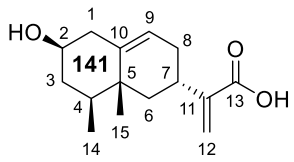
Figure 3.3.71 Structure of **141 and 142.**

Both compounds exhibit an olefinic proton at H-9 (**141**: δ_H 5.42, m; **142**: δ_H 5.38, m). Additionally, they display two methyl signals at C-14 (**141**: δ_H 0.89, 3H, d, *J* 6.4 Hz; **142**: δ_H 0.89, 3H, d, *J* 6.4 Hz) and C-15 (**141**: δ_H 1.03, 3H, s; **142**: δ_H 0.97, 3H, s). Furthermore, both compounds possess four sets of methylene protons at H-1, H-3, H-6, and H-8 (Table 3.3.27 and 3.3.28) and methine groups at H-4 and H-7. The distinguishing features between them are as follows: **141** contains a vinyl group (δ_H 5.55, dd, *J* 1.3, 1.3 Hz; δ_H 6.17, d, *J* 1.2 Hz) and a carboxylic acid group at C-11, while **142** includes a methyl group (δ_H 0.88, 3H, d, *J* 6.4 Hz) and an oxymethyl group (δ_H 3.38, dd, *J* 10.7, 6.9 Hz; δ_H 3.55, dd, *J* 10.7, 5.9 Hz) at C-11(Table 3.3.27 and 3.3.28).

The ^{13}C NMR spectra (Table 3.3.27 and 3.3.28) revealed similar features for both compounds. Specifically, they each displayed two methyl groups (**141**: $\delta_{C\text{-}14}$ 15.8, $\delta_{C\text{-}15}$ 18.3; **142**: $\delta_{C\text{-}14}$ 15.8, $\delta_{C\text{-}15}$ 18.4), four sp^3 methylene groups (**141**: $\delta_{C\text{-}1}$ 42.6, $\delta_{C\text{-}3}$ 40.8, $\delta_{C\text{-}6}$ 44.9, $\delta_{C\text{-}8}$ 32.9; **142**: $\delta_{C\text{-}1}$ 42.7, $\delta_{C\text{-}3}$ 40.9, $\delta_{C\text{-}6}$ 43.6, $\delta_{C\text{-}8}$ 29.0), and three sp^3 methine groups, one of which was an oxymethine group (**141**: $\delta_{C\text{-}2}$ 71.6, $\delta_{C\text{-}4}$ 42.7, $\delta_{C\text{-}7}$ 33.3; **142**: $\delta_{C\text{-}2}$ 71.7, $\delta_{C\text{-}4}$ 42.8, $\delta_{C\text{-}7}$ 32.8). Additionally, each compound featured a single sp^3 quaternary carbon (**141**: $\delta_{C\text{-}5}$ 39.3; **142**: $\delta_{C\text{-}5}$ 38.8) and two olefinic carbons (**141**: $\delta_{C\text{-}9}$ 122.2, $\delta_{C\text{-}10}$ 142.9; **142**: $\delta_{C\text{-}9}$ 122.7, $\delta_{C\text{-}10}$ 142.8). The distinguishing characteristics between them include **141**'s vinyl group ($\delta_{C\text{-}11}$ 147.5, $\delta_{C\text{-}12}$ 123.1) and carboxylic acid group ($\delta_{C\text{-}13}$ 170.7), while **142** possesses a methyl group ($\delta_{C\text{-}12}$ 13.3) and an oxymethyl group ($\delta_{C\text{-}13}$ 66.4) on C-11 ($\delta_{C\text{-}11}$ 41.5, Table 3.3.27 and 3.3.28).

The stereochemical configuration of H-2 in both **141** and **142** was determined by analyzing the coupling constants between key protons situated around the stereocenters. In the case of **141**, H-2 was established to occupy the axial position (H-2$_{ax}$), as supported by the proton coupling constants H-1$_{eq}$/H-2$_{ax}$ (^3J 5.1). NOESY correlations between H-2$_{ax}$ and H-4$_{ax}$ further confirmed that both H-2$_{ax}$ and H-4$_{ax}$ were situated on the same side of the (+)-aristolochene backbone (Figure 3.3.72). Consequently, the relative configuration at C-2 in **141** was determined to be 2*R*.

Similarly, in **142**, H-2 was also identified in the axial position (H-2$_{ax}$), substantiated by proton coupling constants H-1$_{eq}$/H-2$_{ax}$ (^3J 5.5). NOESY correlations between H-2$_{ax}$ and H-4$_{ax}$ likewise indicated that both H-2$_{ax}$ and H-4$_{ax}$ resided on the same face of the (+)-aristolochene backbone (Figure 3.3.73). So, the configuration at C-2 in **142** was determined to be 2*R* as well.



Chemical Formula: $C_{15}H_{22}O_3$
Exact Mass: 250.1569

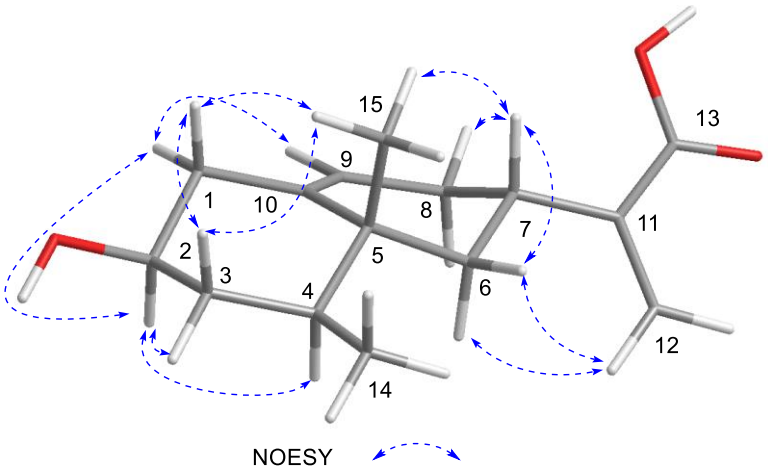


Figure 3.3.72 Structure and a stick model of **141** for the NOESY.

Table 3.3.27 Summarized NMR signals (^{1}H-400 MHz, ^{13}C-100 MHz) for **141** recorded in CD_3OD.

Compound 141				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	42.6	2.10, 1H, m;	1, 2	2, 3, 9, 10
		2.28, 1H, ddd (12.9, 5.1, 2.3)	1, 2, 3	2, 3, 5, 9, 10
2	71.6	3.45, 1H, m	1, 3	1, 3
3	40.8	1.39, 1H, m;	3, 2, 4	1, 2, 4, 10,
		1.70, 1H, m	3, 2, 4, 1	1, 2, 4, 5, 10
4	42.7	1.31, 1H, m	14, 3	3, 5, 6, 14, 15
5	39.3			
6	44.9	1.13, 1H, m;	6, 7	4, 5, 7, 15
		1.84, 1H, m	6, 7	5, 7, 15
7	33.3	2.79, 1H, m	6, 8	5, 6, 8, 11, 12, 13
8	32.9	1.88, 1H, m;	8, 7	7, 9, 10, 11
		2.14, 1H, m	8, 7, 9	7, 9, 10, 11
9	122.2	5.42, 1H, m	8	1, 5, 7
10	142.9			
11	147.5			
12	123.1	5.55, 1H, dd (1.3, 1.3);	12, 7	7, 13
		6.17, 1H, d (1.2)	12	7, 11, 13
13	170.7			
14	15.8	0.89, 3H, d (6.4)	4	3, 4, 5
15	18.3	1.03, 3H, s		4, 5, 6, 10

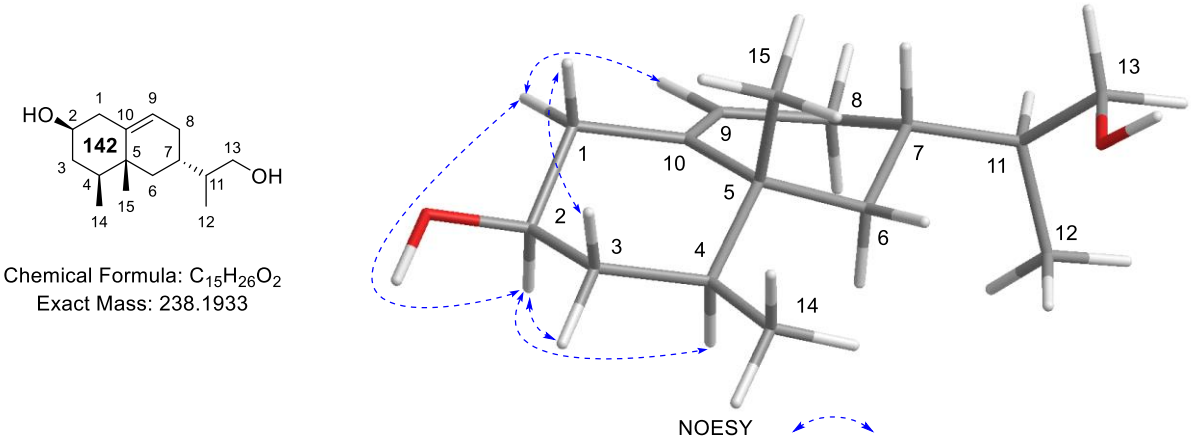


Figure 3.3.73 Structure and a stick model of **142** for the NOESY.

Table 3.3.28 Summarized NMR signals (^{1}H-400 MHz, ^{13}C-100 MHz) for **142** recorded in CD_3OD.

Compound 142				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	42.7	2.12, 1H, m;	1, 2	2, 3, 9, 10
		2.26, 1H, ddd (15.5, 5.5, 2.3)	1, 2, 3	2, 3, 5, 9, 10
2	71.7	3.45, 1H, m	1, 3	1, 3
3	40.9	1.39, 1H, m;	3, 2	2
		1.70, 1H, m	3, 2, 1	2
4	42.8	1.31, 1H, m	14	3, 5, 6, 14, 15
5	38.8			
6	43.6	0.99, 1H, m;	6, 7	5, 7, 15
		1.71, 1H, m	6, 7	5, 7, 15
7	32.8	1.75, 1H, m	6, 11	6, 8, 11
8	29.0	1.71, 1H, m;	8, 9	7, 9, 10, 11
		1.90, 1H, m	8, 9	7, 9, 10, 11
9	122.7	5.38, 1H, m	8	1, 5
10	142.8			
11	41.5	1.51, 1H, m	7, 13	7, 8, 12, 13
12	13.3	0.88, 3H, d (6.4)	11	7, 13
13	66.4	3.38, 1H, dd (10.7, 6.9);	13, 11	7, 11, 12
		3.55, 1H, dd (10.7, 5.9)	13, 11	7, 11, 12
14	15.8	0.89, 3H, d (6.4)	4	3, 4, 5
15	18.4	0.97, 3H, s		5, 6, 10

In LCMS analysis, when subjected to positive ion mode (ES^+) ionization, **141** exhibited distinctive mass fragmentation patterns, producing peaks at *m/z* 233 and 251, as well as 273. Conversely, under negative ion mode (ES^-), **141** yielded peaks at *m/z* 249 (Figure 3.3.74B). The appearance of the 233 fragment can be attributed to the removal of a water molecule $[M + H - H_2O]^+$, while the presence of the 273 peak indicates the formation of $[M + Na]^+$ ions, and 251 corresponds to $[M + H]^+$. The signals at 249 align with $[M - H]^-$ ions. These findings provided partial confirmation of the anticipated mass of 250 (Figure 3.3.74B). Furthermore, through HRMS analysis, we conclusively determined that **141** possesses a molecular formula of $C_{15}H_{22}O_3$. Theoretical calculations for the $[M - H]^-$ ion indicated an expected *m/z* value of 249.1491, which closely matched the observed mass of 249.1493 (Figure 3.3.75).

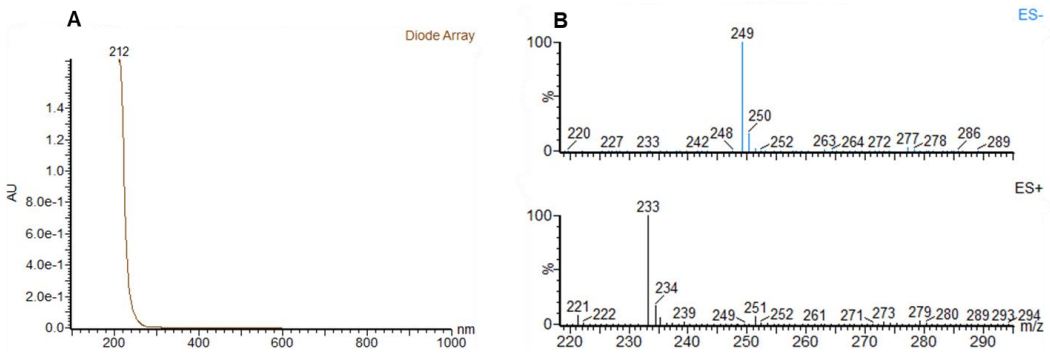
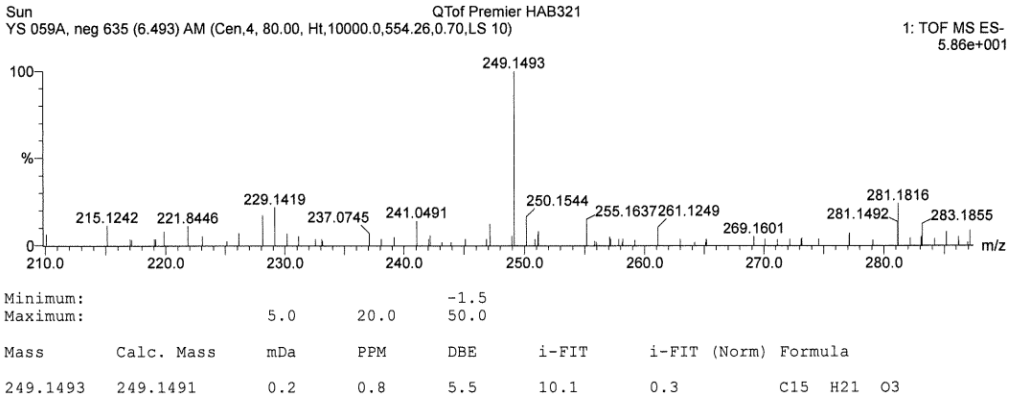


Figure 3.3.74 UV-absorption **(A)** and fragmentation pattern **(B)** of **141** in ES^+ (bottom) and ES^- (top) by LR-LCMS.



Minimum:				-1.5			
Maximum:		5.0	20.0	50.0			
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	i-FIT (Norm)	Formula
249.1493	249.1491	0.2	0.8	5.5	10.1	0.3	C15 H21 O3

Figure 3.3.75 HRMS data for **141**; m/z $[M - H]^-$ calc. mass is 249.1491, 249.1493 was found by HRMS.

In LCMS analysis of **142**, when subjected to positive ion mode (ES^+) ionization, **142** displayed distinctive mass fragmentation patterns, resulting in the appearance of peaks at *m/z* 203, 221, and 239 (Figure 3.3.76B). The negative ion mode (ES^-) did not yield clear results. The presence of the *m/z* 203 fragment can be attributed to the removal of a water molecule $[M + H - 2H_2O]^+$, while the 219 fragment is associated with the removal of a water molecule $[M + H - H_2O]^+$, and 239 corresponds to $[M + H]^+$. Furthermore, through HRMS analysis, we ascertained that **142** possesses a molecular formula of $C_{15}H_{26}O_2$ using HRGC-MS. Theoretical calculations for the *m/z* indicated an expected value of 238.1933, which closely matched the observed mass of 238.1934 (Figure 3.3.77).

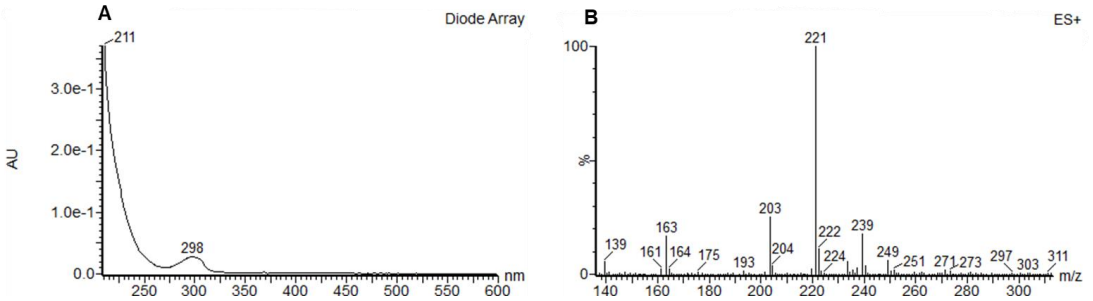
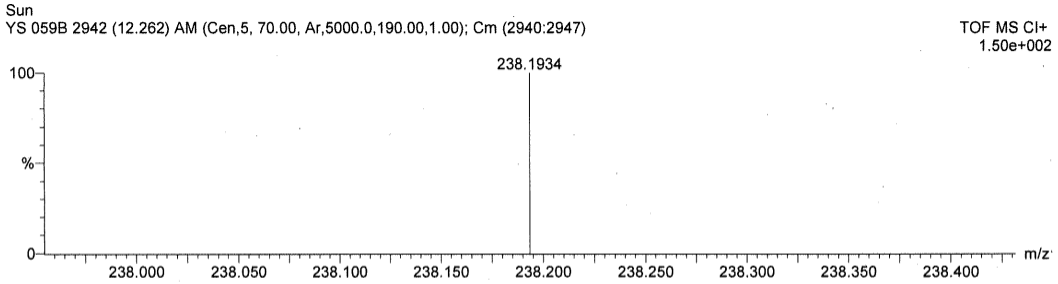


Figure 3.3.76 UV-absorption **(A)** and fragmentation pattern **(B)** of **142** in ES^+ by LR-LCMS.



Minimum:				-1.5		
Maximum:		5.0	40.0	50.0		
Mass	Calc. Mass	mDa	PPM	DBE	i-FIT	Formula
238.1934	238.1933	0.1	0.4	3.0	5546081.5	C15 H26 O2

Figure 3.3.77 HRMS data for **142**; *m/z* [M] calc. mass is 238.1933, 238.1934 was found by HRGC-MS.

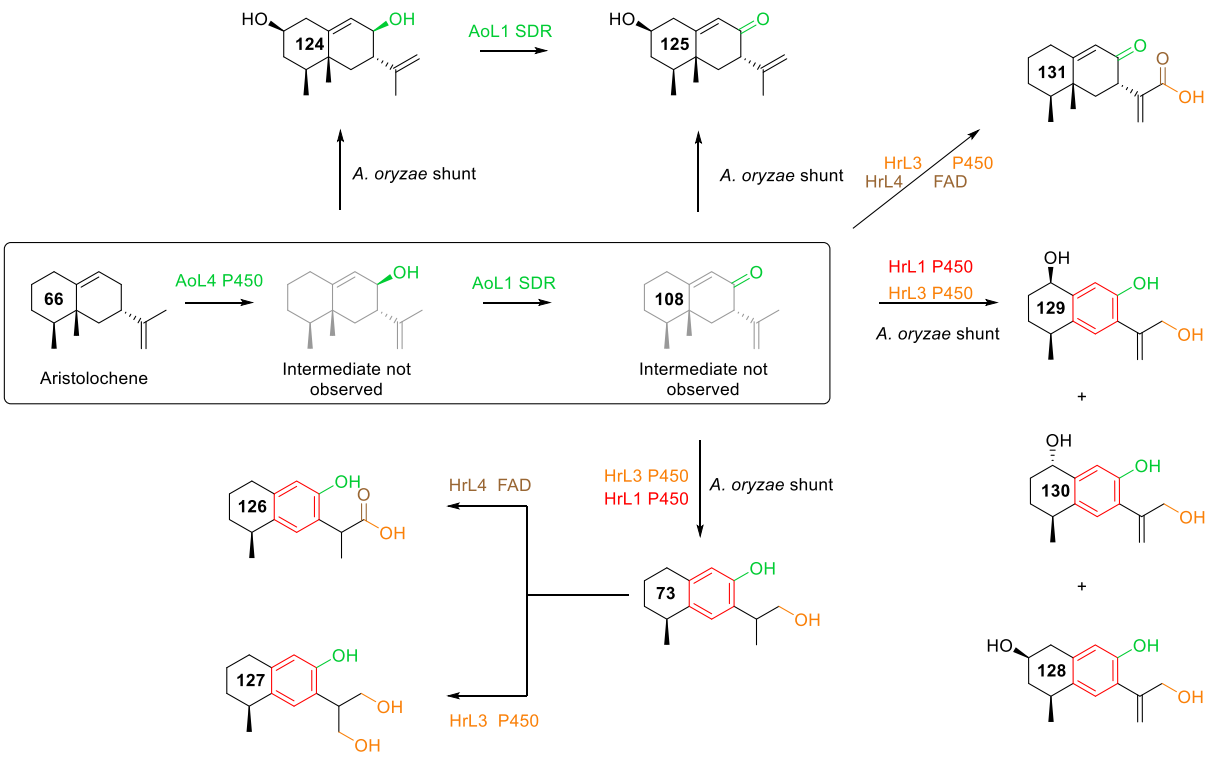
3.4 Discussion and Conclusion

In our research, we utilized *Aspergillus oryzae* as our primary expression host. *A. oryzae* is a domesticated variant derived from *A. flavus* in which most of the specialized metabolic biosynthetic gene clusters are rendered inactive. This modification proves advantageous for conducting heterologous expression experiments. The key benefit lies in the fact that newly introduced biosynthetic pathways do not need to compete with the host's endogenous pathways for precursor molecules. Consequently, any newly synthesized compounds can be more easily detected against a clean metabolic backdrop. Under our typical fermentation conditions, *A. oryzae* NSAR1 fails to produce sporogen AO-1 **27**. This lack of production could potentially stem from a terpene cyclase malfunction or inadequate expression. Nevertheless, our RT-PCR analysis has demonstrated that genes *aol4* (encoding P450) and *aol1* (encoding SDR) exhibit active transcription.

3.4.1 Biosynthesis of Hypoxylan A

As a result, we introduced the *H. rickii* terpene cyclase gene (*hrtc*) into *A. oryzae* NSAR1, regulating its expression using the inducible *A. oryzae amy*B promoter (P_{amyB}). Upon subjecting the transformed strains to LCMS analysis after fermentation under conditions that induced gene expression, we detected the production of two previously unknown compounds. These compounds were subsequently isolated

and subjected to comprehensive characterization *via* NMR spectroscopy, revealing their identities as 2, 8-dihydroxy aristolochene **124** and 2-hydroxy, 8-oxo-aristolochene **125** (Exp1).



Scheme 3.4.1 Compounds isolated from the expression of genes from *H. rickii*.

Additionally, we confirmed the presence of aristolochene **66** by employing GCMS and comparing the obtained data with literature references. It is important to note that we presumed the absolute stereochemistry of these compounds to align with that of **66** itself, while we established their relative stereochemistry through the analysis of coupling constants and NOESY correlations.

The first key observation is that in these experiments we did not need to introduce genes that encode oxidation at C-8. RT-PCR experiments (Section 3.3.2) had already shown that *aol4* and *aol1* from the native putative sporogen AO1 BGC, encoding a cytochrome P450 oxygenase and an SDR respectively, are transcribed under the fermentation conditions. These results show that the protein products are active, giving (the unobserved) 8-alcohol and 8-ketone **108** intermediates thet must be precursors of the later compounds **124** - **131** (Scheme 3.4.1).

Hypoxylan A **73**, sourced from *H. rickii*, exhibits a structure characterized by the absence of oxidation at positions 1-3, accompanied by the presence of a phenolic ring and the absence of C-15. To gain insights into the biosynthetic reactions responsible for this framework, we employed a comprehensive approach by coexpressing all the biosynthetic genes within the hr BGC. This yielded both **73** itself and its corresponding carboxylic acid **126** (Exp2).

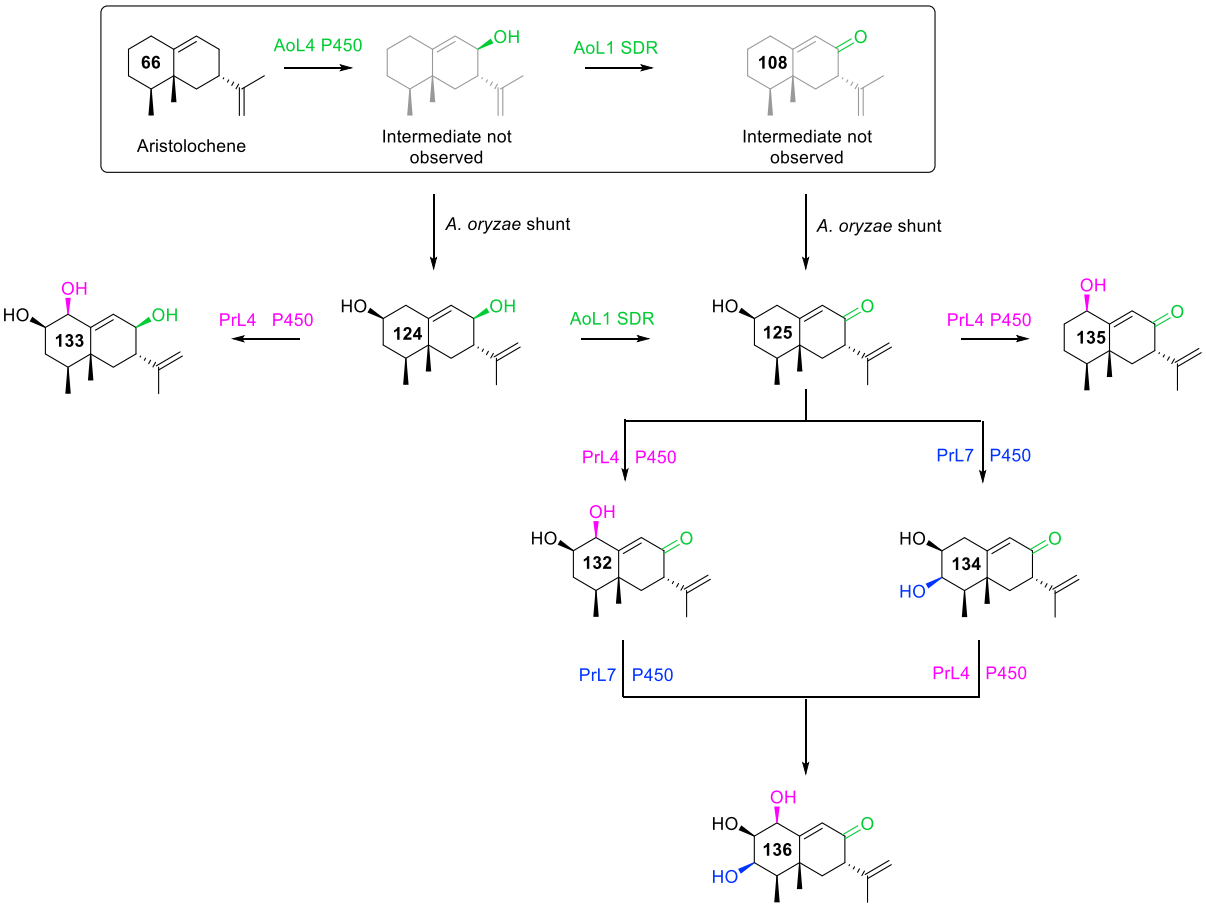
Exclusion of the SDR enzyme encoded by *hrl5* resulted in the production of **124** and **125** (Exp3). This result suggests that *hrl5* does not play a role in the biosynthesis of hypoxylan A **73**. Likewise, the omission of the FAD-dependent dehydrogenase encoded by *hrl4* led to the formation of the diol **127**, with no generation of more highly oxidized analogs (Exp4). Furthermore, the absence of the P450 enzyme encoded by *hrl3* (Exp5) resulted in the production of compound **128**, which retains the phenolic moiety. Compound **128** was also formed by coexpressing *hrtc* and *hrl1* alone (Exp8), accompanied by the generation of C1-alcohol epimers **129** and **130** (Exp7). Coexpression of *hrtc* with *hrl3* and *hrl4* led to the synthesis of the carboxylic acid **131** (Exp6).

Collectively, our investigations have substantiated the role of *hrtc, hrl1, hrl3*, and *hrl4* in the biosynthesis of hypoxylan A **73**. Enzymes within the *H. rickii* hypoxylan A **73** pathway play a pivotal role in catalyzing redox processes on the isopropenyl side-chain carbons 11-13. In a noteworthy discovery, the cytochrome P450 monooxygenase Hrl1 showcases activity of aromatase. It achieves this by catalyzing the oxidative elimination of the C-15 methyl group, resulting in the formation of the phenolic nucleus seen in hypoxylan A **73**. While aromatase is a well-documented enzyme in vertebrates and plays a pivotal role in estrogen biosynthesis, its existence in other organisms remains relatively uncharted territory. As far as our knowledge extends, this marks the first documented case of fungal aromatase-like activity.

3.4.2 Co-expression with PR-toxin genes

PR-toxin **107**, originating from *P. roquefortii* (pr), undergoes oxidation at both the C-1 and C-3 positions. When we coexpressed *hrtc* with *prl3, prl4, prl7*, and *prl9*, it resulted in the transformation of previously observed **124** and **125** into their 1-hydroxylated congeners **133** and **132**, alongside the formation of 3-hydroxylated congener **134** and 1, 3-dihydroxylated congener **136**, in addition to **135** (Exp9). When only *prl3* was coexpressed with *hrtc*, exclusively **124** and **125** were observed (Exp10), suggesting that *prl3* does not play a role in product formation. Similarly, when only *prl4* was coexpressed with *hrtc*, **132** and **133** were detected (Exp11), notably indicating that both were oxidized at C-1.

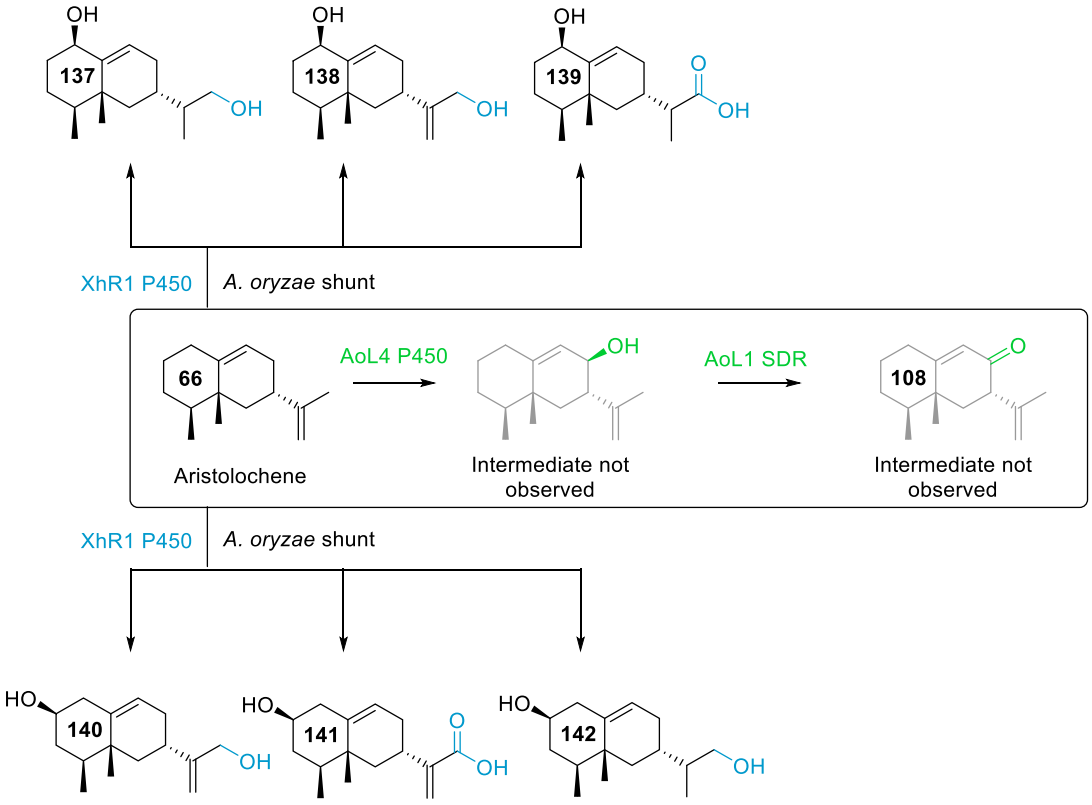
The introduction of *prl7* into this system further yielded the 3-hydroxylated congener **134** and the 1, 3-dihydroxylated congener **136** (Exp12). This evidence definitively establishes the role of Prl4 in oxidizing C-1 and the role of Prl7 in oxidizing C-3. Intriguingly, they both demonstrate the ability to work at both C-1 and C-3 simultaneously in the presence of the substrate 2-hydroxy, 8-oxo-aristolochene **125**. In summary, the oxidative enzymes within the PR-toxin BGC display the capability to hydroxylate at both the C-1 and C-3 positions.



Scheme 3.4.2 Compounds isolated from the expression of *hrtc* and genes from *P. roquefortii*.

3.4.3 Co-expression with *xhr1*

In the final phase, when we coexpressed *xhr1* with *hrtc* (Exp13), we observed the formation of six previously unreported aristolochene congeners, namely **137-142** (Scheme 3.4.3). Notably, among these, **137, 138**, and **139** exhibited oxidations at C-1, while **140, 141**, and **142** displayed oxidations at C-2. These specific oxidation patterns at C-1 and C-2 were likely influenced by an *A. oryzae* shunt mechanism. It is worth highlighting that these congeners primarily underwent oxidation at C-13 but remained unoxidized at C-8. This observation suggests that early oxidation at C-13 may hinder recognition by *A. oryzae* C-8 oxidases, rendering these compounds less likely to serve as substrates (Scheme 3.4.3).



Scheme 3.4.3 Compounds isolated from the expression of *hrtc* and *xhr1* from *Xylaria hypoxylon*.

The approach of exploring four BGCs, all involved in processing the same terpene framework, proved remarkably effective in providing valuable tools for future endeavors in total biosynthesis projects. Interestingly, the host organism, *A. oryzae*, appears to naturally possess catalysts, namely Aol1 and Aol4, responsible for the formation of the 8-hydroxyl and 8-oxo functional groups. This aligns with the requirements of all four BGCs, which necessitate the formation of the common intermediate, 8-ketone **108**, at the early stages of their respective biosynthetic pathways. Nonetheless, it is worth noting that **108** is known to be exceptionally unstable, and regrettably, during our experiments, we were unable to directly detect the presence of this compound, despite its prior identification as an intermediate in the PR-toxin pathway by Dickschat et al.[91]

These experiments demonstrate the feasibility and appeal of achieving total synthesis using the readily available genetic resources of fungi. Notably, we presented the inaugural instance of a one-step total synthesis of hypoxylan A **73**. Furthermore, these efforts yielded more intricate compounds, such as triol **136**, which is a C-1 epimer of the fungal natural product guignaderemophilane C **37**. Guignaderemophilane C **37** had previously been synthesized through an eleven-step total chemical synthesis process. Remarkably, the synthesis of epimer **136** in a single step was made possible through the combination of genes sourced from *H. rickii*, *P. roquefortii* and the host organism *A. oryzae*.

A. oryzae is commonly used as a host organism for pathway expression and genetic engineering, typically considered a clean host with minimal interference in engineered pathways. However, in this

project, *A. oryzae* unexpectedly exhibits heightened activity as a host organism. Notably, we have observed the active involvement of Aol4 (P450) and Aol1 (SDR). This proves advantageous, as all pathways leading to **27**, **73**, **123** and **107** require the presence of the 8-oxo functionality. Interestingly, the occurrence of 2-hydroxylation was unanticipated, and many of the observed products exhibit this particular modification. Additionally, there are other unexpected side reactions, exemplified by the putative route to compound **128**, which entails unintentional hydroxylation at both C-2 and C-13, in conjunction with the aromatase chemistry catalyzed by Hrl1.

These findings demonstrate the effectiveness of a strategy centered around the identification and screening of tailoring genes within pathways sharing the same carbon structure. Impressively, the manipulation of only six genes yielded twenty distinct compounds. It is highly likely that many more compounds can be synthesized using this same approach, given that only a fraction of available genes and their combinations were coexpressed.

Considering that aristolochene-type terpenes are prevalent among fungi, it's reasonable to assume that numerous other BGCs could be harnessed to further enhance diversity. This approach is expected to be equally successful for different categories of specialized metabolites. As we continue to discover and associate BGCs with specific compounds, the development of efficient total biosynthetic pathways for target compounds will become increasingly straightforward. In conclusion, these results strongly support the notion that the acquisition and loss of biosynthetic genes play a pivotal role in shaping the diversity and evolution of fungal specialized metabolites.

4. Total Biosynthesis of Fungal Tetraketide Pyrones

4.1 Introduction

4.1.1 Tetraketide Pyrones and Their Bioactivities

Tetraketide pyrones represent a substantial category of biologically active metabolites produced by fungi, with the structures characterized by a diverse oxygenation and acylation patterns. To illustrate, multiforisins A, B, and E (**143a, 143b** and **143e**) were isolated from the Ascomycete *Gelasinospora multiforis* IFM4498.[184] Their structures were elucidated as 2, 3, 4, 5-tetrasubstituted tetraketide *α*-pyrones (Figure 4.1.1). These three compounds feature an *O*-methyl group at C-3 and a formyl group at C-4, while their substitutions at C-2 vary, encompassing hydroxymethyl, acetoxymethyl, and *O*-methyl groups, respectively.[184,185]

Multiforisins A, B, and E (**143a, 143b** and **143e**) possess distinct levels of immunosuppressive activity. Their IC_{50} values were determined as follows: 0.6, 24, and 5 μg/ml against concanavalin A (ConA) - induced proliferation, and 0.6, 22, and 4 μg/ml against lipopolysaccharide (LPS) - induced proliferation of mouse spleen lymphocytes, respectively. Furthermore, multiforisin A exhibited an IC_{50} value of 10 μg/ml against human KB cells.[184,185]

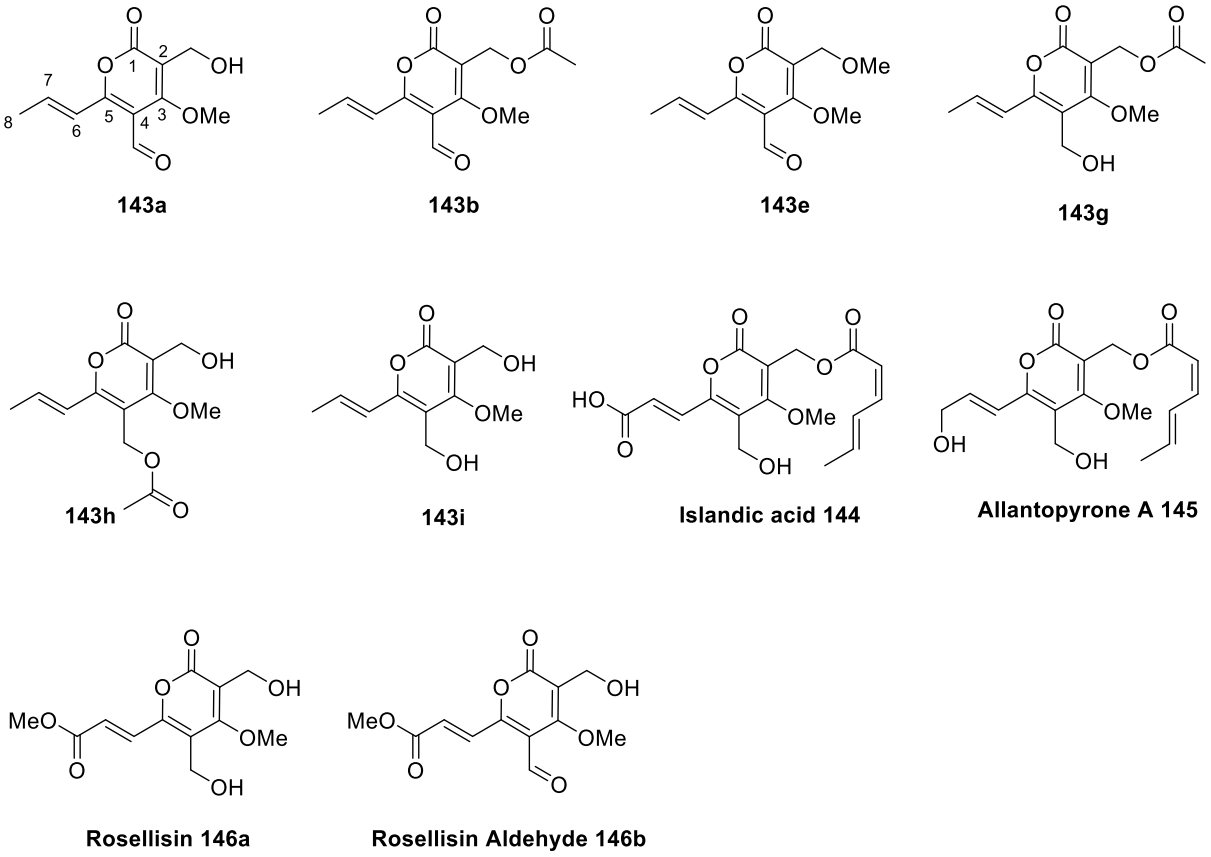


Figure 4.1.1 Examples of fungal tetraketide pyrones.

Multiforisins G, H, and I (**143g, 143h** and **143i**) were isolated from three Ascomycete fungi: *G. multiforis* IFM4498, *G. heterospora* 74-T-542-1, and *G. longispora* IFM4617.[185] These compounds are categorized as 2, 3, 4, 5-tetrasubstituted tetraketide α-pyrones (Figure 4.1.1). Multiforisin G **143g** bears an acetoxymethyl group at C-2, whereas multiforisin H **143h** features this group at C-4, and multiforisin I **143h** lacks any acetyl group altogether. The immunosuppressive properties of **143g, 143h** and **143i** were assessed against concanavalin A (Con A)-induced proliferation (T cells) and lipopolysaccharide (LPS)-induced proliferation (B cells) of mouse splenic lymphocytes. The IC_{50} values for **143g, 143h** and **143i** were determined as 0.9 and 1.2 μg/ml, 1.8 and 0.9 μg/ml, and >25 and 19 μg/ml, respectively, against these two cell types. Notably, multiforisin I **143i**, devoid of the acetoxymethyl group, exhibited significantly diminished activity. Furthermore, when tested against human leukemic HL-60 cells, multiforisin I **143i** displayed weaker activity compared to multiforisin G **143g** and H **143h**, with IC_{50} values of 5, 5, and >50 μg/ml, respectively.[184,185]

Isandic acid **144**, allantopyrone A **145**, rosellisin **146a**, and rosellisin aldehyde **146b** share a common structural feature, namely a substitution at C-8 (Figure 4.1.1). Among these compounds, **144** and **145** exhibit an almost identical structure, differing only in their C-8 substituents; **144** features a carboxylic acid group, while **145** possesses a hydroxyl group. The structures of **146a** and **146b** exhibit a distinction at C-4, with **146a** featuring a hydroxymethyl group, while **146b** incorporates a formyl group at this position.

Isandic acid **144** was originally extracted from *Penicillium islandicum* and displays potent cytotoxicity against *Yoshida sarcoma* cells, with an IC_{50} of 1 μg/ml for its methyl ester derivative.[186] Similarly, allantopyrone A **145**, isolated from the fungus *Allantophomopsis lycopodina* KS-97, acts as a potent cytotoxin with an IC_{50} of 0.32 μM, primarily through the inhibition of the NF-kB signaling pathway.[187] In addition, **145** was found to effectively inhibit the TNF-α-induced expression of ICAM-1 in human lung carcinoma A549 cells. On another note, the antibiotics rosellisin **146a** and rosellisin aldehyde **146b** were both isolated from *Hypomyces rosellus* and exhibited activity against *Staphylococcus aureus* in serial dilution tests, with effectiveness observed at a concentration of 30 ppm.[188,189]

4.1.2 Biosynthesis of Tetraketide Pyrones

The production of islandic acid **144** and related compounds has previously been achieved through multi-step total chemical synthesis.[190] However, these methods are inefficient, and there is a potential advantage in producing these compounds through total biosynthesis. Although some biogenetic studies have involved the incorporation of 1-^{13}C and 2-^{13}C acetates, providing information that rosellisin **146a** is a polyketide derived from four acetate units, there is still room for more efficient biosynthesis approaches.[188]

Additionally, recent work by Yuan and colleagues unveiled a fungal biosynthetic gene cluster from *Amphichorda felina*, which encodes the biosynthesis of tetraketide pyrones.[191] In their research, the

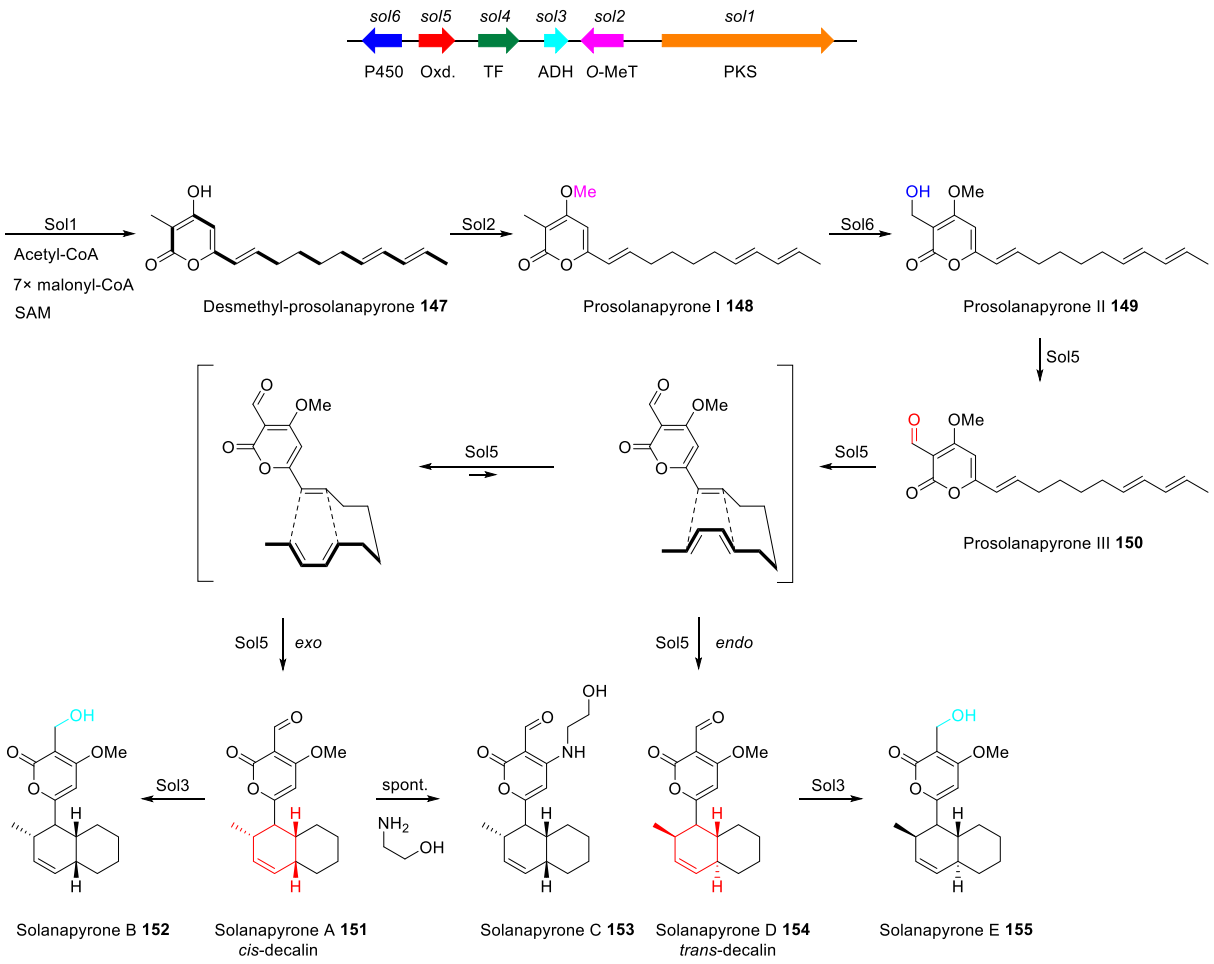
expression of the initial catalysts from this pathway led to the production of α-pyrone derivatives, but further steps in the pathway have not been investigated. Despite the notable bioactivities of compounds like **144** and **145**, along with the presence of an intriguing *E/Z* diene, the biogenetic aspects of this class of bioactive tetraketide pyrones, including the genes responsible for the tetraketide backbone and those governing modifications, remain relatively unexplored.

4.1.3 Biosynthesis of Solanapyrones

Octaketide pyrone **147** is synthesized by PKS Sol1. Prosolanapyrone I **148** is yielded through the addition of *O*-methyl by *O*-MeT Sol2. Transformation of **148** into prosolanapyrone II **149** is catalysed by P450 enzyme Sol6. The hydroxymethyl moiety of **149** is oxidized to an aldehyde by Sol5 to form **150**.[192] **150** is engaged in conjugation with the dienophile moiety, resulting in a reduction of its LUMO energy and a consequent lowering of the energy barrier for the subsequent [4+2] cycloaddition (Scheme 4.1.1).[193]

Synthesized prosolanapyrone III **150** was used to investigate the spontaneous non-enzymatic Diels–Alder cyclization.[194] The *endo* diastereoselectivity of the Diels–Alder cyclization increased with solvent polarity. In water, the observed selectivity for *trans*-decalin product solanapyrone D **154** was up to 23:1. Conversely, in the presence of Sol5, Diels–Alder cyclization showed exoselectivity (7:1), producing *cis*-decalin solanapyrone A **151**, with possible *endo* product formation in an aqueous environment (Scheme 4.1.1).[194]

The mechanism by which Sol5 catalyzes *exo*-selective [4+2] cycloaddition, despite the natural *endo* preference, remains unclear. It is proposed that the enzyme provides a stereochemical template, fixing **150** in an *exo*-transition state by forming hydrogen bonds with amino acid residues upon introducing the essential aldehyde function.[195] Pending final experimental data, the biosynthetic pathway likely involves aldehyde reduction (Sol3) to yield solanapyrones B **152** and E **155**.[195,196] This shows how nature conceals the aldehyde's role, seemingly enabling decalin formation through Diels–Alder cycloaddition. Solanapyrone A **151** undergoes spontaneous conversion to solanapyrone C **153** *via* 2-aminoethanol nucleophilic attack, observed both in *vitro* and in the fungal host.[197] Engineering efforts could involve supplementing host culture media with diverse nucleophiles to screen for novel solanapyrone derivatives, a strategy successful in producing new congeners in fungal azaphilones (Scheme 4.1.1).[198]



Scheme 4.1.1 Biosynthesis of solanapyrones in *Alternaria solani*.[192,193,198,199]

4.2 Project Aims

Firstly, we will utilize transcriptome data (*e.g* RNA-seq) from the native fungi capable of producing tetraketide pyrones. By comparing the growth conditions under which the native strain produces tetraketide pyrones with conditions where it does not, we can evaluate the expression levels of crucial genes. This transcriptome data will play an important role in defining the limits of biosynthetic gene clusters. Furthermore, RNA sequence will aid in the accurate pinpointing of intron locations within these genes. Following this, we can clone intron-free DNA fragments and effectively assemble these fragments using yeast-mediated recombination techniques to construct expresssion plasmids.

We also will conduct heterologous expression experiments in *A. oryzae* to elucidate the specific functions of the candidate gene cluster. To isolate the newly synthesized compounds, we will employ preparative LCMS. The structures of these compounds will be elucidated using NMR spectroscopy. Moreover, we plan to co-express genes from different clusters, such as islandic acid BGC, to generate a wider array of pyrone natural products.

4.3 Results

4.3.1 Bioinformatic Analyses

The previous group members conducted fermentation experiments with *Hypomontagnella monticulosa* as part of their investigations into sporochartine biosynthesis.[200] They also identified multiforisins H **143h** in *H. spongiphila*, a species closely related to *H. monticulosa* based on sporochartine research carried out.[201] Furthermore, they conducted transcriptome experiments on *H. monticulosa* under both producing and non-producing conditions, as well as the data analysis including processing the RNASeq. In this project, we've conducted bioinformatic analyses to uncover a potential biosynthetic gene cluster responsible for the multiforisin H pathway, as well as the islandic acid biosynthetic gene cluster sourced from a public database.

4.3.1.1 Identification of mfn and ila Biosynthetic Gene Clusters

To identify potential mfnBGC candidates, the genomic sequences of three closely related organisms (*H. monticulosa, H. submonticulosa*, and *H. spongiphila*) were utilized to establish a dedicated database using the Geneious software package. In our comparative analysis, the solanapyrone synthase (D7UQ44) was chosen as a reference template due to its pivotal role in solanapyrone biosynthesis.[202] This approach resulted in the identification of three polyketide synthase (PKS) candidates with notable similarities for each fungus (Figure 4.3.2).

We manually identified a candidate BGC in each of the three *Hypomontagnella* species. The three BGCs were very similar (Figure 4.3.2), containing the same number of genes in the same order and orientations. We then annotated all the genes in the *H. monticulosa* BGC (Table 4.3.1). Within these three presumed biosynthetic gene clusters, a diverse array of genes was identified, encoding enzymes such as DNA ligase (encoded by *mfnL9*), epimerase (encoded by *mfnL8*), transcriptional regulators (encoded by *mfnL7*, *mfnR1*), short-chain dehydrogenases/reductases (SDRs) (encoded by *mfnL6, mfnR4*), DNA polymerase (encoded by *mfnL5*), hydrolase (encoded by *mfnL4*), highly reducing polyketide synthases (hrPKS) (encoded by *mfnPKS2, mfnPKS1*), *O*-acetyltransferase (encoded by *mfnL3*), cytochrome P450s (encoded by *mfnL2, mfnR3*), flavin-dependent monooxygenase (FMO) (encoded by *mfnR2*), *O*-methyltransferase (encoded by *mfnL1*), and membrane proteins (encoded by *mfnR6*).

Table 4.3.1 Gene annotations of mfnBGC (blue area). Average expression levels from conditions of non-production (a) production (b).

Gene	AA	Putative Function	Predicted Cofactor	a non-producing	b producing	Log_2-fold change b/a
mfnL9	906	DNA ligase		-	-	-
mfnL8	343	Epimerase		462.24	286.16	-0.69
mfnL7	364	Transcriptional regulator		32.13	26.61	-0.27
mfnL6	186	SDR	NAD(P)	-	-	-
mfnL5	2248	DNA polymerase		-	-	-
mfnL4	163	Hydrolase		1.04	0.40	-1.39
mfnPKS2	2504	hrPKS	NAD(P)	101.57	290.21	1.51
mfnL3	487	*O*-acetyltransferase		-	-	-
mfnL2	537	P450		100.53	147.30	0.55
mfnL1	427	*O*-methyltransferase	SAM	169.97	307.35	0.85
mfnPKS1	2591	hrPKS	NAD(P)	60.11	86.35	0.52
mfnR1	654	Transcriptional regulator		-	-	-
mfnR2	526	FMO	FAD	203.14	703.65	1.79
mfnR3	580	P450		23.84	50.29	1.08
mfnR4	274	SDR	NAD(P)	386.59	529.38	0.45
mfnR5	279	Unknown		-	-	-
mfnR6	333	membrane protein		9.33	9.33	-0.72

We used transcriptome data to refine the boundaries of the mfnBGC. The expression patterns of functional genes within the mfnBGC were assessed by plotting the Log_2-fold change, which was calculated based on the expression levels of the strain growing under producing and non-producing conditions by previous co-workers. Significant upregulation was observed in genes from *mfnPKS2* to *mfnR4* under conditions conducive to production (Figure 4.3.1). In contrast, downregulation or remaining unexpressed was noted in genes located beyond this defined region. Finally, a well-defined and organized mfnBGC was obtained by these findings.

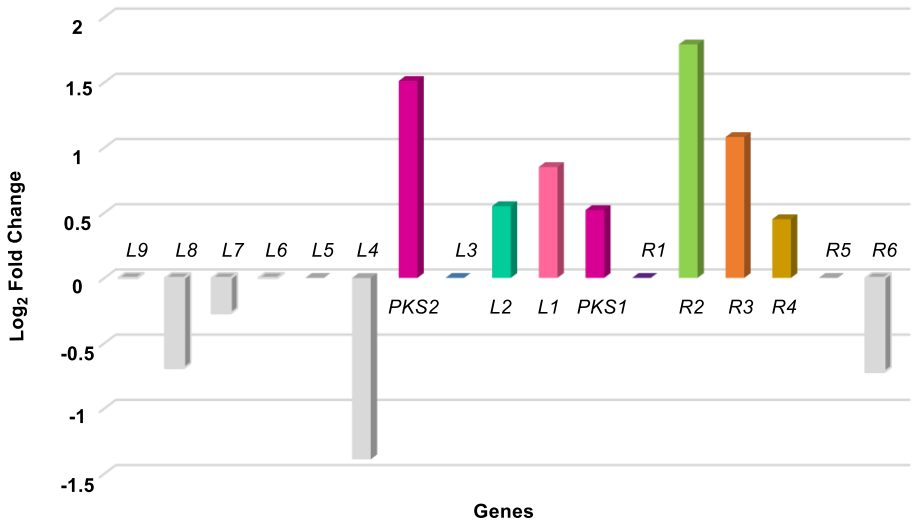


Figure 4.3.1 Bar chat of Log_2-fold changes obtained from producing vs non-producing conditions shows the expression level of the predicted genes of mfnBGC.

We used the mfnBGC as a reference to conduct a BLAST search against the genome of *Penicillium islandicum*, the producer of islandic acid.[186]. We identified a gene cluster (ilaBGC) that exhibited a high degree of identity to the mfnBGC. While IlaPKS1 and most of tailoring enzyme proteins shared high similarities with that of mfnBGC, IlaPKS2 did not show any similarity to MfnPKS2.

A detailed analysis of ilaBGC was undertaken and a range of probable functions of proteins were identified including two hrPKS (encoded by *ilaPKS1*and *ilaPKS2*), two P450 oxygenases (encoded by *ilaR4* and *ilaR6*) two SDRs (encoded by *ilaR7* and *ilaR8*), along with an *O*-MeT (encoded by *ilaR1*), an *O*-AcT (encoded by *ilaR2*), an FMO (encoded by *ilaR5*), a transcription factor enzyme (encoded by *ilaR3*), and a transporter enzyme (encoded by *ilaR9*).

Table 4.3.2 Proposed functions of ilaBGC.

Gene	Locus_tag (PISL3812_)	AA	Protein BLAST	Putative Function	Predicted cofactor	Best hit accession
IlaPKS1	09789	2630	Prosolanapyrone synthase	hrPKS	NAD(P), SAM	D7UQ44
IlaR1	09788	427	*O*-methyltransferase	*O*-MeT	SAM	D7UQ43
IlaR2	09787	468	Probable acetyltransferase	*O*-AcT		Q0CS99
IlaR3	09786	649	Probable transcription factor	TF		D7UQ41
IlaR4	09785	473	Cytochrome P450 monooxygenase	P450		M2UJ60
IlaR5	09784	466	FAD-linked oxidoreductase	FMO	FAD	W6QEK0
IlaR6	09783	581	Cytochrome P450 monooxygenase	P450		Q9C1I4
IlaR7	09782	273	Short-chain dehydrogenase/reductase family 32C member 1	SDR	NAD(P)	Q0IH28
IlaPKS2	09781	2972	Highly reducing polyketide synthase	hrPKS	NAD(P), SAM	A0A084API3
IlaR8	09780	249	Short-chain dehydrogenase/reductase	SDR	NAD(P)	A0A455LLX2
IlaR9	09779	563	MFS-type transporter	Transporter		A0A1V6PBC8

Subsequently, we conducted an extensive comparison that encompassed the solBGC, mfnBGCs from the three *Hypomontagnella* organisms, ilaBGC, and ampBGC.[191] This comparative analysis was visually represented through a cluster map, which effectively illustrated all six gene clusters (Figure 4.3.2).[17]

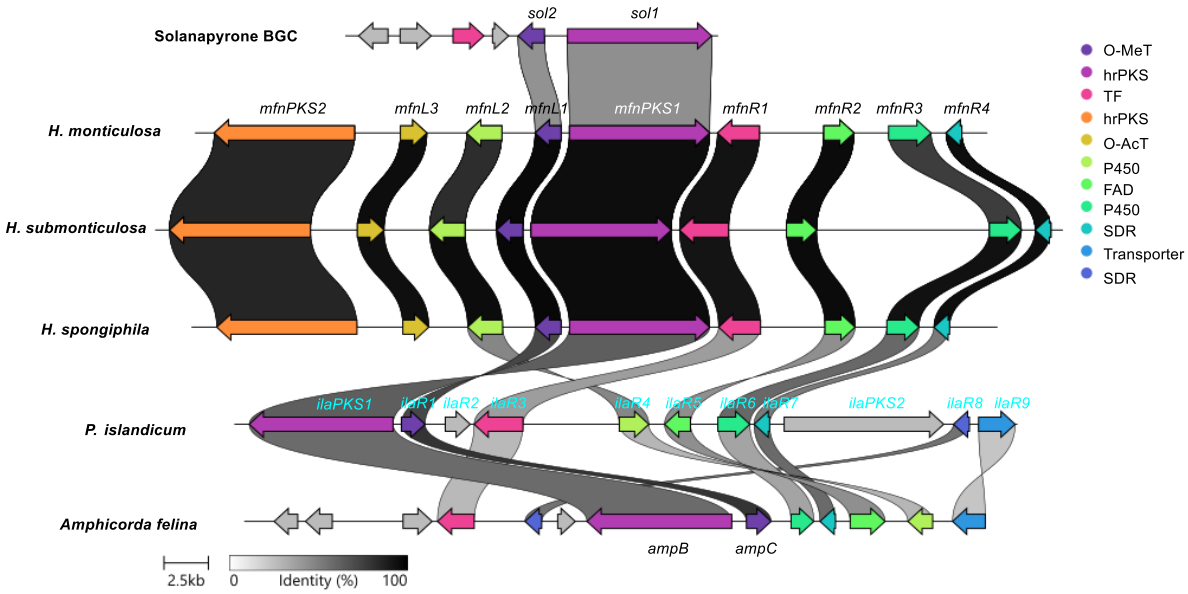


Figure 4.3.2 BGC alignment of solBGC, mfnBGCs from the three *Hypomontagnella* organisms, ilaBGC, and ampBGC by Clinker.[17]

As the protein sequence of MfnPKS2 exhibited no identity with that of IlaPKS2, we proceeded to analyze the domains of both. Firstly, we found that the active site cysteine on the KS domain of MfnPKS2 was mutated to serine (Figure 4.3.3). Mutation of the active site cysteine to serine would result in a KS that is unable to elongate carbon units, thus suggesting that this PKS is inactive.[203,204]

We observed the presence of a carnitine acyl transferase (cAT) C-terminal domain in IlaPKS2 (Figure 4.3.4). Such domains are uncommon in fungal PKS. However, in a particular case studied by Tang and colleagues, a cAT domain played a role in the reversible transfer of the completed polyketide to a polyol.[205]

Figure 4.3.3 Blastp multiple sequence alignment of the KS domain of mfnPKS2 with that of other PKS enzymes indicated that the amino acid residues in all three active sites were mutated. # Active site in red.[206]

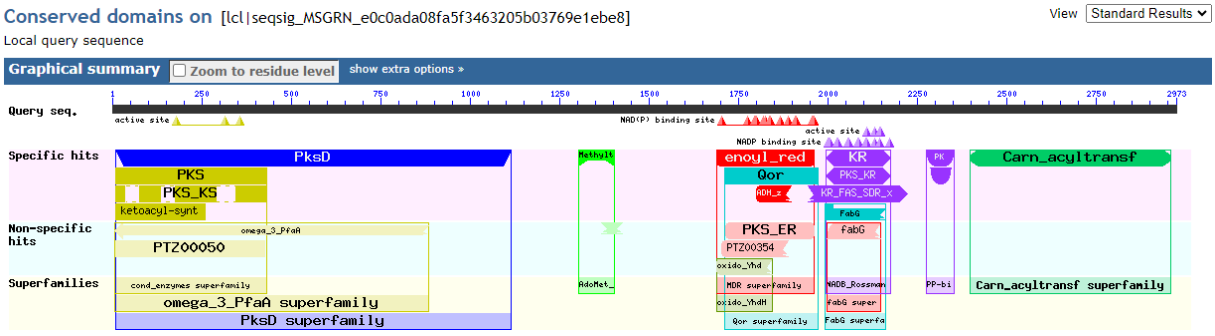


Figure 4.3.4 Conserved domains analysis of IlaPKS2 showed a cAT domian.

4.3.1.2 Intron Analysis of mfnBGC

The processed RNA-Seq data was aligned with the mfnBGC reference using tools in integrated in the Geneious software package. Through the examination of the RNA reads, we achieved precise identification of intron positions, benefiting from the high quality of the reads. For instance, regarding *mfnL1*, *mfnL2*, *mfnR4*, *mfnR2*, and *mfnL3* (Figure 4.3.5 A, B, C, D, E), the superior quality of the reads facilitated accurate intron position determination. However, when it came to *mfnPKS1*, *mfnPKS2*, and *mfnR3* (Figure 4.3.5 H, G, F), although some reads offered insights into intron localization, we found it necessary to utilize prediction tools such as antiSMASH[15] and FGENESH[207] for definitive confirmation of intron positions in specific segments.

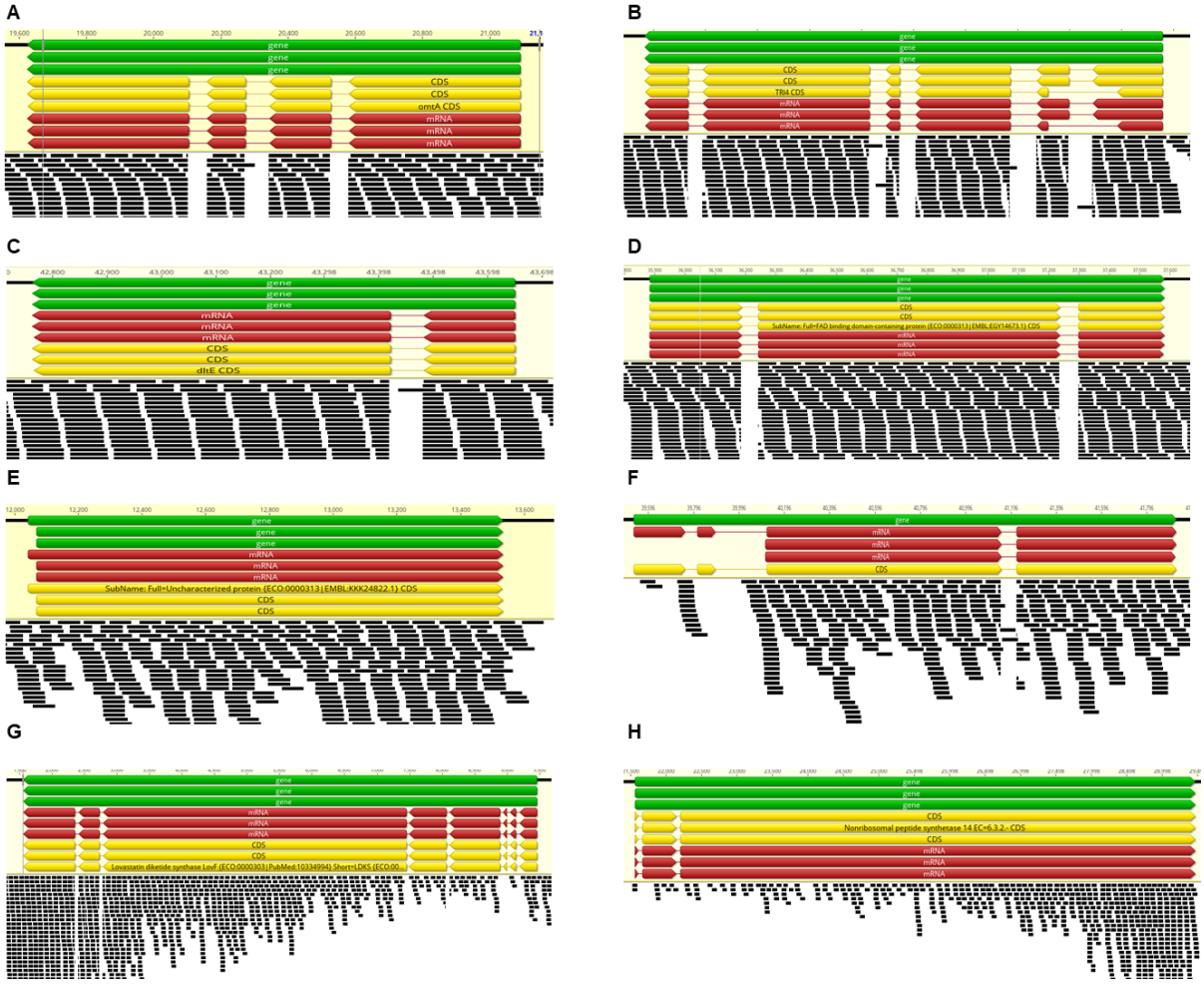


Figure 4.3.5 The RNASeq mapping of *mfnL1* **(A)**, *mfnL2* **(B)**, *mfnR4* **(C)**, *mfnR2* **(D)**, *mfnL3* **(E)**, *mfnR3* **(F)**, *mfnPKS2* **(G)**, *mfnPKS1* **(H)**. Intron positions are indicated by gaps, mapped reads are represented by black bars, genes are depicted with green bars, mRNA with red bars, and CDS with yellow bars.

4.3.2 Plasmid Construction and Transformation of *A. oryzae*

4.3.2.1 Overview of Constructed Plasmids

Subsequent experiments employed three modified vectors (pTYGs), each tailored with distinct selection markers (*argB*, *sC*, *adeA*) to facilitate targeted selection in *A. oryzae* NSAR1.[130,208] The genes encoding MfnPKS1, MfnPKS2 and other tailoring enzymes were cloned through intron-free fragments from gDNA based on transcriptome data, and these fragments were subsequently inserted into the PE-YA vector using yeast recombination to generate the entry plasmids. Concurrently, tailoring genes, including genes encoding P450 oxygenases, were integrated into pTYGs through yeast recombination involving various promoters and terminators (P/T_{amyB}, P/T_{adh}, P/T_{gpdA}, *and* P/T_{eno}) to produce the destination plasmids. The entry plasmid and destination plasmid were combined and subjected to LR recombination, resulting in the final plasmid containing both PKS and tailoring genes, denoted as PL_ (Figure 4.3.6, 4.3.7).

The genes encoding IlaPKS2, IlaR2, and IlaR8 were cloned through intron-free fragments from the genomic DNA of *Penicillium islandicum* strain iBT20602. All DNA inserts were subjected to sequencing, revealing a complete correspondence with the respective sequences found in *Penicillium islandicum* WF-38-12 [209] (ATCC 26535), which was recognized as the source organism for producing islandic acid [186].

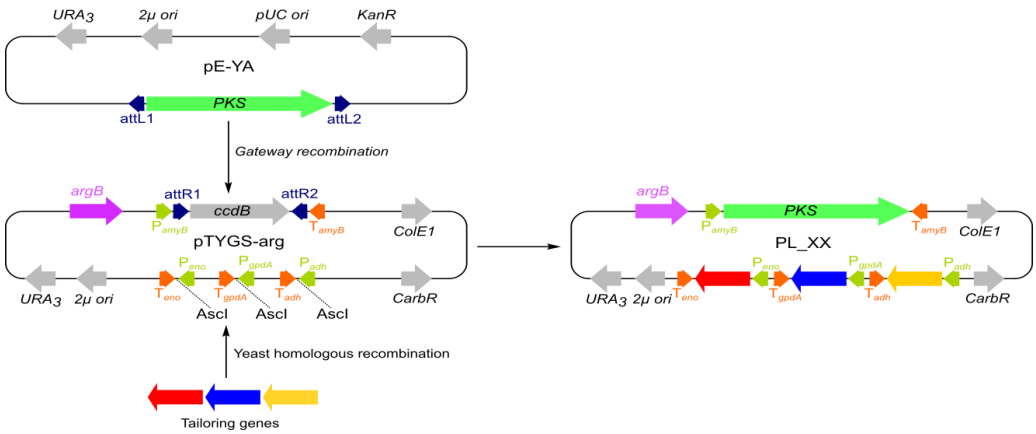


Figure 4.3.6 Workflow of construction of plasmids.

In this project, 14 plasmids were built (Table 4.3.3). The subunits of each plasmid were plotted in a cartoon map (Figure 4.3.7), which show the postitons of genes and selection markers.

Table 4.3.3 Plasmids constructed in this study.

Construct ID	Plasmids	Features
PL01	pTYGS_*arg-mfnPKS1*	P_{amyB} promotes *mfnPKS1*
PL02	pTYGS_*arg-mfnPKS1-mfnL2*	P_{amyB} promotes *mfnPKS1*, P_{adh} promotes *mfnL2*
PL03	pTYGS_*arg-mfnPKS1-mfnR3*	P_{amyB} promotes *mfnPKS1*, P_{adh} promotes *mfnR3*
PL04	pTYGS_*arg-mfnPKS1-mfnL1*	P_{amyB} promotes *mfnPKS1*, P_{adh} promotes *mfnL1*
PL05	pTYGS_*arg-mfnPKS1-mfnL2-mfnL1*	P_{amyB} promotes *mfnPKS1*, P_{adh} promotes *mfnL2*, P_{gpdA} promotes *mfnL1*
PL06	pTYGS_*arg-mfnPKS1-mfnL1-mfnR3*	P_{amyB} promotes *mfnPKS1*, P_{gpdA} promotes *mfnL1*, P_{eno} promotes *mfnR3*
PL07	pTYGS_*arg-mfnPKS1-mfnL2-mfnL1-mfnR3*	P_{amyB} promotes *mfnPKS1*, P_{adh} promotes *mfnL2*, P_{gpdA} promotes *mfnL1*, P_{eno} promotes *mfnR3*
PL08	pTYGS_*ade-mfnR2*	P_{eno} promotes *mfnR2*
PL09	pTYGS_*ade-mfnR4*	P_{gpdA} promotes *mfnR4*
PL10	pTYGS_*ade-mfnL3*	P_{adh} promotes *mfnL3*
PL11	pTYGS_*ade-mfnPKS2-mfnL3*	P_{amyB} promotes *mfnPKS2*, P_{adh} promotes *mfnL3*
PL12	pTYGS_*met-ilaPKS2*	P_{amyB} promotes *ilaPKS2*
PL13	pTYGS_*met-ilaPKS2-ilaR2*	P_{amyB} promotes *ilaPKS2*, P_{gpdA} promotes *ilaR2*
PL14	pTYGS_*met-ilaPKS2-ilaR8-ilaR2*	P_{amyB} promotes *ilaPKS2*, P_{adh} promotes *ilaR8*, P_{gpdA} promotes *ilaR2*

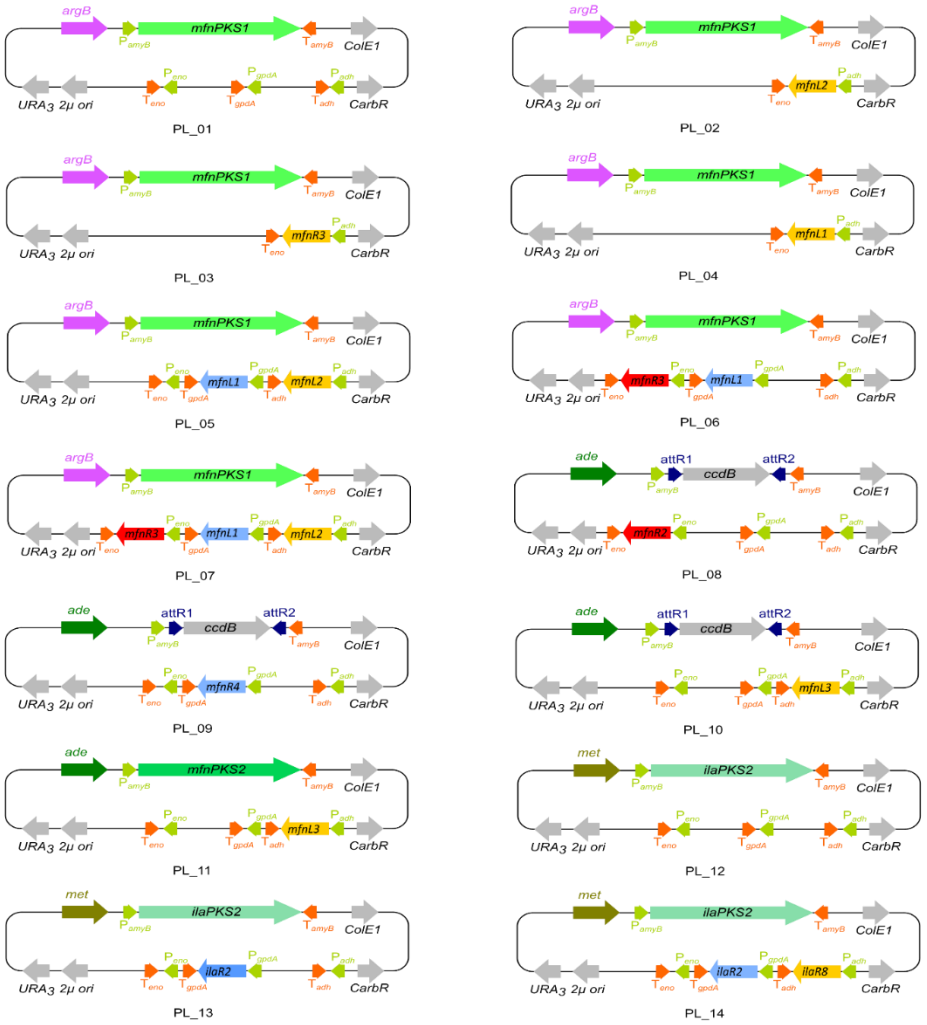


Figure 4.3.7 Plasmids for heterologous expression experiments in *A. oryzae*.

4.3.2.2 Gene Combinations and Transformation

We also employed heterologous expression in *A. oryzae* to achieve two objectives: to uncover the functions of genes derived from mfnBGC and ilaBGC, and to generate additional compounds of interest. As the host strain, we also selected the *A. oryzae* NSAR1 quadruple auxotrophic mutant (*niaD*$^{-}$, *sC*$^{-}$, △ *argB*, *adeA*$^{-}$). The introduction of the constructed vectors into *A. oryzae* NSAR1 was followed by their selection, utilizing various combinations (Table 4.3.4). After conducting two rounds of screening, we successfully obtained transformants on the CZD/S selection medium. These transformants were transferred to DPY medium to facilitate subsequent fermentation. The culture broth of each transformant underwent extraction and analysis using LCMS for further characterization.

Table 4.3.4 Combinations of plasmids for each experimental group

Gene	*mfnPKS2*	*mfnL3*	*mfnL2*	*mfnL1*	*mfnPKS1*	*mfnR2*	*mfnR3*	*mfnR4*	*ilaPKS2*	*ilaR2*	*ilaR8*	Plasmids
Exp	hrPKS	O-AcT	P450	O-MeT	hrPKS	FMO	P450	SDR	hrPKS	O-AcT	SDR	
1					✓							PL01
2			✓		✓							PL02
3					✓		✓					PL03
4				✓	✓							PL04
5			✓	✓	✓							PL05
6				✓	✓		✓					PL06
7			✓	✓	✓		✓					PL07
8			✓	✓	✓	✓	✓					Exp 7+ PL08
9			✓	✓	✓		✓	✓				Exp 7+ PL09
10		✓	✓	✓	✓		✓					Exp 7+PL10
11	✓	✓	✓	✓	✓		✓					Exp 7+ PL11
12			✓	✓	✓		✓		✓			Exp 7+ pTYGS-ade + PL12
13	✓	✓	✓	✓	✓		✓		✓	✓		Exp 11+ PL13
14			✓	✓	✓		✓		✓	✓	✓	Exp 11+ pTYGS-ade + PL14

4.3.3 Expression of *mfnPKS1*

Based on bioinformatic analyses, the identity between MfnPKS1 and solanapyrone synthase [202] (D7UQ44) suggests that *mfnPKS1* likely encodes a highly-reducing polyketide synthase (hrPKS) responsible for synthesizing a polyketide-type pyrone, similar to the backbone of prosolanapyrone **148**. To investigate this, we cloned the CDS sequence of *mfnPKS1* and inserted it into the pTYGs-*arg* vector under the control of the maltose-inducible P_{amyB}. The resulting plasmid, PL01, was introduced into *A. oryze* NSAR1, which was selected on CZD/S agar lacking arginine.

The transformed strains were cultured in DPY medium, and the culture medium was collected and subjected to ethyl acetate extraction. The crude extract was then analyzed using LCMS, with a crude extract from the wild-type strain used as a control. A comparison of the DAD chromatograms and TIC

spectra of the wild-type and Exp1 strains revealed the presence of a new peak **156**, which was not observed in the wild-type strain (Figure 4.3.8).

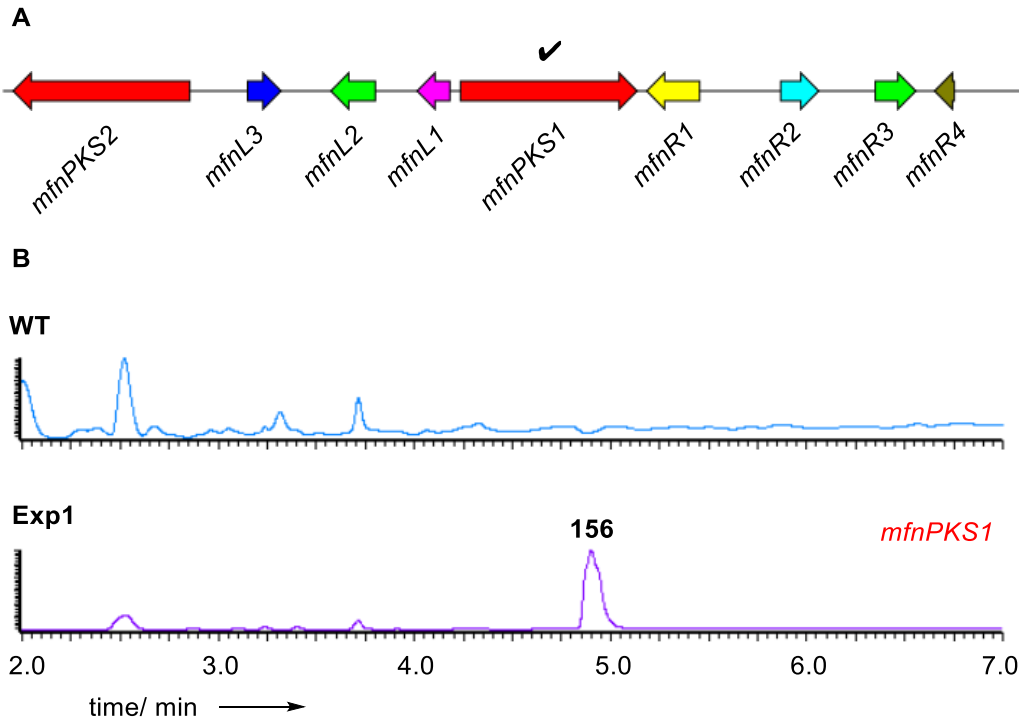


Figure 4.3.8 A, mfnBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of WT, Exp1 LCMS analysis.

4.3.3.1 Characterization of 156

Based on the LCMS data for **156**, when analyzed in positive ion mode (ES^+), it displayed some mass fragmentation patterns, resulting in relatively stronger peaks at *m/z* 181 and 203, along with *m/z* 361. Conversely, when analyzed in negative ion mode (ES^-), **156** generated a strong peak at *m/z* 179 (Figure 4.3.9A). The appearance of the *m/z* 181 fragment can be attributed to the $[M + H]^+$ ion. The presence of the *m/z* 203 peak indicates the formation of $[M + Na]^+$ ion, while the *m/z* 361 peak corresponds to $[2M + H]^+$. Additionally, the signals at *m/z* 179 align with $[M - H]^-$ ions. These findings partially confirm the anticipated mass of 180 (Figure 4.3.9A).

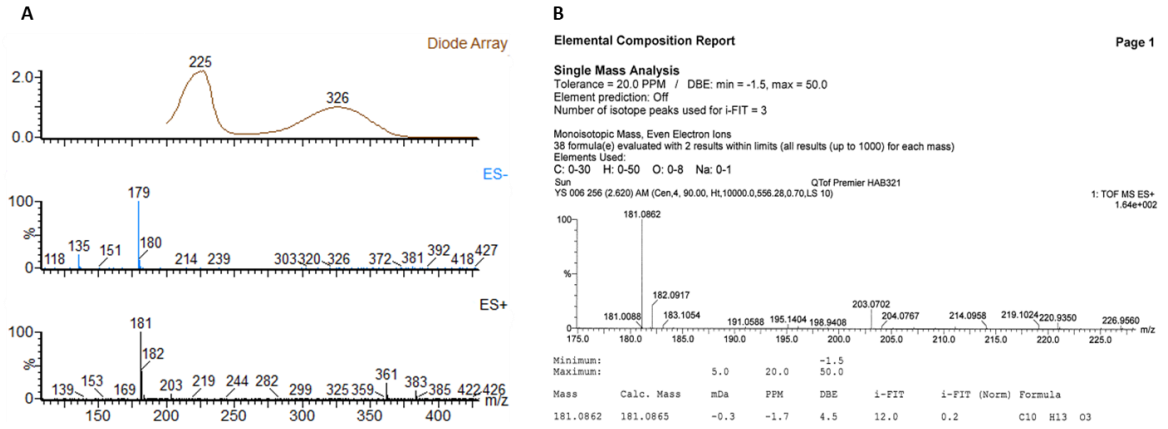


Figure 4.3.9 A, UV-absorption (top) and fragmentation pattern of **156** in ES^- (middle) and ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **156.**

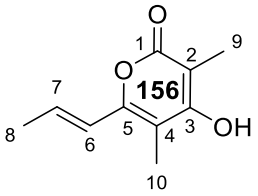
Through HRMS analysis, we established that **156** possesses a molecular formula of $C_{10}H_{12}O_3$. Theoretical calculations for the $[M + H]^+$ ion indicated an expected *m/z* value of 181.0865, which closely matched the observed mass of 181.0862 (Figure 4.3.9B). Furthermore, an absorption peak was observed within the UV maxima at 225 and 326 nm, suggesting the presence of an α-pyrone substructure.[210]

A 1-liter culture of DPY medium containing fresh cultures of the Exp1 strain was incubated, and the crude extract was subjected to preparative LCMS. Approximately 123 mg of **156** was isolated, and a portion of it underwent 1D and 2D NMR analysis. The ^1H NMR spectrum revealed two olefinic proton signals at δ_H 6.41 (dddd, *J* 15.3, 1.7, 1.7, 1.7 Hz) and δ_H 6.6 (dddd, *J* 15.4, 6.9, 6.9, 6.9 Hz), in addition to three methyl groups at δ_H 1.92 (3H, m), δ_H 1.92 (3H, m), and δ_H 2.0 (3H, s, Table 4.3.5).

The assignment of the above-mentioned five carbons was achieved through HSQC and ^{13}C NMR spectra, corresponding to δ_C values of 121.4, 134.2, 18.6, 9.0, and 9.4, respectively. Beyond these protonated carbons, an additional five non-protonated carbons were attributed to the 2, 3, 4, 5-tetrasubstituted α-pyrone skeleton (δ_C 99.9, 167.8, 108.6, 153.5, and the carbonyl carbon at δ_C 167.6, Table 4.3.5).

We utilized HMBC correlations to determine the positions of the methyl groups. Specifically, H-9 correlated with C-1, C-2, and C-3, H-10 correlated with C-3, C-4, and C-5, and H-8 correlated with C-6 and C-7 (Table 4.3.5). These correlations revealed that C-9 is positioned at C-2, C-10 is located at C-4, and C-8 is situated at C-7 (Table 4.3.5). The HMBC correlations from H-6 to C-5, as well as from H-7 to C-5 strongly suggested that the propenyl chain was attached at the C-5 position of the α-pyrone ring (Table 4.3.5).

We confirmed the 6*E* geometric configuration by observing a coupling constant of H-6/H-7 (*J* 15.3 Hz) between the two olefinic protons. In the end, we verified that the elucidated structure was identical to the known compound, amphichopyrone A **156**. This confirmation was supported by a comparison of the ^1H and ^{13}C NMR data (Table 4.3.5).



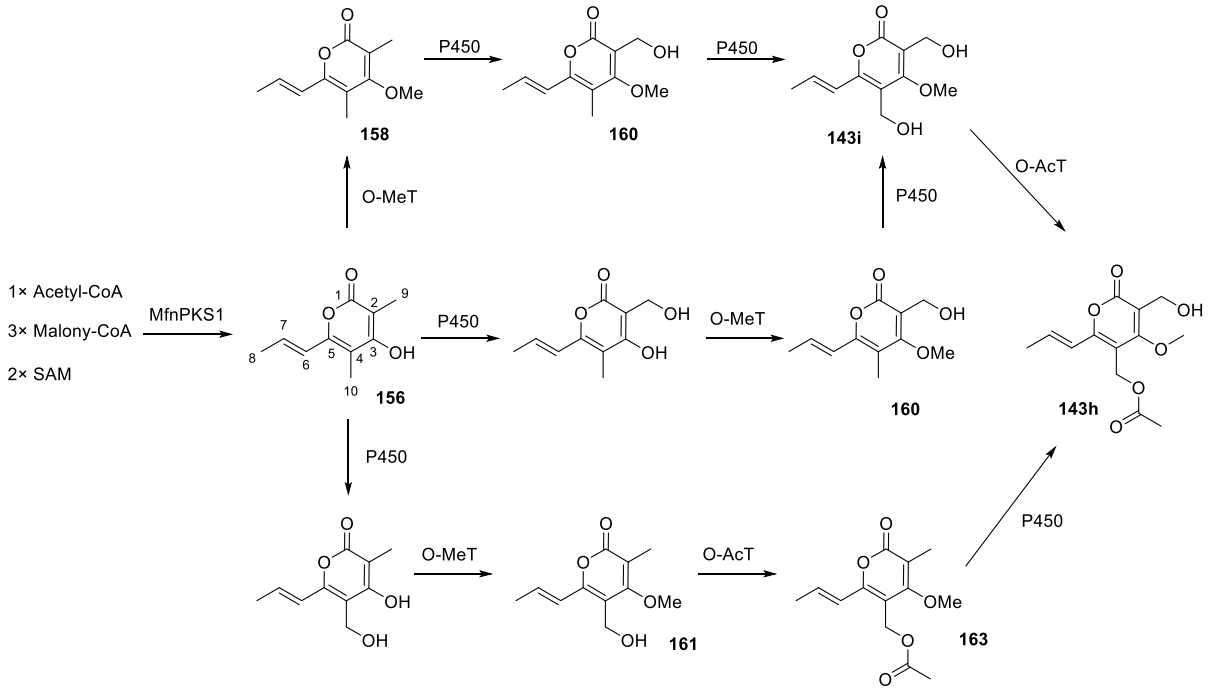
Chemical Formula: $C_{10}H_{12}O_3$
Exact Mass: 180.0786

Table 4.3.5 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **156** recorded in CD_3OD, Literature [211] data was measured in acetone-d6.

Compound 156						
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**	**δ_C / ppm literature** [211]	**δ_H / ppm (J/Hz) literature** [211]
1	167.6				163.6	
2	99.9				99.0	
3	167.8				164.3	
4	108.6				106.3	
5	153.5				151.6	
6	121.4	6.41, 1H, dddd (15.3, 1.7, 1.7, 1.7)	7, 8	5, 8	120.6	6.42, dq (15.4, 1.3)
7	134.2	6.6, 1H, dddd (15.4, 6.9, 6.9, 6.9)	6, 8	5, 8	132.0	6.50, dq (15.4, 6.0)
8	18.6	1.92, 3H, m	6, 7	6, 7	17.6	1.90, d (6.0)
9	9.0	1.92, 3H, m		1, 2, 3	8.6	1.94, s
10	9.4	2.0, 3H, s		3, 4, 5	8.4	2.01, s

4.3.4 Expression of *mfnL2*, *mfnR3*, *mfnL1* with *mfnPKS1* Respectively

Once the structure of **156** was elucidated, we confirmed the activity of MfnPKS1 in synthesizing a tetraketide pyrone, which forms the backbone of the product multiforisin H **143h**. According to the annotation of the mfnBGC (Table 4.3.1), we also identified: two genes, *mfnL2* and *mfnR3* encoding P450 oxygenases; the gene *mfnL1* encoding an *O*-methyltransferase (*O*-MeT); and a gene likely to be responsible for *O*-acetylation (*O*-AcT). These four genes were proposed to be involved in modifying the substituent units at positions C-9, O-11, and C-10, but the specific order of their reactions needed clarification.



Scheme 4.3.1 Proposed pathways for the biosynthesis of multiforisin H **143h.**

Several possible scenarios were considered for the biosynthetic pathway. The initial action of one P450 on C-9, followed by *O*-MeT, another P450 catalyzing the formation of multiforisin I **143i**, and finally *O*-AcT leading to the production of multiforisin H **143h** (Scheme 4.3.1). Alternatively, one P450 could initiate the process by acting on C-10. Subsequently, *O*-MeT would add a methyl group to O-11, and *O*-AcT would then convert it to **163**. Another P450 could be responsible for adding a hydroxyl group at C-9 to form multiforisin H **143h**. *O*-MeT might be the starting point, followed by the oxidation of C-9 and C-10 by two P450 enzymes to form multiforisin I **143i**. This would then be followed by the finalization of the conversion to multiforisin H **143h** by *O*-AcT. Another possible scenario involves both P450 enzymes initially generating two hydroxyl groups, followed by *O*-MeT and then *O*-AcT. Considering the presence of two P450 enzymes, various steps and pathways are feasible within this biosynthetic process (Scheme 4.3.1). Further research is required to determine the precise sequence and mechanisms involved in the pathway.

To elucidate the functions and sequence of reactions involving the two P450 and *O*-MeT, we conducted a series of heterologous experiments. Initially, we introduced the *mfnL2* gene (encoding P450) along with co-expression of *mfnPKS1* (Exp2). Compound **156** was observed as the product of MfnPKS1, as expected. Interestingly, upon comparing the data with the wild-type and the *mfnPKS1* transformant, we observed a novel peak **157**, appearing earlier than peak **156** (Figure 4.3.10). This compound was isolated following a 1-liter fermentation in DPY with a titre of 33 mg/L and subsequent chemical extraction.

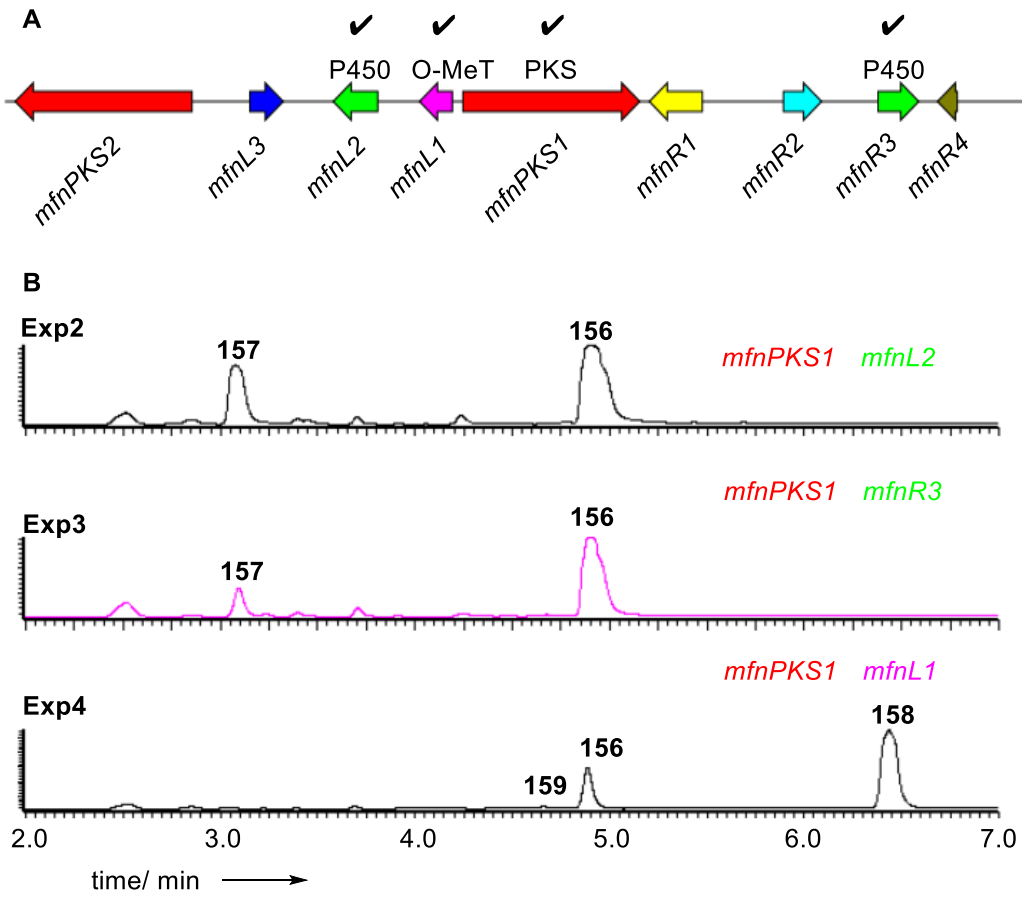


Figure 4.3.10 A, mfnBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of Exp2, Exp3, Exp 4.

4.3.4.1 Characterization of 157

We propose that compound **157** is an oxidative congener at either C-9 or C-10 of substrate **156**. This hypothesis is based on the observation that multiforisin I **143i** and H **143h** both exhibit oxidation at C-9 and C-10. Compound **157** has a molecular mass of 196, as indicated by mass fragmentation patterns that yield relatively strong peaks at *m/z* 197 and 219 in positive ion mode (ES^+), and a peak at *m/z* 195 in negative ion mode (ES^-, Figure 4.3.11A). The *m/z* 197 fragment can be attributed to the $[M + H]^+$ ion, while the *m/z* 219 peak suggests the formation of $[M + Na]^+$ ion. Additionally, the *m/z* 393 peak corresponds to $[2M + H]^+$ ion, and the signals at *m/z* 195 align with $[M - H]^-$ ions (Figure 4.3.11A). These findings provide partial confirmation of the expected mass of 196 of **157**. HRMS further supports the determination that **157** has a mass of *m/z* 196, as the observed $[M + H]^+$ ion at *m/z* 197.0816 closely matches the calculated $[M + H]^+$ ion at *m/z* 197.0814. Moreover, the UV absorption spectrum of **157** closely resembles that of **156**, displaying absorption peaks at 227 and 325 nm (Figure 4.3.11A).

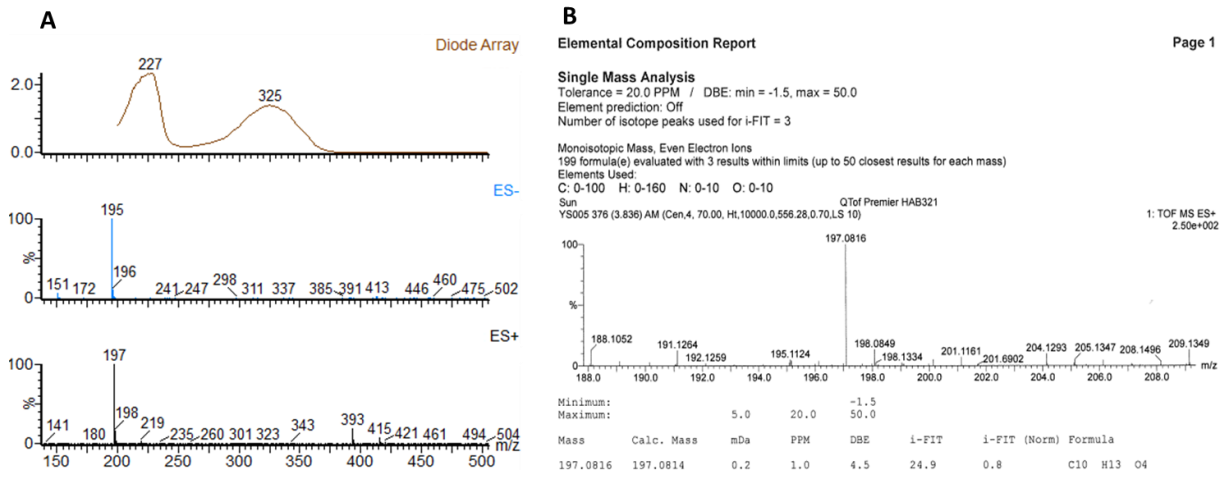


Figure 4.3.11 A, UV-absorption (top) and fragmentation pattern of **157** in ES^- (middle) and ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **157.**

We speculated that the structure of **157** could potentially resemble that of **156** with an additional hydroxyl group based on mass information and UV absorption data. However, the precise location of this hydroxyl group within the backbone remains undetermined. So, we employed both 1D and 2D NMR to elucidate the absolute structure of **157**. The ^{1}H-NMR spectrum of **157** revealed the presence of two methyl group signals at δ_H 1.94 (3H, m) and δ_H 2.03 (3H, m, Table 4.3.6).

The ^{13}C and HSQC spectra revealed the presence of two methyl group signals at δ_C 9.1 and δ_C 9.5, as well as an additional hydroxymethyl signal at δ_H 4.29 (2H, d, J 2.9 Hz), corresponding to δ_C 62.7. In the HMBC spectrum, H-9 exhibited correlations with C-1, C-2, and C-3, while H-10 showed correlations with C-3, C-4, and C-5. Furthermore, H-8 displayed correlations with C-5, C-6, and C-7. Based on the HSQC data, it was apparent that there were only two protons attached to C-8. Thus, we deduced that C-8 must be a hydroxymethyl group. So, the structure of **157** was confirmed to have the hydroxyl group located at C-8, consistent with the backbone of **156**. However, this structure did not align with our initial proposal, which suggested a connection to the formation of multiforisin H **143h** or I **143i**. As a result, we speculated that this might be a metabolic shunt in *A. oryzae*. We searched for the mass of compound **157** in both ES^+ and ES^- modes in Exp1 by comparing it with the standard compound **157** purified from Exp2. We observed the presence of compound **157** in Exp1, albeit in a small amount (Figure 4.3.12). So, this P450 (MfnL2) does not act on C-9 or C-10, which probably can act in the later stage.

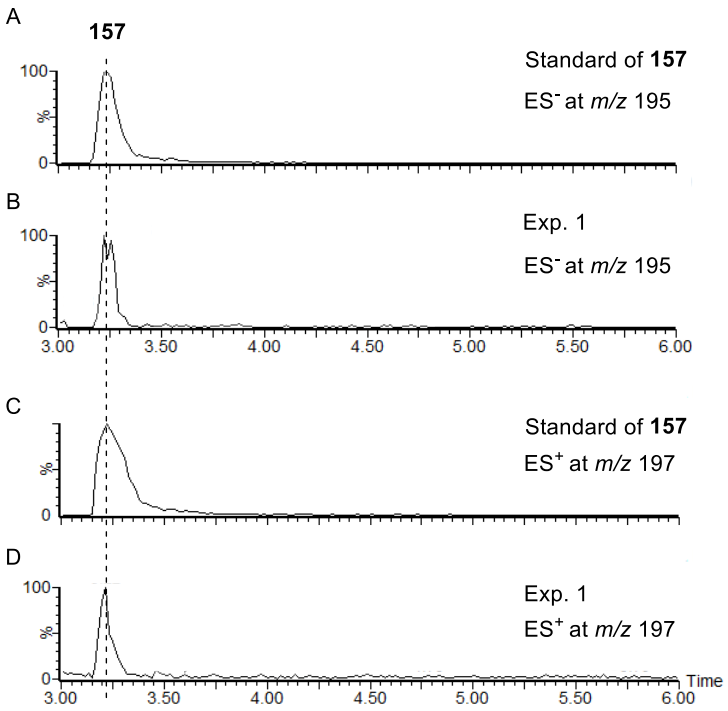
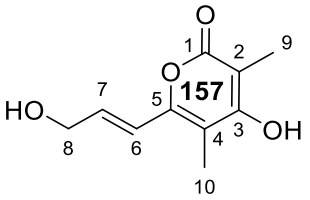


Figure 4.3.12 A, the mass of standard **157** scanned by ES- at *m/z* 195; **B**, the mass of **157** from exp.1 scanned by ES- at *m/z* 195; **C**, the mass of standard **157** scanned by ES+ at *m/z* 197; **D**, the mass of **157** from exp.1 scanned by ES+ at *m/z* 197.

We then expressed another P450 (MfnR3) which is the Exp3 (Figure 4.3.10). From the peak retention time in the chromatogram and mass information, the new peak is identical with **157**. Exp3 also told that P450 (MfnR3) could not react with **156**, which probably act in the later stage.



Chemical Formula: $C_{10}H_{12}O_4$
Exact Mass: 196.0736

Table 4.3.6 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **157** recorded in CD_3OD.

Compound 157				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	168.1			
2	100.3			
3	167.5			
4	110.1			
5	153.1			
6	118.8	6.65, 1H, m	7, 8	5, 7, 8
7	137.4	6.65, 1H, m	6, 8	5, 6, 8
8	62.7	4.29, 2H, d (2.9)	6, 7	5, 6, 7
9	9.1	1.94, 3H, s		1, 2, 3
10	9.5	2.03, 3H, s		3, 4, 5

Since the two P450 enzymes, MfnL2 and MfnR3, did not exhibit reactivity with substrate **156**, we hypothesized that they probably catalyse steps later in the pathway. To explore this further, we conducted Exp4 (Figure 4.3.10) by co-expressing *O*-MeT (MfnL1) and MfnPKS1 to investigate the potential occurrence of *O*-methylation in a later step.

Upon analyzing the crude extract of the transformants involved in Exp4 using LCMS, we observed the emergence of a new peak **158**, along with a minor new peak **159**. For the fermentation process aimed at isolating peaks **158** (140 mg/L) and **159** (12 mg/L), we utilized a 1-liter DPY medium. Our proposition is that the major peak **158** corresponds to the product resulting from the reaction of *O*-MeT (MfnL1) and MfnPKS1. It is likely that the methyl group is added to the hydroxyl group at O-3, resulting in a mass of 194.

4.3.4.2 Characterization of 158

The mass spectrum of **158** revealed several peaks in positive ion mode (ES^+), specifically at *m/z* 195, 217, 389, and 412, while the negative ion mode (ES^-) did not exhibit significant signals (Figure 4.3.13A). The *m/z* 195 fragment can be attributed to the $[M + H]^+$ ion, while the *m/z* 217 peak suggests the formation of $[M + Na]^+$ ion. Furthermore, the *m/z* 389 peak corresponds to $[2M + H]^+$ ion, and signals at *m/z* 412 align with $[2M + Na]^+$ ion (Figure 4.3.13A). These offer partial confirmation of the anticipated mass of *m/z* 194 for **158**. HRMS provides further support for the determination of compound **158**'s mass as *m/z* 194. This is evident from the closely matching values, with the observed $[M + H]^+$ ion at *m/z* 195.1019 closely aligning with the calculated $[M + H]^+$ ion at *m/z* 195.1021 (Figure 4.3.13B). Moreover, the UV absorption spectrum of **158** closely resembles that of compounds **156** and **157**, exhibiting absorption peaks at 226 and 333 nm (Figure 4.3.13A).

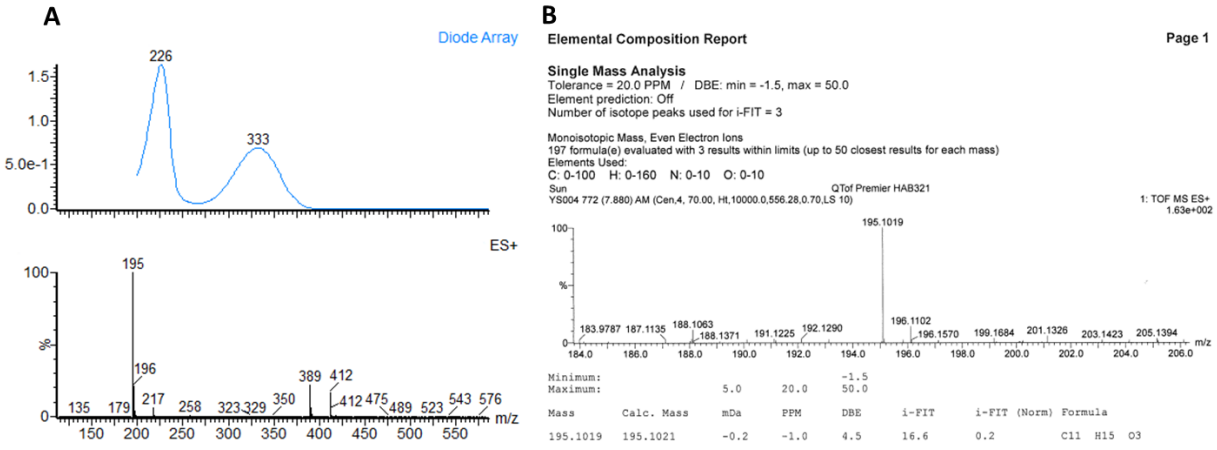
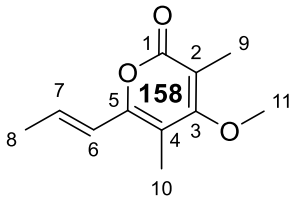


Figure 4.3.13 A, UV-absorption (top) and fragmentation pattern of **158** in ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **158.**

To determine the absolute structure of **158**, we employed ^1H, ^{13}C and 2D NMR. In the ^1H-NMR spectrum of **7**, we observed four distinct methyl group signals at δ_H 1.93 (3H, dd, *J* 7.0, 1.7 Hz), δ_H 2.01 (3H, s), δ_H 2.0 (3H, s), and δ_H 3.85 (3H, s). These proton signals correlated with carbon signals at δ_C 18.7, 10.3,

9.6, and 61.1, respectively, as determined through HSQC analysis (Table 4.3.7). We deduced that the proton signal at δ_H 3.85 (3H, s) and the carbon signal at δ_C 61.1 can be attributed to the *O*-methyl group introduced by *O*-MeT (MfnL1).

Other proton signals, such as those at δ_H 6.41 (1H, dddd, *J* 15.30, 1.71, 1.71, 1.69 Hz) and δ_H 6.61 (1H, dddd, *J* 15.41, 6.92, 6.92, 6.90), correspond to the two olefinic protons and are associated with carbon resonances at δ_C 121.3 and δ_C 134.5, respectively. The keto carbon C-1 resonates at δ_C 167.4, while the olefinic carbons within the ring include C-2 (δ_C 110.7), C-3 (δ_C 170.7), C-4 (δ_C 111.1), and C-5 (δ_C 154.0). Furthermore, HMBC analysis provided crucial evidence by revealing a correlation between H-11 and C-3 (Table 4.3.7), thereby confirming that **158** is indeed a methylated derivative of **156**.



Chemical Formula: $C_{11}H_{14}O_3$
Exact Mass: 194.0943

Table 4.3.7 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **158** recorded in CD_3OD, Literature data was measured in acetone-d6.

Compound 158						
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**	**δ_C / ppm literature** [211]	**δ_H / ppm (J/Hz) literature** [211]
1	167.4				164.5	
2	110.7				111.2	
3	170.7				168.7	
4	111.1				109.4	
5	154.0				153.0	
6	121.3	6.41, 1H, dddd (15.30, 1.71, 1.71, 1.69)	7, 8	5, 8	121.4	6.42, dq (15.4, 1.3)
7	134.5	6.61, 1H, dddd (15.41, 6.92, 6.92, 6.90)	6, 8	5, 8	133.1	6.51, dq (15.4, 6.5)
8	18.7	1.93, 3H, dd (7.0, 1.7)	7, 6	6, 7	18.6	1.91, d (6.5)
9	10.3	2.01, 3H, s		1, 2, 3	10.4	1.96, s
10	9.6	2.0, 3H, s		3, 4, 5	9.5	1.98, s
11	61.1	3.85, 3H, s		3	60.7	3.83, s

4.3.4.3 Characterization of 159

A minor product, peak **159**, was also isolated and purified (Figure 4.3.10). The retention time of **159** is earlier than that of both **158** and **156** indicating that it is more polar (Figure 4.3.10). Mass spectrometry data for **159** indicated peaks at *m/z* 225, 247, and 449 in the positive ion mode (ES^+), and *m/z* 223 and 447 in the negative ion mode (ES^-, Figure 4.3.14A). These mass patterns suggest that *m/z* 225 likely corresponds to the $[M + H]^+$ ion, *m/z* 247 may represent the $[M + Na]^+$ ion, and *m/z* 449 could be indicative of the $[2M + H]^+$ ion (Figure 4.3.14A). Correspondingly, the $[M - H]^-$ ion at *m/z* 223 and the

[2M - H]$^-$ ion at *m/z* 447 provide some evidence to suggest that the mass of **159** is approximately *m/z* 224 (Figure 4.3.14A).

To further corroborate these findings, **159** underwent HRMS analysis in the ES$^-$ mode, yielding a mass of *m/z* 223.0600 (Figure 4.3.14B). The calculated [M - H]$^-$ mass was determined to be *m/z* 223.0606, and this closely matched the observed mass, providing additional confirmation. In terms of UV absorption characteristics, **159** exhibited similarities with **158**, displaying peaks at 237 and 342 nm (Figure 4.3.14A). These observations indicate that **159** shares a structural basis with **158**.

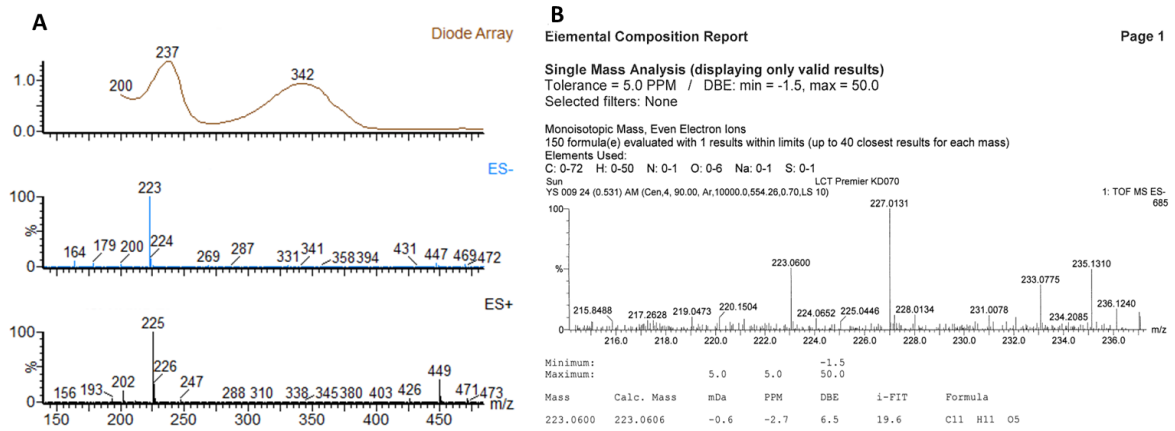
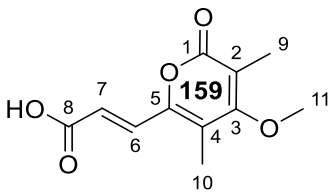


Figure 4.3.14 A, UV-absorption (top) and fragmentation pattern of **159** in ES$^-$ (middle) and ES$^+$ (bottom) by LR-LCMS; **B,** HRMS in ES$^-$ data for **159.**



Chemical Formula: $C_{11}H_{12}O_5$
Exact Mass: 224.0685

Table 4.3.8 Summarized NMR signals for ^{13}C, ^1H, ^1H-^1H COSY, HMBC for **159** recorded in CD_3OD.

Compound 159				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^1H-^1H COSY**	**HMBC (H-C)**
1	166.1			
2	115.1			
3	169.2			
4	118.1			
5	151.1			
6	124.6	6.55, 1H, d (15.3)	7	5
7	132.1	7.52, 1H, d (15.4)	6	5, 6, 8
8	169.6			
9	10.8	2.07, 3H, s		1, 2, 3
10	10.1	2.13, 3H, s		3, 4, 5
11	61.4	3.9, 3H, s		3

The ^{1}H NMR spectrum of **159** reveals two olefinic protons: one at δ_H 6.55 (1H, d, *J* 15.3 Hz) and the other at δ_H 7.52 (1H, d, *J* 15.4 Hz), corresponding to carbon resonances at δ_C 124.6 and 132.1, as indicated by HSQC analysis. Additionally, three distinct methyl signals at δ_H 2.07 (3H, s), δ_H 2.13 (3H, s), and δ_H 3.9 (3H, s) were observed, and these are correlated with carbon resonances at δ_C 10.8, δ_C 10.1, and δ_C 61.4, respectively (Table 4.3.8). Notably, the proton at δ_H 3.9 (3H, s) is likely attributed to the *O*-methyl group, which is analogous to the one found in **158**. The HMBC correlations from H-9 to C-1, C-2, and C-3, as well as from H-10 to C-3, C-4, and C-5, in conjunction with the HSQC data, strongly suggest the presence of two methyl groups at C-9 and C-10 (Table 4.3.8).

Furthermore, the carbonyl carbon C-1 resonates at δ_C 166.1, while the carbons involved in the ring's olefinic structure are labeled as follows: C-2 (δ_C 115.1), C-3 (δ_C 169.2), C-4 (δ_C 118.1), and C-5 (δ_C 151.1, Table 4.3.8). HSQC analysis did not reveal the presence of a proton at C-8, and the carbon shift at C-8 is observed at δ_C 169.6, suggesting that it likely corresponds to a carboxylic acid carbon (Table 4.3.8). The mass spectrometry data, with a mass of *m/z* 224, further supports the presence of a carboxylic acid group at C-8. Consequently, the structural configuration of **159** closely resembles that of **158**, with the additional incorporation of a carboxylic acid group at C-8.

4.3.5 Expression of *mfnL2*, *mfnR3* with *mfnPKS1* + *mfnL1* Respectively

Based on the findings from experiments 2, 3, and 4 (section 4.3.4) two P450 oxygenases (MfnL2, MfnR3) do not initiate the first stage of the pathway. Instead, the first reaction appears to be catalyzed by *O*-MeT, leading to the formation of **158**, which likely serves as the precursor for the final products. Subsequently, we conducted an additional set of three heterologous experiments, namely Exp5, Exp6, and Exp7, to further investigate this process.

In all three experiments, we included MfnPKS and MfnL1, both of which were confirmed to yield **158** as a substrate. To begin, Exp5 involved the introduction of *mfnL2* for co-expression with *mfnPKS1* and *mfnL1* (Figure 4.3.15). Following the transformation of *A. oryzae* NSAR1, fermentation, and subsequent chemical extraction, LCMS analysis revealed the presence of a new peak **160**, when compared to the untransformed strain (Figure 4.3.15).

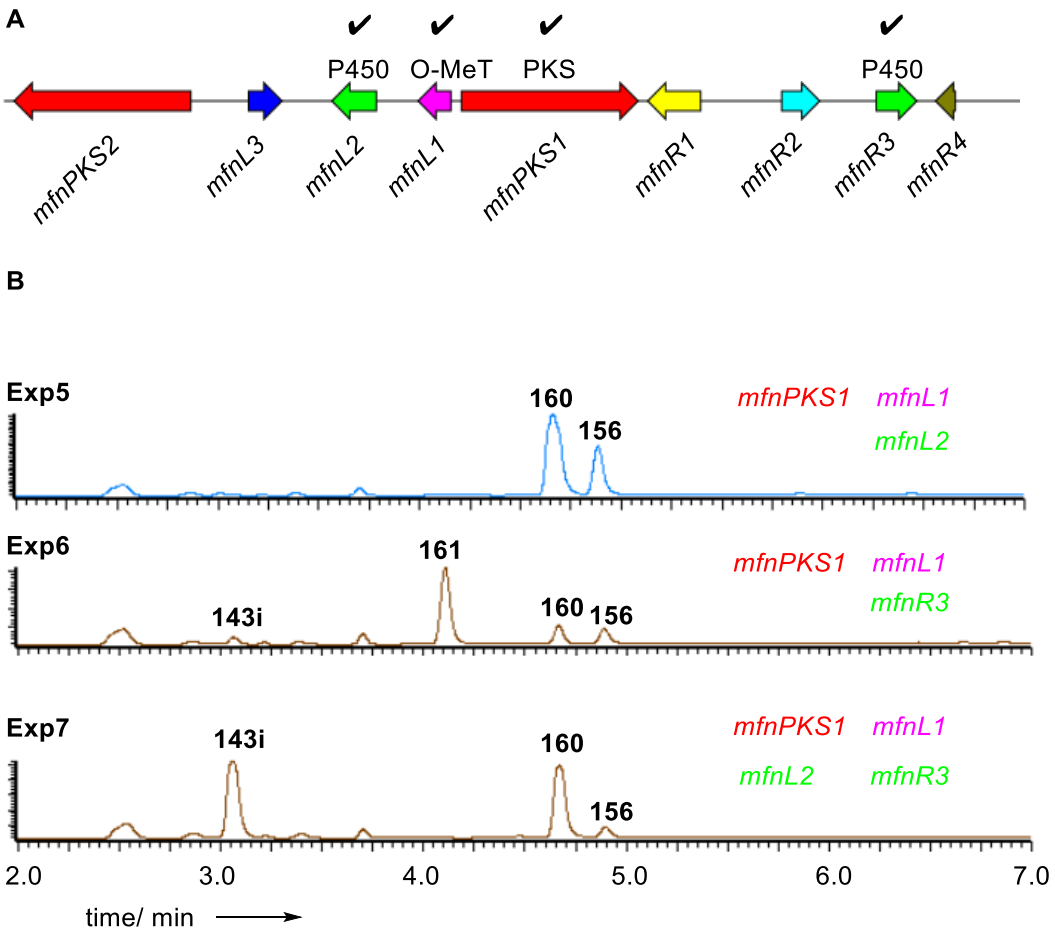


Figure 4.3.15 A, mfnBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of Exp5, Exp6, Exp7.

4.3.5.1 Characterization of 160

The emergence of the new peak **160**, has earlier retention time than peak **156**. The UV-absorption spectrum of this peak exhibited similarities with **158**, displaying two prominent absorption peaks at 226 and 334 nm. This similarity suggests that **160** shares the same pyrone skeleton as **158**. As anticipated, we hypothesized that the product of P450 (either MfnL2 or MfnR3) involves the addition of a hydroxyl group to either C-9 or C-10 of the **158**, resulting in a mass of approximately 210.

To verify this, we examined the mass fragment patterns, which revealed a peak at *m/z* 193, along with 211, 233, and 443 in the ES^+ ion mode (Figure 4.3.16A). It is plausible that *m/z* 211 represents the $[M + H]^+$ ion, while *m/z* 193 likely results from the loss of a water molecule from *m/z* 211. Similarly, *m/z* 233 corresponds to $[M + Na]^+$, and *m/z* 443 suggests $[2M + Na]^+$. The ES^- ion mode did not yield observable spectra. Consequently, we deduced that the mass of **160** is 210, aligning with our initial expectations. The precise mass of the anticipated new compound is 210.0892, while the exact mass of $[M + Na]^+$ is 233.0790. When we conducted mass spectrometry *via* HRMS, we observed a very close mass of 233.0792. This robustly supports the assertion that the mass of **160** is indeed 210.0892.

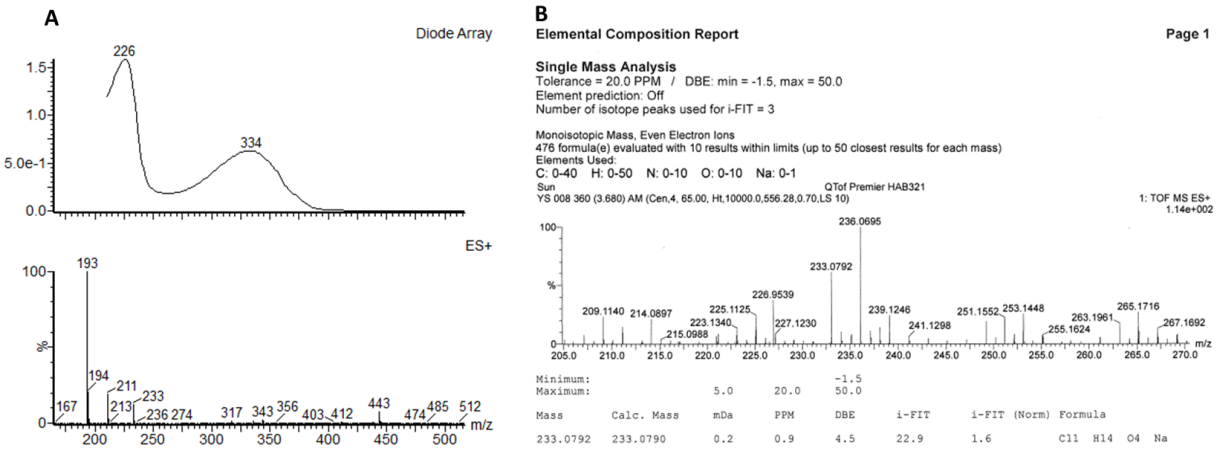
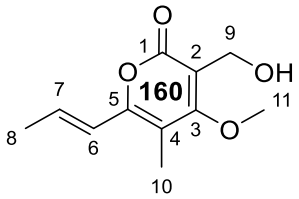


Figure 4.3.16 A, UV-absorption (top) and fragmentation pattern of **160** ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **160.**

We proceeded to conduct ^1H, ^{13}C and 2D NMR analyses of **160**, revealing two olefinic protons at δ_H 6.45 (1H, dddd, *J* 15.3, 1.7, 1.7, 1.7 Hz) and δ_H 6.67 (1H, dddd, *J* 15.3, 6.9, 6.9, 6.9 Hz), corresponding to protons situated at C-6 (δ_C 121.4) and C-7 (δ_C 135.9), respectively. The carbon resonances of the pyrone ring were identified as follows: C-1 (δ_C 166.9), C-2 (δ_C 111.5), C-3 (δ_C 172.2), C-4 (δ_C 110.5), and C-5 (δ_C 111.5).

Further analysis of the ^1H NMR and HSQC data unveiled three distinct proton signals related to three methyl groups, specifically δ_H 1.95 (3H, dd, *J* 7.48, 1.73 Hz), δ_H 2.0 (3H, s), and δ_H 4.07 (3H, s), which correlated with carbon resonances at δ_C 18.7, δ_C 9.6, and δ_C 62.4, respectively (Table 4.3.9). Beyond the *O*-methyl group, we identified the presence of another two methyl groups. The third methyl group had undergone oxidation, resulting in the formation of a hydroxyl group. This carbon was identified at δ_C 55.1, accompanied by two protons at δ_H 4.53 (2H, s, Table 4.3.9).

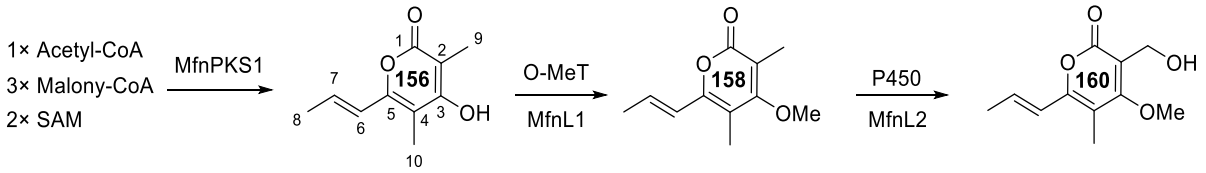
We employed HMBC analysis to pinpoint the position of the hydroxyl group, revealing correlations between the protons δ_H 4.53 (2H, s) and C-1, C-2, and C-3 (Table 4.3.9). This strongly supported the conclusion that the oxidation occurred at C-9. Importantly, C-8 and C-10 exhibited no alterations based on the evidence from HSQC and HMBC data, particularly evident in the *J* values between H-8 and H-7, as well as H-6 (Table 4.3.9). Consequently, we have successfully elucidated the structure of **160**, which represents **158** with an oxidation resulting in a C-9 alcohol group. This structural elucidation aligns with the expected mass of 210, providing robust confirmation.



Chemical Formula: $C_{11}H_{14}O_4$
Exact Mass: 210.0892

Table 4.3.9 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **160** recorded in CD_3OD.

Compound 160				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	166.9			
2	111.5			
3	172.2			
4	110.5			
5	155.6			
6	121.4	6.45, 1H, dddd (15.3, 1.7, 1.7, 1.7)	7, 8	5, 8
7	135.9	6.67, 1H, dddd (15.3, 6.9, 6.9, 6.9)	6, 8	5, 8
8	18.7	1.95, 3H, dd (7.48, 1.73)	6, 7	6, 7
9	55.1	4.53, 2H, s		1, 2, 3
10	9.6	2.0, 3H, s		3, 4, 5
11	62.4	4.07, 3H,s		3



Scheme 4.3.2 A scheme including early steps involved PKS, *O*-MeT and P450.

Based on the findings from experiment 5, we concluded that the P450 enzyme encoded by *mfnL2* is responsible for catalyzing the oxidation of C-9, resulting in the formation of a hydroxyl group. At this juncture, our understanding of the biosynthetic pathway is as follows: MfnPKS1 is responsible for the generation of the tetroketide pyrone, the *O*-MeT enzyme functions at the early stage to produce **158**, and one of the P450 enzymes (MfnL2) acts on **158** to introduce a hydroxyl group at C-9, yielding **160** (Scheme 4.3.2).

4.3.5.2 Characterization of 161

We proceeded to conduct experiment 6, introducing another P450 enzyme (MfnR3) in the presence of MfnPKS1 and MfnL1. Following the transformation, fermentation, chemical extraction, and subsequent LCMS analysis, we observed the presence of isolated intermediates **156** and **160**, albeit in relatively small quantities. A new peak appearing at 4.1 minutes, exhibited a spectrum at *m/z* 211, accompanied by signals at *m/z* 233 and *m/z* 443 in the ES^+ ion mode (Figure 4.3.17A). These mass fragments closely resembled those of **160**, with *m/z* 211 representing the $[M + H]^+$ ion, *m/z* 233 representing the $[M + Na]^+$ ion, and *m/z* 443 representing the $[2M + Na]^+$ ion. As a result, we inferred that **161** possesses a hydroxymethyl group at C-4 with a mass of 210, implying that the MfnR3 is responsible for oxidizing C-10. The exact calculated mass of **161** is 210.0892, and HRMS yielded an observed mass of 211.0974 in the ES^+ ion mode (Figure 4.3.17B). This observed mass is close to the calculated value of 211.0970.

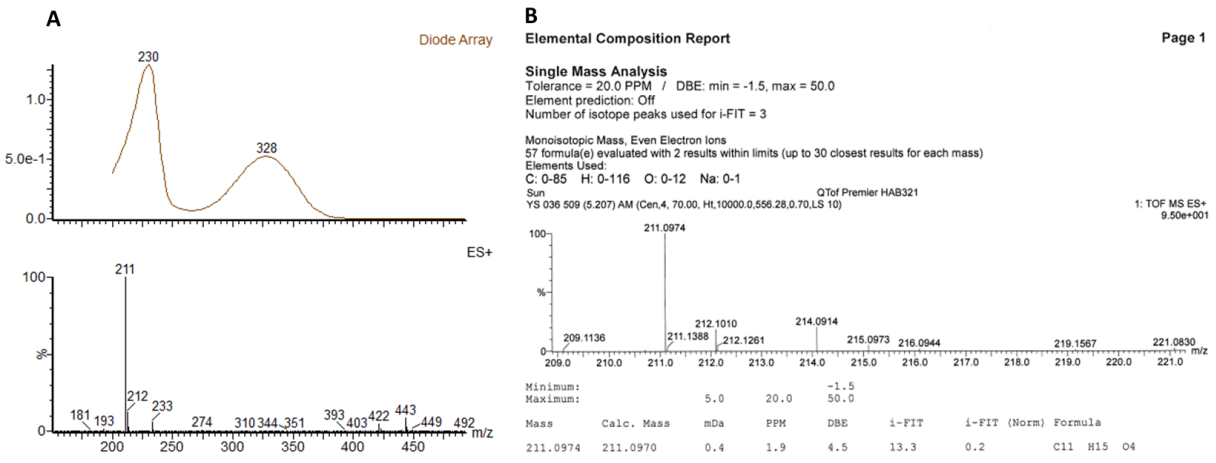
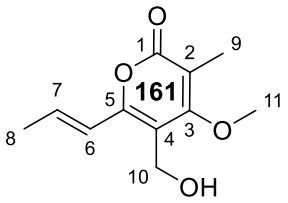


Figure 4.3.17 A, UV-absorption (top) and fragmentation pattern of **161** ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **161.**

We performed a 1-liter DPY fermentation and isolated 31 mg of **161**. We then conducted 1D and 2D NMR analysis of **161**, revealing two olefinic protons at δ_H 6.56 (1H, dddd, *J* 15.3, 1.6, 1.6, 1.6 Hz) and δ_H 6.71 (1H, dddd, *J* 15.3, 6.8, 6.8, 6.8 Hz). These protons can be assigned to C-6 (δ_C 121.2) and C-7 (δ_C 136.1), respectively. Moreover, the carbon resonances of the pyrone ring were also identified: C-1 (δ_C 167.1), C-2 (δ_C 111.0), C-3 (δ_C 170.3), C-4 (δ_C 114.2), and C-5 (δ_C 156.9).

Upon further scrutiny of the ^{1}H NMR and HSQC data, we discerned three sets of proton signals associated with three methyl groups, at δ_H 1.95 (3H, dd, *J* 6.8, 1.6 Hz), δ_H 2.03 (3H, s), and δ_H 4.46 (3H, s). These proton signals correlated with carbon resonances at δ_C 18.8, δ_C 10.4, and δ_C 62.1, respectively (Table 4.3.10). The *O*-methyl group most likely corresponds to the carbon at δ_C 62.1. Interestingly, one of the methyl groups underwent an oxidation, leading to the formation of a hydroxyl group (Table 4.3.10).

To precisely locate the position of the hydroxyl group, we employed HMBC analysis, which unveiled correlations between the protons δ_H 4.46 (2H, s) and C-3, C-4, and C-5 (Table 4.3.10). These correlations provide compelling evidence that the oxidation occurred at C-10. Consequently, we have successfully deduced the structure of **161**, which represents a modified congener of **158** due to an oxidation event resulting in a C-10 alcohol group. This structural determination agrees with the expected mass of 210, thus providing the confirmation.



Chemical Formula: $C_{11}H_{14}O_4$
Exact Mass: 210.0892

Table 4.3.10 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **161** recorded in CD_3OD.

Compound 161				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	167.1			
2	111.0			
3	170.3			
4	114.2			
5	156.9			
6	121.2	6.56, 1H, dddd (15.3, 1.6, 1.6, 1.6)	7, 8	5, 7, 8
7	136.1	6.71, 1H, dddd (15.3, 6.8, 6.8, 6.8)		
8	18.8	1.95, 3H, dd (6.8, 1.6)	6, 7	5, 6, 7
9	10.4	2.03, 3H, s		1, 2, 3
10	54.5	4.46, 2H, s		3, 4, 5
11	62.1	3.94, 3H, s		3

4.3.5.3 Characterization of 143i

In experiment 6, we observed the presence of intermediate **160**, even though MfnL2 was not part of this experiment. This suggests that MfnR3 may have the capability to react with C-9 in some manner, leading to the formation of a small quantity of **160**. If this is indeed the case, it raises the possibility that MfnR3 can act on both C-9 and C-10, ultimately yielding a diol compound known as multiforisin I **143i**. This diol compound has a mass of 226, prompting us to search for peaks with an ion of 227 in ES^+ or 225 in ES^-.

We identified a peak at 3.05 minutes, exhibiting mass fragmentation patterns of *m/z* 209, 227, and 249 in ES^+ ion mode, as well as *m/z* 225 in ES^- ion mode. We assumed that *m/z* 227 corresponds to the $[M + H]^+$ ion, *m/z* 209 represents the $[M + H - H_2O]^+$ ion, *m/z* 249 signifies $[M + Na]^+$ ion, and *m/z* 225 is indicative of the $[M - H]^-$ ion. Furthermore, this compound exhibited similar UV-absorption characteristics to those of compounds **158** and **160**, featuring two maxima at 231 and 329 nm. Collectively, this information lends further support to our initial hypothesis, and this alignment is further substantiated by HRMS data. The observed $[M + H]^+$ mass is 227.0919, closely matching the calculated $[M + H]^+$ mass of 227.0933 (Figure 4.3.18B).

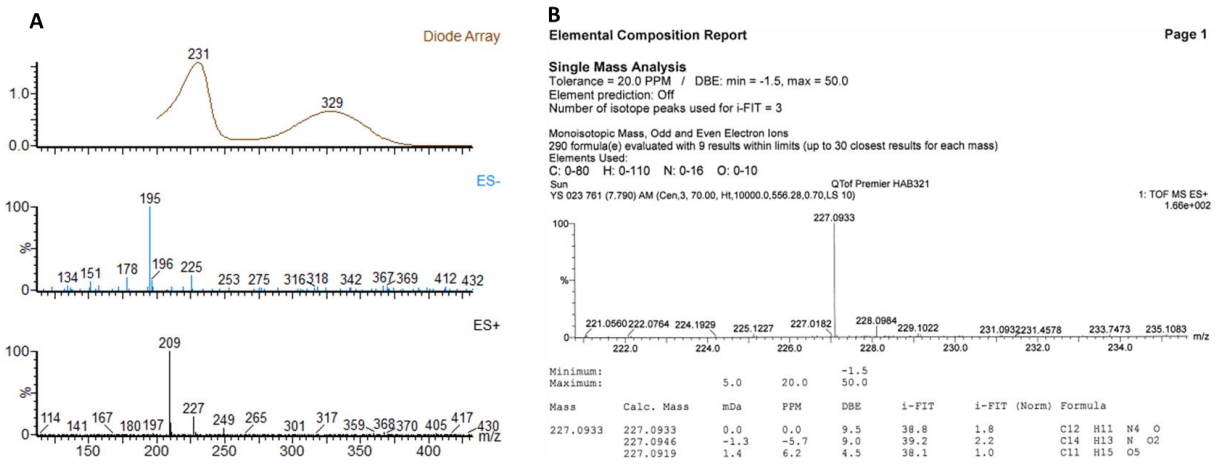
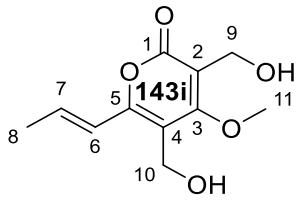


Figure 4.3.18 A, UV-absorption (top) and fragmentation pattern of **143i** ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^+ data for **143i.**

The final validation of the structure was achieved through comprehensive ^1H, ^{13}C, and 2D NMR analyses. In addition to the olefinic protons at δ_H 6.57 (1H, m) and δ_H 6.57 (1H, m), we detected two distinct methyl signals at δ_H 1.92 (3H, m) and δ_H 4.05 (3H, s). Furthermore, there were two hydroxymethyl signals at δ_H 4.33 (2H, d, *J* 5.1 Hz) and δ_H 4.29 (2H, d, *J* 4.8 Hz), indicating the presence of a diol structure (Table 4.3.11).

To precisely find the locations of these two hydroxyl groups, HMBC analysis revealed correlations between the protons δ_H 4.33 (2H, d, *J* 5.1 Hz) and the carbons C-1, C-2, and C-3, as well as δ_H 4.29 (2H, d, *J* 4.8 Hz) and the carbons C-3, C-4, and C-5 (Table 4.3.11). These correlations, ascertained through HSQC, provide evidence that the oxidations took place at C-9 and C-10. Consequently, we have successfully deduced the structure of **143i**, which represents a modified analogue of **158** due to the occurrence of two oxidation events, resulting in the introduction of C-9 and C-10 alcohol groups. This structural determination agrees with the expected mass of 226, thus providing unequivocal confirmation.



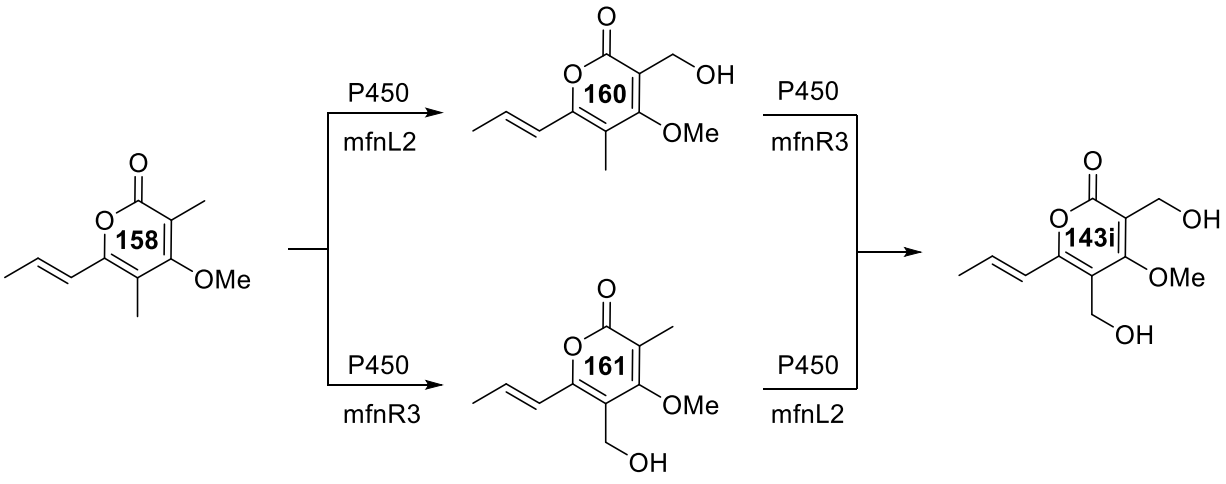
Chemical Formula: $C_{11}H_{14}O_5$
Exact Mass: 226.0841

Table 4.3.11 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **143i** recorded in DMSO-d6, Compound from literature [212] was measured in $CDCl_3$.

Compound 143i						
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**	**δ_C / ppm literature** [212]	**δ_H / ppm (J/Hz) literature** [212]
1	163.2				164.9	
2	111.0				110.8	
3	169.1				168.8	
4	113.0				111.8	
5	155.7				156.8	
6	120.6	6.57, 1H, m	7, 8	5, 8	119.3	6.42 (dq, 15.3, 1.8)
7	134.6	6.57, 1H, m	6, 8	5, 8	137.5	6.84 (dq, 15.3, 7.0)
8	18.4	1.92, 3H, m	6, 7	5, 6, 7	18.9	1.96(3H, dd, 7.0, 1.8)
9	53.2	4.33, 2H, d (5.1)		1, 2, 3	55.9	4.6 (2H, s)
10	52.6	4.29, 2H, d (4.8)		3, 4, 5	55.1	4.52 (2H, s)
11	62.2	4.05, 3H, s		3	63.1	4.12 (3H, s)

To investigate whether both P450 enzymes can work together in the presence of **158**, we co-expressed *mfnL2* and *mfnR3* along with *mfnPKS1* and *mfnL1* (Exp7, Figure 4.3.15). We observed the formation of intermediates **156** and **160**. The product **143i** exhibited higher peak intensities compared to Exp6, indicating that both MfnL2 and MfnR3 can interact with substrate **158** independently, leading to the formation of **143i** without specific order. The titer of **143i** in Exp7 was measured at 36 mg/L.

So, we concluded that one P450 enzyme (MfnL2) introduces a hydroxy group at C-9, yielding intermediate **160**, while another P450 enzyme (MfnR3) oxidizes C-10 to produce **143i**. Additionally, P450 enzyme MfnR3 can also oxidize C-10 to form intermediate **161**, followed by the subsequent addition of a hydroxy group at C-9 by P450 enzyme MfnL2 to generate **143i** (Scheme 4.3.3). These findings suggest that both MfnL2 and MfnR3 exhibit flexibility in accommodating various substrates.



Scheme 4.3.3 A scheme including the *O*-MeT and two P450 enzymes in the presence of **158.**

4.3.6 Expression of *mfnR2, mfnR4* in the Transformant from Exp7

In Exp7, we expressed four genes: *mfnPKS1*, *mfnL1*, *mfnL2*, and *mfnR3*. The resulting transformant exhibited the capability to produce multiforisin I **143i** in conjunction with intermediate **160**. We proceeded to investigate the activities of MfnR2 (FMO) and MfnR4 (SDR) by individually incorporating their genes into the *A. oryzae* strain that produces **143i** and **160** from Exp7. Our hypothesis was that FMO and SDR could potentially perform additional oxidization at the hydroxyl groups of C-9 and C-10, form C-9-keto or C-10-keto.

In Exp8, we utilized the pTYG-*ade* vector and inserted the *mfnR2* gene under the control of the P_{eno}. The *mfnR2* gene was introduced into the strain previously engineered in Exp7 *via* a transformation process. Following fermentation, chemical extraction, and LCMS analysis, we identified the presence of intermediates **156**, **160**, and **143i** (Figure 4.3.19). No novel peaks were observed in this experiment, leading us to conclude that MfnR2 does not demonstrate any activity towards substrates **160** and **143i**.

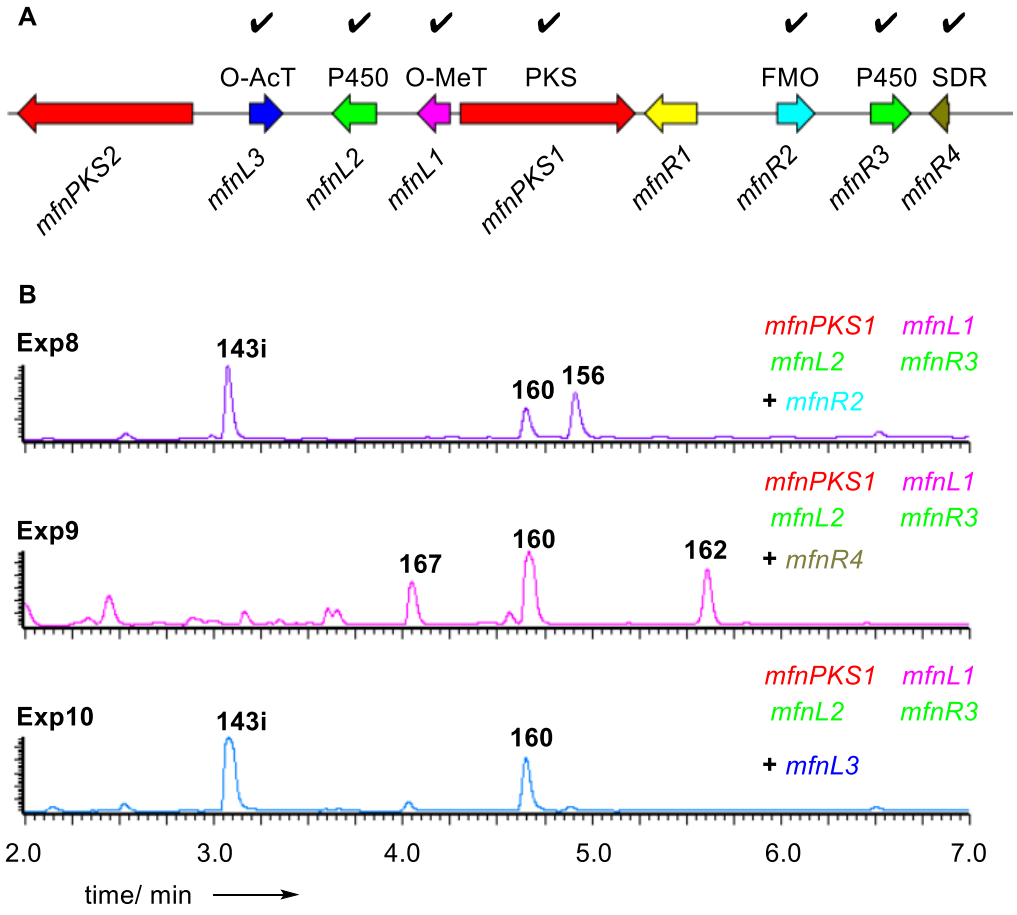


Figure 4.3.19 A, mfnBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of Exp 8, 9, 10.

We proceeded with Exp9, utilizing the pTYG-*ade* vector that contained the *mfnR4* gene, which was positioned under the control of the P_{gpdA}. We introduced this plasmid into the strain previously modified in Exp7. Following fermentation, chemical extraction, and LCMS analysis, we observed the emergence

of two new peaks: peak **162** at a retention time of approximately 5.7 minutes; and peak **167** at a retention time of about 3.1 minutes, alongside the presence of intermediate **160** (Figure 4.3.19).

We then scaled up our fermentation to a larger volume using 1-liter DPY medium and performed preparative LCMS to isolate **162** and **167**. These isolated fractions were subjected to LCMS analysis, followed by purity assessment using an ELSD detector. Finally, they were submitted for NMR analysis.

We hypothesized that the formation of the new products occurred *via* the oxidation of hydroxyl groups situated at positions C-9 or C-10. In the case of substrates **160** or **161**, the resulting product would contain an aldehyde group with an identical mass, approximately 208. On the other hand, if the substrate was **143i**, the product could have aldehyde groups at either C-9 or C-10, both having an expected mass of about 224. Therefore, our initial approach involved searching for peaks with mass fragmentation patterns corresponding to 208 and 224.

4.3.6.1 Characterization of 162

The mass spectrum of **162** exhibited some mass fragmentation patterns in both ES^+ and ES^- ion modes, including *m/z* 209, 231, 439 (ES^+), as well as *m/z* 207 and 523 (ES^-). We reasonably attributed *m/z* 209 to the $[M + H]^+$ ion, *m/z* 231 to the $[M + Na]^+$ ion, *m/z* 439 to the $[2M + H]^+$ ion, *m/z* 207 to the $[M - H]^-$ ion, and *m/z* 523 to, most likely, the $[2(M + 3H_2O) - H]^-$ ion (Figure 4.3.20). Additionally, this compound displayed notable UV-absorption characteristics, featuring two absorption maxima at 229 and 370 nm, which closely resembled those of **160**.

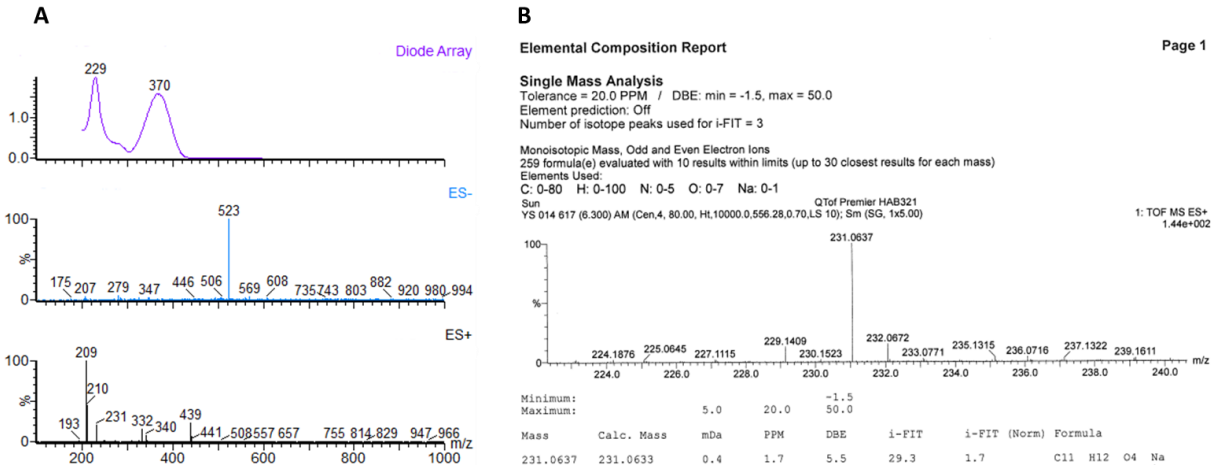
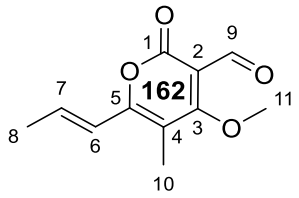


Figure 4.3.20 A, UV-absorption (top) and fragmentation pattern of **162** ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^+ data for **162.**

HRMS analysis revealed that the observed $[M + Na]^+$ mass was 231.0637, in close agreement with the calculated $[M + Na]^+$ mass of 231.0633, corresponding to the molecular formula $C_{11}H_{12}O_4$ (Figure 4.3.20B). Based on this information, we infer that the mass of **162** is likely 208, aligning with our earlier assumption of a structure featuring an aldehyde group at C-9 or C-10.

To elucidate the structure of **162**, we employed ^{1}H, ^{13}C and 2D-NMR analysis. We identified olefinic protons at δ_H 6.61 (1H, dddd, *J* 15.3, 1.5, 1.5, 1.5 Hz) and δ_H 6.72 (1H, m). Moreover, we observed three distinctive methyl signals: one at δ_H 4.02 (3H, s) attributed to the *O*-methyl group (δ_C 64.6), another at δ_H 1.95 (3H, m) corresponding to δ_C 18.7, and a third at δ_H 1.95 (3H, m) associated with δ_C 8.9 (Table 4.3.12). Furthermore, we identified a proton signal at δ_H 9.94 (1H, s) in the low-field region, indicating the presence of an aldehyde group.

To pinpoint the exact position of the aldehyde group, we conducted HMBC analysis, which revealed correlations between H-9 and C-2 (δ_C 104.9), as well as H-9 and C-3 (δ_C 175.1). These correlations provided conclusive evidence that the aldehyde group is situated at C-9, as further supported by its presence at δ_C 187.5, as determined by HSQC analysis (Table 4.3.12).



Chemical Formula: $C_{11}H_{12}O_4$
Exact Mass: 208.0736

Table 4.3.12 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **162** recorded in DMSO-d6.

Compound 162				
Pos.	**δ_C / ppm**	**δ_H / ppm (*J*/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	162.3			
2	104.9			
3	175.1			
4	109.3			
5	157.8			
6	120.6	6.61, 1H, dddd (15.3, 1.5, 1.5, 1.5)	7, 8	5, 7, 8
7	138.7	6.72, 1H, m	6, 8	5, 6, 8
8	18.7	1.95, 3H, m	6, 7	6, 7
9	187.5	9.94, 1H, s		2, 3
10	8.9	1.95, 3H, m		3, 4, 5
11	64.6	4.02, 3H, s		3

4.3.6.2 Characterization of 167

Our subsequent focus was on determining the structure of **167**. Initially, the mass spectrum of **167** exhibited distinct mass fragmentation patterns in both ES^+ ion mode, including *m/z* values of 193, 285, 303, 325, 605, and 627, and ES^- ion mode, featuring *m/z* 301 and 603. The presence of the *m/z* 193 pattern was consistent with intermediates such as **159**, **160**, **161**, and **162**, indicating it likely corresponds to a fragment of the pyrone backbone. Given the presence of *m/z* 301 and 603 in ES^- ion mode, we hypothesized that the molecular weight of **167** is *m/z* 302. This assumption finds support in the mass fragmentation patterns observed in ES^+ ion mode: *m/z* 285 may represent $[M + H - H_2O]^+$, *m/z* 303

could signify $[M + H]^+$ ion, *m/z* 325 likely corresponds to $[M + Na]^+$ ion, *m/z* 605 appears to be $[2M + H]^+$ ion, and *m/z* 627 might be $[2M + Na]^+$ ion (Figure 4.3.21).

We focused on the $[M + H]^+$ *m/z* 303 value in ES^+ ion mode as our target for determining the precise mass by employing HRMS and calculating potential molecular formulas. We detected a mass of 303.0895, which led to the proposal of six candidate formulas, all exhibiting exact masses close to 303.0895 (Figure 4.3.21B). The formula that came closest, with a value of 303.0902, was identified as $C_{13}H_{19}O_6S$. Consequently, we infered that the likely formula for **167** is $C_{13}H_{18}O_6S$, possessing an exact mass of 302.0824.

In the UV spectrum, we observed absorption maxima (λ_{max}) at 207 and 301 nm for **167** (Figure 4.3.21A). These values are notably lower than those observed for intermediates such as **158** (226, 333 nm), **160** (226, 334 nm), **161** (230, 328 nm), and **162** (231, 329 nm). The decrease in λ_{max} is likely a result of a reduction in the number of conjugated π bonds, suggesting the possibility that one of the double bonds has been disrupted.

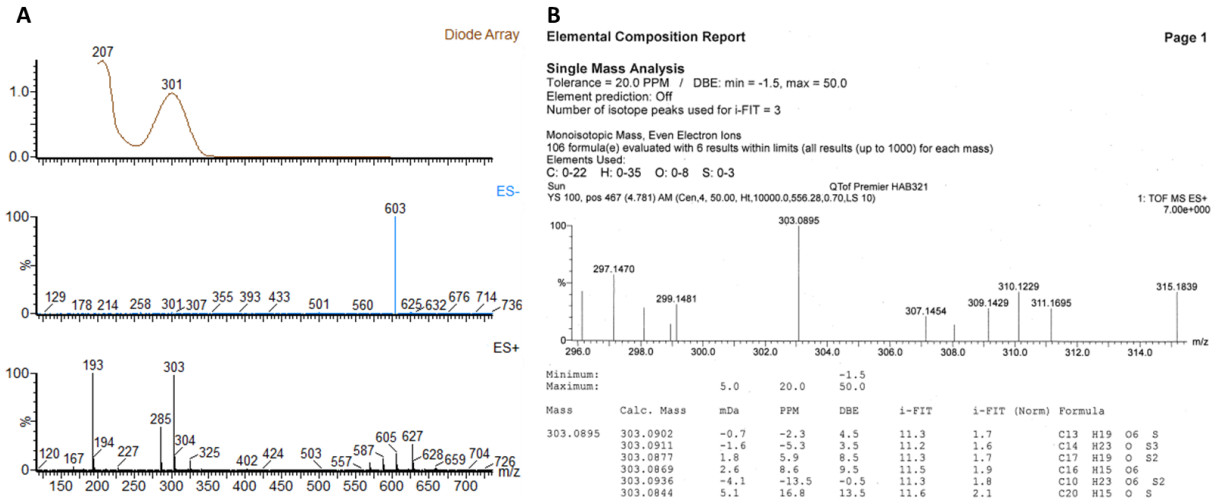
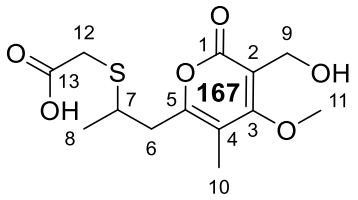


Figure 4.3.21 A, UV-absorption (top) and fragmentation pattern of **167** ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^+ data for **167.**

Subsequently, we carried out the analysis of **167** using ^1H, ^{13}C and 2D-NMR. In the ^{13}C and HSQC spectra, we identified the carbons forming the pyrone backbone, notably C-1 at δ_C 164.1, as well as the olefinic carbons, which include C-2 (δ_C 109.4), C-3 (δ_C 168.8), C-4 (δ_C 110.3), and C-5 (δ_C 157.1). From the ^1H-NMR and HSQC data, we observed the presence of an *O*-methyl group with a chemical shift of δ_H 4.04 (3H, s), indicating its connection to a carbon at δ_C 61.2. The hydroxymethyl group at C-9 (δ_C 53.3) was conclusively verified through HMBC correlations from H-9 (δ_H 4.34, 2H, s) to C-1, C-2, and C-3. Additionally, we observed proton signals at δ_H 1.89 (3H, s) corresponding to the methyl group at C-10 (δ_C 10.2), while signals at δ_H 1.24 (3H, d, *J* 6.8 Hz) were attributed to the methyl group at C-8 (δ_C 20.5).

Notably, when compared to the ^{1}H-NMR spectra of the intermediates like **158** or **160**, we noticed a substitution where two protons at a lower field were replaced by three protons at δ_H 2.68 (dd, *J* 14.5, 7.9 Hz), δ_H 2.84 (dd, *J* 14.5, 6.9 Hz), and δ_H 3.24 (ddd, *J* 7.9, 6.9, 6.9 Hz). Through HSQC, we verified that the protons at δ_H 2.68 and δ_H 2.84 were associated with a single carbon (δ_C 37.4), and HMBC correlations from δ_H 2.68 to C-4 and from δ_H 2.84 to C-4 indicated their connection to C-6. Furthermore, the proton at δ_H 3.24 was assigned to C-7 (δ_C 38.3, Table 4.3.13).

Interestingly, an HMBC correlation from H-7 to a previously unseen carbon, C-12 (δ_C 32.2), was observed. HSQC data indicated the presence of two protons (δ_H 3.31, 2H, s) at C-12. Additionally, an HMBC correlation from H-12 to C-7, along with another new carbon, C-13 (δ_C 171.6), was noted. C-13 is likely the carboxylic acid carbon. Therefore, based on the exact mass and the predicted formula $C_{13}H_{18}O_6S$, the structure of **167** is confirmed. This thioether shunt structure likely arises from the formation of **160** and mercaptoacetic acid.



Chemical Formula: $C_{13}H_{18}O_6S$
Exact Mass: 302.0824

Table 4.3.13 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **167** recorded in DMSO-d6.

Compound 167				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	164.1			
2	109.4			
3	168.8			
4	110.3			
5	157.1			
6	37.4	2.68, 1H, dd (14.5, 7.9) 2.84, 1H, dd (14.5, 6.9)	6, 7 6, 7	4, 5, 7, 8 4, 5, 7, 8
7	38.3	3.24, 1H, ddd (7.9, 6.9, 6.9)	6, 8	5, 6, 8, 12
8	20.5	1.24, 3H, d (6.8)	7, 8	6, 7
9	53.3	4.34, 2H, s		1, 2, 3
10	10.2	1.89, 3H, s		3, 4, 5
11	61.2	4.04, 3H, s		3
12	32.2	3.31, 2H, s		7, 13
13	171.6			

In Exp10, we introduced a plasmid containing the *mfnL3* gene, which was positioned under the control of the P_{adh}, into the strain previously modified during Exp7. The *mfnL3* gene was postulated to encode *O*-acetyltransferase, probably adding an acetate group to C-10 in the presence of substrate **143i**, thereby yielding multiforisin H **143h**. Following fermentation, chemical extraction, and LCMS analysis, we observed the peaks of intermediates **160** and **143i**. However, no new peaks were detected when compared to the untransformed strain (Figure 4.3.19).

4.3.7 Expression of *mfnPKS2* and *mfnL3* in the Transformant from Exp7

To elucidate the roles of *mfnPKS2* and *mfnL3* genes and to expand our repertoire of multiforisin derivatives, we constructed a plasmid, PL11. This plasmid contained *mfnPKS2*, responsible for encoding a highly reducing PKS, and *mfnL3*, which encodes *O*-methyltransferase (*O*-MeT), both under the control of the P_{amyB} and P_{adh}, respectively. We subsequently introduced the built plasmid into the strain capable of producing **143i**, as previously established in Exp7.

Following the isolation of positive transformants, we initiated fermentation. During the extraction of crude chemicals from submerged liquid cultures, we observed variations in the product profiles under different conditions. When we performed direct extraction without pH adjustment, maintaining a normally neutral pH due to the DPY medium, using EtOAc (Exp11a), we identified the emergence of two new peaks, specifically **143h** and **163**, upon subjecting the crude extract to analytical LCMS (Exp11a).

However, when we acidified the submerged liquid cultures to a pH of approximately 3 using 2 M HCl, and subsequently analyzed the crude extract *via* analytical LCMS, we detected the presence of four additional peaks, namely **143h**, **164**, **165**, and **166**.

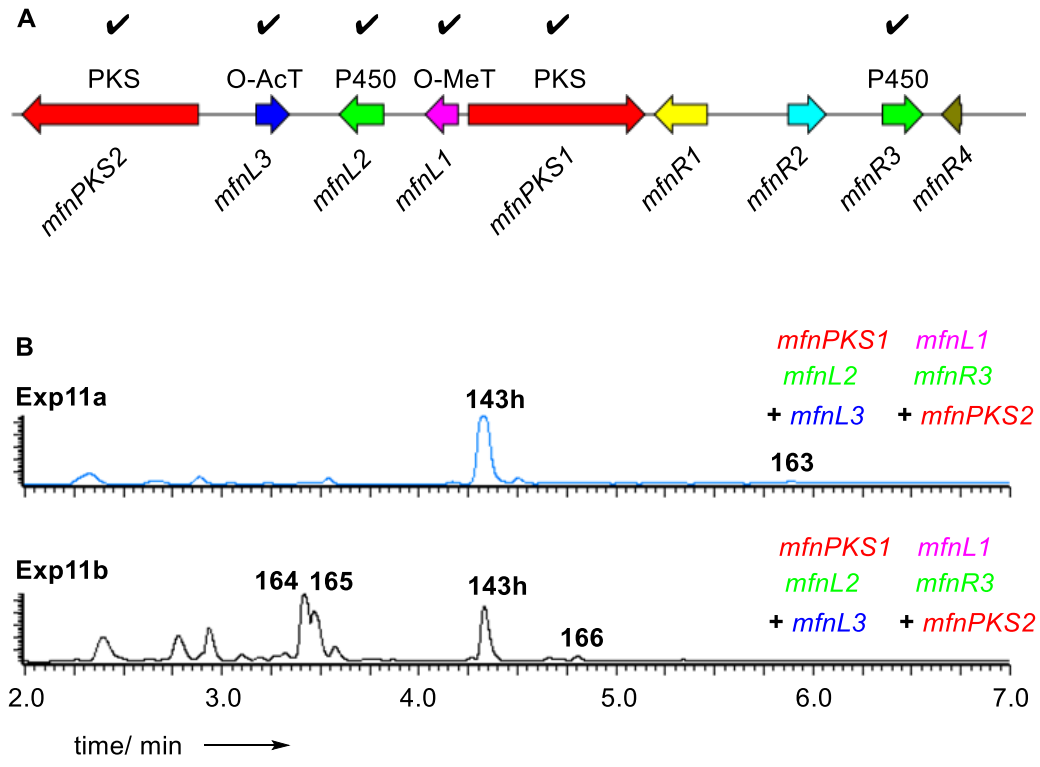


Figure 4.3.22 A, mfnBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of Exp11; Exp11a represents the DAD chromatogram of the crude extract processed from a neutral environment; Exp11b represents the DAD chromatogram of crude extract processed from an acidic environment (pH = ~3).

To elucidate the structures of these newly discovered compounds from Exp11a and Exp11b, we scaled up the fermentation process by using a 2-liter DPY medium. We then extracted one liter of liquid culture

using the same procedure as in Exp11a, and another liter of liquid culture using the same procedure as in Exp11b. Next, we conducted preparative LCMS to isolate these novel compounds. The isolated fractions were further subjected to LCMS analysis and assessed for purity using an ELSD detector. Finally, they were submitted for NMR analysis to determine their precise chemical structures.

4.3.7.1 Characterization of 143h

The mass spectrum of compound **143h** displayed various mass fragmentation patterns in both ES^+ and ES^- ion modes. In ES^+ mode, we observed peaks at *m/z* 251, 269, 291, and 560, while in ES^- mode, we identified a peak at *m/z* 313. We attributed *m/z* 251 to the $[M + H - H_2O]^+$ ion, *m/z* 269 to the $[M + H]^+$ ion, *m/z* 291 to the $[M + Na]^+$ ion, *m/z* 560 to the $[2M + Na]^-$ ion, and *m/z* 313 in the ES^- mode most likely corresponds to the $[M - H - HCOOH]^-$ ion (Figure 4.3.23A). Furthermore, this compound exhibited UV-absorption peaks at 231 and 327 nm, closely resembling those of compound **160**.

HRMS analysis confirmed that the observed $[M + H]^+$ mass was 269.1028, which closely matches the calculated $[M + H]^+$ mass of 269.1025 (Figure 4.3.23B). This result corresponds to the molecular formula $C_{13}H_{17}O_6$. Based on this information, we can reasonably conclude that the mass of compound **143h** is likely 268, consistent with the structure reported in the literature. [212]

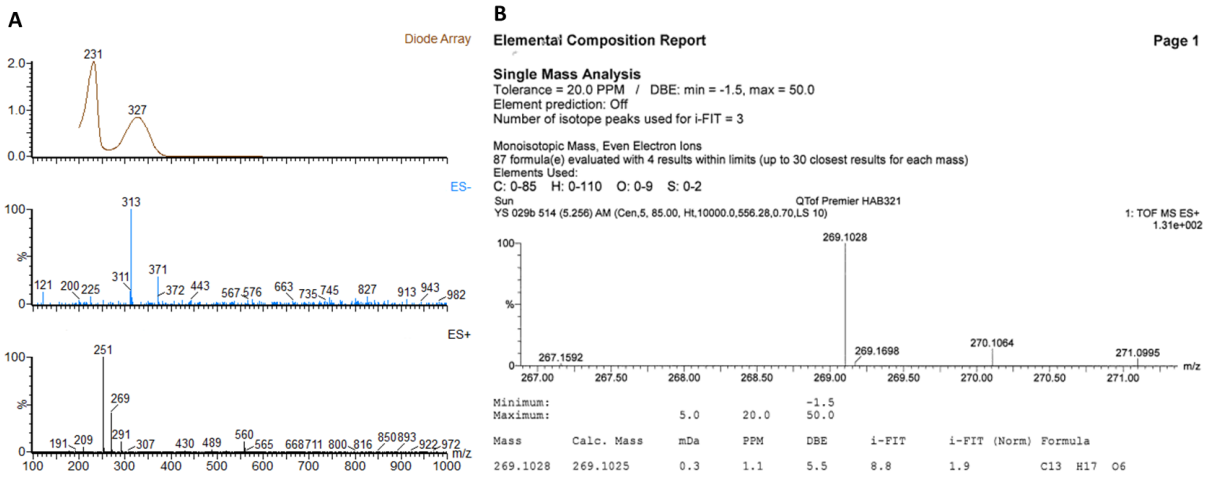
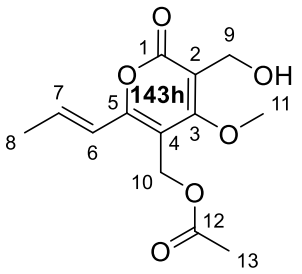


Figure 4.3.23 A, UV-absorption (top) and fragmentation pattern of **143h** ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^+ data for **143h.**

We also conducted NMR analyses, including ^1H, ^{13}C, and 2D-NMR, on compound **143h**. The chemical shifts of all the protons and carbons closely matched the values reported in the literature [212] (Table 4.3.14). Moreover, we utilized HMBC correlations to further elucidate the structure. Specifically, the HMBC correlation from H-10 (δ_H 4.97, 2H, s) to C-12 (δ_C 170.9) aided in locating the position of the acetyl group (Table 4.3.14). Additionally, the methyl group within the acetyl moiety was further corroborated through HMBC correlation from H-13 (δ_H 2.08, 3H, s) to C-12, which distinguished it from other methyl groups. Furthermore, the presence of the hydroxymethyl group at C-9 was validated by

HMBC correlations observed from H-9 (δ_H 4.59, 2H, s) to C-1 (δ_C 164.8), C-2 (δ_C 111.0), and C-3 (δ_C 168.9).



Chemical Formula: $C_{13}H_{16}O_6$
Exact Mass: 268.0947

Table 4.3.14 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **143h** recorded in $CDCl_3$.

Compound 143h						
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**	**δ_C / ppm literature [212]**	**δ_H / ppm (J/Hz) literature [212]**
1	164.8				164.7	
2	111.0				110.8	
3	168.9				168.8	
4	108.0				107.9	
5	158.1				158.0	
6	119.5	6.39, 1H, dddd (15.3, 1.7, 1.7, 1.7)	7, 8	5, 7, 8	119.3	6.39 (dq, 15.1, 1.7)
7	138.4	6.86, 1H, dddd (15.3, 7.0, 7.0, 7.0)	6, 8	5, 6, 8	138.2	6.86 (dq, 15.1, 7.0)
8	19.0	1.96, 3H, dd (7.0, 1.7)	6, 7	6, 7	18.9	1.97(3H, dd, 7.0, 1.7)
9	56.1	4.59, 2H, s		1, 2, 3	55.9	4.59, (2H, s)
10	56.3	4.97, 2H, s		3, 4, 5, 12	56.2	4.97, (2H, s)
11	63.3	4.06, 3H, s		3	63.2	4.07 (s)
12	170.9				170.8	
13	21.0	2.08, 3H, s		12	20.9	2.08, (3H, s)

4.3.7.2 Characterization of 163

While isolating compound **143h**, we also obtained a minor product **163**. **163** displayed some mass fragmentation patterns in the ES^+ ion mode, revealing peaks at *m/z* 253, 275, 505, and 527 (Figure 4.3.24A). We hypothesized that *m/z* 253 corresponds to the $[M + H]^+$ ion, *m/z* 275 to the $[M + Na]^+$ ion, *m/z* 505 to the $[2M + H]^+$ ion, and *m/z* 527 to the $[2M + Na]^+$ ion. Interestingly, this compound exhibited UV-absorption peaks at 227 and 327 nm, closely resembling the absorption pattern of **143h** (Figure 4.3.24A).

For precise mass determination and the calculation of potential molecular formulas, we selected the $[M + Na]^+$ ion as the target for HRMS analysis. The HRMS analysis yielded a mass of 275.0898, which closely matched the proposed formula of $C_{13}H_{16}O_5Na$ with an exact mass of 275.0895 (Figure 4.3.24B). Based on this information, we deduced that the likely formula for **163** is $C_{13}H_{16}O_5$, with an exact mass of 252.0998.

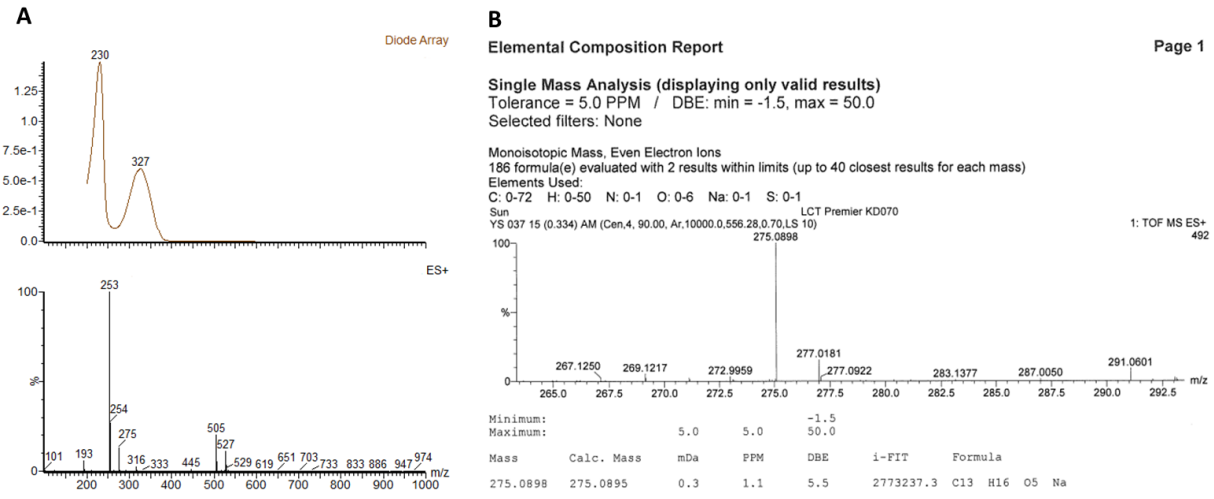
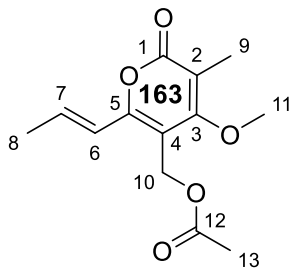


Figure 4.3.24 A, UV-absorption (top) and fragmentation pattern of **163** ES^+ (bottom) by LR-LCMS; **B,** HRMS in ES^+ data for **163.**

We conducted ^1H, ^{13}C, and 2D-NMR analysis of compound **163** (Table 4.3.15). When comparing the ^1H-NMR spectrum of **163** to that of **143h**, we observed many similarities, including the presence of an acetyl group at C-10 (δ_C 57.3), because there were still HMBC correlations from H-10 (δ_H 5.01, 2H, s) to the carbonyl C-12 (δ_C 172.4), as well as from H-13 (δ_H 2.05, 3H, s) to C-12. However, a discrepancy arose in the number of protons at C-9, where the ^1H-NMR and HSQC data indicated the presence of three protons (δ_H 2.04, 3H, s), as opposed to the two observed in **143h**. Therefore, based on this analysis, we confirmed that the structure of **163** closely resembles that of **143h** but lacks a hydroxyl group at C-9.



Chemical Formula: $C_{13}H_{16}O_5$
Exact Mass: 252.0998

Table 4.3.15 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **163** recorded in $CDCl_3$.

Compound 163				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	166.7			
2	110.9			
3	169.8			
4	110.0			
5	157.7			
6	120.9	6.54, 1H, dddd (15.3, 1.7, 1.7, 1.7)	7, 8	5, 7, 8
7	137.2	6.75, 1H, dddd (15.3, 7.0, 6.9, 6.9)	6, 8	5, 8
8	18.8	1.95, 3H, dd (7.0, 1.7)	6, 7	6, 7
9	10.6	2.04, 3H, s		1, 2, 3
10	57.3	5.01, 2H, s		3, 4, 5, 12
11	62.1	3.92, 3H, s		3
12	172.4			
13	20.7	2.05, 3H, s		12

When we isolated the compounds from Exp11, we acidified the submerged liquid cultures to a pH of approximately 3 using 2 M HCl. Subsequently, we analyzed the crude extract *via* analytical LCMS and detected the presence of four new peaks: **143h**, **164**, **165**, and **166**. Compounds **164**, **165**, and **166** are distinct from the compounds isolated from the crude extract without the acidification process. We purified compounds **164**, **165**, and **166** through preparative LCMS from a crude extract obtained from a 1-liter fermentation using DPY medium. We elucidated their structures through ^{1}H, ^{13}C, and 2D-NMR analysis.

4.3.7.3 Characterization of 164

Firstly, in the ES^+ ion mode, we observed specific mass fragmentation patterns for **163**, which manifested as peaks at *m/z* 313, 331, 354, and 643. Similarly, in the ES^- ion mode, we detected peaks at *m/z* 329 and 659 (Figure 4.3.25A). Our hypothesis is that *m/z* 313 corresponds to the ion $[M + H - H_2O]^+$, *m/z* 331 to the ion $[M + H]^+$, *m/z* 354 to the ion $[M + H + Na]^+$, and *m/z* 643 to the ion $[2M + H - H_2O]^+$. Furthermore, the peaks at *m/z* 329 and 659 are indicative of the ions $[M - H]^-$ and $[2M - H]^-$, respectively. Based on these observations, we infer that the molecular mass of **164** is approximately 330.

To achieve accurate mass determination and calculate potential molecular formulas, we opted to focus on the $[M - H]^-$ ion for HRMS analysis. The HRMS analysis returned a mass value of 329.0693, closely aligning with the suggested formula of $C_{14}H_{17}O_7S$, which has an exact mass of 329.0695 (Figure 4.3.25B). Based on these findings, we conclude that the probable formula for **164** is $C_{14}H_{18}O_7S$, with an exact mass of 330.0773. Interestingly, **164** displayed UV-absorption peaks at 226 and 332 nm (Figure 4.3.25A), aligning closely with the absorption profile observed in previously identified compounds, including **156, 157, 158, 159, 160, 161, 162, 163**. It reasonably suggests that **164** also possesses the pyrone tetroketide backbone. However, the mass of **164** is greater than that of **167**, with a mass of 302.

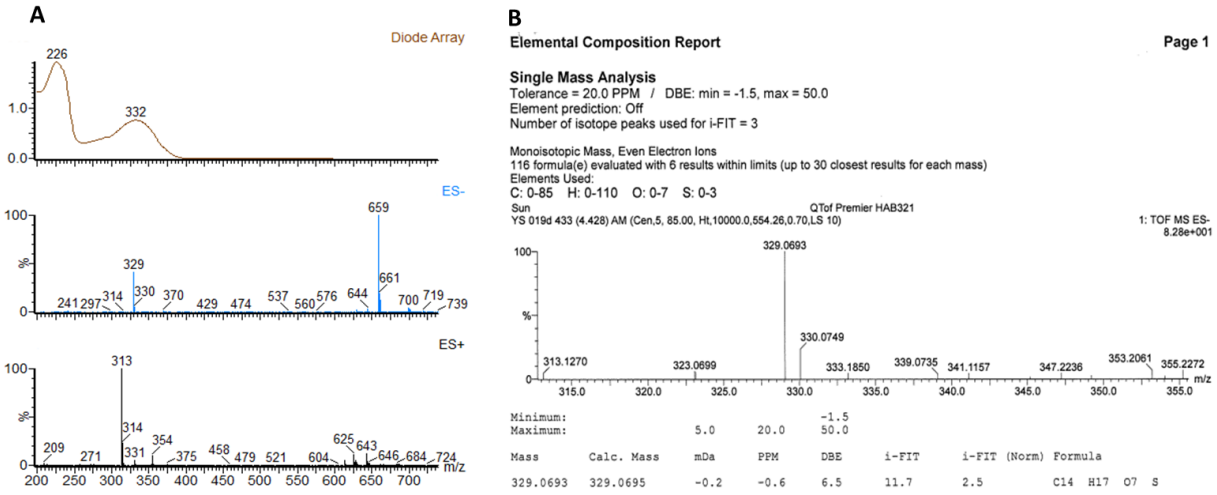
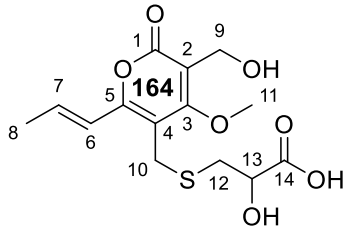


Figure 4.3.25 A, UV-absorption (top) and fragmentation pattern of **164** in ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^- data for **164.**

From the ^1H and ^{13}C NMR spectra, we observe the presence of two distinctive olefinic protons. One at δ_H 6.62 (1H, dddd, *J* 15.2, 1.7, 1.6, 1.6 Hz), corresponding to a proton positioned at C-6 (δ_C 121.5), and another at δ_H 6.73 (1H, dddd, *J* 15.3, 6.8, 6.8, 6.8 Hz), corresponding to a proton situated at C-7 (δ_C 137.1, Table 4.3.16). The carbon resonances associated with the pyrone ring were identified as follows: C-1 (δ_C 166.5), C-2 (δ_C 111.2), C-3 (δ_C 171.3), C-4 (δ_C 112.2), and C-5 (δ_C 156.8). Further examination of the ^1H-NMR and HSQC data reveals two distinct proton signals corresponding to two methyl groups, δ_H 1.97 (3H, dd, *J* = 6.9, 1.6 Hz) at the C-8 position (δ_C 18.8) and δ_H 4.21 (3H, s) at the C-11 position (δ_C 63.7, Table 4.3.16). This information essentially supports our previous proposal that **164** possesses a pyrone backbone.

The presence of a hydroxyl group was confirmed through HMBC correlations between H-9 and C-1, C-2, and C-3. Additionally, two protons with chemical shifts at δ_H 4.54 (2H, s) were observed at the C-9 position (δ_C 55.3) according to the HSQC spectrum. Consistently, HMBC correlations from H-10 were observed with C-3, C-4, and C-5. A new correlation was detected between H-10 and a previously unidentified carbon, C-12 (δ_C 26.8), accompanied by two protons at δ_H 3.77 (2H, s), as indicated by the HSQC data (Table 4.3.16). Furthermore, two additional HMBC correlations emerged from H-12 to C-13 and C-14, which correspond to two new carbon atoms at δ_C 72.4 and δ_C 176.2. The HSQC analysis revealed the presence of two protons at C-12 with chemical shifts of δ_H 2.86 (1H, dd, *J* 14.1, 6.4 Hz) and δ_H 3.0 (1H, dd, *J* 14.2, 4.1 Hz), along with one proton at C-13 with a chemical shift of δ_H 4.38 (1H, dd, *J* 6.5, 4.1 Hz). These findings suggest the possibility of a carboxylic acid group at the C-14 position.

The ^1H - ^1H COSY correlations observed between H-12 and H-13 provide evidence that C-13 is in closer proximity to C-12 compared to C-14. An HMBC correlation from H-13 to C-14 further supports this spatial arrangement. Therefore, we deduced the following sequential arrangement: C-10/C-12/C-13/C-14. When this information is combined with the HRMS data, it strongly suggests that the structure of **164** is derived from the quenching of **143i** by 3-mercaptolactate at the C-10 position.



Chemical Formula: $C_{14}H_{18}O_7S$
Exact Mass: 330.0773

Table 4.3.16 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **164** recorded in CD_3OD.

Compound 164				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	166.5			
2	111.2			
3	171.3			
4	112.2			
5	156.8			
6	121.5	6.62, 1H, dddd (15.2, 1.7, 1.6, 1.6)	7, 8	5, 7, 8
7	137.1	6.73, 1H, dddd (15.3, 6.8, 6.8, 6.8)	6, 8	5, 6, 8
8	18.8	1.97, 3H, dd (6.9, 1.6)	6, 7	5, 6, 7
9	55.3	4.54, 2H, s		1, 2, 3
10	26.8	3.77, 2H, s		3, 4, 5, 12
11	63.7	4.21, 3H, s		3
12	37.3	2.86, 1H, dd (14.1, 6.4) 3.0, 1H, dd (14.2, 4.1)	12, 13 12, 13	10, 13, 14 10, 13, 14
13	72.4	4.38, 1H, dd (6.5, 4.1)	12	12, 14
14	176.2			

4.3.7.4 Characterization of 165

We proceeded to elucidate the structure of **165**, which exhibited certain mass fragmentation patterns in both the ES^+ ion mode and ES^- ion mode. In the ES^+ ion mode, we observed peaks at *m/z* 251, 313, 331, 373, 391, 745, 763, in addition to *m/z* 329, 389, and 779 in the ES^- ion mode (Figure 4.3.26A). It is likely that the peak at *m/z* 251 corresponds to the mass of the pyrone backbone combined with the *O*-methyl and acetyl groups at C-10, akin to the structure of **163**. Due to the nearly identical retention times of **165** and **164**, we propose that the mass fragmentation patterns at *m/z* 313 and *m/z* 331 may come from **164**. We hypothesized that *m/z* 373 is indicative of the $[M + H - H_2O]^+$ ion, *m/z* 391 corresponds to the $[M + H]^+$ ion, *m/z* 745 relates to the $[2M + H - 2H_2O]^+$ ion, and *m/z* 763 signifies the $[2M + H - H_2O]^+$ ion. Additionally, we attributed the presence of *m/z* 329 to contamination from **164**, while *m/z* 389 represents the $[M - H]^-$ ion, and *m/z* 779 signifies the $[2M - H]^-$ ion. Consequently, we proposed that the mass of **165** is 390.

We chose the $[M - H]^-$ ion as our focus for HRMS analysis to achieve precise mass determination and calculate potential molecular formulas. The HRMS analysis resulted in a mass measurement of 389.0909, which closely aligned with the suggested formula of $C_{16}H_{21}O_9S$ with an exact mass of 389.0906 (Figure

4.3.26B). So, we inferred that the probable formula for **165** is $C_{16}H_{22}O_9S$, possessing an exact mass of 390.0985.

Interestingly, **165** displayed distinctive UV-absorption peaks at 214 and 294 nm, a pattern that notably differs from the absorption profiles of most characterized compounds but closely resembles that of **167** (λ_{max} 207 and 301 nm, Figure 4.3.21A, 4.3.26A). We can reasonably conjecture that **165** also possesses the pyrone tetroketide backbone, with the alteration likely occurring at either C-6 or C-7.

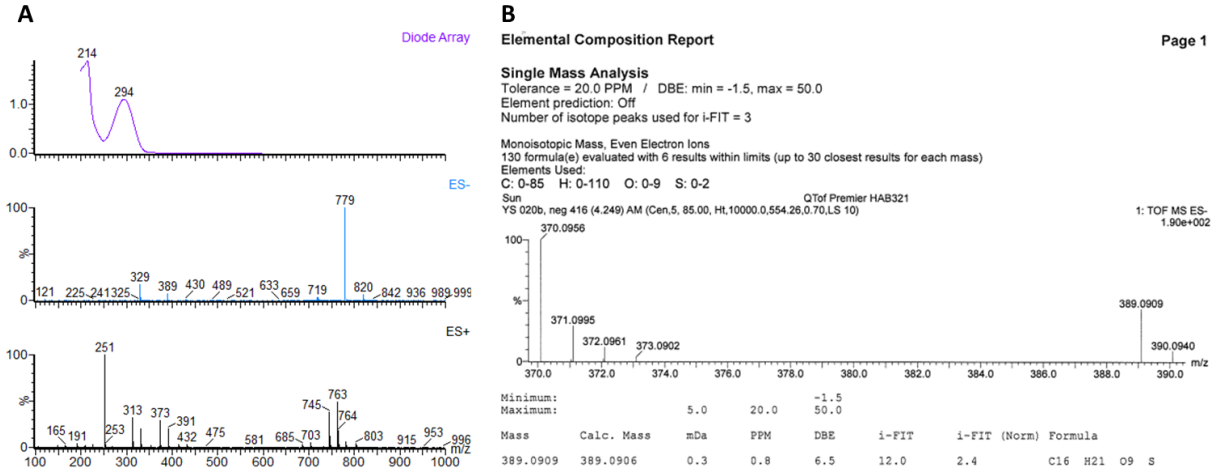
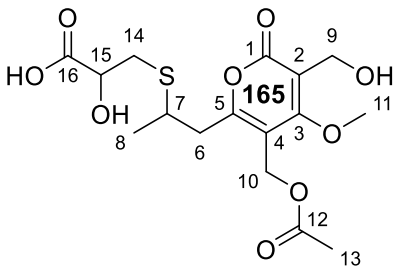


Figure 4.3.26 A, UV-absorption (top) and fragmentation pattern of **165** in ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^- data for **165.**

The comprehensive utilization of ^1H, ^{13}C and 2D NMR spectroscopy played a pivotal role in the elucidation of the structure of **165**. To begin with, similar to **167**, the two olefinic protons originally present were replaced by three protons observed at δ_H 2.83 (dd, *J* 14.7, 7.2 Hz), δ_H 3.01 (dd, *J* 14.7, 7.7 Hz), and δ_H 3.36 (1H, m). Subsequent HSQC analysis provided confirmation that the protons at δ_H 2.83 and δ_H 3.01 were associated with the carbon C-6 (δ_C 39.6), as substantiated by HMBC correlations linking δ_H 2.83 to C-4 and δ_H 3.01 to C-4 (Table 4.3.17). In addition, the proton at δ_H 3.36 was definitively attributed to C-7 (δ_C 40.3). The presence of an acetyl group (C-12 and C-13) at C-10 and a hydroxyl group at C-9 was further affirmed through a comparative examination of the ^1H- and ^{13}C- NMR spectra of **167** in comparison with those of **165**.

Furthermore, an HMBC correlation was observed from H-7 to C-14 (δ_C 36.1). The ^1H - ^1H COSY correlations between H-7 and H-14 (δ_H 2.85, m; δ_H 2.97, m) strongly suggest a four-bond coupling involving a sulfur atom positioned in the middle, reminiscent of **164**. H-14 exhibited a ^1H - ^1H COSY correlation with H-15 (δ_H 4.28, dd, *J* 6.7, 4.0 Hz) and HMBC correlations with C-15 (δ_C 72.2) and C-16 (δ_C 176.2). H-15 also displayed an HMBC correlation with C-16, where a carboxylic acid group is present (Table 4.3.17). Consequently, based on the exact mass and the predicted formula $C_{16}H_{22}O_9S$, the structure of **165** has been confirmed. This thioether shunt structure likely originates from the formation of **143h** and 3-mercaptolactate at C-7.



Chemical Formula: $C_{16}H_{22}O_9S$
Exact Mass: 390.0985

Table 4.3.17 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **165** recorded in CD_3OD.

Compound 165				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	166.8			
2	110.2			
3	170.8			
4	112.6			
5	164.2			
6	39.6	2.83, 1H, dd (14.7, 7.2) 3.01, 1H, dd (14.7, 7.7)	6, 7 6, 7	4, 5, 7 4, 5, 7
7	40.3	3.36, 1H, m	6, 8, 14	5, 6, 14
8	21.8	1.34, 3H, d (6.8)	7	7
9	55.2	4.56, 2H, s		1, 2, 3
10	58.1	4.99, 2H, d (1.8)	13	3, 4, 5, 12
11	63.2	4.19, 3H, s		3
12	172.4			
13	20.8	2.04, 3H, s		10, 12
14	36.1	2.85, 1H, m 2.97, 1H, m	7, 15 7, 15	7, 15, 16 7, 15, 16
15	72.2	4.28, 1H, dd (6.7, 4.0)	14	14, 16
16	176.2			

4.3.7.5 Characterization of 166

The structural attributes of **166** were also thoroughly investigated. In the ES^+ ion mode, we discerned distinctive mass fragmentation patterns, characterized by peaks at *m/z* 193, 315, 337, and 651, whereas in the ES^- ion mode, we identified peaks at *m/z* 313 and 627 (Figure 4.3.27A). The *m/z* 193 peak is likely attributed to the pyrone backbone, a recurring feature observed in prior structures such as **159** and **160**. We hypothesized that *m/z* 315 corresponds to the $[M + H]^+$ ion, *m/z* 337 to the $[M + Na]^+$ ion, and *m/z* 651 to the $[2M + Na]^+$ ion. Conversely, *m/z* 313 represents the $[M - H]^-$ ion, and *m/z* 627 signifies the $[2M - H]^-$ ion. So, we proposed that the molecular mass of **166** is 314.

In order to achieve precise mass determination and potential molecular formula calculation, we targeted the $[M - H]^-$ ion for the HRMS analysis. The HRMS analysis yielded a mass measurement of 313.0744, which closely aligned with the suggested formula of $C_{14}H_{17}O_6S$ with an exact mass of 313.0746 (Figure 4.3.27B). This information leads us to infer that the likely formula for **166** is $C_{14}H_{18}O_6S$, having an exact mass of 314.0824.

This compound exhibited UV absorption peaks at 228 and 340 nm (Figure 4.3.27A), closely mirroring the absorption profiles observed in well-characterized compounds, specifically compounds **156, 157, 158, 159, 160, 161, 143i, 143h**, and **163**. These findings lend strong support to the idea that **166** also incorporates the pyrone backbone.

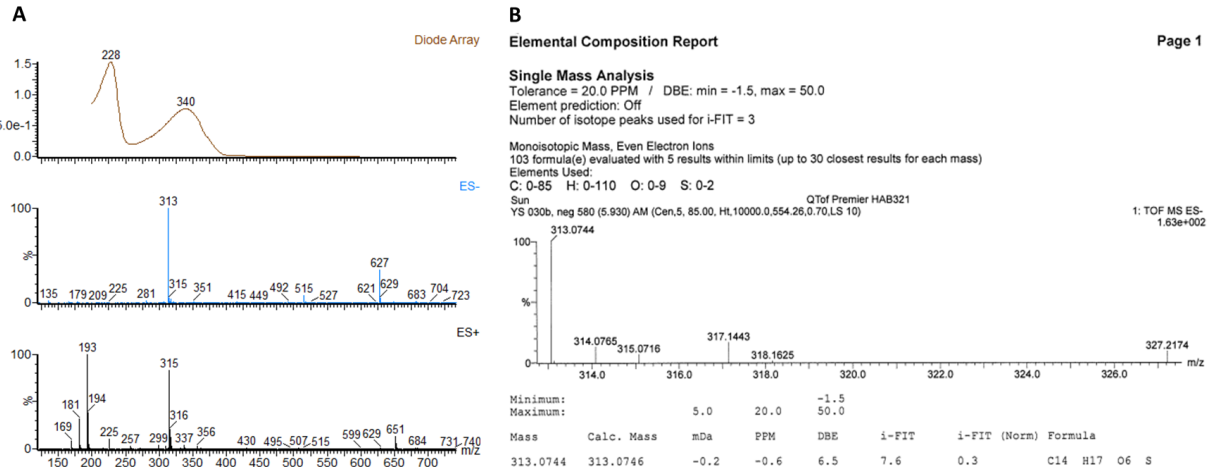
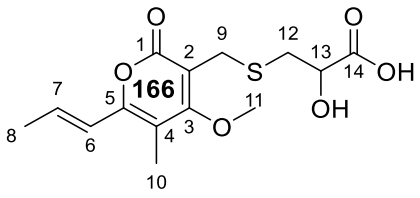


Figure 4.3.27 A, UV-absorption (top) and fragmentation pattern of **166** in ES^+ (bottom) and ES^- (middle) by LR-LCMS; **B,** HRMS in ES^- data for **166.**

Comprehensive ^1H, ^{13}C, and 2D NMR spectra provided clear information for the elucidation of the structure of **166**. The hydroxyl group or acetyl group at C-10 (δ_C 9.4) was not observed; instead, only a methyl group (δ_H 1.97, 3H, s) was present. There were HMBC correlations from H-9 (δ_H 3.61, 2H, s) to C-1 (δ_C 162.7), C-2 (δ_C 111.6), C-3 (δ_C 168.5), and a newly discovered carbon, C-12 (δ_C 36.6). The HSQC spectrum aided in identifying the protons (δ_H 2.77, 1H, dd, J 13.5, 6.8 Hz; δ_H 2.89, 1H, dd, J 13.5, 4.8 Hz) associated with C-12 (Table 4.3.18).

Furthermore, HMBC correlations were observed from H-12 to C-13 (δ_C 70.4) and C-14 (δ_C 174.1). A ^1H - ^1H COSY correlation was also found from H-12 to H-13 (δ_H 4.14, 1H, m). As a result, it can be concluded that C-12, C-13, and C-14 constitute a sub-structure identical to that found in **164** and **165**. Therefore, based on the exact mass and the predicted formula $C_{14}H_{18}O_6S$, the presence of a thioether group in the structure of **166** is confirmed. This thioether shunt structure likely originates from the formation of **160** and 3-mercaptolactate at C-9.



Chemical Formula: $C_{14}H_{18}O_6S$
Exact Mass: 314.0824

Table 4.3.18 Summarized NMR signals for ^{13}C, ^{1}H, ^{1}H-^{1}H COSY, HMBC for **166** recorded in CD_3OD.

Compound 166				
Pos.	**δ_C / ppm**	**δ_H / ppm (J/Hz)**	**^{1}H-^{1}H COSY**	**HMBC (H-C)**
1	162.7			
2	111.6			
3	168.5			
4	108.7			
5	153.1			
6	120.5	6.5, 1H, m	7, 8	5, 7, 8
7	133.8	6.5, 1H, m	6, 8	5, 6, 8
8	18.4	1.91, 3H, m	6, 7	6, 7
9	26.2	3.61, 2H, s		1, 2, 3, 12
10	9.4	1.97, 3H, s		3, 4, 5
11	61.2	3.87, 3H, s		3
12	36.6	2.77, 1H, dd (13.5, 6.8) 2.89, 1H, dd (13.5, 4.8)	12, 13 12, 13	9, 13, 14 9, 13, 14
13	70.4	4.14, 1H, m	12	12, 14
14	174.1			

4.3.8 Adding Genes from Islandic acid BGC

Heterologous expression experiments allowed us to both decipher the biosynthetic pathway to multiforison H **143h** and I **143i**, and to generate good quantities of the metabolites and their intermediates. Islandic acid **144** bears a close structural resemblance to multiforisin I **143i**, where a triketide molecule is attached at C-9, and C-8 is oxidised to a carboxylate. In an effort to manipulate the production of islandic acids **144**, we introduced genes from the *P. islandicum* ilaBGC into the strains established in our prior experiments.

First, in this Exp12, we amplified DNA fragments of ilaPKS2 without introns using the genomic DNA of *Penicillium islandicum* as the template. Subsequently, we inserted the sequenced *ilaPKS2* into the plasmid pTYGs-*met* under the regulation of P_{amyB}. This plasmid was then introduced into the strain capable of producing **143i**, as previously observed in Exp7. Some transformations were validated *via* PCR using extracted gDNA as templates to confirm the insertion of the gene into the gDNA (Figure 4.3.28A) Following this, a fermentation process was conducted. However, the analysis of the crude extract using LCMS did not reveal any new peaks (Figure 4.3.29).

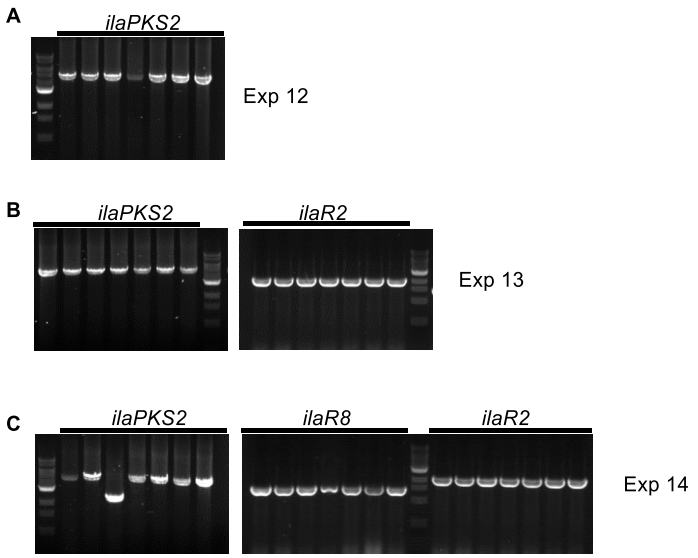


Figure 4.3.28 PCR examination for some transformants; **A**, primers PamyB-F/PiPKS2-R2 were used for Exp12; **B**, primers PamyB-F/PiPKS2-R2 and PgpdA-plugF/ PgpdA-plugR were used for Exp13; **C**, primers PamyB-F/PiPKS2-R2, PgpdA-plugF/ PgpdA-plugR and Padh-plugF/ Padh-plugR were used for Exp14.

We proceeded with Exp13, wherein we co-expressed *ilaPKS2* and *ilaR2* (encoding *O*-AcT) in the strain capable of producing **143h**, as generated in Exp11. Some transformations of Exp 13 were also validated *via* PCR using extracted gDNA as templates to confirm the insertion of the gene into the gDNA (Figure 4.3.28B). Fermentation was carried out, and the LCMS analysis of the crude extract did not reveal any new peaks (Figure 4.3.29). Additionally, Exp14 was conducted by introducing a predicted SDR, encoded by *ilaR8*, for co-expression with *ilaPKS2* and *ilaR2* in the strain that can produce **143i**, as generated in Exp7. Some transformations of Exp 14 were also validated *via* PCR using extracted gDNA as templates to confirm the insertion of the gene into the gDNA (Figure 4.3.28C). Fermentation for Exp14 was performed, and similarly, the LCMS analysis of the crude extract did not yield any new peaks (Figure 4.3.29).

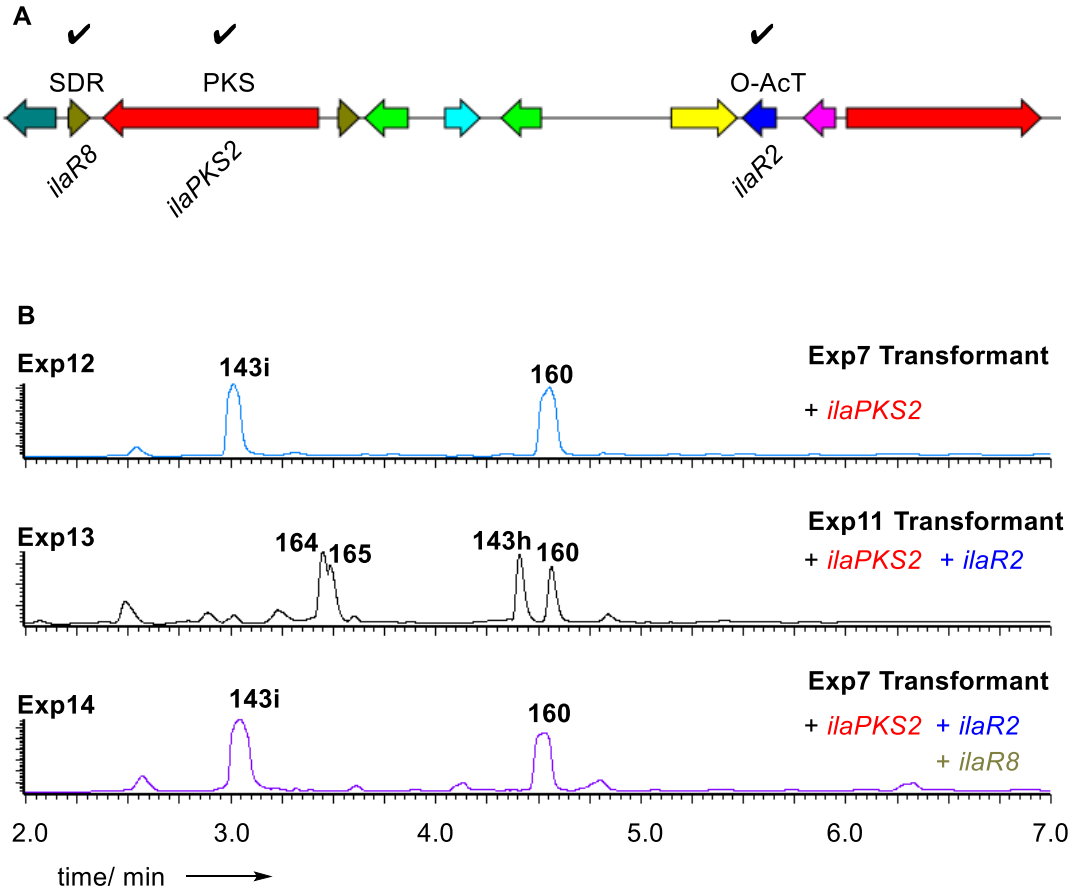


Figure 4.3.29 A, ilaBGC; **B**, DAD chromatograms of crude extracts from *A. oryzae* transformant of Exp 12, 13 and 14.

4.4 Discussion and Conclusion

4.4.1 Total Biosynthesis of Multiforisins

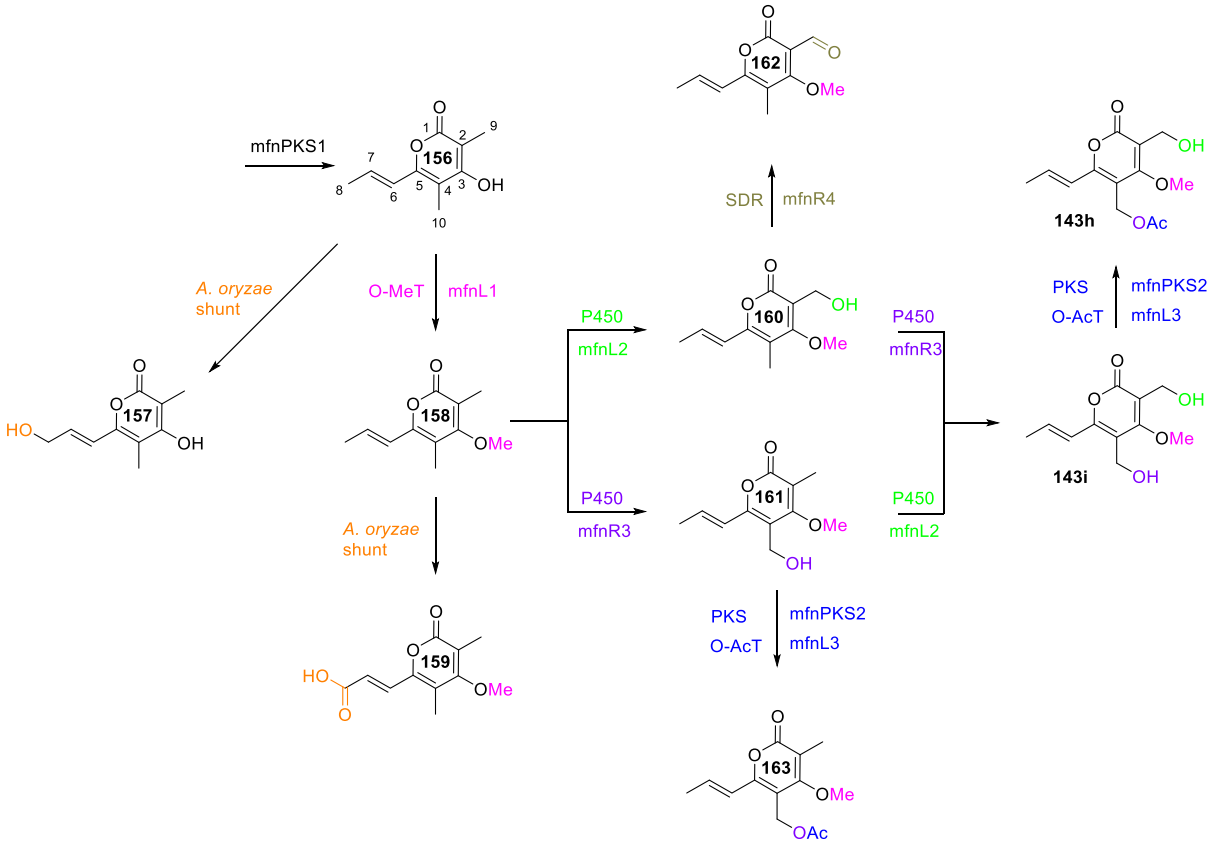
Fungal tetraketide pyrones comprise a large class of bioactive fungal specialized metabolites, often with potent biological activities, particularly in the realm of immunosuppression.[184] However, the precise biosynthetic pathways for their generation are unknown. To address this gap, we conducted a comprehensive investigation into the biosynthesis of multiforisins H **143h** and I **143i**. Our approach involved transcriptomic studies, BGC mining, and heterologous expression experiments.

We conducted a genome screening of *H. monticulosa*, along with two closely related species, *H. spongiphila* and *H. submonticulosa*, with the primary objective of identifying a BGC responsible for this particular pathway. Our attention was focused on a specific BGC from *H. monticulosa,* denoted as mfnBGC, which encodes a pair of hrPKSs and many tailoring enzymes. The boundaries of mfnBGC were established based on transcriptomic studies in *H. monticulosa*, which revealed that this region exhibits transcriptional co-regulation. Almost identical BGCs were identified in the closely related species *H. spongiphila* and *H. submonticulosa*. In addition, we used transcriptomic data to pinpoint the intron positions in two PKS genes and tailoring genes. To further investigate biosynthesis, we employed heterologous expression in the versatile fungal host *Aspergillus oryzae*, establishing a comprehensive biosynthesis platform.

To begin, we started the expression of *mfnPKS1* using the P_{amyB}. The transformed strains yielded a previously unobserved compound **156**, with a titer of 123 mg/L. This compound was absent in the unaltered *A. oryzae*. Through extensive NMR analysis, we unequivocally identified this compound as amphicopyrone A **156** (Scheme 4.4.1).

Subsequently, we undertook a series of gene expression experiments involving the mfn genes (Exp 2-11), followed by LCMS analysis. The introduction of P450 monooxygenase encoding genes, specifically mfnL2 (Exp 2) or mfnR3 (Exp 3), yielded minor quantities of 8-alcohol **157**, while the majority of **156** remained unaltered. Importantly, the alcohol **157** does not align with the multiforisins biosynthetic pathway, leading us to postulate it as an *A. oryzae*-specific metabolic shunt (Scheme 4.4.1)

In contrast, the coexpression of mfnL1, which encodes an *O*-MeT (Exp4), led to a substantial conversion of **156** into the well-established methyl ether **158** (140 mg/L), appearing to be an apparent pathway intermediate (Scheme 4.4.1). Additionally, minor amounts of the presumed shunt product, 8-carboxylate **159**, were also observed. Compound **158** was fully transformed into the known C-9 alcohol **160** (134 mg/L) upon the introduction of the MfnL2 P450 monooxygenase (Exp5). On the other hand, compound **158** undergoes conversion to **161** (31 mg/L), **160** and a minor quantity of diol **143i**, facilitated by MfnR3, which encodes the second P450 monooxygenase (Exp 6). Co-expression of both P450 monooxygenases at this stage (Exp7) yields the anticipated diol, multiforisin I **143i**, in a quantity of 36 mg/L.



Scheme 4.4.1 Deduced pathway for multiforisin H biosynthesis.

At this stage, the expression of *mfnR2*, responsible for encoding the FMO (Exp 8), did not yield any new compounds. Nevertheless, coexpression of *mfnR4*, which encodes the SDR, resulted in the oxidation of intermediate **160** to produce aldehyde **162** (Exp 9, 10 mg/L), along with a shunt product **167**. The addition of *mfnL3*-encoded *O*-AcT alone at this point did not produce any new compounds (Exp 10). However, when *mfnL3* was coexpressed with *mfnPKS2*, it led to the direct synthesis of multiforisin H **143h** (Exp11a, 28 mg/L). The MfnPKS2, featuring a mutation of the active site cysteine to serine on the KS domain, is expected to maintain its ability to initiate with an acetyl starter unit but lose its capacity for subsequent elongation. Our experiment showed that the PKS and *O*-AcT make up a novel acetylene transferase system.

In the end, a set of thioether shunts (**164-166**) was obtained following the acidification of the media during the extraction process (Exp11b, Figure 4.4.1**)**. These compounds are most likely formed through the generation of quinomethide-type structures, which are subsequently quenched by 2-hydroxyl-3-thiopropionate or 2-thioacetate.

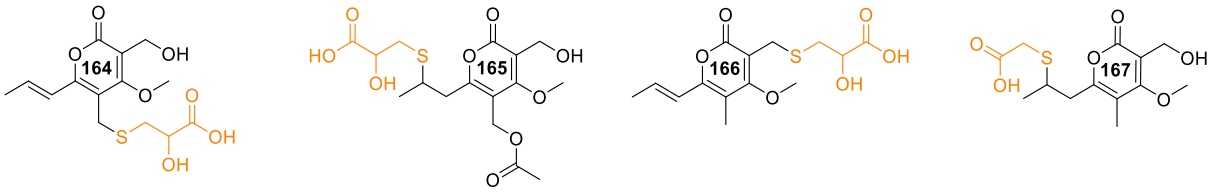
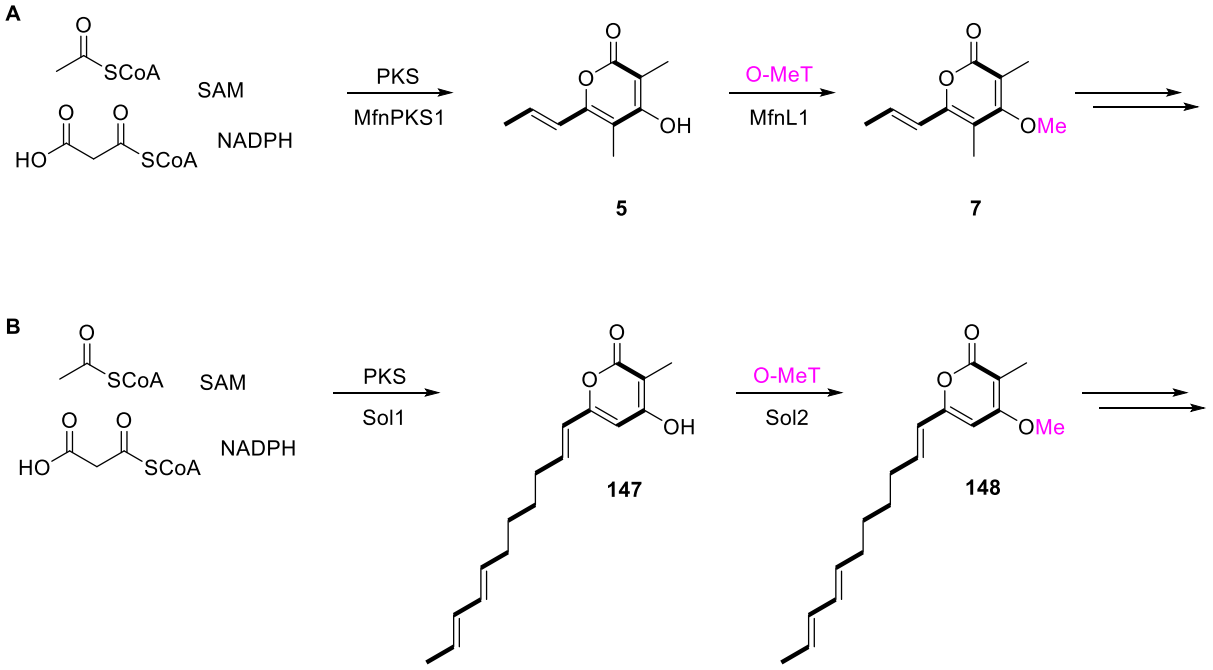


Figure 4.4.1 Products of Exp11b while the extraction process was in acidification.

Our study revealed that MfnPKS1 encodes a tetraketide pyrone **156**, while Sol1, despite more than 50% protein sequence identity, encodes an octaketide pyrone **147**.[192] Tetraketide pyrone **156** undergoes *O*-methylation by MfnL1, yielding compound **158** (Scheme 4.4.1). This agrees with the early steps of the solanapyrones pathway, where Sol1 encodes octaketide **147**, and a methyl group is added to form **148** (Scheme 4.4.2).[192] This result is further supported by BGC alignments, which show higher sequence identities between the protein sequences of MfnPKS1 and Sol1, as well as between MfnL2 and Sol2 (Figure 4.3.2). However, starting from the third step, these two pathways diverge in different directions.[192] This is also consistent with BGC alignment, revealing no sequence identities in other protein sequences within these two BGCs (Figure 4.3.2).



Scheme 4.4.2 Two early steps of multiforisin H **143h** (**A**) and solanapyrones (**B**) pathways

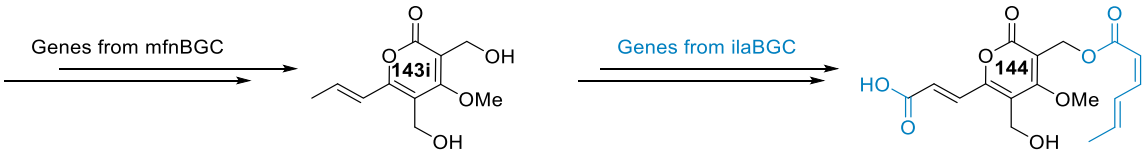
4.4.2 Combinatorial Biosynthesis of Multiforisins and Islandic acid

Islandic acid **144** and its related compounds have been previously synthesized through a multistep chemical process,[190] but these methods are known to be inefficient. Employing total biosynthesis for the production of these compounds offers significant advantages. Furthermore, despite the intriguing presence of an *EZ* diene in compounds **144** and **145**, along with their potent and diverse bioactivities, this class of bioactive tetraketide pyrone has not received comprehensive investigation at the biogenetic level.

In addition, the systematic heterologous expression of the mfn genes offers a convenient biosynthetic platform for production of multiforisins, yielding good overall titres. Therefore, we aimed to leverage the mfn genes in combination with ila genes for co-expression in *A. oryzae*, with the goal of obtaining compounds **144** and **145** or their intermediates (Scheme 4.4.3).

We initially conducted a BLAST search within the publicly available genome of *Penicillium islandicum*, the organism responsible for producing islandic acid **144**. We used the mfn genes as a reference, which led to the discovery of a very similar BGC. This cluster encodes homologs of several key proteins, including MfnPKS1 (IlaPKS1), two parallel cytochrome P450 enzymes (IlaR4 & IlaR6), the *O*-methyltransferase (IlaR1), the transcriptional regulator (IlaR3), the SDR (IlaR7), and FMO (IlaR5). However, the second PKS (IlaPKS2) and the acyl transferase (IlaR2) in the ilaBGC do not exhibit significant similarity to the corresponding proteins encoded by the mfnBGC. This observation is consistent with the idea that they are involved in the formation and attachment of distinct PKS-derived acyl groups. Additionally, the ilaBGC contains a second SDR (IlaR8) with limited similarity to MfnR4.

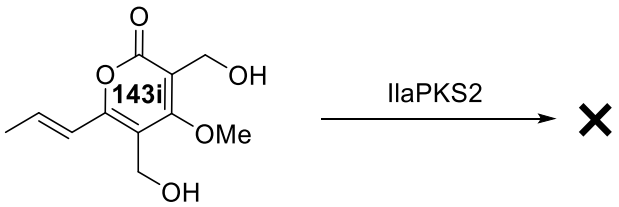
Our hypothesis is that IlaPKS2 is responsible for synthesizing the unusual *EZ* triketide, which might be subsequently transferred to the C-2 hydroxymethyl group by the Ila-AT enzyme, similar to the multiforisin chemistry. Interestingly, in the case of *H. monticulosa*, the role of MfnPKS2 is primarily to provide an acetyl group for attachment at the 10-OH position.



Scheme 4.4.3 Combinatorial biosynthesis of multiforisins and islandic acid **144**.

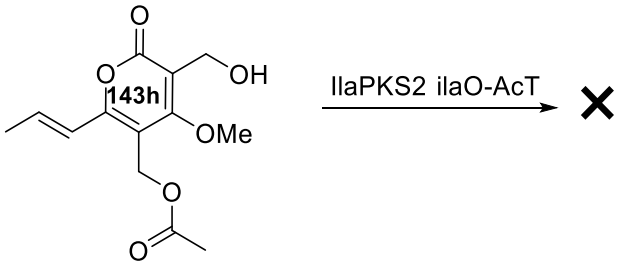
Three experiments were carried out. In the first experiment (Exp12), we co-transformed *A. oryzae* from Exp7, known for producing compounds **143i** and **160**, with *ilaPKS2* alone. We chose not to employ *O*-AcT due to the presence of a rare carnitine acyl transferase (cAT) C-terminal domain in IlaPKS2. This domain was previously reported to be responsible for the reversible transfer of the completed polyketide to a polyol in fungal systems.[205] Regrettably, this specific experiment did not yield successful outcomes (Scheme 4.4.4). We postulated that a compound such as **143i** might potentially function as the acceptor,

but the coexpression of *ilaPKS2* in cultures producing **143i** did not result in the formation of compound **144**.



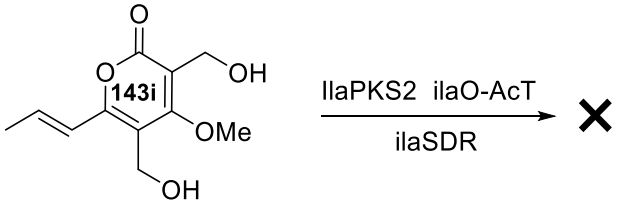
Scheme 4.4.4 Expression of *ilaPKS2* using **143i** as the sunstrate

In Exp13, we co-transformed *A. oryzae* from Exp11, which produces compounds **143i**, **143h**, **160**, and **163**, with *ilaPKS2* and *ilaR2*, the latter encoding a putative acyl transferase. However, the addition of the putative acyl transferase encoded by *ilaR2* did not yield the desired results either (Scheme 4.4.5). We hypothesized that the presence of the terminal carboxylate on compound **144** might be necessary for recognition, but the ilaBGC lacks any additional monoxygenases that could oxidize position 8. Oxygenation at C-8 in *A. oryzae* was inconsistently catalyzed by a native *A. oryzae* enzyme, resulting in alternative pathways leading to the formation of compounds **157** and **159**.



Scheme 4.4.5 Co-expression of *ilaPKS2* and *ilaR2* using **143h** as the sunstrate

In our final experiment (Exp14), we co-transformed *A. oryzae* from Exp7, which produces compounds **143i**, **156**, and **160**, with *ilaPKS2*, *ilaR2*, and *ilaR8*, the latter encoding a SDR (Scheme 4.4.6). None of these transformants yielded any additional compounds beyond those already produced by the *A. oryzae* hosts. It is possible that the oxygenation of carbon-8 during islandic acid biosynthesis may involve an oxygenase enzyme not accounted for in the ilaBGC. Additionally, ilaSDR-R8 may have the potential to catalyze the conversion of the resulting product to a carboxylic acid. However, the introduction of IlaSDR-R8 did not facilitate this chemical transformation.



Scheme 4.4.6 Co-expression of *ilaPKS2*, *ilaR2* and *ilaR8* using **143i** as the sunstrate.

In conclusion, we have accomplished the efficient total biosynthesis of numerous fungal tetraketide pyrones, including both previously identified and newly discovered compounds. We have elucidated the pathway leading to multiforisins. This achievement paves the way for strategically engineering forthcoming members of this compound family through synthetic biology, thereby circumventing the need for multi-step synthetic chemistry methods. Nevertheless, further research is imperative to fully understand the biosynthesis and attachment of the EZ triketide side chain in compounds **144** and **145**.

5. Experimental

5.1 Growth Media, Solutions and Antibiotics

5.1.1 Growth Media

Growth media used in this study was kindly prepared by the technical staff of the BMWZ and their ingredient is listed in Table 5.1.1.

Table 5.1 Growth media and their ingredient used in this study.

Media / buffer	Ingredient
YPAD Agar (w/v)	1.00 % Yeast extract; 2.00 % Tryptone; 2.00 % D (+)-Glucose Monohydrate; 0.03 % Adenine; 1.50 % Agar.
YPAD (w/v)	1.00 % Yeast extract; 2.00 % Tryptone; 2.00 % D (+)-Glucose Monohydrate; 0.03 % Adenine.
SM-URA Agar (w/v)	0.17 % Yeast nitrogen base; 0.50 % Ammonium sulfate; 2.00 % D (+)-Glucose Monohydrate; 0.077 % Complete supplement mixture minus Uracil; 1.50 % Agar.
LB Agar (w/v)	0.50 % Yeast extract; 1.00 % Tryptone; 0.50 % Sodium chloride, 1.5 % agar.
LB (w/v)	0.50 % Yeast extract; 1.00 % Tryptone; 0.50 % Sodium chloride.
SOC (w/v)	0.50 % Yeast extract; 2.00 % Tryptone; 0.06 % Sodium chloride; 0.02 % Potassium chloride; 25 mM final concentration Magnesium chloride hexahydrate 2M; 1.0 % final concentration D (+)-Glucose Monohydrate 20 %.
DPY agar (w/v)	2.00 % Dextrin from potato starch; 1.00 % Polypeptone; 0.50 % Yeast extract; 0.50 % Monopotassium phosphate; 0.05 % Magnesium sulfate hexahydrate
DPY (w/v)	2.00 % Dextrin from potato starch; 1.00 % Polypeptone; 0.50 % Yeast extract; 0.50 % Monopotassium phosphate; 0.05 % Magnesium sulfate hexahydrate; 2.5 % agar
PDB (w/v)	2.40 % Potato dextrose broth
GN (w/v)	2.00 % D (+)-Glucose Monohydrate; 1.00 % Nutrient broth;
CZD/S Agar (w/v)	3.50 % Czapek Dox broth; 18.22 % D-Sorbitol; 0.10 % Ammonium sulfate; 0.05 % Adenine; 0.15 % L-Methionine; 1.50 % Agar; or 0.80 % Agar for soft agar
CZD/S soft agar (w/v)	3.50 % Czapek Dox broth; 18.22 % D-Sorbitol; 0.10 % Ammonium sulfate; 0.05 % Adenine; 0.15 % L-Methionine; 0.80 % Agar
CZD/S1 Agar	CZD/S Agar without Adenine
CZD/S1 soft Agar	CZD/S soft Agar without Adenine
CZD/S1 Agar/ w/o Methionine	CZD/S Agar without Adenine and Methionine
CZD/S1 soft Agar/ w/o Methionine	CZD/S soft Agar without Adenine and Methionine

5.1.2 Solutions and Antibiotics

Antibiotics were prepared with stock concentration and stored at - 20 °C for up to one year. Antibiotics were dissolved in the indicated solvents and sterilized by disposable syringe filters. Solutins were were prepared with ddH_2O and sterilized by disposable 0.2 µm Fisherbrand™ PES Bottle Top Filters or by sterilization at 121°C for 20 minutes.

Table 5.2 Solutions and antibiotic used in this study

FCC solution (v/v)	5% glycerol; 10% DMSO; ddH_2O
PEG solution (w/v)	50% polyethylene glycol 3350; ddH_2O
ssDNA	2 mg/mL salmon sperm DNA; TE buffer
Solution 1	0.8 M NaCl; 10mM $CaCl_2$; 50 mM Tris-HCl; pH 7.5.
Solution 2	60% (w/v) PEG3350; 0.8 M NaCl; 10 mM $CaCl_2$; 50 mM Tris-HCl; pH 7.5
gDNA extraction buffer	For 100 mL: 5 mL 1M Tris (pH 7.2); 10 mL 0.5 M EDTA; 3 g SDS; 1 mL Mercaptoethanol
50x TAE	2 M Tris acetate, 0.05 M EDTA, pH 8.3
Carbenicillin	Stock concentration is 50 mg/mLin ddH_2O, working concentration is 50 µg/mL
Kanamycin	Stock concentration is 50 mg/mL in ethanol, working concentration is 30 µg/mL

5.2 Microbiology Methods

All bacterial strains, yeast and *A. oryzae* NSAR1, *Xylaria hypoxylon, Hypoxylon rickii* were obtained from the BMWZ inhouse strain collection.

Table 5.3 Strains used in this work

Strain	**Genotype**
One Shot™ TOP10 chemically competent *E. coli*	F - *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*Lac*ZΔM15 Δ*Lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
One Shot™ ccd B Survival™ 2 T1 R competent cells	F - *mcr* A Δ (*mrr - hsd* RMS- *mcr* BC) Φ80 *lac* ZΔM15 Δ *lac* X74 *rec* A1 *ara* Δ139 Δ (*ara - leu*)7697 *gal* U *gal* K *rps* L (Str R) *end* A1 *nup* G *fhu* A:: *IS2*
One Shot™ OmniMAX™ 2 T1R Chemically Competent *E. coli*	F′ [*pro*AB + *lacIq*lacZΔM15 *Tn*10(TetR) Δ(*ccd*AB)] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *end*A1 *rec*A1 *sup*E44 *thi*-1 *gyr*A96 (NalR) *rel*A1 *ton*A *pan*D
Saccharomyces cerevisiae CEN.PK	MATa/α ura3-52/ura3-52, trp1-289/trp1-289, leu2-3_112/leu2-3_112 his3 Δ1/his3, Δ1MAL2-8C/MAL2-8C, SUC2/SUC2
Aspergillus oryzae NSAR1	Δ*argB sC- adeA niaD-*
Xylaria hypoxylon	wildtype
Hypoxylon rickii	wildtype
Penicillium islandicum strain iBT20602	wildtype

5.2.1 *E. coli*

To prepare for ligation or transformation, the first step is the thawing of 50 µL of One Shot® cells on ice. Next, pipette 1-5 µL from each ligation reaction into a vial containing competent cells, gently mixing by tapping. Place the vial on ice for 30 minutes. Subsequently, incubate the vial for 60 seconds in a 42°C water bath without stirring or shaking. After removing it from the bath, return the vial to the ice for 3 minutes. The next step involves adding 400 µL of S.O.C medium to each vial, followed by shaking at 37°C for one hour at 220 rpm. Finally, spread 50-200 µL from each transformation vial onto LB agar plates with added antibiotics. The plate is inverted and incubated at 37°C overnight, and colonies are selected for further analysis, including plasmid isolation, PCR, or sequencing.

5.2.2 *S. cerevisiae*

S. cerevisiae cells were cultured on YPAD agar at 30 °C for three days. A single colony was incubated overnight in 10 mL of YPAD media at 30 °C with agitation at 200 rpm. Then this 10 mL YPAD culture was transferred to 40 mL of fresh YPAD in a 250 mL Erlenmeyer flask. The combined culture was then incubated at 30 °C with continuous shaking at 200 rpm for an additional 4 hours. Following the incubation, cells were harvested through centrifugation at 3,000 g for 5 minutes. The resulting pellet underwent two rounds of washing with 25 mL of double-distilled H_2O, with centrifugation after each wash. The pellet was suspended in 5 mL of the FCC solution and divided into 50 µL, each placed in separate 1.5 mL Eppendorf tubes and then stored at -80 °C. To thaw the samples, they were incubated on ice and subsequently centrifuged for 15 seconds at 21,000 g. After removing the FCC solution, the cells were ready for use in the yeast transformation process.

The next steps are addition of components to the competent cells: 50 µL of ssDNA, followed by 36 µL of 1 M LiOAc, and then 34 µL of a DNA mixture containing the linearized plasmid and corresponding inserts. Subsequently, 240 µL of the PEG solution was added with thorough mixing to ensure a homogeneous blend. The resulting transformation mixture underwent incubation at 30 °C for 30 minutes at 300 rpm, followed by further incubation at 42 °C for 40 minutes. After these incubation steps, the cells were subjected to centrifugation at 13,000 g for 60 seconds to obtain a pellet, from which the supernatant was separated. The resulting pellet was then suspended in 200 µL of double-distilled H_2O before being plated onto selective SM-Ura plates, which were subsequently incubated at 30 °C for three days. The yeast plasmid extraction was performed using the ZymoprepTM Yeast Plasmid Miniprep II kit (Zymo Research).

5.2.3 *A. oryzae* NSAR1

A. oryzae NSAR1 was cultured on a DPY agar plate for 5-10 days. Conidia were then introduced into 50 mL of GN medium in a 250 mL flask. The flask underwent overnight incubation at 28 °C with agitation at 110 rpm. The mycelia were harvested using a sterile Mira-cloth filter. These mycelia were immersed in a 25 mL solution of 0.8 M NaCl containing 15 mg/mL of lysing enzyme within a 50 mL Falcon tube. This setup was positioned on a Stuart SB3 rotator and allowed to incubate at room temperature for 4 hours. Protoplasts were liberated from the hyphal strands through gentle pipetting. The resultant supernatant was then filtered through another sterile Mira-cloth filter and collected in a new 50 mL Falcon tube. This collected solution underwent centrifugation at 3000 x g for 5 minutes to gather the protoplasts.

The supernatant was discarded, and the pellet containing the protoplasts was reconstituted in 1 mL of solution 1. This reconstitution was divided into 10 individual tubes, each placed within a 15 mL Falcon tube. Plasmids were introduced into the protoplast solution following a specific protocol: for a single plasmid, 1 µg of the plasmid was utilized per tube; for two plasmids, 3 µg of each plasmid was added

to a single tube; for three plasmids, 6 µg of each plasmid was included in a single tube. The combination of the protoplast solution and plasmids underwent ice incubation for 2 minutes. Following this, 1 mL of solution 2 was introduced into each tube. The tubes were gently inverted multiple times to ensure thorough mixing of protoplasts, solutions, and plasmids. Subsequently, the tubes were incubated at 28 °C for 30 minutes. Afterward, 12 mL of pre-warmed CZD/S soft agar was added to each tube and mixed. The resulting mixture was overlaid onto two prepared CZD/S agar plates. These plates were then subjected to incubation at 28 °C for 4-5 days. All the transformants were transferred to another CZD/S selection plate and then this step was repeated. Subsequent to this, the colonies were cultured for 5-7 days on DPY agar.

5.3 Molecular Biological Methods

5.3.1 Genomic DNA Extraction

The mycelia are harvested, dried with paper towels, and a small amount is transferred into a 2 mL Eppendorf tube. Next, 500 µL of Lysis Buffer (Table 5.2) is added, and it is vortexed thoroughly. Following that, incubation is conducted for 15 minutes at 65°C, with vortexing every 5 minutes, succeeded by transferring to ice for 5 minutes. Subsequently, 200 µL of 8M potassium acetate solution is added and thoroughly mixed. The tubes are centrifuged for 15 minutes at 13,000 rpm. Following centrifugation, 300 µL of the liquid phase is transferred into a fresh 1.5 Eppendorf tube, and 300 µL of isopropanol is added with 8-10 inversions. The tubes are then centrifuged for 3 minutes at 13,000 rpm, and the supernatant is removed. The pellet is washed with 1 mL of 70% ethanol, and the tube is centrifuged for 3 minutes at 13,000 rpm and the supernatant was removed. The pellet is dried at 65°C for 10 minutes. Finally, 100 µL of water is added (optional: 2 µL of RNAase), and the gDNA is stored at 4°C.

5.3.2 RNA Extraction and Reverse Transcription

The mycelia of both *A. oryzae* NSAR1 and its transformants were collected after 5 days of cultivation in DPY medium. Subsequently, they were frozen in liquid nitrogen and pulverized into a fine powder. RNA extraction utilized the Quick-RNA Fungal/Bacterial Microprep Kit (ZYMO). Converting the extracted RNA into cDNA involved the use of the High Capacity RNA-to-cDNA™ Kit. The entire process for RNA extraction and cDNA reverse transcription strictly followed the manufacturer's instructions. The resulting cDNA served as the template for amplifying *aol4* and *aol1*, generating fragments ranging from 450 to 500 base pairs, as outlined in Chapter 3.

5.3.3 Oligonucleotides and PCR

Oligonucleotides designed for polymerase chain reaction (PCR) were generated using the Geneious software platform and synthesized by Sigma Genosys and Eurofins (Table 5.4). PCR was conducted to amplify DNA fragments, employing two distinct types of thermostable polymerases. For amplifying

DNA fragments intended for heterologous expression, the high-fidelity DNA polymerase Q5® (New England Biolabs) was utilized. In the case of colony PCR, the enzyme of choice was the OneTaq® DNA polymerase (New England Biolabs), following the protocol from manufacturer.

Table 5.4 Oligonucleotide sequences for chapter 2

Primer	Sequence (5'- 3')	Purpose
F1941A1-F	TCGAGGAGTCCATCATGGGAGCCGAGTGTGCTGGAGTTGTGAG	F1941A
F1941A1-R	CTCACAACTCCAGCACACTCGGCTCCCATGATGGACTCCTCGA	
F2157A1-F	GTAGAGACAGGAGCTATATGGCCGCAGAAGCACTGCCCAAGAT	F2157A
F2157A1-R	ATCTTGGGCAGTGCTTCTGCGGCCATATAGCTCCTGTCTCTAC	
I2147A-F2157V1-F	TCGCCGCTATTGGTAGAGACAGGAGCTATATGGTCGCAGAAGCACTGCCCAAGAT	I2147A/F2157V
I2147A-F2157V1-R	GCGACCATATAGCTCCTGTCTCTACCAATAGCGGCGAGATCGACCGCAGTGAAGG	
F1941A1-F	TCGAGGAGTCCATCATGGGAGCCGAGTGTGCTGGAGTTGTGAG	F1941A/F2157A
F1941A1-R	CTCACAACTCCAGCACACTCGGCTCCCATGATGGACTCCTCGA	
F2157A1-F	GTAGAGACAGGAGCTATATGGCCGCAGAAGCACTGCCCAAGAT	
F2157A1-R	ATCTTGGGCAGTGCTTCTGCGGCCATATAGCTCCTGTCTCTAC	
F1941A1-F	TCGAGGAGTCCATCATGGGAGCCGAGTGTGCTGGAGTTGTGAG	F1941A/I2147A/F2157V
F1941A1-R	CTCACAACTCCAGCACACTCGGCTCCCATGATGGACTCCTCGA	
I2147A-F2157V1-F	TCGCCGCTATTGGTAGAGACAGGAGCTATATGGTCGCAGAAGCACTGCCCAAGAT	
I2147A-F2157V1-R	GCGACCATATAGCTCCTGTCTCTACCAATAGCGGCGAGATCGACCGCAGTGAAGG	
I1938A1-F	TGGGCCAGCTCGAGGAGTCCGCCATGGGATTCGAGTGTGCTGG	I1938A
I1938A1-R	CCAGCACACTCGAATCCCATGGCGGACTCCTCGAGCTGGCCCA	
L2146A1-F	CCTTCACTGCGGTCGATGCCATCGCTATTGGTAGAGACAG	L2146A
L2146A1-R	TCTCTACCAATAGCGATGGCATCGACCGCAGTGAAGGTTG	
L2146V1-F	CAACCTTCACTGCGGTCGATGTCATCGCTATTGGTAGAGACAG	L2146V
L2146V1-R	CTGTCTCTACCAATAGCGATGACATCGACCGCAGTGAAGGTTG	
I2147A1-F	CCTTCACTGCGGTCGATCTCGCCGCTATTGGTAGAGACAGGAG	I2147A
I2147A1-R	CTCCTGTCTCTACCAATAGCGGCGAGATCGACCGCAGTGAAGG	
L2146A-I2147A1-F	CAACCTTCACTGCGGTCGATGCCGCTGCTATTGGTAGAGACAGGAG	L2146A/I2147A
L2146A-I2147A1-R	TCTCTACCAATAGCAGCGGCATCGACCGCAGTGAAGGTTG	

Table 5.5 Oligonucleotide sequences for chapter 4

Primer	Sequence (5'- 3')	Purpose
HSHE15-P1	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGGCGCCTCGAGACGAACA	Cloning for *mfnPKS1* into PEYA
HSHE15-P2	CTCGCATCTCCATCTGCGAGAAATGCGCCGCGCATCGCCATGGTTCCCGG	
HSHE15-P3	CTCTAACCGCCCGGGAACCATGGCGATGCGCGGCGCATTTCTCGCAGATG	
HSHE15-P4	GATGACGAGCAAGCTGTGTCGACCGTGAGACTCGGCCCTTGAAGGTTGAA	
HSHE15-P5	AAGCTACATATTCAACCTTCAAGGGCCGAGTCTCACGGTCGACACAGCTT	
HSHE15-P6	TGCCAACTTTGTACAAGAAAGCTGGGTCGGTCACGAACTCTCAGCTGGAG	
PKS3655seq-F1	TCTTCAGACCACGCTTTCAG	
PKS3655seq-R1	ATTCGAAAAAGTGACCCACG	
3653-1F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGACAGTTCAGGACCCCAT	Cloning for *mfnL2* into pTYGs-*arg* under Padh
3653-2F	GACGGATTCAAGAGTTACATGAAATTTACGGACCAATTGTCCGTATTAGC	
3653-1R	CTTCATTCGGGCTAATACGGACAATTGGTCCGTAAATTTCATGTAACTCT	
3653-3F	GCCTGCCCACCAGGTTAGAGGCAGCGGCTCGGTAGCCCACGCGATCTTCT	
3653-2R	TCGGTTGTTGAGAAGATCGCGTGGGCTACCGAGCCGCTGCCTCTAACCTG	
3653-4F	TTTCACCAAGCACATGTCGGATGATATGGCGCTGCTGATCAAAACTCTCACC GTTGATATGCCGAATCAAAT	
3653-3R	GGTGAGAGTTTTGATCAGCAGCGCCATATCATCCGACATGTGCTTGGTGAAA ACACTAGCCGCAAGACCGAC	
3653-5F	CAAGGGAAGCCGATCTTGCCTTGGCATGAACTTGGCATTCTGTGAGCTGT	
3653-4R	AGTGAGAGATACAGCTCACAGAATGCCAAGTTCATGCCAAGGCAAGATCG	
3653-5R	CAGGTTGGCTGGTAGACGTCATATAATCATTTATACGATAATCGCCCGCA	
3653Tadh-R	TCGTAGCTGTTTTCATTCTATGCGTTATGAACATGTTCCCTTATACGATAAT CGCCCGCA	
3654-1F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGTCCAGCCAAAACACCAC	Cloning for *mfnL1* into pTYGs-*arg* under Padh and PgpdA
3654-2F	TGAGAACCTGCGATCCGTTGTACACTACTCACTGGATGAGATGTTGAAGG	
3654-1R	TCTGCGGCCGCCTTCAACATCTCATCCAGTGAGTAGTGTACAACGGATCG	
3654-3F	TTGCGAAGGCAATGGCTGGCTTACGGCAAATGGACTACCACTTGGACTAC	
3654-2R	CTTTGAGCAGGTAGTCCAAGTGGTAGTCCATTTGCCGTAAGCCAGCCATT	
3654-4F	AACGGCCACATCTCCAGGTCTTTGGCCAAGCAATTCCCCGACCTGAACTT	
3654-3R	TTGAACAACAAAGTTCAGGTCGGGGAATTGCTTGGCCAAAGACCTGGAGA	
3654-4R	CAGGTTGGCTGGTAGACGTCATATAATCATTTAGGCTCGCTTCAGGTAAA	
Pgpd3654-F	AGCTTGACTAACAGCTACCCCGCTTGAGCAGACATCACCGATGTCCAGCCAA AACACCAC	
3654Tgpd-R	TCTTGCGAACTACGACAATGTCCATATCATCAATCATGACTTAGGCTCGCTT CAGGTAAA	
3658-1F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGTCAGTAAACGTACACGAAG TTCCC	Cloning for *mfnR3* into pTYGs-*arg* under Padh and Peno
3658-F2	GGTAAAAGCCTCTAAACTCTTCGACCGGATCCGTTTTCTGTCTACCAACG	
3658-R1	TTATCGTCGACGTTGGTAGACAGAAAACGGATCCGGTCGAAGAGTTTAGA	
3658-F3	ACACTAGGGACAAGGCTAATGTCTCACCAATAAACGTACACGAAGTTCCC	
3658-R2	CCTGTGGTAAGGGAACTTCGTGTACGTTTATTGGTGAGACATTAGCCTTG	
3658-F4	AATTTCCAGAGTAACGTTCTAGTCGATGAGATCTTCGGCGATATAATTGG	
3658-R3	ATGGTGACCGCCAATTATATCGCCGAAGATCTCATCGACTAGAACGTTAC	
3658-R4	CAGGTTGGCTGGTAGACGTCATATAATCATTTATAGTTTCTTCGGTCTTA	
007corect2-F	CGACTGACCAATTCCGCAGCTCGTCAAAGGATGTCAGTAAACGTACACGAAG TTCCC	
padh3652-F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGACGGTGAAAACACCAAA	Cloning for *mfnL3 into* pTYGs-*ade*
padh3652-R	TTTCATTCTATGCGTTATGAACATGTTCCCTTAGGAAGATACATAGGAGG	
pgpda3659-F1	ACAGCTACCCCGCTTGAGCAGACATCACCGATGGCTTCGAGAATCAACAC	Cloning for *mfnR4* into pTYGs-*ade* under PgpdA
pgpda3659-F2	GCTGAACTGAAGGGTGTTGAAACCCGGCAGTTCGACATCGGCGACATTGC	
pgpda3659-R1	CGGAAGAGCGGCAATGTCGCCGATGTCGAACTGCCGGGTTTCAACACCCT	
pgpda3659-R2	TACGACAATGTCCATATCATCAATCATGACTTAAATGGTCATGCCGATAG	
3657-Peno-P1	CGACTGACCAATTCCGCAGCTCGTCAAAGGATGAAGTTCTCTGCTCAGCA	Cloning for *mfnR2* into pTYGs-*ade* under Peno
3657-Peno-P3	CCCGCCACCGAGAAGGATGTTTCAACTATTGTCAAATACTGCAATGACAA	
3657-Peno-P2	CTCAATGCTGTTGTCATTGCAGTATTTGACAATAGTTGAAACATCCTTCT	
3657-Peno-P5	CTTCCCCTTCCGTGCGGACCGTCATCTCATGCTCTTCGATGTCCAGATCC	
3657-Peno-P4	TTCTCAGTGGGGATCTGGACATCGAAGAGCATGAGATGACGGTCCGCACG	
3657-Peno-P6	CAGGTTGGCTGGTAGACGTCATATAATCATTTACGCGGTGGTAGCTTCCG	

Table 5.5 Oligonucleotide sequences for chapter 4 (continue)

HSHE13-P1	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGGCGCGCATCTCAAGGCC	Cloning and sequencing for *mfnPKS2* in PEYA
HSHE13-P2	CTTTAGGAAATGGCCGCTATGTGTGTGATTTCTCTCTGTGCTCTCCTCTT	
HSHE13-P3	GTAGATGGGCAAGAGGAGAGCACAGAGAGAAATCACACACATAGCGGCCA	
HSHE13-P4	GCATCGGTGCTATTGTATTGCTGGTCATAGCCTCCCAGGAAAACAGCGGT	
HSHE13-P5	GGATTCGAATACCGCTGTTTTCCTGGGAGGCTATGACCAGCAATACAATA	
HSHE13-P6	TCGTGGCCAAGGAATTAACTTGTTCGGCACCTTAACGTGCCATTGGCTTG	
HSHE13-P7	AATATCCCCACAAGCCAATGGCACGTTAAGGTGCCGAACAAGTTAATTCC	
HSHE13-P8	GGCTTGAGGACTCATCCCTTAGTAGTTCGTCATATAGGGACCATGTTGCA	
HSHE13-P9	ATGGGTACGGTGCAACATGGTCCCTATATGACGAACTACTAAGGGATGAG	
HSHE13-P10	CAAAGAAAGTAGAGATGATCGGGGGATGACATTGACTTGATCATCCTCAG	
HSHE13-P11	ATTGTTCCAACTGAGGATGATCAAGTCAATGTCATCCCCCGATCATCTCT	
HSHE13-P12	TGCGGAGGATACGTGTTCGGATGCGGTGGCATTTAGAACCAGATCCGTGT	
HSHE13-P13	GGGGTTATCAACACGGATCTGGTTCTAAATGCCACCGCATCCGAACACGT	
HSHE13-P14	TGCCAACTTTGTACAAGAAAGCTGGGTCGGCTATGCAGCAGCCTCTTTGC	
3651Seq-F3A	TCGTAACACTGGTCTCAACC	
3651Seq-F1	ATGGGCCGAGAGCTCATCG	
3651Seq-F4A	ACGCGATCTTGGAAGGATCA	
3651Seq-R1	GGAGGAACCCTTCGTTTGCG	
3651Seq-F5A	TGGCGCTTTATTCCACTACT	
3651Seq-F6A	CTACGCGGACTTGGAGACAA	
3651Seq-F2	AACAACTCGAAGCGCCTGGA	
PiPKS2-F1	CTGAACAATAAACCCCACAGCAAGCTCCGAATGAGCGGAAGAAATCCTAT	Cloning for *ilaPKS2* into pTYGs-*met* under PamyB
PiPKS2-R1	CGTCTCCAGCAAGAGCCGTTGTTGAGGGTCCATTGCTGCAGCCTCCTGCG	
PiPKS2-F2	AATGTTCTGCCGCAGGAGGCTGCAGCAATGGACCCTCAACAACGGCTCTT	
PiPKS2-R2	TAGTGGAAAGTGTATCAAAC	
PiPKS2-F3	CATAGTTTCTCAGTGACTGC	
PiPKS2-R3	CGAACTCTTGAAGCTCTTTC	
PiPKS2-F4	AGTGGCGTCACATATTTAGT	
PiPKS2-R4	ACTCTCCACCCTTCACGAGCTACTACAGATTCAAGCCTTAGATAGACCAC	
PiPKS2-seq-1	CAACCAGGACGGCCACACGG	
PiPKS2-seq-2	TACGTGGAACCTCGTTGAAG	
PiPKS2-seq-3	ATCGGATCAGGGCTTCGAGC	
PiPKS2-seq-4	CCTATTCAATCTTCCGAGGT	
Pgpd-PiAcT-F	ACAGCTACCCCGCTTGAGCAGACATCACCGATGTCTTTCGCCAAAGCTCA	Cloning of *ilaR2*
PiAcT-Teno-R	CAGGTTGGCTGGTAGACGTCATATAATCATCTATAACACACTAGACGGCC	
Padh-SDR2-F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGGCTTCATATCTCATCAC	Cloning of *ilaR8*
Padh-SDR2-R	TTTCATTCTATGCGTTATGAACATGTTCCCTTACCAGGGAGCATTGGAGC	
SeqPEYA-F	ACGGCCAGTCTTAAGCTCGGG	Public primers located in all plasmids, for sequencing or cloning.
SeqPEYA-R	CTATAGGGGATATCAGCTGGA	
PamyB_S-F1	CATGCTTGGAGGATAGCAACCG	
PamyB_S-R1	ACTCCAACTGTACATCAAACTCA	
Padh plugF	ATTCACCACTATTATTCCCACCCTATAATA	
Padh plugR	GAGACGAAACAGACTTTTTCATCGCTAAAA	
PgdpA plugF	CTTTTCTTTTCTCTTTCTTTTCCCATCTTC	
PgpdA plugR	TACGACAATGTCCATATCATCAATCATGAC	
Peno plugF	CTTCTTAAATATCGTTGTAACTGTTCCTGA	
Peno plugR	CGAAGTATATTGGGAGACTATAGCTACTAG	

Table 5.6 Oligonucleotide sequences for Chapter 3

Primer	Sequence (5'- 3')	Purpose
Pamy-hrtc-F	CTGAACAATAAACCCCACAGCAAGCTCCGAATGGCGCCAATGGTCGACGA	HrTc forward
Pamy-hrtc-R	ACTCTCCACCCTTCACGAGCTACTACAGATTCAGGTGGCCGGCTGCACAT	HrTc reverse
Padh-hrl4-F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGCTCTCCTACATGCTCCC	HrL4 forward
Padh-hrl4-R	TTTCATTCTATGCGTTATGAACATGTTCCCTTAATTCAATGGAAACCCTC	HrL4 reverse
Pgpd-hrl5-F	ACAGCTACCCCGCTTGAGCAGACATCACCGATGGTTAAGTTGGACTTCGA	HrL5 forward
Pgpd-hrl5-R	TACGACAATGTCCATATCATCAATCATGACCTACGTCAACCAGGGCTTGA	HrL5 reverse
Peno-hrl8-F	CGACTGACCAATTCCGCAGCTCGTCAAAGGATGGCATCACTCAGGATTGA	HrL8 forward
Peno-hrl8-R	CAGGTTGGCTGGTAGACGTCATATAATCATTTAGCGGAAGTCTGTCGCCG	HrL8 reverse
Pamy-hrl3-F	CTGAACAATAAACCCCACAGCAAGCTCCGAATGGAGGAAGGCTTGATTGC	HrL3 forward
Pamy-hrl3-R	ACTCTCCACCCTTCACGAGCTACTACAGATTCAGGACGAGAGCGCATCTC	HrL3 reverse
Padh-hrl1-F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGTCGTCGACTCCCGTGCT	HrL1 forward
Padh-hrl1-R	TTTCATTCTATGCGTTATGAACATGTTCCCTTAGACTGCCGCTTTCCGCC	HrL1 reverse
Pgpd-hrl7-F	ACAGCTACCCCGCTTGAGCAGACATCACCGATGTTCGGAACAGATGCTCT	HrL7 forward
Pgpd-hrl7-R	TACGACAATGTCCATATCATCAATCATGACTTAGTGCTCTACCAACTTTG	HrL7 reverse
Padh-prl3-F	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGGCCGTGATCTTCTCCTC	PrL3 forward
Padh-prl3-R	TTTCATTCTATGCGTTATGAACATGTTCCCTCAGGCCTTGAGACGTCTTT	PrL3 reverse
Pgpd-prl4-F	ACAGCTACCCCGCTTGAGCAGACATCACCGATGAACGCCTCAAAGTTGCC	PrL4 forward
Pgpd-prl4-R	TACGACAATGTCCATATCATCAATCATGACCTAGCCAGTCTTGACGTCTG	PrL4 reverse
Pgpd-prl7-F	ACAGCTACCCCGCTTGAGCAGACATCACCGATGACCATATCTCCAATCCC	PrL7 forward
Pgpd-prl7-R	TACGACAATGTCCATATCATCAATCATGACCTATGCCTTGAGTGTGCTAC	PrL7 reverse
Peno-prl9-F	CGACTGACCAATTCCGCAGCTCGTCAAAGGATGGATGCACTCGAGGCTCC	PrL9 forward
Peno-prl9-R	CAGGTTGGCTGGTAGACGTCATATAATCATCTAATCGACTTCTGTGTTTA	PrL9 reverse
Padh-aol3-F	TTTCTTTCAACACAAGATCCCAAAGTCAAAATGTTTGAGGGCGCCTACAC	AoL3 forward
Padh-aol3-R	TTCATTCTATGCGTTATGAACATGTTCCCTCTATGTGCACCTTCTCCGTT	AoL3 reverse
Pgpd-aol2-F	TAACAGCTACCCCGCTTGAGCAGACATCACATGACTCTAATATCGCTGTC	AoL2 forward
Pgpd-aol2-R	ACGACAATGTCCATATCATCAATCATGACCCTACTGCGCACTTAATCTCA	AoL2 reverse
RT-PCR-aol4-F	TGTTTTCAATACAATGGAACGATTC	RT-PCR for AoL4
RT-PCR-aol4-R	GGCGAATGTGGCATACTTCCACTGG	RT-PCR for AoL4
RT-PCR-aol1-F	AGGAGGAAGCTCAGGGATTGGCTGG	RT-PCR for AoL1
RT-PCR-aol1-R	GCCTAATTTCGTTGCACTATAGACC	RT-PCR for AoL1
Padh-2XhR1-F1	TTTCTTTCAACACAAGATCCCAAAGTCAAAATGACAGCCAAAATGTTCGA	CDS part1 forward
Padh-2XhR1-R1	TAGAGTCTATCAACAAACGA	CDS part1 reverse
Padh-2XhR1-F2	TGCAAGAAAGTCGTTTGTTGATAGACTCTATCGTCGGTTTGGTGATACGC	CDS part2 forward
Padh-2XhR1-R2	GCATAGCACTCTTGGGCATT	CDS part2 reverse
Padh-2XhR1-F3	AGATGTTTGGAATGCCCAAGAGTGCTATGCACATATATCGCAACGACAAG	CDS part3 forward
Padh-2XhR1-R3	TCGCGTTGCCGTTCTATAGTGATGGTGCCACACGCGGAGATGAGCTGGTA	CDS part3 reverse
Padh-2XhR1-F4	TGGCACCATCACTATAGAAC	CDS part4 forward
Padh-2XhR1-R4	GCTGTAAGAACGAACAGATC	CDS part4 reverse
Padh-2XhR1-F5	TGCAAGCGAGGATCTGTTCGTTCTTACAGCTGCAACCGCAAATGCTATGC	CDS part5 forward
Padh-2XhR1-R5	AGGTTGGCTGGTAGACGTCATATAATCATACTAAGAATGCTTTACAGTGG	CDS part5 reverse
PamyB_S-F	CATGCTTGGAGGATAGCAACCG	For sequencing, located in PamyB
PamyB_S-R	ACTCCAACTGTACATCAAACTCA	
Padh plugF	ATTCACCACTATTATTCCCACCCTATAATA	For PCR fragments to close the *AscI* cloning site without targeting DNA and also for sequencing
Padh plugR	GAGACGAAACAGACTTTTTCATCGCTAAAA	
PgdpA plugF	CTTTTCTTTTCTCTTTCTTTTCCCATCTTC	
PgpdA plugR	TACGACAATGTCCATATCATCAATCATGAC	
Peno plugF	CTTCTTAAATATCGTTGTAACTGTTCCTGA	
Peno plugR	CGAAGTATATTGGGAGACTATAGCTACTAG	

5.3.4 Enzyme Restriction and Gateway Recombination

The restriction enzymes *Asc*I and *Not*I were employed to linearize the pTYGs and PE-YA vectors. Plasmids, along with the corresponding restriction enzymes and buffers, were combined and incubated at 37 °C overnight. Following this incubation, heat inactivation was carried out at 80 °C for 15 minutes, and the resulting mixture was subsequently purified using a DNA clean-up kit. To transfer the larger genes from the PE-YA vectors to the pTYGs vectors, the PCR Cloning System with Gateway® Technology from Invitrogen was utilized. This process followed the manufacturer's instructions for LR recombination reactions.

5.3.5 DNA Purification and Plasmid Isolation

PCR products were loaded into an agarose gel utilizing a horizontal gel electrophoresis system (Bio-Rad). This process involved the use of 0.5x TAE-buffer and 1% (w/v) agarose gels, which were supplemented with ROTI®-Safe GelStain. The sizes of the PCR products were determined by comparison to a 1 kb DNA ladder from New England Biolabs. Subsequently, images of the gels were captured using the Gel DocTM XR^+ system (Bio-Rad). For further purification, the PCR products or the cut gel were subjected to the NucleoSpin Gel and PCR Clean-up Mini kit for DNA extraction and clean-up, by following the instructions of the manufacturer (Macherey-Nagel).

Plasmid extraction from bacterial cells was accomplished by the NucleoSpin® Plasmid Mini Kit (Macherey-Nagel) according to the instructions of the manufacturer. Then the concentration of the isolated plasmids was determined using a DeNovix spectrophotometer. Elution buffer or ddH_2O, as used in the plasmid isolation process, served as blanks for this measurement. Following concentration determination, the plasmids were sent for sequencing by Eurofins Genomics. Total plasmid DNA from *S. cerevisiae* cells was extracted using the ZymoprepTM Yeast Plasmid Miniprep II kit (ZYMO), following the protocols of manufacturer.

5.4 Chemical Analysis

5.4.1 Liquid Chromatography Mass Spectrometry

Analytical LCMS system includes a Waters 2767 autosampler, a Waters 2545 pump, and a Phenomenex Kinetex column (2.6 µm; C_{18}; 100 Å; 4.6 x 100 mm) equipped with a Phenomenex Security Guard precolumn (Luna, C_5; 300 Å). The solvent flow rate was maintained at 1.0 mL/min. Two instruments were employed for detection: a Waters SQD-2 mass detector, capable of operating in both ES^+ and ES^- modes, with a mass range of 100 to 1000 *m/z*, and a Waters 2998 Diode Array detector covering a wavelength range from 210 to 600 nm, in addition to a Waters 2424 ELSD.

The HPLC system runs using two solvents: water (A) with an additional 0.05% formic acid and acetonitrile (B) containing 0.045% formic acid. The system parameters included a 15-minute gradient

starting at 10% B, gradually increasing to 90% B over 10 minutes, maintaining 90% B for next 2 minutes, and finally, a three-minute gradual return to a 10% B.

For purification, all compounds underwent processing through a Waters mass-directed autopurification system, which comprised a Waters 2767 autosampler, Binary Gradient Module 2545 with 515 HPLC pumps, and a System Fluid Organizer. A Phenomenex Kinetex Axia column (5 µm, C_{18}, 100 Å, 21.2 × 250 mm), along with a Security Guard pre-column (Luna C_5 300 Å), was employed. Compounds were eluted at a flow rate of 20 mL/min, and the process was conducted at ambient temperature. Fraction collection was managed by the Waters Sample Manager 2767, initiating fraction collection either based on mass detection or time intervals.

The collected fractions underwent vacuum evaporation to remove organic solvents, while the aqueous phases were dried using freeze dryers. Preparative LCMS was conducted using high-purity solvents, which included (A) water with 0.05% formic acid and (B) acetonitrile with 0.045% formic acid. The specific solvent gradient settings were determined based on the compound's characteristics, as assessed through analytical LCMS after testing.

5.4.2 Gas Chromatography Mass Spectrometry

An Agilent GCMS system is composed of an HP 6890 gas chromatograph, a 5973-mass detector, and an AG19091J-433 capillary column (30 m, 0.32 mm internal diameter, with a 0.25-µm film). The operating parameters were configured as follows: He was supplied at a flow rate of 2.0 mL per minute, with an injection volume of 1 µL. The transfer line temperature was set at 250 °C, and the electron energy was maintained at 69.9 eV.

The measurement method included an initial temperature of 100 °C for 2 minutes, followed by a temperature gradient from 100 °C to 180 °C at a rate of 2 °C per minute. Subsequently, the temperature was further increased to 300 °C at a rate of 40 °C per minute, with a final temperature hold at 300 °C for 7 minutes. The GCMS operated in splitless mode.

5.4.3 HRMS and NMR

A 1-mg sample was dissolved in 1 ml of solvent and subsequently diluted 100-fold before being subjected to high-resolution mass spectrometry (HRMS) analysis. The HRMS data were acquired by the technical team at the MS center of OCI. The analysis was performed using a high-resolution ultra-performance liquid chromatography mass spectrometry system (HR UPLC MS/MS), the Waters QTof Premier, equipped with both ESI and APCI-MS/MS capabilities. Additionally, the system was paired with a UPLC configuration that included a Waters Acquity system along with a TUV detector.

Nuclear Magnetic Resonance (NMR) measurements were obtained using three different Bruker instruments: the Bruker Ascend 600 (Avance NEO console, sample case and cryo-cooled probe DUL),

Bruker Ultrashield 500 (Avance IIIHD console, Sample Xpress and cryo-cooled probe TCI), and Bruker Ascend 400 (Avance III console, Sample Xpress and Prodigy BBFO probe), each operating at frequencies of 600, 500, and 400 MHz for proton (^{1}H) and 150, 125, and 100 MHz for carbon-13 (^{13}C), respectively. The chemical shifts are reported in ppm with respect to the tetramethylsilane (TMS) standard, and the coupling constants (*J* values) are provided in Hz. Data analysis was conducted using TopSpin 4.2.0 and MestReNova 14.3 softwares. The comprehensive structural elucidation was performed, involving the acquisition of 2D NMR experiments, such as ^{1}H-^{1}H COSY, HSQC, HMBC, as well as NOESY.

5.4.4 Extraction of *A. oryzae* Cultures

1 mL spore suspension was inoculated into a 500 mL baffled flask containing 100 mL of DPY-medium. This mixture underwent incubation at 28 °C with agitation at 110 rpm for 5-7 days. The entire culture was homogenized using a hand blender and then subjected to filtration. A double extraction process was conducted using ethyl acetate, and after separating the organic layers, they were desiccated with $MgSO_4$. The solvent was then evaporated under reduced pressure. The crude extract was dissolved in methanol at a concentration of 10 mg/mL and passed through a glass wool filter before undergoing LCMS testing. For the purification process, it was essential to elevate the crude extract concentration to 50 mg/mL. This was achieved by scaling up the cultivation of transformants to 1 liter in preparation for subsequent LCMS analysis.

Reference

[1] D. Bass, T. A. Richards, *Fungal Biol Rev* **2011**, *25*, 159–164.

[2] D. S. Hibbett, M. Binder, J. F. Bischoff, M. Blackwell, P. F. Cannon, O. E. Eriksson, *Mycol Res,* **2007**, *111*, 509–547.

[3] J. E. Stajich, S. K. Wilke, D. Ahrén, C. H. Au, B. W. Birren, M. Borodovsky, C. Burns, *Proc Natl Acad Sci U S A,* **2010**, *107*, 11889–11894.

[4] D. L. Hawksworth, R. Lücking, *Microbiol Spectr,* **2017**, 5.

[5] B. Wu, M. Hussain, W. Zhang, M. Stadler, X. Liu, M. Xiang, *Mycology,* **2019**, *10*, 127–140.

[6] J. E. Stajich, *Microbiol Spectr,* **2017**, 5(4).

[7] M. Stadler, D. Hoffmeister, *Front Microbiol,* **2015**, *6,* 127.

[8] M. Gill, *Nat Prod Rep* **2003**, *20*, 615–639.

[9] Z.-Y. Zhou, J.-K. Liu, *Natural Product Reports*, **2010**, *20*, 615.

[10] D. D. De Silva, S. Rapior, E. Sudarman, M. Stadler, J. Xu, S. Aisyah Alias, K. D. Hyde, *Fungal Divers,* **2013**, *62*, 1–40.

[11] K. Lorenzen, T. Anke, *Curr Org Chem,* **1998**, *2*, 329–364.

[12] C. Richter, K. Wittstein, P. M. Kirk, M. Stadler, *Fungal Diversity*, **2014**, *71*, 1-15.

[13] E. Skellam, *Trends Biotechnol* **2019**, *37*, 416–427.

[14] K. K. Sharma, *Crit Rev Biotechnol* **2016**, *36*, 743–759.

[15] K. Blin, K Blin, S Shaw, AM Kloosterman, Z Charlop-Powers, GP Van Wezel, MH Medema, T Weber, *Nucleic Acids Research*, **2021**, *49*, 29-35.

[16] M. Yandell, D. Ence, *Nat Rev Genet* **2012**, *13*, 329–342.

[17] C. L. M. Gilchrist, Y. H. Chooi, *Bioinformatics,* **2021**, *37*, 2473–2475.

[18] C. L. M. Gilchrist, T. J. Booth, B. Van Wersch, L. Van Grieken, M. H. Medema, Y. H. Chooi, *Bioinformatics Advances,* **2021**, *1*.

[19] A. David, S. Islam, E. Tankhilevich, M. J. E. Sternberg, *J Mol Biol,* **2022**, *434*.

[20] I. V. Muñoz, S. Sarrocco, L. Malfatti, R. Baroncelli, G. Vannacci, *Front Plant Sci,* **2019**, *10*.

[21] F. Alberti, G. D. Foster, A. M. Bailey, *Appl Microbiol Biotechnol,* **2017**, *101*, 493–500.

[22] Y. M. Chiang, T. S. Lin, S. L. Chang, G. Ahn, C. C. C. Wang, *ACS Synth Biol,* **2021**, *10*, 173–182.

[23] M. Chypre, N. Zaidi, K. Smans, *Biochem Biophys Res Commun,* **2012**, *422*, 1–4.

[24] T. Mahmud, H. B. Bode, B. Silakowski, R. M. Kroppenstedt, M. Xu, S. Nordhoff, G. Höfle, R. Müller, *Journal of Biological Chemistry,* **2002**, *277*, 32768–32774.

[25] J. B. Ohlrogge, J. G. Jaworski, *Annual Review of Plant Biology*, **1997,** *48*, 109-136.

[26] H. G. H van der Lee, M. R. Munday, *Biochem. Biophys. Res. Commun* **2002**, *19*, 155–63.

[27] E. Choi-Rhee, J. E. Cronan, *Journal of Biological Chemistry* **2003**, *278*, 30806–30812.

[28] R. W. Brownsey, A. N. Boone, J. E. Elliott, J. E. Kulpa, W. M. Lee, *Biochem Soc Trans*, **2006,** *34*, 223-7.

[29] L. Buyachuihan, F. Stegemann, M. Grininger, *Angewandte Chemie International Edition,* **2023**, DOI 10.1002/anie.202312476.

[30] T. A. Holak, S. K. Kearsley, Y. Kim, J. H. Prestegard, *Biochemistry* **1988**, *27*, 6135–6142.

[31] J. Crosby, M. P. Crump, *Nat Prod Rep* **2012**, *29*, 1111–1137.

[32] J. Beld, D. J. Lee, M. D. Burkart, *Mol. BioSyst* **2015**, *11*, 38.

[33] P. W. MAJERUS, A. W. ALBERTS, P. R. VAGELOS, *PNAS,* **1964**, *51*, 1231–1238.

[34] J. Elovson, P. R. Vagelos, *Journal of Biological Chemistry,* **1968**, *243*, 3603–3611.

[35] S. Smith, A. Witkowski, A. K. Joshi, *Prog Lipid Res,* **2003**, *42*, 289–317.

[36] S. W. White, J. Zheng, Y.-M. Zhang, C. O. Rock, *Annu Rev Biochem*, **2005**, *74*, 791-831.

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

[38] M. Hashimoto, T. Nonaka, I. Fujii, *Nat Prod Rep,* **2014**, *31*, 1306–1317.

[39] Y. Seshime, P. R. Juvvadi, K. Kitamoto, Y. Ebizuka, T. Nonaka, I. Fujii, *Bioorg Med Chem Lett,* **2010**, *20*, 4785–4788.

[40] Y. Seshime, P. R. Juvvadi, I. Fujii, K. Kitamoto, *Biochem Biophys Res Commun,* **2005**, *331*, 253–260.

[41] I. Soehano, L. Yang, F. Ding, H. Sun, Z. J. Low, X. Liu, Z. X. Liang, *Org Biomol Chem,* **2014**, *12*, 8542–8549.

[42] A. Watanabe, Y. Ebizuka, *Chem Biol*, **2004**, *11*, 1101–1106.

[43] P. Chang, J. W. Cary -J Yu -D Bhatnagar, T. E. Cleveland, *Mol Gen Genet*, **1995**, *248*(3), 270-7.

[44] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature,* **2009**, *461*, 1139–1143.

[45] J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *PNAS*, **2006,** *103* (45) 16728-16733.

[46] J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher, C. A. Townsend, *Science,* **2008**, *320*, 243-246.

[47] A. M. Bailey, R. J. Cox, K. Harley, C. M. Lazarus, T. J. Simpson, E. Skellam, *Chemical Communications,* **2007**, 4053-4055.

[48] N. Schracke, U. Linne, C. Mahlert, M. A. Marahiel, *Biochemistry,* **2005**, *44*, 8507-8513.

[49] N. Gaitatzis, B. Kunze, R. Mü, *PNAS*, **2001,** 98 (20) 11136-11141.

[50] K. M. Fisch, E. Skellam, D. Ivison, R. J. Cox, A. M. Bailey, C. M. Lazarus, T. J. Simpson, *Chemical Communications,* **2010**, *46*, 5331-5333.

[51] I. Fujii, Y. Ono, H. Tada, K. Y. Gomi Ebizuka, U. Sankawa, Y. Ebizuka, K. Gomi, *Molecular and General Genetics MGG*, **1996,** 253, 1-10.

[52] J. S. Parascandolo, J. Havemann, H. K. Potter, F. Huang, E. Riva, J. Connolly, I. Wilkening, L. Song, P. F. Leadlay, M. Tosin, *Angewandte Chemie,* **2016**, *128*, 3524–3528.

[53] J. Hitschler, E. Boles, *Metab Eng Commun,* **2019**, *9*, DOI 10.1016/j.mec.2019.e00093.

[54] T. Moriguchi, Y. Kezuka, T. Nonaka, Y. Ebizuka, I. Fujii, *Journal of Biological Chemistry,* **2010**, *285*, 15637–15643.

[55] Y. Li, Y. Shen, X. Zhu, L. Du, *J Nat Prod,* **2009**, *72*, 1328–1330.

[56] B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Chemical Communications,* **2016**, *52*, 6777–6780.

[57] S. M. Ma, Y. Tang, *FEBS Journal*, **2007**, *274*, 2854–2864.

[58] K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus, T. J. Simpson, *ChemBioChem* **2007**, *8*, 289–297.

[59] E. Skellam, *Nat Prod Rep* **2021**, *39*, 754–783.

[60] J. Barrios-González, & Ailed Pérez-Sánchez, & María, E. Bibián, *Applied Microbiology and Biotechnology,* **2020**, *104*, 8979–8998**.**

[61] D. J. Witter, J. C. Vederas, *J. Org. Chem*, **1996**, *61*, 2613–2623.

[62] W. L. Kelly, *Org Biomol Chem* **2008**, *6*, 4483–4493.

[63] C. R. Hutchinson, J. Kennedy, C. Park, S. Kendrew, K. Auclair, J. Vederas, *Antonie van Leeuwenhoek*, **2000**, *78*, 287–295.

[64] X. Xie, K. Watanabe, W. A. Wojcicki, C. C. C. Wang, Y. Tang, *Chem Biol,* **2006**, *13*, 1161-1169.

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

[66] H. Ogawa, K. Hasumi, K. Sakai, S. Murakawa, A. Endo, *The Journal of Antibiotics*, **1991**, 762-767.

[67] K. Hasumi, K. Tachikawa, K. Sakai, S. Murakawa, N. Yoshikawa, S. Kumazawaf, A. Endo, *The Journal of Antibiotics*, **1993**, 689-691.

[68] J. D. Bergstrom, M. M. Kurtz, D. J. Rew, A. M. Amend, J. D. Karkas, PNAS, **1993**, *90* (1) 80-84

[69] J. L. Goldstein, M. S. Brown, *Nature*, **1990**, *343 (6257)*, 425-30.

[70] C. Hertweck, B. S. Moore, *Journal of Bacteriology,* **2003**, *185*, 399-404**.**

[71] A. V. Qualley, J. R. Widhalm, F. Adebesin, C. M. Kish, N. Dudareva, *Proc Natl Acad Sci U S A* **2012**, *109*, 16383–16388.

[72] E. J. Skellam, D. Hurley, J. Davison, C. M. Lazarus, T. J. Simpson, R. J. Cox, *Mol Biosyst,* **2010**, *6*, 680–682.

[73] K. E. Lebe, R. J. Cox, *Chem Sci,* **2019**, *10*, 1227–1231.

[74] H. L. Ai, X. Lv, K. Ye, M. X. Wang, R. Huang, B. B. Shi, Z. H. Li, *Journal of Fungi,* **2022**, *8*, 492.

[75] A. Bradfield, A. Penfold, J. S.-J, Simonsen*, Journal of the Chemical Society,* **1932.**

[76] R. Ebel, *Mar. Drugs,* **2010**, *8*, 2340–2368.

[77] C. Hou, M. Kulka, J. Zhang, Y. Li. F. Guo, *Mini Reviews in Medicinal Chemistry,* **2014,** *14*, 664-677.

[78] K. T. Yuyama, D. Fortkamp, W. R. Abraham, *Biol Chem,* **2017**, *399*, 13–28.

[79] Y. Moule, S. Moreau, J.F. Bousquet. *Chemico-Biological Interactions*, **1977**, *17,* 185-192**.**

[80] S. Tanaka, K. Wada, *Agricultural and Biological Chemistry,* **1984**, *48*, 3189.

[81] Y. Yao, M. Hausding, G. Erkel, T. A.-M, *Molecular Pharmacology,* **2003.**

[82] C. Daengrot, V. Rukachaisirikul, C. Tansakul, T. Thongpanchang, S. Phongpaichit, K. Bowornwiriyapan, J. Sakayaroj, *ACS Publications,* **2015**, *78*, 615–622.

[83] ZY Hu, YY Li, YJ Huang, WJ Su, YM Shen*, Helvetica Chimica Acta,* **2008**, *91*, 46–52.

[84] C. M. De Oliveira, G. H. Silva, L. O. Regasini, O. Flausino, S. N. López, B. M. Abissi, R. G. D. S. Berlinck, L. D. Sette, R. C. Bonugli-Santos, A. Rodrigues, V. D. S. Bolzani, A. R. Araujo, *J Nat Prod,* **2011**, *74*, 1353–1357.

[85] Y. Che, J. B. Gloer, D. T. Wicklow, *J Nat Prod,* **2002**, *65*, 399–402.

[86] T. Hatakeyama, T. Koseki, T. Murayama, *Phytochemistry Letters*, **2010**, *3,* 148-151**.**

[87] V. Mierau, T. Anke, O. Sterner, *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* **2003**, *58*, 541–546.

[88] D. Zhang, H. Ge, J. H. Zou, X. Tao, R. Chen, J. Dai, *Org Lett,* **2014**, *16*, 1410–1413.

[89] J. Liu, D. Zhang, M. Zhang, J. Zhao, R. Chen, N. Wang, D. Zhang, J. Dai, *J Nat Prod,* **2016**, *79*, 2229–2235.

[90] Y. Liu, Y. Li, J. Qu, S. Ma, C. Zang, Y. Zhang, S. Yu, *J Nat Prod,* **2015**, *78*, 2149–2154.

[91] R. Riclea, J. S. Dickschat, *Angewandte Chemie - International Edition,* **2015**, *54*, 12167–12170.

[92] A. Ilazi, R Liffert, K Gademann*, Helvetica Chimica Acta,* **2018**, *101*.

[93] F. Sugawara, G. Strobel, L. E. Fisher, G. D. Van Duyne, J. Clardy, *Proc Natl Acad Sci U S A* **1985**, *82*, 8291–8294.

[94] D Malviya, UB Singh, B Dehury, P Singh, M Kumar, S Singh, A Chaurasia, *Antioxidants,* **2022**, DOI 10.3390/antiox11091754.

[95] L Wang, M Li, J Tang, X Li, *Molecules*, **2016***, 21(4),* 473.

[96] K Motohashi, J Hashimoto, S Inaba, ST Khan, H Komaki, A Nagai, M Takagi, K Shin-ya*, The Journal of Antibiotics,* **2009**, *62*, 247–250.

[97] Z. Cheng, J. Zhao, D. Liu, P. Proksch, Z. Zhao, W. Lin, *ACS Publications* **2016**, *79*, 45.

[98] Y. Shiono, T. Murayama, *Zeitschrift fur Naturforschung - Section B Journal of Chemical Sciences* **2005**, *60*, 885–890.

[99] A Tchoukoua, T Suzuki, NR Ariefta, T Koseki, Y Okawa, K Kimura, Y Shiono*, The Journal of Antibiotics,* **2017**, *70*, 1129–1132.

[100] L. A. McDonald, L. R. Barbieri, V. S. Bernan, J. Janso, P. Lassota, G. T. Carter, *J Nat Prod,* **2004**, *67*, 1565–1567.

[101] S. B. Singh, P. Felock, D. J. Hazuda, *Bioorg Med Chem Lett* **2000**, *10*, 235–238.

[102] D. C. J. Waalboer, H. A. Van Kalkeren, M. C. Schaapman, F. L. Van Delft, F. P. J. T. Rutjes, *Journal of Organic Chemistry,* **2009**, *74*, 8878–8881.

[103] M. L. Garduño-Ramírez, A. Trejo, V. Navarro, R. Bye, E. Linares, G. Delgado, *J Nat Prod* **2001**, *64*, 432–435.

[104] E. Kuhnert, F. Surup, V. Wiebach, S. Bernecker, M. Stadler, *Phytochemistry*, **2015,** *117*, 116-122.

[105] R. H. Summerville, C. A. Senkler, P. R. Schleyer, T. E. Dueber, P. J. Stang, F. Yuste, E. Diaz, F. Walls, K. Jankowski Syntex, *Journal of Organic Chemistry,* **1976**, *41*, 4103–4106.

[106] J. ; Zhang, D. ; Liu, A. ; Fan, J. ; Huang, W. Lin, G. Wang, H. Dai, J. Zhang, D. Liu, A. Fan, J. Huang, W. Lin, *Marine drugs,* **2022**, *20*(11), 712.

[107] A. Hou, J. S. Dickschat, *Nat Prod Rep,* **2022**, *40*, 470–499.

[108] R. Schoenheimer, D. Rittenberg, *Science (1979),* **1935**, *82*, 156–157.

[109] T. J. Simpson, *Nat Prod Rep* **1987**, *4*, 339–376.

[110] G. F. Bills, F. Peláez, J. D. Polishook, M. T. Diez-Matas, G. H. Harris, W. H. Clapp, C. Dufresne, K. M. Byrne, M. Nallin-Omstead, R. G. Jenkins, M. Mojena, L. Huang, J. D. Bergstrom, *Mycol Res,* **1994**, *98*, 733–739.

[111] J. S. Dickschat, *European J Org Chem,* **2017**, *2017*, 4872–4882.

[112] B. Felicetti, D. E. Cane, *J Am Chem Soc,* **2004**, *126*, 7212–7221.

[113] R. Proctor, T.M. Hohn. *Journal of Biological Chemistry*, **1993**, *268*, 4543-4548**.**

[114] B. Felicetti, D. E. Cane, *J Am Chem Soc,* **2004**, *126*, 7212–7221.

[115] M. J. Calvert, P. R. Ashton, R. K. Allemann, *J Am Chem Soc,* **2002**, *124*, 11636–11641.

[116] D. E. Cane, H. J. Ha, C. Pargellis, F. Waldmeier, S. Swanson, P. P. Murthy, D. E. N ; Cane, R. Croteau, *J. Am. Chem. Soc,* **1989,** *111***,** 8914-8916.

[117] David E. Cane, P. C. Prabhakaran, John S. Oliver, and Douglas B, *J. Am. Chem. Soc*, **1990**, *112*, 8, 3209–3210.

[118] K. Tsukada, S. Shinki, A. Kaneko, K. Murakami, K. Irie, M. Murai, H. Miyoshi, S. Dan, K. Kawaji, H. Hayashi, E. N. Kodama, A. Hori, E. Salim, T. Kuraishi, N. Hirata, Y. Kanda, T. Asai, *Nat Commun,* **2020**, *11*, 1830.

[119] E. Kim, B. S. Moore, Y. J. Yoon, *Nat Chem Biol,* **2015**, *11*, 649–659.

[120] K. Tagami, C. Liu, A. Minami, M. Noike, T. Isaka, S. Fueki, Y. Shichijo, H. Toshima, K. Gomi, T. Dairi, H. Oikawa, *J Am Chem Soc,* **2013**, *135*, 1260–1263.

[121] M. Okada, K. Saito, C. P. Wong, C. Li, D. Wang, M. Iijima, F. Taura, F. Kurosaki, T. Awakawa, I. Abe, *Org Lett*, **2017**, *19*, 3183–3186.

[122] X. Meng, Y. Fang, M. Ding, Y. Zhang, K. Jia, Z. Li, J. Collemare, W. Liu, *Biotechnol Adv* **2022**, *54*, 107866.

[123] R. Fujii, A. Minami, T. Tsukagoshi, N. Sato, T. Sahara, S. Ohgiya, K. Gomi, H. Oikawa, *Biosci Biotechnol Biochem,* **2011**, *75*, 1813–1817.

[124] C. Schotte, L. Li, D. Wibberg, J. Kalinowski, R. J. Cox, *Angewandte Chemie - International Edition,* **2020**, *59*, 23870–23878.

[125] T. Itoh, K. Tokunaga, Y. Matsuda, I. Fujii, I. Abe, Y. Ebizuka, T. Kushiro, *Nat Chem,* **2010**, *2*, 858–864.

[126] Y. Matsuda, T. Wakimoto, T. Mori, T. Awakawa, I. Abe, *J Am Chem Soc,* **2014**, *136*, 15326–15336.

[127] C. J. B Harvey, M. Tang, U. Schlecht, J. Horecka, C. R. Fischer, H.-C. Lin, J. Li, B. Naughton, J. Cherry, M. Miranda, Y. Fuga Li, A. M. Chu, J. R. Hennessy, G. A. Vandova, D. Inglis, R. S. Aiyar, L. M. Steinmetz, R. W. Davis, M. H. Medema, E. Sattely, C. Khosla, R. P. St Onge, Y. Tang, M. E. Hillenmeyer, *Science Advances*, **2018**, *4*, 4.

[128] L. Kahlert, E. F. Bassiony, R. J. Cox, E. J. Skellam, *Angewandte Chemie - International Edition,* **2020**, *59*, 5816–5822.

[129] M. N. Heneghan, A. A. Yakasai, L. M. Halo, Z. Song, A. M. Bailey, T. J. Simpson, R. J. Cox, C. M. Lazarus, *ChemBioChem,* **2010**, *11*, 1508–1512.

[130] K. Pahirulzaman, K. Williams, C. Lazarus, *Methods in enzymology*, **2012**, *517,* 241-260**.**

[131] R. Li, Z. X. Li, K. Ma, G. Wang, W. Li, H. W. Liu, W. B. Yin, P. Zhang, X. Z. Liu, *Sci China Life Sci,* **2019**, *62*, 1087–1095.

[132] KE Lebe, RJ Cox*, Chemical science,* **2019,** 10, 1227-1231**.**

[133] B Bonsch, V Belt, C Bartel, N Duensing, M Koziol, CM Lazarus, AM Bailey, TJ Simpson, *Chemical Communications,* **2016***, 52*, 6777-6780**.**

[134] C. A. Jones, P. J. Sidebottom1, R. J. P. Cannell, D. Noble, B. A. M. Rudd, *The Journal of Antibiotics*, **1992**, *45*, 1492-1498.

[135] J. Wang, J. Liang, L. Chen, W. Zhang, L. Kong, C. Peng, C. Su, Y. Tang, Z. Deng, Z. Wang, *Nat Commun* **2021**, *12*.

[136] R. J. Cox, *Nat Prod Rep* **2022**, *40*, 9–27.

[137] RJ Cox, F Glod, D Hurley, CM Lazarus, TP Nicholson, BAM Rudd, TJ Simpson*, Chemical Communications,* **2004,** 2260-2261**.**

[138] D. Ivison, **2013**.

[139] DM Roberts, C Bartel, A Scott, D Ivison, TJ Simpson, RJ Cox*, Chemical Science,* **2017**, *8*, 1116-1126.

[140] O Piech, RJ Cox*, RSC Advances,* **2020**, *10*, 18469-18476.

[141] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, *Nucleic Acids Res* **2003**, *31*, 3381–3385.

[142] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, *J Chem Inf Model* **2021**, *61*, 3891–3898.

[143] E. Krieger, G. Vriend, *Bioinformatics* **2014**, *30*, 2981–2982.

[144] XL Yang, S Friedrich, S Yin, O Piech, K Williams, TJ Simpson, RJ Cox*, Chemical Science,* **2019,** *10*, 8478-8489**.**

[145] W. Xu, Y.-H. Chooi, J. W. Choi, S. Li, J. C. Vederas, N. A. Da Silva, Y. Tang, *Angewandte Chemie,* **2013**, *125*, 6600–6603.

[146] A. O. Zabala, Y. H. Chooi, M. S. Choi, H. C. Lin, Y. Tang, *ACS Chem Biol,* **2014**, *9*, 1576–1586.

[147] B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Chemical Communications,* **2016**, *52*, 6777–6780.

[148] N. Liu, Y. S. Hung, S. S. Gao, L. Hang, Y. Zou, Y. H. Chooi, Y. Tang, *Org Lett,* **2017**, *19*, 3560–3563.

[149] P. E. Coudron, E. S. Moland, K. S. Thomson, *J Clin Microbiol*, **2000**, *38*, 1791-1796.

[150] M. M. Rahman, H. N. Hunter, S. Prova, V. Verma, A. Qamar, D. Golemi-Kotra, *mBio,* **2016**, *7*.

[151] K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, A. M. Bailey, C. M. Lazarus, T. J. Simpson, R. J. Cox, *ACS Publications* **2011**, *133*, 16635–16641.

[152] AM Bailey, F Alberti, S Kilaru, CM Collins, K de Mattos-Shipley, AJ Hartley, P Hayes, A Griffin*, Scientific reports,* **2016,** 25202**.**

[153] F Alberti, K Khairudin, JA Davies, S Sangmalee, CL Willis, GD Foster, AM Bailey*, Chemical Science,* **2023**, *14*, 3826-3833**.**

[154] N. J. Foy, S. V. Pronin, *J Am Chem Soc* **2022**, *144*, 10174–10179.

[155] R Chávez, I Vaca, C García-Estrada*, Journal of Fungi,* **2023**, *4*, 459**.**

[156] R.-D. Wei, P. E. Still, E. B. Smalley, H. K. Schnoes, F. M. Strong, *Appl Microbiol*,**1973**, *25*, 111–114.

[157] R. dong Wei, H. K. Schnoes, P. A. Hart, F. M. Strong, *Tetrahedron* **1975**, *31*, 109–114.

[158] M. K. Dubey, M. Aamir, M. S. Kaushik, S. Khare, M. Meena, S. Singh, R. S. Upadhyay, *Front Pharmacol,* **2018**, *9*.

[159] S. Moreau, A. Lablach-Combier, J. Biguet, *Appl Environ Microbiol* **1980**, *39*, 770–776.

[160] NL Brock, JS Dickschat, *ChemBioChem,* **2013**, *14*, 1189–1193.

[161] P. Hidalgo, R. Ullán, S. Albillos, O. Montero, *Fungal Genetics and Biology*, **2014,** *62***,** 11-24**.**

[162] C. P. Gorst-Allman, P. S. Steyn, *Tetrahedron Lett*, **1982**, *23*, 5359–5362.

[163] A. B. Chase, D. Sweeney, M. N. Muskat, D. G. Guillén-Matus, P. R. Jensen, *mBio,* **2021**, *12*.

[164] A. Miral, S. Ferron, I. Rouaud, D. Slyambayev, L. Bousarghin, C. Camuzet, S. Belouzard, K. Séron, P. Le Pogam, S. Tranchimand, S. Tomasi, *J Nat Prod,* **2023**, *86*, 730–738.

[165] B. Felicetti, D. E. Cane, *J Am Chem Soc,* **2004**, *126*, 7212–7221.

[166] D. Wibberg, M. Stadler, C. Lambert, B. Bunk, C. Spröer, C. Rückert, J. Kalinowski, R. J. Cox, E. Kuhnert, *Fungal Divers,* **2021**, *106*, 7–28.

[167] E. Kuhnert, J. C. Navarro-Mu~ Noz, K. Becker, M. Stadler, J. Collemare, R. J. Cox, *Studies in Mycology,* **2021**, *99*, 100118.

[168] D. Wibberg, M. Stadler, C. Lambert, B. Bunk, C. Spröer, C. Rückert, J. Kalinowski, R. J. Cox, E. Kuhnert, *Fungal Divers,* **2021**, *106*, 7–28.

[169] M Stanke, M Diekhans, R Baertsch, D Haussler*, Bioinformatics,* **2008,** *24*, 637–644 **.**

[170] V. Ter-Hovhannisyan, A. Lomsadze, Y. O. Chernoff, M. Borodovsky, *Genome Res,* **2008***, 12* 1979-1990*,*

[171] F Meyer, A Goesmann, AC McHardy, D Bartels, T Bekel, JÈ Clausen, JÈ Kalinowski, B Linke, *Nucleic Acids Research,* **2003**, *31*, 2187-2195**.**

[172] O. Rupp, J. Becker, K. Brinkrolf, C. Timmermann, N. Borth, A. Pühler, T. Noll, A. Goesmann, *PLoS One,* **2014**, *9*.

[173] R. Proctor, T.M. Hohn, *Journal of Biological Chemistry*, **1993**, *268,* 4543-4548.

[174] D. Cane, I. Kana. *Archives of Biochemistry and Biophysics*, **2000**, *376,* 354-364**.**

[175] J. Caruthers, I. Kang, M. Rynkiewicz, *Journal of Biological Chemistry*, **2000**, *275,* 25533-25539.

[176] E. Y. Shishova, L. Di Costanzo, D. E. Cane, D. W. Christianson, *Biochemistry,* **2007**, *46*, 1.

[177] S. Habtemar-Iam, A. I. Gray, P. G. J. Waterman, I. Nat; Kubo, M. Himejima, T. K. Experientia; Devon, A. I. Scott, J. D. Connolly, R. A. Hill, L. Ruzicka, D. E. Experientia; Cane, D. E. Cane, T. M. Hohn, R. D. Plattner, S. Tanaka, R. H. Proctor, Z. Wu, *J. Am. Chem. Soc,* **1991**, *57*, 7212–7221.

[178] A. Miral, S. Ferron, I. Rouaud, D. Slyambayev, L. Bousarghin, C. Camuzet, S. Belouzard, K. Séron, P. Le Pogam, S. Tranchimand, S. Tomasi, *J Nat Prod,* **2023**, *86*, 730–738.

[179] T. Srisapoomi, T. Ichiyanagi, H. Nakajima, T. Aimi, S. Boonlue, *Curr Microbiol,* **2015**, *70*, 550–555.

[180] G. A. Payne, W. C. Nierman, J. R. Wortman, B. L. Pritchard, D. Brown, R. A. Dean, D. Bhatnagar, T. E. Cleveland, M. Machida, J. Yu, *Med Mycol*, **2006**, *44*, 9–11.

[181] Z. Y. Huang, Q. Y. Wu, C. X. Li, H. L. Yu, J. H. Xu, *J Agric Food Chem,* **2022**, *2022*, 5868.

[182] Z. Cheng, J. Zhao, D. Liu, P. Proksch, Z. Zhao, W. Lin, *J Nat Prod,* **2016**, *79*, 1035–1047.

[183] E. Kuhnert, F. Surup, V. Wiebach, S. Bernecker, M. Stadler, **2015**, DOI 10.1016/j.phytochem.2015.06.002.

[184] H FUJIMOTO, Y SATOH, M NAKAYAMA, T TAKAYAMA, M YAMAZAKI, *Chemical and Pharmaceutical Bulletin,* **1995***, 43,* 547-522**.**

[185] H Fujimoto, M Sumino, J Nagano, H Natori, E Okuyama, M Yamazaki*, Chemical and Pharmaceutical Bulletin,* **1999***, 47,* 71-76**.**

[186] Y Fujimoto, H Tsunoda, J Uzawa, T Tatsuno*, Journal of the Chemical Society, Chemical Communications,* 1982*,* 83-84.

[187] Y Shiono, M Yokoi, T Koseki, T Murayama, N Aburai, K Kimura, *The Journal of Antibiotics,* **2010***, 63,* 251–253**.**

[188] M.S.R. Nair, *Phytochemistry*, **1976**, *15,* 1090-1091.

[189] M. NAIR, C. ST, *Tetrahedron Letters,* **1975,** *41*, 3517-3518.

[190] T. Shimizu, S. Hiranuma, T. Watanabe, M. Kirihara, *Heterocycles,* **1994**, *38*, 243–248.

[191] S. Yuan, L. Chen, Q. Wu, M. Jiang, H. Guo, Z. Hu, S. Chen, L. Liu, Z. Gao, *Mar Drugs,* **2022**, *20*, 294.

[192] K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, I. Fujii, *ChemBioChem,* **2010**, *11*, 1245–1252.

[193] L. Kahlert, C. Schotte, R. J. Cox, *Synthesis (Germany),* **2021**, *53*, 2381–2394.

[194] J. Kawabata, E. Fukushi, J. Mizutani, R. P. Walker, D. H. Faulkner, D. V Engen, J. Clardy, W. M. Bandaranayake, J. E. Banfield, D. S. C. Black, D. E. Cane, W. Tan, W. R. Ott, J. F. Sanz-Cervera, T. Glinka, R. Williams, H. Oikawa, Y. Murakami, A. Ichihara, J. E. Baldwin, R. C. Whitehead, A. Ayukawa, T. Nomura, S. Ueda, T. Naturwis-Senschaften, *J. Am. Chem. Soc,* **1994,** 600-602.

[195] K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, I. Fujii, *ChemBioChem,* **2010**, *11*, 1245–1252.

[196] L. Kahlert, C. Schotte, R. J. Cox, *Synthesis (Germany),* **2021**, *53*, 2381–2394.

[197] R. Fujii, T. Ugai, H. Ichinose, M. Hatakeyama, T. Kosaki, K. Gomi, I. Fujii, A. Minami, H. Oikawa, *Biosci Biotechnol Biochem,* **2016**, *80*, 426–431.

[198] W. Chen, Y. Feng, I. Molnár, F. Chen, *Nat Prod Rep,* **2019**, *36*, 561–572.

[199] J. Kawabata, E. Fukushi, J. Mizutani, R. P. Walker, D. H. Faulkner, D. V Engen, J. Clardy, W. M. Bandaranayake, J. E. Banfield, D. S. C. Black, D. E. Cane, W. Tan, W. R. Ott, J. F. Sanz-Cervera, T. Glinka, R. Williams, H. Oikawa, Y. Murakami, A. Ichihara, J. E. Baldwin, R. C. Whitehead, A. Ayukawa, T. Nomura, S. Ueda, T. Naturwis-Senschaften, *J. Am. Chem. Soc*, **1994,** 600-602.

[200] D. S. Tian, E. Kuhnert, J. Ouazzani, D. Wibberg, J. Kalinowski, R. J. Cox, *Chem Sci,* **2020**, *11*, 12477–12484.

[201] F Sklenář, K Glässnerová, Ž Jurjević, J Houbraken, RA Samson, CM Visagie, N Yilmaz*, Studies in Mycology,* **2022**, *102*, 53–93.

[202] K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, I. Fujii, *ChemBioChem,* **2010**, *11*, 1245–1252.

[203] T. Robbins, J. Kapilivsky, D. E. Cane, C. Khosla, *Biochemistry,* **2016**, *55*, 4476–4484.

[204] W. Cheng, M. Chen, M. Ohashi, Y. Tang, *Angewandte Chemie - International Edition,* **2022**, *61*, DOI 10.1002/anie.202116928.

[205] L. Hang, M.-C. Tang, C. J. B. Harvey, C. G. Page, J. Li, Y.-S. Hung, N. Liu, M. E. Hillenmeyer, Y. Tang, *Angewandte Chemie,* **2017**, *129*, 9684–9688.

[206] W. Cheng, M. Chen, M. Ohashi, Y. Tang, *Angewandte Chemie - International Edition,* **2022**, *61*, DOI 10.1002/anie.202116928.

[207] V. Solovyev, P. Kosarev, I. Seledsov, D. Vorobyev, *Genome Biol,* **2006**, *7 Suppl 1*, DOI 10.1186/GB-2006-7-S1-S10.

[208] F J Jin, J Maruyama, PR Juvvadi, M Arioka, K Kitamoto, *FEMS microbiology letters,* **2004,** *239*, 79-85.

[209] T. Schafhauser, D. Wibberg, C. Rückert, *Journal of Biotechnology*, **2015**, *211,* 101-102**.**

[210] A. Macabeo, A. G. Patrick, A. C. Jean, L. E. Agustin, K. M. Yasmin, *Phytochemistry Letters*, **2020**, 35, 147-151.

[211] S Yuan, L Chen, Q Wu, M Jiang, H Guo, Z Hu, S Chen, L Liu, Z Gao, *Marine Drugs,* **2022**, *294*, 20(5)**.**

[212] H Fujimoto, M Sumino, J Nagano, H Natori, E Okuyama, M Yamazaki , *Chemical and Pharmaceutical Bulletin*, **1999**, 263-8522**.**

Appendix

Multiple alignment of SQTKS with FAS and TENS.

```
                    AT                                                                                AT
MF_SQTKS    702     GEIAAAYAAGALDFRSAMAVTYFRGEVGLACQDKIVGK----GGMIAVGLGPEEAED--RIARVQSGKIVIACINSQSSV  775
C2_SQTKS    700     GEIAAAYAAGALDFKSAMAVTYFRGEVGLACQDKIVGK----GGMIAVGLGPEDAED--RIARVQSGKIVVACINSQSSV  773
MFAS        583     GEVACGYADGCLSQREAVLAAYWRGQ---CIKDAHLPP----GSMAAVGLSWEECKQ--RCP----AGVVPACHNSEDTV  649
PFAS        583     GEVACGYADGCLTQEEAVLSSYWRGY---CIKEANVLP----GAMAAVGLSWEECKQ--RCP----PGIVPACHNSKDTV  649
TENS        686     GEIAATYASGIINLKAAMQIAYYRGL----YAKLARGQSDEAGGMMAAGLSMDDAVKLCRLPEFE-GRIQVAASNAPQSV  760

                    AT                                                                                AT
MF_SQTKS    776     TVSGDLAGIVELEEGLKAEGVFARRVKV-QAAYHSHHMQVIANGYLTSLKDI---LK-PGKKFGE---IIYSSPTTGK--  845
C2_SQTKS    774     TVSGDLSGIVELEDLLKAEGVFARRVKV-QAAYHSHHMQVIANGYLTSLKDM---LK-PTKKFGK---IIYSSPTTGR--  843
MFAS        650     TISGPQAAVNEFVEQLKQEGVFAKEVRTGGLAFHSYFMEGIAPTLLQALKKV---IREPRPRSAR---WLSTSIPEAQ--  721
PFAS        650     TISGPQAAMSEFLQQLKREDVFVKEVRTGGIAFHSYFMESIAPTLLRQLRKV---ILDPKPRSKR---WLSTSIPEAQ--  721
TENS        761     TLSGDKEAIKAAKAKLDADGVFARELKV-DTAYHSHHMLPCAEPYLKALLACDIQVSAPTKTPGRKCMWSSSVRGDAELL  839

                    AT                                                                                AT
MF_SQTKS    846     RETSAKLMASAQHWVNNMLSPVRFAESFQNMCFPTQKVSRSGELEQDVDIILEVGPHGMLQGPIQQMMSLPRFESARMPY  925
C2_SQTKS    844     RETNAKLMASAQHWVNNMLSPVRFAESFQNMCFSNRNSSQSEEIFQDVDIVLEVGPHGMLQGPIQQMMSLPIFERARLPY  923
MFAS        722     WQSSLARTSSAEYNVNNLVSPVLFQEALWHI--------------PEHAVVLEIAPHALLQAVLKRGVK------SSCTI  781
PFAS        722     WQGSLARTFSAEYSVNNLVSPVLFQEALQHV--------------PAHAVVVEIAPHALLQAVLKRSLE------SSCTI  781
TENS        840     RRDRNLDSLKGPYWVANMVQTVQFSRAIQSTIWHG----------GPFDLAVEVGPHPALKGPTEQTLKA--VYGSAPLY  907

                    AT                    AT
MF_SQTKS    926     LSCLLRGQSA-VYTMQSLAAGLMG---------WGYRVDMAA----VNFPQGTHGARILHDLPSYPWNHDNSHWWEPRLN  991
C2_SQTKS    924     ISCLLRGQSA-VHTMQTVAAGLMG---------WGYRVDMVA----VNFPQGTYGVKILHDLPSYPWNHDNSHWWEPRLN  989
MFAS        782     IPLMKRDHKDNLEFFLTNLGKVHL---------TGINVNPNALFPPVEFP-APRGTPLIS--PHIKWDHSQT-WDVPVAE  848
PFAS        782     IPLMKKDHRDNLEFFLSNVGRLHL---------AGVSVNPNGLFPPVEFP-APRGTPLIS--PHXKWDHSQA-WDVPSAA  848
TENS        908     TGVLSRGAND-AVAFSTAIGNIWSHLGPAFVDITGYQSIFSGTCE----GHGGSEAPFISDLPLYPWDHDEEYWRESRIS  982

                    DH                                                                                DH
MF_SQTKS    992     KAHRQRVHPPHDLLGSLIPGR-DLREPTWRHFIRVQDIPWIRDHVVQSQLVYPGAGFICMAIEAMVQLHDLKDSQSKKIA  1070
C2_SQTKS    990     KAHRQRVHPPHDLLGSLIVGR-DLREPTWRHFIRVQDIPWIRDHVVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVA  1068
MFAS        849     D------FPNGSSSSSATVYSIDASPESPDHYLV--------DHCIDGRVIFPGTGYLCLVWKTLARSLGLSLEETPVV-  913
PFAS        849     D------FPSGSSCSSVAVYKFDVSPESPDHYLV--------DHCIDGRVLFPGTGYLWLTWKTLARALSQNLEETPVV-  913
TENS        983     RRYRTGKDESHELLGRRMPDD-NEREIRWRNLLKVSELPWTQGHRVLGEVLLPGAAYISMAIEAGRR---LALDQGREVS  1058

                    DH                                                                                DH
MF_SQTKS    1071    GYRLADVDILRAMLIPDTSEG---------LEAHISLRPCST--KLLLTNEWYDFCVSSVGEDDKFVDHCRGRIAVEFNT  1139
C2_SQTKS    1069    GYRLAEVDILRAMLIPDTSEG---------LEAHISLRPCST--KLLLTNEWYDFCVSSVGDDDKFVDHCRGRITIEFDT  1137
MFAS        914     ---FENVSFHQATILPKTGTVA---LEVRLLEASHAFEVSDT-GNLIVSGKVYLW----EDPNSKL-----------FDH  971
PFAS        914     ---FEDVTLHQATILPKTGTVS---LEVRLLEASHAFEVSDSNGSLIASGKVYQW----ESPDPKL-----------FDT  972
TENS        1059    LLEVSDVDILRPVVVADNKEGTETLFTVRLLDEYASTGKKSD-ELMTASFSFYIY---NSPASTSIVHTCEGRIAVHLGA  1134

                    DH                                                                                DH
MF_SQTKS    1140    SSLSDAPKTTSRE---RSRGAGLTRSVDPSNLYSFLRAQGIYHGSIFQ-----NLKTISSRKNYSESSFVVADT---ASV  1208
C2_SQTKS    1138    SGSADTPRTSLRE---RSRSTGLMRSVDPSNLYSFLRAQGIYHGPIFQ-----NLKTISSRKDHSESSFVVANT---ASV  1206
MFAS        972     PEVPTPPESASVS---R---------LTQGEVYKELRLRGYDYGPQFQGICEATLEGEQGKLLWKDNWVTFMDTMLQVSI  1039
PFAS        973     RAAVDPADSTAEF---R---------LSQGDVYKDLRLRGYDYGPFFQLVLESDLEGNRGRLQWNDSWVSFLDAMLHMSI  1040
TENS        1135    KLGSEAAANSTPQLPPREPSVSNLQQLDCEKLYSVFETIGLEYSGAFRRIV--------------SSSRCLGHATATASW  1200

                    DH                                                                                DH
MF_SQTKS    1209    MPDGFQSAHVVHPTTLDSIFQGAYTALPSAGLDQ-KTAMIPRSIQEIYLSSALTSEVGQCLVSDTSLIRYDGQSFT-VNV  1286
C2_SQTKS    1207    MPNGFQSPHVIHPTTLDSIFQGAYTALPGAGLDQ-NTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFT-VNV  1284
MFAS        1040    LGSSQQSLQL--PTRVTAIY-----------IDP-ATHR-----QKVYRLKEDTQVADVTTSRCLGITVSGGIHISRLQT  1100
PFAS        1041    LAPGQLGLYL--PTRFTSIR-----------IDP-VTHR-----QKLYTLQDTTQAADVVVDRNLNTVVAGGALFLGAHS  1101
TENS        1201    PTADLNDCYLVHPAILDVAFQTIFVARAHPDSGQLSSALLPSRIERVRVVPSLAMGSKLQNNENFNAAI-DSWALNQTAS  1279

                    DH                    DH          CMeT                                    CMeT
MF_SQTKS    1287    GISSK-----ADSEC-TPVLE-IKGLRNQSVGQMAPQQGDSGNNDLCFKLEWALDISSVKQERLKEKFGFPLDPAEADII  1359
C2_SQTKS    1285    DVSSK-----ADSEH-TPVLE-IKGLRNQSVGQMAPQPGDSSNNDLCFKLDWAPDISSVKQERLKEKFGFPLDPTEADII  1357
MFAS        1101    TATSR-----RQQEQLVPTLEKFVFTPHMEAECLSESTALQKELQLCKGLARALQTKATQQ-------------------  1156
PFAS        1102    SVAPR-----RPQEHLKPILEKFCFTPHVESGCLAGNTALQEELQLCRGLAQALQTKVAQQ-------------------  1157
TENS        1280    SLTGNINVYDADSER---ALIQVEGFEVRAVG----EPDASKDRLLFYETVWGRDISIMGLSDPIRDE--TSDAMVQNLS  1350

                    CMeT                                                                              CMeT
MF_SQTKS    1360    MGLRQACLYYIRQALTSLTPSARDQLDWHQKRFYDWMMLQMHLAEEDRLAPNSSAWLQCTSSDEQKLLENVRAASVNGQM  1439
C2_SQTKS    1358    MGLRQACIHFIHRSLQSLTAPDRDQLDWHQKRFYDWMVLQIQLAEEDRLAPNSSAWLQCSSSDEQKLLENVRASSVNGQM  1437
```

```
MFAS        1157  -GLKAA-------------------------------MLGQE------DPPQHGL---------PRLLAAACQLQLNGNL  1189
PFAS        1158  -GLKMV-------------------------------VPGLDGAQAPREAPQQSL---------PRLLAAACQLQLNGNL  1196
TENS        1351  EAIERVSLFYVRQLMGELSTADRRQANWYHTRMLAAFDHHLAKVHEETHLHLRPEWL-ADDWTVIQTIDEAYPDAVELQM  1429

                  CMeT                                                                          CMeT
MF_SQTKS    1440  VVHVGESILAILRHEIAPLELMLQDKLLYRYYTDAIKWDRSYQQIDQLVKLHAHKCPSAKIIEIGAGTGGCTRAVLDALS  1519
C2_SQTKS    1438  VVHVGKSMLAILRHEIAPLELMLQDKLLYRYYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTGGCTRAVLDALS  1517
MFAS        1190  QLELGEAL--------AQERLLLPEDPLISGLLNSQALKACVD-----TALENLSTLKMKVAEVLAGEGHLYSRIPALLN  1256
PFAS        1197  QLELGQVL--------AQERPLLCDDPLLSGLLDAPALKACVD-----TALENMASPKMKVVEVLAGDGQLYSRIPALLN  1263
TENS        1430  LHAVGQNVADVIRGKKHLLEVLRVDNLLDRLYTEDKGMHMANLFLANALKEITFKFPRCKILEIGAGTGATTWAALSAIG  1509

                  CMeT                                                                          CMeT
MF_SQTKS    1520  THGAARCAQYDFTDVSSGFFEAAQQKFTAFADVIRFQKLDIEKDIETQG---------FECGSYDLVIASQVLHATGKIE  1590
C2_SQTKS    1518  NQGIARCAQYDFTDVSSGFFEAAQQKFAAFDDVIRFQKLDIEKDIEMQG---------FECGSYDLVIASQVLHATGKME  1588
MFAS        1257  TQPMLQ-LEYTATDRHPQALKDVQTK--------------LQQHDVAQGQWNPSDPAPSSLGALDLLVCNCALATLGDPA  1321
PFAS        1264  TQPVMD-LDYTATDRNPQALEAAQAK--------------LEQLHVTQGQWDPANPAPGSLGKADLLVCNCALATLGDPA  1328
TENS        1510  EA----FDTYTYTDLSVGFFENAVERFSAFRHRMVFRALDIEKDPASQS---------FDLNSYDIIIATNVLHATRNLG  1576

                  CMeT                                                                          CMeT
MF_SQTKS    1591  DTMANVRRLLKPGGKLLL-VETTRDEMDLQLVFGLLPGWWLSSEEERKMSPSLSTSSWEKVLKKTGFNGLDVELRDCDSD  1669
C2_SQTKS    1589  HTMANVRKLLKPGGKLLL-VETTRDEMDLQLVFGLLPGWWLSSEEERQMSPSLSTNSWEKVLKKTGFDGLDIELRDCDSD  1667
MFAS        1322  LALDNMVAALKEGG-FLL-VHTVLKGHALGETLACLPS------EVQPAPSLLSQEEWESLFSRKALHLVGLKR------  1387
PFAS        1329  VAVGNMAATLKEGG-FLL-LHTLLAGHPLGEMVGFLTS------PEQGGRHLLSQDQWESLFAGASLHLVALKR------  1394
TENS        1577  VTLGNVRSLLKPGGYLLLNEKTGPDSLRATFNFGGLEGWWLAEEKERQLSPLMSPDGWDAQLQKAQFSGVDHIVHDVQED  1656

                                YKR                                                         YKR
MF_SQTKS    1670  Q-----------------FYSFSVIMATASPTVPMNPVDFIILHGKSS-----------IPDQWMNDLRTATSPF-TKSD  1720
C2_SQTKS    1668  E-----------------FYSFSVMMATASSTIASSSMAFAIVYGEVP-----------LPDQFLDDMKTAISSS-AVSD  1718
MFAS        1388  S-----------------FYGTALFLCRRA--IPQEKPIFLSVEDTSF--------------QWVDSLKS-TLAT-SSSQ  1432
PFAS        1395  S-----------------FYGSVLFLCRQQ--TPQDSPVFLSVEDTSF--------------RWVDSLKD-ILAD-ASSR  1439
TENS        1657  QQDKQQNSMIMSQAVDDTFYARLSPLSEMANLLPMNEPLLIIGGQTTATLKMIKEIQKLLPRQWRHKVRLIASVNHLEAE  1736

                  YKR                                                                           YKR
MF_SQTKS    1721  PVVGHINNADPTGKFCIFLEDPEEDILFHPDEKSYASIKRVITQCKGLLWISRGGSMHGTLPTSSLKTGLLRTLRLEYAE  1800
C2_SQTKS    1719  PVVGHLDSIDATGKFCIFIEDPETDILSSPDEKSYASIQKLVTRCKGLIWVSRGGAMHGTRPNSSLKTGLLRTLRLEYTE  1798
MFAS        1433  PV-------------------------------------------------WLTAMDC-----PTSGV-VGLVNCLRKEPGG  1459
PFAS        1440  PV-------------------------------------------------WLMAVGC-----STSGV-VGMVNCLRKEPGG  1466
TENS        1737  GVPAHSNVI------CL-QELDRGLFTTAMTSKCLDALKTLFINTRNLLWVTNAQHSSSMTPRASMFRGITRVLDGEIPH  1809

                  YKR                                                                           YKR
MF_SQTKS    1801  KRF----ISLDLNPA-------------RAPWAHESISTIREVLRGALAQTAEIPIRDSEFAENDGQLYVPRISSDIARN  1863
C2_SQTKS    1799  KRF----ISLDLDSA-------------RPQWNHDSITTINEVLCGALAQNADSSIKDSEFAEQDGQLFVPRISCDIARN  1861
MFAS        1460  HRIRCILLSNLSNTS-------------HAPKLDPGSPELQQVLKHDLVMN-------------------------VYRD  1501
PFAS        1467  HRIRCVLVSNLSSTS-------------PAPEMHPSSSELQKVLQGDLVMN-------------------------VYRD  1508
TENS        1810  IRTQVLGIEPRATSSATARNLLEAFLRLRSDDGRHAANVDEDGADGSSQQVLWLHEPEAELLSN-GTMMIPRVKARKSLN  1888

                                                ER                                          ER
MF_SQTKS    1864  EALSSNSHSP--------AQTEPFHQPGKLLQMGIKTPGLIDTLQFSKTDAPDHLPADYIEIEPKAF--GLNFRDVMVAM  1933
C2_SQTKS    1862  EDLSSDSNSP--------AQMEPFHQPGKLLQMGIKTPGLIDTLQFSKTDATDNLPNDYIEIEPKAF--GLNFRDVMVAM  1931
MFAS        1502  GAWGAFRHFQ--------LEQDKPKEQTAHAFVNVLTRGDLASIRWVSSPLKHTQPSSSGAQLCTVYYASLNFRDIMLAT  1573
PFAS        1509  GAWGAFRHFP--------LEQDRPEKQTEHAFVNVLSRGDLSSIRWVCSPLHYALPASCQDRLCSVYYTSLNFRDVMLAT  1580
TENS        1889  DTYLASTRAISTTVDARCVSVQAVAGPAKMLLRPVEDFAVEHAISSQSTDSKVHIQVESTLHIPEAL-------------  1955

                  ER                                                                            ER
MF_SQTKS    1934  GQLEESIMGFECAGIVRRVGPSSAGHNIKVGDRVCALLGGQWTNT-VRVHWHAVAPIPQAMGWETAASIPIVFVTAYISL  2012
C2_SQTKS    1932  GQLEESIMGFECAGVVRRVGPSSAGHNIKVGDRVCALLGGQWTNT-VRVHWHSVAPIPQAMDWETAASIPIVFVTAYISL  2010
MFAS        1574  GKLSPDAIPGKWASRDCMLGMEFSGRD-RCGRRVMGLVPAEGLATSVLLSSDFLWDVPSSWTLEEAASVPVVYTTAYYSL  1652
PFAS        1581  GKLSPDSIPGKWLTRDCMLGMEFSGRD-ASGRRVMGMVPAEGLATSVLLLQHATWEVPSTWTLEEAASVPIVYTTAYYSL  1659
TENS        1956  --------------------------------DGTCLYLVCGWTRT-------AETSVPVIALSTSNASIVAVESKA-VAM  1996

                  ER                                                                            ER
MF_SQTKS    2013  VKIAKLQAKETVLIHAASGGVGQAAIILAKYAGAEIFATVGTEEKRELLIKEY-KIPDDHIFSSRNALFAKSIRQRTNGK  2091
C2_SQTKS    2011  VKIARMQAGETVLIHAASGGVGQAAIILAKHVGAEIFATVGTDEKRDLLIKEY-KIPDDHIFSSRNALFAKSIRQRTNGK  2089
MFAS        1653  VVRGRIQRGETVLIHSGSGGVGQAAISIALSLGCRVFTTVGSAEKRAYLQARFPQLDDTSFANSRDTSFEQHVLLHTGGK  1732
PFAS        1660  VVRGRMQPGESVLIHSGSGGVGQAAIAIALSRGCRVFTTVGSAEKRAYLQARFPQLDETCFANSRDTSFEQHVLRHTAGK  1739
TENS        1997  IDEADVKPETLFRVFQHMAMQALDSAVGRHGQGQSTALIYGADEELAKLTSERFAVRESKVYFASTRTSAPGDWLKVQPL  2076

                  ER                                                                            ER
```

```
MF_SQTKS    2092  GVDVVLNCLAGGLLQESFDCLADFGRFIEIGKRDIELNHCLNMGMFARSATFTAVDLIAIGRDRSYMVAEALPKVMALLQ  2171
C2_SQTKS    2090  GVDVVLNCLAGGLLQESFDCLADFGRFIEIGKRDIELNHCLNMGMFARSATFTAVDLIAIGRDRSYMFAEALPKIMTLLQ  2169
MFAS        1733  GVDLVLNSLAEEKLQASVRCLAQHGRFLEIGKFDLSNNHPLGMAIFLKNVTFHGILLDALFEEANDSWREVAALLKAGIR  1812
PFAS        1740  GVDLVLNSLAEEKLQASVRCLAQHGRFLEIGKFDLSNNHALGMAVFLKNVTFHGILLDSLFEEGGATWQEVSELLKAGIQ  1819
TENS        2077  LSKFALSQMMPADVEVFIDCLGDTESFDACRTLE---------SCLSTTSTVHRLDACLLSRMSQCSPDTLADAYSHAKT  2147

                  ER                                  ER                                KR              KR
MF_SQTKS    2172  QKAVRPVTPISIYKIGDIETAFRLMQAGKHMGKIVITAPEDAMVPVVTQPPKLQLR--------SDASYLIVGGLGGIGR  2243
C2_SQTKS    2170  EKAIRPVTPISIYKIGDIETAFRLMQAGKHMGKIVITAPEDAMVPVITRPPKLQLR--------PDASYLIVGGLGGIGR  2241
MFAS        1813  DGVVKPLK-CTVFPKAQVEDAFRYMAQGKHIGKVLVQVREEEPEAVLPGAQPTLISAISKTFCPAHKSYIITGGLGGFGL  1891
PFAS        1820  EGVVQPLK-CTVFPRTKVEAAFRYMAQGKHIGKVVIQVREEEQGPAPRGLPPIALTGLSKTFCPPHKSYVITGGLGGFGL  1898
TENS        2148  QSNAEFSWNGNVQTFTAAELAGKLSHSLMHS---VYMTDWQEKDSILVTVPPLQTRGLFK----SDRTYLMVGAAGGLGT  2220

                  KR                                                                              KR
MF_SQTKS    2244  SLCKNFVENGARSLVLLSRNANVSRQSGEFLDELRSTGCVVSVVDCDISNKTQVESTMLRLKEEKLPIRGIVHAGMVLQD  2323
C2_SQTKS    2242  SLCKNFVENGARSLVLLSRNANVSQQSGEFLDELRSTGCIVGVVDCDISSKTQVEATMLRLKKDMLPIRGIVHAGMVLQD  2321
MFAS        1892  ELARWLVLRGAQRLVLTSRSGIRTGYQAKHIREWRRQGIQVLVSTSNVSSLEGARALIAEATK-LGPVGGVFNLAMVLRD  1970
PFAS        1899  QLAQWLRLRGAQKLVLTSRSGIRTGYQARQVREWRRQGVQVLVSTSNASSLDGARSLITEATQ-LGPVGGVFNLAMVLRD  1977
TENS        2221  SICRWMVRNGARHVVVTSRN---PKADPEMLNEARRYGAAVKVVPMDACSKDCVQTVVDMIRDTMPPIAGVCNAAMVLRD  2297

                  KR                                                                              KR
MF_SQTKS    2324  SVFEHMTLEDYNTATRPKVRGSWNLHSALSDC--DLDFFIMLSSLAGVSGSASQANYTAGGAYQDALATYRRSRGLAAVS  2401
C2_SQTKS    2322  SVFERMSLDDYNTAIRPKVQGSWNLHSGLSDC--DLDFFIMLSSLAGVSGSASQANYTAGGAYQDALAKYRRAQGLSAVS  2399
MFAS        1971  AMLENQTPELFQDVNKPKYNGTLNLDRATREACPELDYFVAFSSVSCGRGNAGQTNYGFANSTMERICEQRRHDGLPGLA  2050
PFAS        1978  AVLENQTPEFFQDVSKPKYSGTANLDRVTREACPELDYFVIFSSVSCGRGNAGQANYGFANSAMERICEKRRHDGLPGLA  2057
TENS        2298  KLFLDMNVDHMNNVLGPKMQGTEHLDSIFAQE--PLDFFVLLSSSAAILNNTGQSNYHCANLYMDSLVTNRRSRGLAASI  2375

                  KR                                                                              KR
MF_SQTKS    2402  IDLGMVQSVGYVAETKGVAERLVRMGYSPISEMEVLKIVEHAIT-----NPPPETSSGQIITGIS--TKPGRHWTESSWL  2474
C2_SQTKS    2400  IDLGMVQSVGYVAETKGVAERLVRMGYSPISEMEVLKIVEHAIT-----NPPPEASSAQIITGIS--TKPGRHWTESSWL  2472
MFAS        2051  VQWGAIGDVGIVLEAMGTNDTVIG-GTLPQRISSCMEVLDLFLN-----QPHAVLSSFVLAEKKA--VAHGDGDTQRDLV  2122
PFAS        2058  VQWGAIGDVGVVLETMGTNDTVIG-GTLPQRIASCLEVLDLFLS-----QPHPVLSSFVLAEKKA--AAPRDGSSQKDLV  2129
TENS        2376  IHVGHVCDTGYVARLVDDSKVQMSLGTTRVMSVSETDVHHAFAEAVRGGQPDSRSGSHNIIMGIEPPTKPLDVAKRKPVW  2455

                  ACP                                                                             ACP
MF_SQTKS    2475  ----QDARFATLRERARDVKEQSNAQGGGQDKQIGAGQELSMATSLVEAIDVVGRAITAKLATMFLIAAESIIASKSLSE  2554
C2_SQTKS    2473  ----QDARFATLRERARDVKELSNSQGGAQDKQLAAGQELSMATSLVEAIDVVGRAITAKLATMFLIAAESIIASKSLSE  2552
MFAS        2123                                                        KAVAHILGIRDLAGINLDSTLAD  2202
PFAS        2130                                                        KAVAHILGIRDVASINPDSTLVP  2209
TENS        2456  ISDPRLGHMLPFSTLENQMVASEQAAASAADSLAQQVSEATTDEEAAAAALKGFATKLEGILLLPLGSIGEDSAGRPVTP  2535

                  ACP                                                ACP
MF_SQTKS          YGVDSLVAVELRNWLAAQLSSDVSVFDVTQSQSLTALATTVATKSSRIDKSLLVA    2605
C2_SQTKS          YGVDSLVAVELRNWLAAQLSSDVSVFDVTQSQSLTALATTVATKSSRIDKSLLVA    2603
MFAS              LGLDSLMGVEVRQILEREHDLVLPMREVRQLTLRKLQEMSSKTDSATDTTAPKSRSD  2202
PFAS              LGLDSLMGVEVRQILEREHDLVLSMREVRQLSLRKLQELSSKTSTDADPATPTSHED  2209
TENS              LGIDSLVAVEIRTWFLKQLRVDVPVMKILGGSTVGQ                       2571
```

Active sites are in green

Curriculum Vitae

Oct 2019 - present	**PhD student**, Organic Chemistry, Centre for Biomolecular Drug Research (BMWZ), Institute of Organic Chemistry, Leibniz Universität Hannover, Germany.
Sep 2016 - Jul 2019	**lixue shuoshi**, Organic Chemistry, Jiangxi Science and Technology Normal University, China.
Sep 2010 - May 2014	**Bachelor**, Pharmaceutical Engineering, Wannan Medical College, China.

List of Publications

[1] **Y. Sun**, J. Gerke, K. Becker, E. Kuhnert, B. Verwaaijen, D. Wibberg, J. Kalinowski, M. Stadler, R. Cox, *Chemical Science* **2023**.

[2] **Y. Sun**, D. Tian, E. Kuhnert, G. Le Goff, G. Arcile, J. Ouazzani, R. J. Cox, *Chemical Communications* **2023.**

[3] **Y. Sun**, Y. Niu, H. Huang, B. He, L. Ma, Y. Tu, V. T. Tran, B. Zeng, Z. Hu, *Frontiers in Microbiology* **2019**, 10.

[4] **Y. Sun**, Y. Niu, B. He, L. Ma, G. Li, V. T. Tran, B. Zeng, Z. Hu, *Journal of Microbiology and Biotechnology* **2019**, 29, 230–234.

[5] Z. Hu, H. Huang, **Y. Sun**, Y. Niu, W. Xu, Q. Liu, Z. Zhang, C. Jiang, Y. Li, B. Zeng, *Microorganisms* **2019**, 7.

[6] Z. Hu, G. Li, **Y. Sun**, Y. Niu, L. Ma, B. He, M. Ai, J. Han, B. Zeng, *Brazilian Journal of Microbiology* **2019**, 50, 43–52.

[7] Z. Hu, B. He, L. Ma, **Y. Sun**, Y. Niu, B. Zeng, *Indian J Microbiol* **2017**, 57, 270–277.