

Biosynthesis of Fungal Alkyl Citrates and Polyketides

Von der Naturwissenschaftlichen Fakultät
Gottfried Wilhelm Leibniz Universität Hannover

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

genehmigte Dissertation

von

Sen Yin, Master (China)

2020

Referent: Prof. Dr. Russell Cox

Korreferent: Prof. Dr. Andreas Kirschning

Tag der Promotion: 02.10.2020

Abstract

The main focus of the presented work concentrated on understanding the biosynthetic pathway of viridofungin, exploring the enzymatic properties of the key enzymes from the byssochlamic acid pathway, and engineering the PKS of tenellin. In a combined genetic and chemical approach, the biosynthetic pathway of viridofungin and functions of enzymes from of byssochlamic acid pathway and programming of the tenellin PKS (TENS) system were elucidated.

Trichoderma viride MF5628 was firstly genome sequenced by Illumina and Oxford nanopore sequencing. Then, using a combination of targeted gene knockout and RNA interference-based silencing in the native organism, a novel unreported tRNA ligase like enzyme involved in the biosynthetic pathway was identified. A citrate synthase like enzyme and a tRNA ligase like enzyme were found to be essential to the biosynthesis of viridofungin.

The biosynthesis of byssochlamic acid was investigated by protein expression and *in vitro* study. A hydrolase, citrate synthase, two 2-methyl citrate dehydratases, two KSI and two PEBP proteins were expressed in *E. coli* or yeast, in addition, *in vitro* assay was carried out by using these proteins. During maleic anhydride monomer biosynthesis, a hydrolase can hydrolyze the hexaketide from ACP. Citrate synthase can use hexaketide CoA and oxaloacetate as substrates to form (2*S*, 3*R*) citrate. 2-Methyl citrate dehydratase takes (2*S*, 3*R*) citrate diastereoisomer as substrate to form a 2, 3-alkene product, and the product can spontaneously form a cyclised maleic anhydride monomer. Two KSI finally used maleic anhydride monomer to form the dimerised product byssochlamic acid. Two PEBP enzymes are not thought to be involved in catalysis of dimerization.

Based on KR domain swap experiment on TenS, chain-length programming in TenS was elucidated. Six KR domain subfragments were swapped with the homologous fragments from hexaketide desmethylbassianin synthase (DMBS) and three with the heptaketide militarinone synthase (MILS) resulted in the synthesis of different chain length of polyketide products. Particularly, the MILS KR domain swap resulted in the synthesis of penta, hexa and heptaketides. The results of these and previous experiments in our group are rationalised by considering the existence of competition between the C-MeT and KR domains.

Keywords: viridofungin, byssochlamic acid, biosynthesis, PKS, domain swap

Zusammenfassung

Das Hauptaugenmerk der vorgestellten Arbeit lag auf dem Verständnis des Biosynthesewegs von Viridofungin, der Untersuchung der enzymatischen Eigenschaften der Schlüsselenzyme aus dem Byssochlaminsäure-Weg und der Entwicklung der PKS von Tenellin. In einer Kombination aus genetischen und chemischen experimenten Ansatz wurden der Biosyntheseweg von Viridofungin und die Funktionen von Enzymen aus dem Byssochlaminsäure-Weg sowie die Programmierung des Tenellin-PKS (TENS) -Systems aufgeklärt.

Das *Trichoderma-virid* MF5628 wurde zunächst durch Illumina- und Oxford-Nanoporen-Sequenzierung genomsequenziert. Es wurde unter Verwendung einer Kombination aus gezielter Gen-Knockout- und RNA-Interference-basierter Stummschaltung im nativen Organismus ein neuer, nicht veröffentlicht Biosyntheseweg identifiziert, an dem eine tRNA-Ligase beteiligt ist. Es wurde festgestellt, dass ein Citrat-Synthase-ähnliches Enzym und ein tRNA-Ligase-ähnliches Enzym für die Biosynthese von Viridofungin essentiell sind.

Die Biosynthese von Byssochlaminsäure wurde durch Proteinexpression und *in-vitro*-Studie enukleiert. Eine Hydrolase, Citrat-Synthase, zwei 2-Methylcitrat-Dehydratase, zwei KSIs und zwei PEBPs-Proteine wurden in *E. coli* und Hefe exprimiert. Während der Maleinsäureanhydridmonomer Biosynthese kann eine Hydrolase die Hexaketidform vom ACP hydrolysieren. Die Citrat-Synthase kann Hexaketid-CoA und Oxalacetat als Substrat verwenden, um (2*S*, 3*R*)-Citrat zu bilden. 2-Methylcitrat-Dehydratase nimmt (2*S*, 3*R*)-Citrat-Diastereoisomer als Substrat, um ein 2, 3-Alken im Produkt zu bilden, und das Produkt kann ein spontan cyclisiertes Maleinsäureanhydridmonomer bilden. Zwei KSI verwenden schließlich Maleinsäureanhydridmonomer, um das dimerisierte Produkt Byssochlaminsäure zu bilden. PEBP-Enzyme sind nicht der Katalyse der Dimerisierung beteiligt.

Basierend auf dem KR-Domain-Swap-Experiment mit TenS wurde die Kettenlängenprogrammierung in TenS aufgeklärt. Sechs KR-Domänen-Subfragmente, die mit den homologen Fragmenten der Hexaketid-Desmethylbassianin-Synthase (DMBS) und drei mit der Heptaketid-Militarinon-Synthase (MILS) ausgetauscht wurden, führten zur Synthese unterschiedlicher Kettenlängen von Polyketid-Produkten. Insbesondere der MILS KR-Domänenaustausch führte zur Synthese von Penta, Hexa und Heptaketiden. Die Ergebnisse dieser und vorheriger Experimente in unserer Gruppe werden einbezogen, indem die Existenz einer Konkurrenz zwischen den CMeT- und KR-Domänen berücksichtigt wird.

Schlüsselwörter: viridofungin, byssochlamic acid, Biosynthese, PKS, domain-swap

Acknowledgement

Thanks, Prof. Russell Cox for giving me the opportunity to finish my PhD in this fantastic group. I appreciate your professional supervision and kind help throughout the last four years.

Thanks, all cooperation partners the CeBiTEC Bielefeld, especially Prof Jörn Kalinowski and Dr Daniel Wibberg; thanks Prof. Jianqiang Kong for yeast expression strain and vectors.

Thanks, BMWZ media kitchen team, especially to Katja; OCI colleague Dr Jörg Fohrer and Dr Gerald Dräger for their help with NMR and mass related matters.

Thanks, our group members (Liz, Dongsong, Chongqing, Steffen, Jin, Lei, Yunlong, Slawik, Erik, Hao, Carsten, Karen, Lukas, Oliver, Mary, Eman, Raissa, Verena ...).

Thanks, China Scholarship Council (CSC) for the foundation.

Thanks, my parents who always covered my back in my life.

Submitting this thesis is only possible because of you.

Abbreviations and Units

AA	amino acid	KR	β -ketoreductase
ACP	acyl carrier protein	Ψ KR	pseudo-ketoreductase
acetyl-CoA	acetyl-coenzyme A	LCMS	liquid chromatography mass spectrometry
AntiSMASH	antibiotic and Secondary Metabolite Analysis Shell	DAD	diode-array detection
AT	acyltransferase	MCoA	malonyl-CoA
<i>att</i>	site-specific attachment	MeOD	deuterated methanol
bp	base pair	MeOH	methanol
BLAST	basic local alignment search tool	MS	mass spectrometry
CDCl ₃	deuterated chloroform	MeT	methyltransferase
cDNA	copy deoxyribonucleic acid	MAT	malonyl/acetyl transferase
C-Met	C-methyltransferase	mFAS	mammalian fatty acid synthase
CoASH	coenzyme A	mRNA	messenger ribonucleic acid
CS	citrate synthase	NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
COSY	homonuclear correlation spectroscopy	NOESY	nuclear overhauser effect spectroscopy
dATP	deoxyadenosine triphosphate	NPRS	nonribosomal peptide synthetase
DH	dehydratase	NMR	nuclear magnetic resonance
DNA	deoxyribonucleic acid	nr-iPKS	nonreducing iterative PKS
ELSD	evaporative light scattering detector	ORF	open reading frame
ESI	electrospray ionization	<i>P_{amyB}</i>	AmyB promoter
ER	enoyl reductase	<i>P_{gpdA}</i>	GpdA promoter
FAS	fatty acid synthase	PCR	polymerase chain reaction
FAD	flavin adenine dinucleotide	PKS	polyketide synthase
gDNA	genomic DNA	pr-iPKS	partially reducing iterative PKS
HMBC	heteronuclear multiple bond correlation	RNA	ribonucleic acid
HRAM	high resolution accurate mass	RT	retention time
¹ H NMR	proton NMR	SAM	S-adenosyl-methionine
HPLC	high performance liquid chromatography	SAT	starter unit: ACP transacylase
hr-iPKS	highly reducing iterative PKS	SM	secondary metabolite
HRMS	high resolution molecular weight	TE	thioesterase
HSQC	heteronuclear single quantum coherence	tf	transformant
ITS	internal transcribed spacer	TIC	total ion current
iPKS	iterative PKS	UV	ultraviolet
KS	ketosynthase	WT	wild-type

Contents

Abstract	i
Zusammenfassung	ii
Acknowledgement	iii
Abbreviations and Units.....	iv
Chapter 1. Introduction and Background of the Projects	1
1.1 General overview of projects	1
1.2 Natural products, secondary metabolites and their major classes.	2
1.3 Polyketides and Fatty acids.....	4
1.3.1 Basic enzymology of polyketide and fatty acid biosynthesis.....	4
1.3.1.1 Acyl transfer.....	4
1.3.1.2 Condensation step.....	5
1.3.1.3 β -processing.....	5
1.3.1.4 Release.....	7
1.3.1.5 Programming of PKS.....	7
1.3.2 Classification of FAS and PKS.....	8
1.3.2.1 Different types of FAS and PKS.....	8
1.3.2.2 Type I iterative PKS and eukaryotic FAS.....	9
1.3.3 Protein structure of FAS and PKS.....	9
1.3.3.1 Fungal FAS.....	9
1.3.3.2 mFAS.....	10
1.3.3.3 hr-PKS.....	11
1.4 Nonribosomal peptide and PKS-NRPS compounds in fungi.....	12
1.4.1 Nonribosomal peptide and NRPS.....	12
1.4.2 PKS-NRPS compounds in fungi.....	13
1.5 The alkyl citrates – an overview.....	15
1.5.1 Maleidrides.....	16
1.5.2 Squalastatin.....	18
1.5.3 Viridiofungin.....	19
1.6 Biosynthesis of Maleidrides.....	20
1.6.1 Early steps for maleic anhydride monomers.....	21
1.6.2 Dimerization step of maleidrides.....	22

1.6.3 Tailoring steps of maleidrides.....	25
1.7 Method for biosynthetic studies.....	27
1.7.1 Genome sequencing.....	27
1.7.2. Bipartite knockout.....	27
1.7.3 CRISPR-Cas9 gene editing.....	28
1.7.4 Silencing.....	29
1.7.5 Heterologous expression.....	31
1.8 Overall aims.....	33
Chapter 2. Exploring the Biosynthetic Pathway of Viridiofungin	34
2.1 Background and aims of the project.....	34
2.1.1 Introduction	34
2.1.2 Project aims.....	39
2.2 Results.....	40
2.2.1 Optimization of production conditions.....	40
2.2.2 Viridiofungin purification and characterization	44
2.2.3 Genome sequencing, bioinformatics and development of a biosynthetic hypothesis	48
2.2.3.1 gDNA extration.....	48
2.2.3.2 ITS PCR.....	49
2.2.3.3 Genome sequencing and gene cluster prediction	49
2.2.3.3 Biosynthetic pathway prediction	54
2.2.4 Transformation condition testing	55
2.2.4.1 Antibiotic test.....	55
2.2.4.2 Preparation of <i>T. viride</i> protoplasts	56
2.2.4.3 Transformation of <i>T. viride</i>	58
2.2.4.4 Promoter exchange.....	59
2.2.5 Gene knockout	61
2.2.5.1 Citrate synthase knockout	61
2.2.5.2 tRNA ligase like gene knockout.....	67
2.2.5.3 KO of the tailoring genes in viridiofungin cluster.....	71
2.2.6 Attempted Gene silencing of TvR3.....	74
2.2.6.1 Vector construction.....	75
2.2.6.2 Transformation, LCMS analysis and PCR confirmation	76
2.2.7 Heterologous expression in <i>A. oryzae</i>	77

2.2.7.1 Plasmid construction.....	77
2.2.7.2 Transformation and LCMS analysis	79
2.3 Discussion.....	81
2.3.1 Oxidation of fatty acid.....	81
2.3.2 Citrate synthase in primary and secondary metabolism	82
2.3.3 tRNA ligase like enzyme in viridiofungin pathway	83
2.3.4 Tailoring genes in viridiofungin pathway	88
2.4 Conclusion and outlook.....	89
Chapter 3. Heterologous Expression of the Enzymes from Nonadride Biosynthetic Pathways and <i>in vitro</i> Assay.....	91
3.1. Background and aims of the project.....	91
3.1.1 Introduction	91
3.1.2 Project aims.....	97
3.2 Results	98
3.2.1 Protein production	98
3.2.1.1 Whole length expression in E. coli	98
3.2.1.2 Bioinformatic analysis	100
3.2.1.3 Removing the signal peptide for expression.....	101
3.2.1.4 Use of chaperones.....	103
3.2.1.5 Yeast expression.....	105
3.2.2 <i>In vitro</i> studies.....	109
3.2.2.1 In vitro study of citrate synthase	109
3.2.2.2 In vitro assay of 2-Methyl citrate dehydratase	118
3.2.2.3 Stereoselectivity of 2-methyl citrate dehydratase.....	120
3.2.2.4 In vitro study of dimerization step	124
3.2.2.5 Regrowing the byssochlamic acid biosynthetic pathway genes A. oryzae strains.....	124
3.2.2.6 Preparing the decarboxylated maleic anhydride monomer substrate and in vitro test	127
3.2.2.7 Preparing the maleic anhydride monomer substrate.....	131
3.2.2.8 In vitro study of yeast expressed KI and PEBP enzymes	132
3.2.2.9 In vitro study of E. coli expressed KI and PEBP enzymes	135
3.3 Discussion.....	136
3.3.1 The signal peptide at the N terminus of the proteins.....	136

3.3.2 The substrate and stereochemistry of citrate synthase	137
3.3.2.1 Substrate of CS	137
3.3.2.1 Stereochemistry of CS	138
3.3.3 Stereochemistry of dehydratase	139
3.3.4 Maleic anhydride monomer	140
3.3.5 Decarboxylation during the biosynthesis of maleidride	142
3.3.6 Dimerization	143
3.3.7 Function of PEBP enzymes	144
3.4 Conclusion and outlook	146
Chapter 4. Molecular Basis of Chain-Length Programming in a Fungal Iterative Highly Reducing Polyketide Synthase	148
4.1. Background and aims of the project	148
4.1.1 Introduction.	148
4.1.2 Aims of the project	152
4.2 Results.	153
4.2.1 Strategy for domain swaps.	153
4.2.2. DMBS KR sub-domain swap.	155
4.2.2.1. TDS1.	155
4.2.2.2. TDS2.	156
4.2.2.3 TDS3.	157
4.2.2.4 TDS4.	157
4.2.2.5 TDS5.	157
4.2.2.6 TDS6.	159
4.2.3 MILS KR full domain swap	160
4.2.3.1 Vector construction of TMS1.	160
4.2.3.2 Fungal transformation of TMS1 and chemical analysis.	162
4.2.4 MILS KR sub-domain swap.	163
4.2.4.1. TMS2.	164
4.2.4.3. TMS3	166
4.3. Discussion	168
4.3.1 Summary of the observations and attempt to explain the result using model structure	168
4.3.2 Domain competition hypothesis	169
4.3.3 <i>In vitro</i> study on hrPKS C-MeT and KR domains	171

4.4. Conclusion and outlook.....	173
Chapter 5. Materials and Methods	174
5.1 Solvents and chemicals	174
5.2 Growth media, buffers and solutions	174
5.3 Antibiotics and enzymes	175
5.4 Strains.....	176
5.5 Microbiology methods	177
5.5.1 <i>E.coli</i> growth and transformation	177
5.5.2 Yeast transformation and expression	177
5.5.2.1 Yeast transformation	177
5.5.2.1 Yeast expression.....	177
5.5.3 Transformation of <i>Aspergillus oryzae</i> NSAR1	178
5.5.4 Transformation of <i>Trichoderma viride</i> MF5628.....	178
5.6 Chemistry methods.....	179
5.6.1 <i>Trichoderma viride</i> extraction.....	179
5.6.2 Extraction of <i>Aspergillus oryzae</i> NSAR1	179
5.6.4 LCMS.....	180
5.6.4.1 Analytical LCMS.....	180
5.6.4.2 Semi-Preparative LCMS and Compound Purification.	180
5.6.5 High Resolution Mass Spectrometry (HRMS).....	180
5.6.6 Nuclear Magnetic Resonance (NMR) Analysis	180
5.7 Molecular biology methods	181
5.7.1 Oligonucleotides	181
5.7.2 DNA isolation and PCR	185
5.7.2.1 Isolation of Small Scale gDNA.....	185
5.7.2.2 Isolation of Large Scale gDNAs.....	185
5.7.2.3 Isolation of Small Scale DNA Plasmid from <i>E. coli</i>	185
5.7.2.4 Isolation of DNA Plasmid from <i>S. cerevisiae</i>	185
5.7.2.5 Analytical or Preparative PCR.....	185
5.7.2.6 Colony PCR or Analytical PCR.....	186
5.7.2.7 Restriction Analysis of Plasmid DNA	186
5.7.2.8 DNA Electrophoresis	186
5.7.3 Cloning procedure.....	186

Contents

5.7.3.1 Restriction enzyme digestion.....	186
5.7.3.2 Dephosphorylation and ligation.....	187
5.7.3.3 Gateway cloning.....	187
5.7.3 Obtained and constructed vectors in this thesis	187
5.7.4 Heterologous Protein Production and Purification.....	189
Reference	191

Chapter 1. Introduction and Background of the Projects

1.1 General overview of projects.

Fungi are excellent producers of various metabolites with a wide range of bioactivities. Some filamentous fungi like *Byssoschlamys fulva* and *Trichoderma viride* can produce alkyl citrates such as byssochlamic acid **1** and viridiofungin **2**. Fungi like *Beauveria bassiana* can produce tenellin **3**. Although these compounds appear to be structurally quite different from each other, they do have interesting similarities. For example, tenellin **3** and viridiofungin **2** both contain a tyrosine unit (blue in **Fig. 1.1.1**); byssochlamic acid **1** and viridiofungin **2** both contain a citrate-derived unit (red in **Fig. 1.1.1**); and tenellin **3** and byssochlamic acid both contain a polyketide or fatty acid moiety (black in **Fig. 1.1.1**)

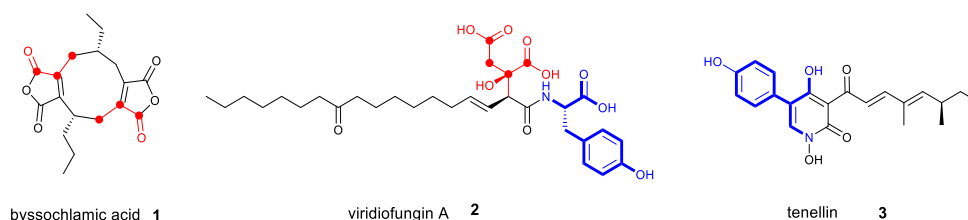


Fig. 1.1.1 Structures of byssochlamic acid **1**, viridiofungin A **2** and tenellin **3**. Citrate blocks highlighted in red, amino acid blocks highlighted in blue, alkyl chains in black.

Byssochlamic acid **1** is known to be derived from polyketide chains modified with oxaloacetyl moieties. Compound **1** features a 9-membered carbocyclic core, with two fused furan-2,5-dione/maleic anhydride moieties.^[1-3] The gene cluster of Byssochlamic acid **1** was first reported by the Cox group in 2016. Genes encoding a polyketide synthase (PKS), a hydrolase, a citrate synthase and a 2-methyl citrate dehydratase are involved in the core early steps in the biosynthetic pathway to form the maleic anhydride moieties.^[3] Ketosteroid isomerase (KI)-like enzymes and phosphatidylethanolamine binding protein (PEBP)-like enzymes appear to control the dimerization steps to form the central carbocyclic ring.

The byssochlamic acid **1** project is focused on exploring mechanisms and selectivity of the enzymes for maleic anhydride monomer and ring-forming enzymes. The project is designed to use heterologous expression in *E. coli* or yeast to obtain the enzymes for *in vitro* assay. The chemical approach involved application of analytical HPLC to analyse crude extracts from the reactions.

Viridiofungin A **2** is a secondary metabolite of *Trichoderma viride*. It is a member of the alkyl citrate family of metabolites and is an inhibitor of squalene synthase.^[4] Viridiofungins have a characteristic citric acid structural element that is alkylated at the 2-position. Viridiofungins were first found in 1993, and a lot of total synthesis work has been conducted by different groups in recent years.^[5, 6] However, the biosynthesis of viridiofungins is still unknown.

The viridiofungin **2** project is aimed to explore the biosynthetic pathway of viridiofungins. The project should start with genome sequencing of *Trichoderma viride* MF 5628. Then different bioinformatics analysis software will be used to find a possible viridiofungin biosynthetic gene cluster. Based on sequencing and bioinformatic analysis, the project is designed to use chemical (analytical) and biological (genetic engineering) methods to explore the gene cluster.

Tenellin **3** is a secondary metabolite of *B. bassiana* and its structure was elucidated by Wat and co-workers.^[7] It is a member of a large class of related compounds including other structurally complex 2-pyridones and acyltetramic acids isolated from fungi. The Cox group found the gene cluster of tenellin **3** in 2008.^[8] The gene cluster of tenellin **3** encodes two cytochrome P450 oxygenases (TenA and TenB), a *trans*-acting enoyl reductase (ER, TenC) and an iterative polyketide synthase nonribosomal peptide synthetase (PKS-NRPS, TenS).^[9] In 2011 using rational domain swaps between the polyketide synthases encoding the biosynthesis of the closely related compounds tenellin **3** and desmethylbassianin, the Cox group found that expression of the hybrid polyketide synthases in *Aspergillus oryzae* led to the production of reprogrammed compounds in which the changes to the methylation pattern and chain length could be mapped to the domain swaps.^[10] The reprogrammed compounds always related to the changes in keto reductase (KR) and C-methyltransferase (C-MeT) domain.

In the tenellin **3** project, efforts will be made to understand programming of fungal PKS by using much smaller sub-domain swaps within the KR domain, to try to find out the molecular basis of methylation and chain-length programming in fungal iterative highly reducing polyketide synthases (hr-PKS).

1.2 Natural products, secondary metabolites and their major classes.

Natural products can be defined as any chemical compounds derived from nature. However, the traditional definition of natural products of organic and medicinal chemistry communities are small organic molecules whose molecular weight is less than 1500 daltons which are generated from conditional metabolic pathways.^[11] Here the conditional metabolic pathway usually refers to secondary metabolite pathways of organisms such as bacteria, algae, fungi or plants.^[12]

Different from primary metabolites, secondary metabolites are not present in all organisms and are not essential for life. Secondary metabolites are usually thought to confer some form of advantage or protection to the producers. Natural products and secondary metabolites are a great source for the pharmaceutical industry to find the new molecules for medicine (50% of the drugs approved between 1981 and 2010 were from natural origin).^[13]

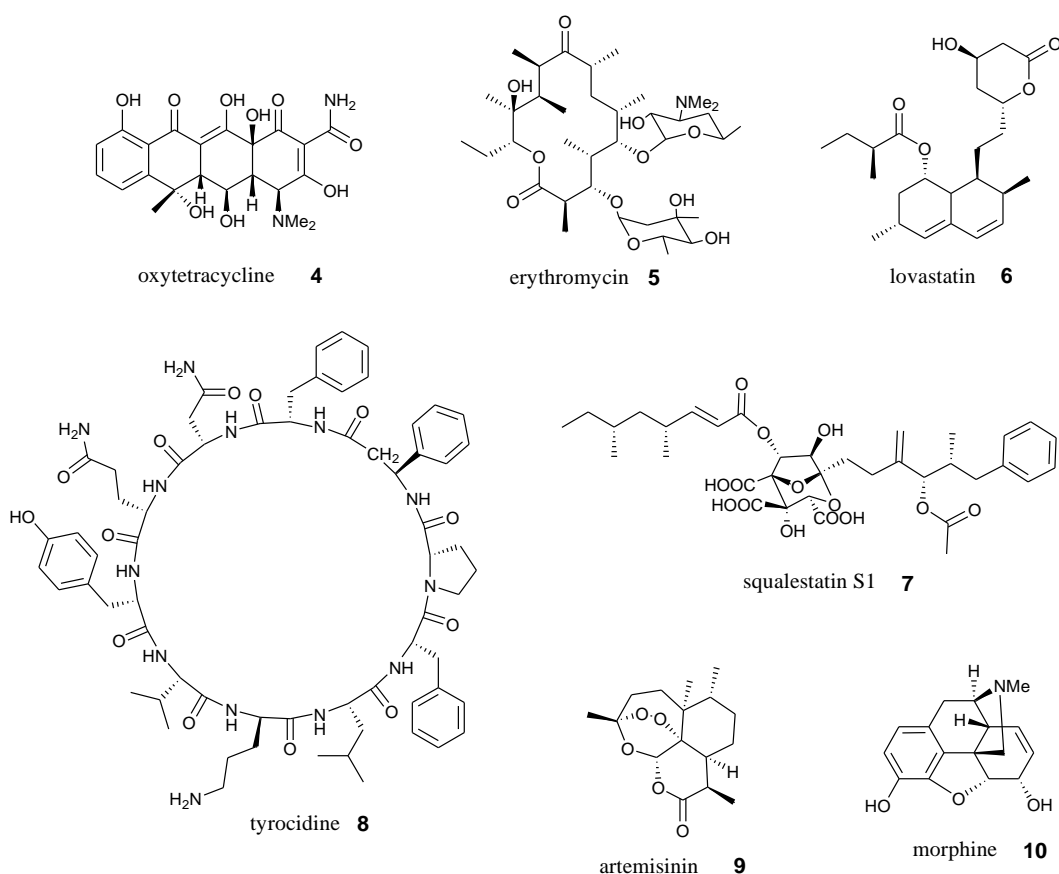


Fig. 1.2.1 Examples of different classes of natural products (polyketide, peptide, terpene and alkaloid).

Based on their biosynthetic origin, natural products are classified as polyketides (and some fatty acid derived compounds), peptides, terpenes and alkaloids. Some secondary metabolites are hybrids of more than one class. Most of these compounds' building blocks come directly from primary metabolism, and these natural products are derived from common precursors. For example: polyketides are derived from acetyl-CoA; peptides and alkaloids are derived from amino acids; and mevalonic acid is the precursor of terpene biosynthesis. Oxytetracycline **4**, erythromycin **5**, squalestatin S1 **7** and lovastatin **6** are polyketides; tyrocidine **8** is peptide; tenellin **3** is a mixed compound from a polyketide and peptide hybrid system; artemisinin **9** is a terpene; and morphine **10** is an alkaloid (**Fig. 1.2.1**).

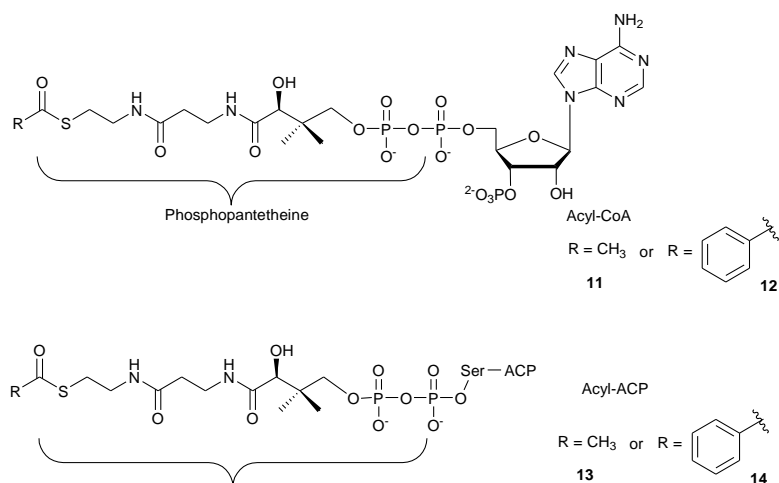
1.3 Polyketides and Fatty acids.

1.3.1 Basic enzymology of polyketide and fatty acid biosynthesis.

Polyketides are a very diverse group of natural products. They constitute one of the major classes of fungal metabolites. Polyketides are secondary metabolites which are built by polyketide synthase enzymes (PKS). Most fatty acids are primary metabolites, these fatty acids are synthesised by a complex biosynthetic machinery named fatty acid synthase (FAS). Polyketides and fatty acids share a common pattern of biosynthesis. Polyketides and fatty acids are generated from the same acyl and malonyl CoA building blocks by repeated condensation reactions.^[14] The general biosynthetic process of both polyketides and fatty acids is the same, but polyketide biosynthesis is more highly programmed.

1.3.1.1 Acyl transfer.

Both fatty acid and polyketide biosynthetic pathways start with the acyl transfer step. The acyl carrier protein (ACP) plays an important role in the acyl transfer steps. The *starter* and *extender* units, and the growing acyl chain are held by an ACP. The ACP is a small protein (*ca* 10 kDa) which has a phosphopantetheine (PP) prosthetic group attached to a highly conserved serine. The PP has a terminal thiol to which the acyl groups are attached as thioesters. ACP-bound acyl units can then be transported to all the functional domains of the PKS or FAS as required.^[15, 16]



Scheme 1.3.1.1 Structures of Acyl-CoA and Acyl-ACP.

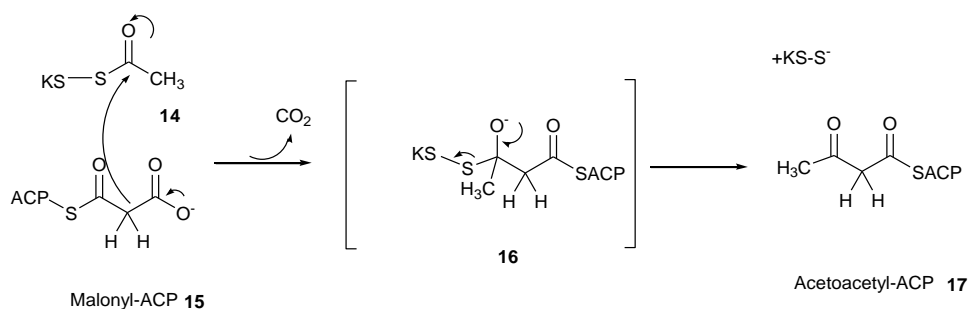
The starter group for FAS is the thiolester acetyl-CoA **11**. The starter group in PKS biosynthesis can be other acyl CoAs such as benzoyl-CoA **12**, for example as used by the hexaketide synthase of squalestatin **7**.^[4] During this reaction an acyl-transferase (AT) domain transfers the acyl group

form CoA **11** to ACP **13**. The acyl-CoA is connected as a thioester to the terminal thiol of a phosphopantetheine which is linked to ACP through the side-chain –OH of a serine residue in the protein (**Scheme 1.3.1.1**). The acyl group of **13** is then transferred to a cysteine residue in the β -ketosynthase (KS) domain giving an acyl-KS **14** which is used in the next condensation step.^[17]

Malonyl-CoA is the common extender unit for both FAS and PKS. Malonyl-CoA is synthesized from acetyl-CoA **11**. In this step, acetyl-CoA **11** is carboxylated by transferring CO₂ to it with the help of the coenzyme biotin to form malonyl-CoA. Malonyl-CoA is then transferred to ACP by the AT to give malonyl-ACP **15**.^[18]

1.3.1.2 Condensation step.

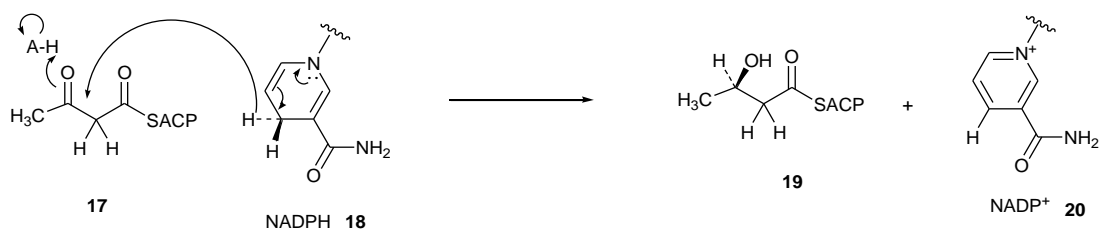
Malonyl-ACP **15** as extender, and acetyl-KS **14** as starter, then undergo a Claisen-like Condensation to extend the chain. The mechanism of the condensation involves decarboxylation of malonyl-ACP **15** to give an enolate ion, followed by immediate addition of the enolate ion to the carbonyl group of acetyl-KS, to form acetoacetyl-ACP **17**. In this step, acetyl-KS **14** is an electrophilic acceptor and malonyl-ACP **13** is a nucleophilic donor (**Scheme 1.3.1.2**).



Scheme 1.3.1.2 Claisen condensation step catalysed by the KS domain of PKS or FAS.

1.3.1.3 β -processing.

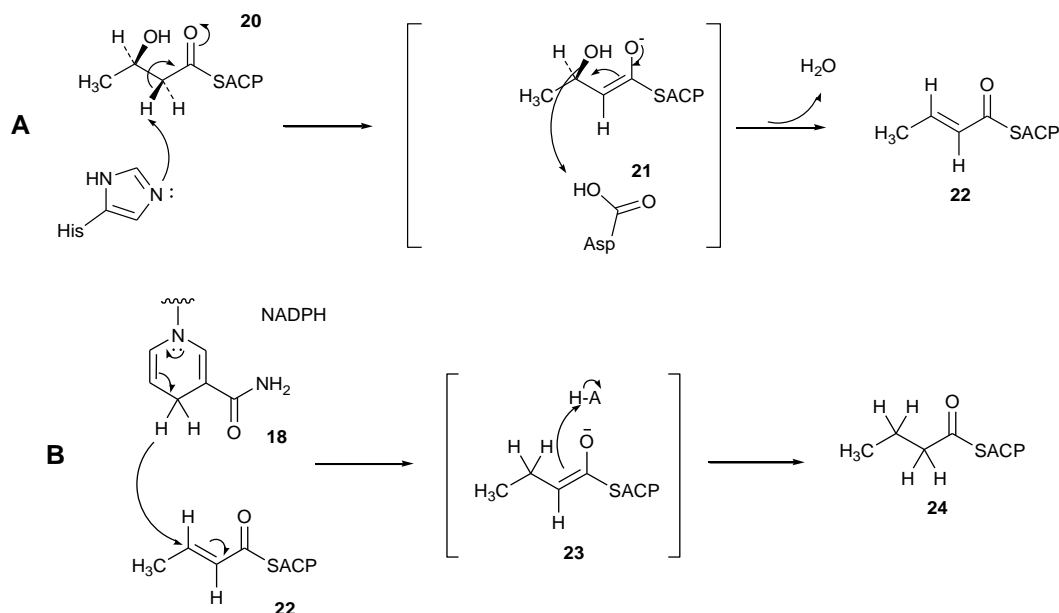
The KS thus creates a β -ketone. This ketone is then reduced by FAS. The first step is catalysed by the keto-reductase (KR) domain. The β -ketone carbonyl group is reduced to an alcohol forming β -hydroxybutyryl-ACP **19** (**Scheme 1.3.1.3**). The cofactor in this step is NADPH **18**.^[19]



Scheme 1.3.1.3 Ketone reduction step catalysed by KR domain of PKS or FAS.

The next steps of β -processing are dehydration and enoyl reduction which are catalysed by dehydratase (DH) and enoyl reductase (ER) domains respectively. During the dehydration step, the *2-pro-S* hydrogen of compound **20** is stereoselectively removed in an E1cB elimination mechanism to obtain olefin **22** (**Scheme 1.3.1.4 A**). The elimination is thought to proceed through an intermediate thiolester enolate **21**.^[20] Catalysis is thought to proceed by histidine abstraction of a proton from the α position, and aspartic acid protonation of the departing hydroxyl from the β position.^[21]

The enoyl reduction step occurs by conjugate addition of a hydride ion from NADPH **18** to compound **22**, the product **24** is then obtained (**Scheme 1.3.1.4 B**).^[17]



Scheme 1.3.1.4 Mechanisms of dehydration and reduction of PKS and FAS intermediates: **A**, DH domain; **B**, ER domain.

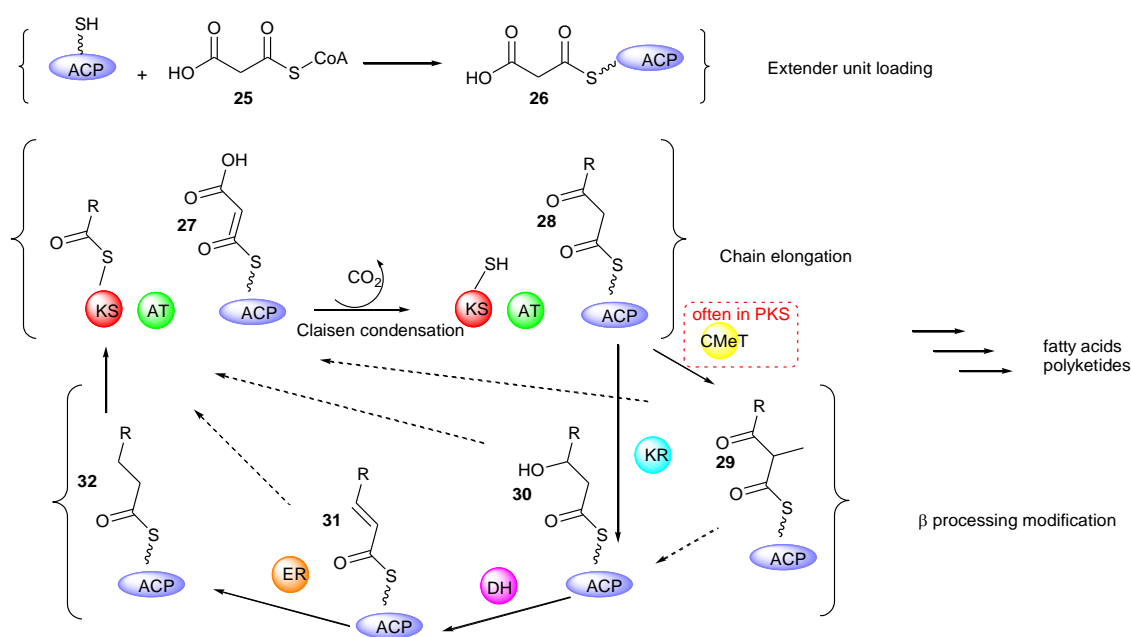
1.3.1.4 Release.

After further elongation rounds by C_2 units and β -processing, the complete fatty acid chain is released by a thioesterase (TE) domain. The product of fatty acid biosynthesis is a carboxylic acid. The fatty acid undergoes a hydrolytic release.

In the case of polyketide biosynthesis, the most well-studied chain release mechanism involves macrolactonization catalysed by a TE domain appended to the C terminus of the PKS (e.g. biosynthesis of erythromycin).^[22] Examples of TE domains that catalyse chain release *via* thioester hydrolysis or macrolactamization are also known.^[23, 24] In addition, several other types of catalytic domains have been reported to catalyse chain release. Examples include thioester reductase (TR) domains, which catalyse reductive release to form an aldehyde or alcohol.^[25, 26] Furthermore, α -oxoamine synthase (OAS) domains, release the polyketide chain *via* decarboxylative condensation with the α -carbon of an amino acid.^[27]

1.3.1.5 Programming of PKS

Polyketides and fatty acids bear strong resemblance in biosynthesis, which is divided in three phases: extender unit loading, chain elongation, and β -processing (Scheme 1.3.1.5). In fungal PKS, malonyl-CoA **25** is used as extender unit and loaded to ACP forming **26**. Chain elongation undergoes a claisen condensation from **27** to **28**. β -processing includes C-methylation **29**, keto-reduction **30**, dehydration **31** and enoyl reduction **32** (Scheme 1.3.1.5).



Scheme 1.3.1.5 General mechanisms involved in fatty acid and polyketide biosynthesis.

The key difference between FAS and PKS is that of programming. Full reduction of the β -carbon chain always takes place during fatty acid biosynthesis and without methylation. However, PKS is sophisticatedly programmed so that β -processing occur optionally give an alcohol **30**, an alkene **31** or an unmodified ketone group **28** on particular carbons. It can additionally control starter and extender unit selection.^[28] In addition, fungal PKS often have a functional C-methylation (C-MeT) domain which can methylate the α -carbon at the beginning of the β -processing cycle.

1.3.2 Classification of FAS and PKS.

FAS and PKS can be classified into different types based on the architecture of the proteins, whether they are iterative or non-iterative, and the degree of reduction at the β position.

1.3.2.1 Different types of FAS and PKS.

Based on the structure of the enzyme, FAS are divided into type I and II systems. Type I FAS systems are multi-domain enzymes which contain all domains in one or two multifunctional proteins. Type II FAS systems consist of non-covalently linked enzyme complexes that are characterised as a set of dissociated monofunctional proteins. Type I FAS is usually found in animals and fungi, and type II FAS is usually found in bacteria and plants.

According to the differences in architecture, PKS can be classified into three types: type I, type II and type III PKS. Type I PKS are multifunctional proteins containing all active domains for biosynthesis of polyketide. Type I PKS can be sub-classified into two types: Type I modular PKS and the type I iterative PKS. Type I modular PKS are usually found in bacteria, but the biosynthesis of fungal polyketides is usually driven by type I iterative PKS. In type I modular PKS, each single module of the PKS catalyses one cycle of polyketide chain elongation, such as the deoxyerythromycin B synthase (DEBS) during the biosynthesis of erythromycin A **5**. In contrast, type I iterative PKS, use a single set of catalytic domains in a single module which is highly programmed and works iteratively.^[29]

Type II PKS are non-covalent aggregates of monofunctional proteins, and they are involved in the synthesis of aromatic polyketides in bacteria. Type II PKS are not found in fungi. Type III PKS do not contain acyl carrier protein (ACP) domain, and are exemplified by systems such as chalcone synthase.^[28, 30] Type III PKS are homodimeric KS that synthesize smaller aromatic compounds in bacteria, fungi and plants, and which are very different from type I and type II PKS systems.^[31]

1.3.2.2 Type I iterative PKS and eukaryotic FAS.

Type I iterative PKS can be classified into three types thus: highly reducing PKS (hr-PKS); partial reducing PKS (pr-PKS); and non-reducing PKS (nr-PKS). hr-PKS are a large subgroup of type I iterative PKS which are involved in the biosynthesis of highly reduced compounds such as lovastatin **9** and squalestatin S1 **13**. The basic domains of hr-PKS include KS, AT and ACP. hr-PKS also include the β -processing domains like KR, DH, ER and C-MeT. The domain architecture and assembly line of the hr-PKS closely resembles that found in mammalian fatty acid synthases (mFAS). However, there are also many differences between mFAS and hr-PKS. For example, mFAS only contains a broken C-MeT domain (known as ψ C-MeT), and hr-PKS use the special set of β -processing domains in different extension cycle, which result in the different degree of reduction and methylation at individual α - and β - position in the final product. Additionally, hr-PKS usually have no thiolesterase (TE) domain at the C-terminus of the megasynthase.^[32]

In summary, the type I iterative PKS and most of eukaryotic FAS (like fungal FAS and mFAS) are multifunctional proteins which catalyse multiple rounds of precursor elongation by the same set of enzymatic domain in one multienzyme.^[33] Eukaryotic FAS are highly optimized molecular machines specifically for producing fatty acids. While multienzymes like type I iterative PKS might have the overlapping set of functional domains not belong to primary fatty acid metabolism. The typical catalytic domains of type I iterative PKS are closely related to mFAS which include KS, AT, DH, ER, KR, C-MeT and ACP domains.

1.3.3 Protein structure of FAS and PKS.

In recent years, several protein structures of fungal FAS and mFAS have been reported. The similarities and differences between FAS and PKS in protein structures are also discussed in various studies.

1.3.3.1 Fungal FAS.

Fungal FAS from *Thermomyces lanuginous* is a 2.6 MDa $\alpha_6\beta_6$ heterododecamer with a barrel shape enclosing two large chambers which contain three sets of active sites separated by a central wheel-like unit (**Fig. 1.3.3.1 A**). There are three distinct units in structure of the fungal FAS, the central wheel formed with 6 α units which form the equatorial disk of the particle, and two domes formed by 3 β units build up the reaction chambers (**Fig. 1.3.3.1 B**).^[34]

There are interactions between the subunits. The α chains in the wheel like central unit will contact with β units (**Fig. 1.3.3.1 C**). At the same time, the interaction between α - α subunits mainly involve homodimeric interfaces and interlocking modules arranged along the two-fold axis of symmetry. Two α helices constitute a thin spoke in the central of the wheel (**Fig. 1.3.3.1**

C A). One of the sequence insertions in the KS core domain form the second spoke (**Fig. 1.3.3.1 C C**). In the position of **Fig. 1.3.3.1 C B**, peripheral domain forms which exhibits flexibility with the body of the FAS. Another peripheral domain is formed by expansion segment helices at the outside of the KS. The β - β contacts in the dome like structure are related with the formation of the trimeric apical connection of the arches (**Fig. 1.3.3.1 D K**). The three different α - β connections are all contributed by expansion segments (**Fig. 1.3.3.1 D J, H, I**). In this research, ACP is attached in the reaction chamber together with the spatial distribution of active sites which indicate that iterative substrate shuttling is achieved by a restricted circular motion of the carrier domain in fungal FAS.^[35]

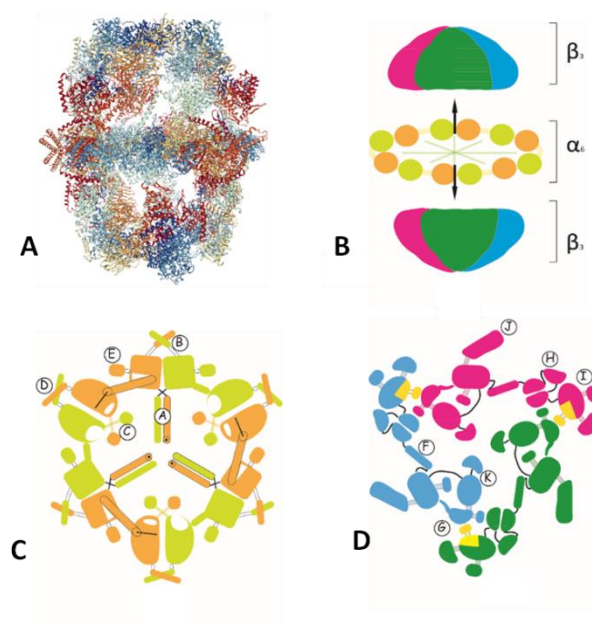


Fig. 1.3.3.1 Structure and diagrams of fungal FAS: **A** protein structure; **B** $\alpha_6\beta_6$ structure; **C** wheel like structure; **D** dome like structure

1.3.3.2 mFAS

Animal FAS (the crystal structures of natively purified mFAS from pigs) is a single protein which forms a 540 kDa X-shaped homodimer with two lateral reaction clefts (**Fig. 1.3.3.2 A**). mFAS can be segregated into two parts, the condensing part with KS and malonyl-CoA ACP transacylase AT domains and an upper part, and a modifying part, containing all of the domains for β -processing.^[36] The arrangement of domains in the 3D structure actually does not follow the linear order at the sequence level. mFAS is segregated into a lower condensing portion, containing the chain-building KS and the AT domains, and an upper portion including the β -processing domains. The KR domain divides into a functional (catalytic) part and a “structural” part known as ψ KR. There is also a non-functional C-MeT domains known as ψ C-MeT (**Fig. 1.3.3.2 B**). Dimerization of mFAS is mainly mediated through interactions of the lower KS domain part with ER domain

upper part. Compared with fungal FAS, mFAS does not have tightly embedded core enzymatic domains. The linkers between the upper and lower domains are only short stretches of extended polypeptide which is only wrapped around the surface of the catalytic domains. Only the KS-AT linker domain acts as a true adapter between two domains and adopts a folded core structure.^[37]

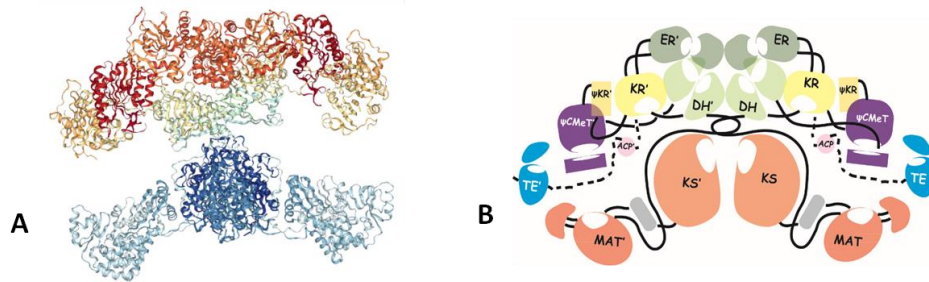


Fig. 1.3.3.2 Structure of mFAS: A, crystal structures; B, Schematic architectural drawing of the mFAS domain connectivity.

1.3.3.3 hr-PKS.

hr-PKS are closely related to mFAS with a sequence identity of approximately 20% and the same domain order overall.^[38] However, there is no structural information about fungal hr-PKS except the *trans* acting ER of LovC which is not structurally related to the ER domains of mFAS or hr-PKS.^[39] Due to the fact that mFAS and fungal hr-PKS show end-to-end sequence homology, it seems highly likely that hr-PKS are structurally similar to mFAS. They probably share the same overall dimeric structure in which the elongation domains (KS and AT) form a discrete structural unit known as the “elongation block,” and the modifying domains (C-MeT, KR, DH and ER) form a separate structural unit, known as the “modifying block”. The organization in the modifying region is likely to be particularly similar between hr-PKS and mFAS too. mFAS maintain degraded non-functional C-MeT remnants (Ψ CMeT). But hr-PKS usually contain a fully function C-MeT domain at the same position.

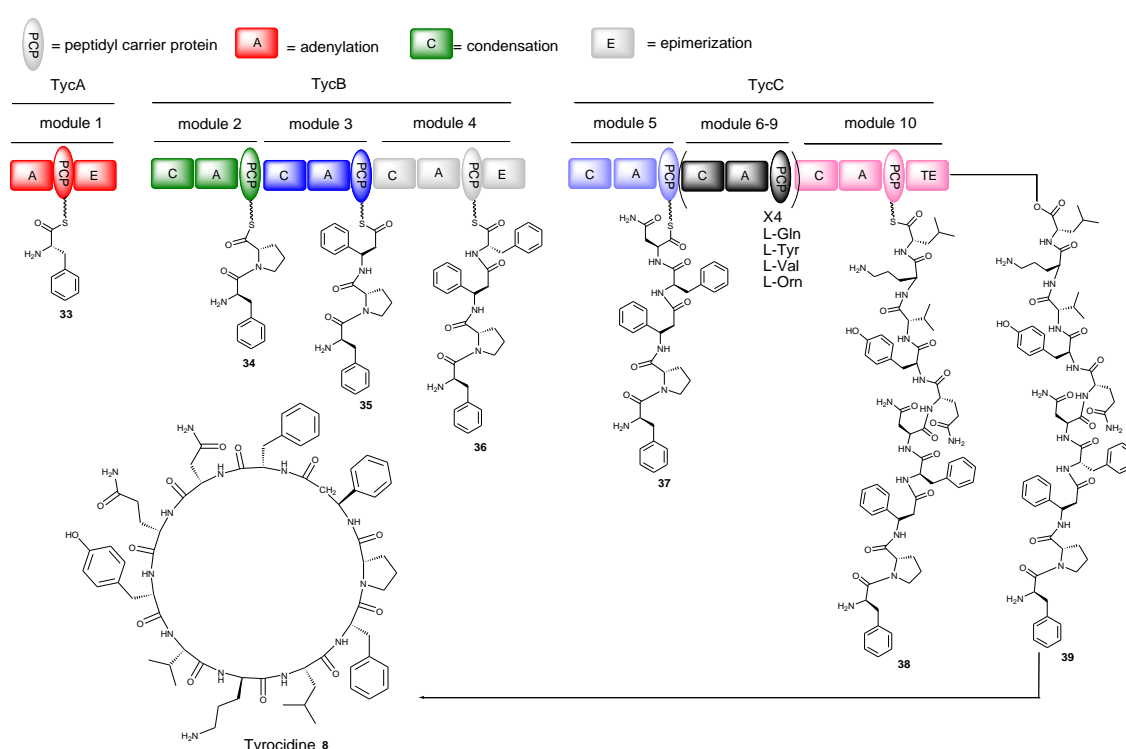
Differences among the mFAS and PKS modifying blocks are certain to arise as a result of the absence of either the ER or both the DH and ER domains in many PKS modules. The ACP domains, which are fused to the C terminus of the modifying block, also differ in PKS and mFAS systems.^[38, 40] In summary, the organization of hr-PKS modifying regions remains unclear, but current data suggest a considerable similarity to mFAS.

1.4 Nonribosomal peptide and PKS-NRPS compounds in fungi.

1.4.1 Nonribosomal peptide and NRPS.

Both prokaryotic and eukaryotic organisms can produce small peptides, many of which are nonribosomally constructed. In general, the peptides which belong to secondary metabolites are nonribosomal peptides, however, there are also exceptions, for example cyanobactins are small ribosomally derived cyclic-peptides from species of cyanobacteria.^[41, 42]

Biosynthesis of the nonribosomal peptide is independent of any mRNA or other RNA templates. Nonribosomal peptide synthetases (NRPS) are the key enzymes involved in the biosynthesis of nonribosomal peptides. NRPS assembly lines share a similar chemical logic with PKS assembly lines. There are initiation modules, elongation modules, and termination modules in NRPS assembly lines and in some cases additional tailoring domains. NRPS enzymes are modular and act in a stepwise assembly from amino acid monomers. NRPS enzymes can introduce functional group diversity in their products by utilizing all 20 naturally occurring amino acids as well as a number of unnatural amino and aryl acid substrates.



Scheme 1.4.1.1 Example of NRPS biosynthetic pathway.

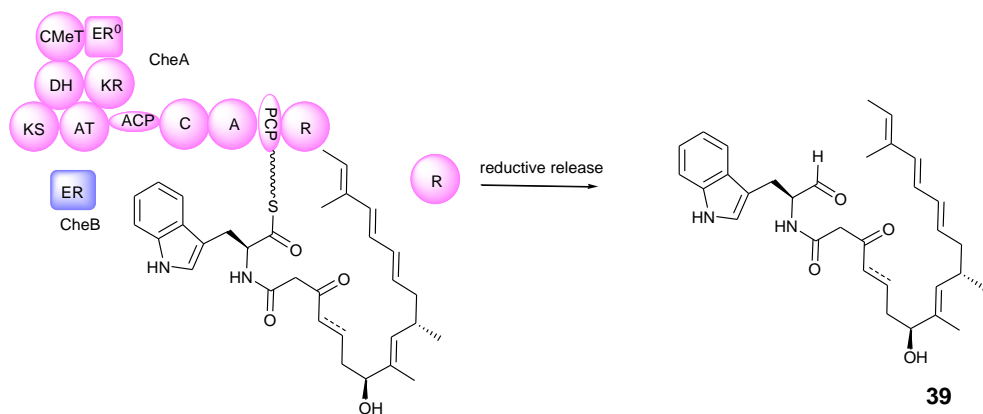
Four main domains are necessary for complete synthesis:^[43] the adenylation domain (A) selects and activates the monomer, transforming it into the adenylate form; the thiolation (T) or peptidyl carrier protein (PCP) domain binds the activated monomer to the synthetase; the condensation (C) domain catalyses the formation of peptide bonds; the thiolesterase domain, releases the peptide from the synthetase.

Taking the biosynthesis of Tyrocidine **8** as an example, there are 10 NRPS modules involved in the biosynthetic pathway. Module 1 is the initiation module which activates L-Phe, then a PCP domain binds the activated monomer to the synthetase to obtain **33**. The epimerization (E) domain of module 1 then convert L-Phe to D-Phe. The module 2 A domain activates the second amino acid, PCP domain binds the second monomer to the synthetase, then C domain catalyses the formation of peptide bonds to form **34**. Module 3 and 4 follow the next two cycles to obtain **35** and **36**. After 10 rounds of reaction, the compound **38** is formed. At the last step, the TE domain of module 10 releases the peptide chain using a macrocyclization mechanism and **8** forms. Tyrocidine undergoes a head to tail macrolactamization as the free amine of Phe₁ captured the carbonyl of Val₁₀ (**Scheme 1.4.1.1**).^[42]

1.4.2 PKS-NRPS compounds in fungi.

Some natural products are also synthesized by a hybrid PKS-NRPS systems. In bacteria, there are many known modular systems which place the PKS module and NRPS module together in the assembly lines.^[43] However, in fungi, PKS and NRPS usually combine as a single module of an iterative PKS followed by a single NRPS module. The architecture of the PKS part in the iterative hybrid PKS-NRPSs closely resembles that found in mFAS.^[42, 44] The PKS module in iterative PKS-NRPS has a general domain organisation of KS, AT, DH, C-MeT, ψ KR, ER, KR and ACP. Normally in PKS-NRPS the ER is broken (ER⁰) and replaced by a *trans*-acting ER typified by LovC and TenC.^[9, 39] The NRPS modules of the iterative hybrid PKS-NRPS contain C (condensation), A (adenylation) and T (thiolation) domains. The C-terminal domain normally catalyse either reductive release (R) or Dieckmann cyclisation (DKC).

The intermediate compound **39** of cytochalasin is synthesised by a PKS-NRPS (CheA) with a *trans*-ER (CheB). CheA contains an hr-PKS and a single module of NRPS. The hr-PKS produces a C₁₈ highly reduced polyketide after 8 rounds of elongation. The polyketide firstly loads on the ACP domain of PKS. The A domain of NRPS activates the Trp, and the active Trp bind to PCP domain. The C domain of NRPS catalyses the condensation of amino acid and polyketide. Finally the reduction (R) domain catalyzes a reductive release (**Scheme 1.4.2.1**).^[45]



Scheme 1.4.2.1 Example of PKS-NRPS.

Here, the natural products from fungi such as tenellin **3** and its related compounds bassianin **40** and militarinone C **42**, cytochalasins **43**, equisetin **41** and pseurotin A **44** are examples to explain how the PKS-NRPS work (amino acid part in red Fig. 1.4.2.1).

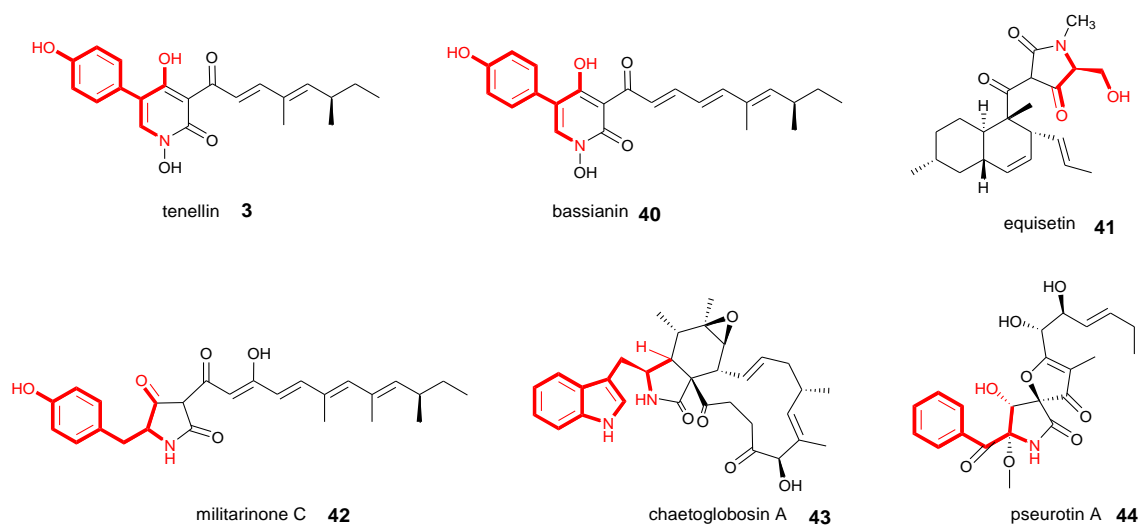
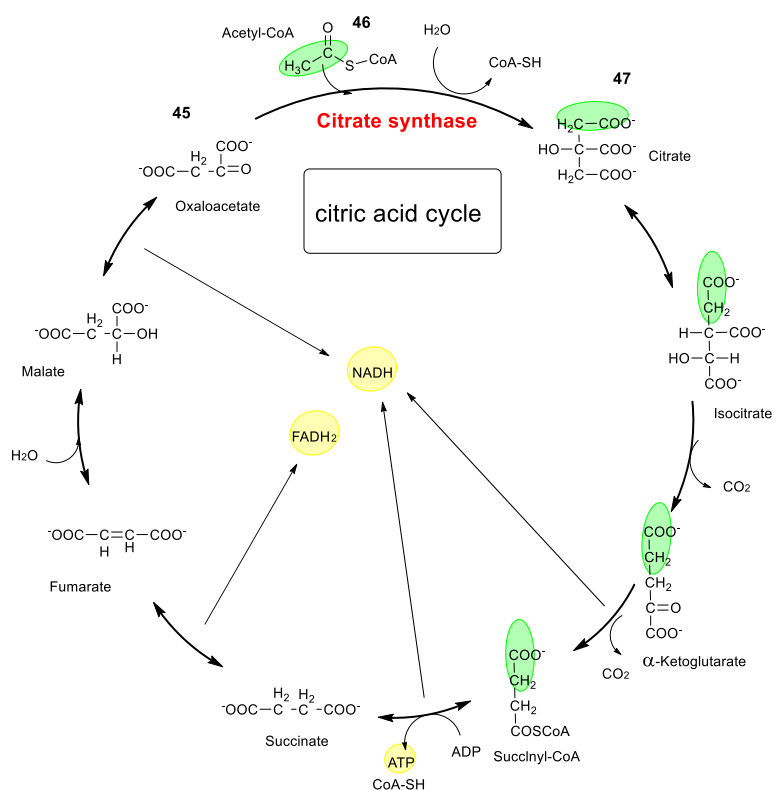


Fig. 1.4.2.1 Examples of hybrid PKS-NRPS compounds, the NRPS part is highlighted in red.

1.5 The alkyl citrates – an overview.

The generic term ‘alkyl citrate’ is proposed to encompass biosynthetically-related compounds where a citrate synthase (CS) like enzyme is involved in the biosynthetic pathway. These CS enzymes can combine oxaloacetic acid **45** and polyketides (or fatty acids) to form an alkyl citrate structure.



Scheme 1.5.1 Citric acid cycle.

Citrate synthase is a key enzyme in the primary metabolism citric acid cycle catalysing an irreversible step during primary metabolism. It catalyses the condensation of oxaloacetic acid **45** and acetyl-CoA **46** forming a citrate **47** (Scheme 1.5.1).^[46, 47] Natural product pathway such as those involved in alkyl citrate and maleidride biosynthesis also contain CS-like enzymes. These CS-like enzymes from secondary metabolism have high similarities with the CS from primary metabolism.^[1-3]

1.5.1 Maleidrides.

Maleidrides are a large group of alkyl citrates which contain a maleic anhydride in the structure. The collective name “nonadrides” was introduced by Barton and Sutherland for fungal metabolites which containing the core structure of at least one maleic anhydride unit fused on a nine-membered carbocyclic ring.^[48]

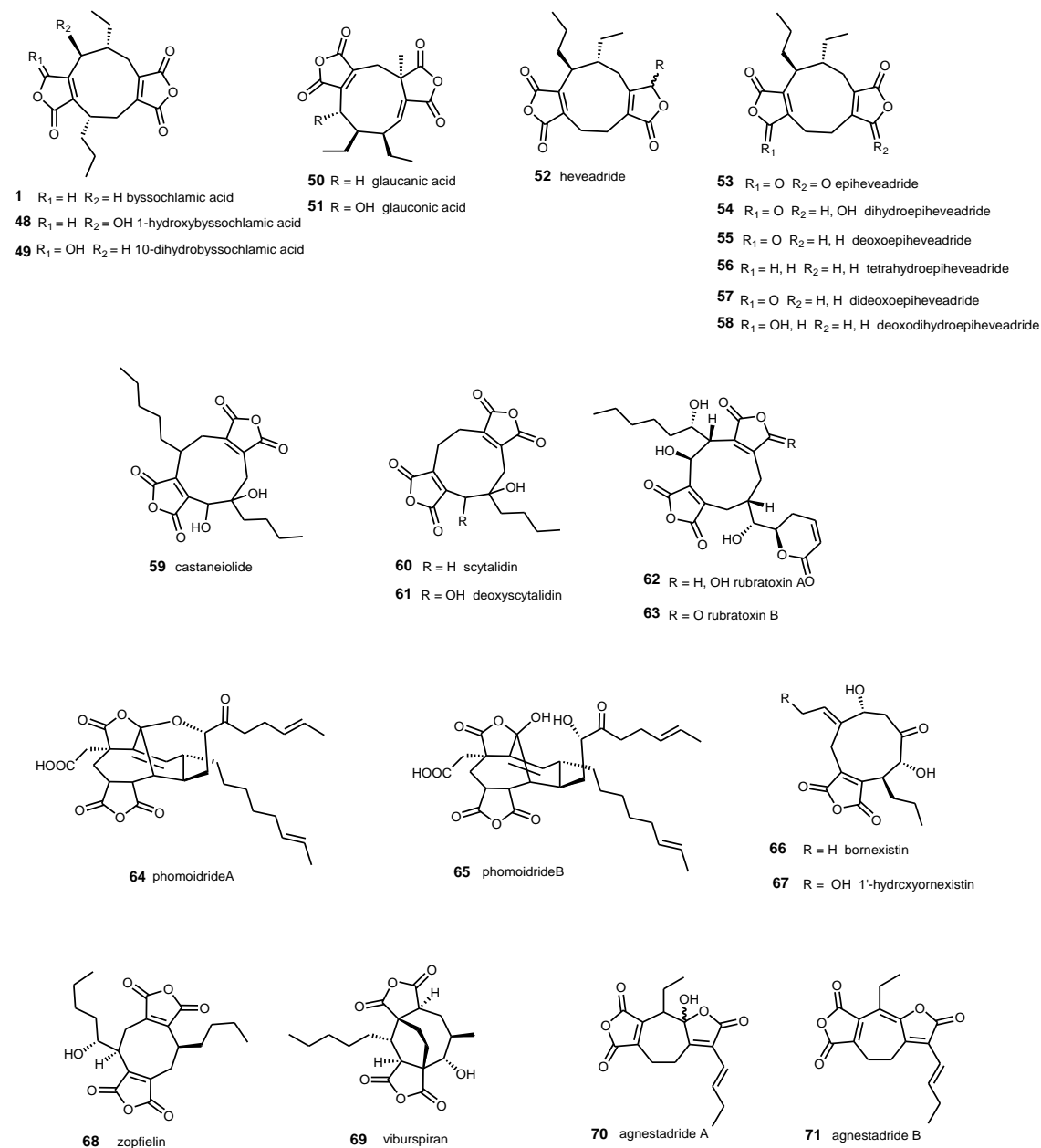


Fig. 1.5.1.1 Maleic anhydride units in structures of dimerized carbocyclic rings.

Glaucanic **50** and gluconic **51** acids were the first two compounds isolated from the fungus *Penicillium glaucum* by Wijkamn in 1931.^[49] Byssochlamic acids (**1**, **48** and **49**) were found by Raistrick and Smith in the fungus *Byssochlamys fulva* in the 1930. Its structure was elucidated in the 1960s,^[50] and the structures of its isomers glaucanic acid **51** and heveadride **52** were also elucidated at that time. Later, other nonadrides related to heveadride were also identified, for example, epiheveadride **53**, dihydroepiheveadride **54**, deoxoepiheveadride **55**, tetrahydroepiheveadride **56**, dideoxoepiheveadride **57** and deoxodihydroepiheveadride **58**.^[51, 52] From *Macrophoma castaneicola* castaneiolide **59** was isolated, and from *Scytalidium* sp. scytalidin **60** and its dehydroxyanalogue **61** also identified.^[53-55] In 1999, Baldwin and coworkers isolated phomoidrides A **64** and B **65** at the Pfizer Central Research Laboratories from an unidentified fungus. These two compounds are inhibitors of squalene synthase and the Ras-farnesyl transferase.^[56, 57] Rubratoxins ^[58, 59] A **62** and B **63** are produced by *P. rubrum*. Cornexistin **66** and 1'-hydroxycornexistin **67** produced by *Paecilomyces variotii*, are potent wide-spectrum herbicides against weeds but with low activity against maize.^[60] In addition to 9 membered carbocyclic rings fused to one or two maleic anhydrides, there are also 8-membered rings (octadrides) like zopfiellin **68**^[61] and viburspiran **69**.^[62, 63] In 2015, Cox and coworkers found two novel 7-membered heptadrides, agnestadride A **70** and B **71** from *Byssochlamys fulva*.^[2] In summary, the compounds from **48** to **71** use two anhydrides to form 7, 8 and 9- membered carbocyclic rings.

In nature, maleic anhydride units are not only seen from the dimerized carbocyclic rings. Some compounds are linear compounds with a few maleic anhydride units in the structures like cordyanhydrides A **72**, B **73** and C **74**.^[64] Furthermore, maleic anhydride units sometimes combine with other natural product structures, for example Lienhwalide A **75** and B **76** are the compounds which combine one maleic anhydride unit with a tropolone structure. Two sporothriolide-related compounds (**77**, **78**, **79**) were obtained from an extract of the fungus *Hypoxylon monticulosum* CLL- 205, isolated from a *Sphaerocladina* sponge collected from the Tahiti coast in 2017. These sporothriolide **80** related compounds also contain maleic anhydride units.^[65, 66]

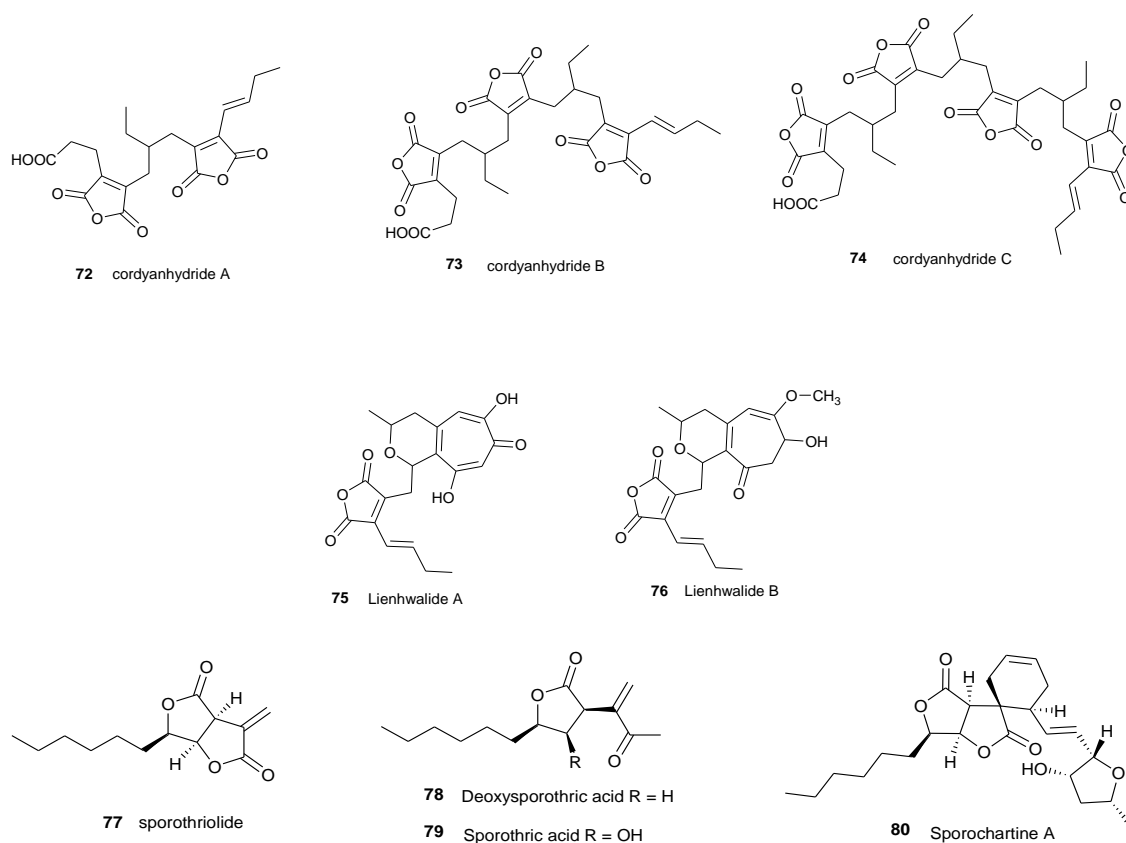


Fig. 1.5.1.2 Maleic anhydride units in different non-cycling structures. Stereochemistry of **72-76** is not known.

1.5.2 Squalestatin.

Squalestatin is an alkyl citrate compound which contains a citrate unit in its structure. Squalestatin S1 **7** first discovered in 1992, is a heavily oxidized fungal polyketide that offers potent cholesterol lowering activity.^[67] Many different squalestatins like zaragozic acid C **81** and zaragozic acid B **82** have been isolated from more than ten taxa of filamentous fungi, mostly varying in the attached 1-alkyl and 6-*O*-acyl side chains (**Fig. 1.5.2.1**).^[68] Squalestatins share a 2,8-dioxobicyclic octane-3,4,5- tricarboxylic acid core. The biosynthesis involves the production of two polyketides: a hexaketide initiated by benzoate; and a tetraketide produced by a hr-PKS. The other carbon atoms are derived from a citric acid cycle intermediate.^[69]

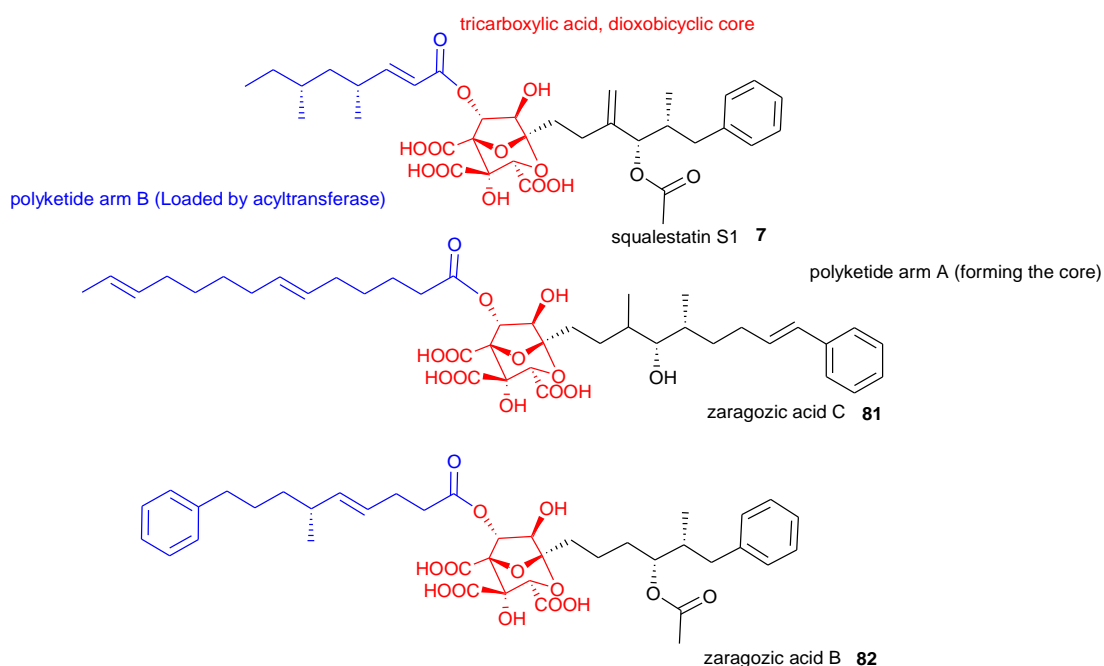


Fig. 1.5.2.1 Structures of squalestatin 2 and zaragozic acid.

1.5.3 Viridifungin

The viridifungins also belong to the family of alkyl citrates. The structural element of citric acid is present in viridifungins A-C (**2**, **84**, **85**), A1-4 (**86**, **87**, **88** and **89**), B2 **90** and Z2 **91**, obtained from the solid fermentation of *T.viride* (Fig. 1.5.3.1).^[6, 70] Viridifungins have different numbers of carbon atoms on the lipophilic tail attached at the 2-position of citric acid. At the same time, viridifungins are distinguished by the presence of an aromatic amino acid moiety in their structures.^[71]

The viridifungins have potent, broad spectrum antifungal activity and were found to inhibit the yeast and rat squalene synthases at the micromolar level. Viridifungins are also nanomolar inhibitors of the first enzyme in the sphingolipid biosynthetic pathway, serine palmitoyltransferase, and that inhibition of sphingolipid synthesis accounts for their antifungal activity.^[4, 72]

The previous research on viridifungins has been focused on natural product isolation and structure elucidation, biological activities and total synthesis. However, the biosynthetic pathway of viridifungins is still unknown.

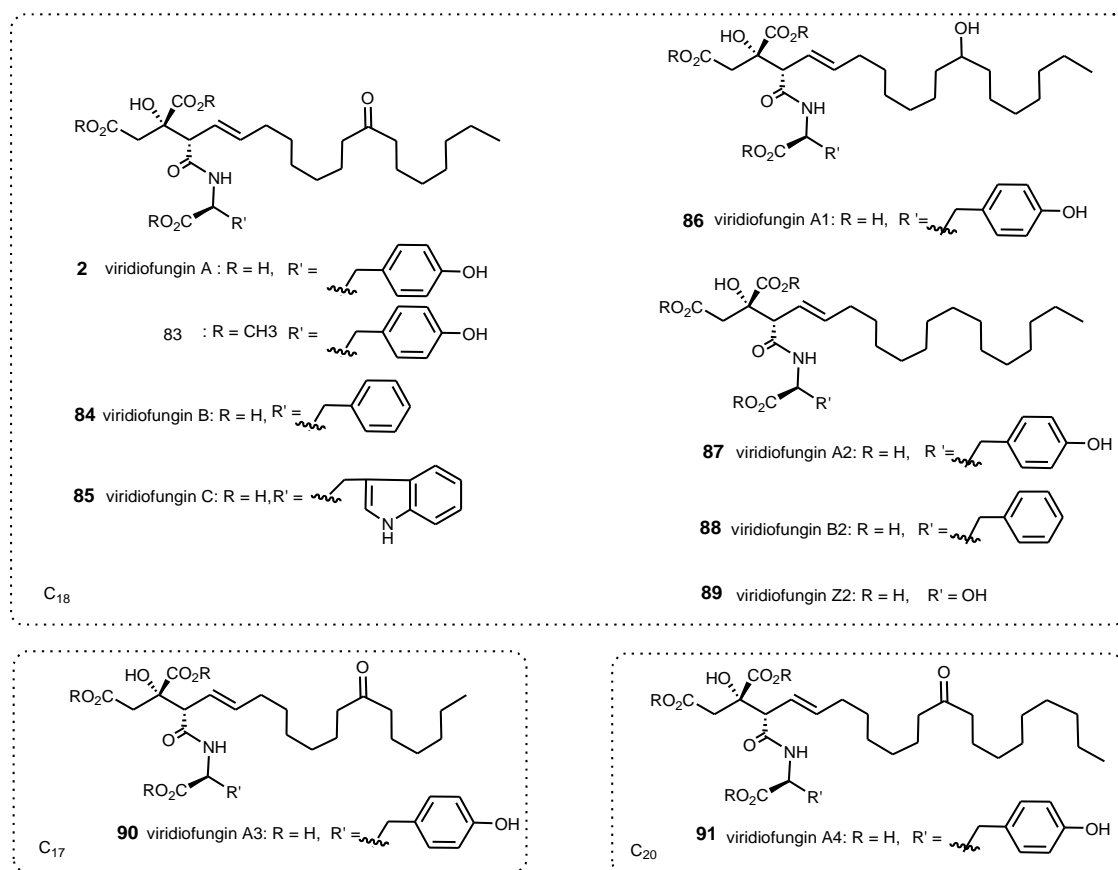


Fig. 1.5.3.1 Structures of viridiofungins.

1.6 Biosynthesis of Maleidrides

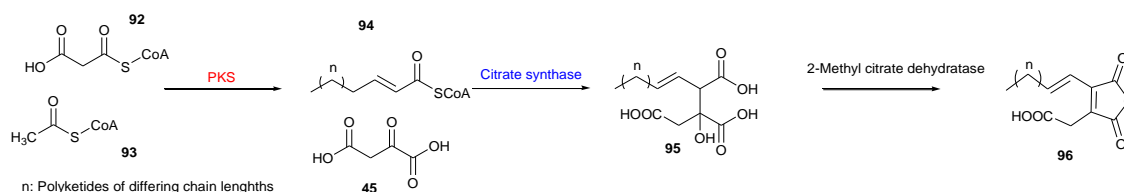
The biosynthesis of maleidrides is known to proceed in three basic steps, namely: early steps for maleic anhydride monomers; middle step for dimerization and the late tailoring steps. In 1965, Barton and Sutherland described that nonadrides such as byssochlamic acid **1** and glaucanic acid **50** are constructed from two identical C₉ units.^[73] The C₉ units are known to be maleic anhydride monomers. Barton *et al* also proposed that the combination of the two C₉ units may involve an anionic-type coupling mechanism which would furnish glaucanic acid **50** with the two ethylenic linkages in the correct positions.^[74] However, compounds like phomoidrides **64**, **65**, rubratoxins A and B (**62** and **63**) and castaneiolide **59** are not dimerised by C₉ units, but these compounds are also proposed to use similar dimerization mechanism as glaucanic acid **50** with different chain length of maleic anhydride monomers.

The earliest evidence for the biosynthesis of the monomers and the dimerization steps mostly come from feeding experiment and chemical synthesis of the putative substrates.

With the development of genome sequencing technology, gene clusters of byssochlamic acid **1**, cornexstin **66**, rubratoxins A and B (**62** and **63**) and phomoidrides **64** and **65** have been found, and much more information about the biosynthesis of maleic anhydride monomers (section **1.6.1**) and dimerization (section **1.6.2**) have been obtained from gene knock-out experiments, homologous expression experiments, and *in vitro* assays.

1.6.1 Early steps for maleic anhydride monomers.

Before the availability of genome information, labelling experiments to introduce small labelled molecules into the pool of primary metabolites in producing strains was a good way to explore the biosynthesis. Barton *et al* used [1-¹⁴C] and [2-¹⁴C]-acetate in feeding experiments in *P. purpurogenum*, which resulted in the production of radiolabelled glauconic acid **50**. Furthermore, the result indicated that C₉ maleic anhydride was a hybrid unit containing two chains of C₃ and C₆ respectively. The C₆ unit was confirmed to come from a polyketide or fatty acid. The C₃ unit is derived from oxaloacetate.^[73, 75, 76] Holker carried out a feeding experiment by adding [2,3-¹³C₂]-succinic acid to *P. purpurogenum*. The results confirmed that the succinic acid was incorporated into the C₃ unit.^[77] Another feeding experiment carried out by Sulikowski *et al.* on phomoidride B **64**, also showed that the longer chain of the monomer unit was built by a polyketide/fatty acid synthase.^[78]



Scheme 1.6.1.1 Biosynthetic pathway of maleic anhydride monomers.

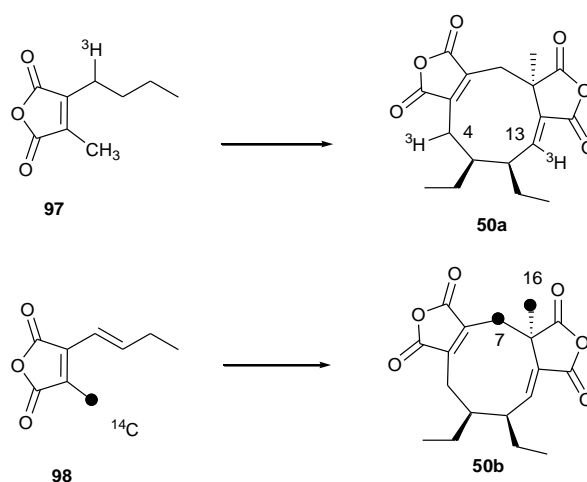
Although feeding experiments gave hypotheses that the monomers are formed from the combination of polyketide/ fatty acid and oxaloacetate, the biosynthetic details were still unclear. In 2015 the Oikawa group reported a possible gene cluster for phomoidride B **64** from the unidentified fungus ATCC 74256. The BGC encodes one highly reducing polyketide synthase (*phiA*), a citrate synthase (*phiJ*) and a 2-methylcitrate dehydratase like enzyme (*phiI*). These three genes were transformed into *A. oryzae*. The transformant *AO-phiAIJ* produced a new metabolite which have an identical UV absorption (λ_{max} 312 nm) for maleic anhydride conjugated with an olefin.^[61]

The Oikawa group also expressed *phiJ* and *phiI* in *E. coli* and carried out an *in vitro* tests using polyketide CoA and oxaloacetate as substrates. They added PhiJ and PhiI at the same

time, which resulted in a maleic anhydride product. The *in vivo* and *in vitro* results from the phomoidride B **64** gene cluster show that the PKS synthesizes a polyketide **94**, the citrate synthase uses oxaloacetate **45** and polyketide as substrates to synthesise the citrates **95** and the dehydratase forms the alkene citrate **96** (Scheme 1.6.1.1).^[61]

1.6.2 Dimerization step of maleidrides.

The study of dimerization steps during nonadrides biosynthesis started with feeding experiments. Sutherland *et al* carried out a feeding experiment on *P. purpurogenum* to explore glaucanic acid **50** dimerization. They directly used labelled maleic anhydrides [1-³H]-2-n-butyl-3-methylmaleic anhydride **97** and [7-¹⁴C]-2,3-n-butenyl-3-methylmaleic anhydride **98** to feed the fungus, and **50a** and **50b** were formed (Scheme 1.6.2.1).^[73] The results demonstrated that the incorporation happened at the positions that were expected for the product of head-to-head dimerization of the monomer units.

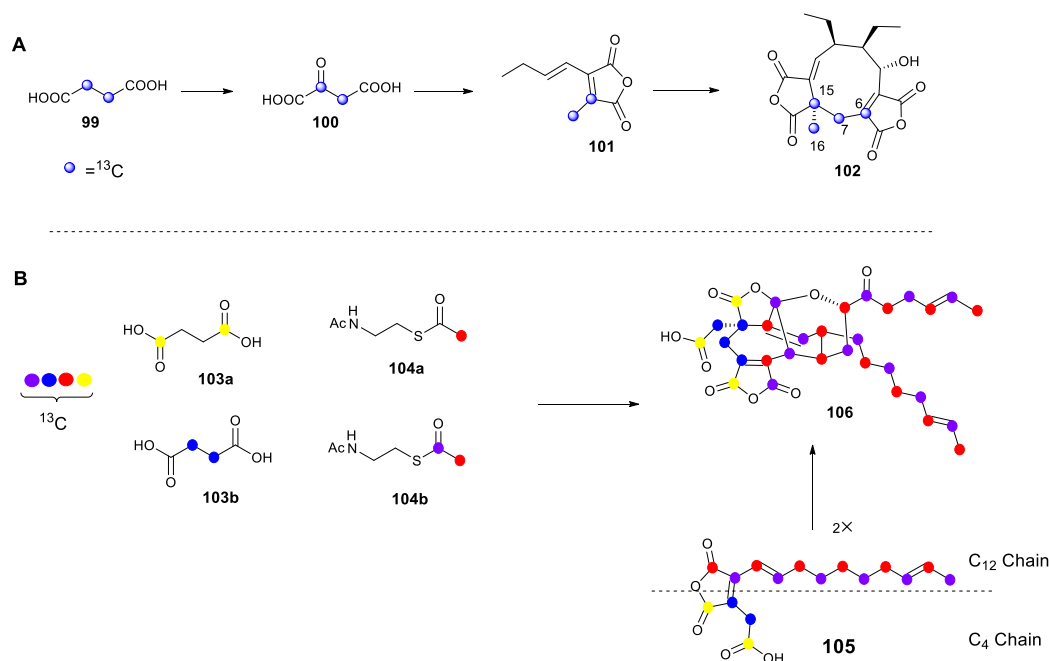


Scheme 1.6.2.1 Labelled maleic anhydride feeding experiment on.

Holker also carried out a labelling experiment using [2,3-¹³C₂]-succinate **99** in *P. purpurogenum*. The ¹³C NMR spectrum of glaucanic acid **102** showed the presence of two pairs of split and enhanced signals corresponding to ¹³C-15, ¹³C-16 (*J*= 33 Hz) and ¹³C-6, ¹³C-7 (*J*= 48 Hz). This result suggested that the intact succinic acid **99** or oxaloacetate **100** was incorporated into the C₃ chain of the nonadride precursor (Scheme 1.6.2.1 A). Furthermore, Sulikowski and co-workers also carried out a labelling experiment on phomoidride B **64** in fungus ATCC 74256 with [2,3-¹³C₂]-succinate **103a** and [1,4-¹³C₂]-succinate **103b**. The ¹³C NMR showed the enhanced intensities of C-12, 13, 14 and 18 resonances, which were the positions expected to be derived from **103a**. The enhanced intensities of C-27, 29 and 30 positions is from **103b** (Scheme 1.6.2.2).

I Dimerization step of meleidrides

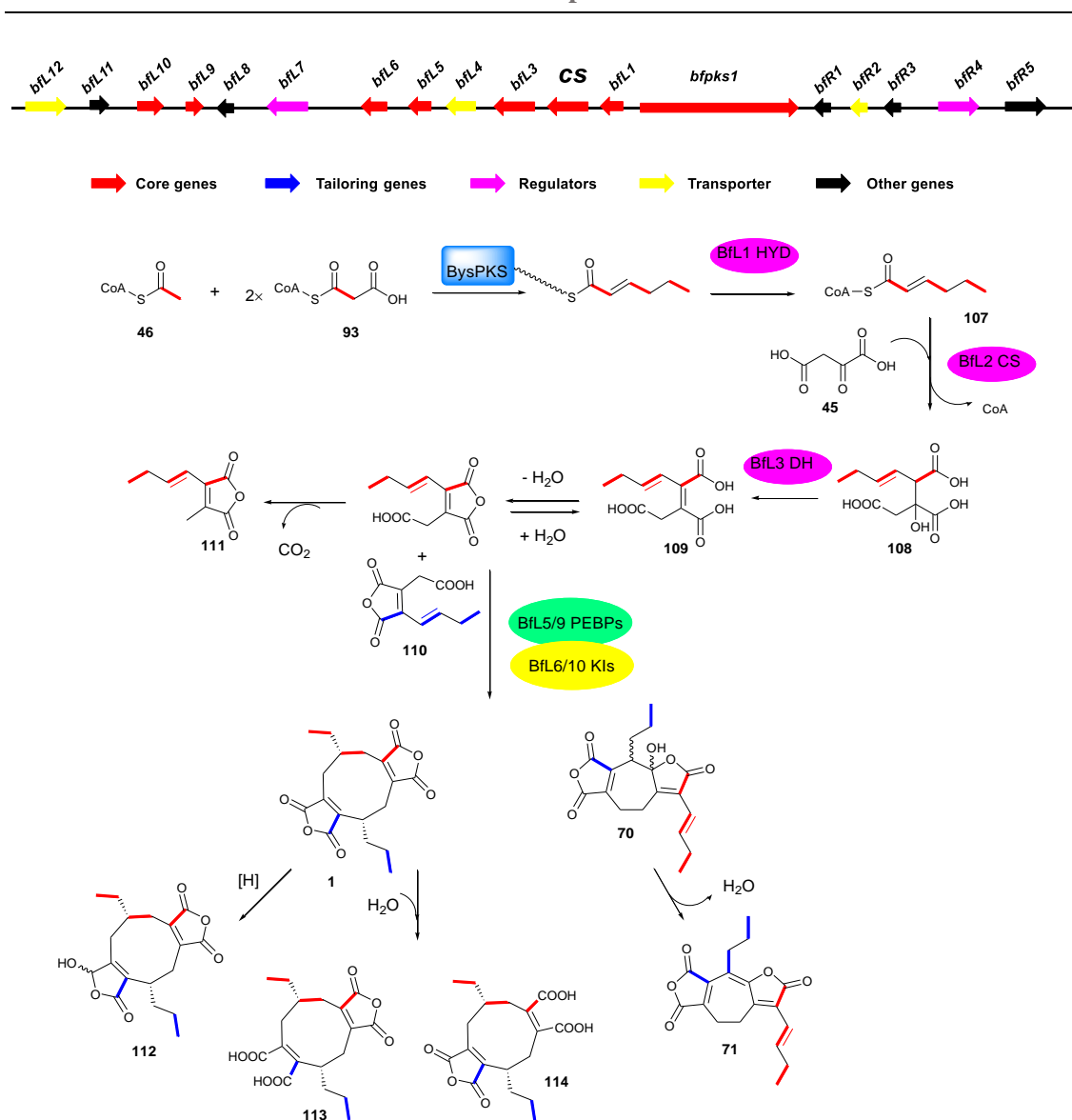
In addition, results from feeding with singly [2-¹³C]-**104a** or doubly labelled [1,2-¹³C]-**104b** (as SNAC thiolesters) were in full accordance with the longer C₁₂-chain of the monomer **105**.^[77]



Scheme 1.6.2.2 Labelling studies of meleidrides: **A**, gluconic acid **50**; **B** phomoidride **B 64**.

The feeding experiments give some evidence for the dimerization process, but not enough to explain the biosynthesis and enzymology of the dimerization step. In 2016 the Cox group expressed the *B. fulva* byssochalmic acid genes in *A. oryzae*. *Byssochlamys fulva* also produces a novel 7-membered heptadride, agnestadride **70**. By comparing different nonadride BGC from different organisms, they found that the BGC always includes genes encoding putative ketosteroid isomerase (KI) like enzymes and putative phosphatidylethanol-amine-binding proteins (PEBP). In the case of byssochalmic acid **1** there are two sets of KI and PEBP genes. When core early genes were expressed, only compounds **110** and **111** were formed. But when they co-expressed core genes with the two KI-like enzymes in *A. oryzae*, byssochalmic acid **1** and agnestadride **70** were produced. In addition, when both KI like genes and both PEBP like enzymes were expressed, the production increased more than 10 times. The results indicated that one set of KI and PEBP is not enough for ring formation in this case.^[3]

I Dimerization step of meleidrides

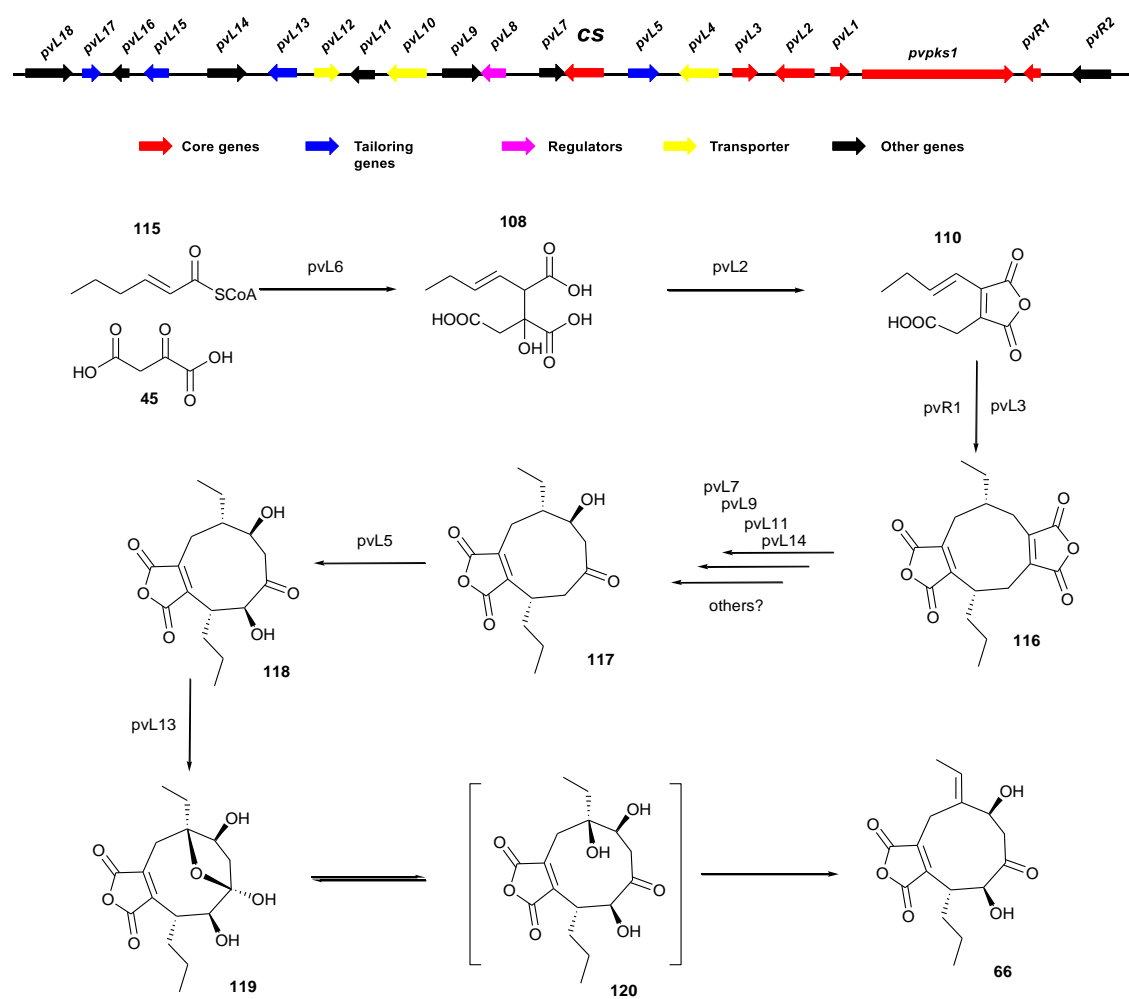


Scheme 1.6.2.3 Gene cluster and proposed biosynthetic pathway of byssochlamic acid **1**.

According to the *in vivo* studies, including knockout experiments in the wild type strain and heterologous expression in *A. oryzae*, it was shown that the biosynthesis of byssochlamic acid **1** and agnestadride A **70** is based on the dimerization of maleic anhydride monomers. The precursors for cyclization were formed by an iterative highly reducing fungal polyketide synthase followed with a hydrolase (Bfl1), a citrate synthase (Bfl2) **108**, and a 2-methyl citrate dehydratase (Bfl3) **109**. In addition, the enigmatic ring formation was catalyzed by two KI (Bfl6 / 10) and assisted by two PEBP (Bfl5 / 9) enzymes. *A. oryzae* transformants did not produce 10-dihydrobyssochlamic acid **112**, or agnestadride B **71**, while wild type *B. fulva* did, so it was likely that some unspecified reductase and dehydratase genes from outside of the currently defined cluster were involved (**Scheme 1.6.2.3**). The *A. oryzae* transformants can also produce **113** and **114** which is thought to from the ring-opening of one maleic anhydride moiety.^[3]

1.6.3 Tailoring steps of maleidrides.

For the simple maleidrides like byssochlamic acid **1**, agnestadride A **70** and glaucanic acid **50**, the dimerization step is the last step to form the final structure. However, there are also more complicated maleidrides whose biosynthetic pathways contain more tailoring steps, such as cornexstin **66** and rubratoxins **62** and **63**.

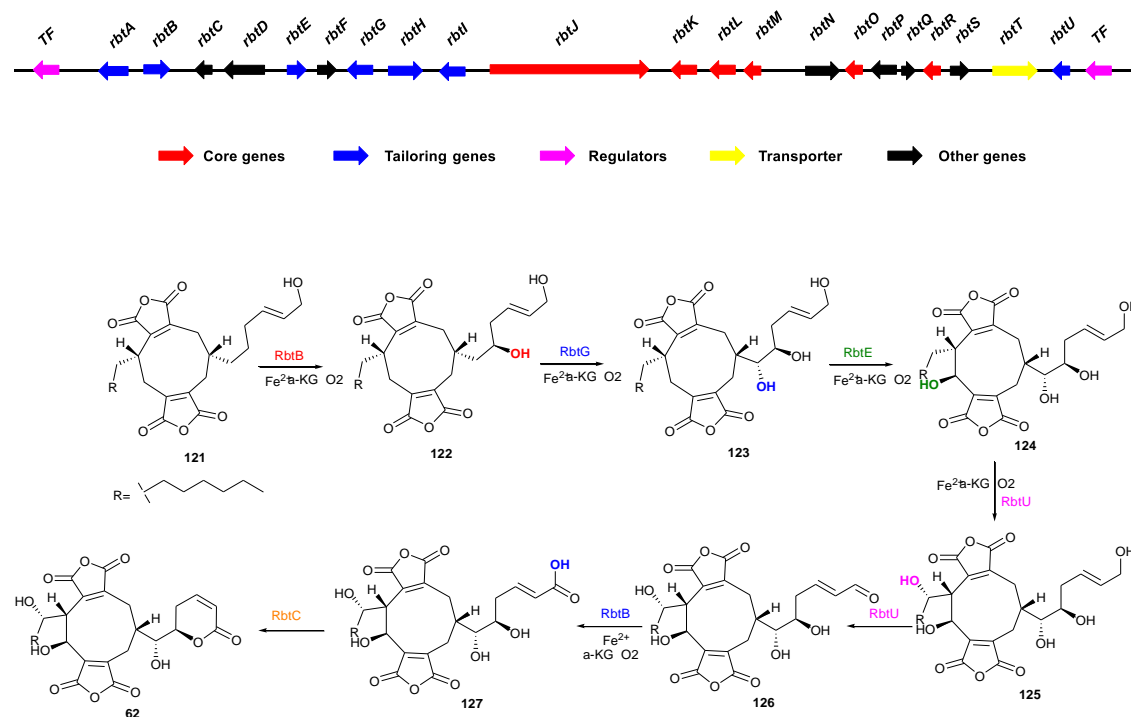


Scheme 1.6.3.1 Gene cluster and proposed biosynthesis of cornexstin **66**.

The biosynthesis of the herbicide cornexistin **66** in the fungus *Paecilomyces variotii* was investigated by full sequencing of its genome in 2017 from the Cox group.^[1, 79] Cornexistin **66** is different from the other nonadrides, because of the removal of one maleic anhydride moiety. But the biosynthesis of the anhydride monomer **110** and the dimerization to form cornexistin **66** should be highly similar to that of byssochlamic acid **1**. By sequencing and bioinformatics analysis, a gene cluster containing core genes *pvL11* (PKS), *pvL17* (citrate synthase), *pvL13* (2-methylcitrate dehydratase), *pvL14* (KI), and *pvR1* (PEBP) was found. At the same time, there are also a lot of tailoring genes that might be involved in the pathway. According to the results of knock-out

experiments, the Cox group found the final steps of the pathway involve decoration of the carbon skeleton of cornexistin by oxygenases encoded by *pvL13* and *pvL5*. These steps involve selective hydroxylation at the 6- position of **117** to **118** by a non-heme iron dependent monooxygenase. Dehydrogenation of the 2-2' position of **118** to obtain **119** then occurs (Scheme 1.6.3.1).^[1] However, the details of the steps from compound **116** to **117** are still unknown.

In 2017, the Hu group found the BGC which encodes the biosynthesis of rubratoxins **62** and **63** from *Penicillium dangeardii* Pitt. The core genes for the monomer formation and the KI-like and PEBP-like enzymes for dimerization were found in the gene cluster. There are six encoded redox enzymes including four α -ketoglutarate- and iron(II)-dependent dioxygenases, one flavin-dependent dehydrogenase and the ferric-reductase-like enzyme *RbtH*. All of these six genes take part in the tailoring steps of rubratoxin biosynthesis. *RbtB*, *G*, *E* and *U* are α -ketoglutarate- and iron(II)-dependent dioxygenases which could hydroxylate the different sp^3 carbons from compound **121** to **127**. And they found that *RbtH* can regioselectively reduces one maleic anhydride moiety in rubratoxin B **82** to the γ -hydroxybutenolide (Scheme 1.6.3.2).^[80]



Scheme 1.6.3.2 A summary of gene cluster and reactions catalysed by different redox enzymes in the rubratoxin **62** pathway.

1.7 Method for biosynthetic studies

1.7.1 Genome sequencing.

Whole-genome sequencing of microorganisms is an efficient way of discovering natural-product biosynthetic gene clusters. *Illumina* sequencing and Oxford nanopore sequencing are two common methods for genome sequencing. At the moment, combining the result from *Illumina* sequencing and Oxford nanopore Sequencing is a good way to get longer reads and more precise genome data.

Illumina sequencing methods are based on reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands (**Fig. 1.7.1.1 A**).^[81] *Illumina* gives short reads which are hard to assemble, but great depth which means it is highly accurate. In Oxford nanopore sequencing single molecules of DNA can go through the nanopore and be sequenced without the need for PCR amplification or chemical labelling of the sample. Nanopore sequencing uses electrophoresis to transport an unknown sample through an orifice of diameter 10^{-9} meters (**Fig. 1.7.1.1 B**).^[82] Oxford nanopore sequencing gives low depth, which means high errors, but very long reads. Combining the two methods is better to give overall, fast, accurate and long sequences, which is ideal for finding BGC.

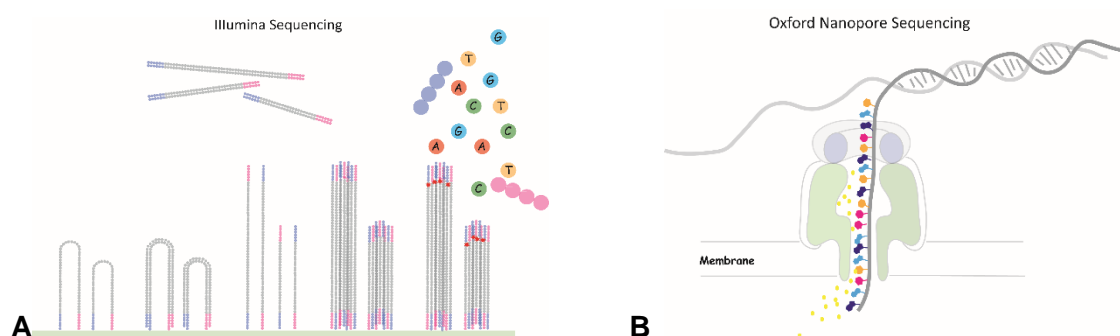


Fig. 1.7.1.1 Diagram of Illumina Sequencing and Oxford Nanopore Sequencing methods: **A**, Illumina sequencing method; **B**, Oxford Nanopore Sequencing.

1.7.2. Bipartite knockout.

Target gene knockout is a useful tool to elucidate the biosynthesis of natural products. A commonly used strategy to disrupt a gene of interest (GOI) from the biosynthetic gene cluster (BGC) is based on stable integration of exogenous DNA into the fungal genome by homologous recombination. The transformation is usually achieved by a polyethylene glycol (PEG)/ CaCl_2 method.^[83]

To increase the level of targeted integration of transformed DNA into the genome, a gene knockout method has been developed based on the bipartite strategy. The method requires transformation with two DNA fragments which overlap in the resistance gene (usually hygromycin resistance gene) to reduce the chance of ectopic integration.^[84]

Each of the two overlapping DNA fragments (left and right fragment) are built up from a region of the gene targeted for disruption, amplified from gDNA (the blue part), and a region of resistance gene (yellow part), amplified from a plasmid which already contains this part (like pTHygGS-eGFP). By yeast recombination, a new plasmid will be constructed which can give a PCR template (**Fig. 1.7.2.1**).

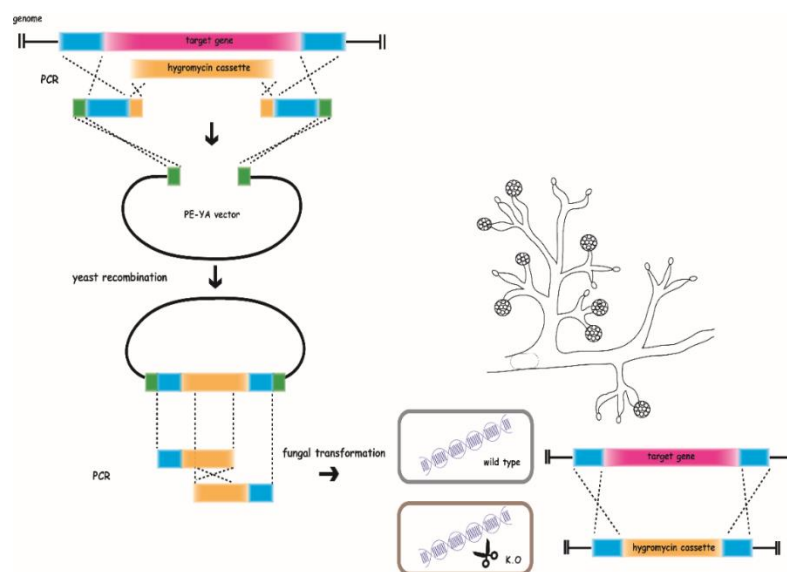


Fig 1.7.2.1 The bipartite gene knockout strategy.

1.7.3 CRISPR-Cas9 gene editing.

Because sometimes bipartite knockout is not an efficient method for the creation of the real K.O transformants, gene editing methods like the CRISPR-Cas9 systems are good choices to try. CRISPR is an abbreviation of Clustered Regularly Interspaced Short Palindromic Repeats. The CRISPR/Cas system is a prokaryotic immune system which protects the cell from the invasion of virus particle.^[84] In bacteria, there are three steps for CRISPR-Cas9 working with virus. First of all, when the virus invades the cell, the Cas protein like CAS1 and CAS2 can cut a DNA fragment (protospacer) on the foreign DNA, and integrate the protospacer into the original CRISPR sequence (**Fig.1.7.3.1 a**). Then the new CRISPR sequence from the first step transcribes into pre-crRNA and tracrRNA (a sequence binds to crRNA), and they will assemble with Cas9 protein finally change into mature CRISPR-Cas9 system (**Fig.1.7.3.1 b**). The complex

will work on the foreign DNA to cut it up (Fig.1.7.3.1 c). Fig.1.7.3.1 d is an example of the use of CRISPR-Cas9 to introduce a donor gene to edit the genome.^[86]

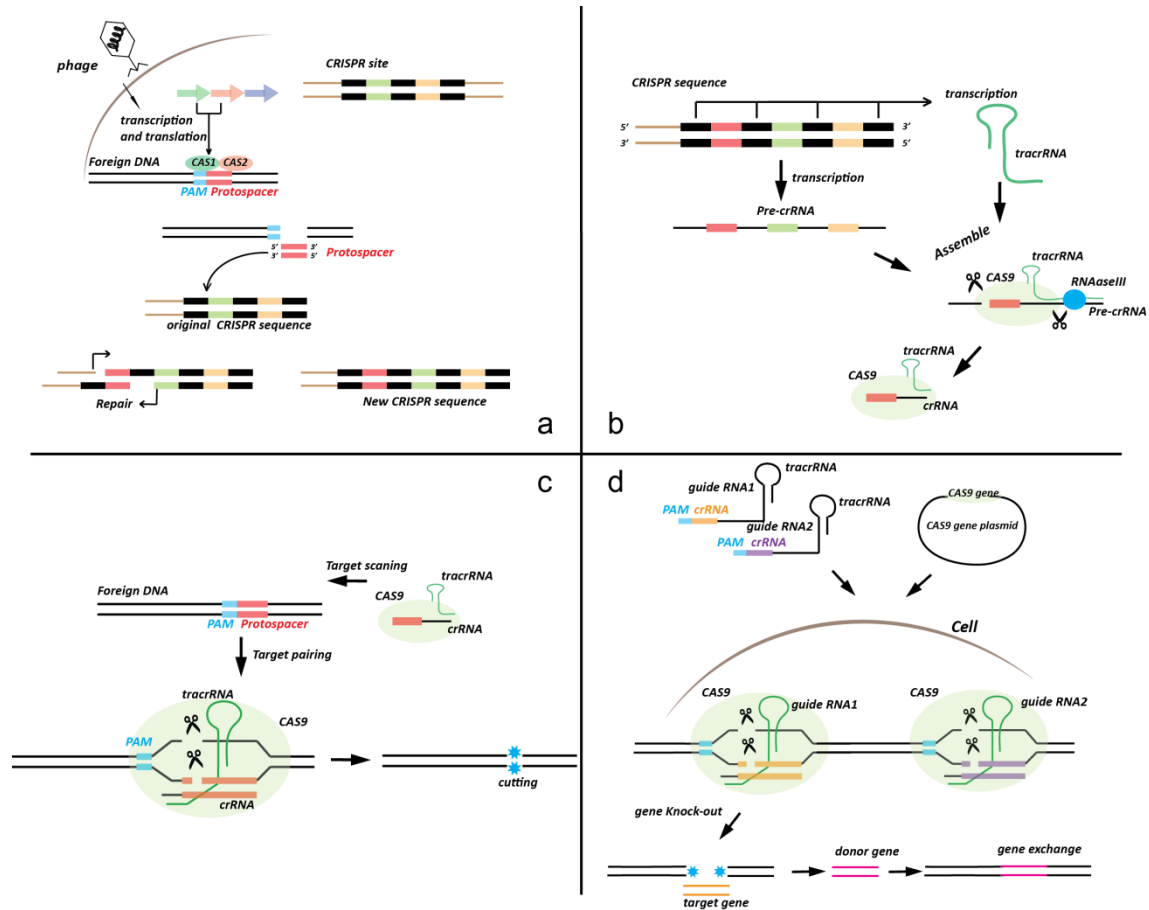


Fig. 1.7.3.1 The working mechanism and the application of CRISPR-Cas9 system.

In recent years, several methods to apply CRISPR-Cas9 technology in filamentous fungi have been published.^[87-89] Amongst the first is a single vector based method developed by Nødvig *et al.* This strategy was chosen for application in *T. viride* due to convenient cloning and transformation methodology.^[89]

1.7.4 Silencing.

Post-transcriptional gene silencing is a method of gene interference which relies on degradation of targeted mRNA, intercepting transfer of genetic information between the transcriptional and translational machinery. Gene silencing has been extensively used in studies of gene function,^[90-92] and has been successfully applied to the characterisation of fungal genes controlling polyketide biosynthesis such as chaetoglobosin A in *Penicillium expansum* and bikaverin in *Fusarium fujikuroi*.^[93, 94]

The most commonly exploited pathway leading to post-transcriptional gene silencing is the RNA interference (RNAi) system (**Fig. 1.7.4.1**). Initiation of RNAi is usually *via* introduction of double stranded RNA (dsRNA), a form of RNA generally aberrant to eukaryotic organisms, but common in viruses,^[95] and from transcription of the inverted repeat regions of transposons.^[90] As a defense against these invasive genetic parasites the RNAi pathway relies minimally on two proteins, Dicer and Argonaute. Dicer is an RNase which degrades dsRNA into short, 21-22 bp fragments known as small interfering RNA (siRNA).^[96] These siRNA fragments bind to Argonaute as part of the RNA-induced silencing complex (RISC)^[97] which, guided by the anti-sense component of the siRNA, seek out and catalytically degrade complementary mRNA.

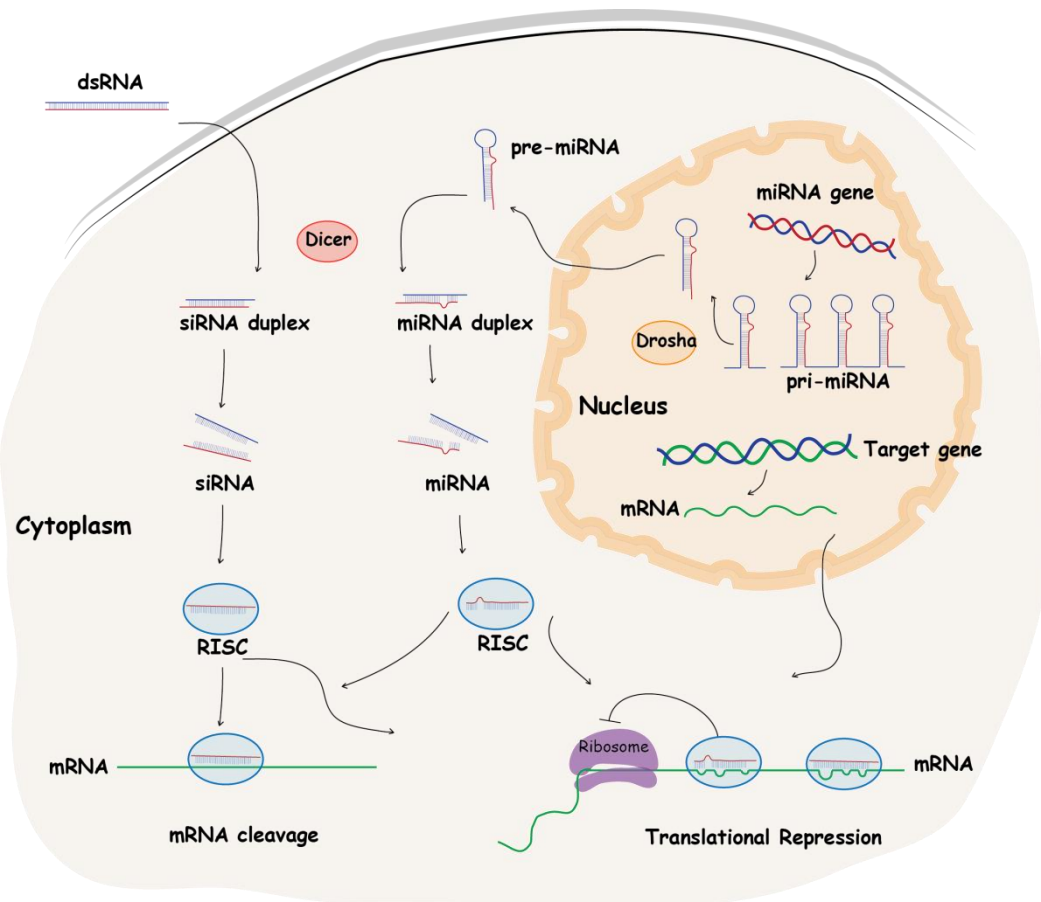


Fig. 1.7.4.1 The RNA interference pathway.

Activation of the RNAi pathway can be achieved directly by introduction of siRNA to the organism of interest or by incorporation of a transgene expressing targeting RNA. The simplest form of targeting RNA is antisense to the sequence to be downregulated (asRNA). If asRNA can diffuse to its sense partner before it reaches the ribosome the two complementary strands anneal, forming dsRNA and removing the target from the mRNA pool.^[98]

Annealed dsRNA will also function as a substrate for Dicer and activate the RNAi pathway for catalytic degradation of targeted mRNA. However, this method is limited by the requirement of asRNA to diffuse to its sense target before annealing, and by potential barriers to annealing

due to formation of strong RNA secondary structure. Strategies have been designed to increase dsRNA levels, including vectors which express both sense and antisense sequences from the same locus, and constructs expressing a hairpin RNA with both sequences on the same transcript.^[99] These advanced methods can be considerably more effective than a simple asRNA approach, but at a cost of more complex vector construction (**Fig. 1.7.4.2**).

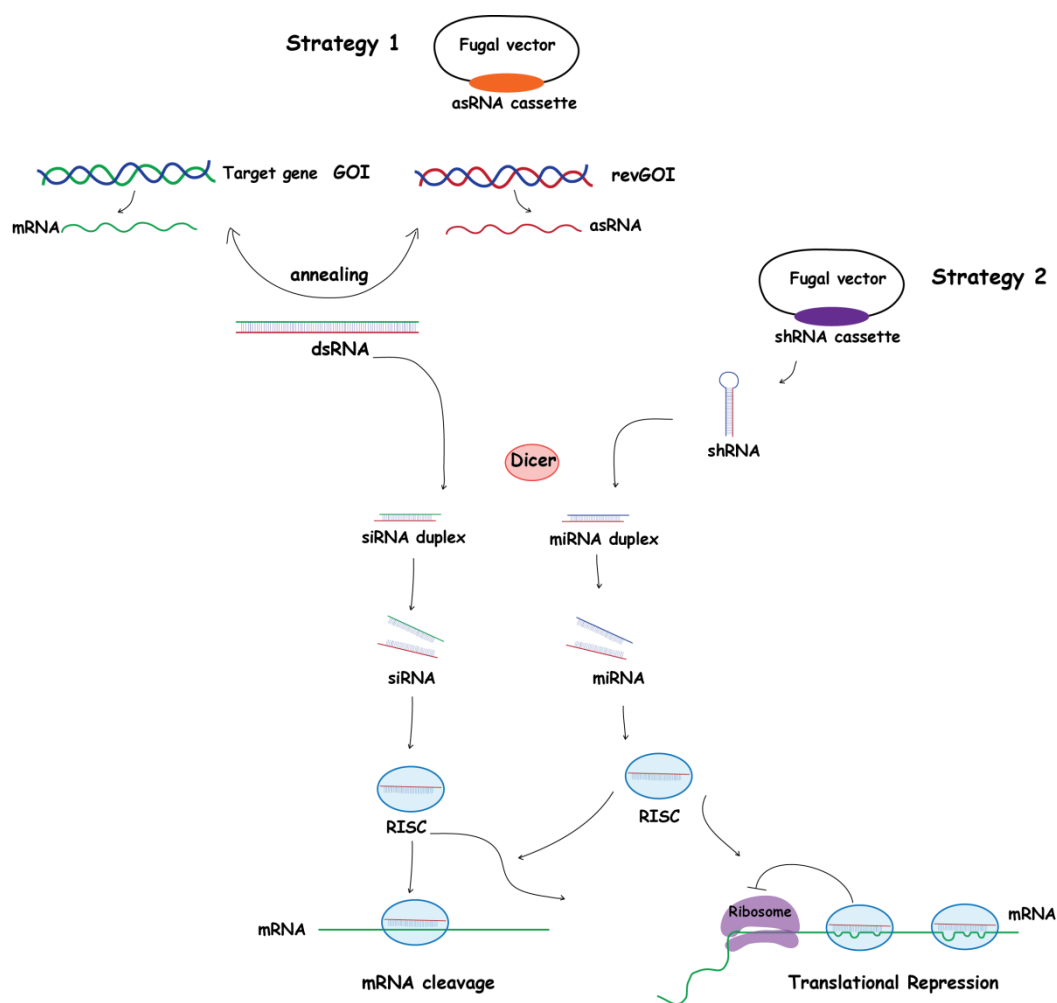


Fig. 1.7.4.2 Two different strategy for silencing.

1.7.5 Heterologous expression.

To explore the biosynthetic pathway of the natural products, heterologous expression is also a very useful tool. Heterologous expression hosts for secondary metabolites or specific enzymes in the biosynthetic pathway include *E. coli*, yeast and some fungi like *A. oryzae* and *T. reesei*. Compared to *E. coli* and yeast expression systems, fungi have the advantage of being able to heterologous express multigene clusters. *E. coli* and yeast have the disadvantage that, they cannot process eukaryotic introns and often possess a significant codon bias. Bacteria also have

difficulty in correctly folding fungal polypeptides. So *E. coli* and yeast are better to overexpress proteins for *in vitro* study.

The first successfully expressed complete multigene fungal cluster was the tenellin **3** gene cluster in 2010.^[100] The biosynthesis of tenellin was reconstituted by expressing four genes responsible for the production of tenellin in an arginine auxotrophic *A. oryzae* M-2-3 strain. Later, more and more complete and partial biosynthetic pathways have been expressed in fungi such as that encoding byssochlamic acid **1** and squalestatin **7** biosynthesis.^[3, 101]

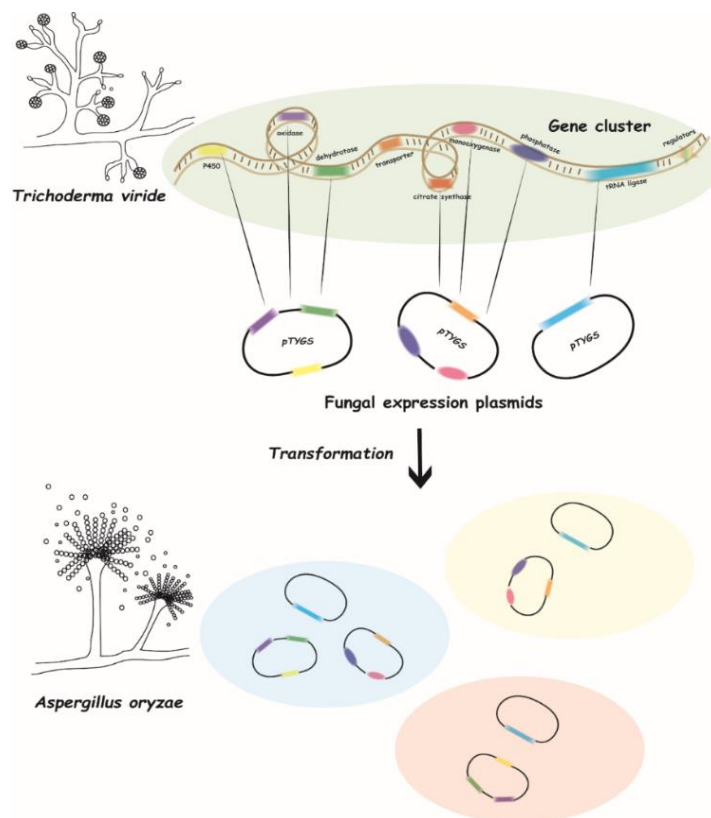


Fig. 1.7.5.1 Cartoon example for heterologous expression in *A. oryzae*.

A. oryzae strain NSAR1 is deficient of arginine ($\Delta argB$), sulphate reductase (sC^-), adenine ($adeA^-$) and ammonium ($niaD^-$) metabolism which enable its use as a multi-selection marker system. In addition, the natural sensitivity of *A. oryzae* NSAR1 towards the antibiotics bleomycin and the herbicide glufosinate (ble^R and bar) can be two more selection markers. The pTYGS vectors are a series of fungal expression vectors which are available with all of the auxotrophic and sensitive selection markers ($argB$, sC , $adeA$, $niaD$, ble^R and bar) containing four sets of fungal promoter/terminator (P/T_{amyB} , P/T_{adh} , P/T_{gpdA} and P/T_{eno}), so *A. oryzae* strain NSAR1 and the vectors are enough for the heterologous expression of up to 16 individual genes.^[102]

1.8 Overall aims.

The main aims of this thesis is trying to understand the biosynthesis of alkyl citrates produced by filamentous fungi: viridofungin **2** and byssochlamic acid **1**. At the same time, programming of hr-PKS (TenS) from tenellin **3** biosynthetic pathway will also be explored by domain swap experiments.

Using most up to date methods for natural product biosynthesis such as gene knockout, RNA-interference based silencing, CRISPR-Cas9 gene editing, *in vivo* study of heterologous expression in *A. oryzae* and *T. reesei* and *in vitro* studies with *E. coli* and yeast expression should uncover the functions and mechanism of genes involved in the three biosynthetic pathways.

In particular, the viridofungin **2** project is focused on finding more evidence on the biosynthesis of the citrate and amino acid moieties in the structure. In addition, the oxidation of C-13 is also unknown.

For the byssochlamic acid **1** project, more effort will be put into understanding the exact substrate and stereochemistry of citrate synthase and 2-methyl citrate dehydratase. The chemical property of the maleic anhydride monomer and the dimerization mechanism of maleidrides will additionally be explored by *in vitro* study.

Finally, the tenellin **3** project will focus on the research of the molecular basis of methylation and chain-length programming in fungal iterative highly reducing polyketide synthase by protein domain swaps.

Chapter 2. Exploring the Biosynthetic Pathway of Viridiofungin

2.1 Background and aims of the project

2.1.1 Introduction

The previous researches on viridiofungins were focused on natural product isolation and structure elucidation, biological activities and total synthesis. However, the biosynthetic pathway of viridiofungins is still unknown. Based on structure analysis (section 1.1.1), viridiofungin **2** belongs to the family of alkyl citrates. The alkyl citrate moiety is made by the condensation of a fatty acid (or polyketide) and oxaloacetate, catalyzed by a citrate synthase-like enzyme (CS). This kind of CS belongs to secondary metabolism (section 1.5.3). Hexylaconitic acid **128**, oryzines **129/130**, piliformic acid **131**, byssoclamic acid **1**, cornexstin **66** and squalestatin **7** are known to be created by alkyl citrate biosynthetic pathways (Fig. 2.1.1.1). The biosynthesis of these compounds should be similar in the early steps with viridiofungin.

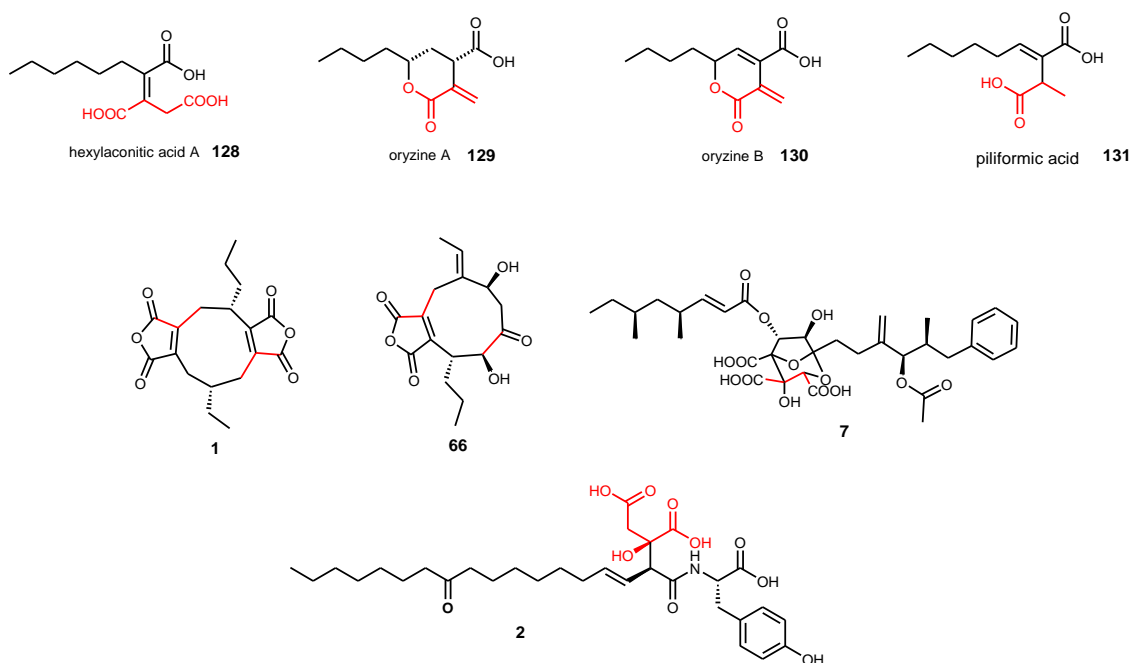
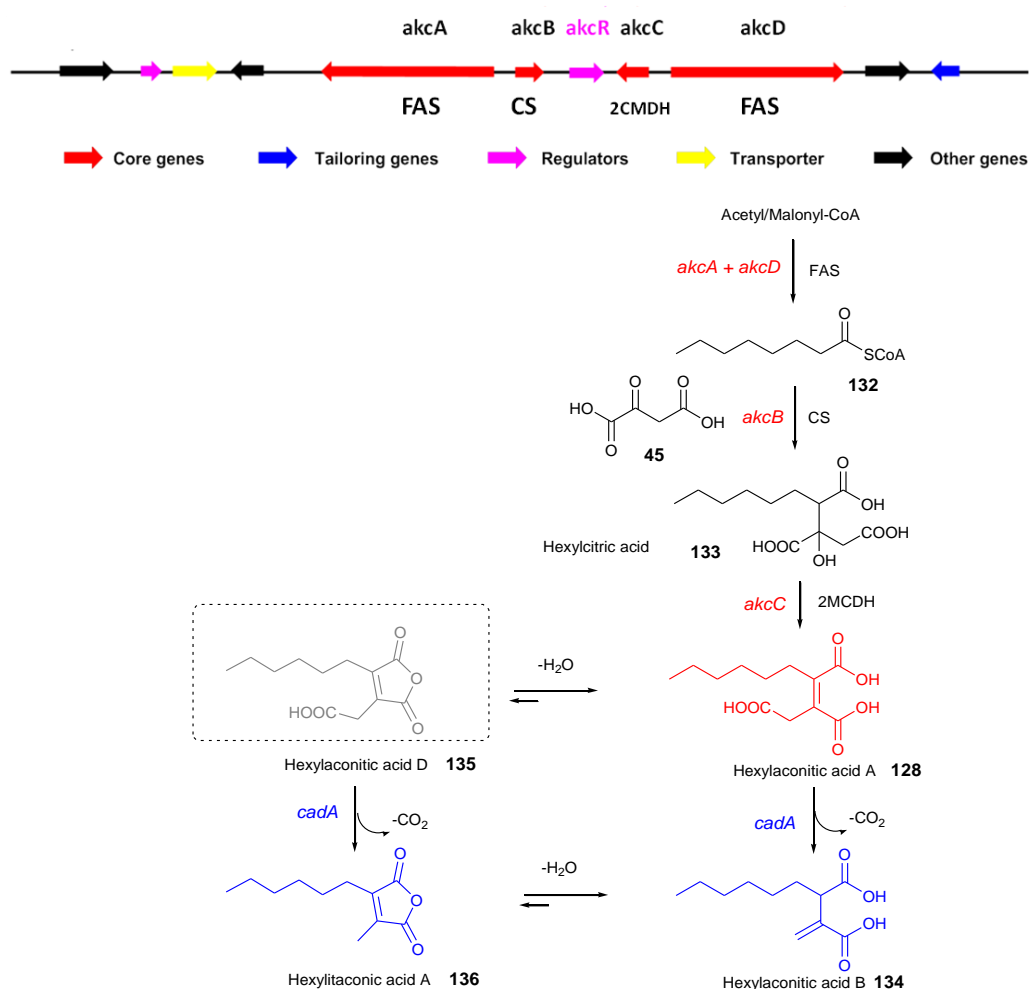


Fig. 2.1.1.1 Examples of alkyl citrate (oxaloacetate part in red).

The Pham group reported a possible hexylaconitic acid **128** biosynthetic gene cluster in *A. niger* NRRL3. They performed BLASTP to locate orthologous gene clusters using the sequences of fungal fatty acid synthase and the citrate synthase. The orthologous gene cluster contained FAS genes (*akcA* and *akcD*) as well as CS (*akcB*) and DH (*akcC*) genes (Scheme 2.1.1.2).^[103, 104] AkcA and AkcD catalyze the formation of a fatty acyl CoA **132**. AkcB then catalyzes the condensation

II Introduction

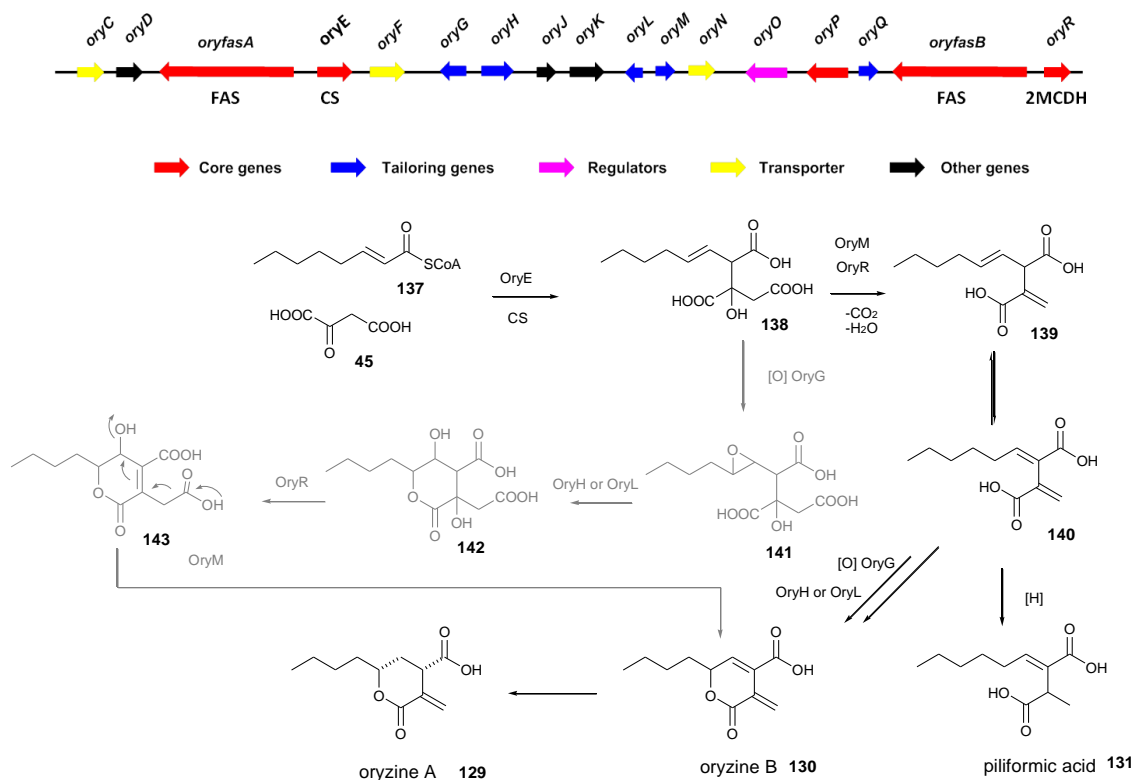
of the fatty acyl CoA **132** and oxaloacetate **45**. Then citrate product **133** is then dehydrated by AkcC to form **128**. By overexpressing the transcription regulator (*akcR*), the Pham group obtained hexylaconitic acids **128**. When they overexpressed both *cis*-aconitate decarboxylase gene *cadA* (this gene is outside the hexylaconitic acid cluster in *A. niger* NRRL3) and *akcR* together, the products shifted the profile of alkylcitrate production from hexylaconitic acids (red) **128** to mainly hexylitaconic acid (blue) **134/136** (Scheme 2.1.1.1). This study showed that phylogenomic analysis together with experimental manipulations (introduction of a decarboxylase) can be used to reconstruct a more complete biosynthetic pathway in generating a broader spectrum of alkylcitric compounds.^[103]



Scheme 2.1.1.1 Proposed biosynthetic pathway of hexylaconitic acid.

Other examples of FAS based alkylcitrate biosynthetic pathway are involved in the biosynthesis of oryzine A **129** and pilifomic acid **131**.^[105, 106] The Cox group discovered two new metabolites in extracts of *A. oryzae* M-2-3 with an unusual maleidride backbone, which were named oryzine A **129** and B **130**. One gene cluster was found to contain fungal FAS genes with α and β module

(*oryfasA* and *B*), citrate synthase (*CS*), 2-methyl citrate dehydratase (*oryR*) and a decarboxylase (*oryM*), with a series of tailoring other genes in the BGC. Like hexylaconitic acid **128**, the biosynthesis of the oryzines **129** also appears to use a C₈ fatty acid **137** to produce the basic structure of the molecule (**Scheme 2.1.1.2**).^[105]



Scheme 2.1.1.2 Proposed biosynthetic pathway of oryzine A **129** and B **130**.

As demonstrated in section **1.3** and **1.4**, the core structures of byssochlamic acid **1**, cornexstin **66** and squalestatin **7** are also derived from the condensation of a secondary metabolism CS (**Fig. 2.1.1.2**).^[1, 3, 100] However, these three compounds were derived from an hrPKS based biosynthetic pathway rather than FAS. The CS is thought to use either a CoA or ACP bound polyketide as substrate combining with oxaloacetate to give a citrate intermediate. It was shown that some hydrolases (*bfLI*/*pvLI*/*Mfm8* and *Mfm10*) were involved in these three BGC (**Fig. 2.1.1.2**). These hydrolases could be involved in the PKS off-loading step. The Cox group heterologous expression study of byssochlamic acid **1** biosynthesis in *A. oryzae* indicated that *bfLI* is important for the production of maleic anhydride.^[3] Compared with PKS, fungal FAS is known to be able to produce fatty acyl CoAs without the need for a separate hydrolase, which could be directly used for the condensation step of CS.^[107] This would explain why there is no hydrolase encoded within the BGCs of oryzine A **129** and hexylaconitic acid A **128**.

II Introduction

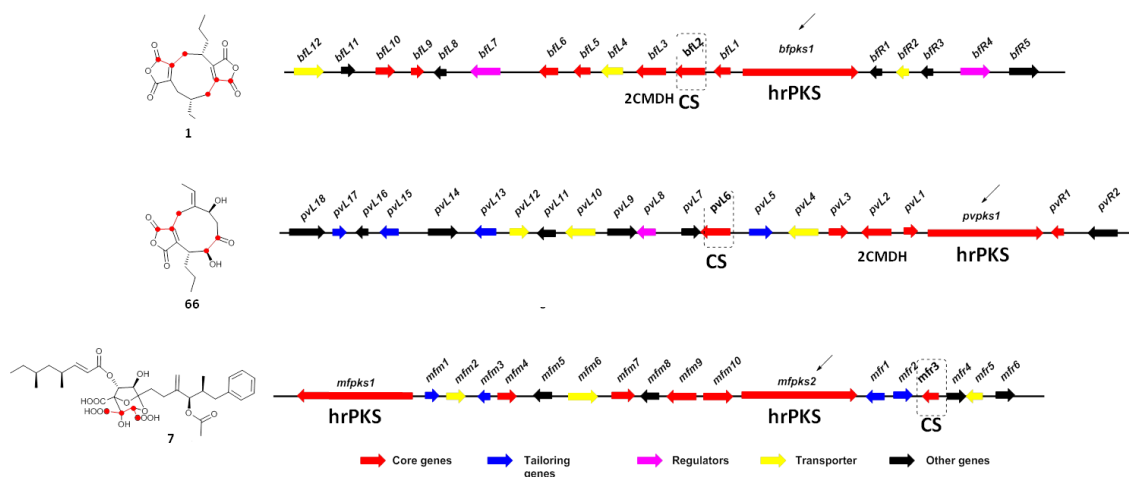
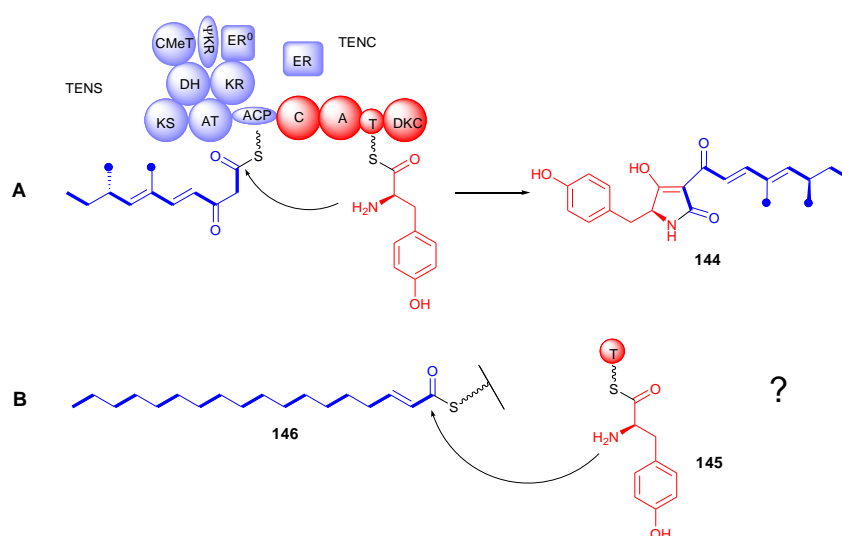


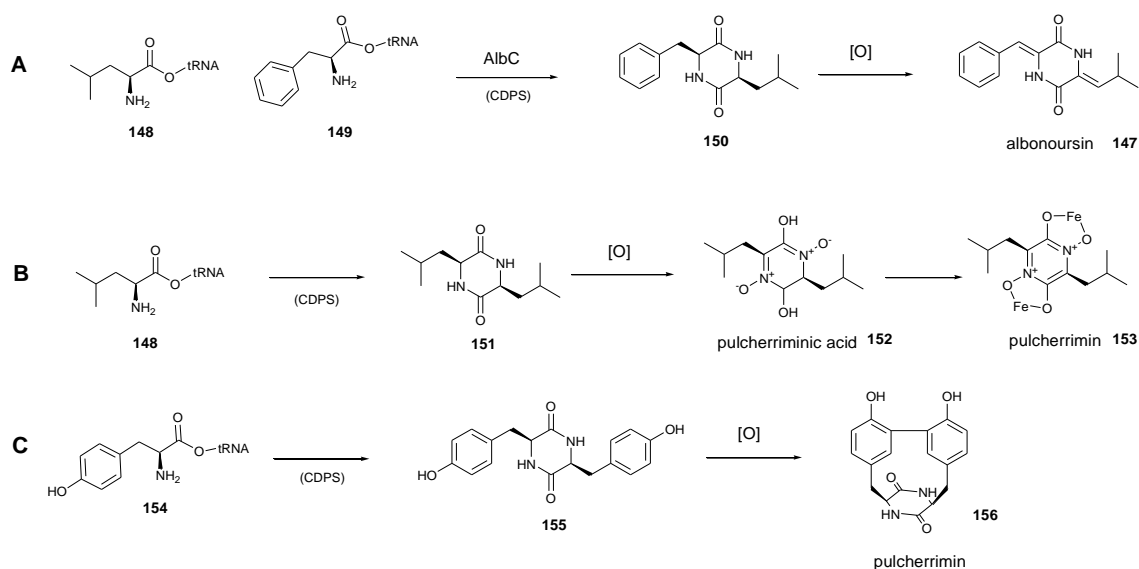
Fig. 2.1.1.2 Gene clusters of different compounds which contain citrate synthase inside.

The most important and interesting part of viridiofungin **2** biosynthesis is the attachment of an amino acid to the citrate structure. As far as we know, NRPS or PKS-NRPS systems are the most common enzymes which can catalyse the attachment of amino acids to polyketides during secondary metabolism.^[108-110] Taking tenellin **3** (a PKS-NRPS derived compound) as an example, the C domain of the NRPS catalyses the condensation of tyrosine with a polyketide, finally producing a pretenellin **144** product (Scheme 2.1.1.3 A).^[100] Therefore, a PKS-NRPS or an NRPS single domain might be also involved in the biosynthetic pathway of viridiofungin **3** which could link amino acid **145** with C₁₈ polyketide or fatty acid **146** (Scheme 2.1.1.3 B).



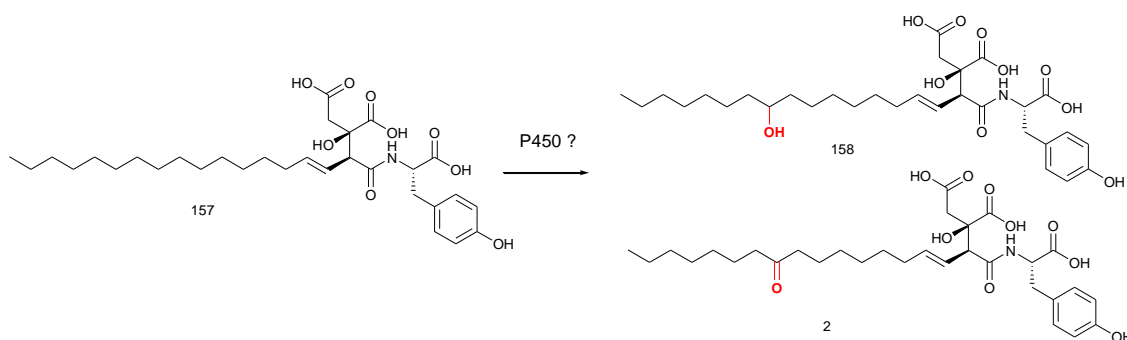
Scheme 2.1.1.3 Biosynthesis of tenellin based on PKS-NRPS and prediction of viridiofungin biosynthesis: A, tenellin **3**; B, viridiofungin **2**.

Other types of NRPS independent amino acid attachment are known. For example, some enzymes have been shown to catalyse the transfer of amino acids from tRNA.^[111] Cyclodipeptides are well known secondary metabolites synthesized predominantly by microorganisms. For a long time, the formation of the cyclodipeptides constituting the diketopiperazines (DKP) scaffolds were thought to be catalysed by NRPS. However, in 2002, the biosynthesis of albonoursin **147** was shown to involve AlbC which is a small enzyme later characterized as a cyclodipeptide synthase (CDPS).^[112] AlbC can catalyse the formation of the cyclo (L-Phe-L-Leu, cFL) intermediate **150**, into which a cyclic dipeptide oxidase (CDO) introduces two α,β -dehydrogenations (**Scheme 2.1.1.4 A**).^[113] In addition, pulcherriminic acid **152** (further converted into pulcherrimin **153**)^[114, 115] and mycocyclosin **156** were also shown that a CDPS involved in their biosynthetic pathways (**Scheme 2.1.1.4 B, C**).^[116]



Scheme 2.1.1.6 Biosynthesis of albonoursin **147**, pulcherriminic acid **153** and pulcherrimin **156** based on CDPS: **A**, albonoursin **147**; **B**, pulcherriminic acid **153**; **C**, pulcherrimin **156**.

The structure of viridifungin also indicates that there should be some tailoring genes included in the viridifungin gene cluster. Oxidation at C-13 of viridifungin might be catalyzed by a P450 monooxygenase, or it could also be a non-heme iron type of oxygenation as observed in the squalstatins **7**^[101] or oryzines **129** (*e.g.* oryG).^[105] These enzymes could catalyze oxidation of intermediate **157** to form a hydroxyl group **158** or a carbonyl group **2** at C-13 (**Scheme 2.1.1.5**).



Scheme 2.1.1.5 Proposed reaction at C-13 position of viridifungin

Based on these analogies to other known biosynthetic pathways we can predict that a biosynthetic gene cluster which encodes the biosynthesis of the viridifungins should include: genes encoding either fatty acid or polyketide production; a gene encoding a citrate synthase; genes encoding either a PKS-NRPS or tRNA-dependent transferase; and genes for tailoring processes such as cytochrome P450 or non-heme iron oxygenases.

2.1.2 Project aims

The overall aim of this project is to explore the biosynthetic pathway of viridifungin. More concretely, we want to achieve the goals as follows:

1. Work on the wild type *T. viride* and find out a suitable condition for reproducing of viridifungin and related compounds.
2. Obtain the genome data of *T. viride* and annotate the genes in the genome.
3. Try to find out the possible gene cluster of viridifungin by using different bioinformatic methods, and analyze the gene cluster to give a prediction of the biosynthetic pathway.
4. Based on the genome data and analysis of the gene cluster, carry out knockout (KO) experiments to confirm the gene cluster of viridifungin.
5. If the gene cluster is confirmed by KO experiment, heterologous expression work will be carried out to express the gene cluster in *A. oryzae* to get more biosynthesis details.
6. On the other hand, gene editing and gene silencing will also be used for knockout or knockdown to analyze the biosynthetic pathway.

2.2 Results

2.2.1 Optimization of production conditions

Trichoderma viride (MF5628, ATCC 74084) was obtained from CBS-KNAW. First of all, different media were used to test the producing conditions of viridiofungin in *T. viride* (section 5.2). Because *Trichoderma* species grow well, all of the cultures were incubated at 28 °C and 160 rpm, in liquid media.

After 7 days growth, different cultures displayed different colours (Fig. 2.2.1.1). In PDB, the fungi produce spores quickly, so the culture turned to green. In YEME, the fungi grow slowly, and there were not so many mycelia in the culture. In DPY, a lot of foam was produced. This might be because of gas-production. In producing medium (cornmeal medium), the culture was not sticky anymore, the cornmeal was digested by fungi.

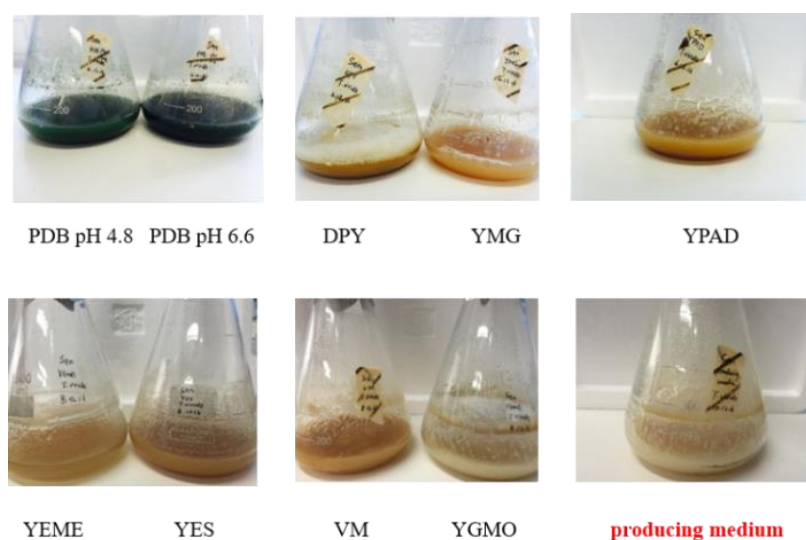


Fig. 2.2.1.1 *T. viride* grown in different media for 7 days (the producing medium is cornmeal medium).

Time-course experiments were carried out by growing *T. viride* in different media. Producing media (cornmeal medium) was the only one which produced the viridiofungin related compounds. The cornmeal in the production medium appears to be important. We found that the strains start to produce viridiofungin after 5 days fermentation, and produce most from 7 days fermentation. However, if the strain grew for a long time, the products would be oxidised (Table 2.2.1.1).

II Optimization of production conditions

Medium	viridiofungin producing			
	3 days	5 days	7 days	10 days
PDB	No	No	No	No
DPY	No	No	No	No
YMG	No	No	No	No
YPAD	No	No	No	No
YEME	No	No	No	No
YES	No	No	No	No
VM	No	No	No	No
YGMO	No	No	No	No
producing medium	No	small amount	Yes	Yes and oxidised

Table 2.2.1.1 Time course of *T. viride* producing condition testing.

100 mL of culture (mixture of media and mycelia) was extracted with 100 mL of ethyl acetate, which was dried ($MgSO_4$) and concentrated. The crude extract was dissolved in 1 mL methanol and analysed by LCMS. According to the literature, viridiofungin A is known to have m/z 591, and viridiofungin B has m/z 575. According to the LCMS results (**Fig. 2.2.1.2**), peaks eluting at 7.9 and 8.9 min corresponded to viridiofungin A and B respectively.

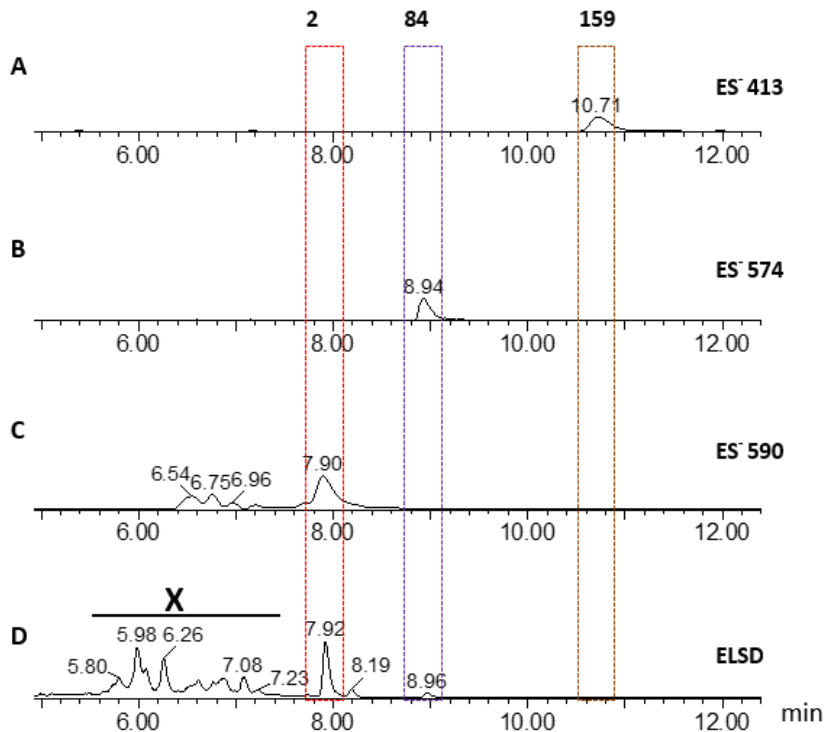


Fig. 2.2.1.2 LCMS result of producing medium culture: **A**, extracted ion chromatogram (EIC) ES⁺ 413; **B**, EIC ES⁺ 574; **C**, EIC ES⁺ 590; **D**, evaporative light scattering detector (ELSD.)

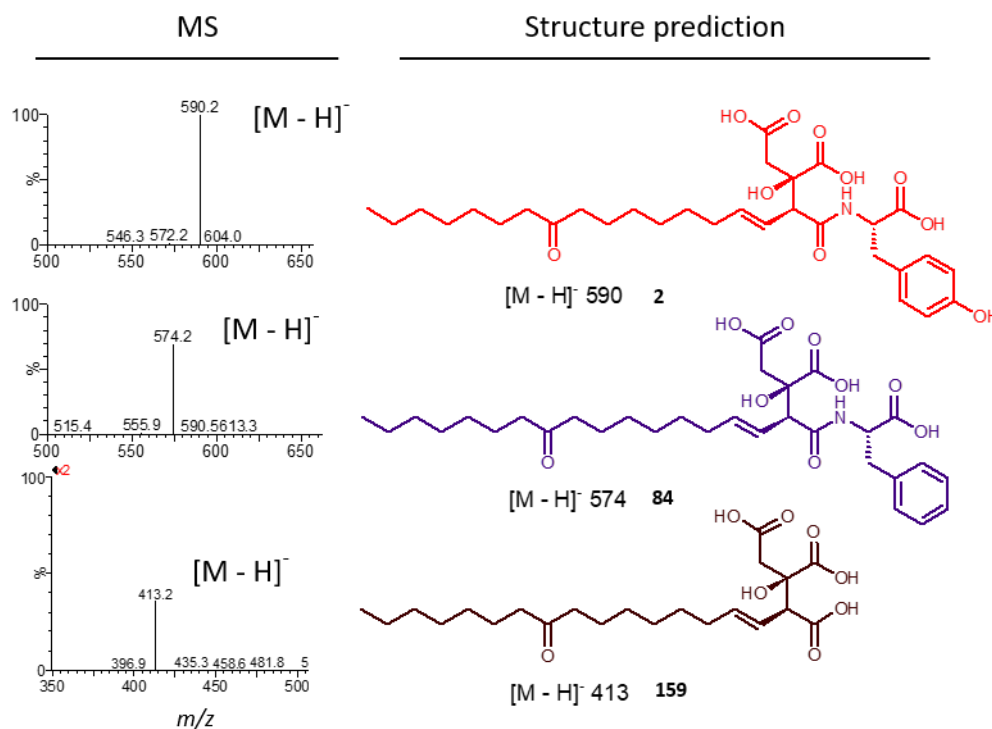


Fig. 2.2.1.3 MS results of the **2**, **84** and **159** peaks.

Compound **2** had $[M - H]^-$ 590 (Fig. 2.2.1.2 C) and compound **84** had $[M - H]^-$ 574 (Fig. 2.2.1.2 B) which were consistent with viridiofungin A **2** and B **84** respectively (Fig. 2.2.1.3). The peak at 10.7 min $[M - H]^-$ 413 (Fig. 2.2.1.2 A and 2.2.1.3) corresponds to citrate intermediate **159**, it had very low concentration which could only be detected by MS. Compounds **X** mixed with each other, but they could also be viridiofungin analogues (Fig. 2.2.1.2 D).

More detailed analysis of the LCMS data showed that there were more than ten compounds eluted in the region between 5 min and 7.5 min (Fig. 2.2.1.4 A). According to the MS result of ten major compounds, **x1**, **x2** and **x7** had $[M - H]^-$ 606. **x3** and **x4** had $[M - H]^-$ 604, **x5** and **x6** had $[M - H]^-$ 590, **x8** and **x10** had $[M - H]^-$ 588, and **x9** had $[M - H]^-$ 592 (Fig. 2.2.1.4 B). These compounds are likely to be derivatives of viridiofungin A and B which are produced by various tailoring enzymes (Fig. 2.2.1.4 C). These compounds are more polar than **2** and **84**, which is consistent with higher levels of oxygenation. The tailoring enzymes could be monooxygenase (+ O atom, + 16 m/z), dehydrogenase ($\pm H_2$, $\pm 2 m/z$), hydratase (+ H_2O , + 18 m/z) and dehydratase ($- H_2O$, - 18 m/z) and combinations.

II Optimization of production conditions

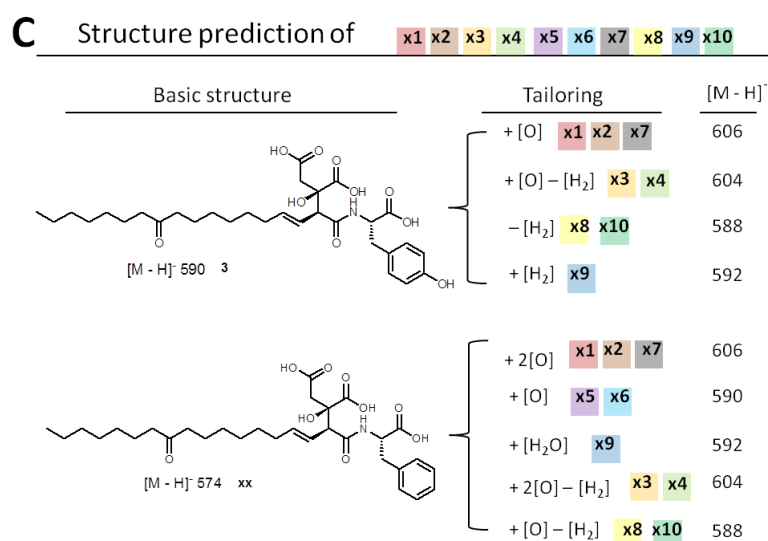
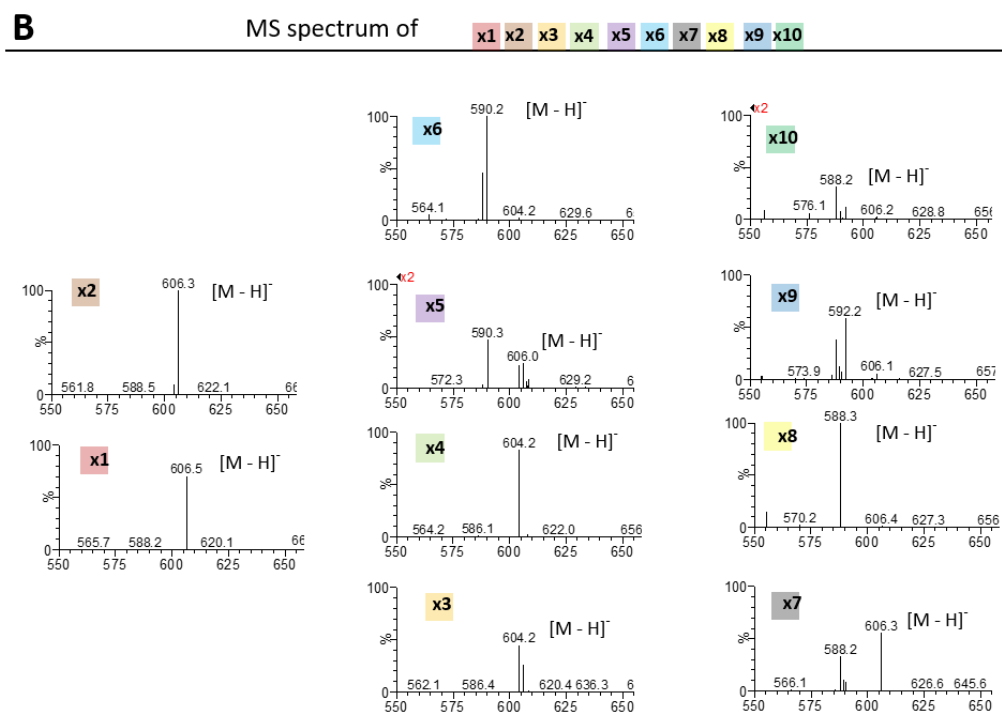
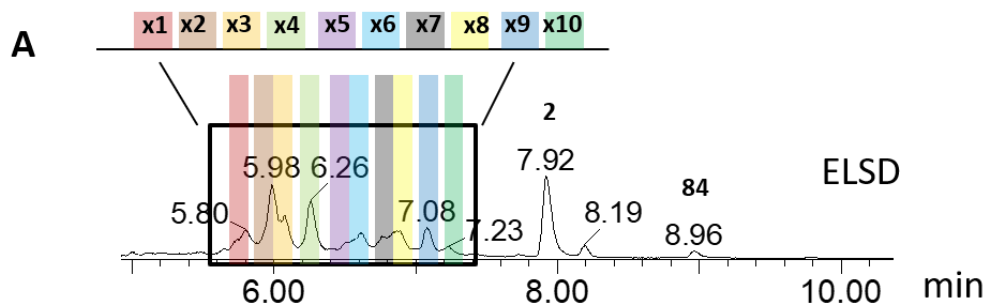


Fig. 2.2.1.4 LCMS result of mixture X: A, ELSD; B, MS spectrum; C, structure prediction.

2.2.2 Viridiofungin purification and characterization

Two liters of culture was fermented for 7 days, and 3 L of ethyl acetate was used for extraction. Finally, 0.9 g of crude extract was obtained. After pre LCMS purification, 1.2 mg of the peak **a** compound was obtained. The UV spectrum of the compound **a** eluting at $t_R = 7.9$ min with a nominal mass of 591 showed maximum UV absorption at 223 nm (**Fig. 2.2.2.1 A**), which is consistent with the literature data for viridiofungin A **2** (**Fig. 2.2.2.2**).

Negative and positive ions provided a nominal mass of 591 ($[M - H]^- = 590$ and $[M + H]^+ = 592$). The ^{13}C NMR spectrum showed 28 unique resonances, with two equivalents by symmetry (C-24 and C-25) and overlapped carbons (C-12/14), for a total of 31 carbons. Four carbons were observed between 172 to 177 ppm which are carboxyl and amide carbons. The acquired ^1H NMR data were identical to the literature data of **2** (**Fig. 2.2.2.3**).

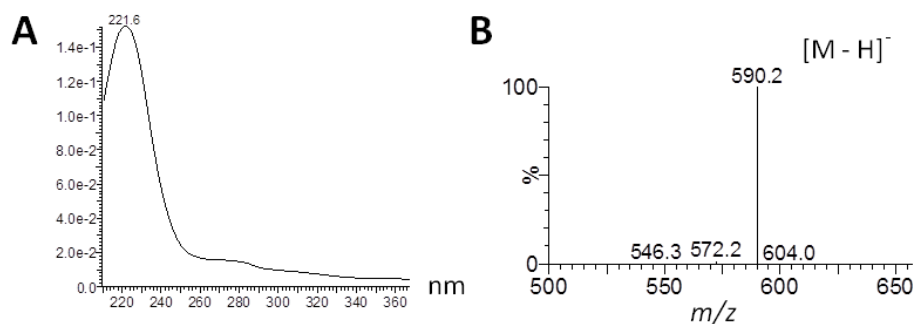


Fig. 2.2.2.1 UV spectrum of peak **2**: **A**, UV; **B**, MS spectrum.

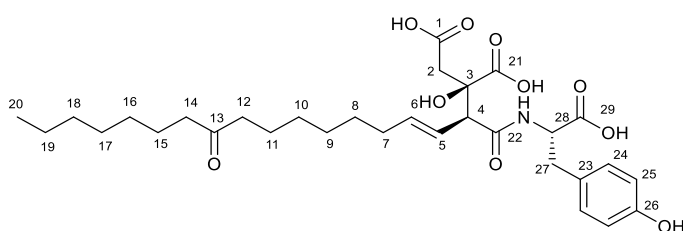


Fig. 2.2.2.2 Structure of viridiofungin A **2**.

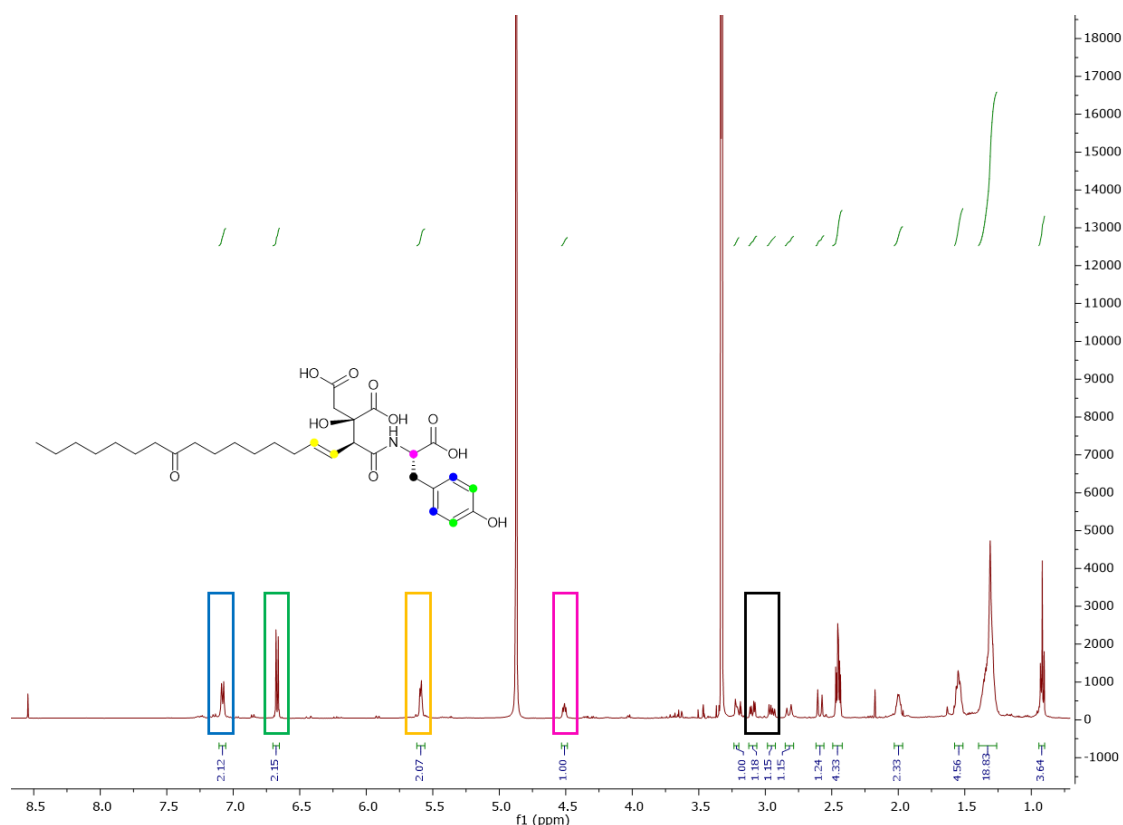


Fig. 2.2.2.3 ^1H NMR spectrum of peak **2**. The most identifiable signal was marked with different colors.

In the literature, the placement of the tyrosine at C-22 was suggested by an EIMS fragment ion from a trimethyl ester of **2** at m/z 296 resulting from neutral loss of the tyrosine and hydroxy-succinic acid moieties. From our experiment, correlation of the tyrosine α -proton to the 172.1 ppm carbonyl (C-22) of **2** using an HMBC experiment optimized for a ^1H - ^{13}C multiple bond coupling of 4.5 Hz allowed the assignment. In the literature, assignment of the side chain carbonyl to C-13 was based upon the observation of an m/z 127 ion in the EIMS of **2** resulting from α -cleavage yielding the odd mass acylium ion.

In summary, by analyzing the HMBC data (**Fig. 2.2.2.4**) and comparing with the data from the literature (**Table 2.2.2.1**), the spectroscopic data of peak **2** was identical with the data of viridifungin A **2**, so compound **2** was viridifungin A **2** produced by *T. viride*.

II Viridifungin purification and characterization

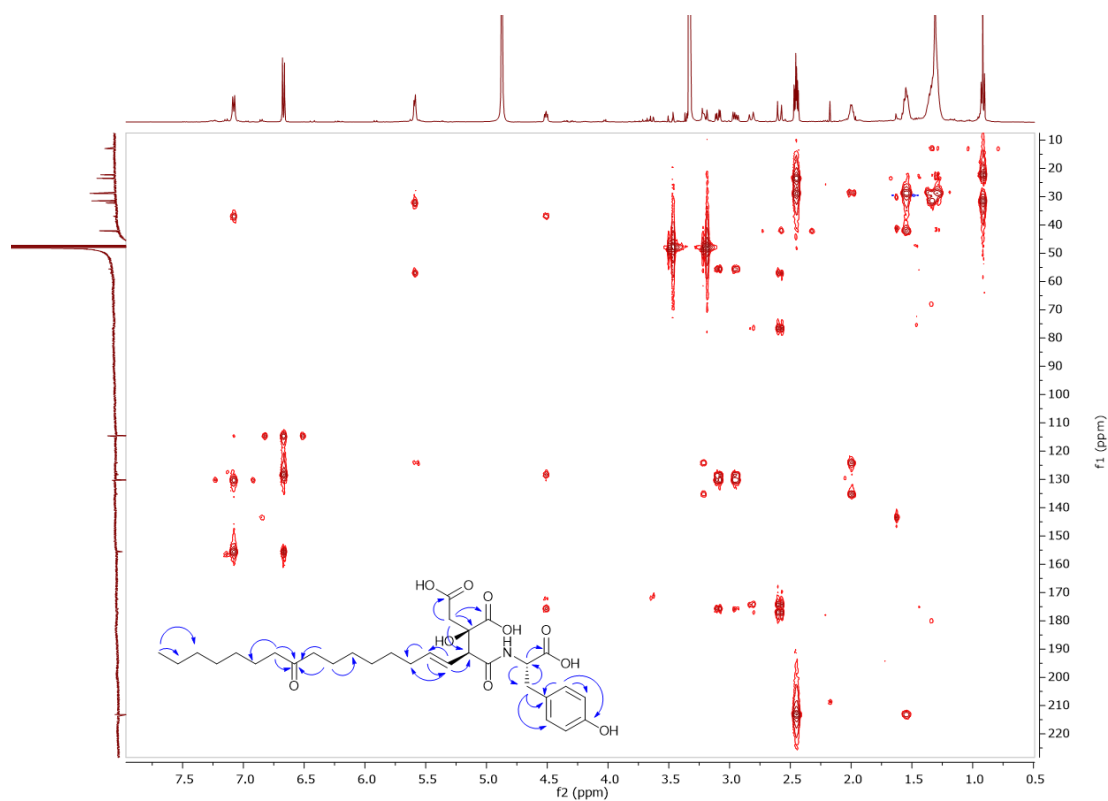


Fig. 2.2.2.4 HMBC spectrum for peak a compound.

position	δ_H /ppm	Mult./J	δ_C /ppm	HMBC (H to C)
1	—	—	174.3	—
2	2.82/2.59	2.82(d, J=15.9Hz, 1H); 2.59(d, J=16.2Hz, 1H)	42.5	1, 3, 4, 21
3	—	—	76.5	—
4	3.21	3.21(d, J=8.4Hz, 1H)	56.9	5, 6
5	5.59	5.59(m, 1H)	135.3	4, 6, 7
6	5.59	5.59(m, 1H)	124.1	5, 7
7	2	2.00(m, 1H)	32.2	5, 6, 8
8	1.31	1.31(m, 2H)	28.7	—
9	1.31	1.31(m, 2H)	28.7	—
10	1.31	1.31(m, 2H)	28.8	—
11	1.55	1.55(m, 2H)	23.4	12, 13, 14
12	2.45	2.45(td, J=7.4Hz, 2.9Hz, 2H)	42.1	13, 14
13	—	—	213.2	—
14	2.45	2.45(td, J=7.4Hz, 2.9Hz, 2H)	42.1	12, 13
15	1.55	1.55(m, 2H)	23.4	13, 16, 17
16	1.31	1.31(m, 2H)	28.8	—
17	1.31	1.31(m, 2H)	28.9	—
18	1.31	1.31(m, 2H)	31.5	—
19	1.31	1.31(m, 2H)	22.3	—
20	0.92	0.92(t, J=7.0Hz, 3H)	13.1	18, 19
21	—	—	177.1	—
22	—	—	172.1	—
23	—	—	155.6	—
24	7.08	7.08(m, 1H)	130.3	25, 26, 27
25	6.67	6.67(m, 1H)	114.6	24, 26
26	—	—	128.2	—
27	3.10/2.95	3.10(dd, J=13.8Hz, 4.9Hz, 1H); 2.95(dd, J=13.8Hz, 6.8Hz, 1H)	37.1	23, 24, 28
28	4.51	4.51(dd, J=8.7Hz, 5.1Hz, 1H)	55.4	27, 29
29	—	—	175.7	—
N-H	—	8.54(s, 1H)	—	—

Table 2.2.2.1 Summary of NMR data of peak a compound and the viridifungin NMR data from literature.

II Viridiofungin purification and characterization

As for the peak **84**, negative and positive ions provided a nominal mass of 575 ($[M-H]^- = 574$ and $[M+H]^+ = 576$) and it had a maximum UV absorption at 224 nm (**Fig. 2.2.2.5**) which is identical with viridiofungin B **84** in literature (**Fig. 2.2.2.6**). However, due to the low concentration (0.1 mg/L, 0.2 mg obtained) of **84**, only ^1H NMR data was obtained, which is consistent with literature data for viridiofungin B (**Fig. 2.2.2.7**).

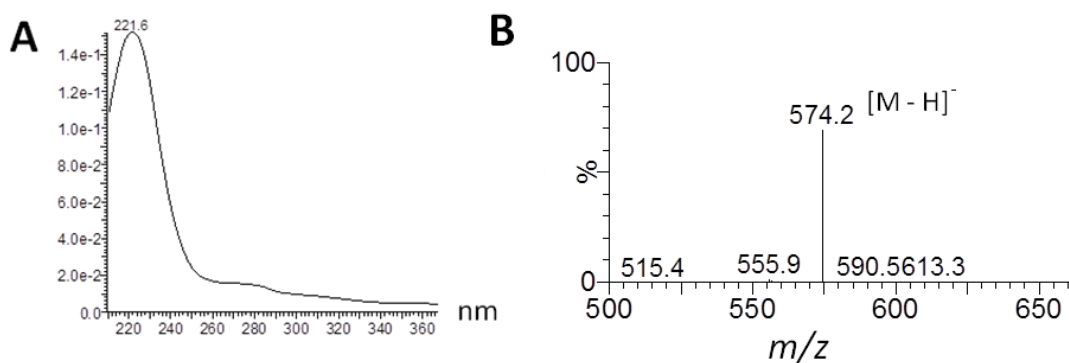


Fig. 2.2.2.5 UV and MS spectrum of peak **84**.

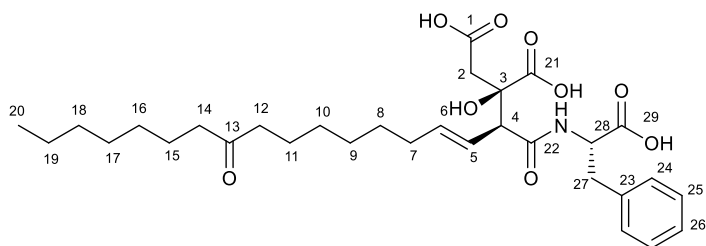


Fig. 2.2.2.6 Structure of viridiofungin B **84**.

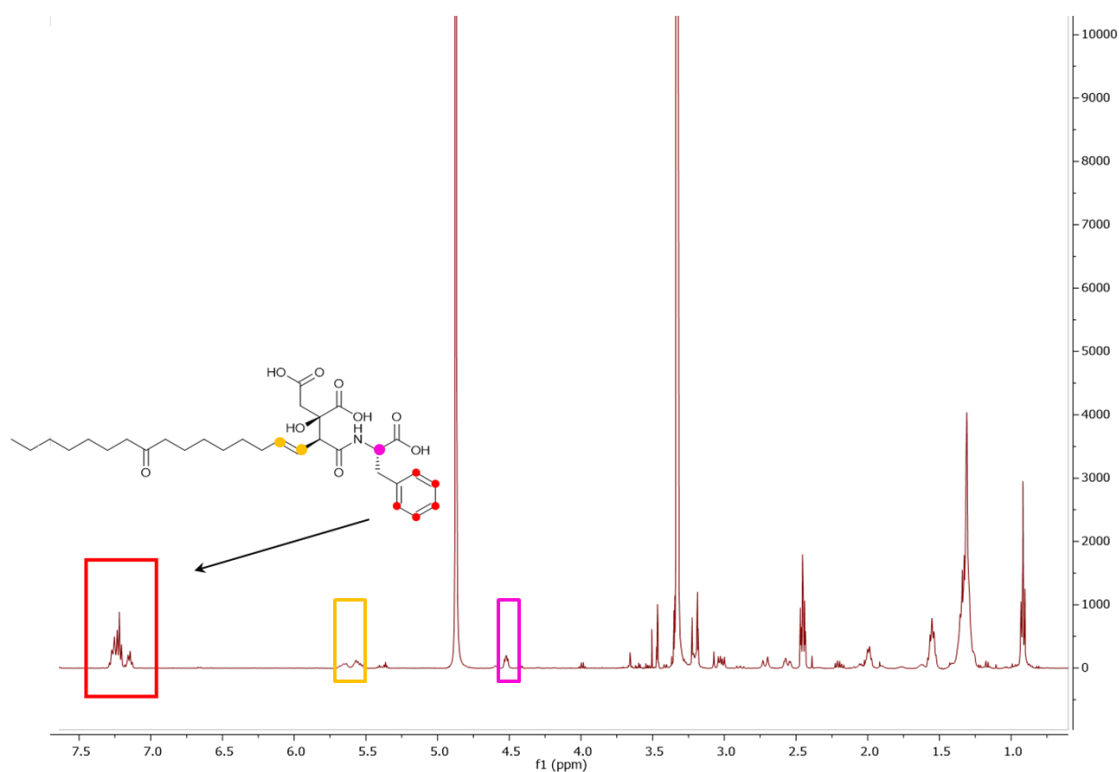


Fig. 2.2.2.7 ^1H NMR spectrum of compound **84**. The signal from benzene was marked with red.

2.2.3 Genome sequencing, bioinformatics and development of a biosynthetic hypothesis

2.2.3.1 gDNA extraction

Studying the biosynthesis of a natural product usually starts with genomic DNA extraction and genome sequencing. In this part of work, GenElute Plant Genomic DNA miniprep kit (Sigma) was used to extract gDNA from 0.5 g frozen dried mycelia of *Trichoderma viride*.

The concentration of the obtained gDNA was 150 ng/ μl (total volume 120 μl). The A_{260}/A_{280} value was 1.79, and the A_{260}/A_{230} value was 2.37. Agarose gel electrophoresis analysis of gDNA showed a single band (Fig. 2.2.3.1). The quality of the gDNA was good enough for sequencing. The gDNA sample was sent for Illumina and Oxford nanopore sequencing.

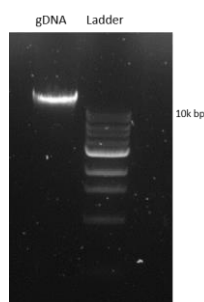


Fig. 2.2.3.1 1% agarose gel electrophoresis analysis of gDNA

2.2.3.2 ITS PCR

Internal transcribed spacer (ITS) refers to the spacer DNA situated between the small-subunit ribosomal RNA (rRNA) and large-subunit rRNA genes in the chromosome or the corresponding transcribed region in the polycistronic rRNA precursor transcript. Because ITS is quite specific from every single species of fungi, it can be considered as a fingerprint of one fungus.^[117, 118] So ITS PCR was performed in this part of work to make sure the species of *Trichoderma viride*. According to ITS sequence alignment, we could find out whether a transformant was from the original strain. 1% agarose gel electrophoresis analysis showed one single ITS PCR product (**Fig. 2.2.3.2 A**).

The PCR product was sequenced and a phylogenetic tree was built (**Fig. 2.2.3.2 B**). The result was compared to ITS sequences from *Aspergillus*, *Penicillium*, *Hypocrea* and *Trichoderma* species, which were downloaded from GeneBank. The ITS sequence of *T. viride* ATCC 74084 has not been previously reported, however, our results show that it clusters closely with the ITS sequences of other known *T. viride* strains (**Fig. 2.2.3.2**).

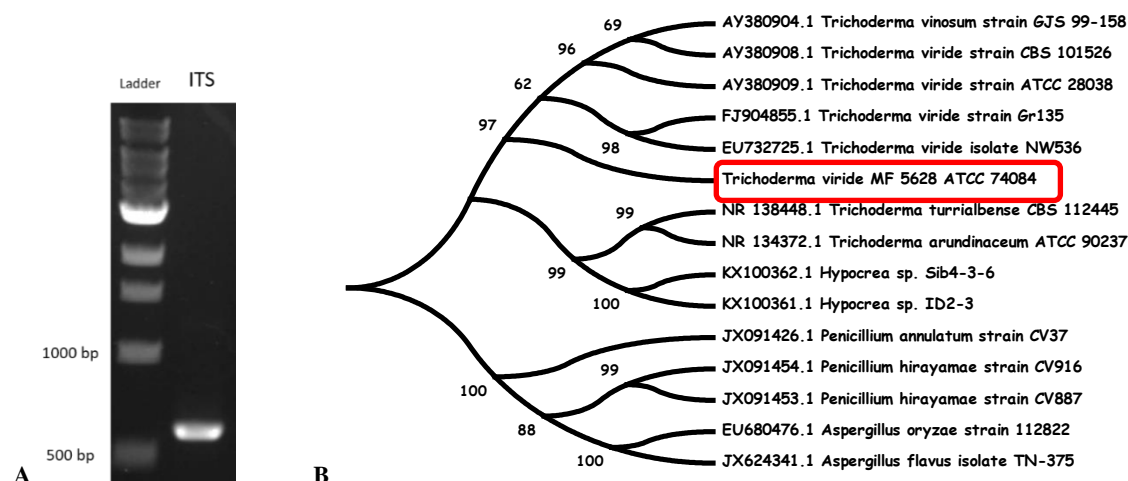


Fig. 2.2.3.2 ITS analysis: **A**, 1% agarose gel electrophoresis analysis of ITS PCR product; **B**, a neighbor-joining tree derived from the nucleotide sequences of ITS sequences from different fungus.

2.2.3.3 Genome sequencing and gene cluster prediction

Genomic DNA (gDNA) of *T. viride* was obtained by GenElute Plant Genomic DNA miniprep kit extraction of freeze-dried fungal mycelia (taken from 3 d old PDB liquid culture). The gDNA was sequenced with Illumina technology. The raw data were processed by Dr Daniel Wibberg using an in-house software platform based on CASAVA 1.8.2.^[119] The sequenced 300 bp reads were assembled with gs Assembler 2.8 with default setting.^[120] From genome assembly data, the genome size of *T. viride* was estimated as 35.9 MB. This is similar to the average size of fungal genomes (31.4 MB). GC content was 48.95%, which was within the normal range for fungal

genomes (32-56%). 100 scaffolds were obtained from the assembly data. N_{50} is a statistic used to assess the quality of genome assemblies. Although, there is no minimum value of N_{50} to determine whether an assembly is of good or bad quality, N_{50} of the *T. viride* sequencing was 108 kb. It means half of the genome was within contigs of sufficient size to encompass complete gene clusters which average around 30 kb in fungi (**Table 2.2.3.1**). However, these values were lower than in some other finished genomes, such as *Sporothrix schenckii* (N_{50} 4.3 MB),^[121] *Sporothrix brasiliensis* (N_{50} 3.8 MB),^[122] *Grosmannia clavigera* (N_{50} 1.2 MB)^[123] and *Penicillium griseofulvin* (N_{50} 2.8 MB).^[124]

Because a secondary metabolism CS is considered to be involved in the biosynthetic pathway of viridifungin, the CS genes from the squalestatin **7** BGC (*mfr3*) and byssocolamic acid **1** BGC (*bfl2*) were chosen to search the genome data (at DNA level). Three citrate synthase – encoding genes were found: Tcirideg1094.t1 was predicted to be a mitochondrial CS; Tcirideg8362.t1 was a predicted 2-methylcitrate synthase which was also likely to be from primary metabolism. Tvirideg7173.t1 was predicted to be the most likely secondary metabolism CS enzyme. It is similar to CS from the Squalestatin **7** and Maleidrides **1** BGC. However, Illumina Sequencing data was not good quality enough to obtain the whole gene cluster of viridifungin **2**. Because the CS gene was exactly at the terminus of one scaffold, and it was difficult to find the other part of the cluster because of the limited information about the biosynthesis of viridifungin **2**.

To get better genome data, the genome was sequenced again. This time the Oxford nanopore method was chosen for sequencing. Therefore higher quality and longer gDNA was needed.^[125] After genome sequencing and assembly, 52 contigs were obtained with an N_{50} of 2145 kb, and 35 of these contigs were more than 50 kb (**Table 2.2.3.2**). Then the Oxford nanopore genome data was polished with Illumina data, finally, a good quality genome data was obtained.

Sequencing method	Total contigs/ scaffolds	N_{50} (kb)
Oxford nanopore	52 contigs	2145
<i>Illumina</i>	100 scaffolds	108

Table 2.2.3.1 Summary of genome assembly in *T. viride* genome

The whole sequencing data was submitted to *antiSMASH* online software and database for gene cluster analysis (<https://fungismash.secondarymetabolites.org>).^[126] There were 32 possible PKS, NRPS and terpene related gene cluster found in the genome of *T. viride* (**Table 2.2.3.2**). But there was no CS in any of these gene clusters.

Type	number	total number
PKS	9	32
NRPS	13	
PKS-NRPS	6	
terpene	4	

Table 2.2.3.2 Summary of gene cluster prediction by antiSMASH.

After bioinformatic analysis and annotation of the genome based on *InterPro*,^[127, 128] one gene cluster was found to be the possible gene cluster of viridifungin. Firstly, CS like enzymes involved in the secondary metabolism were searched in the genome. Three citrate synthase like genes were found in the genome (Tcirideg1094.t1, Tcirideg8362.t1, Tvirideg7173.t1), but only one namely Tvirideg7173.t1 was not from primary metabolism. A 20 kb genomic DNA fragment, centred on Tvirideg7173, was then selected for more detailed manual annotation. Twelve open reading frames were found using the software pBlast (**Table 2.2.3.3**). The cluster was not recognized by Fungi SMASH, because it does not contain genes encoding core synthases such as PKS, FAS, NRPS or terpene cyclase. If the biosynthesis used a fatty acid from primary metabolism then the associated BGC would not have to encode a FAS or PKS.

Original name	Gene	Protein	protein size(AA)	Putative function based on InterPro	Putative cofactor
Tvirideg7169.t1	<i>tvL4</i>	TvL4	435	cytochrome p450 monooxygenase	NAD(P)
Tvirideg7170.t1	<i>tvL3</i>	TvL3	402	acyl-CoA oxidase	FAD
Tvirideg7171.t1	<i>tvL2</i>	TvL2	192	NAD dependent dehydrogenase	NAD(P)
Tvirideg7172.t1	<i>tvL1</i>	TvL1	288	multidrug resistance transporter	/
Tvirideg7173.t1	<i>tvCS</i>	TvCS	366	citrate synthase	/
Tvirideg7174.t1	<i>tvR1</i>	TvR1	247	unknown	/
Tvirideg7175.t1	<i>tvR2</i>	TvR2	401	Flavin monooxygenase	FAD and NADP
Tvirideg7176.t1	<i>tvR3</i>	TvR3	208	nucleotide phosphatase	divalent metal cation
Tvirideg7177.t1	<i>tvR4</i>	TvR4	659	glutamate-tRNA ligase	ATP
Tvirideg7178.t1	<i>tvR5</i>	TvR5	291	translational repressor protein	/
Tvirideg7179.t1	<i>tvR6</i>	TvR6	298	COG complex component	/
Tvirideg7180.t1	<i>tvR7</i>	TvR7	627	transcriptional regulator	/

Table 2.2.3.3 The possible viridifungin gene cluster found in the *T. viride* genome, and gene prediction

Tvirideg7173.t1 is the core gene CS, and a “tRNA ligase” like enzyme was involved in this gene cluster. This “tRNA ligase” like enzyme was considered to be related to the amino acid linkage step. Moreover, some oxygenase (*tvL4* and *tvR2*) and a dehydrogenase (*tvL2*) were also found in the gene cluster. Upstream of the CS, there are three potential regulators (*tvL 5/ 6* and *7*), and a potential transporter (*tvL1*) which might be involved in resistance to viridiodfungin. Moreover, there is a possible acyl-CoA oxidase (*tvL3*) and a putative nucleotide phosphatase (*tvR3*) encoded by the gene cluster (**Table 2.2.3.3**).

The amino acid sequence of the gene CS and “tRNA ligase” like enzyme were submitted to pBLAST online software to analyse the function and feature of the enzymes (**Fig. 2.2.3.3**). A conserved database (CDD) analysis clearly showed that the protein belongs to the CS_ACL-C_CCL superfamily. Sequence alignment revealed the conserved active site and CoA binding site.

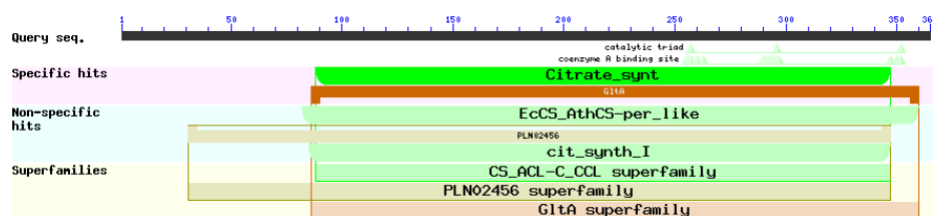


Fig. 2.2.3.3 Graphical summary of CS enzyme BLAST result

Based on pBLAST results, sequences with high similarities were chosen to build an evolutionary tree. Different CS sequences from primary metabolism (*E. coli*, *Aurantimicrobium mintutum*, *Geittrinema sp.*, *T. thermophilus* and *P. furiosus*) and 2-methyl-citrate synthase from *S. typhimurium* were downloaded from GeneBank. In addition, various CS sequences from secondary metabolism (*B. fulva* (BfL2), *P. divaricatus* (PvL7) and *Phoma sp.* MF5453 (MfR3) and TvCS of *T. viride*) were also chosen to build up a phylogenetic tree. According to the neighbor-joining tree of different CS and 2-methyl-citrate synthase, TvCS from *T. viride* had the closest evolutionary relationship with the CS from the squalestatin **7** biosynthetic pathway (green). Two CS (BfL2 and PvL7) from maleidride biosynthetic pathway had closest relationship (blue). Four CS from secondary metabolism had relatively far evolutionary relationship with the ones from primary metabolism (red). The bacterial 2-methyl-citrate synthase (yellow) had the farthest distance with TvCS (**Fig. 2.2.3.4**).

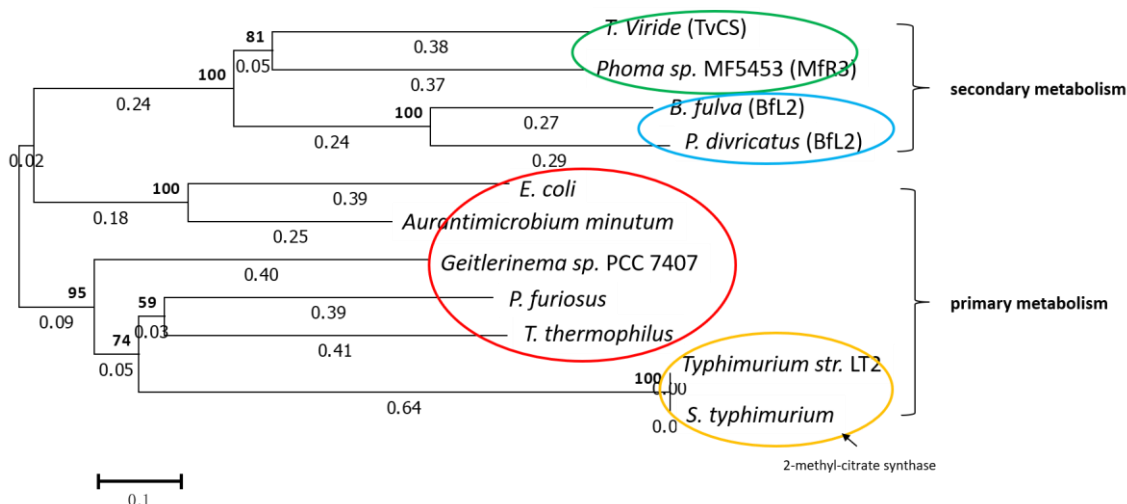


Fig. 2.2.3.4 A neighbor-joining tree derived from the amino acid sequences of CS and 2-methyl-citrate synthase from secondary and primary metabolism.

According to the structural information from the primary metabolism *T. thermophilus*, His-274, His-320 and Asp-375 are the conserved residues in the active site.^[129] The amino acid sequences of primary metabolic CS from *P. furiosus* and *E. coli* (crystal structure already known), as well as secondary metabolic CS from *B. fulva* (BfL2), *P. divaricatus* (PvL7) and *Phoma sp.* MF5453 (MfR3) were aligned using *ClustalX*. The CS have high similarity with each other, especially the residues nearby the active sites (red line). The results indicate that the three residues (Fig. 3.2.8.1 black frame) in the active site are conserved in both primary and secondary metabolic CS.

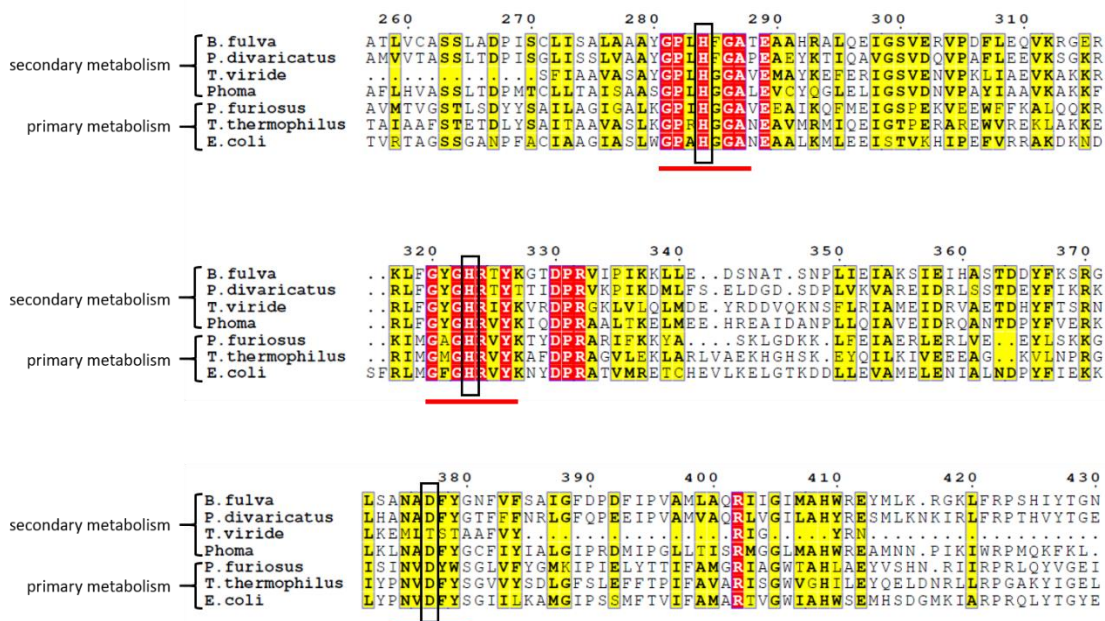


Fig. 3.2.8.1 Multiple alignment of primary metabolic and secondary metabolic CS sequences. Red highlight = 100% similarity; Yellow highlight = more than 60% similarity.

TvR4 is a “tRNA ligase” like enzyme which belongs to the nucleotidyltransferase superfamily. This superfamily includes the class I amino-acyl tRNA synthetases, pantothenate synthetase (PanC), ATP sulfurylase, and the cytidyltransferases, all of which have a conserved dinucleotide-binding domain.^[156] TvR4 also contains a conserved active site HIGH motif, and KMSKS motif for ATP binding (Fig. 2.2.3.4). But TvR4 has a longer amino acid sequence than normal tRNA ligases, meaning that TvR4 might contain other active domains which could catalyse other reactions.

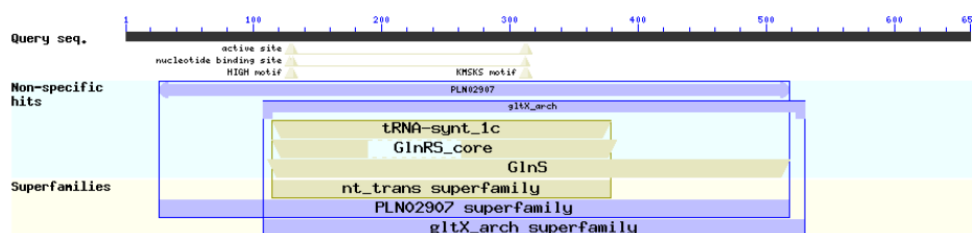


Fig. 2.2.3.4 Graphical summary of TvR4 BLAST result

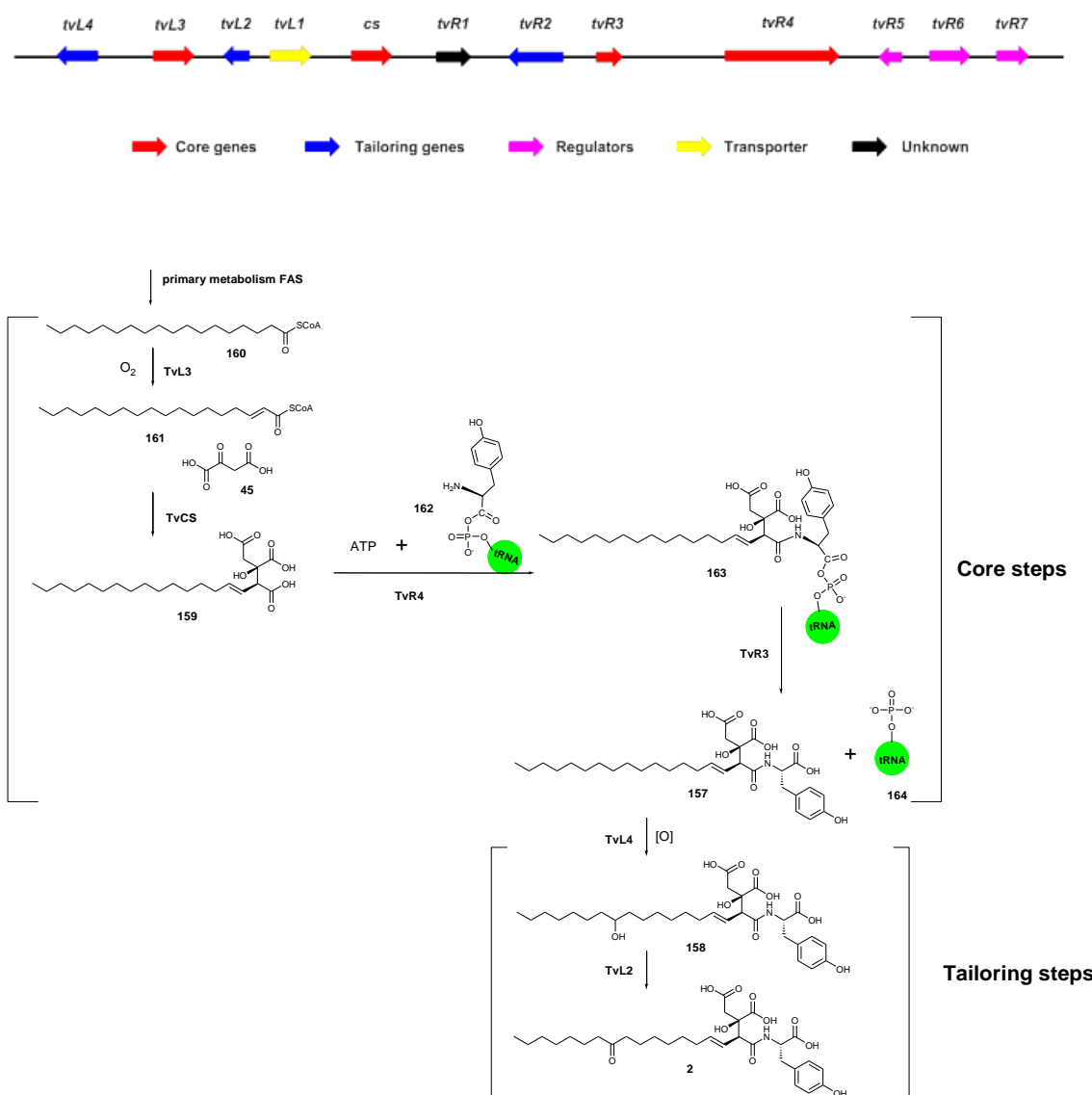
2.2.3.3 Biosynthetic pathway prediction

As is described in section 2.1.1, in nature, there are some NRPS independent processes that add amino acids to other molecules. In the proposed viridifungin gene cluster, a “tRNA ligase” like enzyme (TvR4) and a nucleotide phosphatase (TvR3) could work together with the TvCS. These three genes should be the core genes for connection of the amino acid, primary metabolism fatty acid **160** and oxaloacetate **45**, and additionally release the product from the tRNA (Scheme 2.2.3.1). Acyl-CoA oxidase could catalyse the formation of the α , β – alkene **161** in the fatty acid CoA. The P450 monooxygenase and dehydrogenase are obvious tailoring proteins (Scheme 2.2.3.1). The fatty acyl CoA **160** and aminoacyl tRNA **162** are from the primary metabolism in the *T. viride*.

The reason why TvR4 was thought to be a tRNA related enzyme is because of a nucleotide phosphatase (TvR3) encoded by the BGC. It could be reasonably explained that TvR3 releases a viridifungin intermediate **157** from a tRNA intermediate **163**. Additionally, the P450 monooxygenase (TvL4) and dehydrogenase (TvL2) could be involved in the tailoring steps to form **157** and final product **2** (Scheme 2.2.3.1).

From this general pathway, a variety of different viridifungins (section 1.4.3) could be formed by using different fatty acid chain length and different tRNAs.

II Transformation condition testing



Scheme 2.2.3.1 The proposed biosynthetic pathway of viridifungin **2**.

2.2.4 Transformation condition testing

2.2.4.1 Antibiotic test

Based on the proposed biosynthetic pathway, a series of knockout experiments were planned. In fungi knockouts are normally achieved by inserting a suitable antibiotic resistance cassette into the target gene. Antibiotic testing is the first step of the process. Hygromycin is an antibiotic produced by the bacterium *Streptomyces hygroscopicus*, it is an aminoglycoside that kills fungi and higher eukaryotic cells by inhibiting protein synthesis.^[130] In this part of work, different concentrations of hygromycin were used to test the minimum concentration to kill wild type *Trichoderma viride*.

Seven different concentrations of hygromycin PDB plates were set up: 0 µg/ml, 25 µg/ml, 62.5 µg/mL, 125 µg/mL, 187.5 µg/mL, 250 µg/mL and 375 µg/mL. Wild type *T. viride* was

incubated on these plates at 28 °C for 10 days . *T. viride* could not grow on plates that contained hygromycin at 125 µg/mL or higher (Fig. 2.2.4.1), so these concentrations can be used to select transformants.

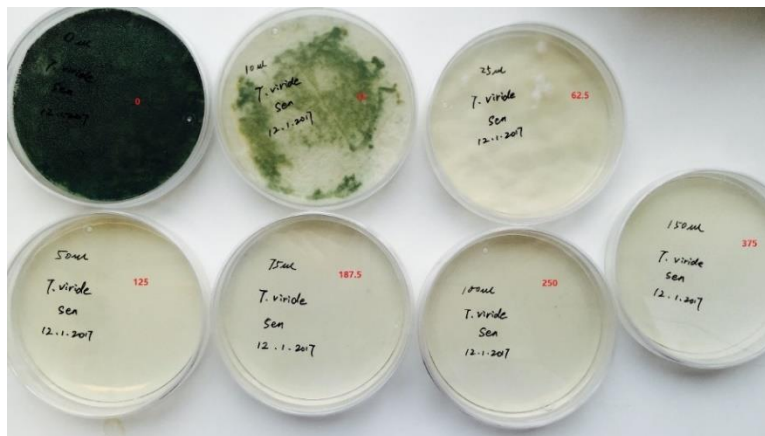


Fig. 2.2.4.1 Result of Testing of antibiotics. different concentration of hygromycin: 0 µg/ml, 25 µg/ml, 62.5 µg/ml, 125 µg/ml, 187.5 µg/ml, 250 µg/ml and 375 µg/ml.

2.2.4.2 Preparation of *T. viride* protoplasts

Protoplasts are cells that have had their cell wall completely or partially removed using either mechanical or enzymatic means. Protoplasts can be used for DNA transformation for making genetically modified organisms.^[131] Therefore, protoplast preparation is a very important step and the first step for knockout experiment and transformation experiment on *T. viride*.

In order to get a good number and quality of the protoplasts, different isotonic solutions (1.2 M sorbitol, 1.0 M sucrose, 0.8 M MgSO₄, 0.8 M NaCl) were tested. The mycelia were collected from 2-day-old PDB culture by filtration using sterile miracloth as a filter, then rinsed with sterilized distilled water and 0.8 M MgSO₄. The mycelia were added to 10 mL of sterilized solutions (1.2 M sorbitol, 1.0 M sucrose, 0.8 M MgSO₄, 0.8 M NaCl) and 300 mg of Trichoderma lysing enzyme and incubated at 30 °C, 120 rpm. After three hours fermentation, 5 µl of digestion solution was checked under the microscope.

II Transformation condition testing

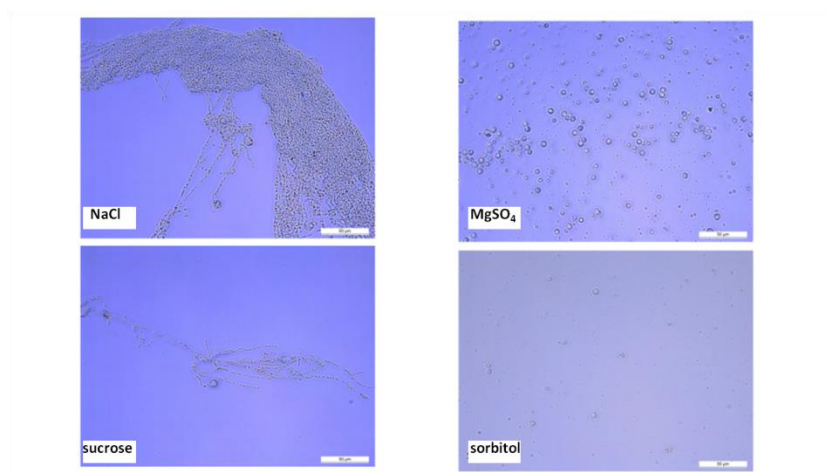


Fig. 2.2.4.2 Result of 3 h of enzyme digestion in different solutions. 1.2 M sorbitol, 1.0 M sucrose, 0.8 M MgSO₄, 0.8 M NaCl

The results showed that in NaCl and sucrose solutions, no protoplasts appeared; in sorbitol solution, only a few protoplasts could be produced; but in MgSO₄ solution, a large number of protoplasts were observed. So the solution: 0.8 M MgSO₄, 20 mM sodium phosphate pH 5.6, was then used for the preparation of protoplasts of *T. viride* (**Fig 2.2.4.2**).

In parallel, different media (PDB, GNB, ME) were used to grow the fungi, and different solutions and different digestion times were also checked. *T. viride* grew most mycelia in PDB medium, and it could produce enough protoplasts in 0.8 M MgSO₄ solution and digested for more than three hours (**Table 2.2.4.1**). The best condition is as follows: 2g mycelia, digesting solution: 0.8 M MgSO₄, 20 mM sodium phosphate pH 5.6, digesting time: 3 h, temperature: 30 °C.

Media	isotonic solution	protopalst producing			cells/mL
		1h incubation	2h incubation	3h incubation	
PDB	0.8 M NaCl	No	No	No	
	0.8 M MgSO ₄	No	small amount	Yes	8.32 × 10 ⁷
	1.2 M sorbitol	No	No	small amount	
	1.0 M sucrose	No	No	No	
GNB	0.8 M NaCl	No	No	No	
	0.8 M MgSO ₄	No	No	No	
	1.2 M sorbitol	No	No	No	
	1.0 M sucrose	No	No	No	
ME	0.8 M NaCl	No	No	No	
	0.8 M MgSO ₄	No	No	No	
	1.2 M sorbitol	No	No	small amount	
	1.0 M sucrose	No	No	No	

Table 2.2.4.1 Best condition for protoplast preparing

2.2.4.3 Transformation of *T. viride*

Trichoderma viride ATCC 74084 has not been reported to have been transformed before. As a control transformation we decided to first transform *T. viride* with a pre-existing eGFPe/*hyg*^R plasmid already in use in our group. This would allow us to test several aspects of the transformation. First it would allow the protoplast mediated protocol itself to be assessed. Second, it would test the *P_{gpdA}* and *P_{amyB}* promoters used for the *hyg*^R and eGFP genes respectively. Finally, observation of green fluorescence would confirm correct translation and protein production

Protoplasts were prepared as section 2.2.4.2, then they were resuspended in 0.5 ml 1M sorbitol-10 mM Tris-HCl, pH 7.5. The suspension was diluted 3 times for counting. The concentration of protoplast was about 8×10^7 / ml. This concentration was enough for transformation (Fig. 2.2.4.3).

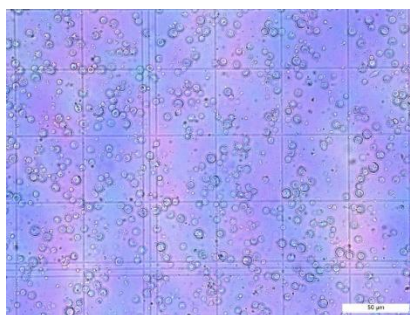


Fig. 2.2.4.3 Calculate the number of protoplasts.

The protoplasts were transformed with pTH-GS-eGFP, and the transformants were selected on HYG plates. The pTH-GS-eGFP plasmid contains a HYG selection marker expressed from *P_{gpdA}*. The eGFP gene was driven by *P_{amyB}* with *T_{amyB}* at the 3' end (Fig. 2.2.4.4).

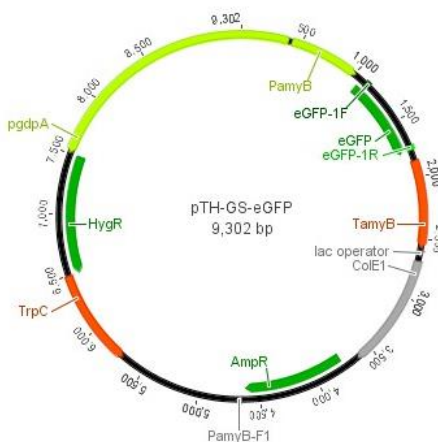


Fig. 2.2.4.4 Plasmid map of pTH-GS-eGFP.

A PEG-mediated transformation method was used for the *T. viride* transformation (**section 5.2.x**). After the first round of selection, the transformants grow on the 125 µg/mL HYG PDA plates. 10 to 20 transformants were observed on each plate (**Fig. 2.2.4.5 A**). These transformants were then transferred to the second and third round of HYG selection to obtain stable transformants. After two rounds of additional selection, 10 out of 40 of the colonies from the first round selection survived on the tertiary 125 µg/mL HYG PDA plates (**Fig. 2.2.4.5 B**).

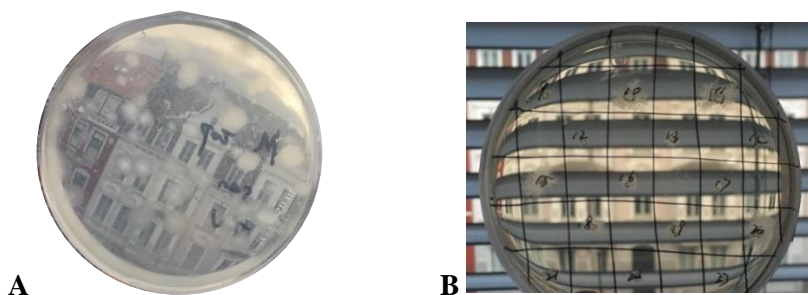


Fig. 2.2.4.5 Transformants on plate: **A**, first round of selection; **B**, third round of selection

Five transformants from the tertiary plate were then grown on PDB agar plates. After 7 days, they were grown in liquid medium for 3 days. The mycelia were observed under a microscope with UV irradiation. But the result showed that there was no difference between wild type and transformants, and there was no detectable fluorescence.

Although transformation was successful, the expression of eGFP was not good enough to see under UV. We assumed that this could be because P_{amyB} is not a strong promoter in *T. viride*. Therefore, the next work on expression of *T. viride* would be change a different promoter to improve the expression level.

2.2.4.4 Promoter exchange

Expression of eGFP in *T. viride* was not successful, the reason might be that P_{amyB} from *A. oryzae* could not be recognised by *T. viride* causing no transcription of *eGFP*. So in this part of the work, a promoter P_{pki} (pyruvate kinase) from *Trichoderma reesei* was combined with eGFP for better expression.

P_{pki} is a constitutive promoter in fungi. The *pki* gene sequences from several other fungal species were download from geneBank. These genes were used to search the *T. reesei* genome. A *pki* gene was found in *T. reesei* by using *artemis*. 4000 bp upstream sequence of *pki* gene was send to *Promoterscan* to predict the promoter sequence. From the prediction tool, 600bp of the

sequence was recognised as the promoter. After finding out the sequence of the P_{pki} , yeast recombination (section 5.5.2) was used to clone P_{pki} upstream of eGFP in the pE-YA plasmid.

The plasmid pEYA- pki -eGFP was then transformed to *E. coli*. eGFP primers were then used for colony PCR. Because P_{pki} contains a *Not* I, five positive colonies of pE-YA- pki -eGFP were also confirmed by *Not* I digestion. All of these 5 colonies were real transformants containing P_{pki} . LR recombination (section 5.7.3) was then performed to get pTYH-GS- pki -eGFP. After LR recombination, colony PCR indicated that the vector had been successfully constructed. Then, three positive colonies were picked for the plasmid extraction and transformation. Using the same method as section 2.1.7, five *T. viride* transformants were obtained after 3 steps of selection (**Fig. 2.2.4.11**).



Fig. 2.2.4.11 pki -eGFP transformants on secondary plate fragment transformation

After 3 days growth in PDB medium, the fluorescence of mycelia was detected under microscope with UV (**Fig. 2.2.4.12**). pki -eGFP transformants had more fluorescence than wild type, but the fluorescent was not so strong. It means P_{pki} from *Trichoderma reesei* can work in *Trichoderma viride* somehow, but this promoter is not a strong promoter, and the expression level of eGFP is low. So, to get better expression, some stronger promoter like P_{gpdA} should be tested in the future.

2.2.5 Gene knockout

Knockout experiments are the most direct way to explore the biosynthetic pathway of a specific natural product. Because the changes in the genome can directly reflect on the production of the secondary metabolites in the native organism. There are two common ways to knock-out the genes. These are: exogenous DNA targeting the gene by recombination; and Crispr-Cas9 mediated gene editing and knockout. Homologous recombination has been very successful previously in the Cox group and will therefore form the main methodology here.

The regular gene knockout is designed to use exogenous DNA to disrupt the targeted gene once the DNA is integrated into the genome. The transformed DNA is designed to contain two sequences homologous to the gene of interest which flank a selection marker on either side. By homologous integration of this KO cassette the targeted gene is disrupted. The common marker for fungal selection is hygromycin B phosphotransferase (hph).^[132] In this part of work, the bipartite KO method (section 1.5.1) will be used to knockout the target genes in the proposed viridiofungin gene cluster.

2.2.5.1 Citrate synthase knockout

2.2.5.1a Gene function analysis

As demonstrated in section 2.1.1, a citrate synthase like gene is likely to be involved in the biosynthetic pathway of viridiofungin 2. According to the analysis of the *T. viride* genome data, one secondary metabolism CS like gene (*TvCS*) was found. The encoded enzyme is similar to the CS involved the biosynthetic pathway of byssochlamic acid 1 and squalestatin 7.^[3, 4] *TvCS* is considered to use oxaloacetate 45 and a C₁₈ fatty acyl CoA 160 as substrates to form an alkylcitrate intermediate 159 (Scheme 2.2.3.1). Because this is likely to be a very early step of viridiofungin biosynthesis, the KO of *TvCS* would abolish the production of viridiofungin.

2.2.5.1b Vector construction and transformation

The *TvCS* gene was amplified from gDNA in two parts (*TvCS-LF* and *TvCS-RF*, from the 5'- and 3'- ends respectively). The primers used (CS-LR and CS-RF) had short tails (30 bp) of sequence homologous to the ends (5' and 3' respectively) of the amplified hygromycin resistance cassette, which was amplified from pTHygGS-Egfp plasmid (using primers HygF and HygR). The specific primers used (CS-LF and CS-RR) contained tails that introduced DNA sequences homologous to the ends of the linearized pE-YA plasmid (Fig. 2.2.5.1).

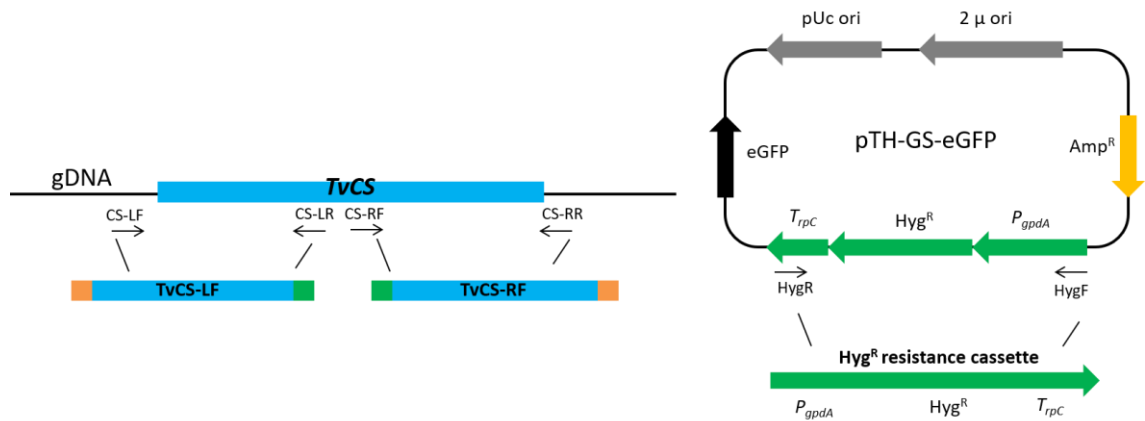


Fig. 2.2.5.1 PCR fragments for yeast recombination.

Upon purification of the PCR products, and using gel electrophoresis to estimate the amounts of DNA to be used, all fragments were transformed into *S. cerevisiae* for recombination (Fig. 2.2.5.2 A), together with the pE-YA vector cut open with *NotI*, using a LiOAc/SS-DNA based transformation protocol (section 5.5.2).

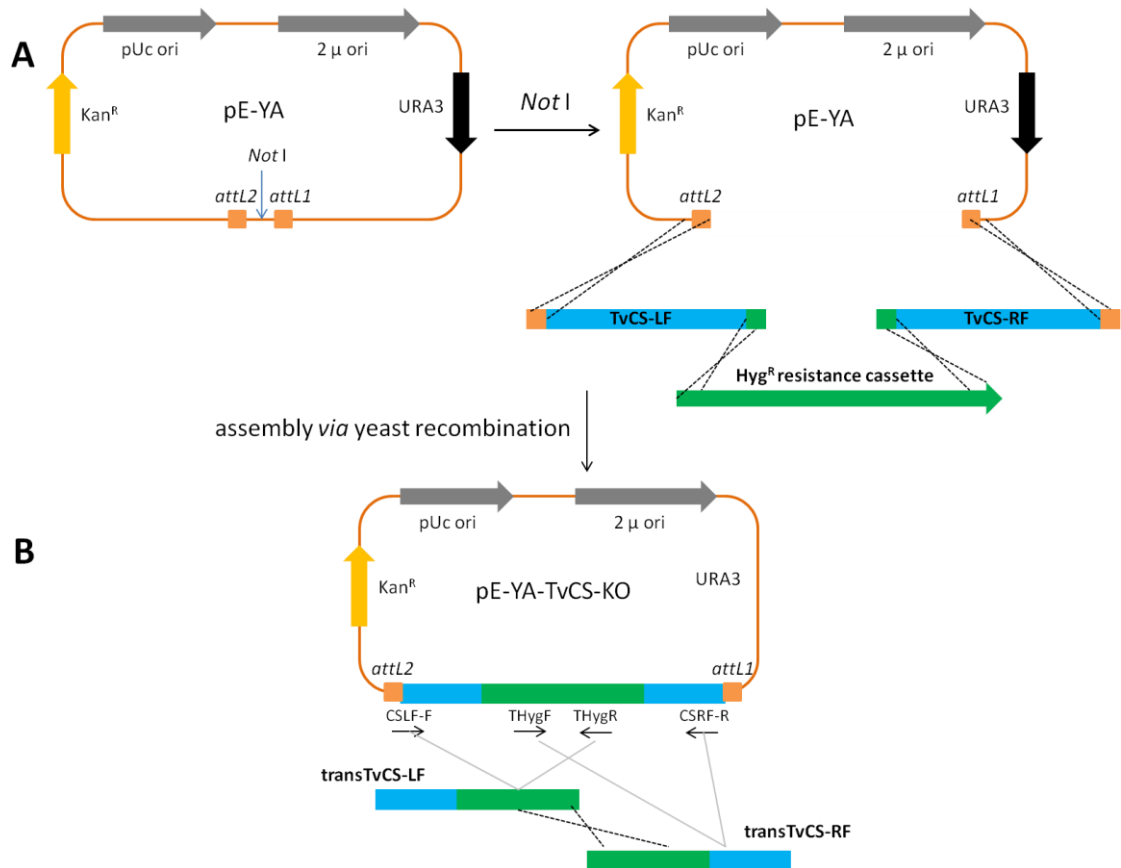


Fig. 2.2.5.2 Yeast assembly and PCR fragment: A, yeast assembly of the pE-YA-TvCS-KO plasmid; B, preparation of the *transTvCS-LF* and *transTvCS-RF* via PCR.

The recombined plasmid was then extracted from yeast and transformed into *E. coli* TOP 10 cells for amplification, which recovered correctly recombined pE-YA-TvCS-KO plasmid (proved by colony PCR). The *transTvCS-LF* and *transTvCS-RF* knockout constructs were then PCR-amplified (CSLF-F/ THygR and CSRF-R/THygF primer pairs) from pE-YA-TvCS-KO plasmid (Fig. 2.2.5.2 B).

Then the left (*transTvCS-LF*) and right (*transTvCS-RF*) PCR products were inserted into *T. viride* by protoplast-mediated transformation (section 5.6.3). Six transformants were obtained after three rounds of selection on hygromycin (Fig. 2.2.5.3).

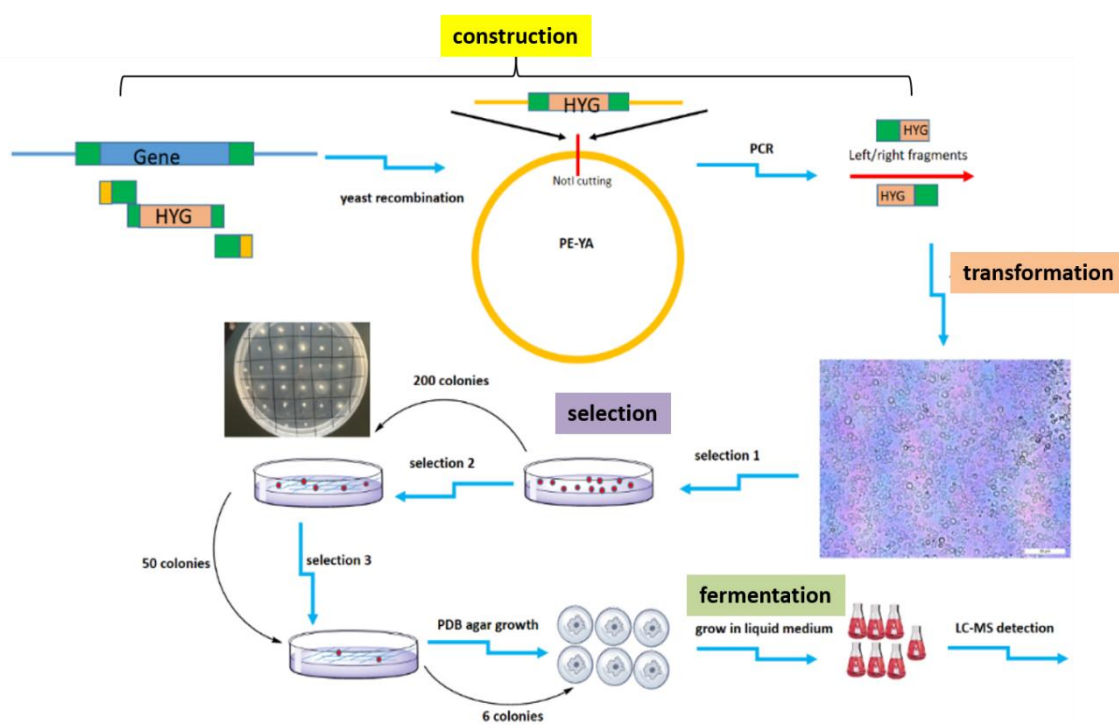


Fig. 2.2.5.3 Work flow of knockout experiment.

2.2.5.1c Knockout result

Six transformants and the wild type strain were grown in producing medium as section 2.1.1. The fermentations were extracted for LCMS analysis. According to the LCMS analysis, one of these six transformants didn't produce any viridiofungins (viridiofungin A **2** and viridiofungin B **84**) compared with wild type strain (Fig. 2.2.5.4 shade in red). So the citrate synthase gene of SY2040-3 might be disrupted, and the biosynthetic pathway was consequently blocked.

According to the total ion chromatogram (TIC), from wild type *T. viride*, peak **2** (t = 7.9 min) is consistent with viridiofungin A **2** and peak **84** (t = 9.0 min) is consistent with viridiofungin B **84**. However, the KO strain SY2040-3 lost both peaks **2** and peak **84** in the ESI trace compared

to the wild type strain and other transformants. Moreover, SY2040-3 also lost the peaks of the proposed viridiofungin congeners from 5.8 min to 7.1 min (**Fig. 2.2.5.4**).

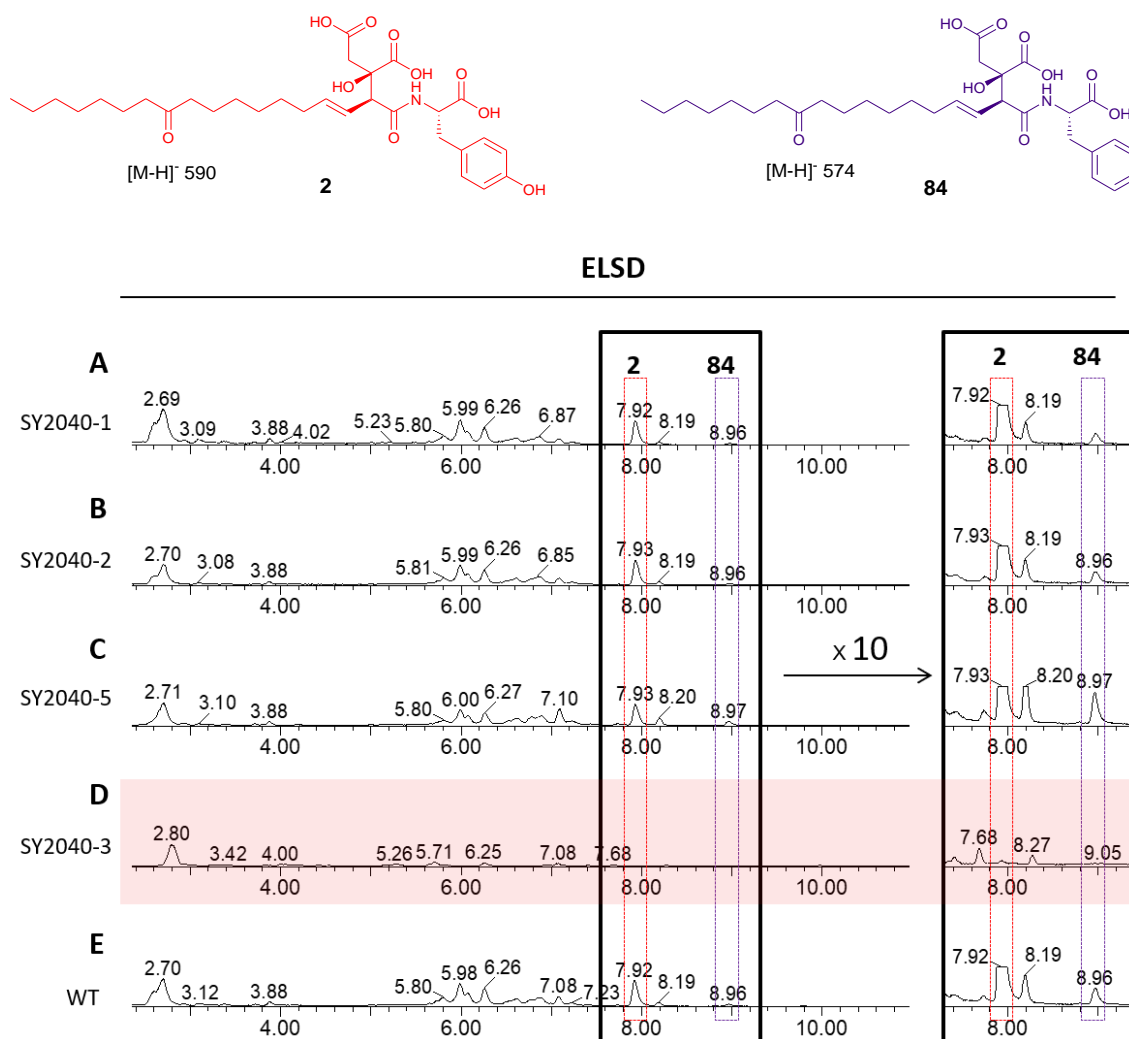
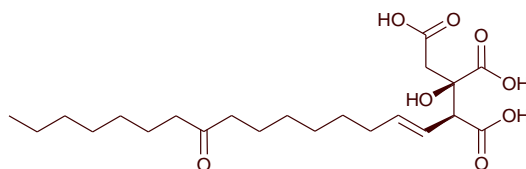


Fig. 2.2.5.4 ELSD analysis of citrate synthase knockout experiment: **A-D**, KO transformants; **E**, wild type.

The extracted ion chromatograms (EIC) of $[M - H]^- 590$ and $[M - H]^- 574$ from the SY2040-3 strain were then analysed. Only noise was observed in the chromatograms (**Fig. 2.2.5.5**). The possible citrate intermediate compound **159** discussed in the section **2.2.1** was also analysed ($[M - H]^- 413$) and shown to be absent (**Fig. 2.2.5.5**). It means that the target gene is probably replaced by a marker gene (resistance gene hygromycin B). Therefore viridiofungin and citrate intermediate are no longer produced.



[M-H]⁻ 413 159

However, the extracted ion chromatogram of the wild type strain showed that viridiofungin A **2**, viridiofungin B **84** and putative citrate intermediate **159** were present (Fig. 2.2.5.5).

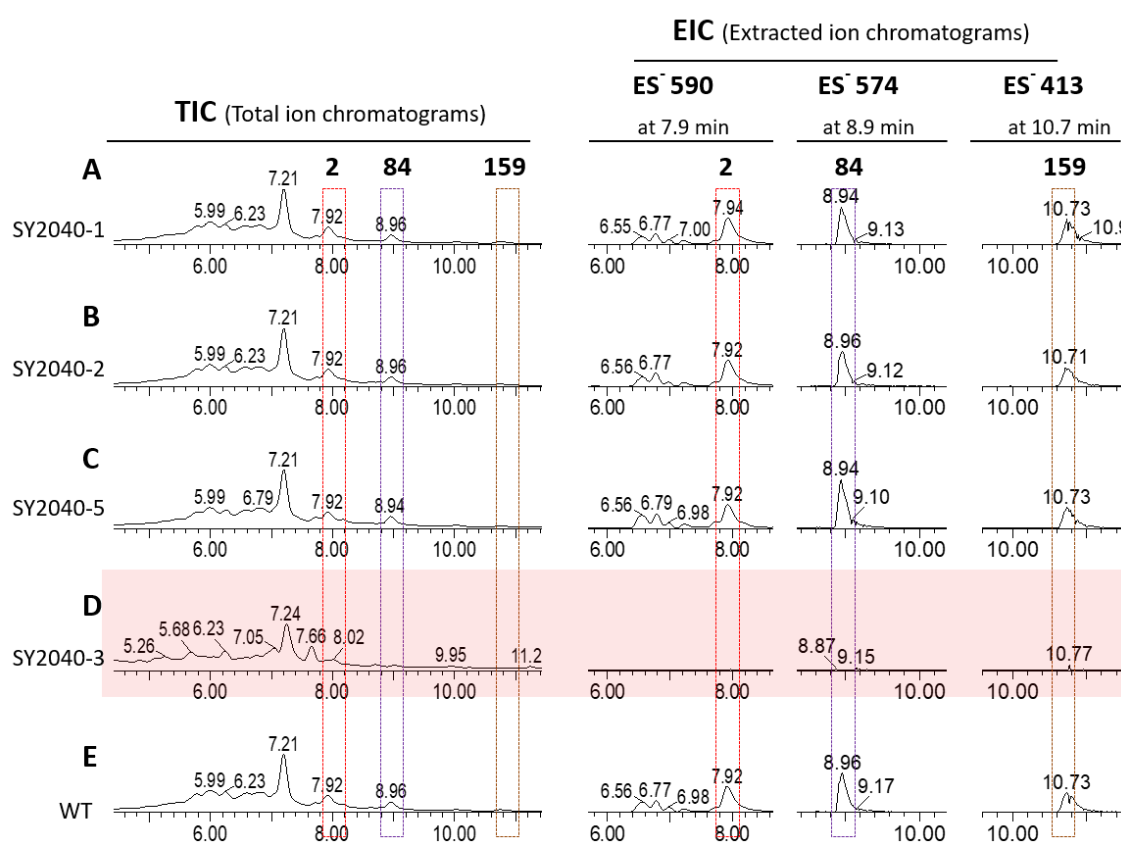


Fig. 2.2.5.5 LCMS analysis of citrate synthase knockout: A-D, KO transformants; E, wild type.

2.2.5.1d Genome confirmation

To confirm that the non-producing transformant SY2040-3 is real knockout strain, genomic DNA of SY2040-3 was isolated for PCR confirmation. Four primers were designed (primer 1 and 4 outside of the gene, primer 2 and 3 inside of the gene Fig. 2.2.5.8).

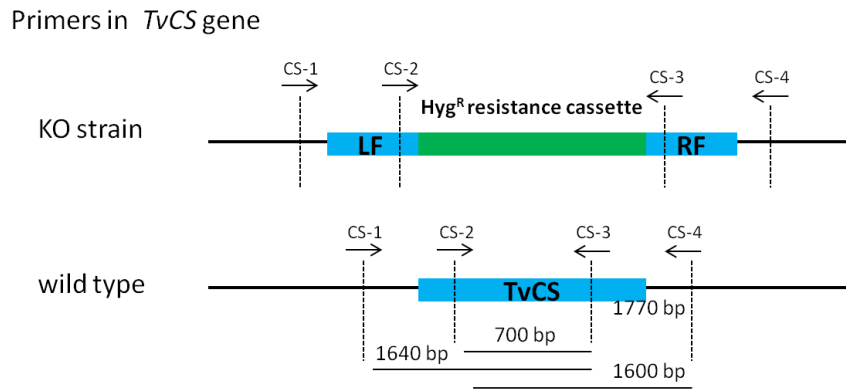


Fig 2.2.5.8 Primer designing for KO confirmation.

In principle, primer CS-2\3, primer CS-1\3 and primer CS-2\4 pairs can produce PCR product 700bp, 1640bp and 1600bp by using wild type gDNA as template. However, K.O strain cannot produce any PCR product. However, the PCR results were not as expected (Fig. 2.2.5.9).

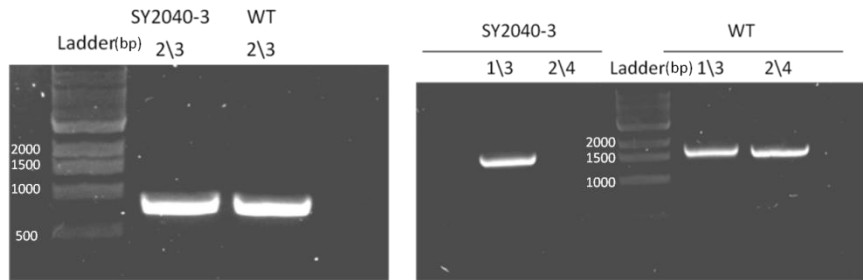


Fig. 2.2.5.9 PCR confirmation of K.O. KO is knock-out transformant, WT is wild type. 2\3, 1\3, 2\4 are primer pairs.

The results indicate that the citrate synthase like gene in the KO transformant is only partially (C terminal) disrupted by integration. This has been observed before in the Cox group. It probably arises because of ectopic integration of the P_{gpdA} -Hyg^R cassette elsewhere in the genome, plus an additional partial intergration into the target sequence. This is supported by the fact that the other transformants are clearly hygromycin resistant, but viridifungin is still produced indicating a high frequency of ectopic integration.

On the other hand, another set of primers was also designed to confirm the transformant (Fig. 2.2.5.10). The PCR result indicated that the wild type can still get the full gene PCR product by primer HCS-1\4, while the CS K.O transformant could not get any PCR product, which means the target gene was somehow broken. However, the left (HCS-1\2) and right (HCS-3\4) fragment confirmation was not successful. Because the wild type and KO strain showed the same PCR results (Fig. 2.2.5.11).

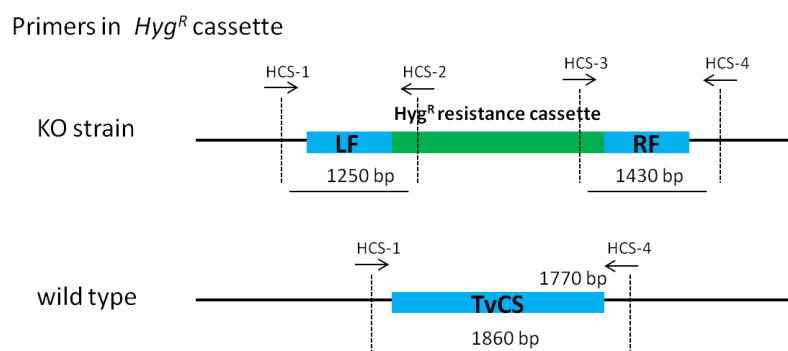


Fig 2.2.5.10 Primer designing for KO confirmation.

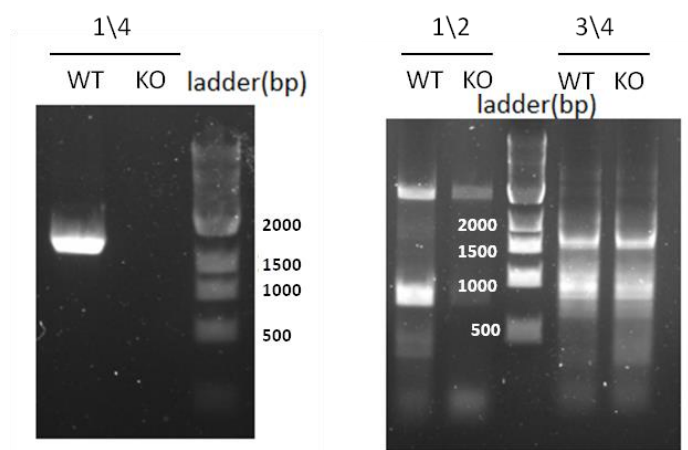


Fig. 2.2.5.11 PCR confirmation of CS K.O

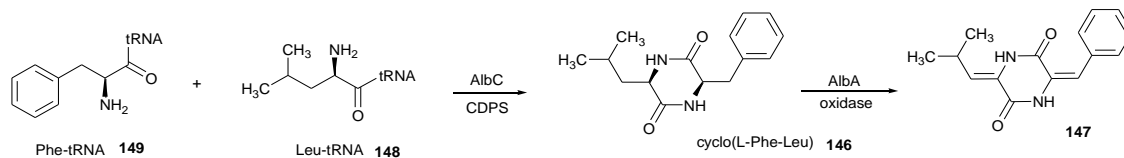
2.2.5.2 *tRNA ligase like gene knockout*

2.2.5.2a *Gene function analysis*

Examination of the structures of viridiofungins shows that tryrosine, phenylalanine and tryptophan are the most common amino acids incorporated. In some rare case, glycine is also found in the viridiofungin structure like viridiofungin Z **89**.^[4]

However, there is no PKS or NRPS module in the proposed viridiofungin BGC. But a tRNA ligase like enzyme (TvR4) is encoded in the gene cluster, close to the citrate synthase in the genome. This is the only enzyme which might be related to the amino acid attaching step (*eg.* **Scheme 2.1.1.6**). Other peptidic secondary metabolites are known to be made by non NRPS systems. For example, albonoursin **147** biosynthesis is mediated by a cyclodipeptide synthetase

(CDPS) using Phe-tRNA **149** and Leu-tRNA **148** as substrates (**Scheme 2.2.6.1**). Crystal structure of this kind of enzyme is quite similar to the tRNA ligases which are involved in primary metabolism.^[112]



Scheme 2.2.6.1 Biosynthesis of Albonoursin mediated by CDPS

Biosynthesis of albonoursin **147** suggests that the tRNA ligase like enzyme (TvR4) in viridifungin BGC might catalyse the reaction between a tRNA amino acid and the fatty acid chain. It is interesting that a nucleotide phosphatase is also encoded in the cluster, which might relate to release of the product from the tRNA. To put the information together, the tRNA ligase might act as the C domain of the NRPS, and tRNA is like the PCP to load the amino acid (**Scheme 2.2.3.1**), then a phosphatase can release the product from tRNA. To test the hypothesis, target knockout on the tRNA ligase (TvR4) was conducted following the method in **2.2.5.1**. If the tRNA ligase like enzyme is the key enzyme of the pathway, deletion of this enzyme will block the whole pathway.

2.2.5.2b Knockout result

After transformation and three rounds of selection, 12 *T. viride* transformants were grown in liquid medium. The fermentations were extracted for LCMS analysis after 7 days. According to the LCMS results, two of the transformants (SY2017-1 and 2, shaded in yellow) lost the ability to synthesise viridifungins. The other transformants still produced viridifungins (**Fig. 2.2.6.1**).

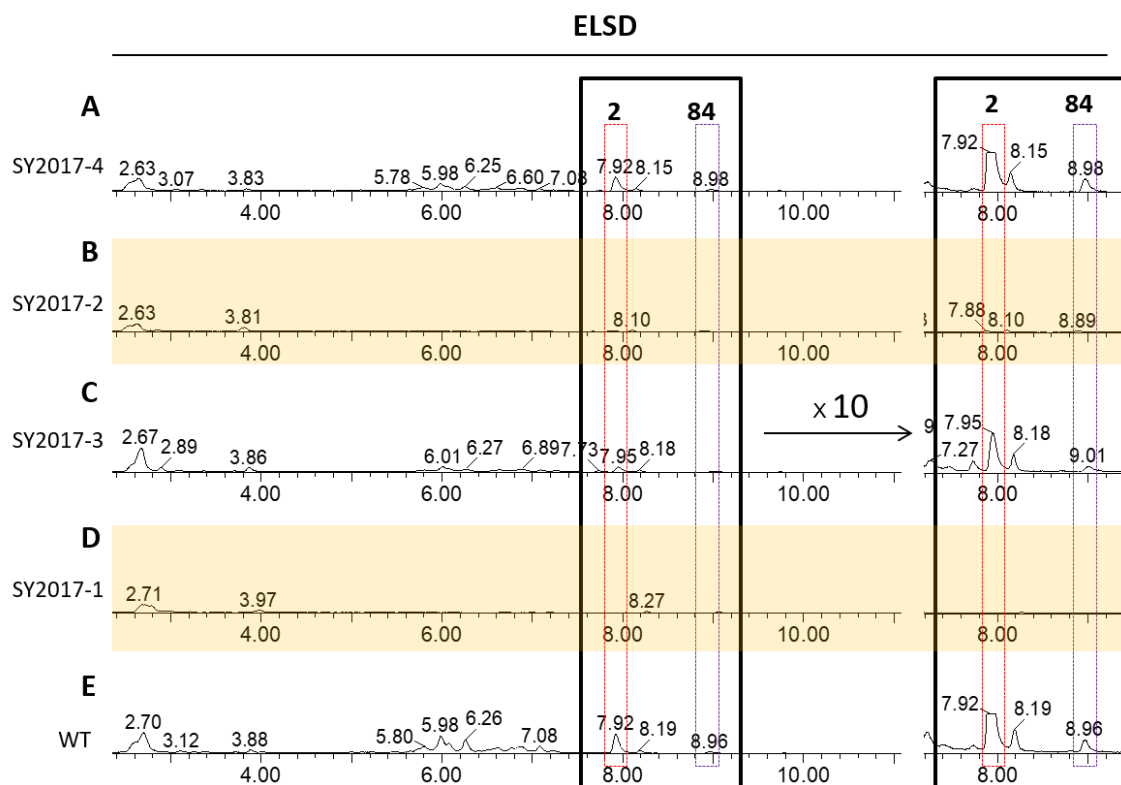


Fig. 2.2.6.1 LCMS analysis of tRNA ligase like enzyme knockout experiment: **A-D**, KO transformants; **E**, wild type.

According to the total ion chromatogram (TIC), the KO strains SY2017-1 and 2 lost the peak **2** and peak **84** in the ESI trace compared to wild type strain and other transformants. Moreover, SY2017-1 and 2 also lost the peaks of proposed viridifungin relatives from 5.8 min to 7.1 min (**Fig. 2.2.5.4**).

The extracted ion chromatograms (EIC) of $[M - H]^-$ 590 and $[M - H]^-$ 574 only show noise (**Fig. 2.2.5.4**). The putative citrate intermediate compound **159** discussed in the section **2.2.1** was also analysed. The EIC for the $[M - H]^-$ 413 ion was examined. There were peaks shown in the chromatograms of the SY2017-1 and 2 transformants. The EIC of wild type strain and other transformants showed viridifungin A **2**, viridifungin B **84** and citrate intermediate **159** peaks (**Fig. 2.2.6.2**). This shows that the **159** compound must be formed before the action of the tRNA ligase. It also shows that the 5,6 alkene is likely to be present from early in the pathway.

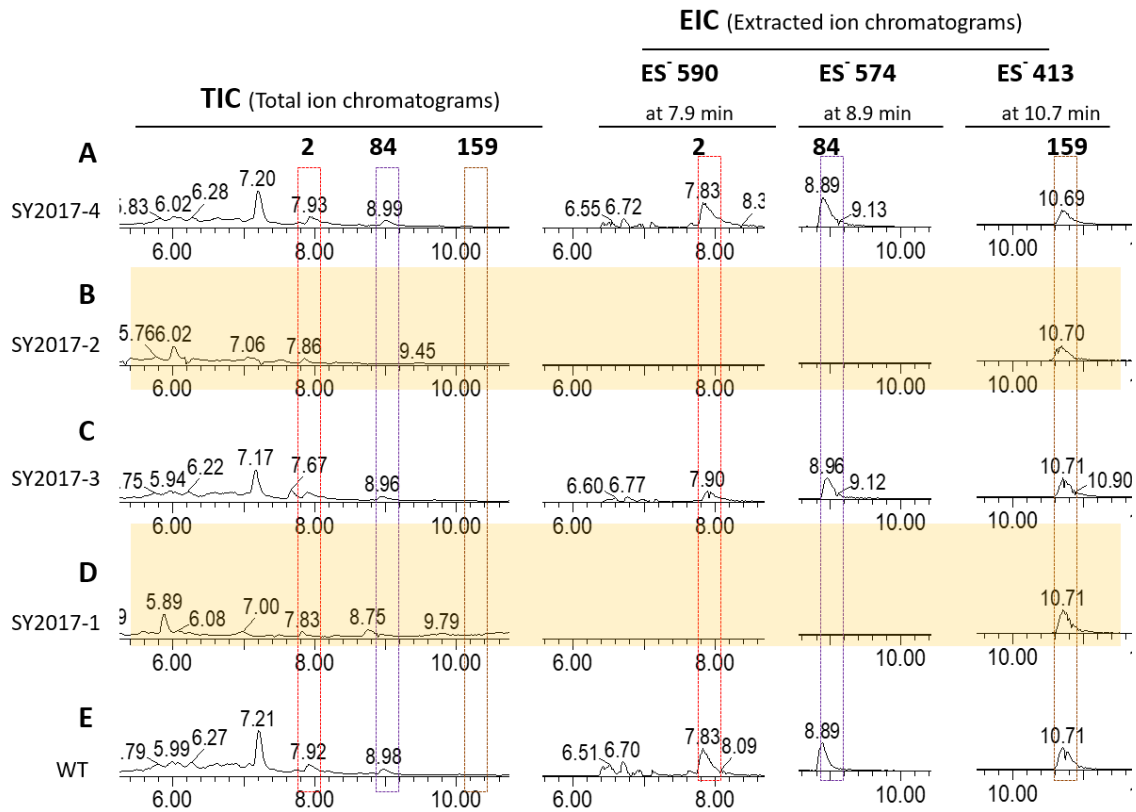


Fig 2.2.6.2 LCMS analysis of *tvR4* knockout experiment: A-D, KO transformants; E, wild type.

2.2.5.2 c Genome confirmation

To confirm that the non-producing transformant was a real knockout strain, the genomic DNA of transformant SY2017-1 and 2 were isolated, primers were designed for PCR (Fig. 2.2.6.3). The PCR result (HR4 - 1\4 pairs) showed that the tRNA ligase gene in KO transformants are broken. For left fragment confirmation (HR4 - 1\2 pairs), the PCR product of the transformants were complex but different from the wild type. The right fragment confirmation (HR4 - 3\4 pairs) showed that transformant SY2017-1 has a weak PCR product bind about 1500 bp (Fig. 2.2.6.4). The situation is similar to the CS KO experiments. The PCR results also indicate that SY2017-1 and 2 transformants also contain ectopic integration of the P_{gpdA} -Hyg^R cassette elsewhere in the genome, plus an additional partial intergration into the target sequence.

II Gene knockout

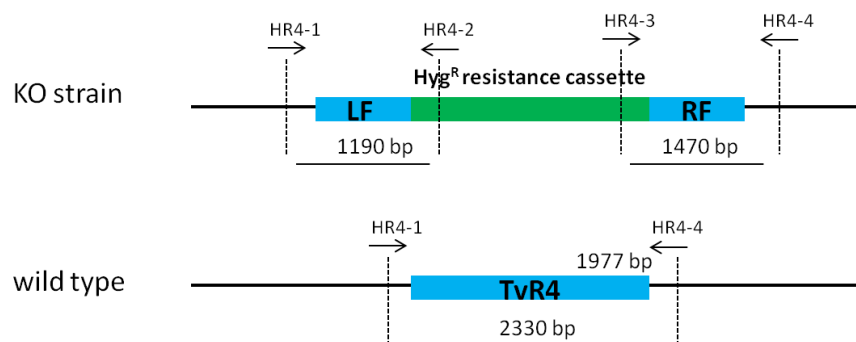


Fig. 2.2.6.3 Primer designing for KO confirmation.

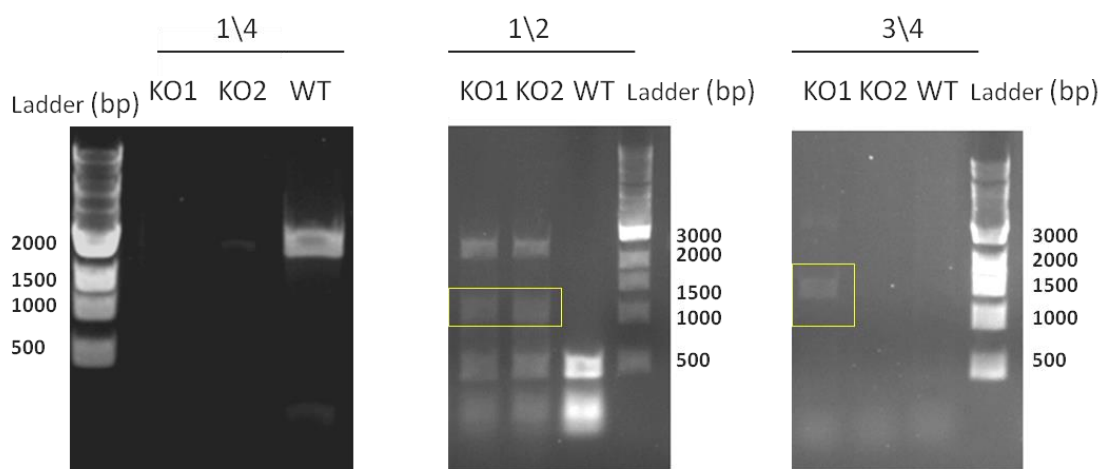


Fig. 2.2.6.4 PCR confirmation of tRNA ligase like enzyme K.O

2.2.5.3 KO of the tailoring genes in viridifungin cluster

According to the knockout experiment of citrate synthase TvCS and tRNA ligase like enzyme TvR4, it is confirmed that these two enzymes are involved in the biosynthetic pathway of viridifungins. This is consistent with a pathway in which the fatty acid chain (blue) and the oxaloacetate (red) are combined by citrate synthase, and the amino acid part is then attached to the fatty acid chain by using tRNA ligase, so the basic skeleton of viridifungin is synthesized by these two enzymes (Fig. 2.2.5.1).

In addition, two positions need tailoring to form the final structure. These are the (5, 6) alkene (purple) and the C-13 ketone (yellow). Correspondently, there are several tailoring genes in the BGC: acyl-CoA oxidase (*tvL3*), P450 oxygenase (*tvL4*), FAD dependent monooxygenase (*tvR2*) and an NAD dependent dehydrogenase (*tvL2*). Acyl-CoA oxidase could be related to the formation of the alkene, while the P450 oxygenase, FAD dependent monooxygenase and an NAD dependent dehydrogenase could be related to the formation of the ketone.

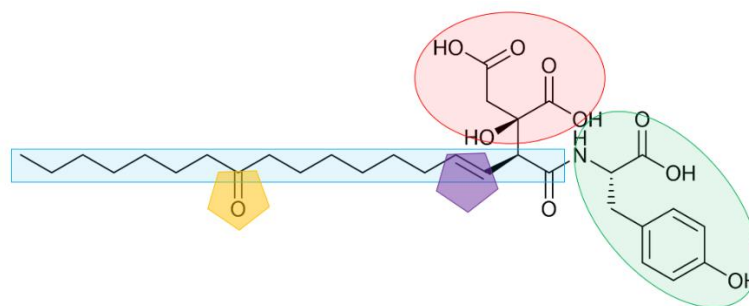


Fig. 2.2.5.1 Structure of viridiofungin A

To explore the function of the tailoring enzymes in the viridiofungin gene cluster, knockout experiments were also conducted on these genes.

By using gDNA of *T. viride* as a template, the left and right fragments (from *tvL4*, *tvR2*, *tvL3* and *tvL2*) for PE-YA vector construction got from PCR (section 2.2.5.1b), and after yeast recombination the fragments together with hygromycin resistance cassette fragment (hyg) built up the pE-YA plasmid for knockout. The PCR product of pE-YA vector was then transformed into the *T. viride* protoplast. After three rounds of selection several transformants were obtained (Table 2.2.5.1). These transformants were then grown in liquid medium. The fermentations were extracted for LCMS analysis after 7 days.

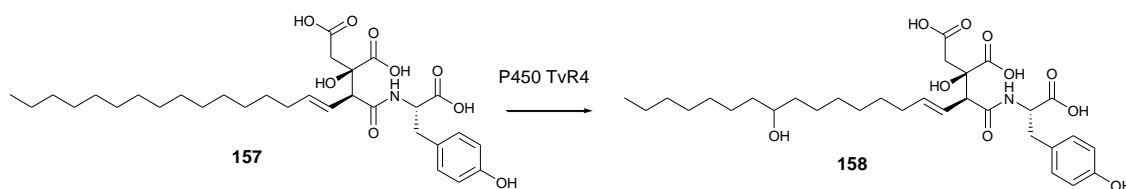
Genes	<i>tvL4</i> (P450 oxygenase)	<i>tvR2</i> (FAD dependent monooxygenase)	<i>tvL3</i> (acyl-CoA oxidase)	<i>tvL2</i> (NAD dependent dehydrogenase)
numbers	18	26	16	30

Table 2.2.5.1 Number of the transformants

2.2.5.3 a Attempted knockout of P450 enzyme genes

TvL4 was predicted as a P450 monooxygenase, it indicated that TvL4 might be involved in the oxidation of C-13 carbon (Scheme 2.2.4.3). If the tvL4 was successfully disrupted, the strain could produce intermediate 157.

II Gene knockout



Scheme 2.3.4.3 Proposed reaction of TVR4

However, the LCMS analysis result shows that all of the transformants can still produce viridiofungin A **2**, B **84** and related compounds. EIC of $[M - H]^-$ 590, $[M - H]^-$ 574 and $[M - H]^-$ 413 also show that all of the transformants can produce viridiofungins **2** and citrate intermediate **159**. The proposed intermediate **157** has m/z 577, but none of the transformant is shown to produce **157** in the ion EIC of $[M - H]^-$ 576 (**Fig. 2.2.5.7**).

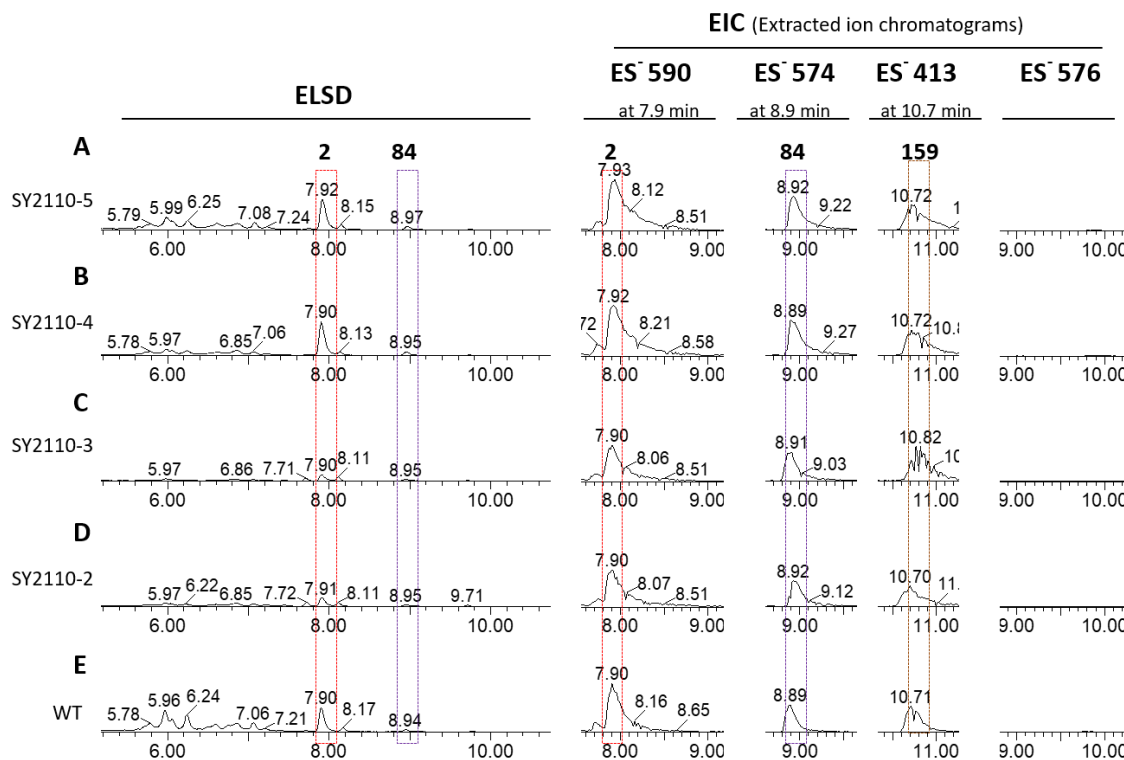


Fig. 2.2.5.7 LCMS analysis of P450 K.O transformants

Several transformants' gDNA were isolated for genome confirming. Primers (HL4 - 1\4) outside of the target gene were used for PCR. A 1430 bp PCR product was obtained from all of the transformants and the wild type strain. This result indicated that the P450 gene *tvL4* was not disrupted in the genomes of the transformants. This is consistent with the LCMS result.

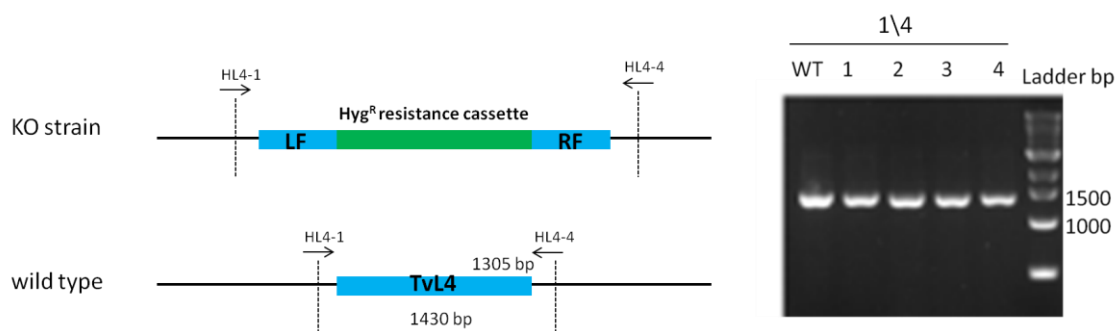


Fig. 2.2.5.7 PCR confirmation of *tvL4* KO.

2.2.5.3 b Attempted knockout of the other genes

Acyl-CoA oxidase (*tvL3*), FAD dependent monooxygenase (*tvR2*), NAD dependent and dehydrogenase (*tvL2*) knockout experiment showed the similar result to the P450 gene K.O. All of the transformants can still produce viridiofungins and related compounds, but none of them can produce any new related intermediate. The genome confirming also show that the target genes are not disrupted.

In conclusion, P450 monooxygenase (*tvR4*), acyl-CoA oxidase (*tvL3*), FAD dependent monooxygenase (*tvR2*) and NAD dependent dehydrogenase (*tvL2*) were not successfully knocked out. It is difficult to knockout the target gene in *T. viride* based on the homologous recombination mechanism.

2.2.6 Attempted Gene silencing of TvR3

Because bipartite knockout is not efficient in *T. viride*, another strategy was tried in this study. Post-transcriptional gene silencing is a method of gene interference which relies on degradation of targeted mRNA, intercepting transfer of genetic information between the transcriptional and translational machinery.^[90] In this part of the work, an RNA interference (RNAi) system (section 1.7.4) will be used to knockdown the genes in viridiofungin cluster.

TvR3 is predicted as a nucleotide phosphatase. The proposed activity of this enzyme is to catalyze the release of viridiofungin **2** from tRNA-viridiofungin complex **163** (Scheme 2.2.6.1). If TvR3 is successfully silenced, the biosynthetic pathway of viridiofungin would be blocked, and the product should be heavily decreased.

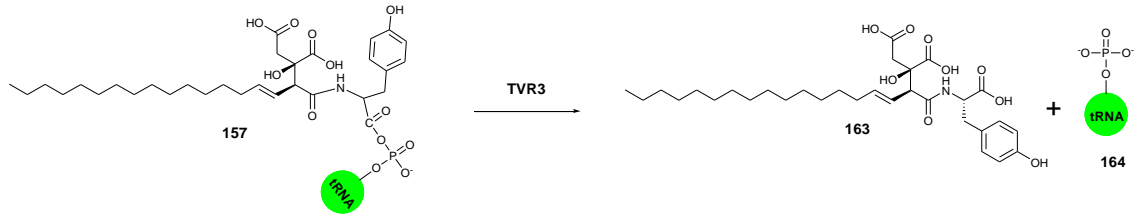


Fig. 2.2.6.1 The proposed reaction of TvR3

2.2.6.1 Vector construction

Based on the mechanism of RNAi (section 1.7.4), the P_{pki} promoter from the *Trichoderma reesei* (section 2.2.4.4) was used to promote the transcription of the reverse gene (revGOI). The revGOI is cloned from target gene (GOI) by using a pair of reverse primers. If the revGOI can be successfully transcribed in *T. viride*, the dsRNA would anneal with mRNA to form a siRNA. The siRNA can degrade the targeted mRNA (Fig. 2.2.6.1 A).^[91]

The revGOI of *tvR3* was cloned from the cDNA of *T. viride*. The P_{pki} promoter was cloned from the *Trichoderma reesei* genome. These two fragments were combined together and inserted into the vector pTYH-GS-pki which contain a HYG marker. The vector pTYH-GS-pki-reGOI was constructed by yeast recombination (Fig. 2.2.6.1 B).

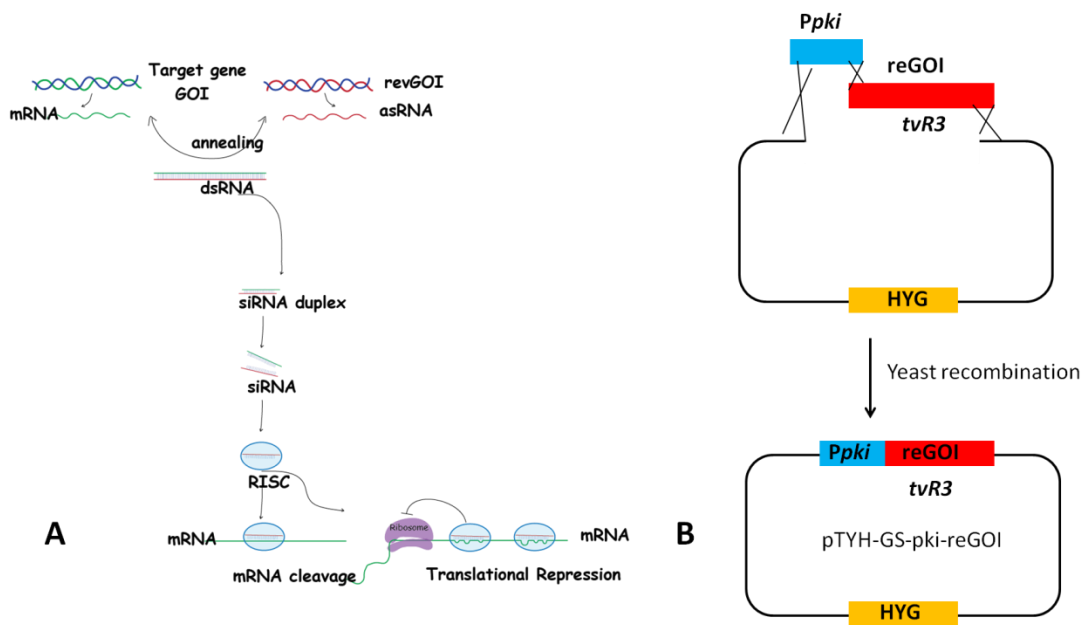


Fig. 2.2.6.2 Mechanism of RNAi and vector construction: A, mechanism; B, RNAi vector construction.

2.2.6.2 Transformation, LCMS analysis and PCR confirmation

The vector pTYH-GS-pki-reGOI was then transformed into *T. viride* by using PEG-mediated transformation. After three rounds of selection, 10 transformants were obtained. The transformants were grown in liquid medium for 7 days, and the compounds were extracted for LCMS analysis.

According to the ELSD chromatograms, the transformants do not show significant difference with the wild type strain. The extracted ion chromatograms trace of [M - H]⁻ 590, [M - H]⁻ 574 and [M - H]⁻ 413 also show no significant difference between wild type and transformants (**Fig. 2.2.6.3**).

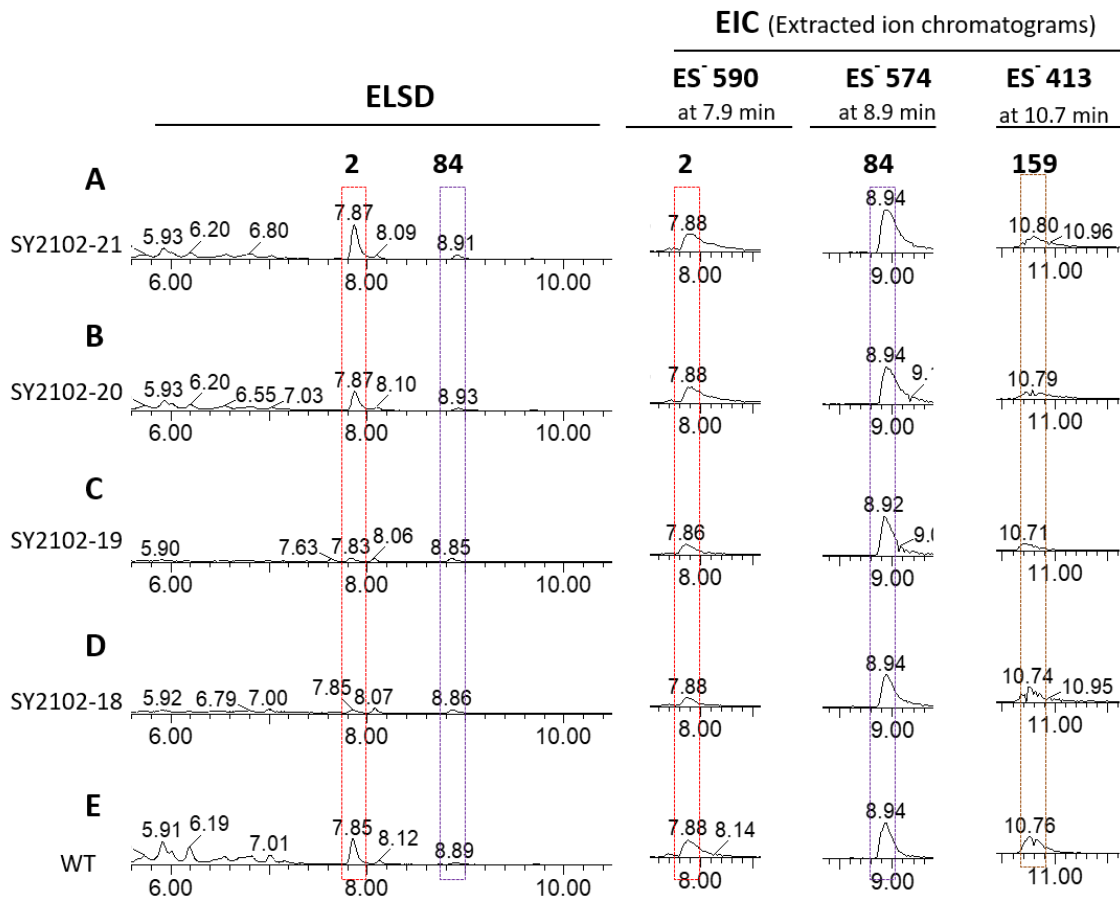


Fig. 2.2.6.3 LCMS analysis of TvR3 RNAi transformants: **A-D**, KO transformants; **E**, wild type.

Genomic DNA of the transformants was also extracted for PCR confirmation. A pair of primers (1\2) was designed for the fragment of *P_{pki}* promoter part. The transformants DNA template could produce the PCR products (650 bp), while the wild type not (**Fig. 2.2.6.4**). It means the RNAi vector has been successfully transformed into the fungi. The reason why RNAi transformants didn't make any difference of the production might because of the weakness of the *pki* promoter.

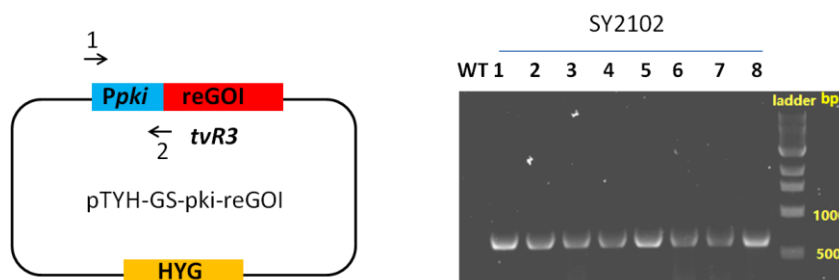


Fig. 2.2.6.4 PCR confirmation of RNAi transformants

2.2.7 Heterologous expression in *A. oryzae*

To explore the biosynthesis pathway of the natural products, heterologous expression is also a very useful tool. Heterologous expression hosts for secondary metabolites or specific enzymes in the biosynthesis pathway are usually known to be *E. coli*, yeast and some fungi like *A. oryzae* and *T. reesei*.^[100] Because of the difficulties of knockout and silencing in *T. viride*, the viridifungin BGC will express in *A. oryzae* in this section. The original vectors and strains used in this part of work have been described in section 1.7.8.

2.2.7.1 Plasmid construction

All PE-YA entry and pTYGS (*arg/ ade*) vectors for expression of genes in the pathway were constructed by *in vivo* homologous recombination in *S. cerevisiae* and *in vitro* LR recombination. Genes were amplified by PCR using *T. viride* cDNA as a template with oligonucleotides which were designed to introduce the 30 bp homologous overlaps. Assembly of genes in pTYGS expression vectors was achieved by designing oligonucleotides with an overlap to T_{adh} / P_{gpdA} or T_{gpdA} / P_{eno} and P_{eno} , which resulted in the removal of the 2480 bp T_{adh} / P_{gpdA} or the respective 800 bp T_{gpdA} / P_{eno} DNA fragment

According to the proposed biosynthesis of viridifungin, the citrate synthase (TvCS) and tRNA ligase (*tvR4*) are the core genes in the pathway, so these two genes were cloned into the pTYGSarg vector, then nucleotide phosphatase (*tvR3*) and acyl-CoA oxidase (*tvL3*) were also cloned into this vector to confirm the function of these two genes. The three tailoring genes (P450 monooxygenase *tvL4*, monooxygenase *tvR2* and NAD dependent dehydrogenase *tvL2*) were cloned into the pTYGSade vector. Finally, ten vectors were constructed with different combinations (Fig. 2.2.7.1).

II Heterologous expression

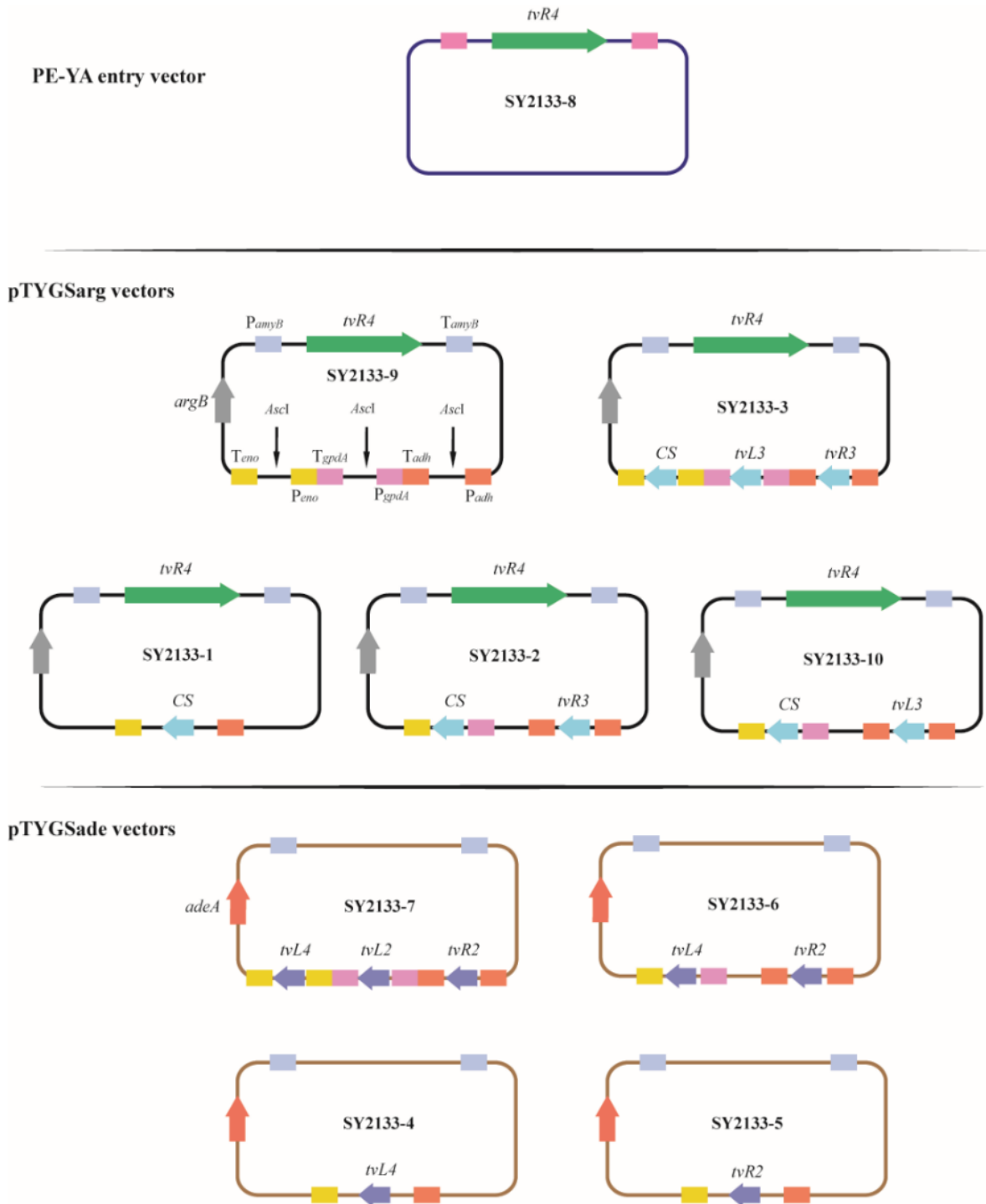


Fig. 2.2.7.1 A. *oryzae* expression vector construction.

2.2.7.2 Transformation and LCMS analysis

To test if the viridiofungin biosynthetic pathway can be rebuilt in *A. oryzae*, the plasmid SY2133-3 and SY2133-7 which contains all of the possible genes (*i.e* the complete pathway) from the viridiofungin BGC were used for fungal transformation (**Fig. 2.2.7.2**).

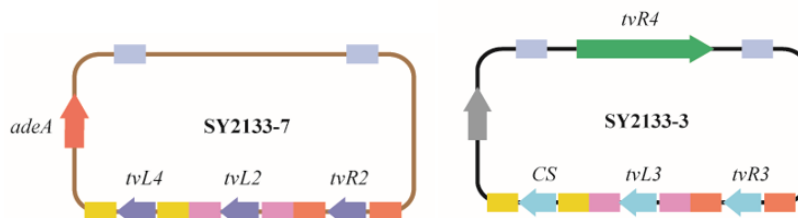


Fig. 2.2.7.2 The plasmid used for *A. oryzae* transformation

CaCl₂/PEG mediated protoplast protocol was used to transform these two plasmids to *A. oryzae* NSAR1. Arginine and adenine auxotroph minimal medium was used for selection. After three rounds of selection, 20 transformants survived (**Fig. 2.2.7.3**). The colorless transformants are more likely to be the real transformants. Because the adenine biosynthesis pathway blocked strain (like *A.oryzae* NSAR1) can produce pink pigments which are intermediates of adenine. This phenomenon usually happens when the environment is short of adenine. So pTYGS-Ade plasmid transformation can provide adenine auxotroph strain with enough adenine, the transformants should not produce the pink pigments. Five colorless transformants, wild type, and four dark color transformants were selected for compound extraction and LCMS analysis

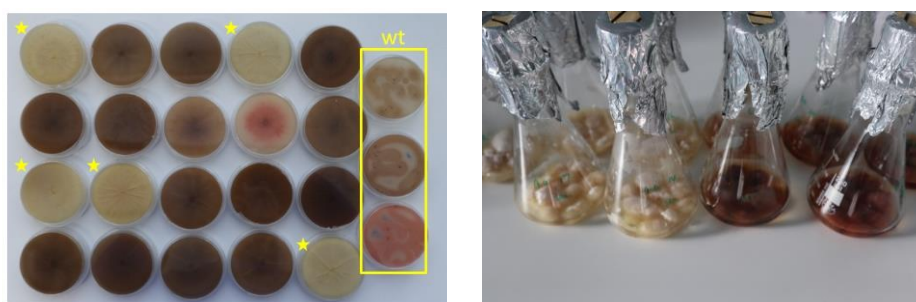


Fig. 2.2.7.3 Results of transformation

The LCMS result indicated that there was no transformant had obvious changes compared to *A.oryzae* NSAR1 wild type, but the *T. viride* wild type strain produced viridiofungins (**Fig. 2.2.7.4**).

II Heterologous expression

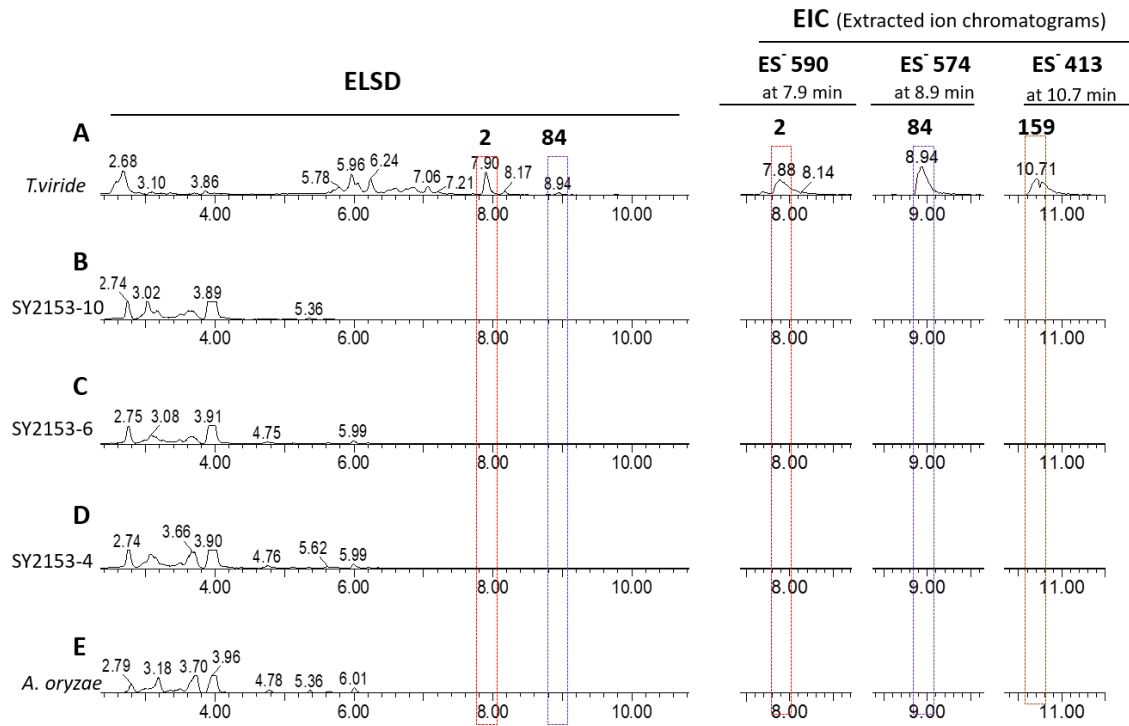


Fig. 2.2.7.4 LCMS analysis of *A. oryzae* transformants and *T. viride* wild type strain: **A**, *T. viride* wild type strain; **B-D**, *A. oryzae* transformants; **E**, *A. oryzae* NSAR1.

The genomic DNA of *A. oryzae* transformants was extracted for PCR confirmation. Seven pairs of primers were designed to amplify the seven target genes. Transformant SY2153-6 was the real transformant which contain all of the genes inside. Transformant SY2153-4 also contains all of the genes. However, there were two bands in the *tvL4* lane, it might be because of the contamination of gDNA samples (**Fig. 2.2.7.5**).

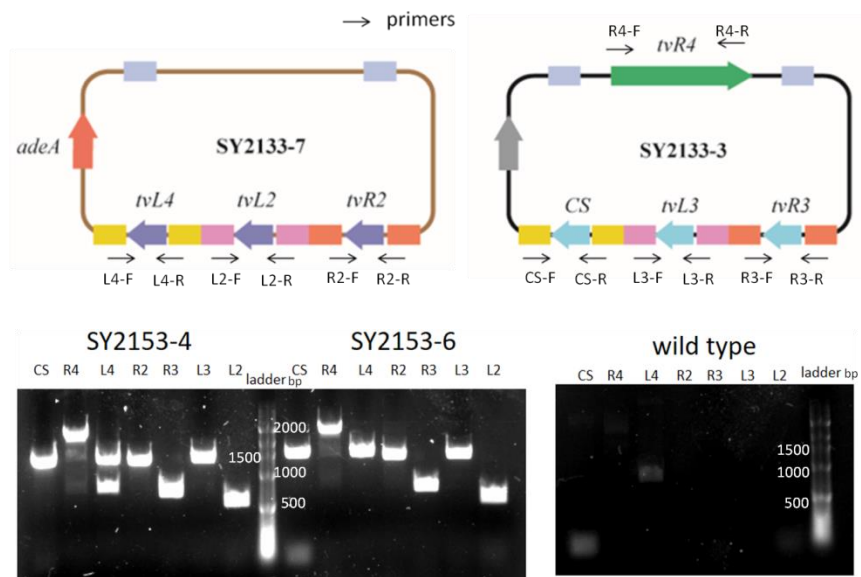


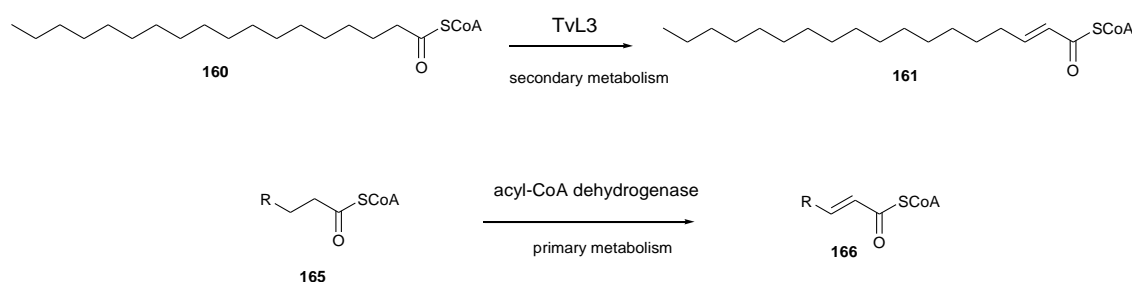
Fig. 2.2.7.5 PCR confirmation of transformants (the arrows means primers, seven pairs of primers were used).

Even the real transformants cannot produce viridiofungin. It might be because the fatty acid substrate (fatty acid CoA) in *A. oryzae* is not enough for viridiofungin synthesis. Furthermore, the tRNA ligase like enzyme form *T. viride* might not be able to recognise the amino acid tRNAs in *A. oryzae*.

2.3 Discussion

2.3.1 Oxidation of fatty acid

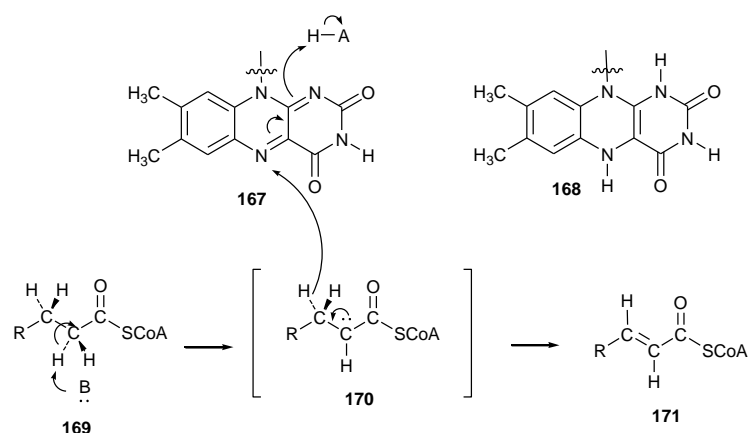
Viridiofungin BGC encodes an acyl-CoA oxidase (TvL3). This acyl-CoA oxidase is thought to introduce the alkene in C₁₈ fatty acid chain of viridifungins (**160** to **161**). Based on the *in vitro* assay of the CS from byssochlamic acid pathway (**Chapter 3**, section **3.2.2**), BfL2 can take 2, 3-alkene as a substrate and isomerise it to a 3, 4 -alkene. This is also observed in the structure of the viridiofungins. Therefore, oxidation of fatty acid is thought to be the first step of the pathway. Also the proposed early intermediate **159** most likely already possesses the 3,4-alkene.



Scheme 2.3.1.1 Proposed reaction of TvL3, and function of acyl-CoA dehydrogenase.

Acyl-CoA dehydrogenase from the primary metabolism involved in fatty acid β -oxidation.^[133-135] The function of this dehydrogenase is to introduce the double bond which is the first step of fatty acid β -oxidation.^[136] Acyl-CoA dehydrogenase is an FAD dependent enzyme which removes two hydrogen atoms from carbons 2 and 3 in the fatty acyl CoA **165** (**Scheme 2.3.1.1**). The X-ray crystal structure of the active site in an enzyme-substrate complex of acyl-CoA dehydrogenase shows the flavin ring and the disposition of the two hydrogens that are removed. Glu-376 removes the *pro-R* hydrogen of the fatty acyl-CoA, The *pro-R* β hydrogen is simultaneously donated to the *Re* face of the N-5 position of FAD **165**. Finally, the α , β -unsaturated acyl-CoA **167** results as a trans double bond **169** (**Scheme 2.3.1.2**).^[137] Because of the similar activity of acyl-CoA dehydrogenase in the primary metabolism, the KO experiment on the acyl-CoA oxidase gene of viridiofungin might not totally block the production of the compounds. Because there is no fatty acid CoA ligase involved in the viridofungin cluster, the C₁₈ fatty acid CoA should directly come

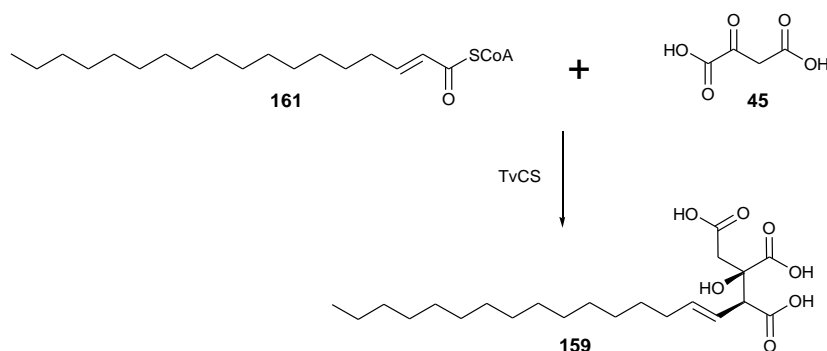
from the other primary pathway in the biosynthesis. It is known that fungal FAS produces CoAs directly.



Scheme 2.3.1.2 Reaction of acyl-CoA dehydrogenase mediate FAD

2.3.2 Citrate synthase in primary and secondary metabolism

The knockout experiment of TvCS indicates that this secondary metabolism CS is involved in the biosynthesis of viridifungin. The TvCS KO strain cannot produce viridifungins or the putative citrate intermediate **159** (section 2.2.5.1c). Therefore, TvCS is thought to take C₁₈ fatty acyl CoA **161** and OAA **45** as substrates to form a citrate precursor **159** (Scheme 2.3.2.1). The alkene of C₁₈ fatty acid CoA substrate **161** is produced by an acyl-CoA oxidase (section 2.3.1). The stereochemistry of these CS is explored in section 3.3.2 and is consistent with the C-3 stereochemistry of the viridifungins.



Scheme 2.3.2.1 Proposed reaction of TvCS.

In primary metabolism CS catalyzes the condensation of acetyl coenzyme A (AcCoA) and oxaloacetate (OAA) to form citrate and CoA, which is the first step in the citric acid cycle (section

1.5.1). While in secondary metabolism, CS catalyzes the condensation of OAA and fatty acid (or polyketide) CoA (or ACP bond) to form citrate and CoA (or ACP). CS involved in secondary metabolism is the key enzyme for the basic structure of alkyl citrates. More detail discussion of CS can see **Chapter 3** (section **3.3.2**) including substrate selection and stereochemistry of the secondary metabolic CS.

2.3.3 tRNA ligase like enzyme in viridiofungin pathway

During the biosynthesis of viridiofungins, the key step is attaching the amino acid to the alkyl citrate. The first propose of this step is that a reaction catalysed by a PKS and NRPS hybrids system. There are two kinds of assembly line of PKS and NRPS hybrids. When an NRPS module is upstream of a PKS module the growing chain will form a new C-C bond. Correspondingly, when the PKS module is upstream of an NRPS module, the condensation domain recognizes the polyketide thiolester as an electrophilic partner for amide bond formation (**Fig. 2.3.3.1**).^[109, 138, 139] Therefore, if the NRPS module is involved in the biosynthesis of the viridiofungin, the NRPS module should be situated at the downstream of the PKS or FAS.

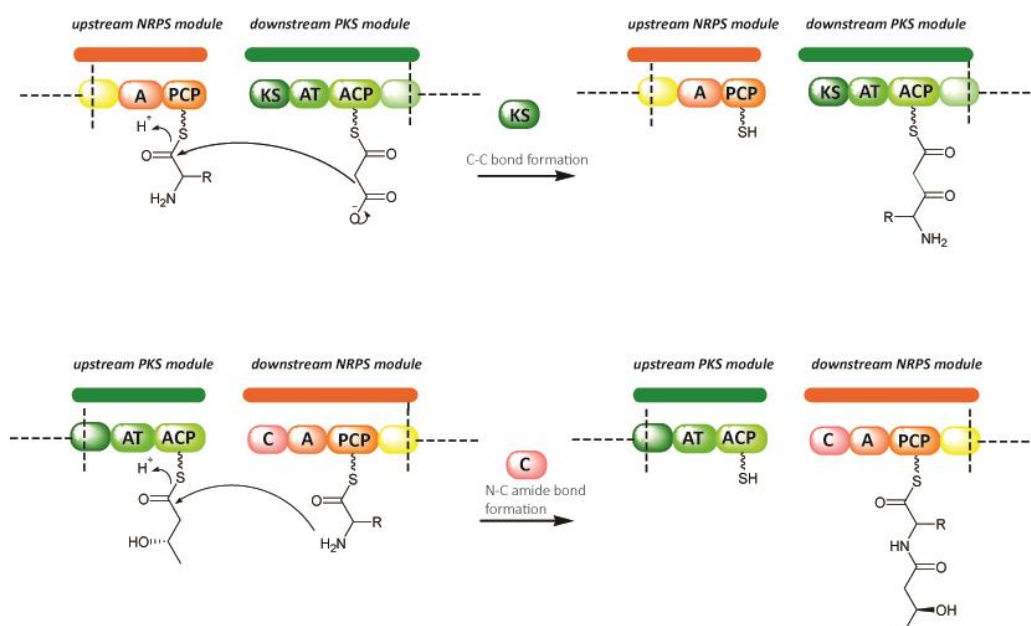
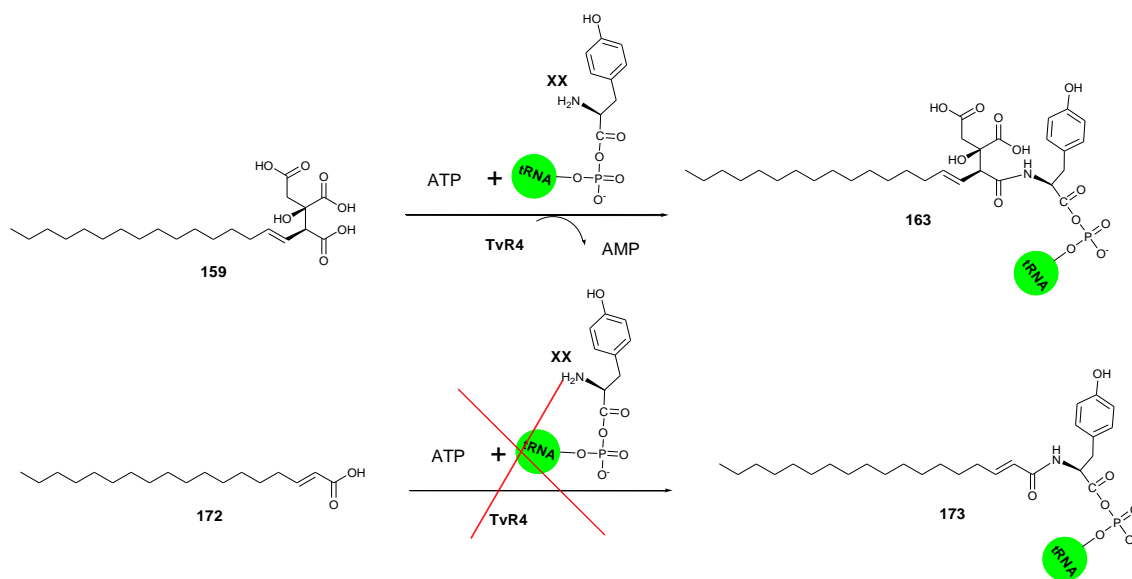


Fig. 2.3.3.1 Recognition issues at PKS and NRPS hybrid module interfaces.

However, the genome annotation (section **2.2.3.3**) and *TvCS* knockout experiment (section **2.2.5.1**) indicated that the viridiofungin BGC does not encode PKS or NRPS. But there is a tRNA^A ligase like enzyme *TvR4* involved in the BGC. According to the knockout results, disruption of *tvR4* leads to abolition of viridiofungins. It means *tvR4* is involved in the biosynthesis of viridionfungins. However, the KO strains can produce putative citrate intermediate **159** (section

2.2.5.2). It means TvR4 should use citrate **159** rather than a fatty acid **161** as a substrate, and the catalytic order of TvR4 is later than TvCS in the biosynthetic pathway (**Scheme 2.3.3.1**).



Scheme 2.3.3.1 Proposed function of TvR4.

Based on the gene annotation, TvR4 might process an aminoacyl-tRNA dependent catalysis. However, the mechanism of this enzyme need more studies..An example of AAtRS is cyclodipeptide synthases (CDPSs).^[111] By using a sensitive sequence profile method, Aravind *et al* found that the CDPSs were members of Rossmannoid domains and are likely to be highly derived versions of the class I AAtRS catalytic domains.^[140] In 2009, Muriel *et al* found that in the cyclodipeptide biosynthetic pathway there was a small protein unrelated to NRPS named AlbC which could use aminoacyl-tRNA as substrates to catalyse the formation of the cyclodipeptides peptide bonds. Moreover, some other bacterial proteins which are similar to AlbC also use aminoacyl-tRNAs to synthesise various cyclodipeptides (**Fig. 2.3.3.2**).^[141, 142]

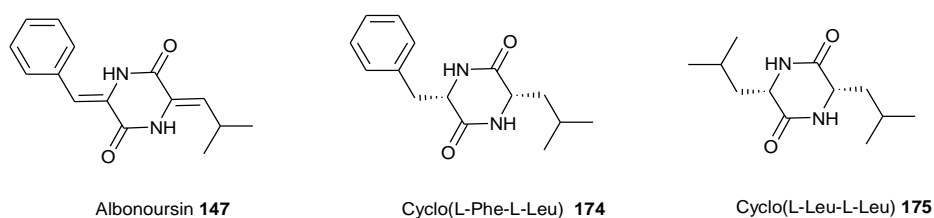
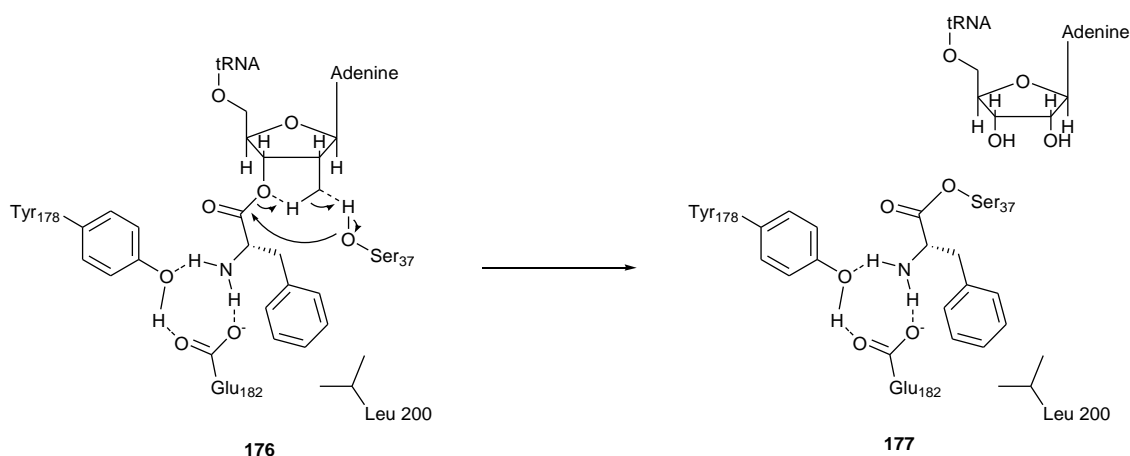


Fig. 2.3.3.2 Structures of cyclodipeptides.

In 2011, Ludovic *et al* described the crystal structure of AlbC, it was surprising that this enzyme is highly similar to the catalytic domain of class-I aminoacyl-tRNA synthetase (or tRNA ligase). AlbC contains a deep pocket which is highly conserved among CDPS. In this study, AlbC was

thought to catalyse its two-substrate reaction *via* a ping-pong mechanism with a covalent intermediate in which L-Phe **176** is shown to be transferred from Phe-tRNA to an active serine **177** (Scheme 2.3.3.3).^[141, 143] This study shows that the molecular basis of the interactions between CDPS and their substrates. All the studies performed on CDPS suggest that these enzymes use the two aa-tRNA substrates in a sequential ping-pong mechanism, with a similar first catalytic step: the binding of the first aa-tRNA and the transfer and storage of its aminoacyl moiety on the conserved serine residue of the enzyme pocket (*e.g.* Ser37, AlbC numbering).



Scheme 2.3.3.3 Proposed mechanism of covalent phenylalanyl-enzyme formation for AlbC.

In general, CDPS can use two aa-tRNA substrates in a sequential ping-pong mechanism, which is a hijack aminoacyl-tRNA pathway from secondary metabolism. First of all, the enzyme can catalyse the transformation of first aa-tRNA to the aminoacyl moiety on the conserved serine residue of the enzyme pocket which forms an amino acyl enzyme. Then the second aa-tRNA to form either a dipeptidyl-enzyme or a dipeptidyl-tRNA intermediate undergo intramolecular cyclization to generate the final cyclodipeptide product. The released tRNA can either go into the next round of condensation of CDPS or go into the normal protein production (**Fig. 2.3.3.4**).^[144]

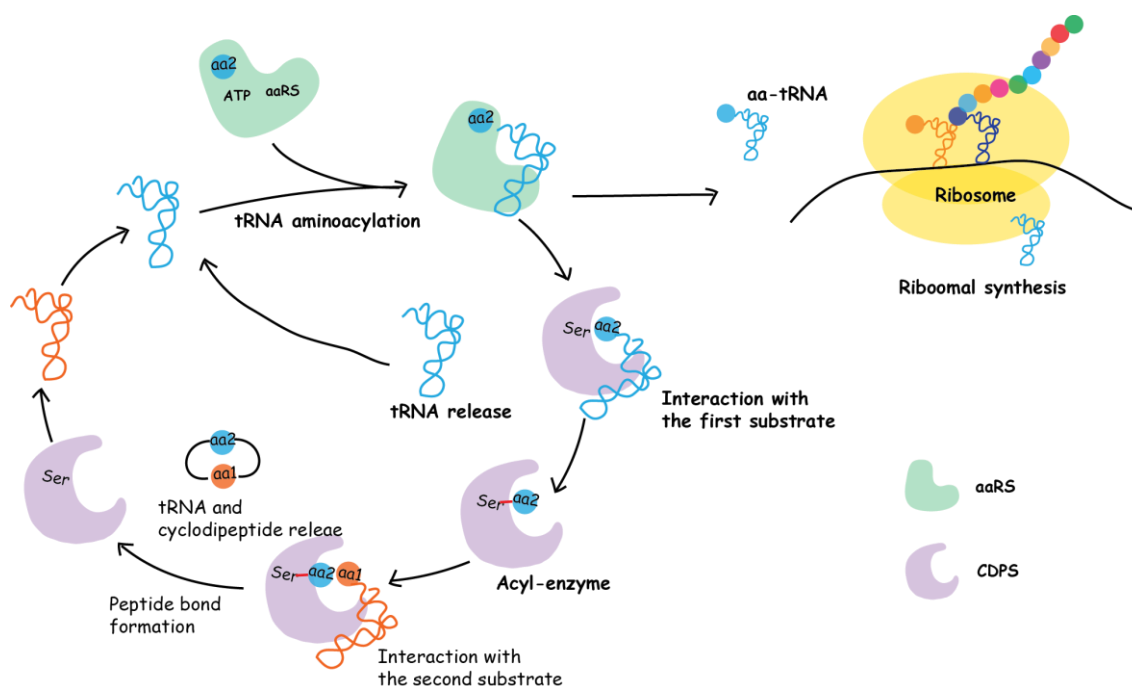


Fig 2.3.3.4 CDPS hijack aa-tRNAs to produce cyclodipeptides. aa-tRNAs are generated from amino acid, ATP and tRNAs by aaRSs (in green). aa-tRNAs are typically delivered to the ribosome (in yellow) for the synthesis of peptide bonds in nascent polypeptides, but CDPSs (in purple) hijack some aa-tRNAs to make cyclodipeptides.

The information of CDPSs suggests that some of the aminoacyl-tRNA synthetases like enzyme are related to the formation of peptide bonds by using aminoacyl-tRNA as substrates. Similar with CDPS, TvR4 might also hijack aminoacyl-tRNA in the biosynthesis of viridifungin.

TvR4 is annotated as a glutaminyl-tRNA synthetase (GlnRS). These enzymes attach Gln to the appropriate tRNA. Like the other class I tRNA synthetases, they aminoacylate the 2'-OH of the nucleotide at the 3' end of the tRNA.^[145] The core domain is based on the Rossmann fold and is responsible for the ATP-dependent formation of the enzyme bound aminoacyl-adenylate. GlnRS contains the characteristic class I HIGH and KMSKS motifs, which are involved in ATP binding. According to the bioinformatic analysis of TvR4, it shows that TvR4 also contains HIGH and KMSKS motifs (section 2.2.3.3).^[146, 147] Moreover, most bacteria lack GlnRS. In these organisms, the "non-discriminating" form of GluRS aminoacylates both tRNA(Glu) and tRNA(Gln) with Glu, which is converted to Gln when appropriate by a transamidation enzyme.^[148, 149]

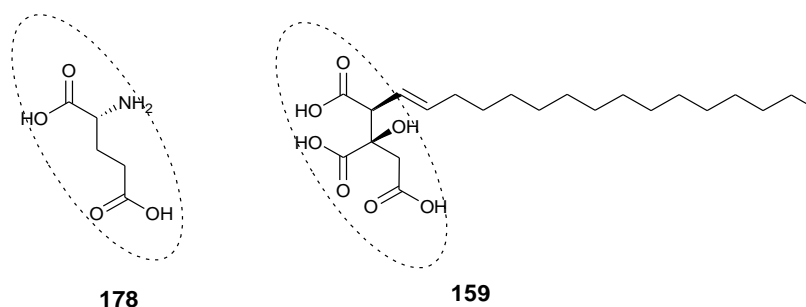


Fig. 2.3.3.5 structure of glutamic acid and proposed viridofungin precursor

Since the structure of Glu **178** is somehow similar with the head of alkyl citrate **159** which is thought to be the precursor of viridofungin (**Fig. 2.3.3.5**), and the function annotations of TVR4 is glutaminyl-tRNA synthetase. So the catalytic mechanism of TvR4 can be proposed based on the knowledge about the normal glutaminyl-tRNA synthetase involved in the primary metabolism. The three substrates of TVR4 should be ATP, alkyl citrate and a Tyr or Phe specific tRNA. The energy of ATP hydrolysis is used to attach each amino acid to its tRNA molecular in a high-energy linkage.

The citrate is first activated with ATP through the linkage of its carboxyl group to form an AMP moiety forming an adenylated citrate **179**. This linkage is driven by the hydrolysis of the ATP molecule that donates the AMP **164**. And this step happens in the activate domain of the protein which contain a ATP binding site inside. Then without leaving the synthetase enzyme, the AMP-linked carboxyl group on the citrate is then transferred to an amino group on the amino acyl tRNA **162**, this transfer joins the citrate by an activated ester linkage to the tRNA and forms the tRNA bound viridofungin **163** (**Fig. 2.3.3.6**). Therefore, there should be a condensation domain which can catalyse the linkage between activated citrate substrate and the amino acyl tRNA.

As we can see, tRNA bound viridofungin is an intermediate. A nucleotide phosphatase gene *tvR3* is located next to the tRNA ligase like gene in the viridofungin BGC (section **2.2.3.3**). It is suggested that the nucleotide phosphatase could be involved in the biosynthetic pathway of viridofungin which could release the viridofungin from tRNA. The exact activity of this nucleotide phosphatase could be to catalyse the hydrolysis of the phosphate bond between the amino acid and 3' terminus of the tRNA.

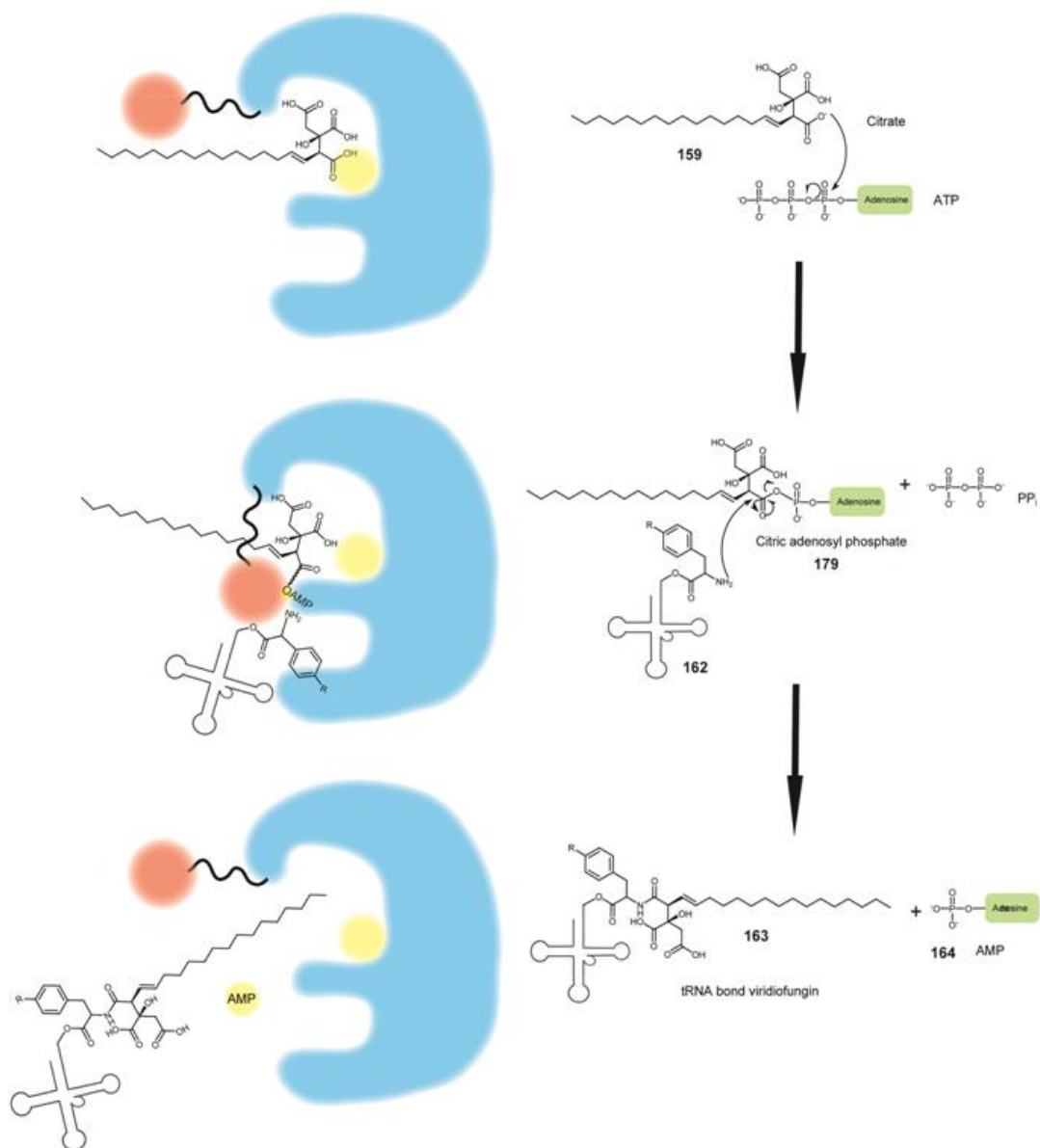


Fig. 2.3.3.6 Proposed tRNA ligase like enzyme. The main protein structure in blue, the condensation domain in red, the active site in yellow.

2.3.4 Tailoring genes in viridifungin pathway

The function of other tailoring steps encoded by the gene cluster like P450 enzyme (*tvL4*), FAD dependent monooxygenase (*tvR2*) and a NAD dependent dehydrogenase (*tvL2*) are still not very clear. According to the bioinformatic analysis, it can be proposed that the P450 enzyme might be related to the formation of the C-13 ketone group on the fatty acid. The oxygenase might add a hydroxyl group and the formation of ketone could be catalysed by the NAD dependent dehydrogenase. or the oxygenase could catalyse a two steps reaction to directly form the ketone group.

If TVL4 is responsible for the oxidation of the C-13 of viridofungin, the FAD dependent monooxygenase TVR2 might be related to the oxidation of other different places in viridofungin, because we could also be able to find other oxidised viridofungin analogous in the extraction of wild type *T. viride* strain.

2.4 Conclusion and outlook

The aim of this project was to elucidate the biosynthesis of viridofungin and to delineate enzymes involved in the novel pathway of the attachment of amino acid Try or Phe with the citrate precursor, and the tailoring enzymes involved in the region selective C-13 carbon oxidation.

First of all, the wild type strain *T. viride* was re-grown in a cornmeal contained producing medium, the suitable condition for reproducing of viridofungin was found. The compounds viridofungin A and B were then isolated and elucidated. The gDNA of *T. viride* was sequenced by using both Illumina and Oxford nanopore methods. According to the bioinformatic analysis of *T. viride* genome data, a CS and tRNA ligase like enzyme containing BGC was found to be the possible for the biosynthesis of viridofungin. In the mean time, the transformation condition of *T. viride* was tested. *T. viride* is hygromycin sensitive, and can be transformed by protoplast PEG-mediated method.

Based on the BGC prediction and fungal transformation condition, seven targeted knockout experiments were carried out. From two of these knockout experiments, citrate synthase gene (*tvCS*) and tRNA ligase like enzyme gene (*tvR4*) was successfully disrupted and the KO strain showed no production of viridofungins. According to the *CS* and *tvR4* knockout, the BGC of viridofungin was confirmed. However, the knockout of *tvL2 /3 /4* and *tvR2 / 3* were not successful.

Because knockout experiments gave were not efficient (less than 1%), other biological methods were used to study this project such as RNA interference based silencing and heterologous expression. Silencing of nucleotide phosphatase like enzyme TvR3 didn't give significant changes in viridofungin production. Heterologous expression of TvR4, TvCS, TvL2/ 3/ 4 and TvR2/ 3 in *A. oryzae* didn't give the production of viridofungin.

In conclusion, a novel non-NRPS dependent biosynthetic pathway of viridofungin was found in *T. viride*, and the core genes CS and tRNA ligase like enzyme *tvR4* were found to be involved in the biosynthesis of viridofungin by knockout experiment. However, the function of the enzymes and the tailoring steps like the oxidation of C-13 carbon are still unclear.

In the future, the transcriptome data should be obtained from cDNA sequencing to get more precise information of the open reading frame information of every gene in the BGC which

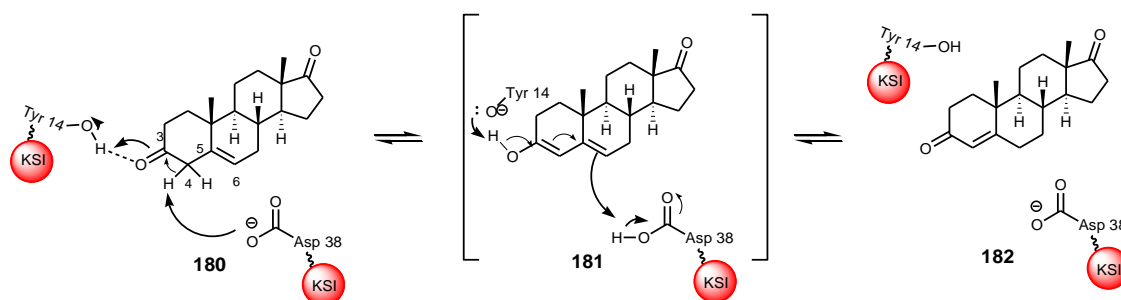
can help to carry out a better heterologous expression experiment. Then *in vitro* study of the novel tRNA ligase like enzyme TVR4 would be the better way to understand how the enzyme TvR4 possibly hijacks the primary metabolite amino acyl tRNA to form the secondary metabolite viridiodfungin. The substrate of TVR4 is thought to be amino acyl tRNA and citrate intermediate. However, the tRNA from different organism should be different, so it is also important to figure out if TvR4 can admit tRNA from different organism. *In vitro* study of Nucleotide phosphatase TvR3 study should also followed with the TvR4 to make sure the mechanism of the releasing steps. Furthermore, the tailoring genes like *tvL2*, *tvL4* and *tvR2* could be good elements for genetic engineering strategy in synthetic biology study.

Chapter 3. Heterologous Expression of the Enzymes from Nonadride Biosynthetic Pathways and *in vitro* Assay

3.1. Background and aims of the project

3.1.1 Introduction

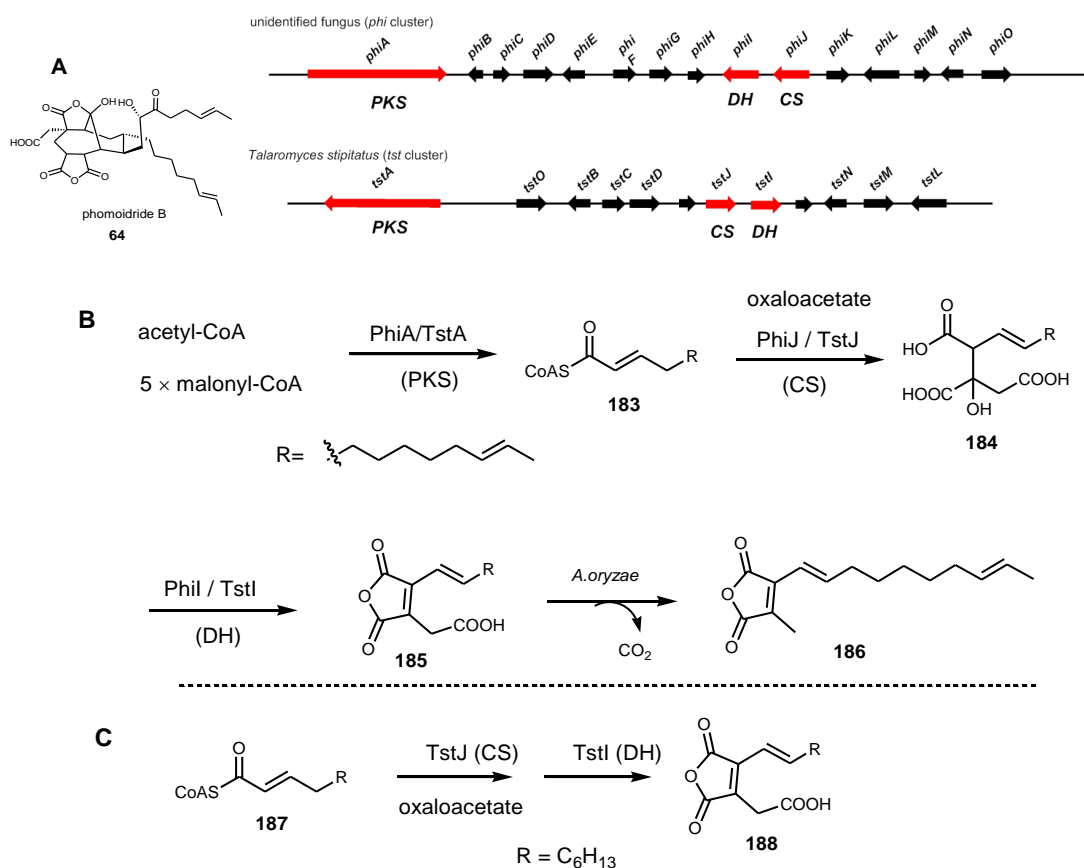
As described in section 1.5.1.2, the core early genes and two sets of KSI-like and PEBP-like genes from the byssochlamic acid pathway were investigated in *A. oryzae* by heterologous expression.^[3] Byssochlamic acid **1** and heptadrides were produced by the host strain. The KSI-like enzymes BfL6 and BfL10 from *B. fulva* appear to act together, to possibly forming a heterodimer that shows flexibility to create both nonadrides and heptadrides. Real ketosteroid isomerase enzymes (KSI) are involved in the biosynthesis of steroids in mammals and bacteria. They can isomerise 3-oxo- Δ^5 ketosteroids to Δ^4 -conjugated isomers.^[150] For example, KSI from *Pseudomonas testosteroni* catalyzes the allylic isomerization of a variety of Δ^5 -3-ketosteroids **180** to Δ^4 -3-ketosteroids **182** by a highly stereoselective intramolecular transfer of the 4 β -proton to the 6 β position (Scheme 3.1.1.1). Asp-38 was identified as the general base abstracting the 4 β -proton of the steroid substrate and Tyr-14 as the general acid protonating or providing a hydrogen bond to the dienol intermediate **181**.^[151, 152]



Scheme 3.1.1.1 General mechanism of the ketosteroid isomerase catalysis.

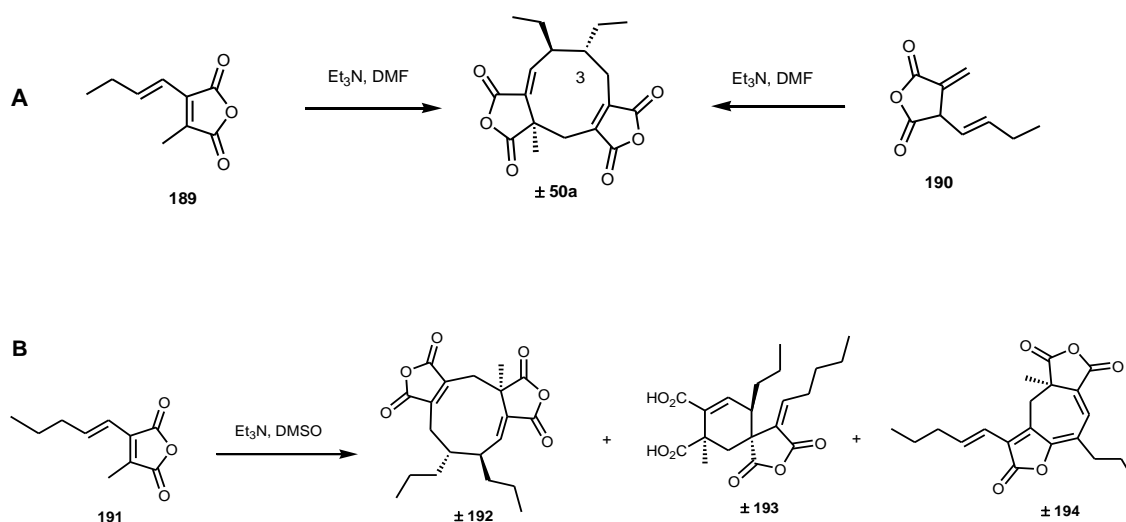
The PEBP-like proteins also appear to take part in the dimerization step by greatly increasing the production of the dimerized compounds. The Cox group speculated that the PEBP-like proteins possibly had an anionic binding ability, therefore PEBP proteins might be the chaperone highly unstable carboxylates to prevent premature decarboxylation prior to dimerization. However, PEBP-like proteins BfL5 and BfL9 are not absolutely necessary for the dimerization, and KI proteins are critical.^[3] To more deeply look into the mechanism of the dimerization step, *in vitro* studies of the KSI-like and PEBP-like proteins should be performed.

As discovered for the core early proteins in the byssochlamic acid pathway, there are likely to be many problems awaiting *in vitro* study. Although both the Cox group and Fujii *et al* confirmed that the hrPKS, (alkyl)citrate synthase and (methyl)citrate dehydratase are used for forming the maleic anhydride monomers **186** (Scheme 3.1.1.2 A), the mechanism and chemo and stereo selectivities of these enzymes are still unknown. Fujii *et al* obtained the anhydride monomer **183** from the incubation of alkylcitrate synthase (*TstJ*) and methylcitrate dehydratase (*TstI*) with oxaloacetate **45** and a synthetic polyketide CoA **187** (Scheme 3.1.1.2 B and C).^[3, 61] However, according to the results of these *in vitro* studies, Fujii *et al* thought the product they obtained was the ring closed and decarboxylated compound **188**, but the UV spectrum (λ_{\max} about 270 nm) does not corresponded to the general UV spectrum of the anhydride monomers (λ_{\max} over 300 nm). Actually, according to the studies of our group, the λ_{\max} 270 nm compounds are the ring opened dicarboxylic acids.^[8] So further work is needed to know how the methylcitrate dehydratase works. Furthermore, there is a hydrolase BfL1 involved in the byssochlamic acid gene cluster, but the precise role of this enzyme is not known.



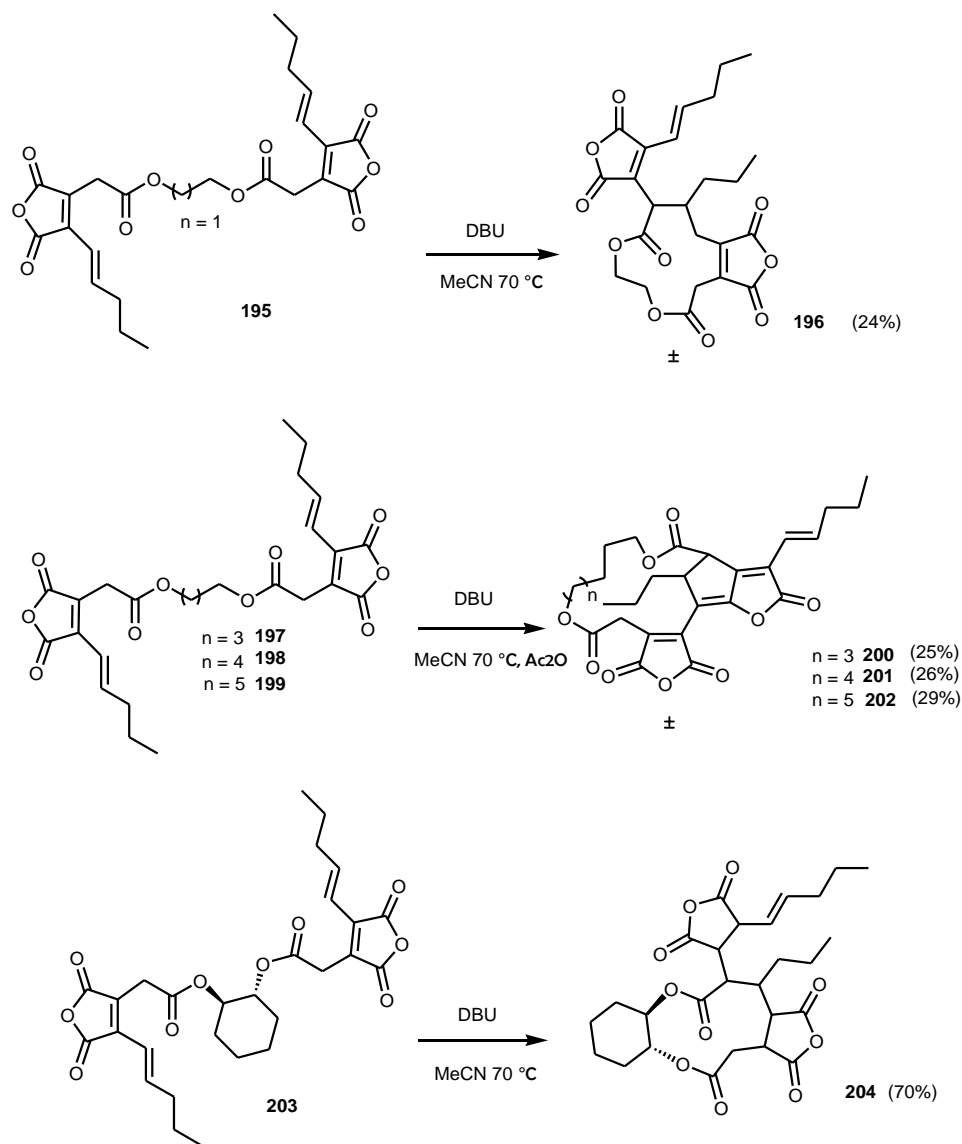
Scheme 3.1.1.2 *in vivo* and *in vitro* studies on maleidride early steps from Fujii group: **A**, gene clusters; **B**, *in vivo* study; **C**, *in vitro* study.

To investigate the dimerization of nonadrines, chemical synthesis work has also been carried out. During the 1960s, Sutherland *et al* synthesized compound **189** as a substrate and treated it with base (Et₃N), and DMF as the solvent. This gave racemic 3-*epi*-gluconic acid **50a** (**Scheme 3.1.1.3 A**) in low yield.^[153] In parallel, they also used compound **190** as a substrate, and the same product was produced as from compound **189**, which might be due to anhydride **190** being converted back to **189** under the reaction conditions. Baldwin *et al* used the anhydride with the longer chain **191** as substrate and also treated it with Et₃N but used DMSO as the solvent. This gave a mixture of products including nonadride **192**, hexadride **193** and heptadride **194**. They found that higher polarity of the solvent was important to stabilize anions and favour dimerization (**Scheme 3.1.1.3 B**). In addition, the type of base not only effected the yields but also the selectivity towards certain products.^[154]



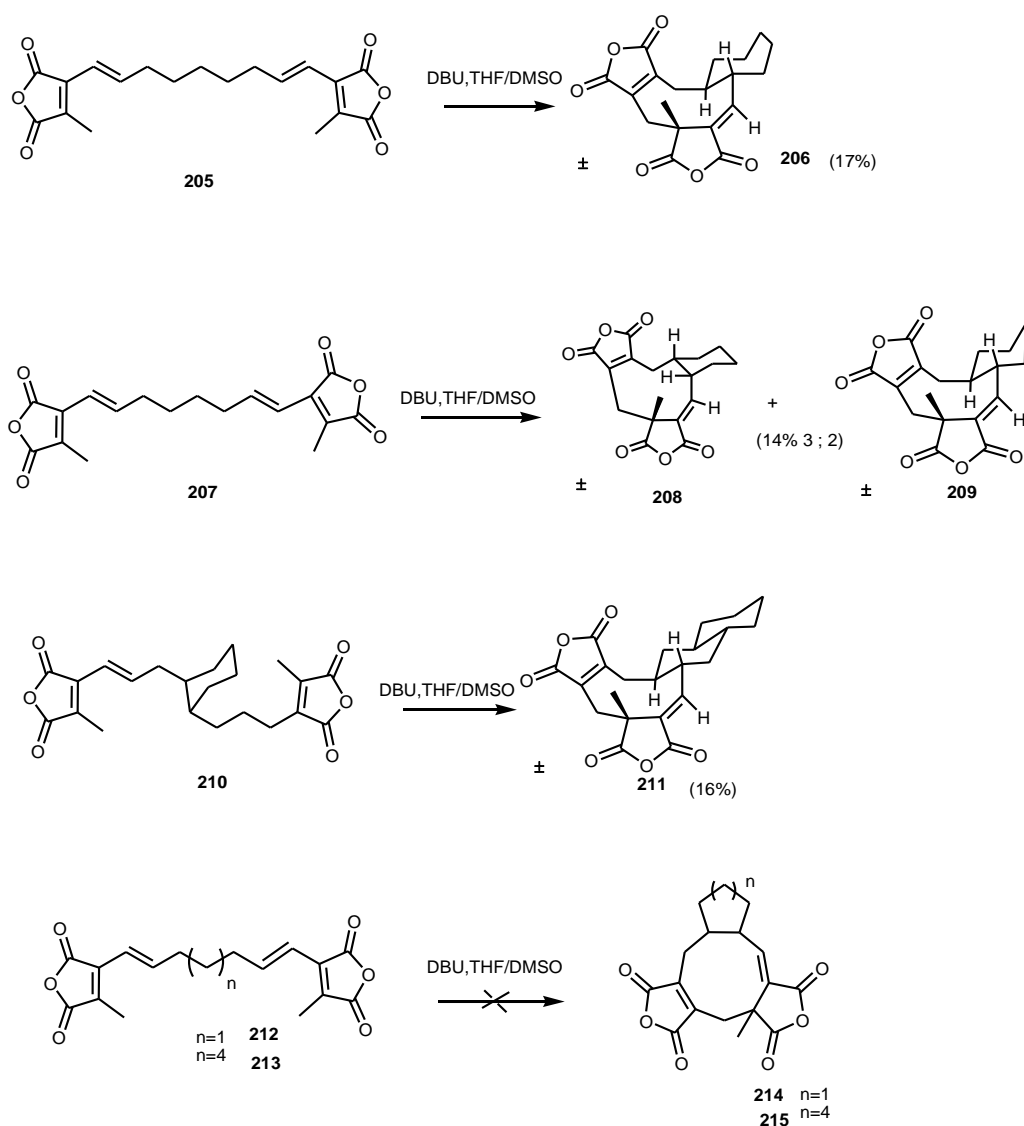
Scheme 3.1.1.3 *In vitro* investigation on dimerization: **A**, by Sutherland *et al*; **B**, Baldwin *et al*.

Later, Sulikowski *et al* changed the method for the reaction. They synthesized compounds containing two maleic anhydride units using different linkers (**195**, **197**, **198**, **199** and **203**) and used DBU and MeCN as base and solvent. The yields of cyclised products were again low. The result indicated that compound **195** underwent a cyclization reaction, while the other substrates were only dehydrated (**Scheme 3.1.1.4**).^[155]



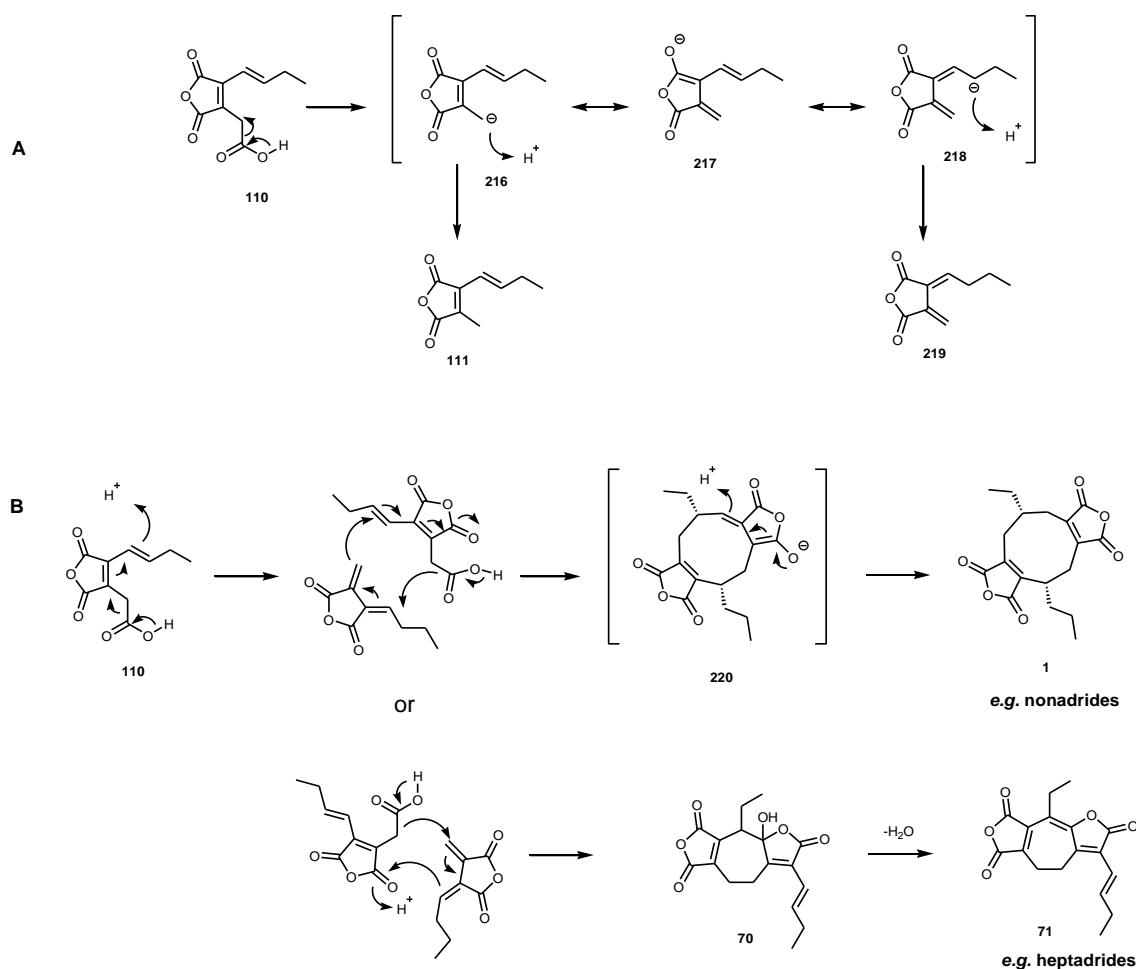
Scheme 3.1.1.4 *In vitro* investigation on dimerization by Sulikowski et al.

Baldwin also used the linked anhydride monomers **205**, **207**, **210**, **212** and **213** as substrates to explore the stereo- and regioselectivity of dimerization. In this work, they used DBU in THF:DMSO (1:4) at 20 °C. Finally, it was shown that compounds **205**, **207** and **210** can be cyclized into **206**, **208/209** and **211** also in low yields (**Scheme 3.1.1.5**).^[156]



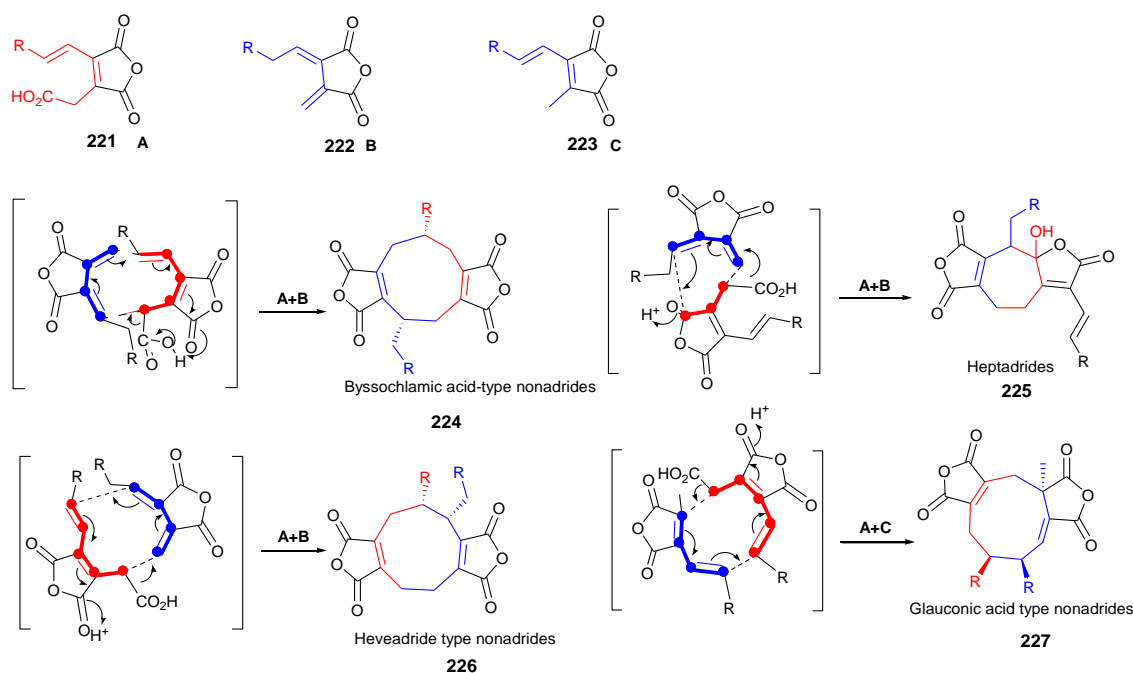
Scheme 3.1.1.5 *In vitro* experiments of Baldwin et al.

In the case of the dimerization step involved during the biosynthesis of byssochlamic acid **1**, our group worked on the purification of maleic anhydride monomer **110** and characterized it by NMR. The result indicated that the maleic anhydride monomer **110** spontaneously decarboxylates to **111** (Scheme 3.1.1.6 A).^[2] The subsequent cyclization step can be explained by a mechanism involving Michael addition of the reactive anionic species **216/217/218**, generated by decarboxylation. The decarboxylation also appears to provide the intermediate **219** necessary for cyclisation with an additional **110** during the biosynthesis of byssochlamic acid **1** (Scheme 3.1.1.6 B). This dimerization step is thought to be catalysed by KSI-like enzymes according to the *in vivo* study on heterologous expression in *A. oryzae*.^[3]



Scheme 3.1.1.6 Possible mechanism of dimerization step: **A**, formation of reactive anionic species; **B**, dimerization step.

According to these *in vivo* results,^[3] the Cox group proposed the general pathway to nonadrides, and three different ways of dimerization by using 3 types of anhydride monomers **221/ 222/223** (**Scheme 3.1.1.7**). The first one was head-to-tail dimerization for byssochlamic acid-type nonadrides **224**, the second was head-to-head dimerization for gluconic acid **227** type nonadrides and heveadride type **226** nonadrides, the third one was head-to-side dimerization for heptadrides **225**.^{[89][3]} The KSI enzymes should react **221** with either **222** or **223**, the orientation of the reacting species, determine the size, substitution pattern, and stereochemistry of the central carbocyclic ring. KSIs and PEBPs were thought to be related to the dimerization. These enzymes might have activities that could protect carbanions from reaction with water and protonation by other situations that can make them unreactive, furthermore, these enzymes should prompt close contact between two monomers, in addition, they could help to ensure the proper species-specific structure and stereochemistry.



Scheme 3.1.1.7 Proposed general pathway to nonadrides and heptadrides.

3.1.2 Project aims

To obtain more information about the mechanisms and selectivities of the nonadride enzymes we plan to study them *in vitro*. We therefore plan to try to express all of the genes, including the early core genes (except the PKS) and the genes for dimerization. The biosynthetic proteins will then be reassembled *in vitro* with suitable substrates. Because the cornexstin **66** biosynthetic pathway is similar to byssochlamic acid **1**, some enzymes from the cornexstin **66** pathway will also be chosen as candidates for expression and *in vitro* study.

For citrate synthase (BfL2), it is important to find out whether it uses a carboxylic acid, a polyketide CoA or an ACP bound polyketide as its substrate, and find out whether it produces a single stereoisomer or mixed stereoisomers of citrate products. 2-Methyl citrate dehydratase (DH BfL3 and PvL2) is thought to be the key enzyme for the biosynthesis of the maleic anhydride, however, whether the DH catalyses *both dehydration and anhydride formation* is still unknown. Furthermore, there is also a stereochemical questions on DH, whether it can select only one citrate isomer as the substrate or different isomers.

According to the *in vivo* heterologous expression studies,^[3] the KSI-like enzymes (BfL6 and BfL10) are involved in the dimerization step, and the PEBP-like (BfL5 and BfL9) enzymes seem to increase the production of dimerized products. Nevertheless, it is still a mystery why *B. fulva* can produce both nonadrides and heptadrides – do the KSI enzymes have a role in ring-size selectivity? We only know that the KSI-like enzymes are related to the dimerization step, but the exact function of the KSI-like enzymes and how they work on the maleic anhydride substrate is

still unknown. So further experiments *in vitro* should help to understand the mechanisms and selectivity of these ring-forming enzymes. Furthermore, other maleidride clusters seem to only encode single copies of KSI-like and PEBP-like enzymes. It is not known why the byssochlamic acid BGC is different and whether two copies of each gene are essential.

The overall work plan will involve the cloning and expression of the relevant proteins in suitable host organisms, followed by protein purification and *in vitro* assay with suitable substrates which could be obtained synthetically or biochemically.

3.2 Results

The project started with gene synthesis. According to the transcriptome sequencing data and the gene cluster information we have obtained, a total of seven genes from the byssochlamic acid **1** pathway were targeted.^[1] These consist of the hydrolase (*bfL1*); citrate synthase (*bfL2*); citrate dehydratase (*bfL3*); two PEBPs (*bfL5* and *bfL9*) and two KSIs (*bfL6* and *bfL10*). One gene from the cornexstin **66** pathway, 2-Methyl citrate dehydratase *pvL2*, was also targeted. Eight codon optimised genes were therefore synthesized for *E. coli* expression. These genes were cloned into *E. coli* expression vector pRSET A.^[165] The genes *bfL1* and *bfL2* were successfully expressed in *E. coli* by our group member Dr. Steffen Friedrich. In this chapter, expression work of *bfL3*, *5*, *6*, *9*, *10* and *pvL2* and their *in vitro* assay will be described.

3.2.1 Protein production

3.2.1.1 Whole length expression in *E. coli*

The project started with the *E. coli* expression of full-length copies *bfL5*, *6*, *9* and *10* using the vector pRSET A. First of all, pRSET A- *bfL5*, *6*, *9* and *10* were transformed into *E. coli* Rossetta (DE3) host strains. It appeared that the BfL5 vector was toxic to the host strain: the colonies grew very slowly on solid media, and they could not grow in liquid medium. However, transformants containing pRSET A- *bfL6*, *9* and *10* grew well in LB medium. After 0.3 mM IPTG induction for 12 h at 25 °C, the cells were lysed by ultrasonication. The cell free extraction was separated into soluble and insoluble fractions by centrifugation. Each fraction was then checked by SDS-PAGE (**Fig. 3.2.1.1**).

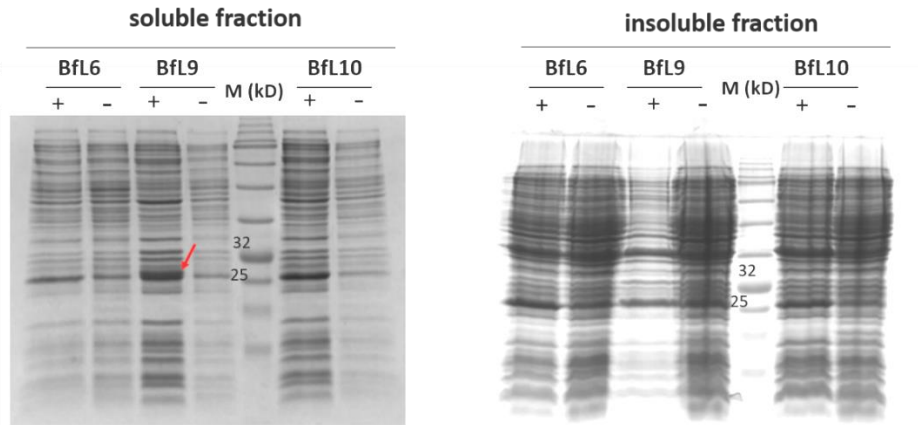


Fig. 3.2.1.1 SDS-PAGE results of pRSET-A- *bfL6*, *9* and *10* in Rossetta DE3. Control (-) is the transformants without the IPTG induction.

BfL9 appeared to be present in the soluble fraction compared with the control, and there was no specific expression in the insoluble fraction. However, neither *BfL6* nor *BfL10* had any visibly expressed protein in either the soluble or insoluble fractions.

BfL5 appeared to be toxic to *E. coli*. Therefore the toxic protein expression host strain BL21 DE3 (pLysS) was tested for the expression. The vectors pRSET A-*bfL5*, *6*, *9* and *10* were transformed into BL21 DE3 (pLysS). The transformants with *bfL5* were able to grow this time, but it still grew slower than the *bfL6*, *9* and *10* transformants. After 0.3 mM IPTG induction for 12 h at 25 °C, the expression was checked by SDS- PAGE (**Fig. 3.2.1.2**). In this case, only *bfL9* showed some expression in the insoluble fraction but without soluble protein. *BfL5*, *6* and *10* again produced no protein in either soluble or insoluble fractions.

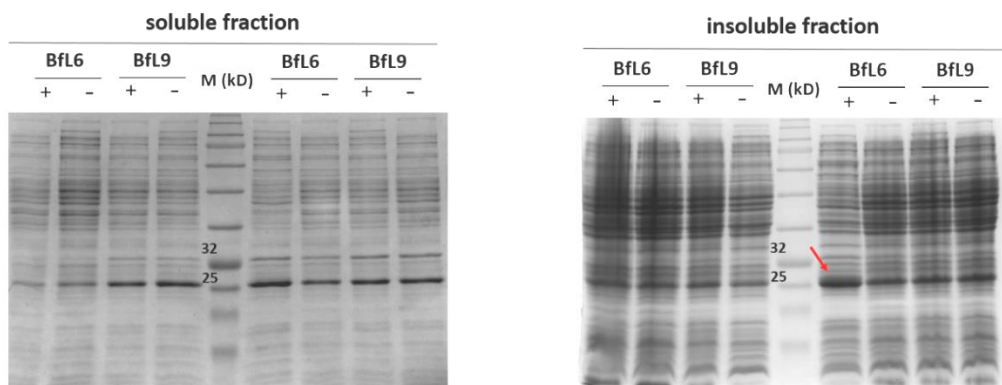


Fig 3.2.1.2 SDS-PAGE results of pRSET-A- *bfL 5*, *6*, *9* and *10* in BL21 DE3 (plysS). Control (-) is the transformants without the IPTG induction.

3.2.1.2 Bioinformatic analysis

Since it was quite difficult to express the KSI-like and PEBP-like enzymes in *E. coli*, bioinformatic analysis of the features and characters of these four proteins was done. First basic parameters were investigated using *expasy*.^[157] This included the number of amino acids, molecular weight, theoretical PI, the number of negatively charged and positively charged residues, formulas of the proteins, instability and aliphatic index, and grand average of hydropathicity (**Table 3.2.1.1**). It should be noted that the instability index of BfL10 was more than 40. This indicated that the protein is likely to be unstable. This might be caused by the hydrophobic amino acid condensed region at the N terminus. The “Grand average of hydropathicity” prediction showed that only BfL5 had a positive value. This revealed that BfL5 could be more hydrophobic compared with the other proteins (**Table 3.2.1.1**).

	BfL5	BfL6	BfL9	BfL10
Number of amino acids	214	242	210	229
Molecular weight (Da)	23223.4	27944.6	23105.5	26602
Theoretical pI	4.7	5.1	6.4	5.5
negatively charged residues (Asp + Glu)	17	30	22	27
positively charged residues (Arg + Lys)	6	18	20	17
Instability index	30.9 (stable)	32.6 (stable)	38.5 (stable)	40.1 (unstable)
Aliphatic index	99.4	79.7	75.3	85.1
Grand average of hydropathicity	0.2	-0.3	-0.4	-0.3

Table 3.2.1.1 Expasy analysis and prediction of BfL5, 6, 9 and 10.

Membrane associated proteins are usually difficult to express in *E. coli*. Proteins which contain transmembrane helices are not properly folded or modified in *E. coli*. Therefore, trans-membrane regions were predicted using the software TMHMM-2.0.^[158] This showed that three of the KSI-like and PEBP-like enzymes contained no obvious transmembrane region. However, BfL10 likely contains some hydrophobic amino acids at the N-terminus of the protein. This region could affect protein expression (**Fig. 3.2.1.3**).

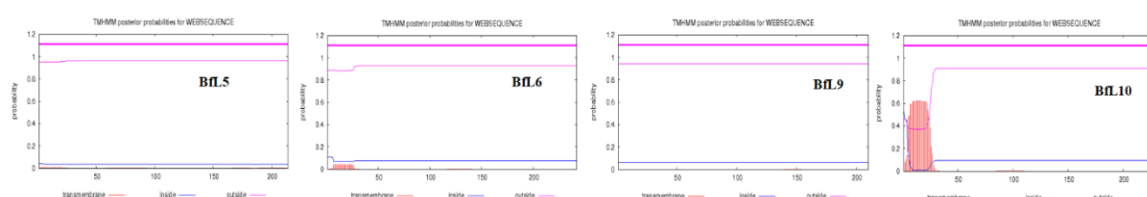


Fig. 3.2.1.3 TMHMM-2.0 prediction result of BfL5, 6, 9 and 10.

We also applied the SignalP 4.1 Server ^[159] to predict the signal peptides. Signal peptides are a short amino acid sequences at the N-terminus of the newly synthesized proteins. They can direct the proteins towards the secretory pathway. Usually, one organism can only recognize its own signal peptide to transport the proteins. In the case of heterologous expression, especially for *E. coli*, eukaryotic signal peptides cannot usually be recognized, and the organism cannot modify the signal peptide after translation. This can block the translation of the proteins. It is therefore important to remove the signal peptides before transforming the genes to *E. coli*.^[160]

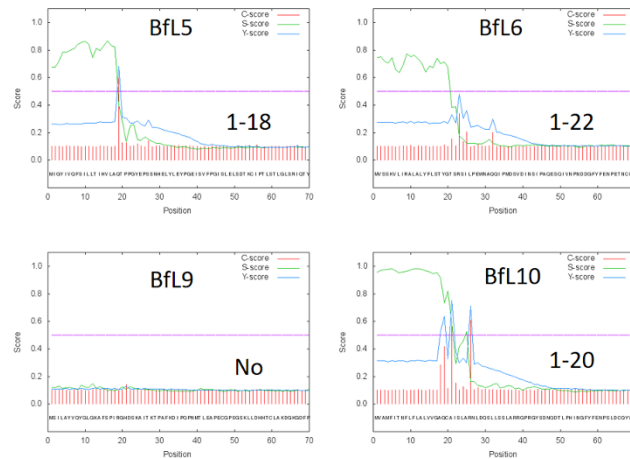


Fig. 3.2.1.4 Signal peptide prediction over residues 1 - 70 of Bfl5, 6, 9 and 10. Score means the possibility to contain a signal peptide. The S-score for the signal peptide prediction is reported for every single amino acid position in the submitted sequence; C-score is the "cleavage site" score; Y-score is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone.

The Bfl5, 6 and 10 N-terminal sequences had a high score in SignalP 4.1 Server which means they probably contain signal peptides (18, 22, 20 amino acids of each) at the N terminus, whereas Bfl9 does not (**Fig. 3.2.1.4**). This result might explain why only Bfl9 can be directly expressed in *E. coli*. However, the other three genes could not be expressed in whole length form.

3.2.1.3 Removing the signal peptide for expression

According to the signal peptide prediction results, Bfl5, 6 and 10 probably include a signal peptide at their N-termii which might affect their expression in *E. coli*. The first 22 AA of Bfl10 are also predicted to encode a *trans*-membrane domain. Therefore, new primers were designed to remove the signal peptides by creating shorter ORFs (**Fig. 3.2.1.5**). The new PCR fragments were cloned into pET-28a vectors and confirmed by colony PCR.

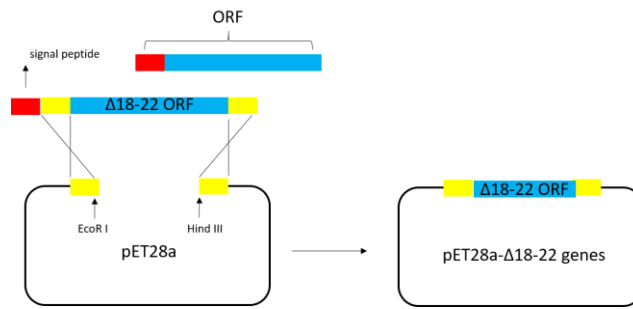


Fig. 3.2.1.5 Vector construction: PCR fragments (*bfl5* ($\Delta 1-18$), *bfl6* ($\Delta 1-22$), *bfl10* ($\Delta 1-20$)) without the signal peptide regions, pET-28a cutting and infusion of PCR products.

The new protein expression vectors pET-28a-*bfl5* ($\Delta 1-18$), *bfl6* ($\Delta 1-22$) and *bfl10* ($\Delta 1-20$) were then transformed into *E. coli* BL21 (DE3). After 0.3 mM IPTG induction for 12 h at 25 °C, the expression was checked by SDS- PAGE (**Fig. 3.2.1.6**).

The SDS-PAGE of the new expression showed that BfL5, 6 and 10 had expressed, but all of the visible expression was in the insoluble fraction. There was no evidence in the SDS-PAGE for any expressed protein in the soluble fraction. Even though the proteins were insoluble, this was the first time that *bfl5*, 6 and 10 were successfully expressed in *E. coli*. It means that removing the signal peptide helped the organism to start the translation properly.

Although *bfl5*, 6 and 10 were expressed after removing the N-terminal signal peptide regions, the expressed protein was still insoluble. It means that the proteins are probably not correctly folding after the translation step. To obtain soluble expression, different conditions of expression were tried many times, including lower IPTG concentration, lower temperature and lower shaking speed. But none of these changes resulted in soluble expression.

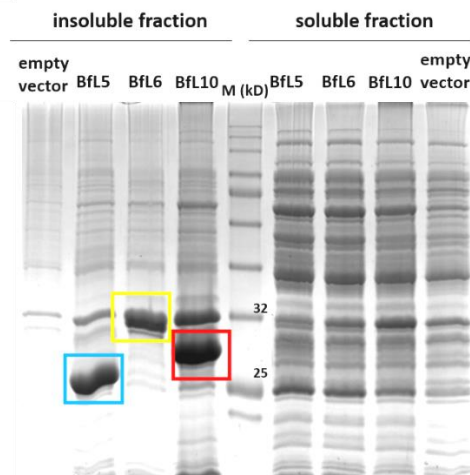


Fig. 3.2.1.6 SDS-PAGE results of pET-28a-*bfl5* ($\Delta 18$), *bfl6* ($\Delta 22$) and *bfl10* ($\Delta 20$). Control is the empty vector PET-28a transformant.

3.2.1.4 Use of chaperones

Insoluble proteins form inclusion bodies. This is usually caused by misfolding of the protein.^[161] In nature, many organisms can produce proteins to help the other proteins fold properly (chaperone proteins have this function).^[162] The major function of chaperones is to prevent newly synthesised polypeptide chains and assembled subunits from aggregating into non-functional forms.^[163]

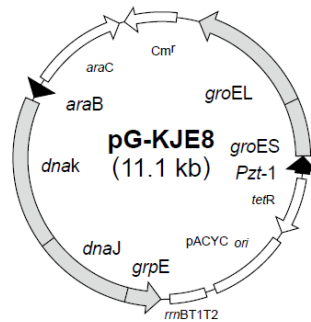


Fig. 3.2.1.7 Chaperone protein expression plasmid.^[x]

We therefore attempted to introduce chaperone proteins into the expression systems. The plasmid used in this experiment was pG-KJE8.^[164] This plasmid can introduce five chaperone proteins which work as two functional groups (DnaK/DnaJ/GrpE and GroEL/GroES). They are induced by arabinose and tetracycline separately. The vector also contains a C_m^r (chloramphenicol) resistance marker for selection (Fig. 3.2.1.7).

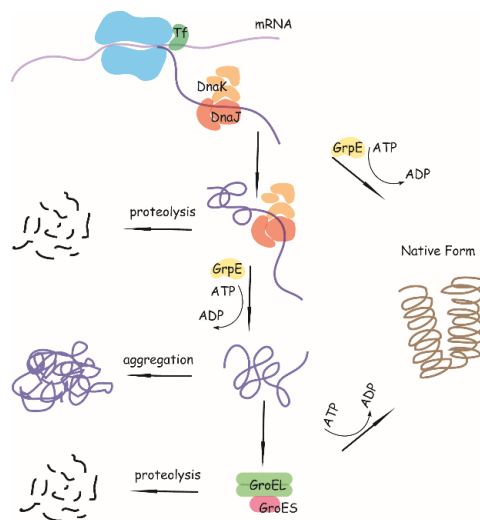


Fig. 3.2.1.8 Diagram to explain how the chaperone proteins work after protein translation.

The chaperones used in the pG-KJE8 system work as ATP-dependent foldases. Both the DnaK/DnaJ/GrpE and GroEL/GroES systems work after translation and are coupled to ATP hydrolysis (Fig. 3.2.1.8).^[161] The pG-KJE8 plasmid and the destination gene expression vectors were co-transformed into *E. coli* BL21 (DE3), then selected by both *Cm* and *Kan* antibiotics. The dnaK/dnaJ/grpE chaperone proteins were induced by addition of 0.5 mg/ml arabinose and groEL/groES were induced by addition of 5 ng/ml of tetracycline. The *bfl5*, 6 and 10 genes were induced by addition of 0.3 mM IPTG. After induction for 12 h at 25 °C, expression was checked by SDS-PAGE.

All three target proteins were observed in both soluble and insoluble fractions. The expression systems still produced quite a lot of protein in the insoluble fraction, however, with the help of the chaperone proteins, some soluble protein of the expected molecular weight was observed (Fig. 3.2.1.9).

An experiment that only co-expressed chaperone proteins groEL/groES with the target genes was also performed. Bfl5, 6 and 10 were only observed in the insoluble fraction, and there was no observed protein in the soluble fraction (Fig. 3.2.1.10). It means the single group of GroEL/GroES chaperone proteins were not enough for soluble expression of the target proteins.

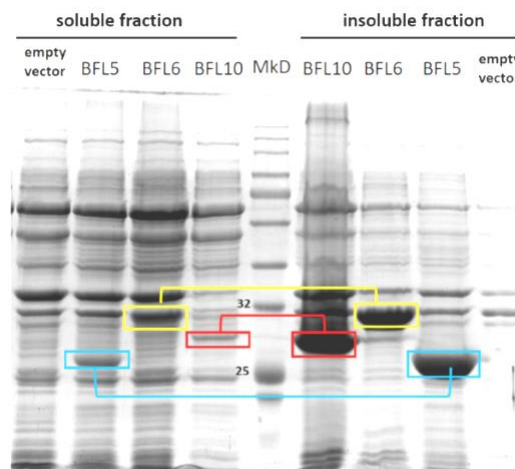


Fig. 3.2.1.9 SDS-PAGE results of PET-28a-*bfl5*($\Delta 18$), *bfl6*($\Delta 22$), *bfl10*($\Delta 20$) co-expression with DnaK/DnaJ/GrpE and GroEL/GroES. Control is the empty vector PET-28a transformant.

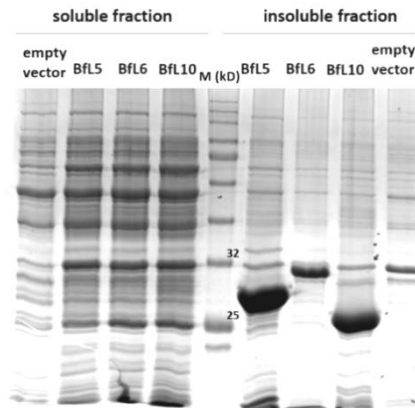


Fig. 3.2.1.10 SDS-PAGE results of PET-28a-*bfl5* ($\Delta 18$), *bfl6* ($\Delta 22$), *bfl10* ($\Delta 20$) co-expression with GroEL/GroES. Control is the empty vector PET-28a transformant.

In summary, in this section, KSI-like and PEBP-like proteins Bfl5, 6, 9 and 10 were expressed in *E. coli*. The *bfl9* gene could be directly expressed as a soluble protein in *E. coli* with the whole length of the gene. The *bfl5,6* and *10* genes were co-expressed with chaperone proteins without the signal peptide region in the genes to give a mixture of soluble and insoluble protein products.

3.2.1.5 Yeast expression

3.2.1.5a Vector construction

According to the results from Dr. Steffen Friedrich, 2-methyl citrate dehydratases Bfl3 and Pvl2 did not expressed well in the *E. coli* system. So the yeast expression system was used to express these two enzymes. Compared with the *E. coli* system, yeast has a major advantage in its ability to catalyze post-translational modification. In addition it is a eukaryotic host which may be more suitable for fungal gene expression. The insolubility of the protein is usually caused by misfolding. This will result in the formation of inclusion bodies. However, the post-translational modification in yeast can help the heterologous protein to fold correctly to make soluble protein. The difficulty of production of soluble Bfl3 and Pvl2 in *E. coli* might be caused by misfolding. At the same time, the KI and PEBP genes *bfl5*, 6 and *10* were also tested in the yeast system, to determine if more soluble protein could be obtained them in the previously optimised *E. coli* system.

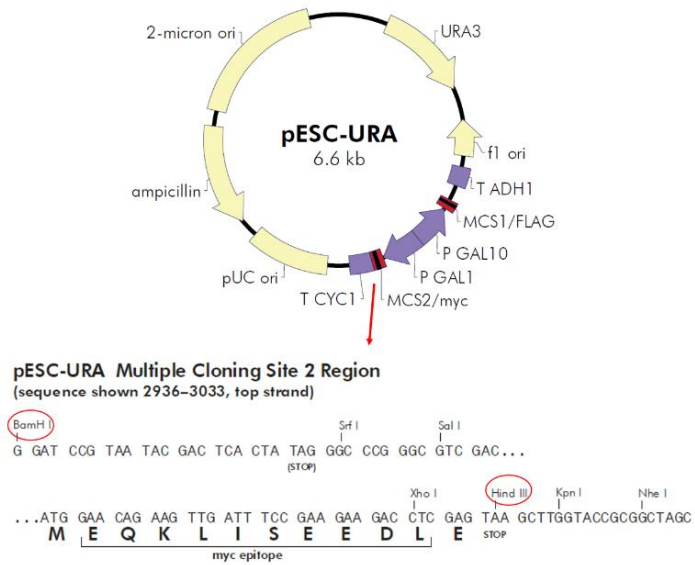


Fig. 3.2.1.11 plasmid map of pESC-URA and the multiple cloning site.

Saccharomyces cerevisiae strain W303-1b and vector pESC-URA (delivered by Prof. Kong) were used to set up the yeast expression system.^[165] W303-1b is a nutrition defective strain which is auxotrophic in Ura, His, Leu, Ade and Trp.^[166] The pESC vectors are a series of epitope-tagging vectors designed for expression and functional analysis of eukaryotic genes in *S. cerevisiae*. These vectors contain the GAL1 and GAL10 yeast promoters in opposing orientations.^[167] With the pESC-URA vector, one or two cloned genes can be introduced into a yeast host strain under the control of a repressible promoter. Genes *bfL3*, *pvL2* and *bfL5*, 6 and 10 were cloned into MCS2 (multiple cloning site 2) between *BamH* I and *Hind* III restriction sites (Fig. 3.2.1.11). The In-fusion cloning method was used for the construction of the plasmid (Fig. 3.2.1.12).^[168] The whole ORF fragments of *bfL3*, *pvL2* and *bfL5*, 6 and 10 were cloned into pESC-URA and confirmed by colony PCR. Plasmid DNA was extracted from positive colonies for yeast transformation.

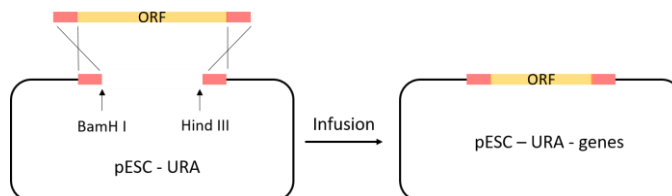


Fig. 3.2.1.12 Yeast expression plasmid construction.

3.2.1.5b Induction and Expression

Individual pESC-URA plasmids containing the genes (*bfl3*, *pvL2* and *bfl5*, 6 and 10) and the empty vector were transformed into yeast using a LiOAc procedure (section 5.2.5). The transformation mixture was plated on SM-URA agar plates. The yeast colonies grew on the plates, but the colour of the colonies changed to pink (Fig. 3.2.1.13), because the host strain W303-1b is an adenine defective strain. When the yeast cells find that there is not enough adenine in the environment, they try to synthesise adenine by themselves.^[169] However, the adenine biosynthetic pathway of the W303-1b strain was blocked, so the cells could only produce the pink intermediate pigment. This phenomenon showed that the pink yeast colonies were not contaminated by other yeast strains. Eight colonies from each transformation were selected for further growth and PCR confirmation.

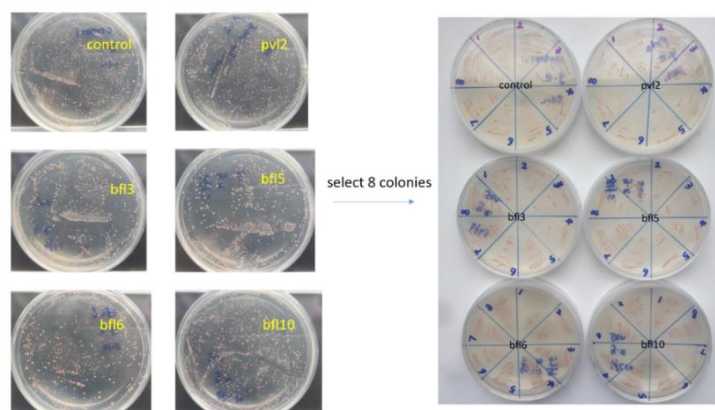


Fig. 3.2.1.13 Yeast transformation results of pESC-URA-*bfl3*, *pvL2* and *bfl5*,6,10.

The yeast cells were collected from the second plates and tested by colony PCR. The positive yeast colonies went into the next inducing and expression step. Because the expression vector contains the *S. cerevisiae GAL1* promoter, it can be induced by galactose. By using media lacking uracil and containing galactose, the yeast was induced for 16 hours at 30 °C (section 5.2.6). Glass beads were then used to break the cells. The crude cell extract was then checked by SDS-PAGE.

PvL2 and Bfl3 have some weak but visible expression on SDS-PAGE compared to the empty vector transformant. The molecular weight of PvL2 and Bfl3 are 55 kDa, and this was consistent with the SDS-PAGE results (Fig. 3.2.1.14). However, Bfl5, 6 and 10 had no visible protein on the SDS-PAGE, but this does not mean that there was no expression of these three genes, because, the promoters of the yeast were not as strong as the *E. coli* T7 promoters, and the growing speed of the yeast was also much slower than *E. coli*. Yeast protein expression is usually difficult to observe on the SDS-PAGE. The genes *bfl5*, 6 and 10 might have been expressed, but the detecting method was not sensitive enough to show the results. Bfl5, 6 and 10 could directly

III Expression in yeast

be used for the *in vitro* assay. The other way around, enzyme activity can reflect on the yeast expression result.

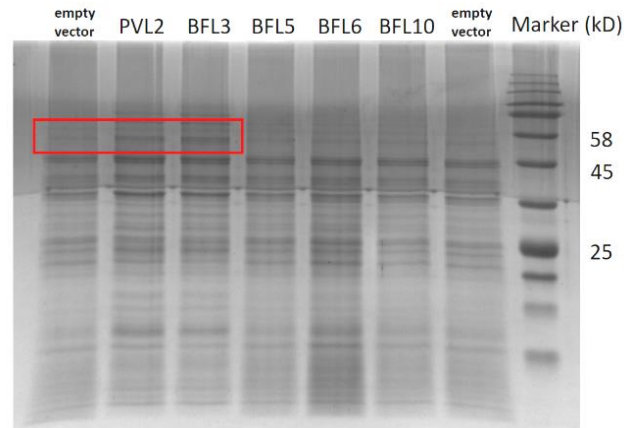


Fig. 3.2.1.14 SDS-PAGE results of yeast expression of *pvL2/ bfl3* and *bfl5, 6* and *10*. The control is empty vector transformant. The red frame highlights the possible expression.

To sum up, by using yeast expression methods, two citrate dehydratase genes (*pvL2* and *bfl3*) were expressed. The genes *bfl5, 6* and *10* were successfully transformed into the yeast, but the SDS-PAGE was not sensitive enough to confirm the expression. Besides, Dr. Steffen Friedrich successfully expressed the hydrolase (*bfl1*) and CS (*bfl2*) in *E. coli*. The gene *bfl5* ($\Delta 18$), *6* ($\Delta 22$) and *10* ($\Delta 20$) were expressed in *E. coli* (**Table 3.2.1.2**).

	in <i>E.coli</i>			in yeast
	whole ORF		with chaperone	
<i>bfl1</i> (hydrolase)	✓	-	-	-
<i>bfl2</i> (CS)	✓	-	-	-
<i>bfl3</i> (DH)	✗	-	✗	✓
<i>pvL2</i> (DH)	×	-	×	✓
	Δ signal peptide			
<i>bfl5</i> (PEBP1)	×	✗	✓	?
<i>bfl6</i> (K11)	×	✗	✓	?
<i>bfl9</i> (PEBP2)	✓	-	-	-
<i>bfl10</i> (K12)	×	✗	✓	?

×	no expression
✓	soluble expression
✗	expressed but insoluble
?	not sure if it is expressed
-	not tried

Table 3.2.1.2 Summary of expression.

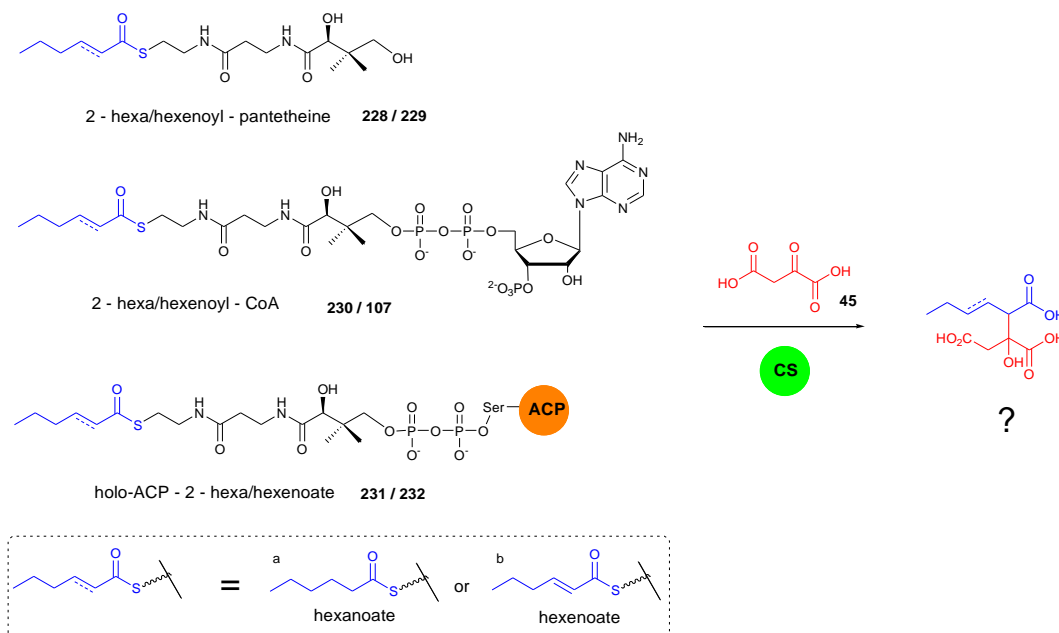
3.2.2 *In vitro* studies

3.2.2.1 *In vitro* study of citrate synthase

As is shown in section 3.2.1, the enzymes BfL1 (hydrolase) and BfL2 (CS) were successfully expressed in *E. coli* by our group member Dr. Steffen Friedrich. He also carried out the *in vitro* study on the BfL1 and BfL2. BfL1 was shown to hydrolyze 2 - hexan/hexen-oyl – pantetheine **228/229**, 2 - hexan/hexen-oyl – CoA **230/107** and also *holo*-ACP - 2 - hexan/hexen-oate **231/232** to form the released hexaketide. To make an overall story of byssoclamic acid biosynthetic pathway, Dr. Steffen Friedrich's *in vitro* works on BfL2 will be briefly demonstrated in this section, and I will try to analyse these *in vitro* data. In addition, to analyse the stereochemistry of CS, I will also try to make the protein modelling of BfL2 to look into the catalysis mechanism of this enzyme.

3.2.2.1a *In vitro* assay of BfL2

CS is known to use oxaloacetate and either polyketide CoA or polyketide bound ACP as substrates to form alkyl citrate products. In this study, a series of hexaketide substrates were synthesized for *in vitro* assay (**Scheme 3.2.2.1**). 2 - Hexan/hexen-oyl – pantetheine **228/229**, 2 - hexan/hexen-oyl – CoA **230/107** and *holo*-ACP - 2 - hexan/hexen-oate **231/232** were used as substrates for BfL2 reactions. The ACP domain of *holo*-ACP - 2 - hexan/hexen-oate **231/232** substrate is from the PKS in the byssolamic acid biosynthetic pathway from *B. fulva*.



Scheme 3.2.2.1 Proposed hexaketide substrates for CS reactions.

1 mM Hexaketide substrates and 1 mM oxaloacetate with purified CS protein BfL2 at 30 °C for 2 hours in 100 µL PBS buffer. The reaction was stopped by addition of 100 µL acetonitrile (section 5.3.5). Protein was removed by centrifugation and the supernatant was analysed by directly LCMS. According to the LCMS analysis of the reactions, BfL2 only showed the activity on 2 – hexan/hexen-oyl – CoA, and formed the citrate product (Fig. 3.2.2.1 and Fig. 3.2.2.2). 2 - Hexan/hexen-oyl – pantetheine 228/229 and *holo*-ACP - 2 - hexan/hexen-oate 231/232 did not showed activity in BfL2 reaction. Hexanoyl – CoA 230 has m/z $[M - H]^-$ 864. After the BfL2 reaction, a new compound was observed by ELSD. The new peak had m/z $[M - H]^-$ 247 which was consistent with the citrate product 233. The boiled enzyme control showed no activity on the substrate (Fig. 3.2.2.1).

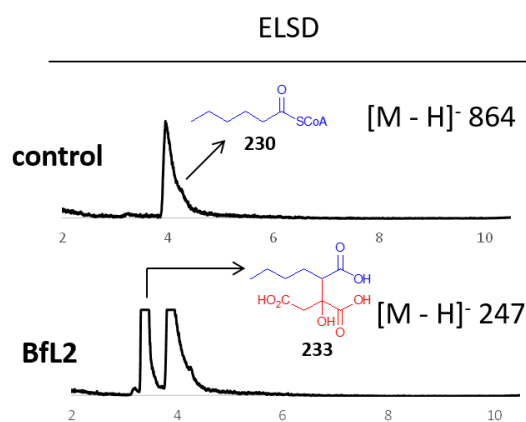


Fig. 3.2.2.1 LCMS analysis of BfL2 reaction on hexanoyl – CoA substrate.

On the other hand, hexenoyl – CoA 107 was also used as substrate for BfL2 reaction. Hexenoyl – CoA 107 ($[M - H]^-$ 862) was catalyzed by BfL2 to form a new peak which has m/z $[M - H]^-$ 245. The molecular weight of this compound was consistent with the citrate product 108. The boiled enzyme control showed no activity on the substrate (Fig. 3.2.2.2).

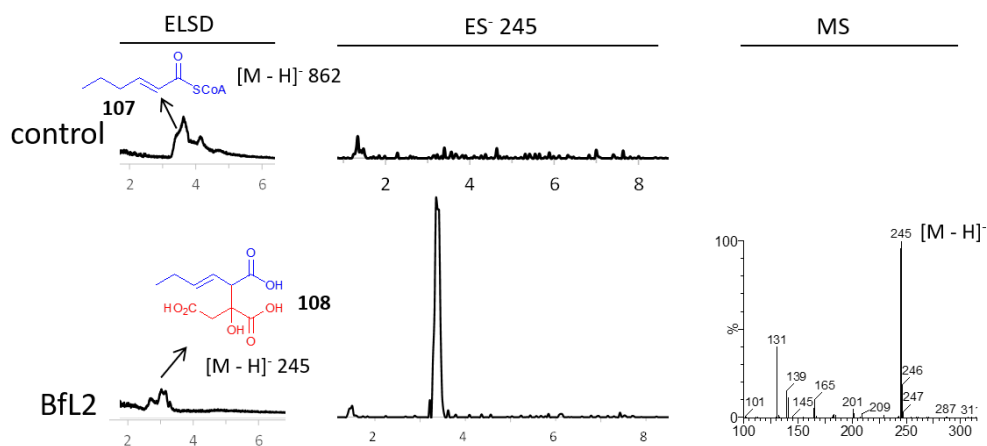
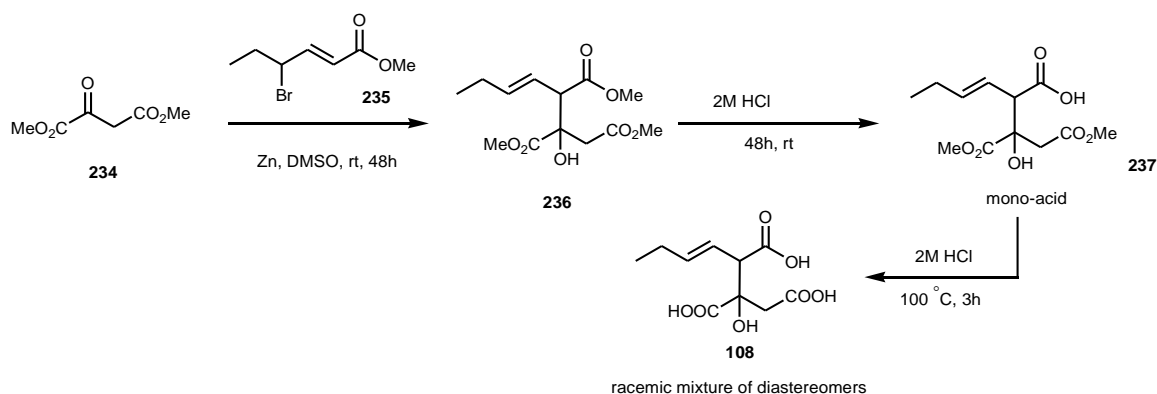


Fig. 3.2.2.2 LCMS analysis of BfL2 reaction on hexenoyl – CoA substrate.

The *in vitro* assay indicated that BfL2 can use both hexanoyl – CoA **230** and hexenoyl – CoA **107** as substrates to form citrate products, but the configuration of the product was not clear. Because the production of the citrate in these analytical scale reactions was not enough for NMR analysis, the chemical synthesis of citrate **108** is a better way to compare with the BfL2 product. The citrate substrate **108** for 2-methylcitrate synthase reaction was synthesized by Steffen Friedrich (Scheme 3.2.6.1).



Scheme 3.2.2.2 Synthesis route to alkyl citrate **108**.

According to proton NMR analysis, alkyl citrate **108** is a mixture of different stereoisomers and ring-open and ring-closed forms (Fig. 3.2.2.3). Comparing with the literature, the proton signals were annotated as follows: the proton indicated in blue was from the major *anti*-diastereomer **108a** ($2R, 3S / 2S, 3R$); and the proton indicated in yellow was the minor *syn*-diastereomer **108b** ($2R, 3R / 2S, 3S$).^[170-172] From HMBC data, we could see the difference between two diastereoisomers (Fig. 3.2.2.4). It appeared that compound **108** seems to cyclise by itself, the proton signals in purple are from the cyclised isomers *anti* **238a** and *syn* **238b**. None of these compounds have any significant UV absorption when the mixture was examined by LCMS.

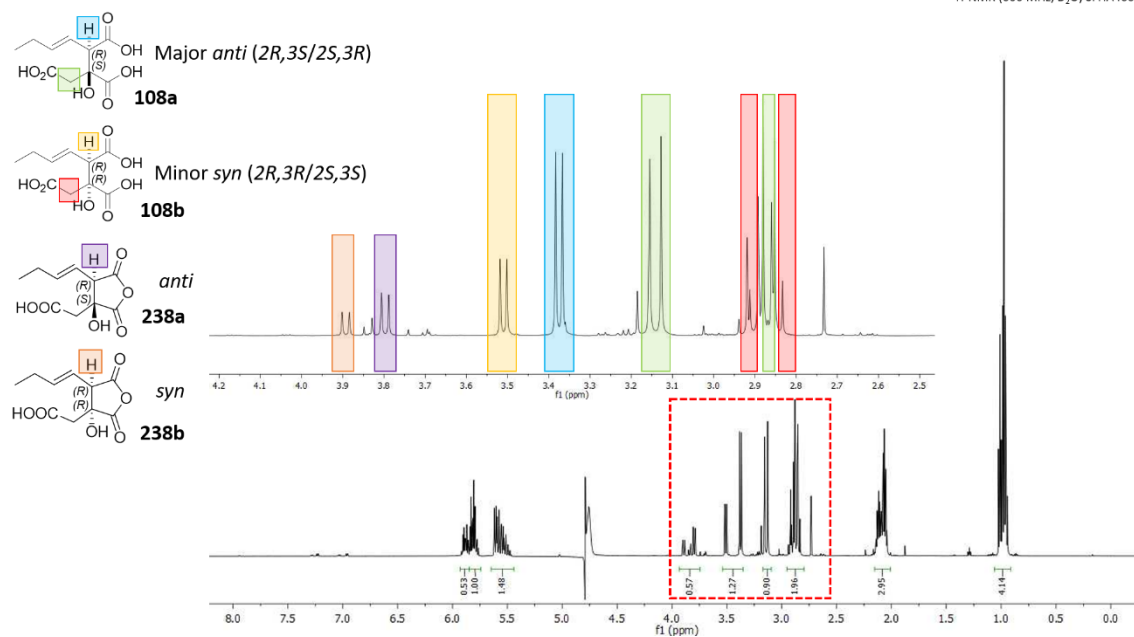


Fig. 3.2.6.2 Proton NMR of the citrate and annotation of the isomers.

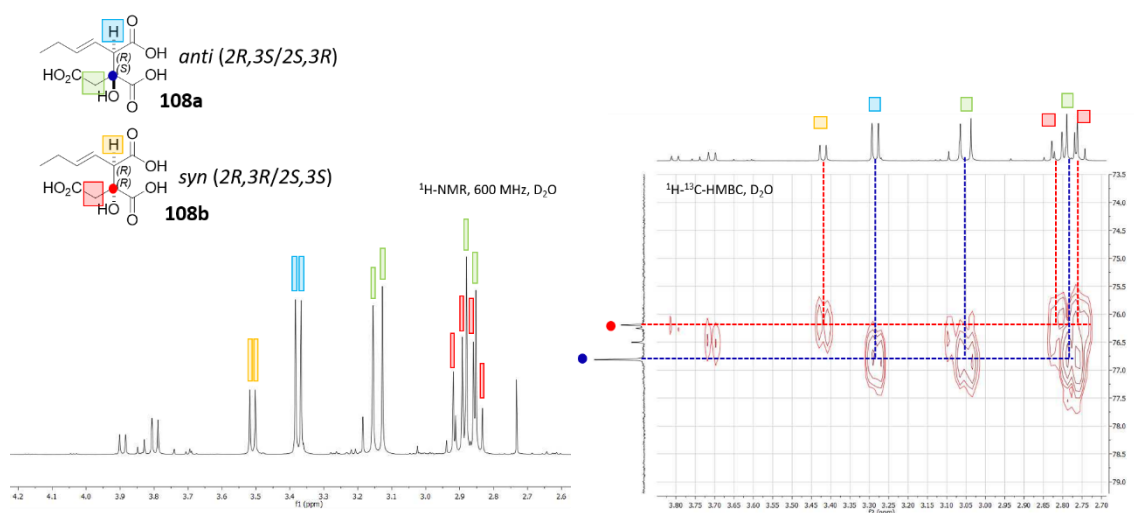


Fig. 3.2.6.3 Proton NMR and HMBC data of two diastereomers.

The mixture of *syn* and *anti* diastereomers of the synthetic citrate **108** was analysed by LCMS. According to the analytical LCMS of **108**, the retention time of **108a** (major compound) is 3.3 min, and **108b** is 3.5 min. Neither of these two isomers have specific UV absorption. Both of these two isomers have [M - H]⁻ 245 (Fig. 3.2.2.5). However, there was no signal (either ELSD or MS) of cyclised compound **238** on LCMS. This is consistent with previous results of Dr. Agnieszka Szwalbe, from the Cox group, who showed that **238** was volatile.

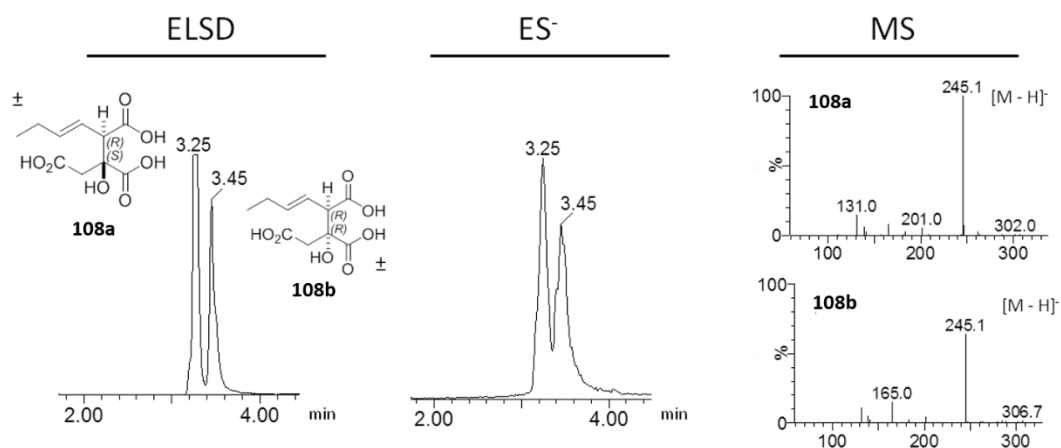


Fig. 3.2.2.5 LCMS analysis of citrate substrate

Extracted ion chromatogram analysis for the 245 ion showed two peaks. However, the BfL2 reaction trace only showed one peak which had the same retention time as the *anti*-diastomer **108a** (Fig. 3.2.2.6). It means BfL2 has a stereoselectivity on the substrate to form a single *anti* isomer. But whether the substrate is (*2R*, *3S*) or (*2S*, *3R*) is still unknown.

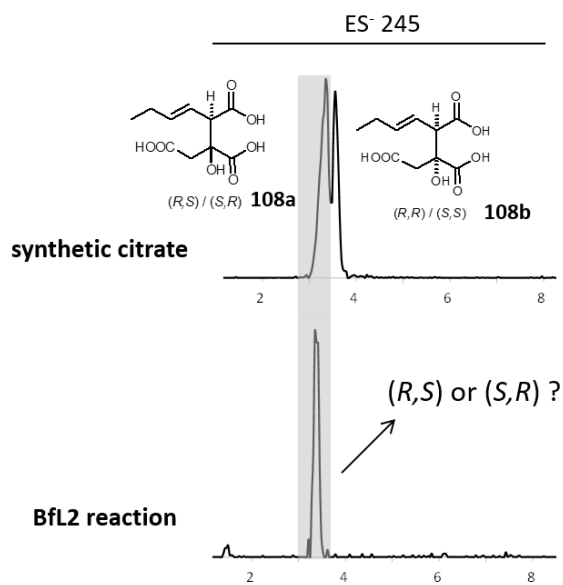
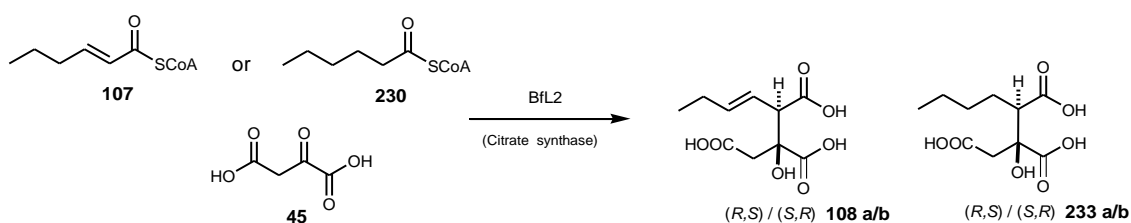


Fig. 3.2.2.6 LCMS analysis of BfL2 reaction and synthetic citrate substrate.

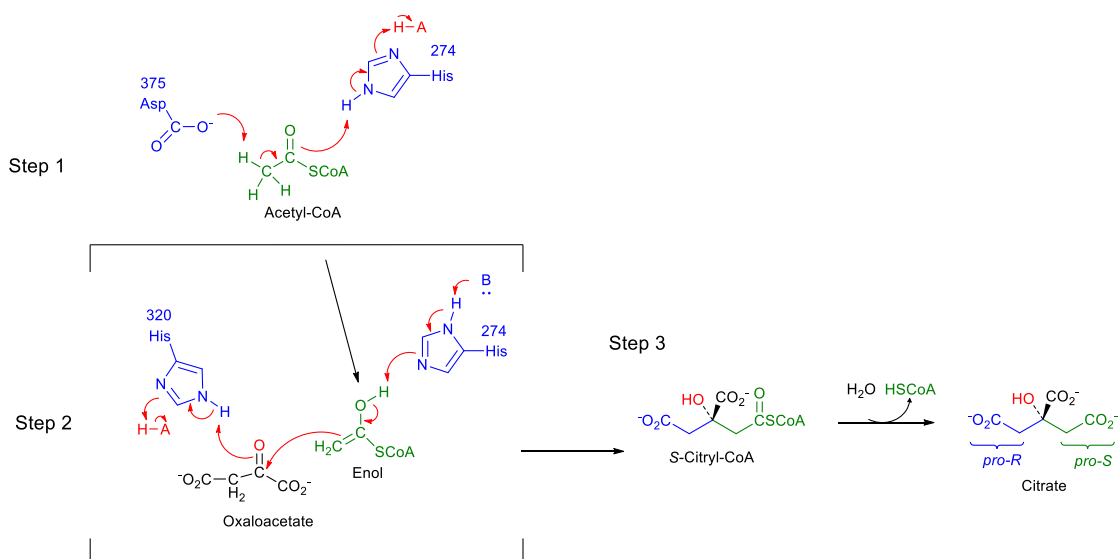
3.2.2.1b Multiple alignment and modelling of citrate synthase

As described in section 1.5.3, BfL2 is a citrate synthase like enzyme encoded by the byssochlamic acid BGC. BfL2 catalyses the formation of an alkyl citrate using oxaloacetate **45** and polyketide CoA as substrate. Dr. Steffen Friedrich's *in vitro* assay of BfL2 showed that BfL2 could use the tetraketide CoA (**107** or **230**) and **45** to produce (*2R*, *3S* or *2S*, *3R*) distereoisomers (**108 a/b** or **233 a/b** Scheme 3.2.2.3). To additionally investigate the absolute configuration of the BfL2 product, protein modelling and multiple alignment of BfL2 will be used to analyse the catalysis mechanism. The known CS crystal structures and sequence will be used as models and templates.



Scheme 3.2.2.3 Proposed BfL2 reactions.

During the 1990s, Karpusas *et al* obtained a crystal structure of *Thermus thermophilus* citrate synthase complexed with oxaloacetate **45** and the unreactive alkyl CoA derivative carboxymethyl CoA.^[129] Based on this structure, the mechanism for the condensation reaction of citrate synthase was deduced.^[173] The overall CS reaction is thought to proceed through three partial reactions and involves both closed and open conformational forms of the enzyme. These steps begin with generation of the carbanion (or equivalent) from acetyl-CoA **46** by base abstraction of a proton. The side-chain carboxylate of Asp-375 removes an acidic α proton from acetyl-CoA **46**, while the side-chain NH of His-274 protonates the carbonyl oxygen, giving an enol. In the second step the nucleophilic attack of this carbanion on the *Si* face of oxaloacetate generates citryl-CoA. His-274 deprotonates the acetyl-CoA enol **239**, which adds to the ketone carbonyl group in an aldol-like reaction. His-320 simultaneously protonates the oxaloacetyl carbonyl oxygen, producing *S*-citryl CoA **240**. Finally, the hydrolysis of citryl-CoA to produce citrate **241** and CoA is catalysed. *S*-citryl CoA **240** is hydrolyzed to citrate by a typical nucleophilic acyl substitution reaction, catalyzed by the same citrate synthase enzyme (Scheme 3.2.2.4).^[173, 174]



Scheme 3.2.2.4 Catalytic mechanism of citrate synthase in primary metabolism.

In order to understand the reaction of the BfL2 citrate synthase better, we compared it to the well-studied primary metabolism citrate synthases from *T. thermophilus*, *Pyrococcus furiosus* and *E. coli* that have been previously studied by crystallography and site direct mutation.^[129, 175, 176]

According to the basic information of the CS from primary metabolism of *T. thermophilus*, His-274, His-320 and Asp-375 are the conserved residues in the active site. So the amino acid sequences of primary metabolic CS from *P. furiosus* and *E. coli* (crystal structure already known), as well as secondary metabolic CS from *B. fulva* (BfL2 of byssoclamic acid **1** pathway) and *P. variotii* (PvL7 of cornexstin **66** pathway) were chosen to make a multiple alignment with *ClustalX*. The result indicated that the three residues (**Fig. 3.2.2.7** highlighted with “*”) in the active site are conserved in both primary and secondary metabolic CS. Furthermore, the residues around the active site are also highly conserved.

According to the result of multiple alignment with primary and secondary metabolic CS, the residues around the active site are conserved between these two kinds of CS. It means that the protein structures of secondary metabolic CS are likely to be similar to the primary ones. In order to look into the protein structure and the active site information of BfL2, protein modelling was carried out in the next step.

III Modelling of CS

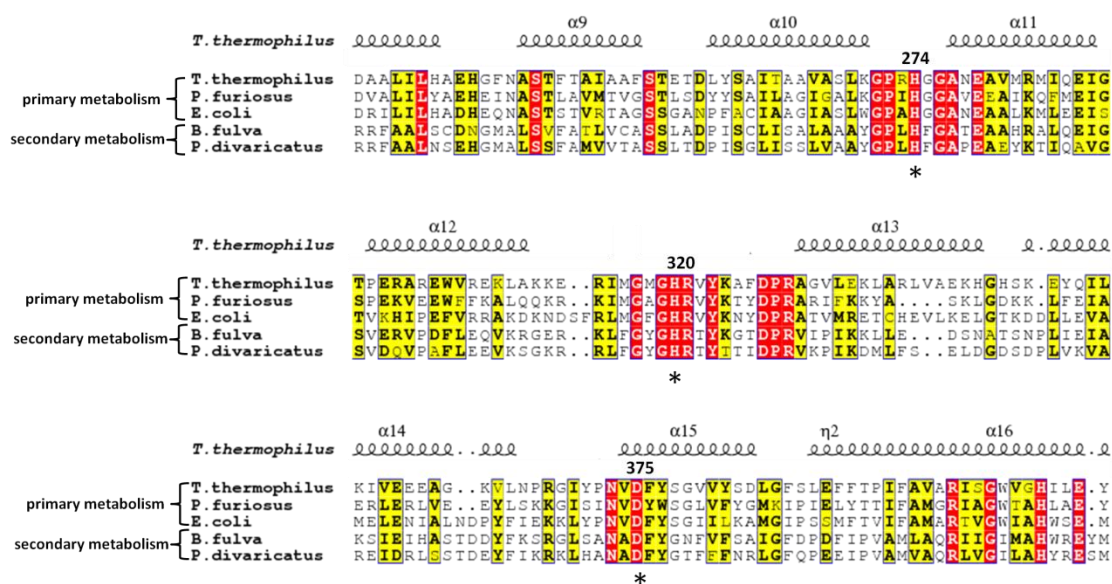


Fig. 3.2.2.7 Multiple alignment of primary metabolic and secondary metabolic CS sequences. Red highlight = 100% similarity; Yellow highlight = more than 60% similarity.

By using online software SWISS-MODEL^[177] and BLASTp, the closest structural model for BfL2 was found. A structure of citrate synthase (PDB ID: 2H12)^[178] complexed with oxaloacetate (OAA) and carboxymethyldethia coenzyme A (CMX) from *Acetobacter aceti* was chosen as the structural model. Taking crystal structure of 2H12 as the template, a BfL2 protein structure model was built. The 2H12 and the BfL2 modelling structure were displayed and analysed on the software *PYMOLO*.^[179] The BfL2 model has a very similar structure as the 2H12 which is a dimer structure. The tertiary structure of the BfL2 (in green) and 2H12 (in blue) were highly similar in general (**Fig. 3.2.2.8 A**). In particular, the active site residues around the bound substrates OAA and CMX were very similarly located between the two structures (**Fig. 3.2.2.8 B**). In addition, the residues around the OAA and CMX within 4 Å were displayed. The result showed that these active site residues were very closely overlapped between the two structures (**Fig. 3.2.2.8 C and D**). The key residues (His-284, His-323 and Asp-377) in BfL2 modelling are also very similarly located with 2H12 structure (**Fig. 3.2.8.2 E**). According to the polar contacts (yellow line) between substrate OAA / CMX and the active site residues, substrates are very likely to be bound in identical orientations. In particular, activation of oxaloacetate is likely to be identical in both cases. So the stereochemistry of BfL2 is extremely likely to be the same as the primary CS 2H12, resulting in an *S*-centre at the newly formed alcohol.

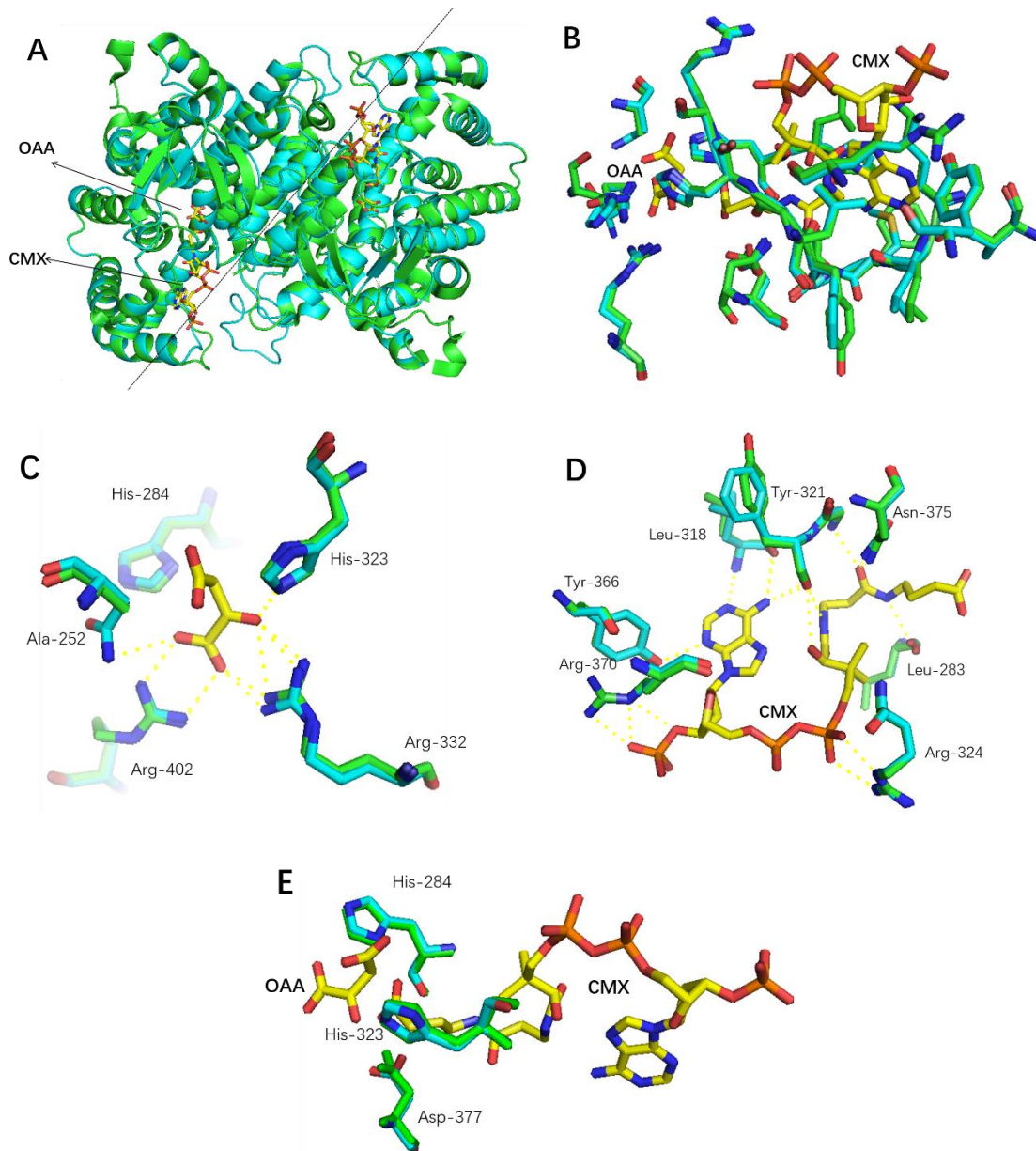


Fig. 3.2.8.2 Modelling of BfL2 (green) in byssochlamic acid pathway and structure of 2H12: **A**, alignment of BfL2 and 2H12 structures; **B**, the residues around the OAA and CMX within 4 Å; **C**, residues around the OAA; **D**, residues around the CMX; **E**, key residues.

To sum up, the catalytic residues are conserved in the protein sequence and are closely overlapped in the structure of both primary metabolic and secondary metabolic CS. This suggests that the two kinds of CS should have the same stereoselectivity. According to the enzymology of 2H12, it catalyses the formation of intermediate *2S*-citryl CoA. Dr. Steffen Friedrich showed that the BfL2 catalyses the formation of *2R*, *3S* or *2S*, *3R* citrate product (section 3.2.1.1a). If BfL2 has the same enzymology as primary CS, it should therefore synthesise the (*2S*, *3R*) product.

3.2.2.2 *In vitro* assay of 2-Methyl citrate dehydratase

BfL3 and PvL2 are 2-methyl citrate dehydratase-like enzymes involved in the biosynthetic pathway of byssochlamic acid **1** and cornexstin **66**. The substrate of both BfL3 and PvL2 was shown to be alkyl citrate **108a** as described in section 3.2.2.1. In this section, the *in vitro* assay of BfL3 and PvL2 was carried out, and the stereoselectivity of these two enzymes were explored.

As described in section 3.2.1, by using yeast expression, 2-methylcitrate dehydratase PvL2 and BfL3 were expressed. The mixture of isomers of **108** was then used as the substrate for *in vitro* assay of the 2MDH proteins. Citrate mixture **108** 1 mM was incubated with cell free extract of the yeast strain expressing PvL2 and BfL3 at 30 °C for 2 hours in 100 μ L PBS buffer. The reaction was stopped by addition of 100 μ L methanol (section 5.3.5). Protein was removed by centrifugation and the supernatant was analysed directly by LCMS.

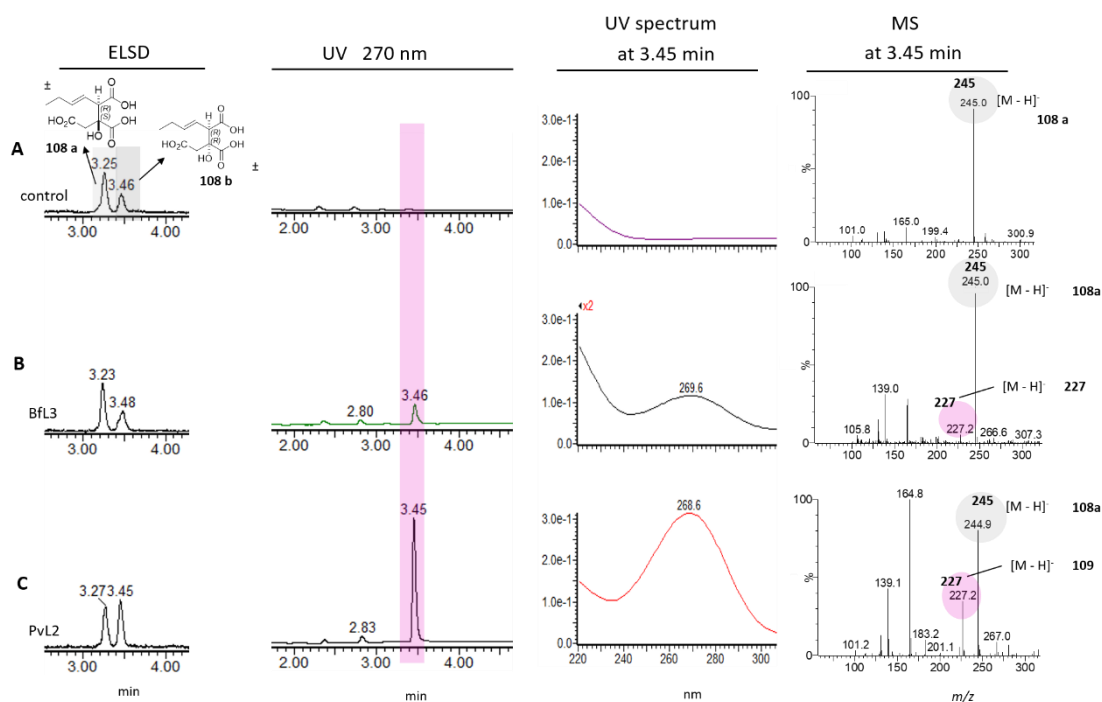
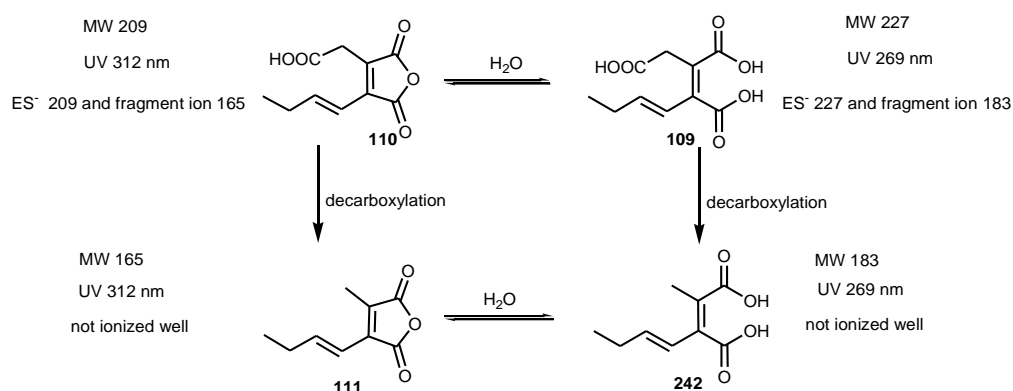


Fig 3.2.2.9 LCMS analysis of PvL2 and BfL3 reactions *in vitro*.

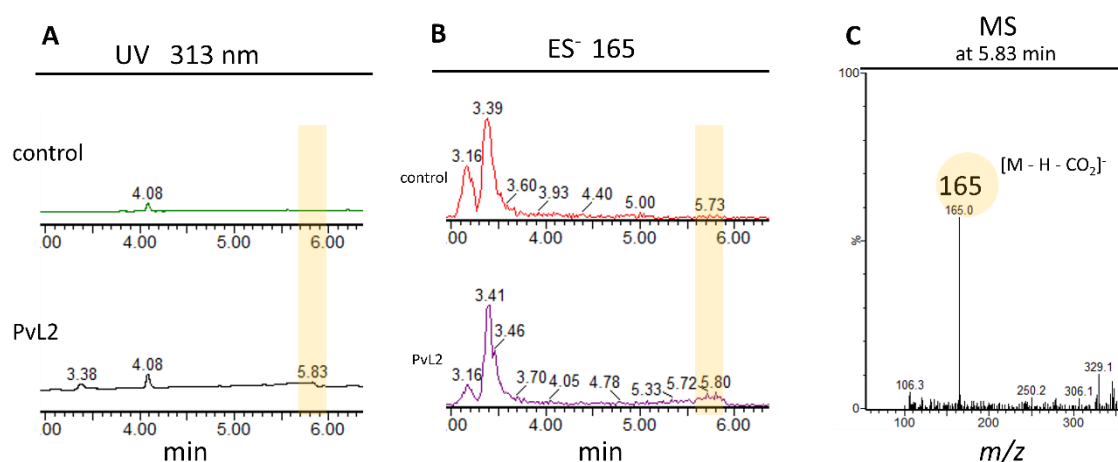
Both PvL2 and BfL3 showed activity on the mixture of citrate diastereomers (Fig. 3.2.2.9). From the ELSD detector, it was obvious that the peak area at 3.5 min (from PvL2 reaction) increased compared with the empty vector control (Fig. 3.2.2.9 A/C), whereas, BfL3 didn't show much difference with the control (Fig. 3.2.2.9 B). But when it comes to the UV detector, both PvL2 and BfL3 formed a new peak with λ_{\max} 269 nm (Fig. 3.2.2.9 B and C). The empty vector control had no UV absorption in the same position (Fig. 3.2.2.9 A).

From the previous study of Dr. Agnieszka Szwalbe,^[29] the maleic anhydride monomer **110** (Scheme 3.2.2.5) is known to decarboxylate spontaneously.^[2] At the same time, **110** and its decarboxylated form **111** also equilibrate with ring-opened forms of **109** and **242**, both of them had similar UV spectra ($\lambda_{\text{max}} = 269 \text{ nm}$) and the expected masses (m/z 227.5 and 183.5 $[\text{M} + \text{H}_2\text{O} - \text{H}]^-$ respectively). The features of the compound **109** were consistent with the new compound synthesised from the PvL2 and BfL3 *in vitro* reactions (Fig. 3.2.6.5 B and C). However, this new product co-elutes with the minor *syn* diastereomer of **108** at 3.45 min.



Scheme 3.2.2.5 Decarboxylation and ring opening of maleic anhydride monomer.

In addition, when we observed the extracted UV chromatogram at 313 nm (the maximum absorption wavelength of maleic anhydride monomer), a small peak showed at 5.83 min (Fig. 3.2.2.10 A). The ES⁻ of this peak was 165 which corresponds to the fragment $[\text{M} - \text{COOH}]^-$ from maleic anhydride monomer **110** (Fig. 3.2.2.10 B and C). However, the empty vector control had no peak in this region. The UV spectrum of this peak was too weak to detect.



Scheme 3.2.2.10 Chromatography of PvL2 reaction: **A**, under UV 313nm; **B**, ES⁻ 165; **C**, MS spectrum of the 5.8 min peak.

In summary, by using yeast expression cell-free extraction of PvL2 and BfL3 and a mixture of citrate isomers to do the *in vitro* reaction, one new compound with UV spectrum ($\lambda_{\max} = 269$ nm) formed. This compound corresponds to the maleic anhydride monomer ring-opened form **109**. In addition, a very weak peak in PvL2 reaction was detected which is thought likely to be maleic anhydride monomer **110**. The LCMS result of the *in vitro* assay on PvL2 and BfL3 indicated that the enzyme probably accepts the isomer **108a** ($2R, 3S / 2S, 3R$) as its substrate. To make sure which isomer the citrate dehydratase uses as its substrate, the single isomer should be used in the reaction.

3.2.2.3 Stereoselectivity of 2-methyl citrate dehydratase

Initial *in vitro* assays showed that both PvL2 and BfL3 were active, but the results did not reveal the stereoselectivity of the enzymes. Therefore, a purification method (prep-column and prep-programming see section 5.5.3) was set up for separating the isomeric **108a anti** ($2R, 3S / 2S, 3R$) and **108b syn** ($2S, 3S / 2R, 3R$) diastereomers by prep-LCMS (Fig 3.2.2.11, section 5.3.3). Both **108a** ($2R, 3S / 2S, 3R$) and **108b** ($2S, 3S / 2R, 3R$) had $[M - H]^-$ 245 (Fig. 3.2.2.11 A), and these two compounds were well-separated on a C_{18} reverse-phase prep-column. The two citrate isomers do not have UV absorption (Fig. 3.2.2.11 C) but can be observed by ELSD (Fig. 3.2.2.11 B).

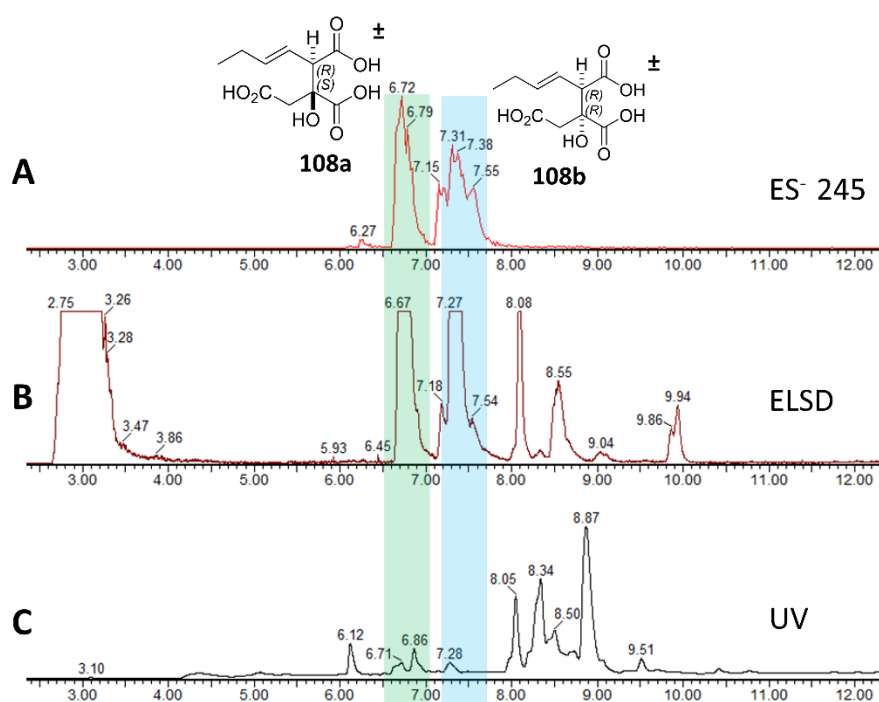


Fig. 3.2.2.11 LCMS purification result of citrate substrates, and the shaded peaks were purified: A, ES⁻ 245 ion extraction; B, ELSD; C, UV chromatogram.

The pure racemic isomers of **108a** (*2R, 3S* / *2S, 3R*) and **108b** (*2S, 3S* / *2R, 3R*) were used for the DH assays *in vitro*. After 3 hours of incubation with substrate and enzymes (PvL2 and BfL3), the *anti*-isomer of **108a** (*2R, 3S* / *2S, 3R*) was clearly converted to a new product (the pink shade **Fig 3.2.2.12 A and B**) compared with the control (**Fig 3.2.2.12 C**). However, *syn* diastereomer **108b** (*2S, 3S* / *2R, 3R*) had no reaction (**Fig 3.2.2.12 D, E and F**). Both PvL2 and BfL3 were selective for the **108a** (*2R, 3S* / *2S, 3R*) *anti* diastereomer.

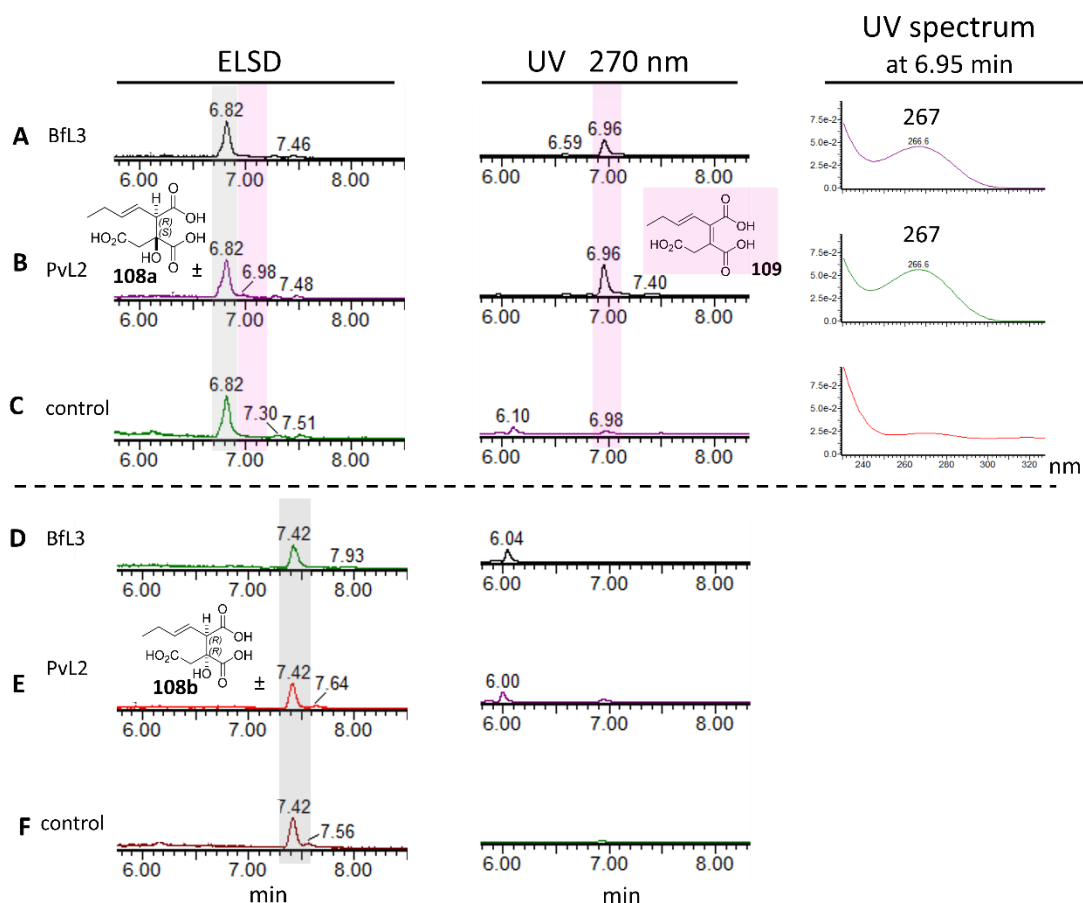


Fig 3.2.2.12 LCMS analysis of the DH reactions: **A, B and C** treated with (*2R, 3S* / *2S, 3R*) *anti* substrate using enzyme BfL3, PvL2 and empty vector control respectively. **D, E and F** treated with (*2R, 3R* / *2S, 3S*) *syn* substrate using enzyme BfL3, PvL2 and empty vector control respectively.

The citrate substrate **108** has no conjugated double bond and therefore no significant UV absorption. The product **109** has a conjugated system, so the maximum absorption wavelength has a blue shift to 267 nm (**Fig. 3.2.2.13 B**). The molecular weight changed from [M - H]⁻ 245 to [M - H]⁻ 227. The fragment ion 165 [M - H₂O - COOH]⁻ is due to decarboxylation (**Fig. 3.2.2.13 A**).

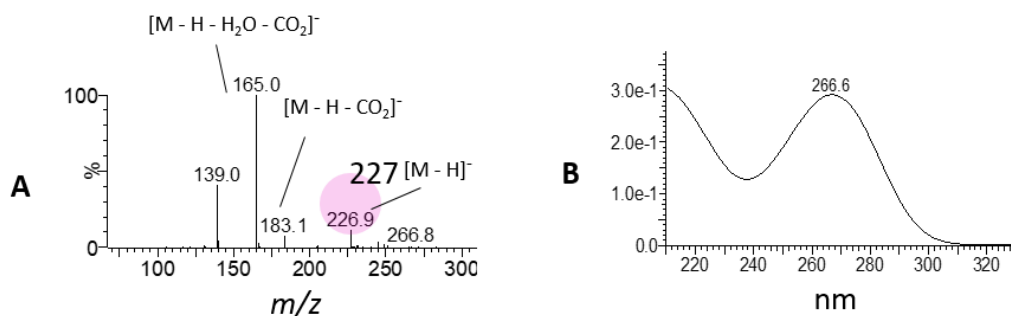


Fig. 3.2.2.13 UV and MS spectrum of product: A, UV spectrum; B, MS spectrum.

The ES⁻ extracted ion chromatogram (EIC) at m/z 245 and 227 were examined, the retention time of *anti* citrate substrate **108a** (2*R*, 3*S* / 2*S*, 3*R*, grey shaded) was at 6.8 min (Fig. 3.2.2.14 A and C), while the double bond product **109** (pink shaded) was at 7.0 min (Fig. 3.2.2.14 D). In this chromatographic method, *anti* citrate substrate **108a** (2*R*, 3*S* / 2*S*, 3*R*) had about 0.1 min time shift comparing with product **109**. As for the reaction of empty vector control, there was no product peak shown on chromatogram (Fig. 3.2.2.14 B).

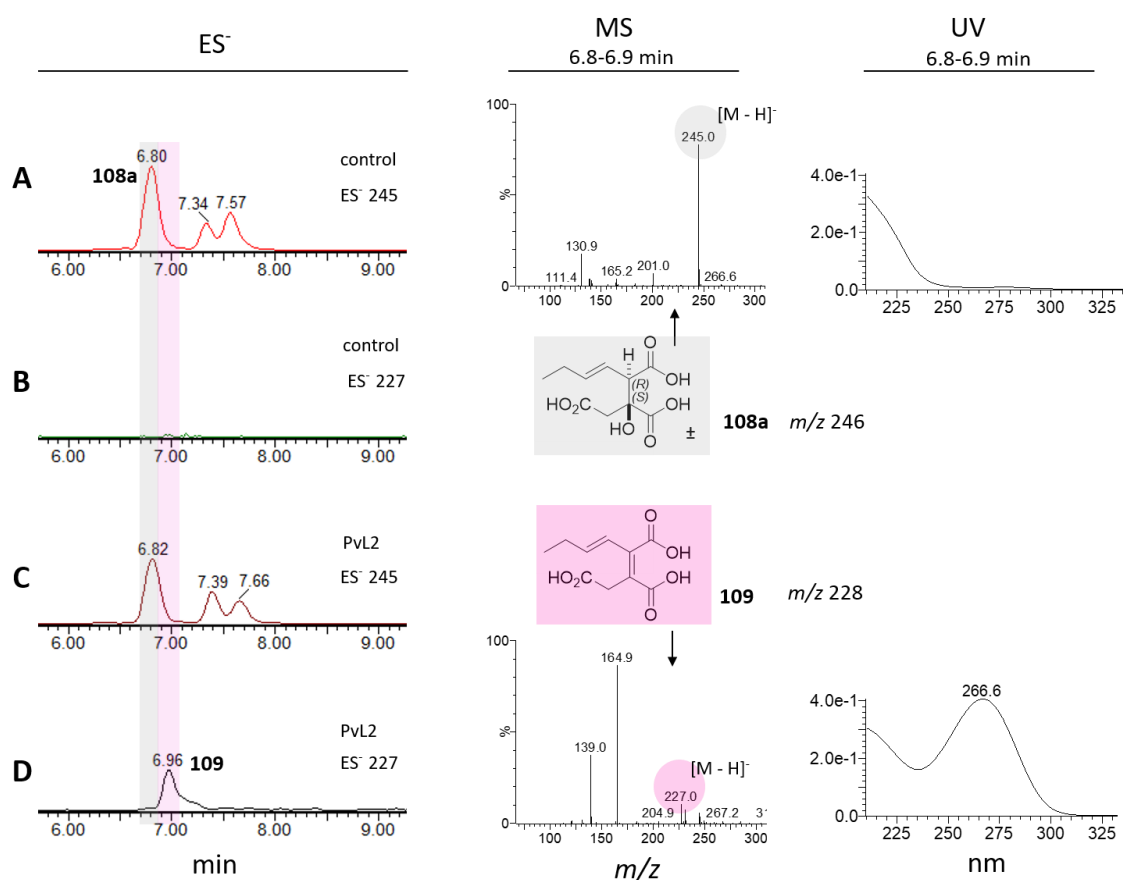


Fig. 3.2.2.14 ES⁻ spectrum of substrate and product: A and B, empty vector control in ES⁻ 245 and 227; C and D, PvL2 reaction in ES⁻ 245 and 227.

It was interesting that when we observed the extracted UV chromatogram at 313 nm (the maximum absorption wavelength of maleic anhydride monomer), a small peak showed at 9.7 min (**Fig. 3.2.2.15 B**). The ES⁻ of this peak was 165 which corresponds to the fragment [M-COOH]⁻ from maleic anhydride monomer **110** (**Fig. 3.2.2.15 B**). The UV spectrum of this peak had a maximum absorption at 312 nm which was consistent with anhydride monomer **110** (**Fig 3.2.7.5 B**). However, the empty vector control had no peak in this region (**Fig. 3.2.2.15 A**). It means that the DH enzyme can dehydrate citrate and at the same time the product **109** can cyclise by itself in the reaction conditions. Finally, we got a mixture of ring-opened and ring closed products.

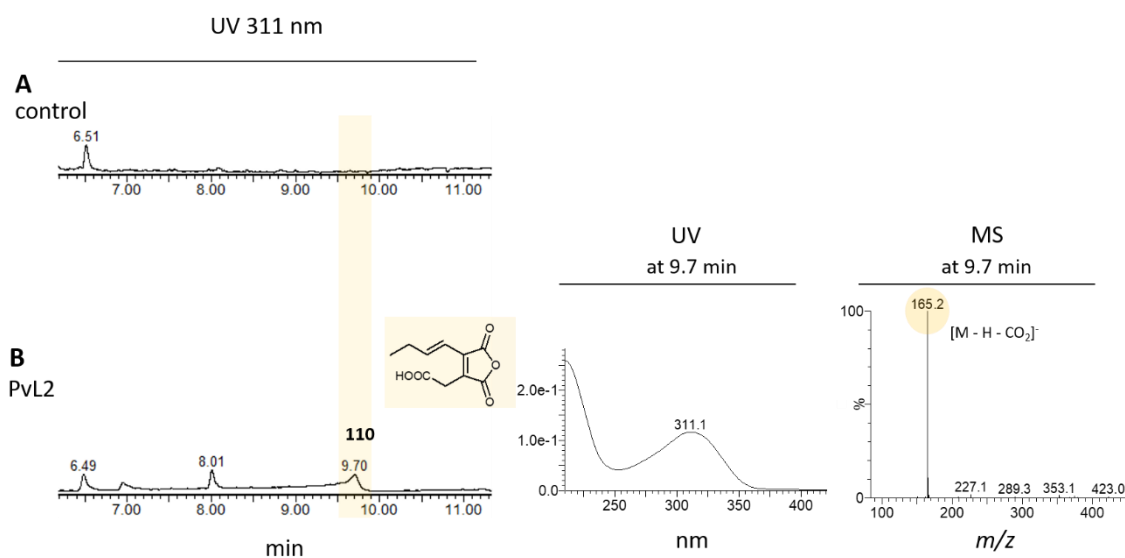
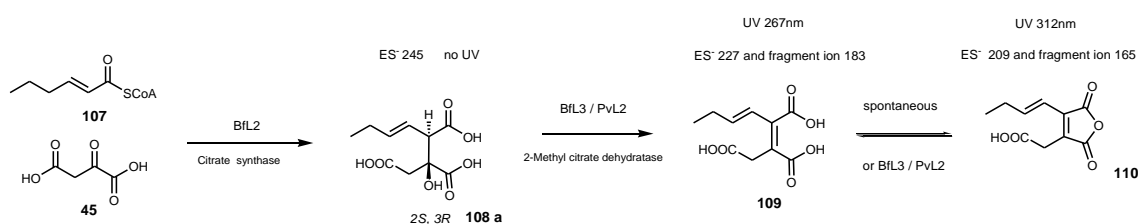


Fig. 3.2.2.15 Chromatography of PVL2 and empty vector control reaction under UV 313nm, and MS and UV spectrum of the maleic anhydride monomer: **A**, control; **B**, PVL2.

To sum up, at this point, it has been determined that the 2-methyl citrate dehydratase can only accept the **108a** *anti* diastereomer (*2R*, *3S* / *2S*, *3R*) as its substrate to form the product **109**. Compound **109** was a ring-opened form of maleic anhydride monomer, at the same time, a little amount of cyclised anhydride product **110** can also be found in the reaction mixture. It means product **109** and anhydride monomer **110** had a reversible reaction. Our previous analysis of the stereoselectivity of the citrate synthase (section **3.2.2.1**) strongly suggested formation of the *2S* enantiomers. This therefore strongly suggests that the true substrate for 2MCDH is the (*2S*, *3R*) isomer (**Scheme 3.2.2.6**).

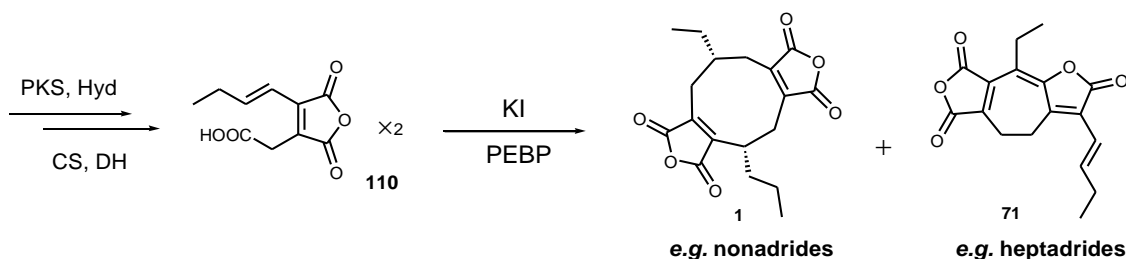
III Maleic anhydride monomer substrate



Scheme 3.2.2.6 A summary of citrate synthase and dehydratase reactions.

3.2.2.4 *In vitro* study of dimerization step

Our group has carried out the *in vivo* study on the biosynthesis of byssochlamic acid **1** through gene disruption and heterologous expression (see section 1.5.2).^[3] The result indicated that the enigmatic ring formation appears to be catalysed by two proteins (BfL6 and BfL10) with homology to ketosteroid isomerases (KSI), and assisted by two proteins (BfL5 and BfL9) with homology to phosphatidyl ethanolamine-binding proteins (PEBP). The substrate of KSI-like and PEBP enzymes is thought to be the maleic anhydride monomer **110**. KSI-like and PEBP-like enzymes catalyse the dimerization of **110** to form nonadrin **1** and heptadrin **71** (**Scheme 3.2.2.7**).



Scheme 3.2.2.7 Proposed dimerization reaction based on KI and PEBP enzymes

The maleic anhydride monomer **110** is the most important substrate for the *in vitro* assay. Our group member Dr. Steffen Friedrich worked on the synthesis of **110**. But it was very difficult to obtain **110** by chemical synthesis due to the instability of this compound. On the other hand, it could be possible to purify **110** from extraction of byssochlamic acid intermediate producing strains of *A. oryzae* (from Dr. K. Williams). In this section, KSI-like genes (*bfl6* and *10*) and PEBP-like genes (*bfl5* and *9*) are expressed in *E. coli* and yeast. The enzymes were then used for *in vitro* assays with the substrate **110**.

3.2.2.5 Regrowing the byssochlamic acid biosynthetic pathway genes *A. oryzae* strains

To obtain maleic anhydride monomer **110** and get better knowledge and information about the biosynthesis of byssochlamic acid **1** and the dimerization step, the *A. oryzae* transformants which

contain core genes for monomer and KSI / PEBP genes inside were regrown (transformants are from Dr. K. Williams). In particular this included the genes PKS, hyd (*bfL1*), CS (*bfL2*), 2MCDH (*bfL3*) plus KSIs (*bfL6* and *10*) and PEBPs (*bfL5* and *9*).

After 7 days' fermentation (28 °C and 130 rpm) in CMP medium, compounds of each transformant were extracted and analysed by LCMS (**Fig. 3.2.2.16**). The transformant with two KSI alone and the transformant with two KI/PEBP co-expressed with core genes for monomer produce the expected final compound byssochlamic acid **1** (**Fig. 3.2.2.16 A and B** red shaded) and agnestadride A **70** (**Fig. 3.2.2.16 A and B** green shaded). Both compound **1** and **70** have a molecular ion of $[M - H]^-$ 331 (**Fig. 3.2.2.16 C and D**). However, the *A. oryzae* NSAR1 wild type and the transformant with only core genes didn't produce any dimerised products (**Fig. 3.2.2.16 C and D**). Furthermore, from experiments **A**, **B** and **C** (**Fig. 3.2.2.16 A/ B/ C**), there was almost no detectable maleic anhydride monomer **109** observed in the chromatograms, and it might be because that the monomer was unstable after a long fermentation time.

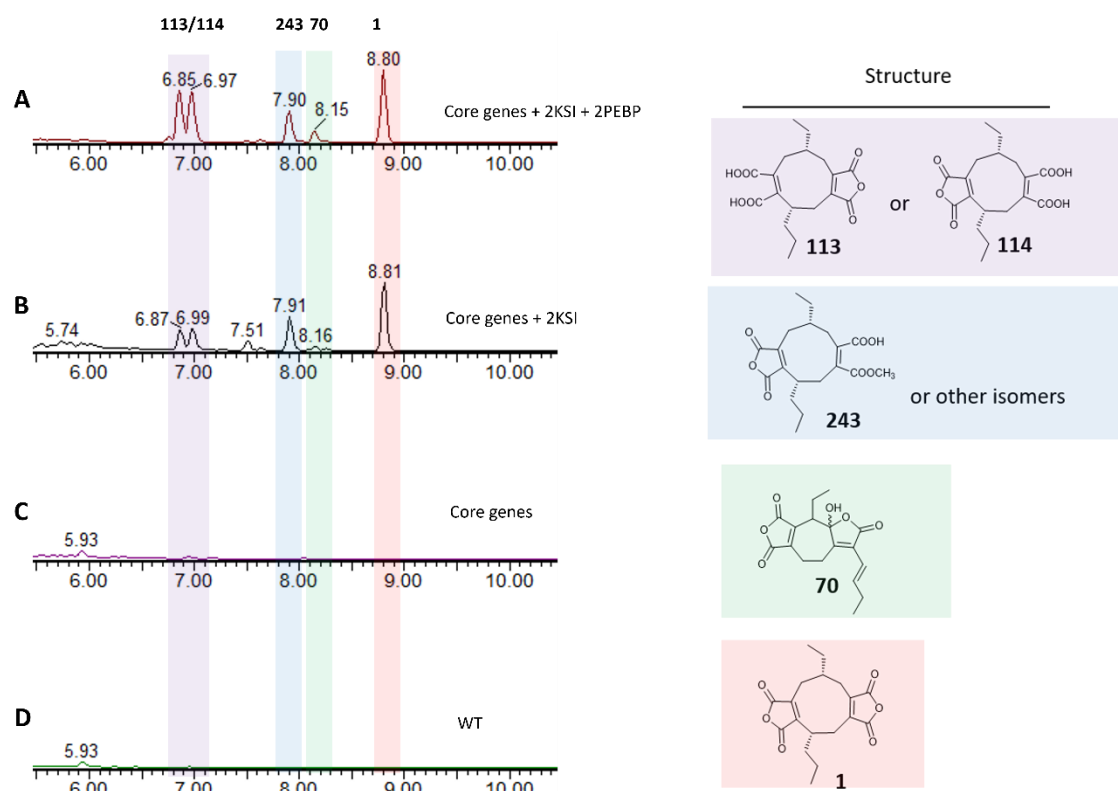


Fig 3.2.2.16 Chromatography of *A. oryzae* transformants with genes from byssochlamic acid pathway: **A**, core genes with 2 KSI and 2 PEBP; **B**, core genes with 2 KSI; **C**, core genes; **D**, wild type. All these chromatograms show at the same scale.

Both experiments **A** and **B** produced the cometabolites which corresponded to the ring-opened form of byssochlamic acid **1** (**Fig. 3.2.2.17 A and B** purple shaded). Compounds **113** and **114** with the $[M - H]^-$ 349 (**Fig. 3.2.2.17 A**) have been previously identified and characterized.^[98]

III Maleic anhydride monomer substrate

Compounds with the retention time 7.9 min (blue shaded) is most likely compound **243** or one of its isomers (**Fig. 3.2.2.17 B**) which is a monomethylated ester from byssochlamic acid **1**. This was because of the addition of methanol to dissolve the extraction mixture. The previous study of our group indicated that the closed anhydride form is more stable in an organic solvent like acetonitrile. However, byssochlamic acid **1** tends to ring-open in water or methanol (**Scheme 3.2.2.8**).

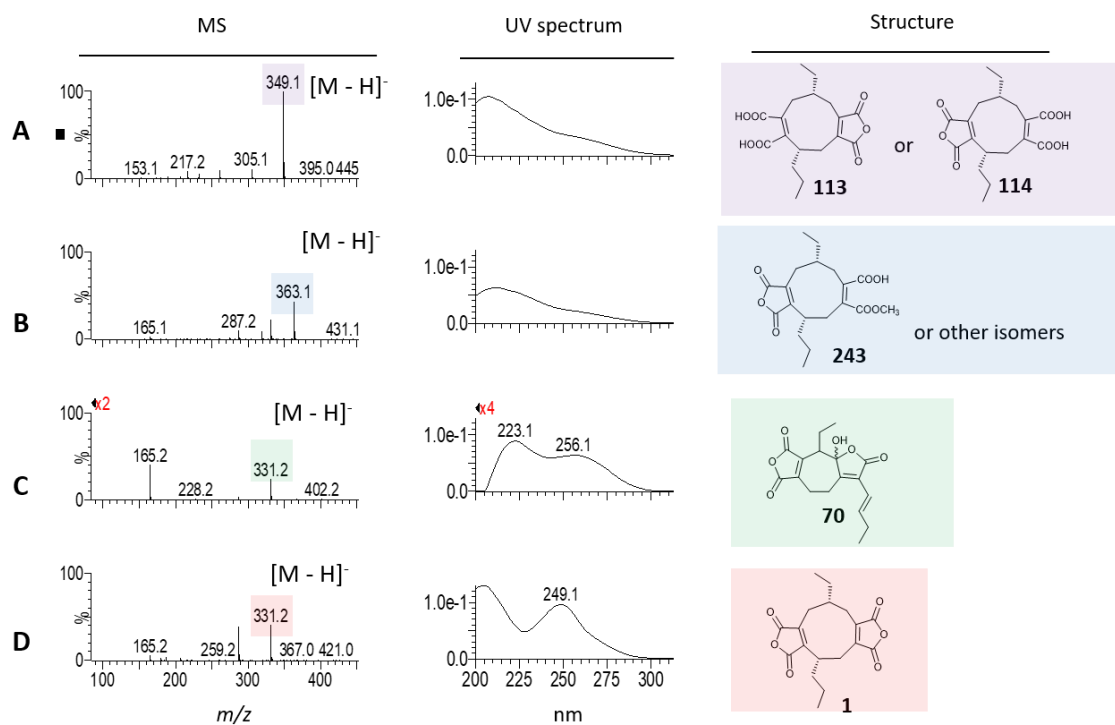
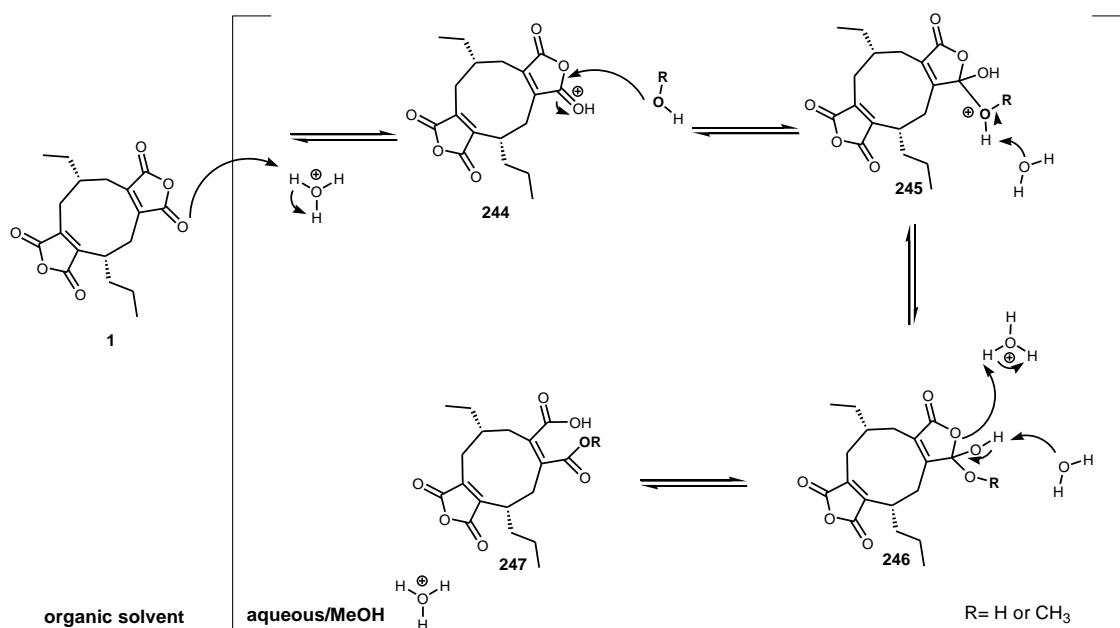


Fig 3.2.2.17 MS spectrum, UV spectrum and structures of the products: **A**, ring opened byssochlamic acid **113/114**; **B**, methylated byssochlamic acid **243**; **C**, agnestadride A **70**; **D**, byssochlamic acid **1**.

In summary, KSI should be the most important enzyme for the dimerization step in the biosynthesis of byssochlamic acid **1**. The production of the dimerised compounds slightly increased when two PEBP enzymes were involved. However, from the previously published data, the PEBP-like enzyme take a very important role to increase production of the final compounds. However, in my case, the two KI transformants and the two KI with two PEBP transformants didn't show significant difference in the titer of the dimerised compounds, showing that the PEBP is not essential.



Scheme 3.2.2.8 Proposed mechanism of the acid-catalysed anhydride opening reaction in byssochlamic acid **1** (R = H in reaction with water; R = CH₃ in reaction with methanol).

3.2.2.6 Preparing the decarboxylated maleic anhydride monomer substrate and in vitro test

The maleic anhydride **110**, the product of the DH reaction, was proposed to be the substrate for the KSI-like enzymes, so the *A. oryzae* transformant containing all of the byssochlamic acid **1** biosynthesis genes was used for monomer production and purification. However, the experiment in section 3.2.10 showed that there was no anhydride monomer produced after seven days' fermentation. A time-course experiment was therefore carried out to test when it would be the best time to get enough anhydride monomer for purification. The result showed that monomer **110** was produced after three days and increased in the next days (**Fig. 3.2.2.18 B and C**), but it disappeared after seven days' fermentation (**Fig. 3.2.2.18 A**).

A large scale of fermentation (2 L, 4 days) was then used for purification. After the prep LCMS purification of compound **110**, and freeze-drying, 1.5 mg of the pure compound was dissolved in DMSO and diluted in acetonitrile. The LCMS analysis showed that maleic anhydride **110** seemed to decarboxylate to **111**. Compound **110** had retention time at 5.7 min, however, the new peak **111** shifted to 7.4 min. In addition, although it had a similar UV spectrum to **110**, the newly formed compound was not ionized well. This is in agreement with the previous observations of Dr. Agnieszka Szwalbe from the Cox group.^[56] There was almost no compound **110** left (**Fig. 3.2.2.19**).

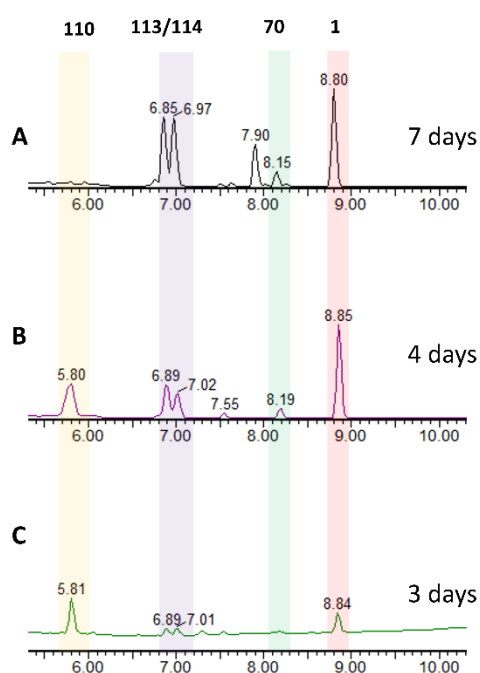


Fig. 3.2.2.18 Time course test for maleic anhydride monomer producing.

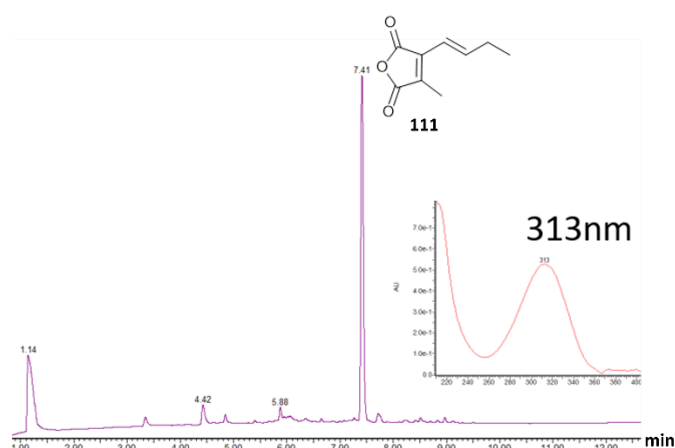
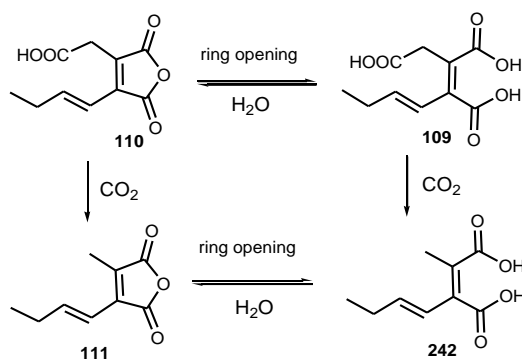


Fig. 3.2.2.19 LCMS analysis of the purified maleic anhydride monomer.

Even though the maleic anhydride monomer substrate decarboxylated, an *in vitro* test was still performed by using the compound **111** as a substrate. BfL6 (KI1) and BfL10 (KI2) from both yeast and *E. coli* expression (cell free extraction) were used in this assay. Empty vector transformants were used as controls. After three hours of incubation, the reaction was stopped by addition of 100 μ L of CH₃CN. Protein was precipitated by centrifugation and the reaction mixture was analysed by LCMS. No dimerization was observed (Fig. 3.2.2.20 A and C), however, the compound **228** was observed. This arises because of a ring-opening reaction (Scheme 3.2.2.9). The compound at retention time 7.6 min was converted to the compound at 4.5 min. And according to the UV spectrum of compound 4.5 min, it was the ring-opened form of **228**. At the

same time, there was still a low concentration of ring-opened aconitate **109** at 3.2 min (**Fig. 3.2.10.3 A/B/C/D**).



Scheme 3.2.2.9 Maleic anhydride ring opening.

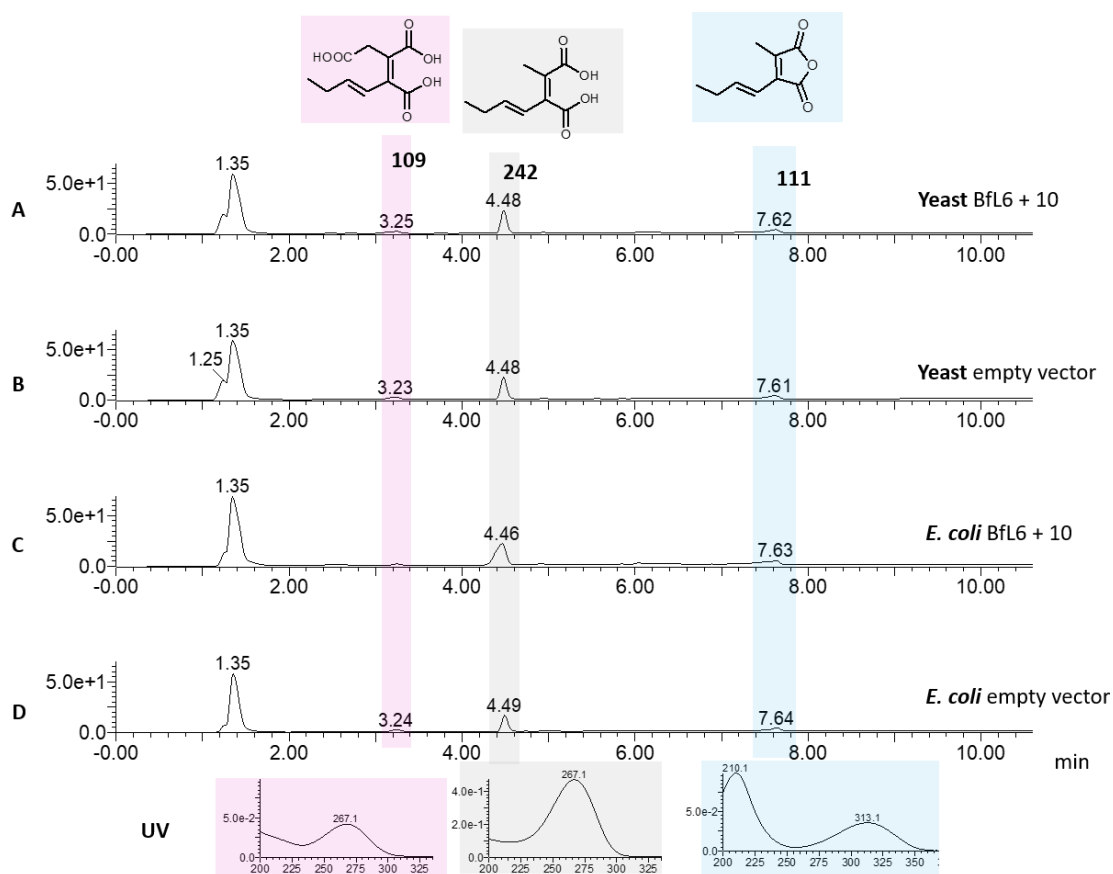


Fig. 3.2.2.20 *In vitro* study on decarboxylated substrate **111**.

In addition, the *in vitro* dimerization test from Baldwin *et al* (section 3.1.1) suggested that the compound **111** could also be treated with DMSO and base (*e.g.* Et₃N). The substrate used by Baldwin was **248** which was one carbon longer than **111**.^[154] In this experiment, 5 mg of compound **111** was incubated with 200 μ L of DMSO and 10 μ L of Et₃N. The reaction system had

III Maleic anhydride monomer substrate

colour changes in 48 h (**Fig. 3.2.2.21**). After 48 h reaction, the sample was analysed by LCMS. The anhydride monomer at 7.4 min had almost disappeared and some new peaks appeared. But these peaks had weak UV absorptions (**Fig. 3.2.2.22 D**). Among these new peaks, ESI MS spectra indicated some putative dimerization products. These compounds gave the $[M + H]^+$ 315 (**Fig. 3.2.2.22 A**), $[M - H]^-$ 349 (**Fig. 3.2.2.22 B**) and $[M + H]^+$ 333 (**Fig. 3.2.2.22 C**). According to the result from Baldwin, these compounds could be the products from different dimerization reactions.

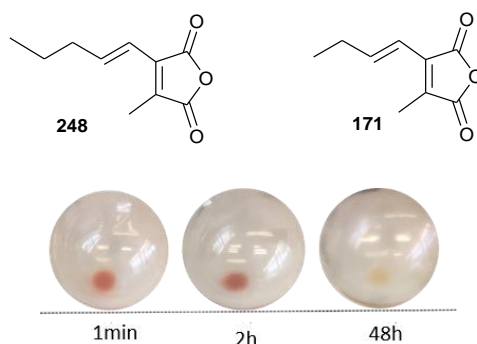


Fig. 3.2.2.21 Colour changes of *in vitro* reaction of the dimerization.

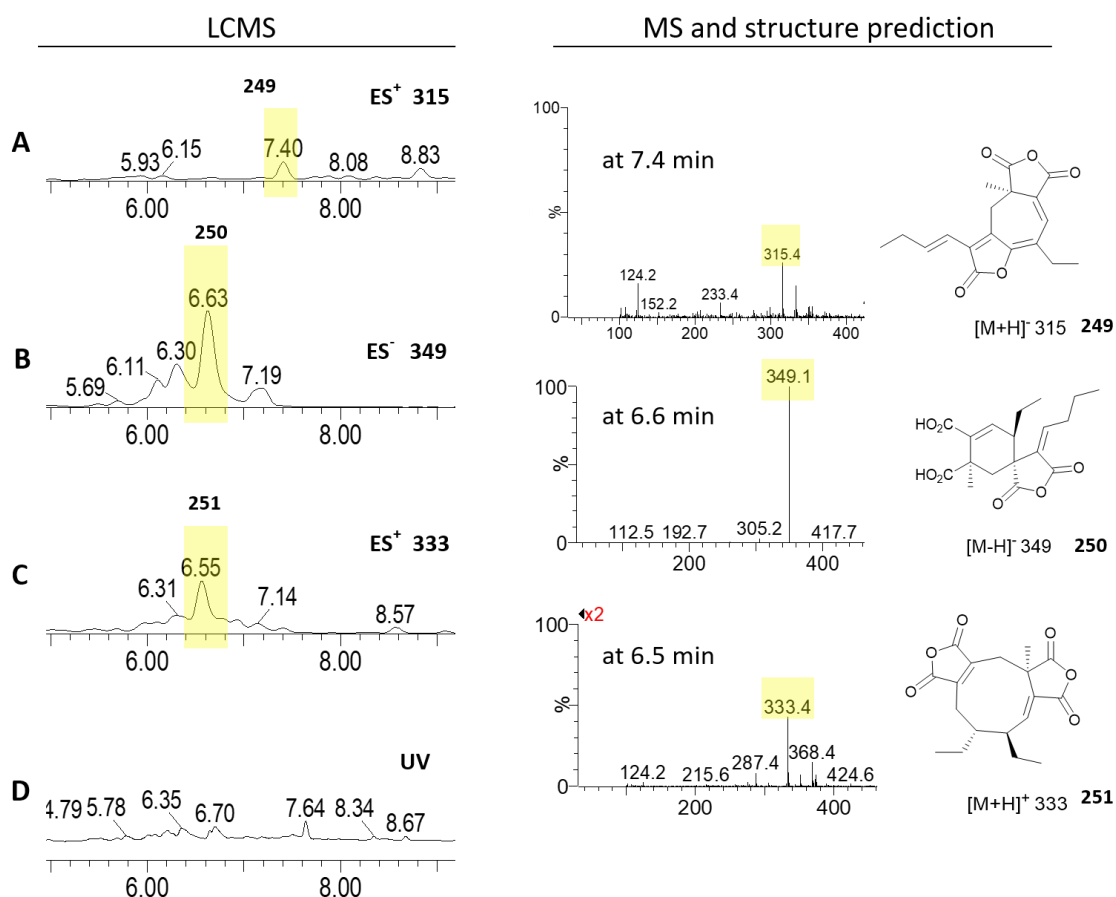


Fig. 3.2.2.22 LCMS analysis of 48h reaction of *in vitro* study.

Our group member Dr. Agnieszka Szwalbe also previously carried out a similar experiment using synthetic sample **110** in her PhD project. The sample was placed in DMSO and colour changes recorded over time. When she analysed the DMSO sample after 48 h, the decarboxylated monomer **111** showed but also some new peaks. And one sample gave a low-resolution molecular ion of m/z 331.7 which could be an isomer of byssochlamic acid **1**. But due to tiny amounts of the sample, the dimerised compound has not yet been purified and identified by NMR.

3.2.2.7 Preparing the maleic anhydride monomer substrate

Because the decarboxylated monomer **111** seemed to be not bioactive, different purification conditions were tested to protect the compound **110** from decarboxylation. Firstly, freeze-drying was found to be one reason for decarboxylation. Every time compound **111** was purified from prep-LCMS and freeze-dried overnight, anhydride **111** would almost totally decarboxylate to **110** (see section 3.2.10.1). Secondly, organic solvent like DMSO and CDCl_3 would also make the compound **110** decarboxylate, and dimerization of the monomers would happen next (see section 3.2.10.1). Thirdly, nucleophilic solvents such as MeOH would make the maleic anhydride monomer ring-open and esterify (see section 3.2.9).

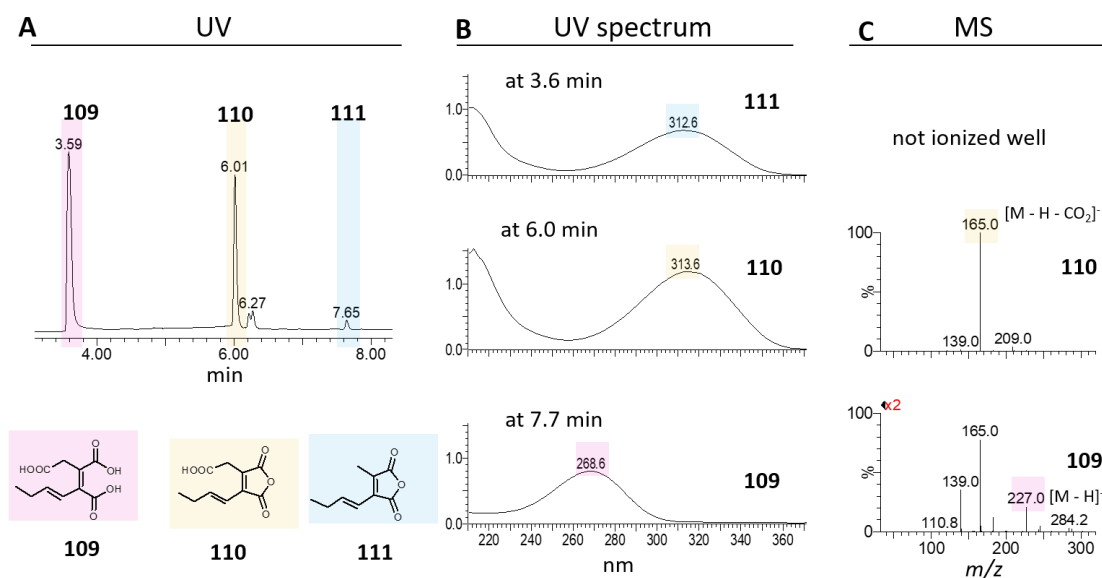


Fig. 3.2.23 LCMS analysis of maleic anhydride monomers: **A**, chromatogram; **B**, UV spectrum; **C**, MS.

The best way to prevent the decarboxylation and ring-opening problems was to purify compound **110** by LCMS. Under these conditions **110** was obtained pure in a mixture of H₂O and CH₃CN. These solvents were then directly removed by rotary evaporation at 5 mbar and 40 °C. After evaporation, 200 µL ddH₂O was added to give a maleic anhydride monomer **110** suspension. The suspension was analysed by LCMS. The compound **110** was still in the sample at 6.0 min. The ring-opened citrate **109** elutes at 3.6 min, and there was some dicarboxylate compound **111** (Fig. 3.2.2.23 A). The UV and MS spectra of these three compounds were identical to the previous study (Fig. 3.2.2.23 B and C). Therefore, this method could be used for making the substrate for the *in vitro* study of KSI and PEBP enzymes.

3.2.2.8 *In vitro* study of yeast expressed KI and PEBP enzymes

In 3.2.10.2, the substrate problem of decarboxylation has been resolved, so KSI and PEBP enzymes can be tested using compound **110**. First of all, the pESC-URA vector with whole ORF of *bfl5/6/9* and *10* were transformed into yeast host *Saccharomyces cerevisiae* w303B, and checked by yeast colony PCR (Fig. 3.2.2.24). Then BfL5-2, BfL6-1, BfL9-2, BfL10-2 were chosen for protein expression.

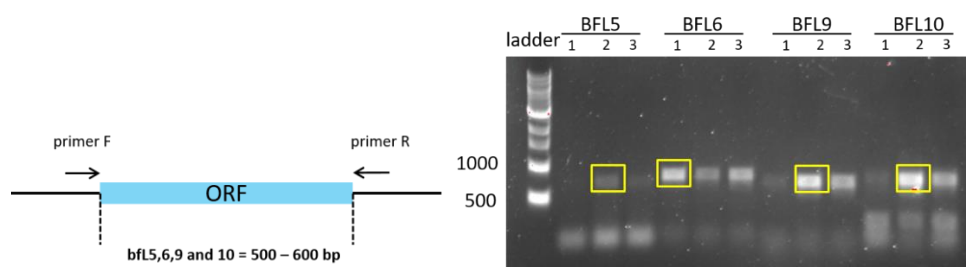


Fig. 3.2.2.24 Yeast colony PCR.

Cell-free extracts of BfL5, 6, 9 and 10 yeast transformants were prepared by using a glass bead vortex method (section 5.6.3). Maleic anhydride monomer suspension **110** (15µL) and each cell extraction BfL5 /6 /9 and 10 (20 µL) then mixed. After 8 hours' reaction at 30 °C, the reaction was stopped by adding 100 µL of acetonitrile. The protein was precipitated by centrifugation and the supernatant was analysed by LCMS.

The result showed that the substrate **110** was converted to three products. These were decarboxylated product (peak **111**) and ring-opened compound (peak **109**). A new peak was observed in the chromatogram (Fig. 3.2.2.25 A). However, the empty vector control did not produce the new peak (Fig. 3.2.2.25 E). Analysis of the MS spectrum and UV spectrum of peak **1** showed that the compound (red shade) was identical with byssochlamic acid **1**. It had molecular ion of *m/z* 331 (Fig. 3.2.2.26) and maximum UV absorption at 248 nm which was the same as

the ES⁻ 331 of LCMS data, it was clearer to see the production of byssochlamic acid **1** product from the KI enzymes (in red shade, **Fig. 3.2.2.25 A/B/C/D and G**)

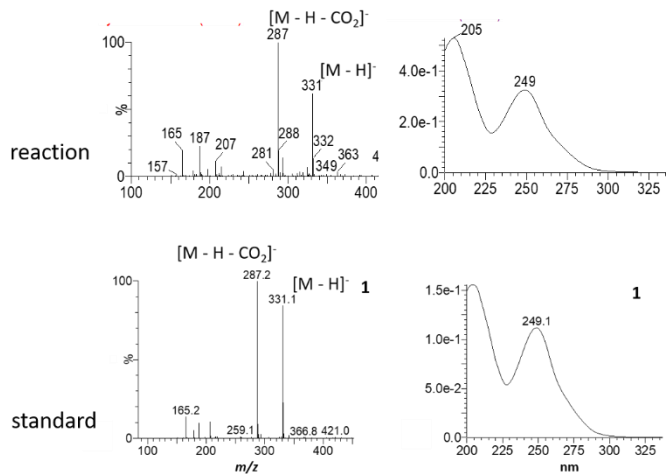


Fig. 3.2.2.26 MS spectrum and UV spectrum of peak **1**

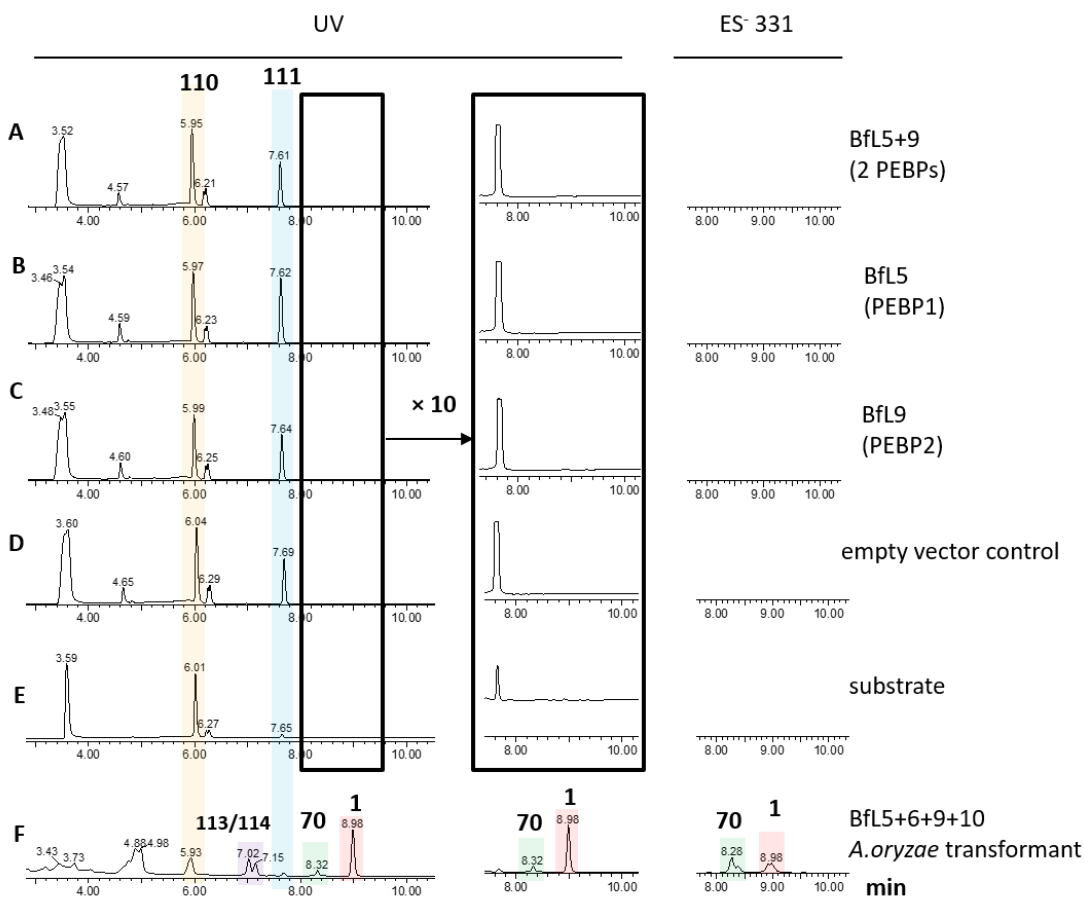


Fig. 3.2.2.27 LCMS analysis the *in vitro* test using yeast expression protein PEBPs alone.

Finally, PEBP enzymes BFL5 and BFL9 were tested in this study to check if they were related to the dimerization reactions. The result indicated that no matter whether both BFL5 and BFL9 (**Fig. 3.2.2.27 A**) were tested together or they worked separately (**Fig. 3.2.2.27 B and C**), there was no byssochlamic acid **1** formation compared with empty vector control (**Fig. 3.2.2.27 D**) and *A. oryzae* transformant (**Fig. 3.2.2.27 F**).

3.2.2.9 *In vitro* study of *E. coli* expressed KI and PEBP enzymes

As it was displayed previously (see section 3.2.2), KI and PEBP enzymes were also expressed in *E. coli*. Due to N terminal signal peptide, it was difficult to express BfL5, BfL6, and BfL10 directly using whole ORF amino acid sequence. The signal peptide sequence was removed by PCR when the expression vector was constructed. To improve the solubility, chaperone proteins were co-expressed with KI and PEBP enzymes. Two KI enzymes were used in this study.

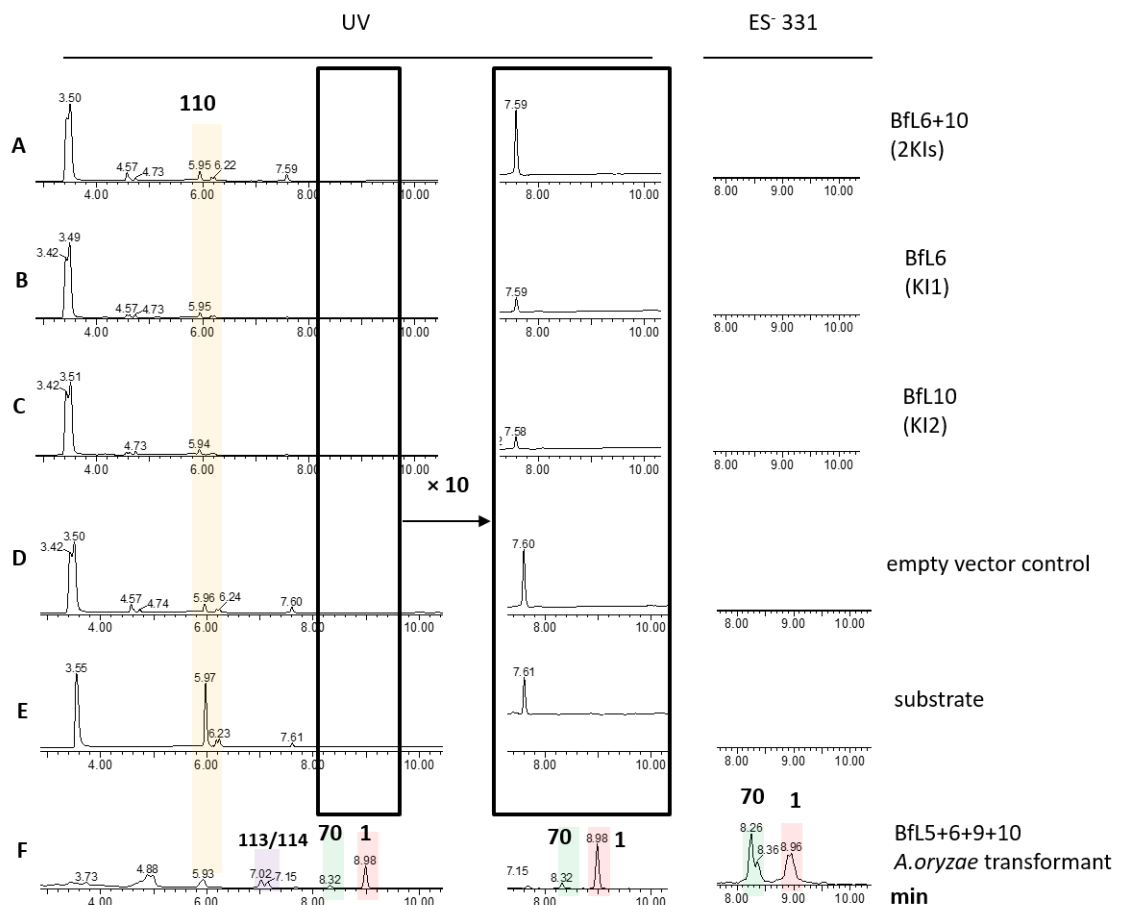


Fig. 3.2.2.28 LCMS analysis the *in vitro* assay using *E. coli* expression protein KSI and PEBP proteins.

The crude *E. coli* cell extractions were used for *in vitro* assay. The experimental groups were set up as BfL6+10 (two KI **Fig. 3.2.2.28 A**), BfL6 (KI1 **Fig. 3.2.2.28 B**), BfL10 (**Fig. 3.2.2.28 C**) and empty vector control (**Fig. 3.2.2.28 D**). After the 8h reaction, reactions were stopped by addition 100 μ L of CH₃CN, then the samples were analysed by LCMS. From ES⁻ analysis, the result indicated that none of this experimental group could give any dimerised product compared with *A. oryzae* transformant extraction (**Fig. 3.2.2.28 F**). It might be because the loss of the N terminal signal peptide region made the proteins not stable or it can affect the proteins to get the right folding.

3.3 Discussion

3.3.1 The signal peptide at the N terminus of the proteins

When the KSI (BfL6 and BfL10) and PEBP (BfL5 and BfL9) enzymes were expressed as complete ORF in *E. coli*, no protein was observed. Bioinformatic analysis suggested that there were signal peptides containing 18 to 20 amino acids at the N-termini of proteins BfL5, 6 and 10. These signal peptides should direct the protein to specific sites within their native organisms.^[180] So these KSI and PEBP enzymes should have protein targeting or protein sorting activities in *B. fulva*.

Protein targeting or protein sorting is the biological mechanism by which proteins are transported to their appropriate destinations in the cell or outside it. Proteins can be targeted to the inner space of an organelle, different intracellular membranes, plasma membrane, or to the exterior of the cell *via* secretion. This delivery process is carried out based on information contained in the protein itself.^[181] A signal peptide (sometimes referred to as signal sequence, targeting signal, localization signal, localization sequence, transit peptide, leader sequence or leader peptide) is a short peptide (usually 16-30 amino acids long) present at the N-terminus of the majority of newly synthesized proteins that are destined towards the secretory pathway.^[180]

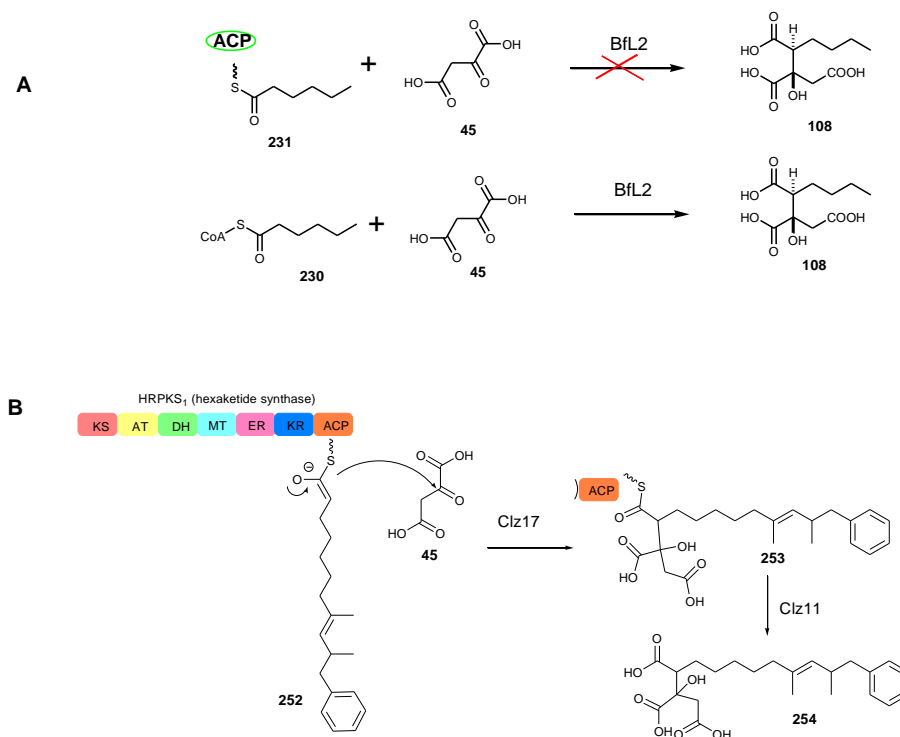
Our *in vitro* study on KI and PEBP enzymes suggested that BfL6 and BfL10 tended to be active at the high concentration of maleic anhydride monomer substrate. So it can be proposed that BfL6 and BfL10 can also be transported to an appropriate destination which contains a high concentration of substrate.

3.3.2 The substrate and stereochemistry of citrate synthase

3.3.2.1 Substrate of CS

As described in section 3.1.1.1, the citrate synthase Bfl2 can use 2-hexan/hexen-oyl-CoA **107/230** as its substrate and catalyses the condensation with OAA. However, *holo*-ACP-2-hexan/hexen-oyl **231/232** are not active for Bfl2 (**Scheme 3.3.2.1 A**). This result shows the CS function does not correspond with the supposition of the Tang group. Tang and co-workers reported the reconstitution studies of **254** in *A. nidulans*. They found that only three enzymes, HRPKS (Clz14), citrate synthase (Clz17) and hydrolase (Clz11) were needed to produce the compound **254**. They inferred that citrate synthase Clz17 was essential in the biosynthesis as removing the gene from *A. nidulans* abolished production. They proposed that the HRPKS synthesized the polyketide, and Clz17 directly took the ACP bound polyketide **252** catalysing the addition to oxaloacetate. Clz11 was then responsible for *directly hydrolysing 253* to yield **254** (**Scheme 3.3.2.1 B**).^[182]

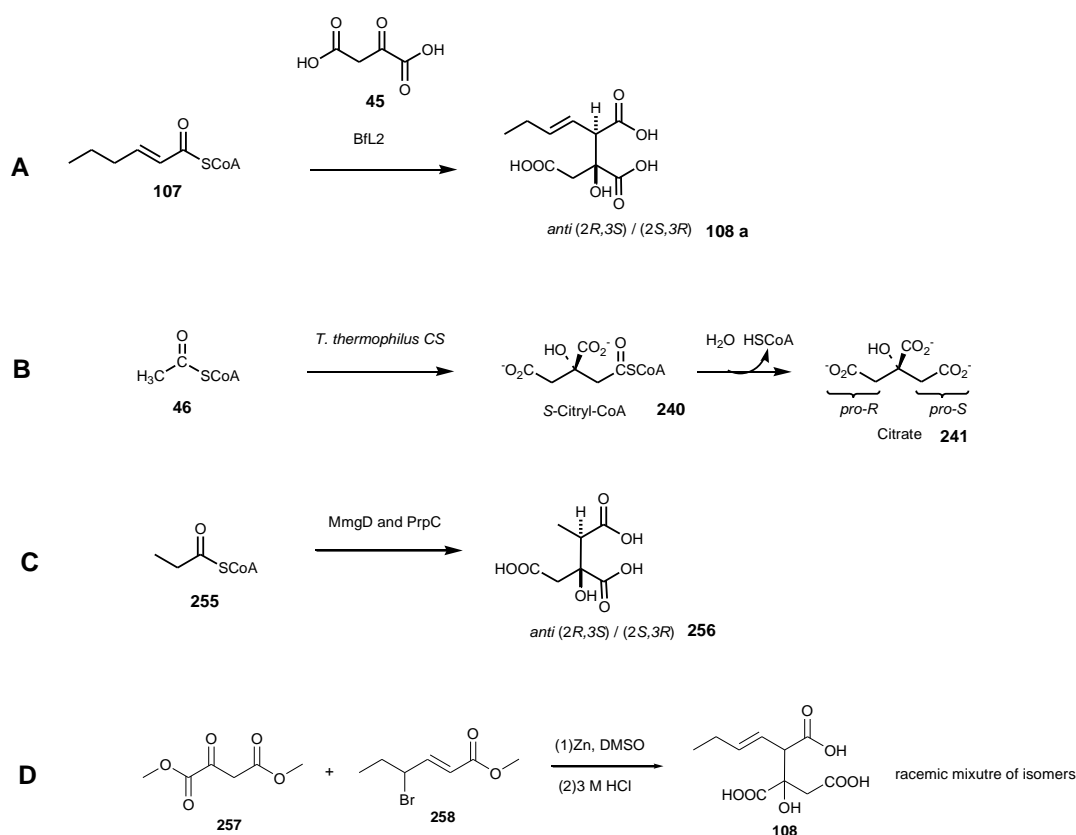
According to our *in vitro* assay on Bfl2, CS uses polyketide CoA as its substrate but not an ACP bound polyketide. Byssochlamic acid BGC does not involve the key enzyme for the formation of polyketide CoA, Therefore, there should be a CoA ligase somewhere in the genome to produce the polyketide CoA substrate.



Scheme 3.3.2.1 Substrate of CS: **A**, Bfl2 reaction with different substrate; **B**, proposed function of the CS and hydrolase in zaragozic acid A pathway.

3.3.2.1 Stereochemistry of CS

According to the LCMS and NMR analysis of BfL2 reaction (section 3.1.1.1), BfL2 is shown to have a stereoselectivity to produce *anti* diastereomers **108a** ($2R, 3S / 2S, 3R$ Scheme 3.3.2.2 A). The primary metabolism CS from *Thermus thermophilus* is known to produce a *S*-citryl CoA **241** (Scheme 3.3.2.2 B).^[129] Two 2-methylcitrate synthases MmgD from *Bacillus subtilis*^[183] and PrpC from *E. coli*^[184] are reported to catalyze the formation of either ($2S, 3R$)- or ($2R, 3S$)- 2-methylcitrate **256** (Scheme 3.3.2.2 C), which is consistent with BfL2. However, different from the enzyme catalysis, chemical synthesis of alkyl citrate can only produce a racemic mixture of diastereomers **108** (Scheme 3.3.2.2 D).



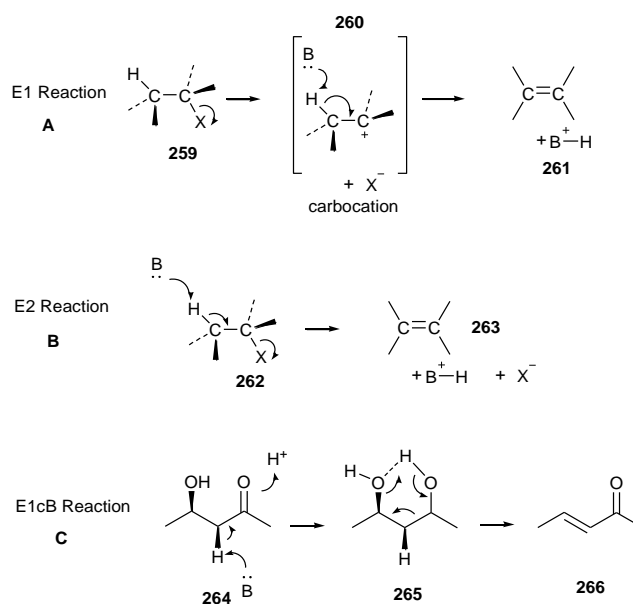
Scheme 3.3.2.2 Stereochemistry of CS and chemical synthesis: **A**, BfL2; **B**, primary metabolic CS; **C**, MmgD; **D**, chemical synthesis.

In this project, an *in silico* protein model of BfL2 was built by using a primary metabolism CS (PDB ID: 2H12) as the template (section 3.1.1.2). Compared with the 2H12 structure, the BfL2 model is highly similar. In particular, the active site residues around the bound substrates are very similarly located between the two structures. Therefore substrates are likely to be bound in identical orientations. The catalytic residues are conserved and closely overlapped in both primary metabolic and secondary metabolic CS. This suggests that the two kinds of CS

are highly likely to have identical stereoselectivity. 2H12 catalyses the formation of 2*S* - citryl CoA. Dr. Steffen Friedrich showed that the BfL2 catalyses the formation of *anti* (2*R*, 3*S* / 2*S*, 3*R*) citrate product (section 3.2.1.1a). If BfL2 has the same enzymology as primary CS, it should obtain a 2*S* product. Because the *in vitro* experiment shows that BfL2 produces *anti* product, it means if C₂ position is *S*, the C₃ position must be *R*. Therefore BfL2 should produce a (2*S*, 3*R*) product. This also corresponds with 2*S* citrates in squalestatin S1 **7** and viridiofungin A **2** which are also installed by CS enzymes.

3.3.3 Stereochemistry of dehydratase

To test the stereoselectivity of the dehydratase, two diastereoisomers produced by synthesis were separated by HPLC and individually tested as substrates. *In vitro* study on dehydratase BfL3 and PvL2 showed that both of these enzymes had stereoselectivity of citrate substrate (section 3.2.2.3). BfL3 and PvL2 could only accept **108a** (2*R*, 3*S* / 2*S*, 3*R*) as substrate rather than (2*R*, 3*R* / 2*S*, 3*S*) **108b** citrate.



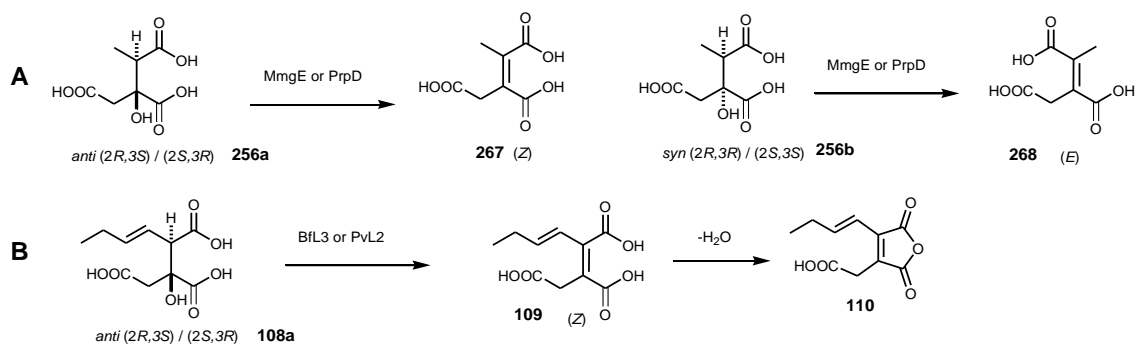
Scheme 3.3.3.1 The three most common mechanisms for elimination: **A**, E1 reaction; **B**, E2 reaction; **C**, E1cB reaction.

The citrate dehydratases PvL2 and BfL3 catalyse elimination reactions, which eliminate H₂O to yield an alkene. There are three common mechanisms of elimination reactions: E1, E2 and E1cB. In the E1 reaction, the C-OH bond breaks first to give a carbocation intermediate **260** (**Scheme 3.3.3.1 A**). However, E2 reaction gives the alkene **263** in a single step without intermediates (**Scheme 3.3.3.1 B**). In the E1cB reaction, the C-H bond breaks first, giving a carbanion or enol intermediate **265** to lose OH⁻ (or H₂O) to form the alkene **266** (**Scheme 3.3.3.1 C**).^[61] Among

three of the reactions, E1cB is the most common type of elimination in biological pathways. An example is the DH domain in the fatty acid synthase catalysis dehydration of the β -hydroxy thiolester to the unsaturated thiolester.^[6] And for this example, the –H and –OH groups are eliminated from the same side of the molecule. So the knowledge of the exact stereochemistry of the substrate can help us to know the mechanism (*i.e. syn*) of the enzyme.

It has been discussed in section 3.3.2 that citrate synthase BfL2 can only produce an *anti* (2*S*, 3*R*) diastereomer **108a**. Dehydratase PvL2 and BfL3 directly take the citrate product **108a** from BfL2 as substrate in the byssochlamic acid pathway. Therefore, PvL2 and BfL3 must eliminate the –H and –OH groups from opposite side (*anti*).

Two 2- methylcitrate dehydratases MmgD from *Bacillus subtilis* and PrpC from *E. coli* are reputed to dehydrate at least two of the four diastereomers of 2-methylcitrate **256** to yield either *E*-2-methyloaconitate **267** or *Z*-2-methyloaconitate **268** (Scheme 3.3.3.2 A).^[171, 183] Either MmgD or PrpC is not a stereospecific enzyme. However, in the case of BfL3 and PvL2, both of these two CMDH show stereospecificity on *anti* (2*R*, 3*S* / 2*S*, 3*R*) diastereomers. According to the BfL2 modelling results, the substrate of BfL3 and PvL2 is more likely to be (2*S*, 3*R*). After the dehydration, the *anti* citrate substrate gives a *Z*-2-hexaconitate **109**, and additional dehydration gives maleic anhydride monomer **110** (Scheme 3.3.3.2 B).



Scheme 3.3.3.3 Stereochemistry of CMDH: A, MmgE and PrpD; B, BfL3 and PvL2.

3.3.4 Maleic anhydride monomer

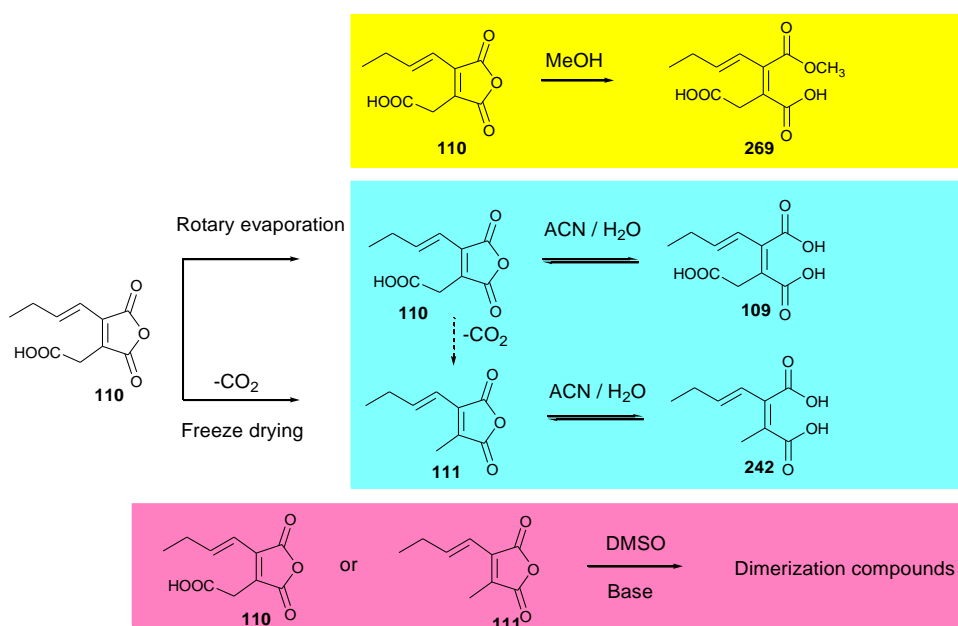
The maleic anhydride **109** is thought to be the basic unit of the maleidrides. The *in vivo* study in *A. oryzae* indicates that introducing PKS, citrate synthase and 2-methylcitrate dehydratase was enough to obtain the maleic anhydride monomer.^[7] However, from the *in vitro* study on citrates and dehydratase, it seemed that we could only get very little cyclised compounds. This

III Stereochemistry of dehydratase

phenomenon might be because of the maleic anhydride monomer with the spontaneous ring opening.
[185]

From Dr. Agnieszka Szwalbe's study, maleic anhydride **110** was chemically synthesized.^[2] However, the route is difficult to follow in our lab, because one reagent is no longer commercially available. To obtain the maleic anhydride substrate, the *A. oryzae* transformant from Dr. Williams containing core genes and two KI and two PEBP genes was grown for monomer purification.^[3]

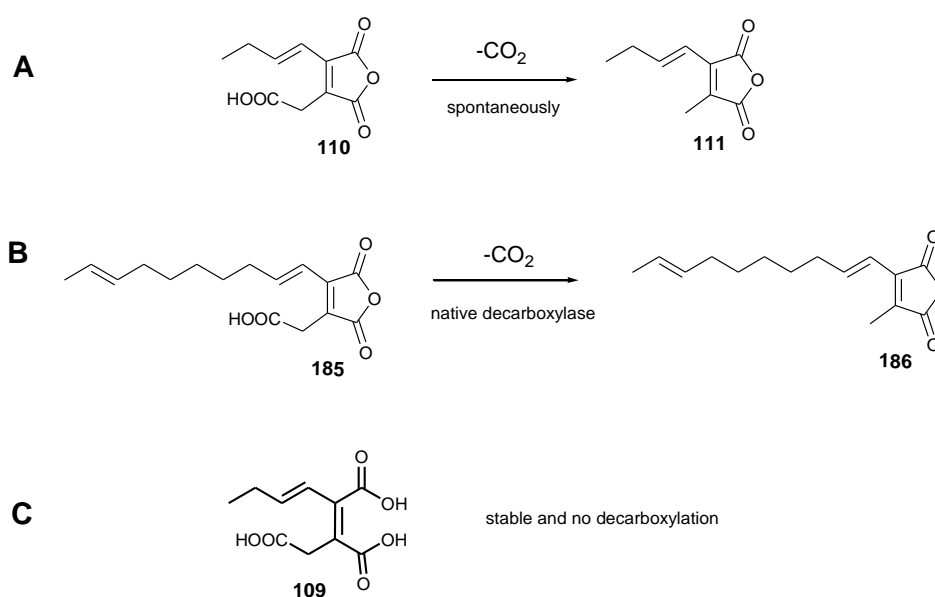
In this project, more detailed studies of maleic anhydride monomer **110** were carried out. From our previous study from Dr. Szwalbe, a maleic anhydride monomer tended to auto-decarboxylate. When the fresh sample of monomer was analysed by ¹H NMR, the presence of a signal at δ_H 3.6 ($\underline{\text{CH}}_2\text{-COOH}$) indicated that the compound was initially carboxylated. However, for the older sample, the presence of δ_H 2.1 ($\underline{\text{CH}}_3$) replaced the δ_H 3.6 peak which was susceptible to auto-decarboxylation.^[2] In this study, we found a way to protect the maleic anhydride monomer from decarboxylation and ring-opening. There are several reasons for decarboxylation and ring-opening. Firstly, freeze-drying turned out to be one reason for decarboxylation. Secondly, organic solvent like DMSO and CDCl_3 would also make the compound **109** decarboxylate, and dimerization of the monomers would happen next. Thirdly, the solvent like MeOH would make the maleic anhydride monomer ring-open and esterify to **269**. The decarboxylated product was not active for KI enzymes, so it was important to use the original maleic anhydride substrate (**Scheme 3.3.4.1**).



Scheme 3.3.4.1 Chemical changes of maleic anhydride in different condition and different solvent.

3.3.5 Decarboxylation during the biosynthesis of maleidride

Our experiment have previously identified that the maleic anhydride monomer **110** could decarboxylate spontaneously to give **111** (Scheme 3.3.5.1 A). This demonstrates that, at least for the shorter-chain maleic anhydrides, no enzymatic decarboxylation was necessary. Oikawa and co-workers suggested that the longer-chain decarboxylated monomer **186** isolated from their heterologous expression experiments was produced by an adventitious native decarboxylase (Scheme 3.3.5.1 B).^[61]

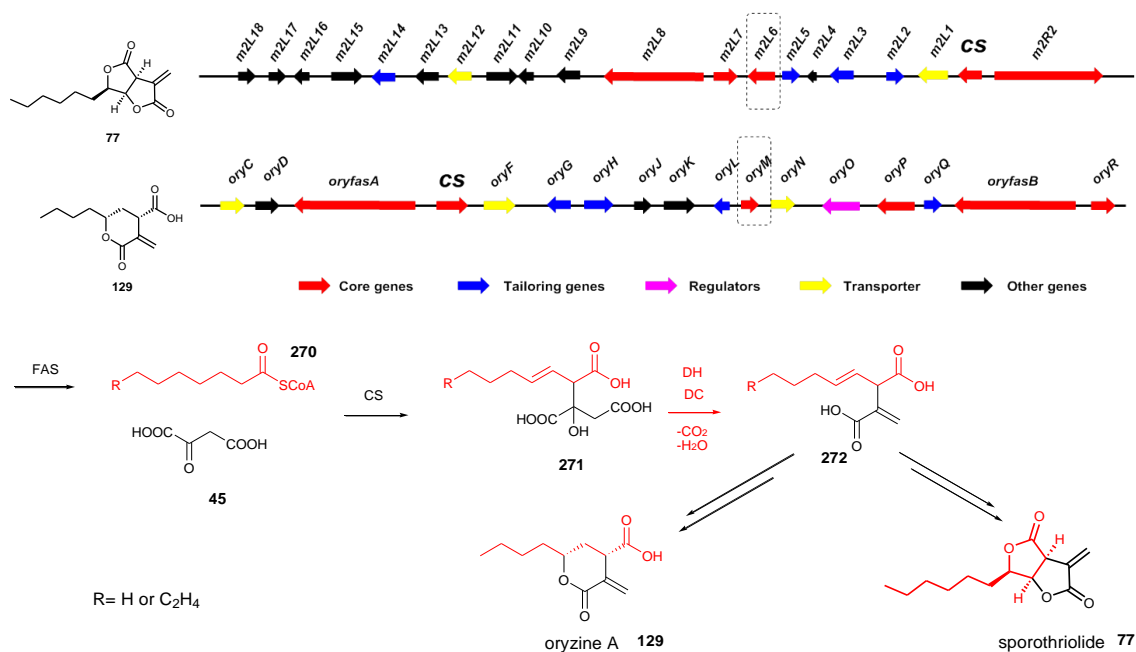


Scheme 3.3.5.1 Decarboxylation of maleic anhydride monomer

On the other hand, our group members Dongsong Tian and Dr. Eric Kuhnert worked on the biosynthesis of sporothriolide **77** (unpublished) and oryzine A **129**^[105]. These two gene clusters encode the core enzymes for alkyl citrate synthesis including citrate synthase, 2-methyl citrate dehydratase, and fatty acid synthase. The BGC also encode one decarboxylase (*m2L6* and *oryM*, Fig. 3.3.5.1). In the structure of sporothriolide **77** and oryzine A **129**, there was no maleic anhydride monomer motif in their structure. The decarboxylation step of sporothriolide and oryzine A should be different from the maleic anhydride **110**.

Experiments show that maleic anhydride monomer **110** can decarboxylate spontaneously after formation of five-membered anhydride ring. However, the five membered ring precursor citrate **227** does not auto-decarboxylate (Scheme 3.3.5.1C). So it probably can be explained: because sporothriolide **77** and oryzine A **129** had decarboxylases involved in the gene clusters, the citrate precursor would decarboxylate in this stage other than cyclise into five-membered anhydride ring (Scheme 3.3.5.1). Our experiments show that the alkyl citrates **109** are stable and

do not easily decarboxylate, but after the 2MCDH step the products **110** decarboxylate easily. Since the sporothriolide and oryzine cluster lack a 2CMDH they require a dedicated decarboxylate to produce intermediates **272**.

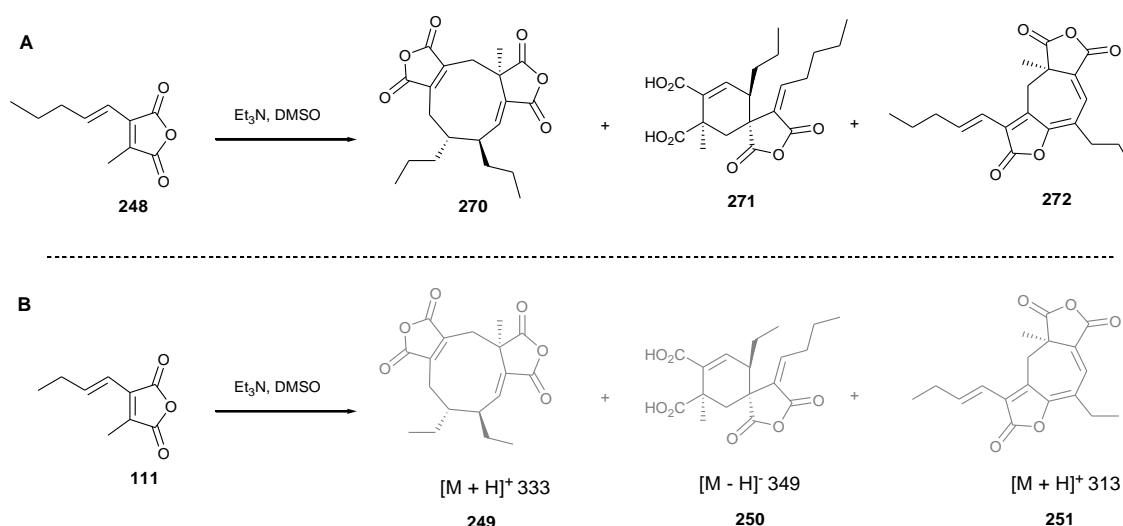


Scheme 3.3.5.1 Proposed gene cluster and biosynthetic pathway of sporothriolide **77** and oryzine A **129**.

3.3.6 Dimerization

The result in this study showed that compound **111** on its own was not a substrate for the KI reaction. It indicated that spontaneous decarboxylation of maleic anhydride monomer is important to drive the reaction. The *in vitro* test on compound **111** treating with DMSO and Et₃N shows that the cyclization reactions were chemically feasible (section 3.2.4.1). Due to the low concentration of the product, it was difficult to purify the dimerized compounds **249-251** for characterization. However, these compounds were corresponded to Baldwin's *in vitro* reaction with compound **248** (scheme 3.3.6.1 A).^[156] The LCMS results indicated that *in vitro* **248** could also be cyclized into different types of dimerized compounds **270-272** which could be related to Baldwin's results (scheme 3.3.6.1 B). As was shown in the chemical reaction of the maleic anhydride monomer **111**, we were not able to observe byssocholamic acid **1** or angestadride **70** formation. This means KI took an important role to control the reaction type in dimerization step in fungus.

III Function of PEBP enzymes



Scheme 3.3.6.1 *in vitro* study of chemical reaction of dimerization. **A**, Baldwin's result; **B**, proposed structure of the dimerised compounds.

The *in vitro* assay showed that both KI enzymes (BfL6 and BfL10) had a weak activity of dimerization. The product is byssochlamic acid **1**. But there was no evidence of heptadride **70** product. No matter in the *B. fulva* strain or KI genes transformed *A. oryzae* strain, heptadride **70** was always a very minor product compared with byssochlamic acid **1**. Heptadride **70** might be a side product of KI enzymes.

3.3.7 Function of PEBP enzymes

According to our *in vitro* study on KI genes, the result suggested that either KI encoded by the byssochlamic acid gene cluster is able to catalyse the dimerization step. This is consistent with the fact that most maleidride BGC encode only a single KI, and the byssochlamic acid BGC probably repeated a gene duplication. However, the exact function of PEBP enzymes is still not clear yet.

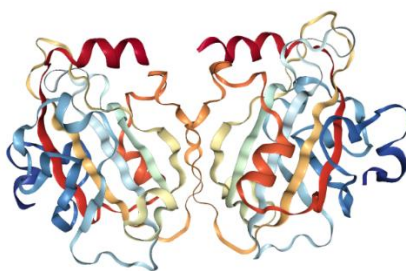
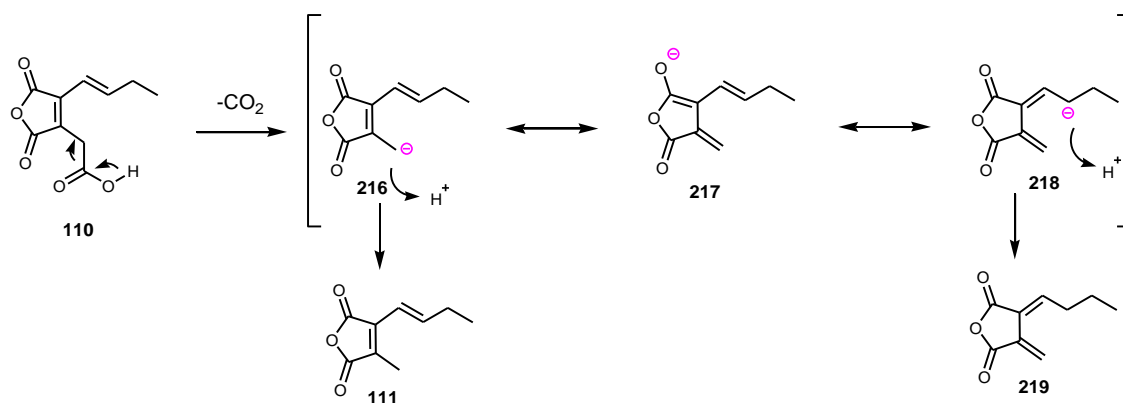


Fig. 3.3.7.1 Protein structure of the phosphatidylethanolamine-binding protein from bovine brain.

Phosphatidyl ethanolamine-binding protein (PEBP) is a basic protein found in numerous tissues from a wide range of species. The screening of gene and protein data banks defines a family of PEBP-related proteins that are present in a variety of organisms, including *Drosophila* and inferior eukaryotes. [186] PEBP binds to phosphatidyl ethanolamine and nucleotides *in vitro*, but its biological function *in vivo* is not yet known. Structures have been determined for several members of the PEBP-like family, all of which show extensive fold conservation. The structure consists of a large central beta-sheet flanked by a smaller beta-sheet on one side, and an alpha helix on the other (**Fig. 3.3.7.1**). [187] Sequence alignments show two conserved central regions, CR1 and CR2, that form a consensus signature for the PEBP family. These two regions form part of the ligand-binding site, which can accommodate various anionic groups. In PEBP, a small cavity close to the protein surface has a high affinity for anions, such as phosphate and acetate, and also phosphorylethanolamine.



Scheme 3.3.7.1 The mechanism of *in vivo* carbanion formation *via* decarboxylation.

As we supposed previously, the maleic anhydride **110** is spontaneously decarboxylated and forms a carbanion intermediate **216-218** (**Scheme 3.3.7.1**). Then PEBP proteins were thought to be able to have the anionic binding ability. Therefore, it could be proposed that, because PEBP had a high affinity for anions, they could attract the carbanion of maleic anhydride monomer to make a relatively high concentration of substrate for the next dimerization step. In addition, it can also be explained that the N terminal signal peptide of KI and PEBP enzymes can target these proteins in a specific place in cells, then these enzymes can catalyse the reaction in a high concentration of substrate (**Fig. 3.3.7.2**).

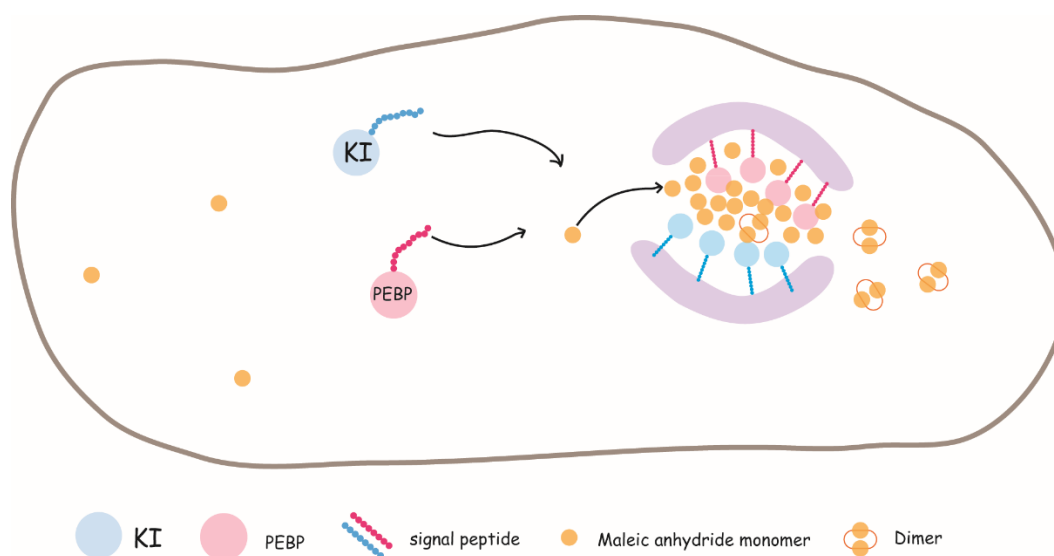


Fig. 3.3.7.2 Proposed function of PEBP based on the signal peptide.

3.4 Conclusion and outlook

The aim of this project was to explore the biosynthesis of byssochlamic acid **1** and angestadrine **70** and try to delineate the enzymes involved in the BGC of these two compound. Based on our previous *in vivo* study of heterologous expression, the BGC in *A. oryzae*, the enzymes related to the biosynthesis of byssochlamic acid **1** were expressed in *E. coli* or in yeast for *in vitro* assay. The stereoselectivities of the enzymes were also explored in this project.

1. Protein expression

By using different organisms (*E. coli* and yeast) and different methods (different induction condition, removing signal peptide and co-expression with charpnones) the hydrolase (BfL1), citrate synthase (BfL2), 2- methyl citrate dehydratase (BfL3 and PvL2) and KI and PEBP proteins (BfL5 / 6 / 9 and 10) were expressed.

2. *In vitro* assay of hydrolase and citrate synthase

Dr. S. Friedrich characterised the function of the hydrolase (BfL1) and citrate synthase (BfL2) using *E. coli* expression and *in vitro* assay. BfL1 shows the activity to hydrolysis both polyketide CoA and ACP bond polyketide.

Citrate synthase BfL2 can take polyketide CoA (but not ACP bond polyketide) as a substrate together with oxaloacetate to form the *anti* citrate product (2*R*, 3*S* / 2*S*, 3*R*).

The citrate synthase BfL2 protein structure modelling result shows that the modelling is very similar with the CS from primary metabolism (stereoselectivity 2*S*). The active site residues

were very closely overlapped between the two structures. Substrates are likely to be bound in identical orientations. Based on this information, BfL2 possibly has a stereoselectivity of (2*S*, 3*R*).

3. *In vitro* assay of 2MCDH

2-methyl citrate dehydratase BfL3 and PvL2 were expressed in yeast and treated with diastereoisomers of citrate substrates. Both BfL3 and PvL2 can only recognise the *anti* diastereoisomer of (2*R*, 3*S* / 2*S*, 3*R*) citrate other than *syn* (2*S*, 3*S* / 2*R*, 3*R*) and dehydrated the substrate to form an alkene **109**. In addition, the product from BfL3 and PvL2 can spontaneously cyclised into a five-membered maleic anhydride **111**, and the anhydride had a balance between the cyclised form and the ring-opening form **109**.

4. *In vitro* assay of dimerization step

The dimerization-related enzymes KI and PEBP were examined by *in vitro*. The maleic anhydride **110** is thought to be the substrate of dimerization related enzymes. The dehydrated compound **110** was very easily decarboxylated to form **111** which was inactive in downstream reactions. However an effective purification protocol allowed **110** to be prepared for *in vitro* assay with the KI and PEBP cell-free preparations. *In vitro* study on KI (BfL6 / 10) and PEBP (BfL5 / 9) suggested that both BfL6 and BfL10 had a weak activity on maleic anhydride monomer to cyclise and form the byssochlamic acid **1**. In the *in vitro* assay, single KI is active to catalysis the dimerization. Whereas, the *in vitro* assay showed that BfL5 and BfL9 were not related to the dimerization,

5. Future work

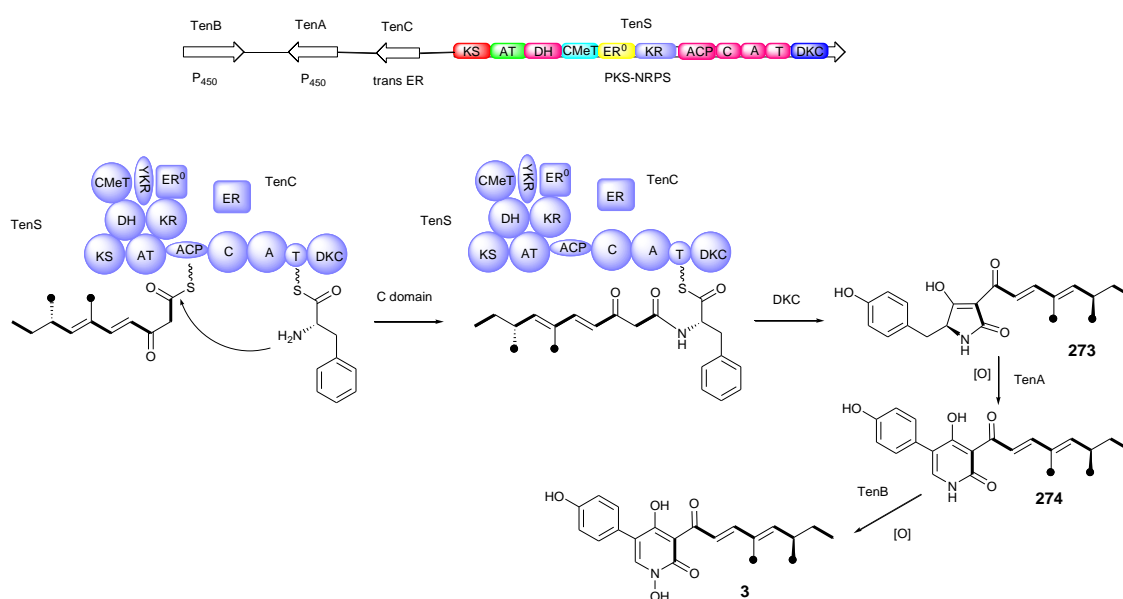
In the future, the heterologous expression system like *Pichia pastoris* or other eukaryotic system should be used to get better producing of BfL6 and BfL10. Because the activity of BfL6 and BfL10 produced by *Saccharomyces cerevisiae* is very weak, the main reason for this should be that *Saccharomyces cerevisiae* didn't produce enough protein for the *in vitro* test. If it is possible to get a good amount of protein, BfL6 and BfL10 could be sent for the identification of protein structure. The structural biology information of BfL6 and BfL10 can help us to get a better knowledge of the mechanism of dimerization function, and it can also help us to set up a strategy for site direct mutation of the KI enzymes. On the other hand, the KI genes in different biosynthetic pathway such as cornexstin **66**, zopfielin **68** and viburspiran **69** can be exchanged by genetic engineering to obtain new maleidride related compounds.

Chapter 4. Molecular Basis of Chain-Length Programming in a Fungal Iterative Highly Reducing Polyketide Synthase

4.1. Background and aims of the project.

4.1.1 Introduction.

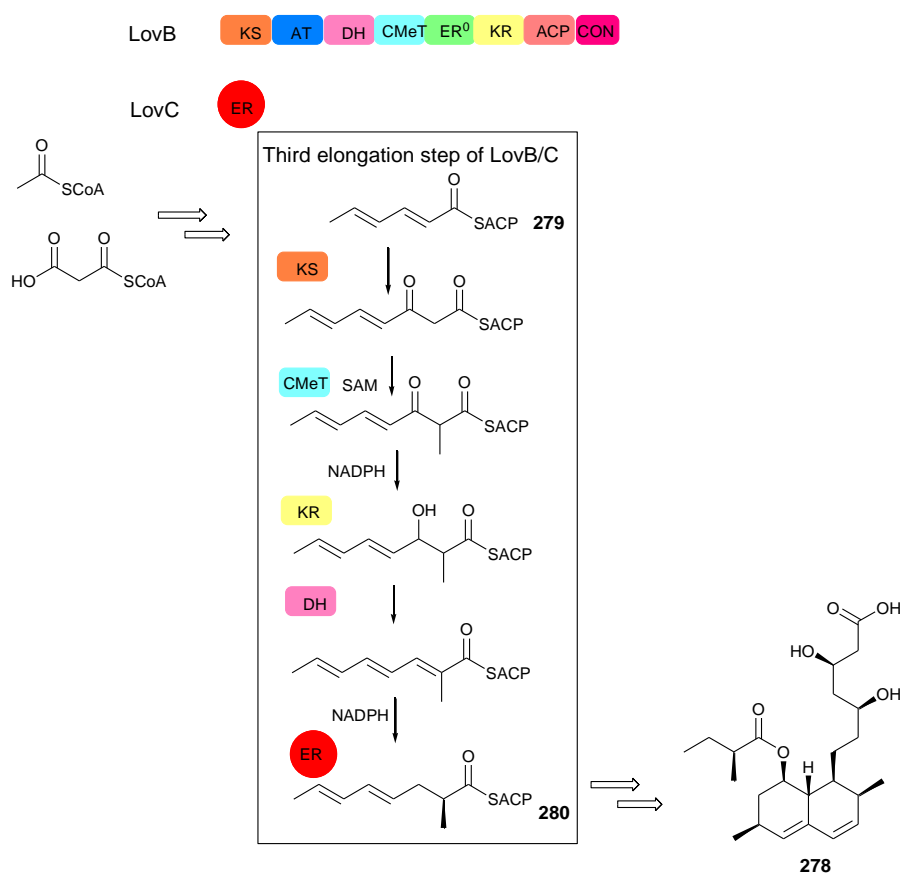
PKS-NRPS systems can be widely found in different fungal genomes. Highly reducing PKS (hr-PKS) have sophisticated programming which control the keto-reduction, dehydration, specific C-methylation and enoyl reduction steps during each cycle (Section 1.2.3),^[109] and the chain-length is also programmed. However, the programming mechanism of hr-PKS is still unknown. The programming in hr-PKS has been rarely investigated. One of the biggest problems is that in an iterative PKS the extension steps are cryptic.^[188] The structure of the polyketide cannot be predicted by sequence analysis. However, the biosynthesis of a few polyketides has been investigated in considerable detail and some information is starting to be discovered. In particular, the cases of tenellin **3**^[10] and lovastatin **6**^[189] have revealed key information about programming.



Scheme 4.1.1.1 Biosynthetic pathway and gene cluster of Tenellin.

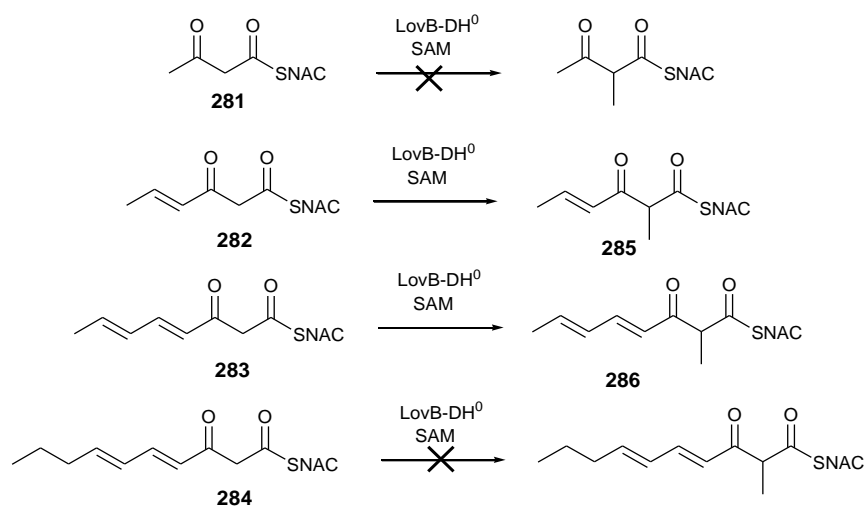
As described in section 1.1, the gene cluster of tenellin **3** encodes four proteins. TenS is a PKS-NRPS enzyme, and TenC is a *trans*-enoyl reductase. TenS and TenC work together to catalyse the formation of pretenellin A **273**. TenA and TenB are P₄₅₀ enzymes. TenA is involved in the ring expansion step from pretenellin A **273** to the pyridone pretenellin B **274**. TenB then catalyses the oxidation at the nitrogen atom to form the final product tenellin **3** (Scheme 4.1.1.1).^[9] TenS

Another example of fungal iterative hr-PKS programming study is from lovastatin biosynthesis. LovB is a hr-PKS which catalyses the biosynthesis of the lovastatin nonaketide precursor dihydromonaconlin L **278**. The C-MeT domain only acts once, after the third elongation of LovB. It has been suggested that the outcome of the modifying steps was determined by the relative activity of each domain towards specific substrates (**Scheme 4.1.1.3 A**). KS, C-MeT, KR, DH and ER domains are all involved in the hr-PKS programming in the third elongation step catalysing the formation of pentaketide **279** from tetraketide **280**.^[189]



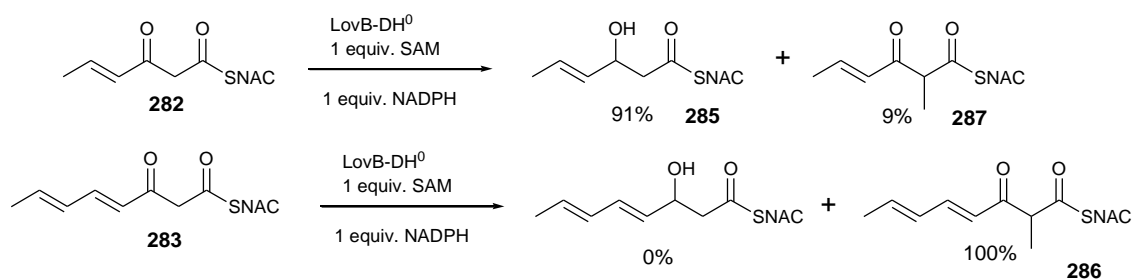
Scheme 4.1.1.3 Third elongation step of LovB and LovC in the biosynthesis of dihydromonaconlin L **278**.

Tang and coworkers showed an example of hr-PKS programming particularly in methyltransferase domain. Full-length LovB was expressed with a mutation in the DH domain which was used for stopping the further modification (LovB-DH⁰). Several acyl-*N*-acetylcysteamine (NAC) substrates (compound **281**, **282**, **283** and **284**) and cofactor *S*-Adenosyl methionine (SAM) were used for the *in vitro* study. The LovB-DH⁰ has no activity on **281** and **284**. But LovB-DH⁰ can catalyse the methylation of triketide **282** and tetraketide **283** to obtain **285** and **286**. The result shows that the C-MeT has much higher activity to substrate **283** than **282** (**Scheme 4.1.1.3**).^[189]



Scheme 4.1.1.3 *in vitro* study of mutant LovB (LovB-DH⁰) on different substrates.

In addition, to test the competition between the KR and C-MeT domains in the presence of the corresponding cofactors SAM and NADPH, triketide **282** and tetraketide **283** were also tested with the LovB-DH⁰. The result indicates that triketide **282** has a higher affinity to the KR (91% **285**) than the C-MeT (9% **287**). While tetraketide **283** shows a high affinity to the C-MeT (100% **286**) domain and no reduced product of KR was observed (**Scheme 4.1.1.4**).^[189]



Scheme 4.1.1.4 Competitive enzyme assays of KR and C-MeT domains.

The result of the LovB programming study shows competition between KR and C-MeT which is consistent with hypothesis about the domain completion among the KR, C-MeT and TE domain in TenS domain swap research.

Because hr-PKS are the key enzymes for the biosynthesis of various bioactive compounds, *e.g.* cytochalasans **43**, lovastatin **6** and squalestatins **7**, it is important to find more information about the mechanism of programming with the long-term aim to combine the biosynthetic pathways and engineer the different hr-PKS systems. To get better understanding of hr-PKS programming, the desmethylbassianin (hexaketide) synthetase (DMBS) and militarinone (a heptaketide compound also belonging to 2-pyridone class) synthetase (MILS) will be used for domain swap experiment with the tenellin **3** synthetase (TenS).

Sub-domain swap to make fewer amino acid changes in the KR and C-MeT domains is a good way to find the key region or even key residues which are responsible for the substrate selection and PKS programming. This project focused on the TenS KR sub-domain swap with DMBS and MILS KR domain.

4.1.2 Aims of the project.

The overall aim of this project is to explore the programming of the hr-PKS. More concretely, we want to achieve the goals as follows:

1. Try to understand the programming mechanism for hr-PKS, especially which individual catalytic domains such as the KR have a component of intrinsic selectivity for their substrates.
2. Try to find the evidence for the hypothesis that competition by different catalytic domain for the ACP-bound intermediates can affect the overall programming.
3. Try to find out the important region or residues in the KR domain protein sequence for programming, and give more information for the further active site mutation work.

4.2 Results.

4.2.1 Strategy for domain swaps.

No structure of a complete fungal hr-PKS has yet been reported. However, there is a good crystal structure of mammalian FAS (mFAS, section 1.3.3). hr-PKS and mFAS are homologous in terms of domain structure and therefore mFAS could be used as a model for hr-PKS. In the previous domain swap study, C-MeT- ψ KR and KR domains of TenS were mapped to the mFAS (Fig. 4.2.1.1). The TenS catalytic KR domain is predicted to span from 1181 aa to 2162 aa containing 281 amino acid residues.

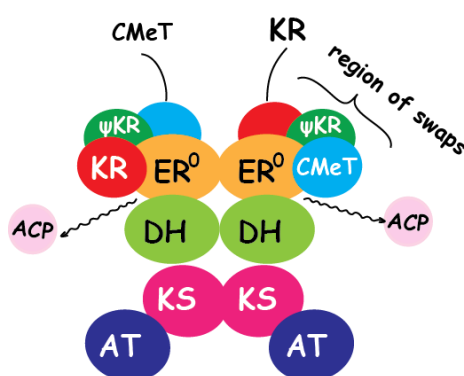


Fig. 4.2.1.1 Cartoon model of the quaternary structure of TenS.

In this project, the previously investigated 281 residues of the TenS KR domain were divided into four approximately equal fragments (Fig. 4.2.1.2) spanning the length of the KR region. These were used to reconstruct a series of 9 chimeric genes using rapid recombination in yeast. Swap sequences are from two halves of *dmbS* KR (experiments TDS1 / 2), four quarter of *dmbS* KR including (experiments TDS3 / 4 / 5 and 6), one full *milS* KR (experiment TMS1) and a third quarter of *milS* KR (experiment TMS2). The final experiment TMS3 involves only 12 residues changes at positions within the KR active site which are introduced by PCR.

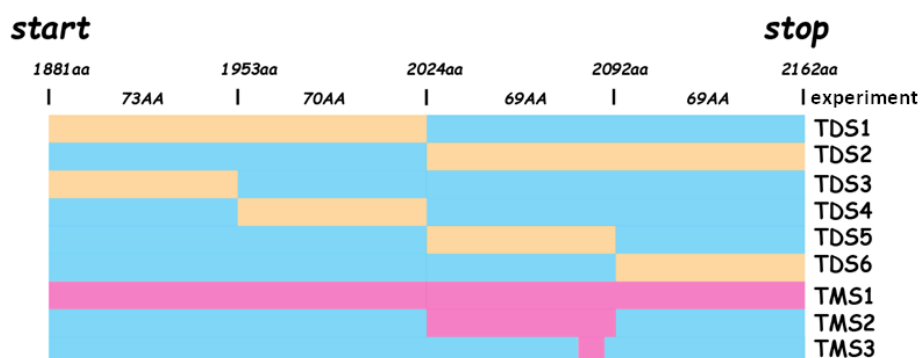


Fig. 4.2.1.2 Protein positions used for swaps: TenS, blue; DMBS, yellow; MILS, red.

IV Strategy for domain swap

The reconstruction procedure consisted simply of combining overlapping DNA fragments constructed from *tenS*, *dmbS* or *milS* templates with the linearised *tenS* cloning vector and transforming the mixture into yeast (**Section 5.3.x**). For example, the exchange of the 281 AA fragment TMS1 into *tenS* was achieved using a 5' PCR fragment obtained from *tenS*, a *ca* 843 bp fragment spanning TMS1 from *milS* with *ca* 30 bp tails matching the *tenS* sequence, and a 3' fragment from *tenS*. The reconstructed plasmids were rescued in *E. coli* and then transferred by *in vitro* recombination (Gateway) methodology into the expression vector (which already contained *tenC*), and this was then transformed into *A. oryzae* NSAR1 (**Fig. 4.2.1.3**). Multiple *A. oryzae* transformants are selected and then grown in parallel on media containing maltose, and extracts are prepared and examined using a standardised extraction and LCMS protocol (**Section 5.3.x**).

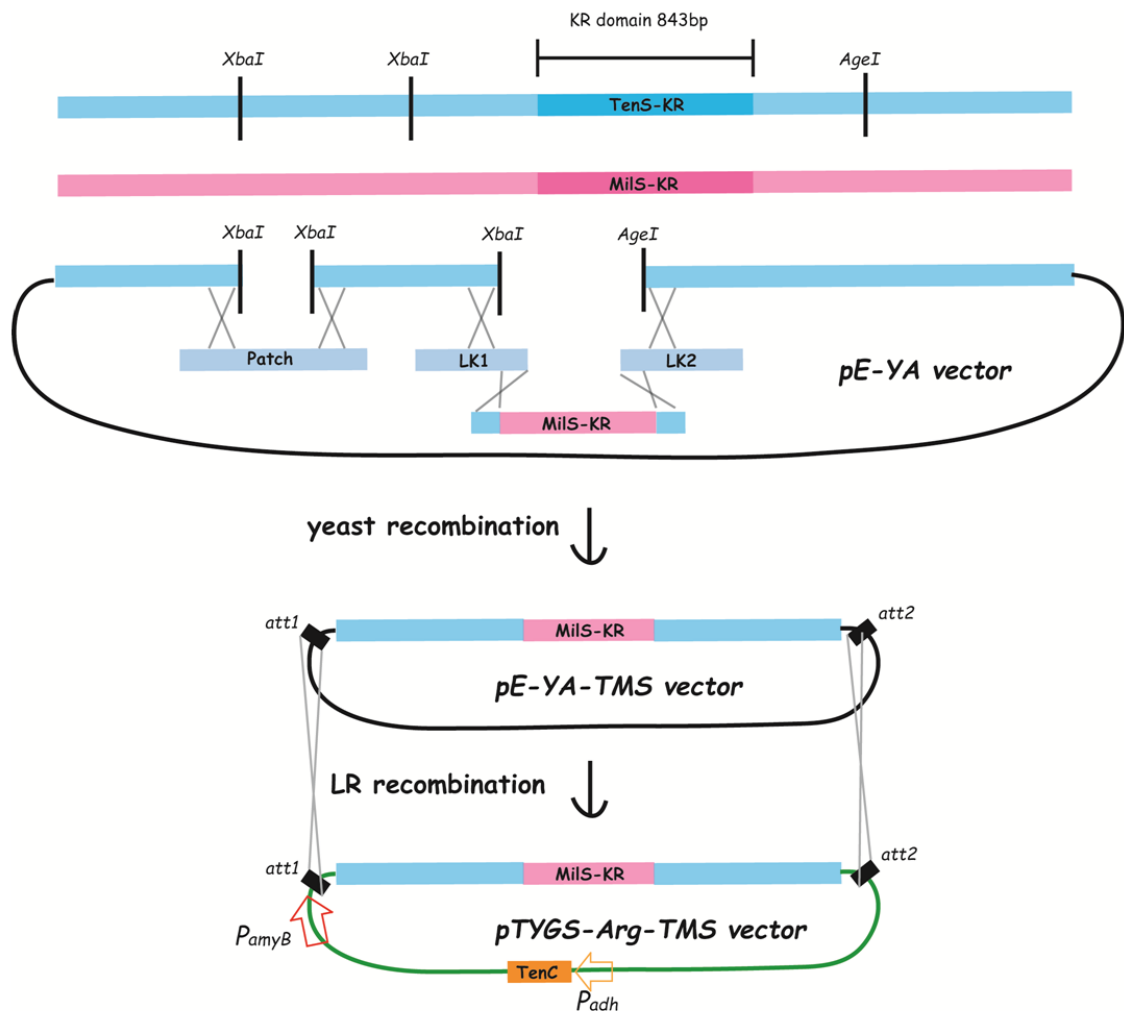


Fig. 4.1.2.3 Cloning strategy for rapid fragment swaps of *tenS* and typical results.

4.2.2. DMBS KR sub-domain swap.

It was previously shown that swap of the entire TenS KR with DMBS resulted in good titres of pentaketides **273** and **276** and the hexaketides **275** and **277**, with dimethylation (70% total product) and hexaketides (80%) predominating.^[197] To get a better understanding about the hrPKS programming of the KR domain, this project is planned to carry out the half and quarter swaps (**Fig. 4.2.2.1**) of the DMBS KR domain. The strategy of the swap experiment has been described in **section 4.2.1**.

In this part of work, the yeast recombination and pE-YA plasmid construction of experiments TDS1-6 was accomplished by Dr. K. Williams. The LR recombination and fungal transformation of experiments TDS1-3 and TDS1 compound extraction were carried out by Dr. S. Friedrich. I worked on the LR recombination and transformation of experiments TDS4-6, and all the extractions and analysis of experiments TDS2-6.

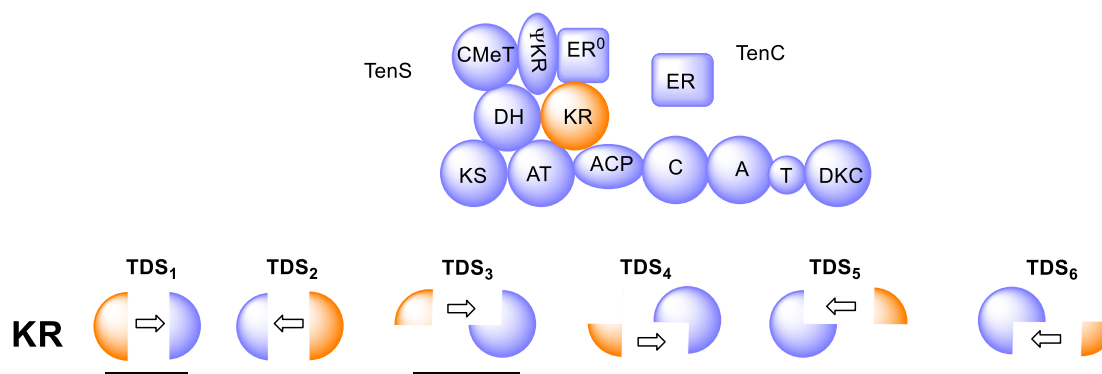


Fig. 4.2.2.1 Cartoon examples of different DMBS KR sub-domain swap. TenS part is in blue, DMBS part in yellow and TenC (ER) in green.

4.2.2.1. TDS1.

The result from Dr. S. Friedrich showed that swap of the first half of the KR (TDS1) was also productive, giving predominantly the previously identified and characterized pentaketides **273** and **288**, and only a little bit of known hexaketide **277** and its isomers. At the same time, the known pretenellin A isomer **289**, proto DMB-B **290**, and desmethyl pretenellin A **276** were also produced (**Fig. 4.2.2.2**).

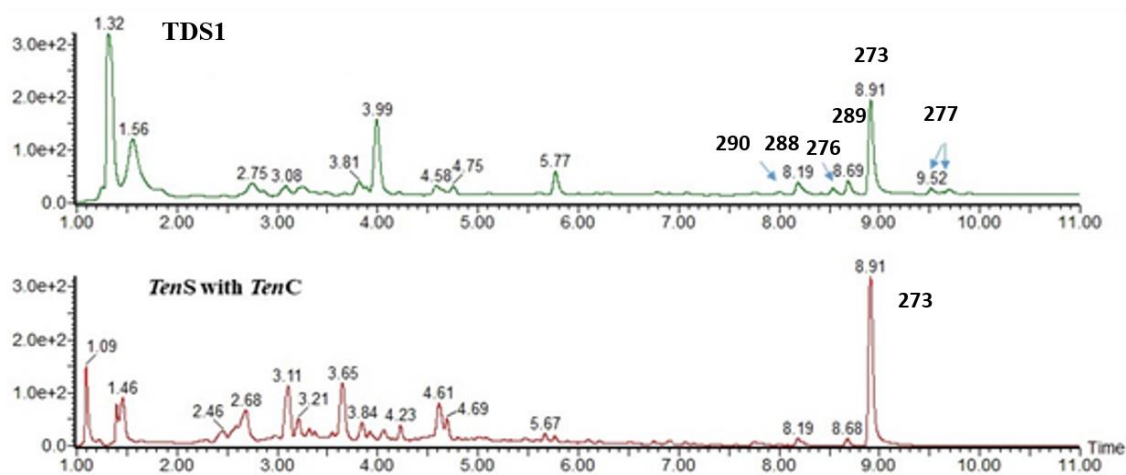
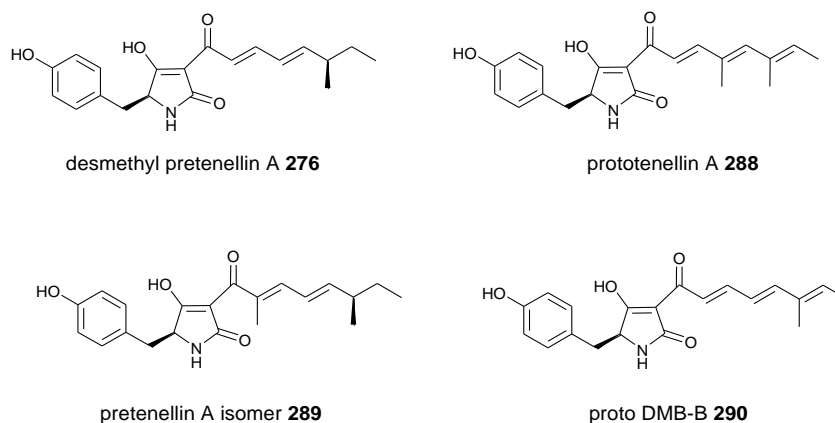


Fig. 4.2.2.2 Analysis of the crude extracts of TDS1 and the control (TenS with TenC).

4.2.2.2. TDS2.

We next swapped the second half of the KR which led to a low yield of predominantly **291**, accompanied by some **277**. Both compound **277** and **291** are hexaketides, and **291** is an oxidised form of compound **277**. In this swap almost no pentaketides were produced (Fig. 4.2.2.3).

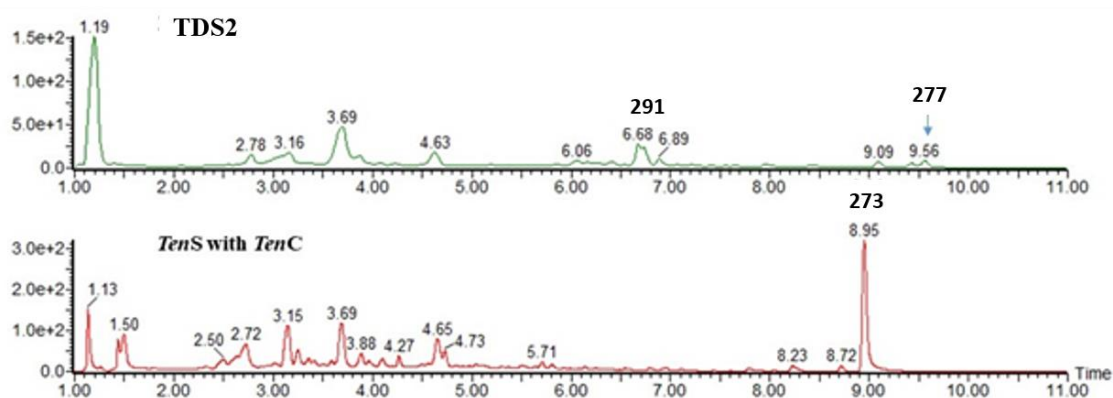


Fig. 4.2.2.3 Analysis of the crude extracts of TDS2 and the control (TenS with TenC).

4.2.2.3 TDS3.

The individual first quarter of the DMBS KR was exchanged. Transfer of fragment TDS3 resulted in production of **273** in titre was similar to the *A. oryzae* NSAR1 system, but with an additional low content of **289** (Fig. 4.2.2.4).

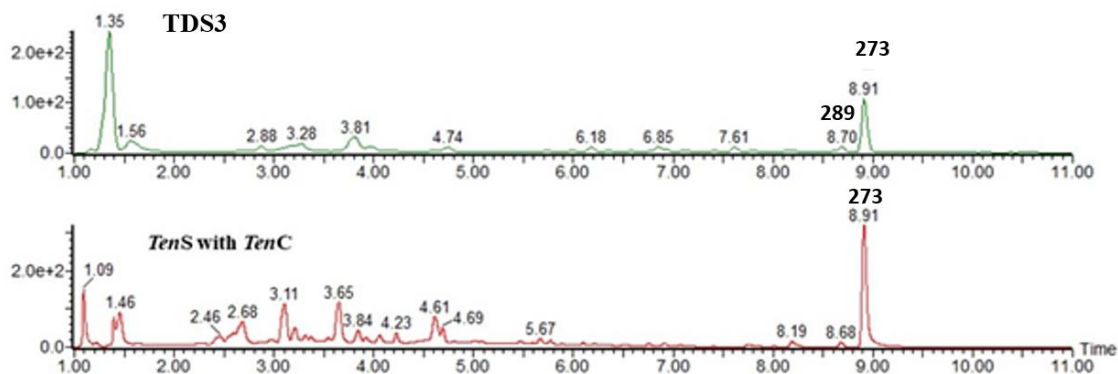


Fig. 4.2.2.4 Analysis of the crude extracts of TDS3 and the control (TenS with TenC).

4.2.2.4 TDS4.

Exchange of the second quarter of the KR, produced protopretenellin A **288** as the major product, and with a low content of compounds **273** and **289** (Fig. 4.2.2.5). However, the overall titre was very low.

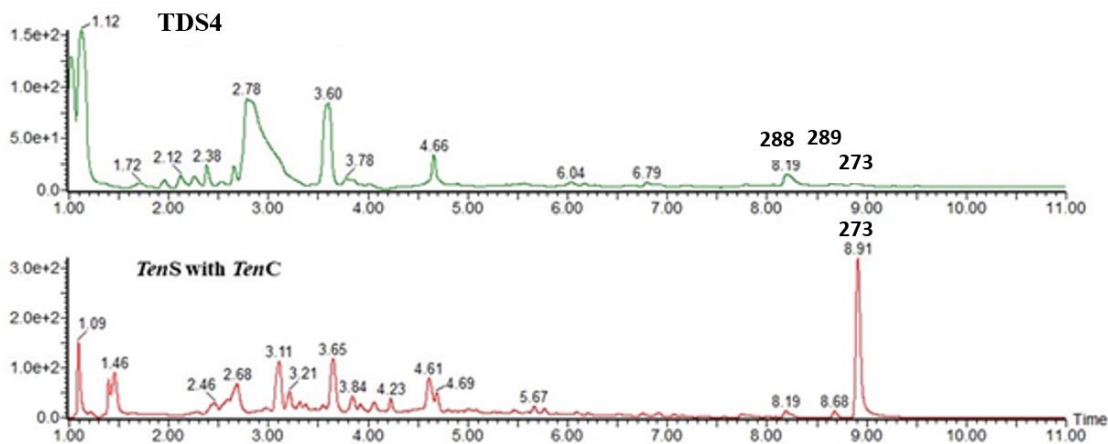


Fig. 4.2.2.5 Analysis of the crude extracts of TDS4 and the control (TenS with TenC).

4.2.2.5 TDS5.

The third quarter of the KR domain was then exchanged. Transfer of TDS5 resulted in a system almost identical to TDS2, but it was much more productive. In overall high titre, TDS5 gave the production of at least four compounds. This consisted of predominantly compound **291**. However, according to the HPLC result, the peak of compound **291** was not a single peak, it means that it

should be a mixture of isomers. TDS5 could also produce low yield of **273**, **290** and **277** (Fig. 4.2.2.6).

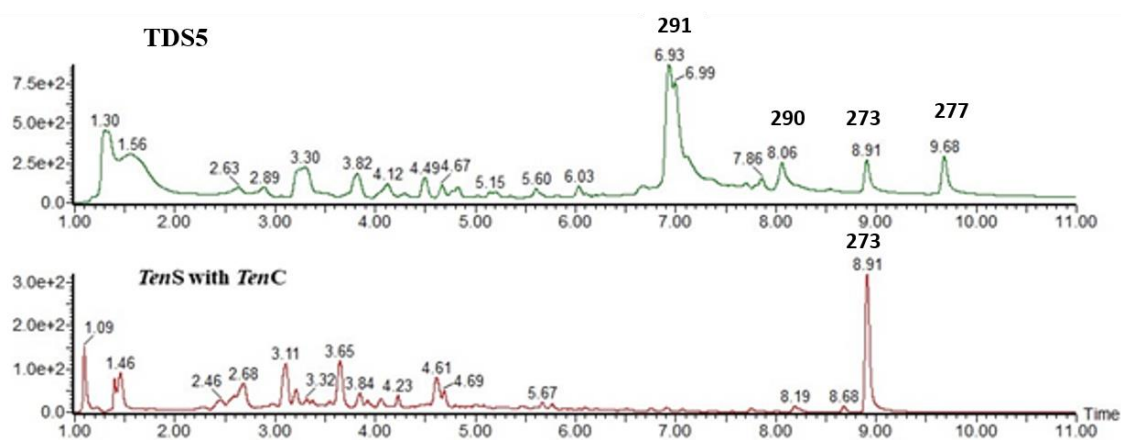


Fig. 4.2.2.6 Analysis of the crude extracts of TDS5 and the control (TenS with TenC).

Because TDS5 gave good production of compound **291**, the fermentation was scaled up which allowed purification of **291**. A crude extract from 5L of CMP medium was purified by preparative-HPLC. 8 mg of purified compound **291** was obtained. Compound **291** has a distinctive UV spectrum with absorption maxima at 225 and 398 nm (Fig 4.2.2.7 A). The compound had a molecular ion of 396.5 [M - H]⁻ (Fig. 4.2.2.7 B). HRMS calculated for C₂₃H₂₆N₂O₅ (396.1811 [M - H]⁻) was obtained as 396.1768.

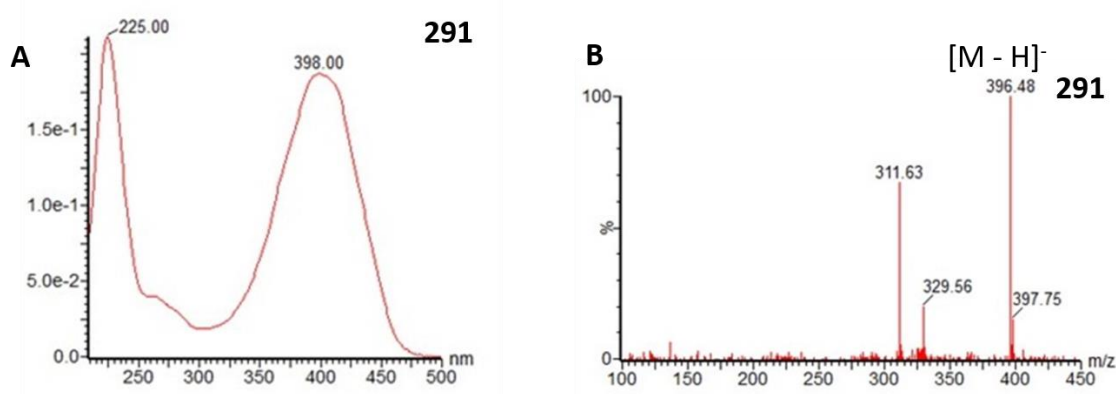
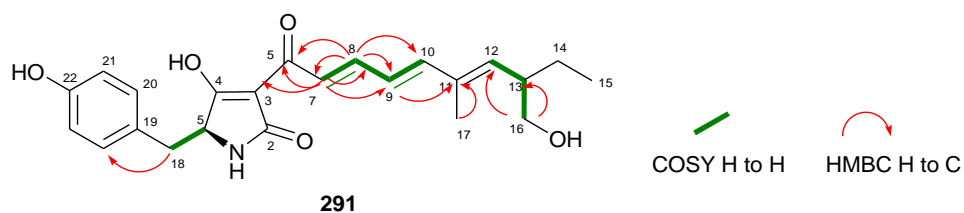


Fig. 4.2.2.7 UV spectrum and MS of compound **13**: A, UV spectrum; B, MS.

Key COSY and HMBC correlations were used to confirm the assignment of the ¹H- and ¹³C NMR spectra (Table 4.2.2.1). The major component of peak **291** turned out to be a hydroxyprebassianin A **291** which was hydroxylated at C-16.



position	δ_H /ppm	Mult./J	δ_C /ppm	COSY	HMBC (H to C)
2	-	-	163.6	-	-
3	-	-	110.4	-	-
4	-	-	194.9 (C)	-	-
5	4.06 (1H)	m	62.4 (CH)	18	4
6	-	-	172.4 (C)	-	-
7	7.04 (1H)	d (15.1 Hz)	119.6 (CH)	8	6, 9
8	7.46 (1H)	dd (15.0, 11.4 Hz)	144.9 (CH)	7, 9	6, 10
9	6.53 (1H)	m	125.7 (CH)	8, 10	7
10	6.88 (1H)	m	149.0 (CH)	9	8, 12
11	-	-	137.1	-	-
12	5.18 (1H)	m	135.1 (CH)	13	12, 15
13	2.35 (1H)	m	42.5 (CH)	12, 14, 16	
14	1.15 (2H)	m	24.3 (CH ₂)	13, 15	15, 17
15	3.29 (2H)	m	64.3 (CH ₂)	14	13, 16
16	1.15 (2H)	m	24.3 (CH ₂)	13	15, 17
17	1.59 (3H)	s	12.9 (CH ₃)	-	13, 10
18	2.83 (2H)	broad s	35.8 (CH ₂)	5	
19	-	-	126.0	-	-
20	6.92 (2H)	m	130.7 (CH)	21	
21	6.61 (2H)	m	114.9 (CH)	20	
22	-	-	155.9 (C)	-	-

Table 4.2.2.1 Summary of NMR data of compound **291**.

4.2.2.6 TDS6.

Finally the fourth quarter of the KR domain was exchanged. Transfer of fragment TDS6 resulted in a system almost identical to the untransformed *A. oryzae* NSAR1 strain. There were no observable new compounds produced (**Fig. 4.2.2.8**).

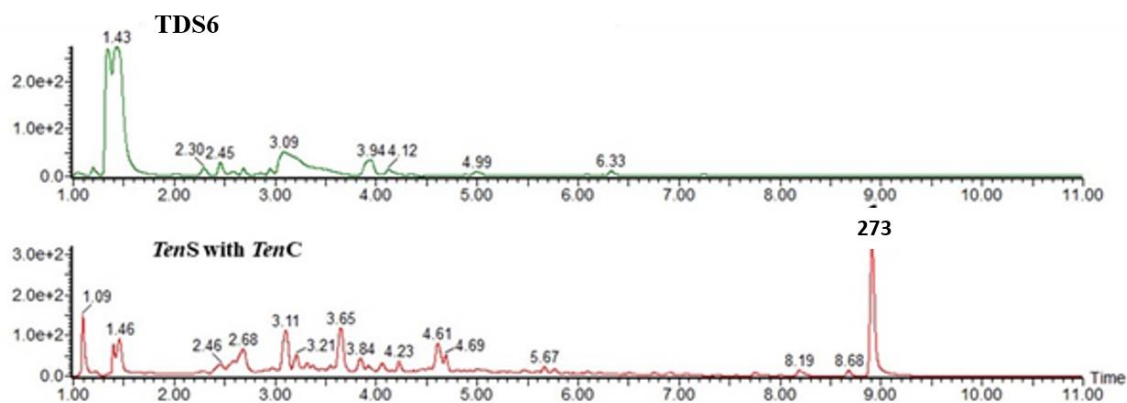
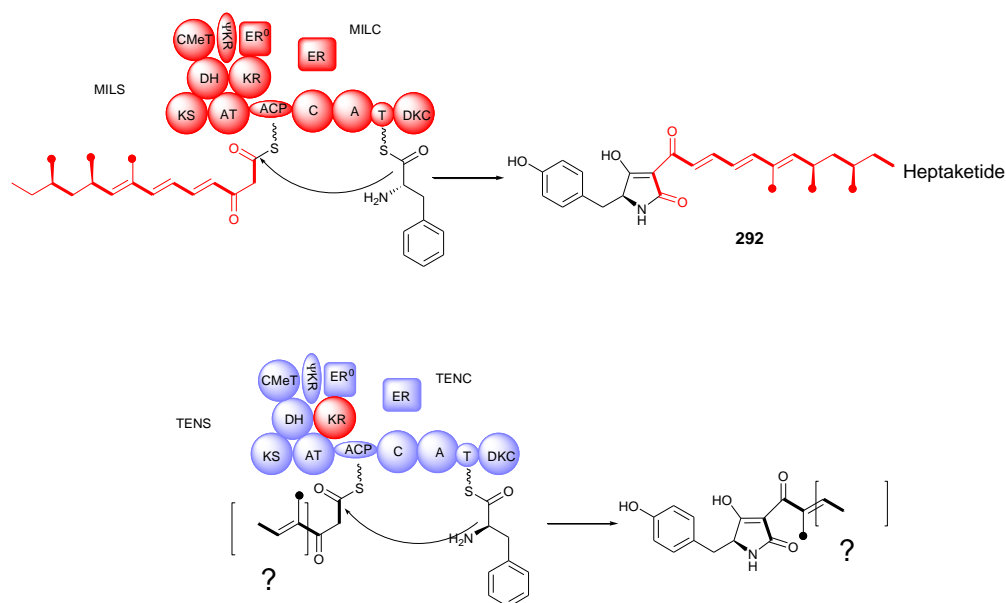


Fig. 4.2.2.8 Analysis of the crude extracts of TDS6 and the control (TenS with TenC).

4.2.3 MILS KR full domain swap

From the previous domain swap experiments, the hexaketide PKS DMBS KR domain swap gives 80% of the hexaketides and 20% of the pentaketides (section 4.1.1). This project is focused on the militarinone synthase (MILS) which synthesises a heptaketide product. Militarinone C **292** is heptaketide congener of pretenellin A **273** which has three methyl groups and is reduced twice by its *trans*-acting ER. The single MILS KR domain swap will be constructed to see if the hybrid PKS system can synthesise heptaketide products (Scheme 4.2.3.1).



Scheme 4.2.3.1 Domain structure and swaps of MILS, cartoon view of a hybrid PKS-NRPS.

4.2.3.1 Vector construction of TMS1.

As described in section 4.2.1, the plasmid pE-YA-*tenS* was digested with *Xba*I and *Age*I to produce the vector fragment VF3 (Fig. 4.2.3.1 A). Fragment F35 (Patch) was amplified by PCR

IV MILS KR full domain swap

(Q5) using *tenSF1* and *tenSR1* as primers and pE-YA-*tenS* as the template to create a patch for VF3. Fragments F51 and F52 were amplified by PCR using TDSLk1-F and TDSLk1-R, T2A-F and TDSLk2-R respectively as primers and pE-YA-*tenS* as the template. Fragment F50 (MILS-KR) harbouring the swap sequence was amplified by PCR using TMS-F and TMS-R as primers with pE-YA1-*milS* (from Dr. K. Williams) as the template (**Fig. 4.2.3.1 B**).

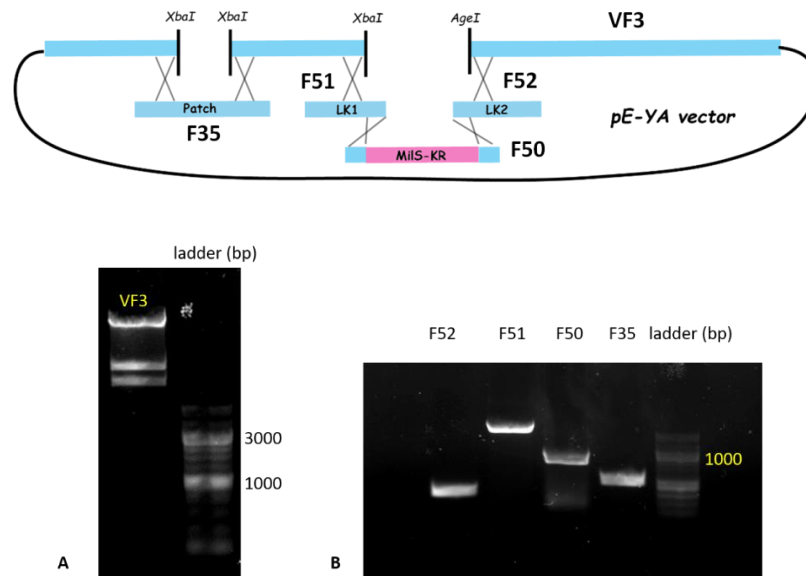


Fig. 4.2.3.1 Fragments for yeast recombination.

Yeast recombination was used to assemble the vector fragment VF3 with F35, F50, F51 and F52. The resulting plasmid pE-YA-*tenS*(Δ milS-KR) harbouring the domain swap was confirmed by restriction cutting with *XbaI* using pE-YA-*tenS* as a control. The pE-YA-*tenS* give two bands with the sizes 14950 and 3198bp. The reconstructed plasmid (2) obtains 3 bands with the sizes 13134, 3198 and 1816bp (**Fig. 4.2.3.2**).

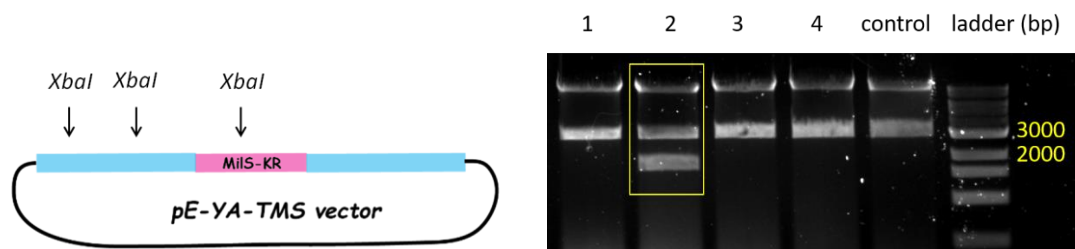


Fig. 4.2.3.2 Restriction cutting conformation.

The positive construction of yeast recombination was then transferred to the expression vector pTYGS-arg-*tenC* using Gateway LR *in vitro* recombination (Invitrogen) leading to pTYGS-arg-*tenC*-*tenS* (Δ milS-KR TMS1).

4.2.3.2 Fungal transformation of TMS1 and chemical analysis.

TMS1 plasmids of KR domain swap were transferred into *A. oryzae* following the general method in section 5.5.3, and eight transformants were obtained. The transformants were then grown in 100 mL CMP medium for 7 days. The compounds from the cultures were then extracted and analysed by LCMS following the general method in section 5.6.4.

The whole MILS KR domain was exchanged with TENS KR (span from 1181 aa to 2162 aa). The chimeric synthetases (in two transformants) produced pentaketide pretenellin A **273**, hexaketide prebassianin A **277** and, surprisingly, the putative heptaketide **293**. Hexaketide prebassianin A **277** and **293** were the major components with pretenellin A **273** as minor component. As a control, TenS with TenC transformant produced only pretenellin A **273** (Fig. 4.2.3.3).

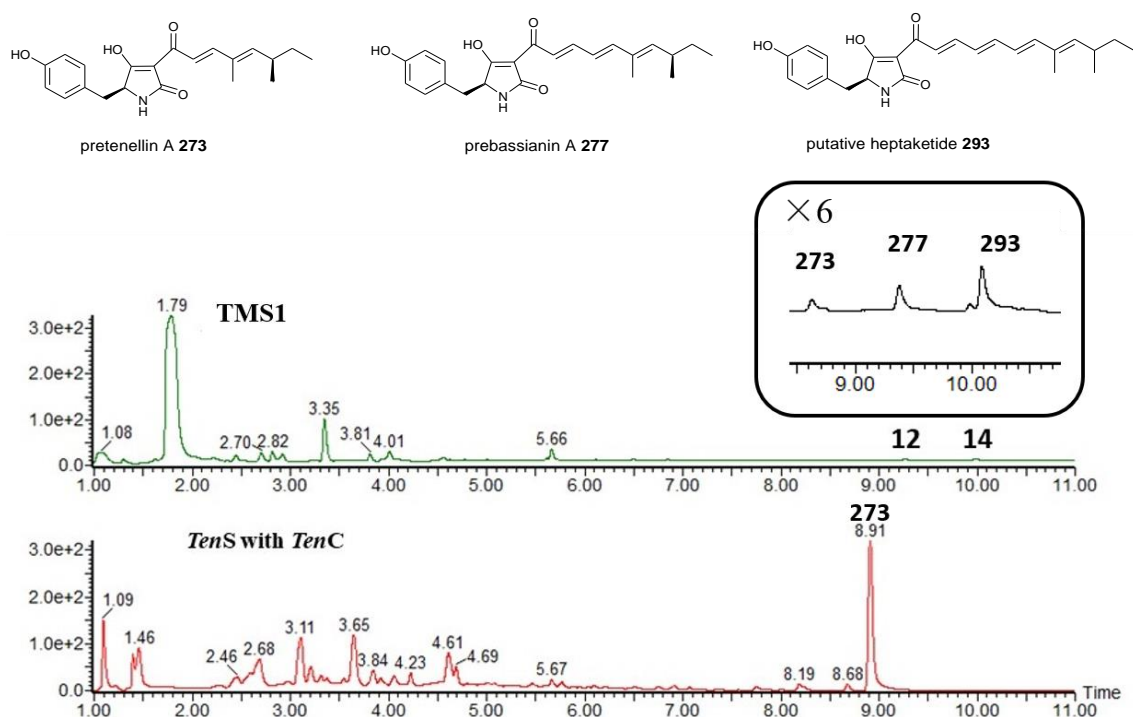


Fig. 4.2.3.3 Analysis of the crude extracts of TMS1 and the control (TenS with TenC) transformants by HPLC (10 - 90% CH₃CN : H₂O gradient, 15 min).

As expected, compounds **273**, **277** and **293** showed an increase λ_{\max} from 373 nm for the dieneone **273** to 426 nm for the tetraeneone **293**. And the molecular weight also showed an increase from 355 for **273** to 407 for **293** (Fig. 4.2.3.4).

The result of TMS1 indicates that the whole MILS KR domain swap gives the mixture of pentaketide, hexaketide and heptaketides. This is again the evidence that KR domain of the hr-PKS controls the programming of the chain length. However the overall titre was very low showing that the chimeric system does not function well.

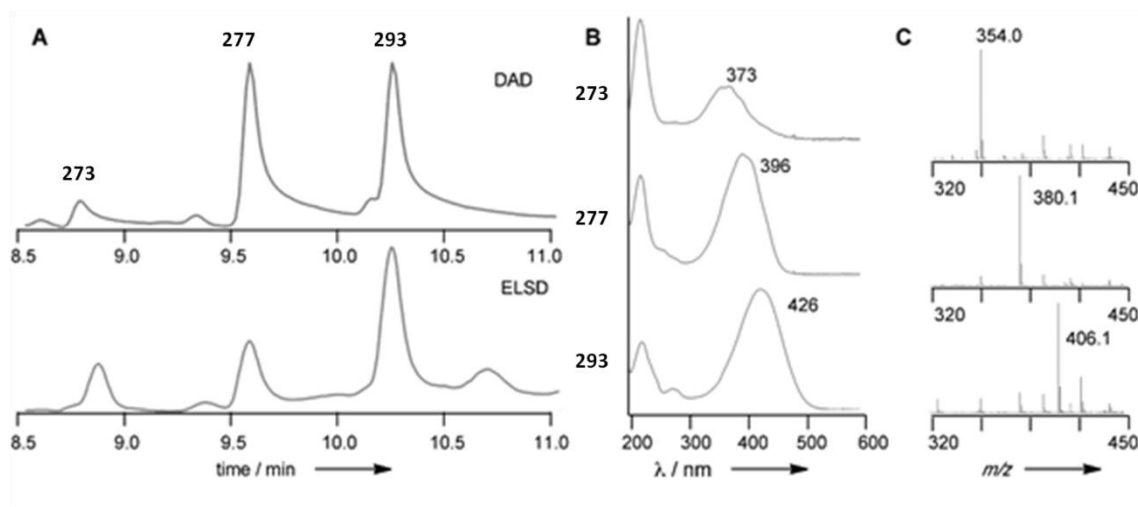


Fig. 4.2.3.4 Results of MILS KR swap: **A**, HPLC chromatograph from 8.5 to 11 min; **B**, UV analysis of indicated compounds; **C**, ESMS- analysis of indicated compounds.

4.2.4 MILS KR sub-domain swap.

According to the result of the TDS2 and TDS5 swap experiment, the chimeric synthetases could produce longer chain compounds as the major components. It means that the third quarter of the KR domain must contain some important regions or residues which take part in the programming of the PKS especially for the chain length control. Therefore, the third quarter of the MILS KR domain were exchanged.

4.2.4.1. TMS2.

Transfer of the third quarter swap of MILS KR (TMS2) resulted in a system almost identical to TMS1, but it was about 10 times more productive. Hexaketide prebassianin A **277** and the putative heptaketide **293** were the major component with pretenellin A **273** as a minor component (**Fig. 4.2.4.1**).

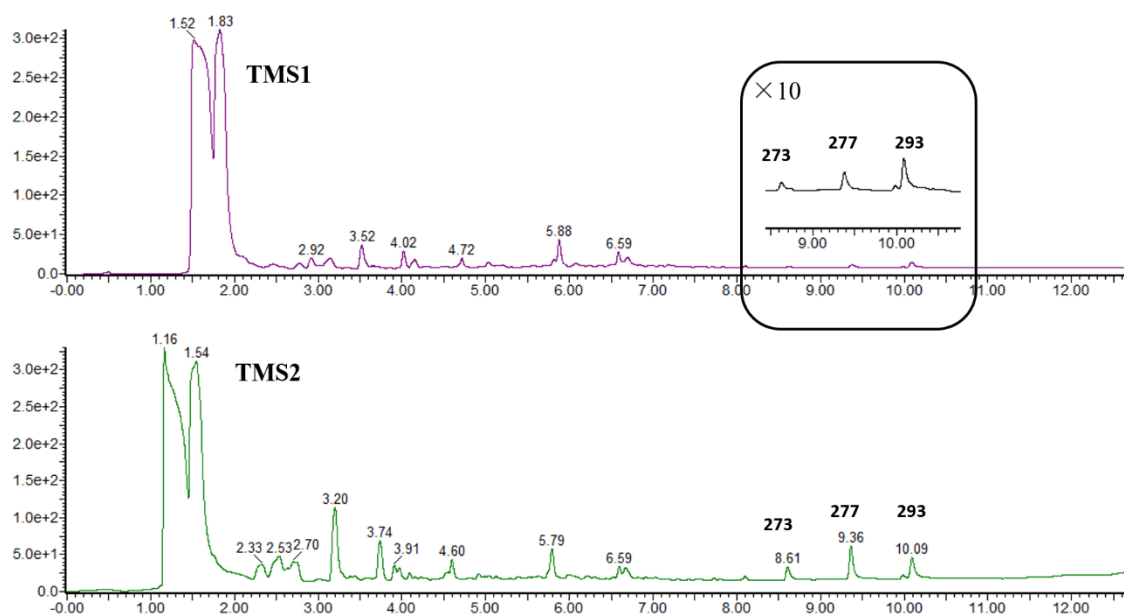


Fig. 4.2.4.1 Analysis of the crude extracts of TMS2 and TMS1.

Because experiment TMS2 gave enough production of compound **293**, large scale of fermentation was carried out for the compound purification. A crude extract from 8 L of CMP medium was purified by preparative-HPLC. 6 mg of purified compound **293** was obtained. Compound **293** has a distinctive UV spectrum with absorption maxima at 224 and 426 nm. This has a molecular ion of 406.2 [M - H]⁻. HRMS calculated for C₂₃H₂₆NO₅ (406.2018 [M - H]⁻) was obtained as 406.2021 (**Fig. 4.2.4.2**).

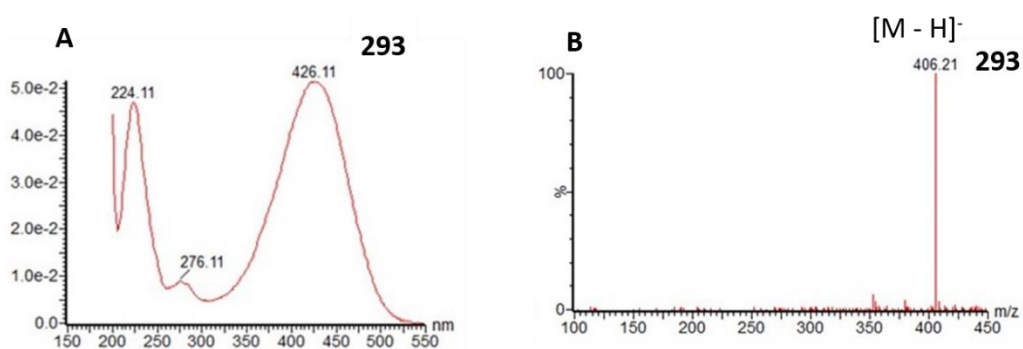
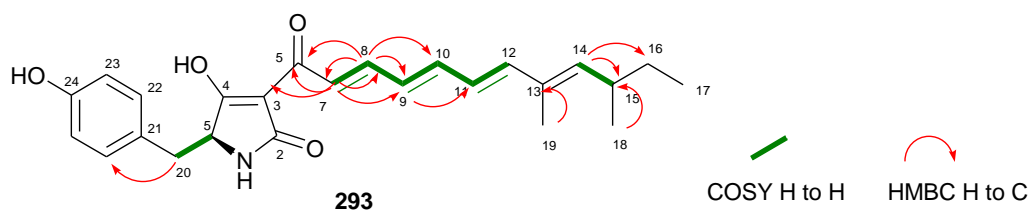


Fig. 4.2.4.2 UV spectrum and MS of compound **14**: **A**, UV spectrum; **B**, MS.

IV Summary of the observation

Key COSY and HMBC correlations were used to confirm the assignment of the ^1H - and ^{13}C NMR spectra (Table 4.2.4.1). The major component of peak **293** turned out to be a new compound named as prefarinosone A **293**.



Pos	δ_{H} /ppm	mult. / Hz	δ_{C} / ppm	COSY	HMBC (H to C)
2	-	-	175.1	-	-
3	-	-	102.5	-	-
4	-	-	191.2	-	-
5	3.68	m	69.9	20	
6	-	-	174.1	-	-
7	7.55	d, $J = 15.8$	129.1	8	6
8	7.18	m	139.5	7, 9	
9	6.47	m	131.3	8, 10	
10	6.63	m	140.2	9, 11	
11	6.34	m	127.0	10, 12	13
12	6.42	m	141.3	11	
13	-	-	133.1	-	-
14	5.41	d, $J = 9.6$	142.2	15, 19	12, 19
15	2.42	m	34.5	14, 18	13, 16
16	1.28, 1.36	m	29.8	15, 17	14, 15, 17, 18
17	0.81	t, $J = 6.9$	12.3	16	15, 16
18	0.95	d, $J = 6.6$	21.1	15	14, 15, 16
19	1.74	s	12.9	14	12, 13, 14
20	2.62, 2.84	m dd, $J = 13.9, 3.9$	37.5	5	22
21	-	-	128.5	-	-
22	6.96	d, $J = 8.2$	130.8	23	21, 23, 24
23	6.60	d, $J = 8.2$	115.2	22	21, 22, 24
24	-	-	156.1	-	-
24-OH	9.11	brs	-	-	23

Table 4.2.4.1 Summary of NMR data of compound **14**.

4.2.4.3. TMS3

Since TMS2 worked quite nicely, so a more focused swap was carried out in the third quarter of the KR domain. By comparing the protein sequences (third quarter) of TENS, DMBS and MILS, a small region was found to contain the most variations in residues (**Fig. 4.2.4.3**).



Fig. 4.2.4.3 Sequence alignment in third quarter (5A part) swap of KR domains. Yellow highlight = identity among PKS; green highlight = identity among mFAS; The red box highlights the most variations part of the domain.

We therefore attempted to swap this focused region from MILS into the TenS KR domain. Exchange of a 12-residue fragment corresponding to Q2398 - V2409 of the TenS KR was introduced by PCR during vector construction, and confirmed by sequencing (**Fig. 4.2.4.4**).

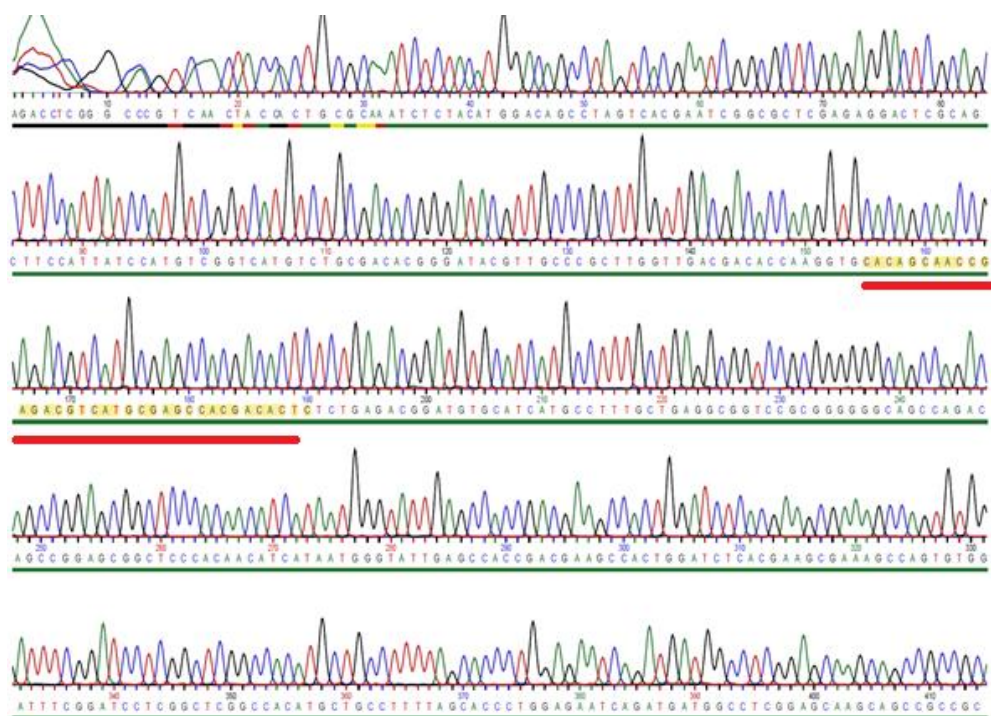


Fig. 4.2.4.4 Sequencing result of the swap. Red line highlights the changed part.

Transfer of TMS3 lead to a similar result as TMS1 although the overall titre was very low. Hexaketide prebassianin A **277** and prefarinosone A **293** were the major components with pretenellin A **273** present as a minor component. However, it was not as productive as TMS2 (Fig. 4.2.4.5).

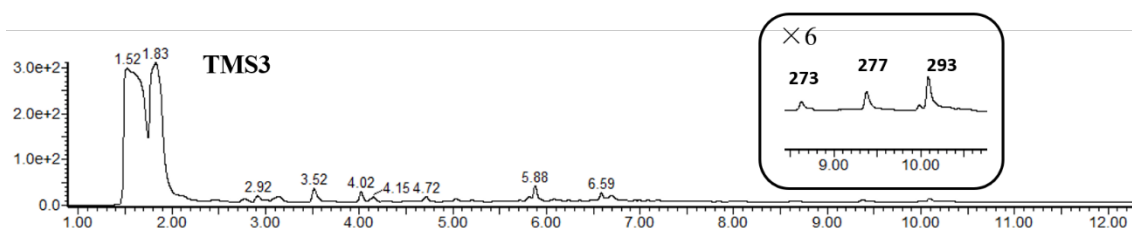


Fig. 4.2.4.5 Analysis of the crude extracts of TMS3.

4.3. Discussion

4.3.1 Summary of the observations and attempt to explain the result using model structure

From the MILS entire KR domain swap (experiment TMS1), the exchange resulted in small amounts of penta- **273**, hexa- **277** and putative heptaketides **293**, all of which are dimethylated and singly reduced, with a majority of the heptaketide. Compared to the MILS whole KR exchange, the previous DMBS KR swap results in good titres of pentaketides **273** and **276** and the hexaketide **275** and **277**, with demethylation (70%) and hexaketides (80%). Both entire KR swaps provide evidence that the KR domain takes part in the chain length control in hr-PKS programming.

Sub-domain swaps of DMBS KR were then carried out. Swaps of the first half of the DMBS KR (experiments TDS1, 3 and 4) were also productive, giving predominantly the pentaketides **273** and **288** but no hexaketides. It indicates that the first half of the KR exchange does not affect the chain length programming. Swap of the second half (experiment TDS2) of the KR gave low titres with few new compounds, while swap of the final quarter (experiment TDS6) resulted in an inactive synthase. Significantly, swap of the third quarter alone (experiment TDS5) was most associated with the production of hexaketide **277** and its hydroxylated derivative **291**, presumably created by adventitious oxygenation in *A. oryzae*.

According to the sub-domain swap of DMBS KR, the third quarter of the KR domain is an important region for the synthase to control the formation of longer chain compounds. The third quarter of MILS KR (experiment TMS2) was also exchanged. It gave a similar distribution of products as TMS1, but in around 5-fold higher titre. In a final experiment, a 36 bp sub-fragment of the KR (experiment TMS3) domain corresponding to residues Q2398 to V2409 of TenS was exchanged. This produced the same distribution of **273**, **277** and **293**, but in overall low titre.

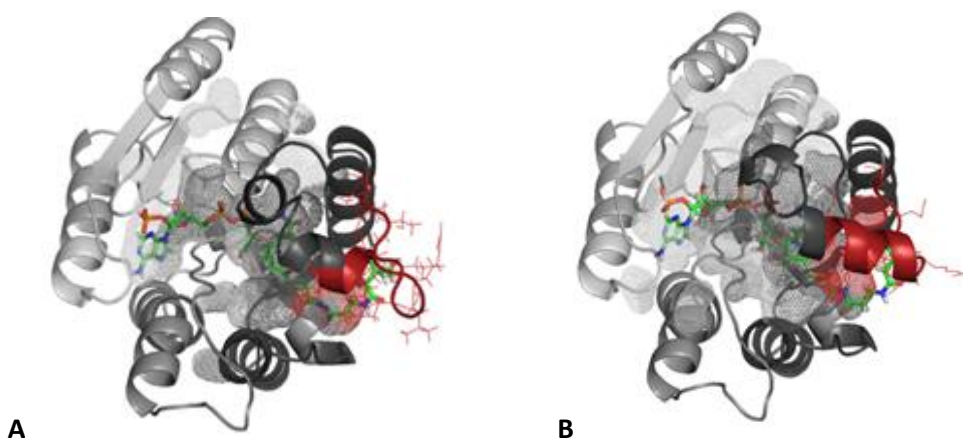


Fig. 4.3.1.1 KR domain modelling: A, TMS3; B, wild type. The changed part highlighted in red.

Our group member Mr Oliver Piech helped me to make two KR domains model structures of experiment TMS3 (**Fig 4.3.1.1 B**) and WT (**Fig 4.3.1.1 A**). Our result shows that experiment TMS3 can produce heptaketide **293** as the dominant product even only with 12 aa exchange in TenS. The wild type TenS however can only produce pentaketide **273**. Compared to the WT KR domain structure, the changes of experiment TMS3 almost focus on one helix (in red). This helix is located exactly in the active site of the KR domain. The modelling of TMS3 KR shows that the helix tends to change into a loop which might make a bigger pocket for active site. The active site therefore can take a longer chain substrate to make a longer chain product.

4.3.2 Domain competition hypothesis

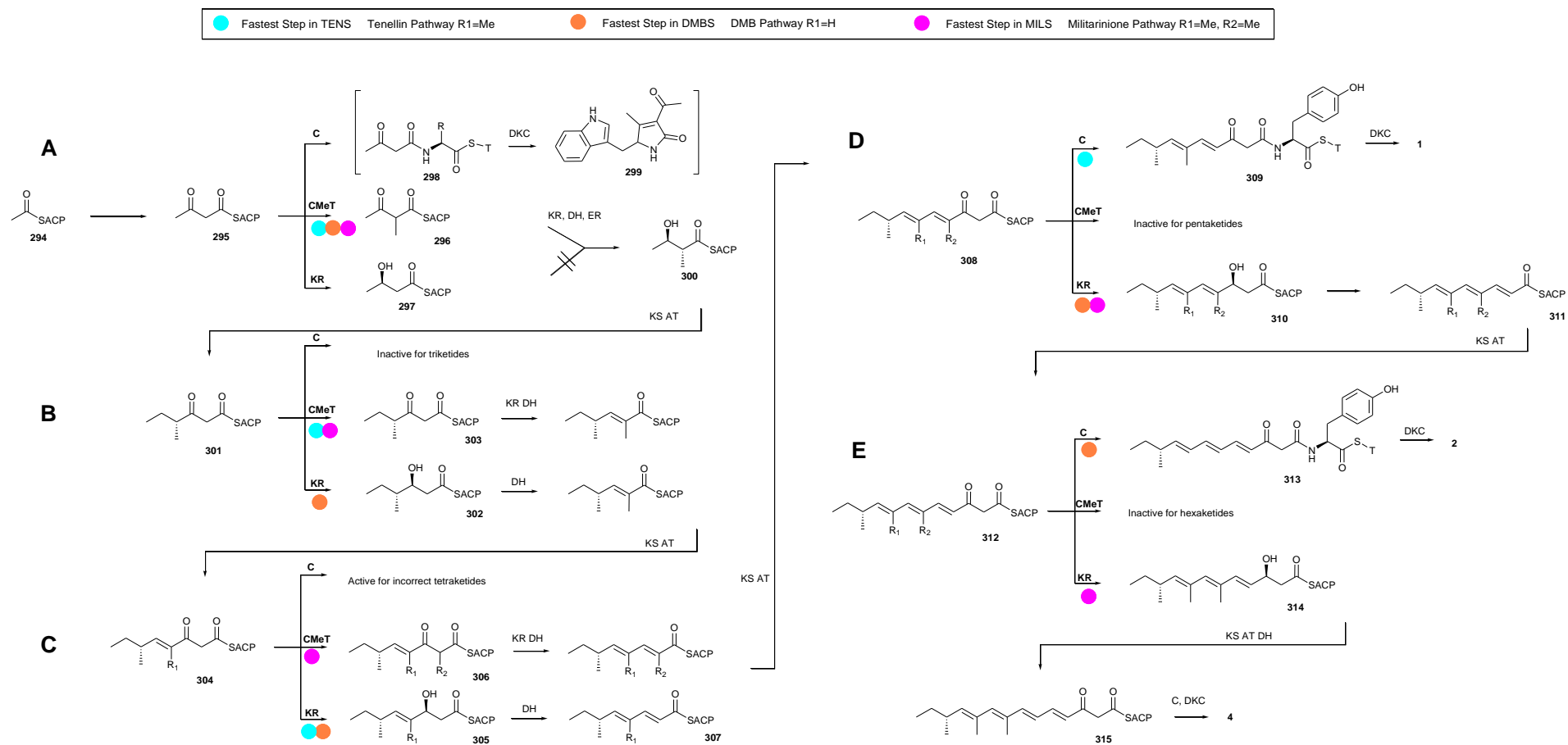
Our observations on KR domain swaps are consistent with a programming mechanism which arises by competition between the selective domains (*C*-MeT, KR and NRPS C domain) for the ACP-bound substrate. For example, in the first condensation (**Scheme 4.3.2.1 A**) of PKS-NRPS hybrid (TenS, DMBS, MILS), the fastest reaction is methylation. In the cases of TENS, DMBS and MILS the *C*-MeT must be significantly faster than KR to give **296**.

In the second condensation (**Scheme 4.3.2.1 B**), the *C*-MeT domains of TenS and MILS is faster than their KR domains after the second round of extension to give **303**. However, in the case of DMBS, KR is faster than *C*-MeT giving **302**. This is consistent with the previous study that the *C*-MeT- ψ KR swap provides mostly monomethylated products. The TenS and DMBS *trans*-acting ERs cannot reduce triketides, so the chain is sequentially extended by KS/AT to give tetraketide **304**. In the third extension (**Scheme 4.3.2.1 C**), both TenS and DMBS *C*-MeT domains cannot methylate therefore giving **305** due to the faster speed of the KR domain. MILS *C*-MeT domain is presumably active in this step to provide **306**.

It then comes into the fourth condensation (**Scheme 4.3.2.1 D**). At this stage the TenS KR cannot reduce the β -carbonyl and the default pathway is release by the NRPS C domain to obtain pentaketide **309** and then pretenellin **273**. On the other hand, the DMBS and MILS KR are still able to reduce pentaketides faster than off-loading to give **310**, in addition to dehydrate to **311**. Triene **311** is extended again by KS/AT to give hexaketide **312**. This is consistent with the DMBS KR swap which can convert TenS to a hexaketide synthase.

Additionally, in the fifth extension (**Scheme 4.3.2.1 E**), the DMBS KR and *C*-MeT domains cannot react with **312** so the default pathway is again *C*-catalysed offloading to give hexaketide **313** and then **275**. MILS KR however can still compete with the off-loading *C*-domain to give **314**. If the KR can reduce hexaketides to give **314**, after dehydration the KS/AT will extend the chain again to heptaketide **315**.

IV Domain competition hypothesis



Scheme 4.3.2.1 Course of polyketide biosynthesis catalysed by hybrid PKS/NRPS: **(A)**, first condensation and subsequent selectivity; **(B)** second condensation and subsequent selectivity; **(C)** third condensation and subsequent selectivity; **(D)** fourth condensation and subsequent selectivity; **(E)** fifth condensation and subsequent selectivity.

4.3.3 *In vitro* study on hrPKS C-MeT and KR domains

From these TENS swap experiments, it was shown that C-MeT and KR domains are the key factors effecting the programming of hr-PKS by controlling the chain length and methylation pattern. However, this *in vivo* study suggests that domain swaps of individual domains, swaps of sub-domains, or even possibly mutations of single amino acid residues, are unlikely to lead to the predictable and clean production of desired compounds. High quality crystallographic data of the TENS or even the specific domain like C-MeT and KR or the *in vitro* study on the domain combining with site direct mutation are needed to get better knowledge of the hrPKS programming.

Stereocontrol is an important aspect of hrPKS programming. At the moment, there are already some studies on stereocontrol of PKS and mFAS. During the 1990s, Cernon Aderson and Gordon Hammes's *in vitro* study showed that stereo-chemistries in fatty acids were strictly controlled by mFAS.^[190] For example, the stereochemistry at C-2 of the malonyl-CoA inverts during chain extension; the KR uses the 4'-*pro-S* hydride of NADPH to attack the 3-*Si* face of the ACP bound polyketide. Due to the close relationship between hrPKS and mFAS, the stereopreferences in mFAS could somehow enlighten the study on hrPKS. However, the hrPKS is usually differently programmed in the β -processing steps while mFAS always do the same. This behaviour of iterative hrPKS made it even more difficult and complicated to understand how the enzyme control the condensation, β -processing and product releasing after one round of condensation.

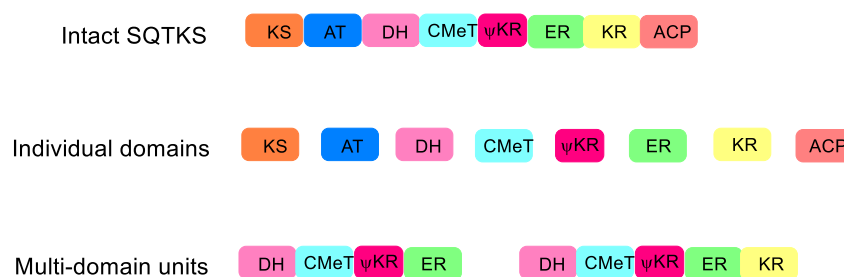
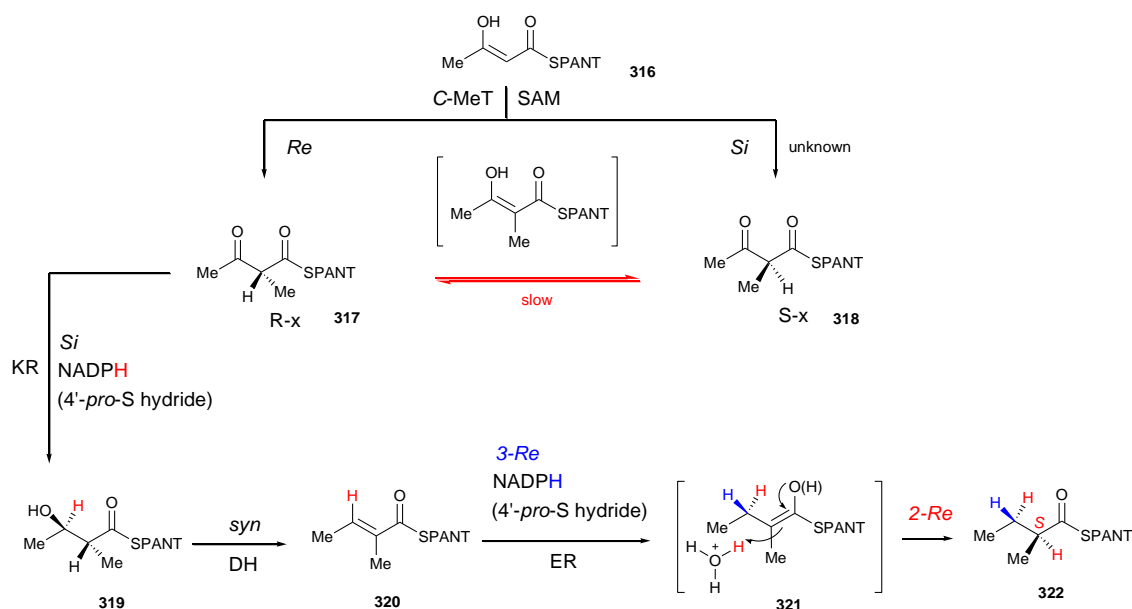


Fig. 4.3.3.1 Strategy of *in vitro* enzyme assay on SQTKS

Our group has already carried out some *in vitro* studies on a hrPKS squalstatin tetraketide synthase (SQTKS). A series of individual domains and multi-domain units of SQTKS were expressed for *in vitro* assay (Fig. 4.3.3.1). According to the *in vitro* results from Dr. H. Yao (unpublished), ER in the DH-KR tetradomain can be used as a general catalyst which was consistent with the previous monodomain study, and all the products of DH, KR and ER were also found by LCMS analysis. Stereoselectivity of ER, KR and C-MeT domains were also tested

by diketide panthethine substrates. For ER in the DH-KR tetradomain, it preferred an *S*-2methyl butanoyl panthethine product **317**. The hydride donor of the KR domain reaction is 4'-*pro*-S proton of NADPH. So far, the stereopreferences involved in KR, DH and ER (**319-322**) of SQTKS are identical between SQTKS and mFAS.



Scheme 4.3.3.1 *In vitro* assay reveals stereochemistry in the first cycle biosynthesis of squalastatin tetraketide.

In addition, Dr. O. Piech in our group also worked on the site direct mutation in the active site of DH and ER domain. Different polyketide panthethine substrates were used for *in vitro* assay of different mutated protein, and the result indicated that mutations could affect the substrate preference of DH and ER domain.^[191]

For the tenellin project, more details of C-MeT and KR domain from TENS can also be explored by *in vitro* study to discuss how active site changes can influence on the chain length and methyl group processing.

4.4. Conclusion and outlook

The overall aim of this project was to explore the programming of the hr-PKS based on the domain swaps between TenS and DMBS/MILS. In this project, we focused on the KR domain and its effect on the chain length control for hr-PKS programming.

Firstly, our results support a programming mechanism for hr-PKS in which individual catalytic domain like KR have a component of intrinsic selectivity for their substrates. The experiments TDS5, TMS2 and TMS3 provide the evidence that the changes taking place in the active site of KR domain can affect the chain length of the product. The amino acid changes around the active site can change the intrinsic selectivity of the hr-PKS and therefore the programming.

Secondly, the results show that the intrinsic selectivity is not the only feature determining overall programming when competition for different catalytic domains is possible. The result of all KR domain or sub-domain swap always tend to give a mixture of products. Experiments TDS2 and 5 give the mixture of pentaketides and hexaketides. Experiments TMS1, 2 and 3 give the mixture of pentaketide, hexaketide and heptaketide. This is the case for the β -keto ACP thiolester intermediates formed by the KS which can be substrates for four different catalytic domains (KR, C-MeT, C and KS) in hybrid PKS-NRPS systems and it must be the ratios of the intrinsic and extrinsic selectivities which determine the final outcome.

Thirdly, we find out the important region in KR domain protein sequence for chain length programming. In this project, experiments TDS5 and TMS2 show that the third quarter of KR domain takes an important role for the chain length programming. Experiment TMS3 is additionally carried out to show that a small number of amino acid changes (12 aa) in the synthase can lead to the production of longer chain compounds. This result provides the evidence for the further active site mutation work.

Chapter 5. Materials and Methods

5.1 Solvents and chemicals

All the chemicals, materials and enzymes were purchased from one of the following companies: Amresco (Solon, OH, USA), Applichem (Darmstadt), Bio-Rad (München), New England Biolabs (Beverly, MA, USA), Roth (Karlsruhe), Sigma Aldrich (Steinheim), Stratagene (Amsterdam, Netherland), and Thermo Fisher Scientific (Waltham, MA, USA).

5.2 Growth media, buffers and solutions

All media, buffers and solutions are summarized in **Table 5.2.1**. All media, buffers and solutions were prepared with millipore water (GenPure Pro UV/UF milipore device, Thermo Scientific) or distilled water. Growth media and transformation solutions were sterilised at 121 °C for 15 min (Autoclave 2100 Classic, Prestige Medical) or by disposable sterile filter (0.45 µm pore size, Roth).

Media	Composition
LB	5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract
LB agar	LB Medium, 15 g/L Agar
YPAD	1 % (w/v) Yeast extract, 2 % (w/v) Tryptone, 2 % (w/v) D(+)-Glucose Monohydrate, 0.03 % (w/v) Adenine
YPAD agar	YPAD Medium, 1.5 % (w/v) Agar
SM-URA agar	0.17 % (w/v) Yeast nitrogen base, 0.5 % (w/v) (NH ₄) ₂ SO ₄ , 2 % (w/v) D(+)-Glucose Monohydrate, 0.077 % (w/v) Complete supplement mixture minus Uracil, 1.5 % (w/v) Agar
glucose-ura	0.67% (w/v) yeast nitrogen base, 0.13% (w/v) of complet supplement mixture minus uracil, 2% (w/v) glucose monohydrate
galactose-ura	0.67% (w/v) yeast nitrogen base, 0.13% (w/v) of complet supplement mixture minus uracil, 2% (w/v) galactose monohydrate

Table 5.2.1 A Media and agar used in this work for *E. coli* and *S. cerevisiae*.

Media	Composition
CZD/S agar	3.5 % (w/v) Czapek Dox broth, 18.22 % (w/v) D-sorbitol, 0.1 % (w/v) (NH ₄) ₂ SO ₄ , 0.15 % (w/v) L-methionine, 0.05 % adenine, 1.5 % (w/v) agar
CZD/S soft agar	3.5 % (w/v) Czapek Dox broth, 18.22 % (w/v) D-sorbitol, 0.1 % (w/v) (NH ₄) ₂ SO ₄ , 0.15 % (w/v) L-methionine, 0.05 % adenine, 0.8 % (w/v) agar
CZD/S1 agar	3.5 % (w/v) Czapek Dox broth, 18.22 % (w/v) D-sorbitol, 0.1 % (w/v) (NH ₄) ₂ SO ₄ , 0.15 % (w/v) L-methionine (optional), 1.5 % (w/v) agar
DPY	2 % (w/v) Dextrin from potato starch, 1 % (w/v) Polypeptone, 0.5 % (w/v) Yeast extract, 0.5 % (w/v) KH ₂ PO ₄ , 0.05 % (w/v) MgSO ₄ * x H ₂ O
DPY agar	DPY Medium, 2.5 % (w/v) Agar
CMP	Czapek Dox, 2% (w/v) Maltose and 1% (w/v) peptone

Table 5.2.1 B Media and agar used in this work for *A. oryzae*.

V Experimental

Media	Composition
PDB	2.4% (w/v) Potato dextrose broth
PDB agar	PDB medium, 2.5 % (w/v) Agar
YMG	0.4% (w/v) D(+)-Glucose monohydrate 0.4% (w/v) Yeast extract 1% malt extract
YPAD	1% (w/v) yeast extract 2% (w/v) tryptone 2% (w/v) D(+)-glucose monohydrate 0.03% (w/v) adenine
YES	4% (w/v) yeast extract 16% (w/v) sucrose
YGMO	0.4% (w/v) D(+)-Glucose monohydrate 0.4% (w/v) Yeast extract 1% (w/v) malt extract 2% (w/v) oat meal fock
producing M	5% (w/v) yellow commeal, 0.1% (w/v) yeast extract, 8% (w/v) surose
CMP	Czapek Dox, 2% (w/v) Maltose and 1% (w/v) peptone
GNB	2 % (w/v) D(+)-Glucose Monohydrate, 3 % (w/v) Nutrient broth Nr. 2 from Oxoid

Table 5.2.1 C Media and agar used in this work for *T. viride*.

Buffer	Composition
Coomassie dye	25 % (w/v) acetic acid, 10 % (w/v) Isopropanol, 0.1 % (w/v) Coomassie
Coomassie bleach	25 % (w/v) acetic acid, 10 % (w/v) Isopropanol
IMAC loading buffer	50 mM Tris-HCl, pH 8, 150 mM NaCl, 10 mM Imidazole
IMAC elution buffer	50 mM Tris-HCl, pH 8, 150 mM NaCl, 40-500 mM Imidazole
50 xTAE buffer	2 M Tris acetate, 0.05 M EDTA, pH 8.3
4X Lämmli buffer	10 % (v/v) β -Mercaptoethanol, 2,5 mg/l Bromophenol Blue 0.25 % (w/v) Bromophenol Blue, 30 % (v/v) Glycerine, 0.25 % (w/v) Xylene Cyanol,
Loading-dye (6 x)	50 mM Na ₂ HPO ₄ , pH 8 25 mM Tris, 192 mM Glycine, 0.1 % (w/v) SDS, pH 8.3
10X SDS running buffer	25 mM Tris, 192 mM Glycine, 0.1 % (w/v) SDS, pH 8.3

Table 5.2.1 D Buffers and Solutions used in this work.

Buffer	Composition
A. oryzae	
Solution 1	0.8 M NaCl, 10 mM CaCl ₂ , 50 mM Tris-HCl, pH 7.5
Solution 2	60 % PEG 3350, 0.8 M NaCl, 10 mM CaCl ₂ , 50 mM Tris-HCl, pH 7.5
T. viride	
protoplast buffer	0.8M MgSO ₄ , 50 mM CaCl ₂ , 50 mM Tris-HCl, pH 7.5
Solution 3	50 % PEG 8000, 1 M Sorbitol, 50 mM CaCl ₂ , 50 mM Tris-HCl, pH 7.5
Solution 4	1M sorbitol, 50 mM CaCl ₂ , 50 mM Tris-HCl, pH 7.5

Table 5.2.1 E Solutions used for fungal transformations.

5.3 Antibiotics and enzymes

Antibiotic stock solutions were prepared in millipore water or ethanol and filter sterilized through 0.45 μ m syringe filters and stored at - 20 °C. Carbenicillin was used in place of ampicillin. Stock and working concentrations are listed in **Table 5.3.1**.

V Experimental

Name	Concentration	Final concentration	Reference
Carbenicillin (Carb)	50 mg/ml in H ₂ O	50 µg/ml	Roth
Chloramphenicol (Cm)	30 mg/ml in EtOH	30 µg/ml	Roth
Kanamycin (Kan)	50 mg/ml in H ₂ O	50 µg/ml	Amresco
Hygromycin B	50 mg/ml in H ₂ O	100-200 µg/ml	Roth

5.3.1 Antibiotics used in this work.

All enzymes used in this study were obtained from NEB or Thermo Scientific and used following the manufacturers' instructions

5.4 Strains

Microbial strains which were kindly provided for this work or purchased from Thermo Fischer Scientific, Stratagene or Novagen are listed in **Table 5.4.1** and **Table 5.4.2**.

Strain	Genotype	Origin
BL21 (DE3)	F- ompT hsdSB (rB ⁻ mB ⁻) dcm gal λ(DE3)	Thermo Fisher Scientific
OneShot Top10	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG	Thermo Fisher Scientific
OneShot ccdB survival 2T1R	F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2	Thermo Fisher Scientific
Rosetta 2	F- ompT hsdSB(rB ⁻ mB ⁻) gal dcm (DE3) pRARE2 (CamR)	Novagen

Table 5.4.1 *E. coli* strains used in this work.

Strain	Feature	Reference
<i>A. oryzae</i> NSAR1	ΔargB sC adeA niaD	Lazarus group, Bristol
<i>T. viride</i> (MF5628, ATCC 74084)	wild type	CBS-KNAW
<i>S. cerevisiae</i> CEN.PK2	MATa/ura3-52/ura3-52 trp1-289/trp1-289 leu23_112/leu2-3_112 his3 D1/his3 D1MAL2-8C/MAL28C SUC2/SUC2	Hahn group, Hannover
<i>S. cerevisiae</i> W303b	ade-, His-, Ura-, Leu-, Trp-	Kong group, Beijing

Table 5.4.2 Fungal strains used in this work.

5.5 Microbiology methods

5.5.1 *E.coli* growth and transformation

E. coli strains were cultivated on liquid or solid LB medium/ agar with appropriate antibiotics. Cultures were incubated at 37 °C for 12 to 18 h shaking at 200 rpm. For long term storage glycerol stocks (25 % glycerol) were stored in cryovial at -80 °C.

Frozen chemically competent *E. coli* cells (kindly provided by technical staff, 50 µL aliquot) were thawed on ice. 1 µl of purified plasmid DNA or 10 µl of ligation mixture were added and mixed gently with the cells. The mixture was cooled on ice for 30 min followed by a heat shock at 42 °C for 10 sec (BL21, Rossetta2 and BL21 codonPlus RP) or for 40 sec (ccdB and Top 10). Afterwards, the cell suspension was placed on ice for 2 min and 250 µL SOC-Medium was added, followed by an incubation of 45 min at 37 °C under shaking (250 rpm) conditions. 50 - 200 µL of bacteria suspension

5.5.2 Yeast transformation and expression

5.5.2.1 Yeast transformation

Transformation of *S. cerevisiae* was conducted based on lithium acetate/single-stranded carrier DNA/PEG protocol. A single colony of *S. cerevisiae* was inoculated into a 10 mL YPAD starter culture and grown overnight at 28 °C with shaking at 200 rpm. The starter culture was then added to 40 mL of YPAD in a 250 ml flask and incubated at 28 °C with shaking at 200 rpm for 5 hours, after which the culture was centrifuged at 3000 x g for 5 min and the supernatant discarded. The cells were washed with 25 ml sterile H₂O and the centrifugation repeated; the pellet was then resuspended in 1 mL 0.1 M LiOAc and transferred to a 1.5 ml microfuge tube. The cells were then pelleted at 14500 x g for 15 sec and the supernatant discarded, after which the cells resuspended with 400 µL 0.1 M LiOAc and transferred 50 µL to a new 1.5 mL microfuge then pelleted again at 14500 x g for 15 sec and the supernatant discarded. Then, it was added 240 µL of PEG solution (50% (w/v) PEG 3350), 36 µL 1M LiOAc, 50 µL SS-DNA (denaturated salmon testis DNA, 2 mg/mL in TE buffer) and up to 34 µL of DNA then mixed by vortexing and incubated at 30 °C for 30 min and 42 °C for 30 min. The cells were pelleted at 6500 rpm for 15 sec then gently resuspended in 1 mL of sterilized H₂O. For each transformant 200 µL aliquots were spread on SM-URA plates and incubated at 28 °C for 3-4 days until colonies appeared.

5.5.2.1 Yeast expression

pESC-Ura vector was transformed to w303B *S. cerevisiae* cell following the procedure of 5.5.2.1. The transformation reaction onto two SM-URA plates with dextrose (D-glucose) as the carbon source. Incubate the plates at 30°C for 2-3 days. When the colony showed on the plates, 8 colonies were picked for yeast colony PCR. The positive colony was then streaked out on an SM-URA

plate, and incubate at 30°C overnight. The transformed colony was then inoculated into 50 mL of glucose-ura medium at 30°C overnight. The yeast cells were harvested and inoculated into 100 mL of galactose-ure medium for 16 hours at 30°C. The cells were harvested by centrifugation at $1000 \times g$ for 5 minutes. The cells were ready for further biochemical analysis.

5.5.3 Transformation of *Aspergillus oryzae* NSAR1

Aspergillus oryzae NSAR1 was grown on DPY agar plates or in DPY medium at 28 °C for 3-7 days (120 rpm). For long term storage *A. oryzae* spores were collected from plates by adding 3 ml distilled water, mixed with glycerol (final concentration 50 %) and stored in cryovials at -80 °C.

Spores of *A. oryzae* NSAR1 were prepared by inoculation of the fungus onto MEA plates and incubating at 30 °C for 1-2 weeks or until sporulation occurred. The spores were harvested from plates and inoculated into GN medium then incubated at 28 °C with shaking 200 rpm overnight. The culture was harvested by centrifugation at $8000 \times g$ for 10 min and the supernatant discarded. The pellet was washed once with sterilized water and once with sterile 0.8 M NaCl. The pellet was resuspended in 10 mL of filter sterilized protoplasting solution (20 mg/mL *Trichoderma* lysing enzyme and 5 mg/mL driselase in 0.8 M NaCl) and incubated at room temperature with rotary shaking. After 1- 1.5 hours, protoplasts were release from hyphae by pipetting with a wide-bore tip then filtered through 2 layers of sterile miracloth. The filtrate was centrifuged at $1000 \times g$ for 5 min and the supernatant discarded. The pellet was washed with solution 1 (0.8 M NaCl, 10 mM CaCl_2 , 50 mM Tris-HCl pH 7.5). The pellet was resuspended in 200-500 μL of solution. Then, 100 μL of protoplast was mixed gently with 5-10 μg (10 μL) of plasmid DNA and incubated on ice 2 min. The mixture was added 1 mL of solution 2 (60% (w/v) PEG 3350, 0.8 M NaCl, 10 mM CaCl_2 , 50 mM Tris-HCl pH 7.5) and incubated for 20 min. Finally, the mixture was added 40 mL of molten (50 °C) CZSB and overlaid onto 4 prepared plates which each plate containing 15 mL of CZDT. The plates were incubated at 28 °C for 3-5 days until colonies appeared.

5.5.4 Transformation of *Trichoderma viride* MF5628

One agar plug of *T. viride* was inoculated into PDB medium and incubated at 28 °C and 200 rpm for 10 days. Then the culture ($\text{OD}_{600} = 0.3$) was subcultured into PDB and incubated with shaking incubator at 200 rpm, 28 °C for 2 days. The mycelia was collected from the culture by filtration using sterile miracloth as a filter then rinsed with sterilized distilled water and 0.8 M MgSO_4 . The mycellia was added to 10 mL of sterilized KCl 0.7 M containing 200 mg of driselase and 300 mg of *Trichoderma* lysing enzyme and incubated at 0.8 M MgSO_4 , 200 rpm. After 3 hours, the

mixture was filtered using sterilized miracloth. The filtrate was centrifuged at 3,000 x g for 3 minutes. The pellet was rinsed with 10 mL of solution STC (STC: Sorbitol 1.2 M, 10 mM of Tris-HCl pH 7.5, 50 mM of CaCl₂·2H₂O) and centrifuged at 3,000 x g for 3 minutes. The pellet was dissolved with 500 µL solution STC then the protoplast contained in the aliquot was counted under phase contrast microscope.

For transformation, 200 µL of the aliquot was added to 20 µL of plasmid and mixed for 2-10 min then added to 200 µL of 25 µL PEG solution and incubated on ice for 40 minutes then added to 200 µL PEG solution (50% (w/v) PEG 8000, 50 mM CaCl₂, 10 mM Tris-HCl pH 7.5). After incubation at 25°C for 40 minute, the aliquot was added to 500 µL of STC and plated out on 15 mL of soft malt extract agar containing 1M sorbitol and incubated at 25°C. After 16-18 hours, it was overlaid with 15 ml of MEA containing 75 µg/mL of hygromycin. After the growth, the transformant was subcultured 2 or 3 times under selection conditions.

5.6 Chemistry methods

5.6.1 *Trichoderma viride* extraction

A plug of fungi (*Trichoderma viride* MF5628, ATCC 74084) was inoculated into 50 mL producing medium in 500 mL Erlenmeyer flask and incubated at 200 rpm and 28°C for 7 days. Each culture was then homogenized by hand blender and twice extracted with ethyl acetate with ratio = 1:1 v/v. The organic extracts were pooled and dried using MgSO₄ and filtered then dried by rotary evaporator. The extract was dissolved in methanol and de-fatted using n-hexane (2 times) then dried. Finally, the extract was dissolved in HPLC grade methanol and analysed by LCMS with solvent system 10-90% acetonitrile containing 0.05 % formic acid for 15min.

5.6.2 Extraction of *Aspergillus oryzae* NSAR1

A. oryzae NSAR1 was grown on malt extract medium until sporulation (3-7 days), after which the spores were harvested and added to 5 ml sterilized H₂O and grown for 6 days at room temperature with shaking at 200 rpm. Cultures was homogenized by hand blender and extracted with ethyl acetate with ratio = 1:1 for 2 times. The organic layer was dried MgSO₄ and filtered then dried by rotary evaporator. The extract was dissolved in methanol and de-fatted using n-hexane (2 times) then dried. Finally, the extract was dissolved in HPLC grade methanol and analysed by LCMS with gradient solvent system 10-90% acetonitrile containing 0.05 % formic acid for 30 min.

5.6.4 LCMS

5.6.4.1 Analytical LCMS

LCMS data were obtained with using a Waters LCMS system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex Kinetex column (2.6 μ , C18, 100 Å, 4.6 \times 100 mm) equipped with a Phenomenex Security Guard precolumn (Luna C5 300 Å) eluted at 1 mL/min. Detection was by Waters 2998 Diode Array detector between 200 and 600 nm; Waters 2424 ELSD and Waters SQD-2 mass detector operating simultaneously in ES+ and ES- modes between 100 m/z and 650 m/z . Solvents were: A, HPLC grade H₂O containing 0.05% formic acid; B, HPLC grade MeOH containing 0.045% formic acid; and C, HPLC grade CH₃CN containing 0.045% formic acid. Gradients were as follows. *Method 1*. Kinetex/CH₃CN: 0 min, 10% C; 10 min, 90% C; 12 min, 90% C; 13 min, 10% C; 15 min, 10% C.

5.6.4.2 Semi-Preparative LCMS and Compound Purification.

Purification of all compounds was generally achieved using a Waters mass-directed autopurification system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex Kinetex Axia column (5 μ , C18, 100 Å, 21.2 \times 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna C5 300 Å) eluted at 20 ml/min at ambient temperature. Solvent A, HPLC grade H₂O + 0.05% formic acid; Solvent B, HPLC grade CH₃CN + 0.045% formic acid. The post-column flow was split (100:1) and the minority flow was made up with HPLC grade MeOH + 0.045% formic acid to 1 ml·min⁻¹ for simultaneous analysis by diode array (Waters 2998), evaporative light scattering (Waters 2424) and ESI mass spectrometry in positive and negative modes (Waters SQD-2). Detected peaks were collected into glass test tubes. Combined fractions were evaporated (vacuum centrifuge), weighed, and residues dissolved directly in solvent for use or analysis.

5.6.5 High Resolution Mass Spectrometry (HRMS)

High Resolution Mass Spectrometry (HRMS) was performed on a Q-ToF Premier mass spectrometer (Waters) coupled to an Acquity UPLC-domain. Positive and negative ions were measured for accurate mass analysis.

5.6.6 Nuclear Magnetic Resonance (NMR) Analysis

NMR measurements were done by the NMR department of Organic Chemistry Department, Leibniz University Hannover. ¹H and ¹³C NMR spectra were recorded on either on a Bruker DRX 400 or 500 MHz. Chemical shifts are given in parts per million (ppm) relative to the TMS (tetramethylsilane) standard and the coupling constants J were shown in Hz. 2D experiments were obtained for complete structural elucidation including Correlation Spectroscopy (COSY),

Heteronuclear Single Quantum Coherence (HSQC), and Heteronuclear Multiple Bond Correlation (HMBC).

5.7 Molecular biology methods

5.7.1 Oligonucleotides

All oligonucleotides used in this work were purchased from Sigma Aldrich (Table 5.7.1.1).

Primer	Template	Direction	sequence 5' - 3'	Amplification	ID
CS-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCTCTGCTCAAGAGCCCTCAACTTTT	tvCS-LF	SYp201
CS-LR	T.v gDNA	rev	CTCGACAGACGTCGCGGTGAGTTCAGGCATCGCCGAAATGCCAGCTAGAAGCATA	tvCS-LF	SYp202
CS-RF	T.v gDNA	fwd	CCCAGCACTCGTCCGAGGGCAAAGGAATAGAATGCTGATCACGGCATGACCAACT	tvCS-RF	SYp203
CS-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGGATATAGTGCATCATCCCATGCCAT	tvCS-RF	SYp204
HygF	pTH-GS-egfp	fwd	TCTAGTGGATCTTTCGACACT	hyg cassette	SYp205
HygR	pTH-GS-egfp	rev	TCGAGTGGAGATGTGGAG	hyg cassette	SYp206
CSLF-F	pE-YA-tvCSKO	fwd	TCCTGCTCAAGAGCCCTCAACTTTT	transtvCS-LF	SYp207
CSRF-R	pE-YA-tvCSKO	rev	GTGTCATCATCCCATGCCAT	transtvCS-RF	SYp208
THygF	pE-YA-tv(x)KO	fwd	AATTCAGCGAGAGCCT	transtv(x)-RF	SYp209
THygR	pE-YA-tv(x)KO	rev	TGTCGTCCATCACAGTTT	transtv(x)-LF	SYp210
CS-1	CSKO gDNA	fwd	AGCGTGATGAGCCCTCTACTG	KO confirming	SYp211
CS-2	CSKO gDNA	fwd	TGCTTGATTGGGAACCCTGACT	KO confirming	SYp212
CS-3	CSKO gDNA	rev	CATTCGACAGCGCCG	KO confirming	SYp213
CS-4	CSKO gDNA	rev	AGGCAAACAGTCTTTGCACT	KO confirming	SYp214
HCS-1	CSKO gDNA	fwd	GGGATACGTACCGGAAA	KO confirming	SYp215
HCS-2	CSKO gDNA	rev	AAGCTGTTGGCAGCCTTAAA	KO confirming	SYp216
HCS-3	CSKO gDNA	fwd	TACTCTGATAGCTTGACTATGAAA	KO confirming	SYp217
HCS-4	CSKO gDNA	rev	AAGAAGCATGGTTCTATTTCTAT	KO confirming	SYp218
tvR4-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCATGAAGCACGAGGAGCT	tvR4-LF	SYp219
tvR4-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAGATCCACTAGAGTGTATTCGGTGGACCT	tvR4-LF	SYp220
tvR4-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGATTCCCTGAACCCCAATACA	tvR4-RF	SYp221
tvR4-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGGTCAAGCAAACGGATGTT	tvR4-RF	SYp222
tvR4LF-F	pE-YA-tvR4KO	fwd	CCGCATGAAGCACGAGGAGCT	transtvR4-LF	SYp223
tvR4RF-R	pE-YA-tvR4KO	rev	CGGTCAAAGCAAACGGATGTT	transtvR4-RF	SYp224
HR4-1	tvR4KO gDNA	fwd	CCCTTCATCTTCATTTG	KO confirming	SYp225
HR4-2	tvR4KO gDNA	rev	AAGCTGTTGGCAGCCTTAAA	KO confirming	SYp226
HR4-3	tvR4KO gDNA	fwd	TACTCTGATAGCTTGACTATGAAA	KO confirming	SYp227
HR4-4	tvR4KO gDNA	rev	TCTTGGAGTCCACCT	KO confirming	SYp228
tvR2-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCGGTGAAGCTCTATCTACAATT	tvR2-LF	SYp229
tvR2-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAGATCCACTAGATTCAACTGTGGTGTGAG	tvR2-LF	SYp230
tvR2-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGACCAACAGTAAGTTGCTCT	tvR2-RF	SYp231

V Experimental

tvR2-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGATACCATACATCATTGAGGC	tvR2-RF	SYp232
tvR2LF-F	pE-YA-tvR2KO	fwd	GCGGTGAAGCTCTATCTACAATT	transtvR2-LF	SYp233
tvR2RF-R	pE-YA-tvR2KO	rev	GATACCATACATCATTTGAGGC	transtvR2-RF	SYp234
tvR3-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCTTCAGCGTCAAATTTTACGG	tvR3-LF	SYp235
tvR3-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAAGTCCACTAGAAAGTGTCTATGCCATATTTTC	tvR3-LF	SYp236
tvR3-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGATGTCCAAGTCCCGTAATAT	tvR3-RF	SYp237
tvR3-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGATGGCACAACCTTCGG	tvR3-RF	SYp238
tvR3LF-F	pE-YA-tvR3KO	fwd	GCTTCAGCGTCAAATTTTACGG	transtvR3-LF	SYp239
tvR3RF-R	pE-YA-tvR3KO	rev	ATGGCACAACCTTCGG	transtvR3-RF	SYp240
tvL2-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCCTCAATCATCAAACCAAC	tvL2-LF	SYp241
tvL2-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAAGTCCACTAGAACATACAAAGTAGTCAGAGG	tvL2-LF	SYp242
tvL2-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGACAATTACCCCTATACACAGAAGC	tvL2-RF	SYp243
tvL2-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGAGCAGTATTTAAAATGGCACA	tvL2-RF	SYp244
tvL2LF-F	pE-YA-tvL2KO	fwd	GCGCAATCATCAAACCAAC	transtvL2-LF	SYp245
tvL2RF-R	pE-YA-tvL2KO	rev	CAGTATTTAAAATGGCACA	transtvL2-RF	SYp246
tvL3-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCCTTGAATAGACTACCGTGA	tvL3-LF	SYp247
tvL3-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAAGTCCACTAGAGGATCTCTCTTTGTTGGC	tvL3-LF	SYp248
tvL3-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGATAAAGCTGTATTTATCCAAATGG	tvL3-RF	SYp249
tvL3-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGCTGGGTAGAATGTAGATTTGT	tvL3-RF	SYp250
tvL3LF-F	pE-YA-tvL3KO	fwd	TTGAATAGACTACCGTGA	transtvL3-LF	SYp251
tvL3RF-R	pE-YA-tvL3KO	rev	CTGGGTAGAATGTAGATTTGT	transtvL3-RF	SYp252
tvL4-LF	T.v gDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCCAATTCATCTACAGATGCAT	tvL3-LF	SYp253
tvL4-LR	T.v gDNA	rev	ACGTATTTTCAGTGTGCAAAAGTCCACTAGACATAGTTATCATCTGGTCAACC	tvL3-LF	SYp254
tvL4-RF	T.v gDNA	fwd	TGTAAGCGCCCACTCCACATCTCCACTCGAAAATGCAGACTCTGCTGC	tvL3-RF	SYp255
tvL4-RR	T.v gDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCCGACTGCCTTCTTGATGAACG	tvL3-RF	SYp256
tvL4LF-F	pE-YA-tvL4KO	fwd	CAATTCATCTACAGATGCAT	transtvL3-LF	SYp257
tvL4RF-R	pE-YA-tvL4KO	rev	CTGCCTTCTTGATGAACG	transtvL3-RF	SYp258
dsRNA-pkiF	T.r gDNA	fwd	GTA AACACAACATATCCAGTCATATTGGCaggaaccgggggggat	PpkiF	SYp259
dsRNA-pkiR	T.r gDNA	rev	CGAGACGGACCAACGGCACAACCACCCATCggttaagagggttctccgg	PpkiF	SYp260
dsRNA-re F	T.v cDNA	fwd	GATGGGTGGTTGTGCCGT	dsRNA	SYp261
dsRNA-reR	T.v cDNA	rev	AGTGTAAGCCTGGGGTGCCTAATGCGGCCTACCGGCTCGGTAAC	dsRNA	SYp262
1argCSP1F	T.v cDNA	fwd	TTTCTTTCAACACAAGATCCCAAAGTCAAATGGATGTCAATAAAACTCTACAGT	tvCS	SYp263
2argCST3R	T.v cDNA	rev	GGTTGGCTGGTAGACGTATATAATCATACTTAGGCGTCATCGTGTGA	tvCS	SYp264
3argCSP2F	T.v cDNA	fwd	TAACAGCTACCCCGCTTGAGCAGACATCACATGGATGTCAATAAAACTCTACAGT	tvCS	SYp265
4argPhaP1F	T.v cDNA	fwd	TTTCTTTCAACACAAGATCCCAAAGTCAAATGTATACCCCGTACATCACA	tvR3	SYp266
5argPhaT1R	T.v cDNA	rev	TTCATTCTATGCGTTATGAACATGTTCCCTCTAACATCTGGGGTGAAAGG	tvR3	SYp267
6argCSP3F	T.v cDNA	fwd	GTCGACTGACCAATTCGCGAGCTCGTCAAATGGATGTCAATAAAACTCTACAGT	tvCS	SYp268
7argAcOP2F	T.v cDNA	fwd	TAACAGCTACCCCGCTTGAGCAGACATCACATGCTGACGGAAGTCG	tvL3	SYp269

V Experimental

8argACoT2R	T.v cDNA	rev	ACGACAATGTCCATATCATCAATCATGACCTTAAGAAACAAGCAGAGGAAAC	tvL3	SYp270
9adeP450P1F	T.v cDNA	fwd	TTTCTTTCAACACAAGATCCCAAAGTCAAATGATCTTGGAAATTGATCGAATT	tvL4	SYp271
10adeP450T3R	T.v cDNA	rev	GGTTGGCTGGTAGACGTCATATAATCATACCTAAATGTATGAGTCTCGCC	tvL4	SYp272
11adeMoP1F	T.v cDNA	fwd	TTTCTTTCAACACAAGATCCCAAAGTCAAATGCTGCGGTTTCAGC	tvR2	SYp273
12adeMoT3R	T.v cDNA	rev	GGTTGGCTGGTAGACGTCATATAATCATACTCAGATAAATCCAAGAATCCGC	tvR2	SYp274
13adeP450T2R	T.v cDNA	rev	ACGACAATGTCCATATCATCAATCATGACCTTAATGTATGAGTCTCGCC	tvL4	SYp275
14adeMoP3F	T.v cDNA	fwd	GTCGACTGACCAATCCGCAGCTCGTCAAATGCTGCGGTTTCAGC	tvR2	SYp276
15adeP450T1R	T.v cDNA	rev	TTCATTCTATGCGTTATGAACATGTTCCTCTAAATGTATGAGTCTCGCC	tvL4	SYp277
16adeDHP2F	T.v cDNA	fwd	TAACAGCTACCCCGCTTGAGCAGACATCACATGGGCGTCCCAACA	tvL2	SYp278
17adeDHT2R	T.v cDNA	rev	ACGACAATGTCCATATCATCAATCATGACCTTATGCGTTTATAGTGATCGAGA	tvL2	SYp279
18argACoP1 F	T.v cDNA	fwd	TTTCTTTCAACACAAGATCCCAAAGTCAAATGCTGACGGAAGTCG	tvL3	SYp280
peyatRNA F	T.v cDNA	fwd	GCCAACTTTGTACAAAAAGCAGGCTCCGCATGAAGCAGGAGGCT	tvR4	SYp281
peyatRNA R	T.v cDNA	rev	TGCCAACTTTGTACAAGAAAGCTGGGTCGGTCAAGCAAACGGATGTTG	tvR4	SYp282

Table 5.7.1.1 Primers used in Chapter 2.

Primer	Template	Direction	sequence 5' - 3'	Amplification	ID
28a-5F	pRSETa-bfL5	fwd	ATGGGTCGCGGATCCGAATTCATGATTGGCTATATTGTG	F28a5	SYp301
28a-5R	pRSETa-bfL5	rev	GCTCGAGTGC GGCCGCAAGCTTTCAGGTGCTGCAGAAA	F28a5, F28asig5	SYp302
28a-6F	pRSETa-bfL6	fwd	ATGGGTCGCGGATCCGAATTCATGTTGAGCAGCAAAGTG	F28a6	SYp303
28a-6R	pRSETa-bfL6	rev	GCTCGAGTGC GGCCGCAAGCTTCTAATAATGGCCATATTT	F28a6, F28asig6	SYp304
28a-9F	pRSETa-bfL9	fwd	ATGGGTCGCGGATCCGAATTCATGAGCATTCTGGCGTAT	F28a9	SYp305
28a-9R	pRSETa-bfL9	rev	GCTCGAGTGC GGCCGCAAGCTTTTAGCGGGCCACGGGCG	F28a9	SYp306
28a-10F	pRSETa-bfL10	fwd	ATGGGTCGCGGATCCGAATTCATGTTGCGATGTTTATT	F28a10	SYp307
28a-10R	pRSETa-bfL10	rev	GCTCGAGTGC GGCCGCAAGCTTTCATCAATATCCGGCGG	F28a10, F28asig10	SYp308
28asignal-5F	pRSETa-bfL5	fwd	ATGGGTCGCGGATCCGAATTCATGAGACCCCGCGGCTAT	F28asig5	SYp309
28asignal-6F	pRSETa-bfL6	fwd	ATGGGTCGCGGATCCGAATTCATGCGCAGCATTCTGCCGAAT	F28asig6	SYp310
28asignal-10F	pRSETa-bfL10	fwd	ATGGGTCGCGGATCCGAATTCATGCGATTAGCCTGGCGCGCAA	F28asig10	SYp311
urabfL3F	pRSETa-bfL3	fwd	GGAGAAAAAACC CGGATCCATGTCTACCGAAAAATACGAC	Fyeast3	SYp312
urabfL3R	pRSETa-bfL3	rev	GCTAGCCGCGGTACCAAGCTTCTACAGCGCCAGGG	Fyeast3	SYp313
urapvL2F	pRSETa-pvL2	fwd	GGAGAAAAAACC CGGATCCATGCAGGCGTATGATCA	Fyeastpv12	SYp314
urapvL2R	pRSETa-pvL2	rev	GCTAGCCGCGGTACCAAGCTTTTACAGTTTCACTTCGCTG	Fyeastpv12	SYp315
urabfL5F	pRSETa-bfL5	fwd	GGAGAAAAAACC CGGATCCATGATTGGCTATATTGTG	Fyeast5	SYp316
urabfL5R	pRSETa-bfL5	rev	GCTAGCCGCGGTACCAAGCTTTCAGGTGCTGCAGAAAT	Fyeast5	SYp317
urabfL6F	pRSETa-bfL6	fwd	GGAGAAAAAACC CGGATCCATGTTGAGCAGCAAAGTG	Fyeast6	SYp318
urabfL6R	pRSETa-bfL6	rev	GCTAGCCGCGGTACCAAGCTTCTAATAATGGCCATATTCGCC	Fyeast6	SYp319
urabfL9F	pRSETa-bfL9	fwd	GGAGAAAAAACC CGGATCCATGAGCATTCTGGCGTAT	Fyeast9	SYp320
urabfL9R	pRSETa-bfL9	rev	GCTAGCCGCGGTACCAAGCTTTAGCGGGCCACGGGCG	Fyeast9	SYp321
urabfL10F	pRSETa-bfL10	fwd	GGAGAAAAAACC CGGATCCATGTTGCGATGTTTATT	Fyeast10	SYp322
urabfL10R	pRSETa-bfL10	rev	GCTAGCCGCGGTACCAAGCTTTCATCAATATCCGGCGG	Fyeast10	SYp323

Table 5.7.1.2 Primers used in Chapter 3.

V Experimental

Primer	Template	Direction	sequence 5' - 3'	Amplification	ID
tenSF1	pE-YA-tenS	fwd	GGTCCTTGCTGAAGAGTTG	F35	SYp401
tenSR1	pE-YA-tenS	rev	GGATATCACAAGCAAGAAGC	F35	SYp402
TDSLk1-F	pE-YA-tenS	fwd	AGATCAGCAGGATAAGCAGC	F36, F41, F43, F48, F51, F53, F58	SYp403
TDSLk1-R	pE-YA-tenS	rev	AAGCCACGGGTCTGGAG	F36, F51	SYp404
TDSLk2-F	pE-YA-tenS	fwd	GCCGCATCGGCGGTAGAC	F44, F52	SYp405
TDSLk2-R	pE-YA-tenS	rev	TCCTTTGGTGGTGGTGATG	F37, F39, F44, F46, F52, F54, F57	SYp406
T2A-F	pE-YA-tenS	fwd	GCCATTCTGAATAATACAGGCC	F37	SYp407
T2A-R	pE-YA-tenS	rev	CTCAGAGACACTCATGACTCG	F48	SYp408
T1B-F	pE-YA-tenS	fwd	ACTGTGGTGGACATGATTCG	F39	SYp409
T1B-R	pE-YA-tenS	rev	AGCGCTCGAGCTTAGCAA	F43	SYp410
T2B-F	pE-YA-tenS	fwd	ACGGATGTGCATCATGCCTT	F46	SYp411
T1A-R	pE-YA-tenS	rev	CTGCACAGAGTCTTTGCTGC	F41	SYp412
D1A-TLk-F	pE-YA1-dmbS	fwd	ACTGTACCGCCCTCCAGACCCGTGGGCTTTTCAAGAGCGACAGGACCTA	F38, F40	SYp413
D1A-TLk-R	pE-YA1-dmbS	rev	CATGGTGGCACGAATCATGTCCACCACAGTCTGCACAGAGTCTTTGTTGC	F40	SYp414
D1B-TLk-F	pE-YA1-dmbS	fwd	ATGGACGCTTGACGAAAGACTCTGTGCAGACTGTCGTGGATACGATTCG	F42	SYp415
D1B-TLk-R	pE-YA1-dmbS	rev	GTTTGACTGGCCTGTATTATTCAGAATGGCAGCGCAGAACCAAGCAGAA	F38, F42	SYp416
D2A-TLk-F	pE-YA1-dmbS	fwd	GACTTTTTTGTCTTGTCTAAGCTCGAGCGCTGCCATCTTGAATAACATGGG	F45, F47	SYp417
D2A-TLk-R	pE-YA1-dmbS	rev	CGCCTCAGCAAAGGCATGATGCACATCCGTCTCAGAGAGCCTCATAGCTC	F47	SYp418
D2B-TLk-F	pE-YA1-dmbS	fwd	GGTACCACGCGAGTCATGAGTGTCTCTGAGACTGACGTGCATCACGCCTT	F49	SYp419
D2B-TLk-R	pE-YA1-dmbS	rev	CGTAGACTGTCTACCGCGATGCGGCGGCTGCTTGCCCGAGGCAATCA	F45, F49	SYp420
TMS-F	PE-YA1-milS	fwd	ACTGTACCGCCCTCCAGCCCGTGGGCTTTCCAGAGCGACAAACCT	F50	SYp421
TMS-R	PE-YA1-milS	rev	CGTAGACTGTCTACCGCATGCGGCGGCGCTTCTCAGAAGTAACCG	F50	SYp422
TDSLK1(3/4)-R	pE-YA-tenS	rev	GTTTGACTGACCTATGTTGTCGGATAGTAGCGCTCGAGCTTAG	F53	SYp423
TDSLK1(3/4)-F	pE-YA-tenS	fwd	GACGTCATGCGAGCCACGACTCTCGGAGACGGATGTGCATCATG	F54	SYp424
TenS(3/4)-F	PE-YA1-milS	fwd	ACTATCGCGAACAACATAG	F55	SYp425
TenS(3/4)-R	PE-YA1-milS	rev	CTCCGAGAGTGTCTG	F55	SYp426
LK2sub-F	pE-YA-tenS	fwd	CACAGCAACCGAGACGTATGCGAGCCACGACTCTCTGAGACGGATGTGC	F57	SYp427
LK1sub-R	pE-YA-tenS	rev	GAGTGTCTGGTCTCGCATGACGTCTCGGTTGCTGTGCACCTTGGTGTCTCA	F58	SYp428

Table 5.7.1.3 Primers used in Chapter 4.

5.7.2 DNA isolation and PCR

5.7.2.1 Isolation of Small Scale gDNA

The GenElute Plant Genomic DNA miniprep kit (Sigma) was used to extract gDNA from fungi. Isolation of gDNA was carried out according to the manufacturer's protocol.

5.7.2.2 Isolation of Large Scale gDNAs

A plug of *S. tenacellus* was inoculated into 100 mL of MEB and incubated on shaker incubator at 200 rpm and 25 °C. After 15 day, the mycelia were collected by filtration using a filter paper then squeezed. Approximately 6 g of squeezed mycelia were ground to a fine powder under liquid nitrogen. The powder was added to 15 mL of DNA extraction buffer (Tris-HCl, pH 8.0, 10 mM; EDTA 10mM; SDS 0.5%) and mixed gently by inversion then added with 10ml of phenol:iso-amyl alcohol (25:24:1) and shaken gently for 15 minutes then centrifuged at 8.500 rpm (biofuge primo R from heraeus) for 10- 20 minutes. This step was carried out two times or until the clear interface between the aqueous and organic phases. Then, 10 mL aqueous phase was extracted with 10 mL phenol:chloroform:iso-amyl alcohol (25:24:1) and shaken gently for 25-30 minutes then centrifuge at 8.500 x g for 10-20 minutes. The 9.5 mL aqueous phase was added 950 µL 3M sodium acetate and 25 mL absolute ethanol (pH 5.2) then centrifuged at 8.000 x g 10 minute. The pellet was rinsed with 10 mL ethanol 70% then centrifuged at 8.000 x g for 2 minutes and then dissolved in 1 mL TE buffer (Tris-HCl, pH 8, 10 mM; EDTA 1mM). Genomic DNA was purified by equilibrium centrifugation in a continuous CsCl-ethidium bromide gradient. Finally, the quality of genomic DNA were analysed by electrophoresis.

5.7.2.3 Isolation of Small Scale DNA Plasmid from E. coli

GeneJET Plasmid Miniprep Kit was used to isolate small scale DNA plasmid from *E. coli*. Isolation was performed based on manufacturer's protocol.

5.7.2.4 Isolation of DNA Plasmid from S. cerevisiae

The Zymoprep II yeast plasmid miniprep kit (Zymo research) was used to extract plasmid DNA from *S. cerevisiae*. Yeast cell walls were digested using the Zymolyase enzyme according to the manufacturer's protocol. Primary transformation plates containing multiple colonies were used as the source of yeast cells for plasmid extraction

5.7.2.5 Analytical or Preparative PCR

Analytical PCR was performed using reagent KOD Hot Start DNA polymerase throughout (Novagen). Reagent composition for 25 µL reaction was as follow 0.5 uL of KOD Hot Star DNA polymerase (Novagen), 2.5 µL of enzyme buffer, 2.5 µL of dNTPs, 1.5 µL of MgSO₄, 2.5 µL of

each primer 3 μ M, 0.5 μ L of DMSO and distilled water until 25 μ L in total. The DNA template was added at 1 pg – 10 ng (low complexity) and 50 - 250 ng (high complexity). The PCR condition was initially denatured by heating at 95 °C for 2 min, followed by 35 cycles of amplification, i.e. denaturation at 95 °C for 30 s, 66 °C for 20 s (depend on the best annealing and long product), 70 °C for 15s and final extension 70 °C for 10 min.

5.7.2.6 Colony PCR or Analytical PCR

Colonies of *E. coli* transformants were screened by PCR. Cells of *E. coli* were picked by sterilized toothpick and resuspended into PCR reaction. PCR reaction consists of 5 μ L of Biomix, 0.5 μ L of each reverse and forward primer, and 4 μ L of deionized water on in 10 μ L volume total. PCR program were performed as follow: 95°C for 5 min then 35 cycles with 95 °C for 15 s, 55° for 30 s, 72 °C for 15-20 s/kb, finally 72°C for 10 min.

5.7.2.7 Restriction Analysis of Plasmid DNA

Approximately 2 μ L of plasmid DNA (0.1- 0.2 μ g) was added 1 μ L of enzyme buffer, 0.25 μ L of certain restriction enzyme and sterilized water until final volume of 10 μ L. The mixture was incubated at 37 °C for an hour. DNA electrophoresis was used to check the digested DNA.

5.7.2.8 DNA Electrophoresis

Agarose gels were prepared in TAE buffer with agarose at concentrations between 0.8 – 1.5% (depend on fragment size) and with Midori green advance added for DNA visualisation. DNA was loaded on gels either using loading dye already present within a reaction (FastDigest Green Buffer (Fermentas)) or after addition of 0.1 vol 10 x loading buffer (0.25% (w/v) bromophenol blue in 10% (v/v) glycerol). Electrophoresis was performed using TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA) and run at 200 volt and 100 mA for 20 min. Gels were visualised by UV light using a transilluminator.

5.7.3 Cloning procedure

5.7.3.1 Restriction enzyme digestion

Restriction enzyme digestions (single or double) of plasmids or PCR products were performed following the manufacturer's guidelines of the enzymes (New England Biolabs or Thermo Scientific). DNA sample were mixed with the enzyme or enzyme mixture and the appropriate buffer. The reaction mixture was incubated at 25 °C or 37 °C (depending on the restriction enzyme) for 1 to 12 h. Enzymes were inactivated by heating the sample at 65 °C or 80 °C for 20 min. Digested DNA or plasmid were purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) as described previously.

5.7.3.2 Dephosphorylation and ligation

Digested plasmids were dephosphorylated using shrimp alkaline phosphatase (SAP) to prevent self ligation. 1 µl of enzyme was added after digestion to the reaction mixture and incubated at 37 °C for 1 h. T4 DNA ligase (New England Biolabs) and its appropriate buffer were used for ligation reaction (final volume 10 µl) following the manufacturer's introductions. Molar ration of 1:5 plasmid to insert was used. Ligation reactions were incubated at 4 °C overnight or at room temperature for 30 min followed by heat inactivation at 65 °C.

5.7.3.3 Gateway cloning

LR-recombination between the entry clone and the destination vector was performed using Gateway LR Clonase enzyme mix II kit (Invitrogen) following the manufacturer's instructions. For transformation of competent *E. coli* T10 cells 1 - 2 µl of the reaction mixture was used.

5.7.3 Obtained and constructed vectors in this thesis

Construction of fungal and bacterial expression vectors was performed using the described *in vitro* cloning procedures (section 5.7.3) and *in vivo* homologous recombination in *S. cerevisiae* (section 5.5.2).

Name	Selective marker	Origin
pE-YA	Kan	Lazarus Group, Bristol
pTYGSargB	Amp, Δ Arg	Lazarus Group, Bristol
pTYGSade	Amp, Δ Ade	Lazarus Group, Bristol
pTYGSmet	Amp	Lazarus Group, Bristol
pTH-GS-egfp	Amp, Hyg	Lazarus Group, Bristol
pET28a (+)	Kan	Hahn group, Hannover
pG-KGE8	Cm	Takara Bio
pESC-URA	Amp, Δ Ura	Prof. Jianqiang Kong

Table 5.7.3.1 Vectors kindly provided or purchased for this work

V Experimental

Name	Insert Fragments	Selective Marker	ID
pE-YA-tvCSKO	tvCS KO cassette	Kan	SYv01
pE-YA-tvL2KO	tvL2 KO cassette	Kan	SYv02
pE-YA-tvL3KO	tvL3 KO cassette	Kan	SYv03
pE-YA-tvL4KO	tvL4 KO cassette	Kan	SYv04
pE-YA-tvR2KO	tvR2 KO cassette	Kan	SYv05
pE-YA-tvR3KO	tvR3 KO cassette	Kan	SYv06
pE-YA-tvR4KO	tvR4 KO cassette	Kan	SYv07
pTYH-GS-pki-reGOI	<i>P_{pki}</i> , reGOI	Amp	SYv08
pTYH-GS-pki-eGFP	<i>P_{pki}</i> , eGFP	Amp	SYv09
pE-YA-SY2133-8	tvL4	Kan	SYv10
pTY-GSarg-SY2133-1	tvL4, tvCS	Amp, Δ Arg	SYv11
pTY-GSarg-SY2133-2	tvL4, tvCS, tvR3	Amp, Δ Arg	SYv12
pTY-GSarg-SY2133-3	tvL4, tvCS, tvR3, tvL3	Amp, Δ Arg	SYv13
pTY-GSarg-SY2133-9	tvL4	Amp, Δ Arg	SYv14
pTY-GSarg-SY2133-10	tvL4, tvCS, tvL3	Amp, Δ Arg	SYv15
pTY-GSade-SY2133-4	tvL4	Amp, Δ Ade	SYv16
pTY-GSade-SY2133-5	tvR2	Amp, Δ Ade	SYv17
pTY-GSade-SY2133-6	tvL4, tvR2	Amp, Δ Ade	SYv18
pTY-GSade-SY2133-7	tvL4, tvL2, tvR2	Amp, Δ Ade	SYv19

Table 5.7.3.2 Vectors constructed in Chapter 2.

Name	Insert Fragments	Selective Marker	ID
pRSETa-bfL1	bfL1	Amp	SYv20
pRSETa-bfL2	bfL2	Amp	SYv21
pRSETa-bfL3	bfL3	Amp	SYv22
pRSETa-pvL2	pvL2	Amp	SYv23
pRSETa-bfL5	bfL5	Amp	SYv24
pRSETa-bfL6	bfL6	Amp	SYv25
pRSETa-bfL9	bfL9	Amp	SYv26
pRSETa-bfL10	bfL10	Amp	SYv27
pET28a-bfL5	bfL5	Kan	SYv28
pET28a-bfL6	bfL6	Kan	SYv29
pET28a-bfL9	bfL9	Kan	SYv30
pET28a-bfL10	bfL10	Kan	SYv31
pET28a-bfL5(Δ 18)	bfL5(Δ 1-18)	Kan	SYv32
pET28a-bfL6(Δ 22)	bfL6(Δ 1-22)	Kan	SYv33
pET28a-bfL10(Δ 20)	bfL10(Δ 1-20)	Kan	SYv34
pESC-URA-bfL3	bfL3	Amp, Δ Ura	SYv35
pESC-URA-pvL2	pvL2	Amp, Δ Ura	SYv36
pESC-URA-bfL5	bfL5	Amp, Δ Ura	SYv37
pESC-URA-bfL6	bfL6	Amp, Δ Ura	SYv38
pESC-URA-bfL9	bfL9	Amp, Δ Ura	SYv39
pESC-URA-bfL10	bfL10	Amp, Δ Ura	SYv40

Table 5.7.3.3 Vectors constructed in Chapter 3.

Name	Insert Fragments	Selective Marker	ID
pE-YA-TDS1	TDS1 swap	Kan	SYv41
pE-YA-TDS2	TDS2 swap	Kan	SYv42
pE-YA-TDS3	TDS3 swap	Kan	SYv43
pE-YA-TDS4	TDS4 swap	Kan	SYv44
pE-YA-TDS5	TDS5 swap	Kan	SYv45
pE-YA-TDS6	TDS6 swap	Kan	SYv46
pE-YA-TMS1	TMS1 swap	Kan	SYv47
pE-YA-TMS2	TMS2 swap	Kan	SYv48
pE-YA-TMS3	TMS3 swap	Kan	SYv49
pTY-GSarg-TDS1	TDS1 swap	Amp, Δ Arg	SYv50
pTY-GSarg-TDS2	TDS2 swap	Amp, Δ Arg	SYv51
pTY-GSarg-TDS3	TDS3 swap	Amp, Δ Arg	SYv52
pTY-GSarg-TDS4	TDS4 swap	Amp, Δ Arg	SYv53
pTY-GSarg-TDS5	TDS5 swap	Amp, Δ Arg	SYv54
pTY-GSarg-TDS6	TDS6 swap	Amp, Δ Arg	SYv55
pTY-GSarg-TMS1	TMS1 swap	Amp, Δ Arg	SYv56
pTY-GSarg-TMS2	TMS2 swap	Amp, Δ Arg	SYv57
pTY-GSarg-TMS3	TMS3 swap	Amp, Δ Arg	SYv58

Table 5.7.3.4 Vectors constructed in Chapter 4.

5.7.4 Heterologous Protein Production and Purification

Each constructed plasmids was transformed into chemically competent *E. coli* BL21 (DE3) cells via heat shock transformation. A Single colony was picked to inoculate an overnight culture (LB medium with kanamycin) which was used to inoculate kanamycin containing 2TY medium. Cultures were grown to an OD₆₀₀ of 0.4-0.6 at 37 °C before protein production was induced with 0.1 mM isopropyl- β -D-thiogalactopyranoside (IPTG). The cultures were further incubated at 16 or 25°C for 20 h to increase the solubility of the produced protein. Cells were harvested by centrifugation (5000 x g, 4 °C, 15 min).

The harvested cell pellets were resuspended in 10 mL cold IMAC loading buffer. Lysis of the cells was done by sonication using a SonoPlus Ultrasonic homogenizer from Bandelin. The cells were intermittently sonicated on ice for 30 s with 30 s allowed for cooling. The total sonication time was 7 min. The cell lysate was centrifuged at 12000 rpm for 45 min at 4 °C. Samples of the supernatant and the pellet were used for SDS PAGE. The protein of interest was purified using immobilized metal ion affinity chromatography. During the whole time the cell samples were kept on ice.

Purification of histidine-tagged recombinant proteins were done by immobilized metal ion affinity chromatography using Ni-NTA agarose (Qiagen). All buffers, washing solutions and samples were filtered through a 0.45 μ m filter prior to usage. The crude lysate was passed through a Ni-NTA column (2 mL bed volume) Firstly, the column was washed with 10 mL loading buffer.

V Experimental

Elution was performed with a stepwise gradient of imidazole (20 - 500 mM). 2 mL fractions were collected over the course of elution and analysed using SDS-PAGE. The eluate was concentrated and the binding buffer was replaced by storage buffer (50 mM potassium phosphate buffer, pH 8.0) using amicon ultra centrifugal filter units, cut off 10000 MW, from Merck. The purified enzyme was stored at -20 °C or immediately used for enzyme assays.

Reference

- [1] Williams, Katherine, et al. "Genetic and chemical characterisation of the cornexistin pathway provides further insight into maleidride biosynthesis." *Chemical Communications* 53.56 (2017): 7965-7968.
- [2] Szwalbe, Agnieszka J., et al. "Novel nonadride, heptadride and maleic acid metabolites from the byssochlamic acid producer *Byssochlamys fulva* IMI 40021—an insight into the biosynthesis of maleidrides." *Chemical Communications* 51.96 (2015): 17088-17091.
- [3] Williams, Katherine, et al. "Heterologous production of fungal maleidrides reveals the cryptic cyclization involved in their biosynthesis." *Angewandte Chemie International Edition* 55.23 (2016): 6784-6788.
- [4] Mandala, Suzanne M., and Guy H. Harris. "Isolation and characterization of novel inhibitors of sphingolipid synthesis: Australifungin, viridiofungins, rustmicin, and khafrefungin." *Methods in enzymology*. Vol. 311. Academic Press, 2000. 335-348.
- [5] Pollex, Annett, et al. "Total synthesis of (3S, 4S, 2' S)-and (3R, 4R, 2' S)-viridiofungin A triester." *Tetrahedron letters* 45.37 (2004): 6915-6918.
- [6] Harris, Guy H., et al. "Isolation and structure elucidation of viridiofungins A, B and C." *Tetrahedron letters* 34.33 (1993): 5235-5238.
- [7] Wat, Chi-Kit, et al. "The yellow pigments of *Beauveria* species. Structures of tenellin and bassianin." *Canadian Journal of Chemistry* 55.23 (1977): 4090-4098.
- [8] Eley, Kirstin L., et al. "Biosynthesis of the 2-pyridone tenellin in the insect pathogenic fungus *Beauveria bassiana*." *ChemBioChem* 8.3 (2007): 289-297.
- [9] Halo, Laura M., et al. "Late stage oxidations during the biosynthesis of the 2-pyridone tenellin in the entomopathogenic fungus *Beauveria bassiana*." *Journal of the American Chemical Society* 130.52 (2008): 17988-17996.
- [10] Fisch, Katja M., et al. "Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite." *Journal of the American Chemical Society* 133.41 (2011): 16635-16641.
- [11] Pickens, Lauren B., Yi Tang, and Yit-Heng Chooi. "Metabolic engineering for the production of natural products." *Annual review of chemical and biomolecular engineering* 2 (2011): 211-236.
- [12] Clardy, Jon, and Christopher Walsh. "Lessons from natural molecules." *Nature* 432.7019 (2004): 829-837.
- [13] David, Bruno, Jean-Luc Wolfender, and Daniel A. Dias. "The pharmaceutical industry and natural products: historical status and new trends." *Phytochemistry Reviews* 14.2 (2015): 299-315.
- [14] Chan, Yolande A., et al. "Biosynthesis of polyketide synthase extender units." *Natural product reports* 26.1 (2009): 90-114.
- [15] Elovson, John, and P. Roy Vagelos. "Acyl carrier protein X. Acyl carrier protein synthetase." *Journal of Biological Chemistry* 243.13 (1968): 3603-3611.
- [16] Prescott, David J., and P. Roy Vagelos. "Acyl carrier protein." *Adv Enzymol Relat Areas Mol Biol* 36 (1972): 269-311.
- [17] McMurry, John, Tadhg P. Begley. *The organic chemistry of biological pathways*. Roberts and Company Publishers, 2005.
- [18] Zhou, Pei, Galina Florova, and Kevin A. Reynolds. "Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis." *Chemistry & biology* 6.8 (1999): 577-584.

- [19] Wakil, Salih J. "Fatty acid synthase, a proficient multifunctional enzyme." *Biochemistry* 28.11 (1989): 4523-4530.
- [20] Bailey, Andrew M., et al. "Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: first observation of a reductive release mechanism during polyketide biosynthesis." *Chemical communications* 39 (2007): 4053-4055.
- [21] Schümann, Julia, and Christian Hertweck. "Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes." *Journal of biotechnology* 124.4 (2006): 690-703.
- [22] Du, Liangcheng, and Lili Lou. "PKS and NRPS release mechanisms." *Natural product reports* 27.2 (2010): 255-278.
- [23] Maier, Timm, et al. "Structure and function of eukaryotic fatty acid synthases." *Quarterly reviews of biophysics* 43.3 (2010): 373-422.
- [24] Bukhari, Habib ST, Roman P. Jakob, and Timm Maier. "Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase." *Structure* 22.12 (2014): 1775-1785.
- [25] Foulke-Abel, Jennifer, and Craig A. Townsend. "Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase." *ChemBioChem* 13.13 (2012): 1880-1884.
- [26] Hendrickson, Lee, et al. "Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene." *Chemistry & biology* 6.7 (1999): 429-439.
- [27] Kennedy, Jonathan, et al. "Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis." *Science* 284.5418 (1999): 1368-1372.
- [28] Gao, Xue, Peng Wang, and Yi Tang. "Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*." *Applied microbiology and biotechnology* 88.6 (2010): 1233-1242.
- [29] Shen, Ben. "Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms." *Current opinion in chemical biology* 7.2 (2003): 285-295.
- [30] Li, Yanran, Wei Xu, and Yi Tang. "Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains." *Journal of Biological Chemistry* 285.30 (2010): 22764-22773.
- [31] Weissman, Kira J. "Introduction to polyketide biosynthesis." *Methods in enzymology* 459 (2009): 3-16.
- [32] Herbst, Dominik A., Craig A. Townsend, and Timm Maier. "The architectures of iterative type I PKS and FAS." *Natural product reports* 35.10 (2018): 1046-1069.
- [33] Bukhari, Habib ST, Roman P. Jakob, and Timm Maier. "Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase." *Structure* 22.12 (2014): 1775-1785.
- [34] Jenni, Simon, et al. "Structure of fungal fatty acid synthase and implications for iterative substrate shuttling." *Science* 316.5822 (2007): 254-261.
- [35] Jenni, Simon, et al. "Architecture of a fungal fatty acid synthase at 5 Å resolution." *Science* 311.5765 (2006): 1263-1267.
- [36] Maier, Timm, et al. "Structure and function of eukaryotic fatty acid synthases." *Quarterly reviews of biophysics* 43.3 (2010): 373-422.
- [37] Maier, Timm, Marc Leibundgut, and Nenad Ban. "The crystal structure of a mammalian fatty acid synthase." *Science* 321.5894 (2008): 1315-1322.
- [38] Dutta, Somnath, et al. "Structure of a modular polyketide synthase." *Nature* 510.7506 (2014): 512-517.

- [39] Ames, Brian D., et al. "Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis." *Proceedings of the National Academy of Sciences* 109.28 (2012): 11144-11149.
- [40] Zhou, Pei, Galina Florova, and Kevin A. Reynolds. "Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis." *Chemistry & biology* 6.8 (1999): 577-584.
- [41] Sivonen, Kaarina, et al. "Cyanobactins—ribosomal cyclic peptides produced by cyanobacteria." *Applied microbiology and biotechnology* 86.5 (2010): 1213-1225.
- [42] Donia, Mohamed S., Jacques Ravel, and Eric W. Schmidt. "A global assembly line for cyanobactins." *Nature chemical biology* 4.6 (2008): 341-343.
- [43] Walsh, Christopher T. "Insights into the chemical logic and enzymatic machinery of NRPS assembly lines." *Natural product reports* 33.2 (2016): 127-135.
- [44] Boettger, Daniela, et al. "Evolutionary imprint of catalytic domains in fungal PKS–NRPS hybrids." *ChemBioChem* 13.16 (2012): 2363-2373.
- [45] Qiao, Kangjian, Yit-Heng Chooi, and Yi Tang. "Identification and engineering of the cytochalasin gene cluster from *Aspergillus clavatus* NRRL 1." *Metabolic engineering* 13.6 (2011): 723-732.
- [46] Williamson, John R., and Barbara E. Corkey. "[65] Assays of intermediates of the citric acid cycle and related compounds by fluorometric enzyme methods." *Methods in enzymology*. Vol. 13. Academic Press, 1969. 434-513.
- [47] Akram, Muhammad. "Citric acid cycle and role of its intermediates in metabolism." *Cell biochemistry and biophysics* 68.3 (2014): 475-478.
- [48] Barton, D. H. R., and J. K. Sutherland. "329. The nonadrides. Part I. Introduction and general survey." *Journal of the Chemical Society (Resumed)* (1965): 1769-1772.
- [49] Yuill, John Lewis. "The acids produced from sugar by a *Penicillium* parasitic upon *Aspergillus niger*." *Biochemical Journal* 28.1 (1934): 222-227.
- [50] Raistrick, Harold, and George Smith. "Studies in the biochemistry of micro-organisms: The metabolic products of *Byssoschlamys fulva* Olliver and Smith." *Biochemical Journal* 27.6 (1933): 1814-1819.
- [51] Hosoe, Tomoo, et al. "A new nonadride derivative, dihydroepihevadride, as characteristic antifungal agent against filamentous fungi, isolated from unidentified fungus IFM 52672." *The Journal of Antibiotics* 57.9 (2004): 573-578.
- [52] Szwalbe, Agnieszka J., et al. "Novel nonadride, heptadride and maleic acid metabolites from the byssochlamic acid producer *Byssoschlamys fulva* IMI 40021—an insight into the biosynthesis of maleidrides." *Chemical Communications* 51.96 (2015): 17088-17091.
- [53] Strunz, G. M., M. Kakushima, and M. A. Stillwell. "Scytalidin: a new fungitoxic metabolite produced by a *Scytalidium* species." *Journal of the Chemical Society, Perkin Transactions 1* (1972): 2280-2283.
- [54] Stillwell, M. A., R. E. Wall, and G. M. Strunz. "Production, isolation, and antifungal activity of scytalidin, a metabolite of *Scytalidium* species." *Canadian journal of microbiology* 19.5 (1973): 597-602.
- [55] Stranks, D. W. "Scytalidin, hyalodendrin, cryptosporriopsin--antibiotics for prevention of blue stain in white pine sapwood." *Wood science* (1976).
- [56] Nicolaou, K. C., et al. "The Absolute Configuration and Asymmetric Total Synthesis of the CP Molecules (CP-263,114 and CP-225,917, Phomoidrides B and A)." *Angewandte Chemie International Edition* 39.10 (2000): 1829-1832.
- [57] Spencer, Paul, Fabio Agnelli, and Gary A. Sulikowski. "Investigations into the Production and Interconversion of Phomoidrides A– D." *Organic Letters* 3.10 (2001): 1443-1445.
- [58] Buechi, George, et al. "Structures of rubratoxins A and B." *Journal of the American Chemical Society* 92.22 (1970): 6638-6641.

- [59] Moss, M. O., F. V. Robinson, and A. B. Wood. "Rubratoxins." *Journal of the Chemical Society C: Organic* (1971): 619-624.
- [60] Nakajima, Mutsuo, et al. "Cornexistin: a new fungal metabolite with herbicidal activity." *The Journal of antibiotics* 44.10 (1991): 1065-1072.
- [61] Fujii, Ryuya, et al. "Biosynthetic study on antihypercholesterolemic agent phomoidride: General biogenesis of fungal dimeric anhydrides." *Organic letters* 17.22 (2015): 5658-5661.
- [62] Saleem, Muhammad, et al. "Viburspiran, an antifungal member of the octadride class of maleic anhydride natural products." *European Journal of Organic Chemistry* 2011.4 (2011): 808-812.
- [63] Saleem, Muhammad, et al. "Viburspiran, an antifungal member of the octadride class of maleic anhydride natural products." *European Journal of Organic Chemistry* 2011.4 (2011): 808-812.
- [64] Isaka, Masahiko, Morakot Tanticharoen, and Yodhathai Thebtaranonth. "Cordyanhydrides A and B. Two unique anhydrides from the insect pathogenic fungus *Cordyceps pseudomilitaris* BCC 1620." *Tetrahedron Letters* 41.10 (2000): 1657-1660.
- [65] Surup, Frank, et al. "Sporothriolide derivatives as chemotaxonomic markers for *Hypoxyylon monticulosum*." *Mycology* 5.3 (2014): 110-119.
- [66] Leman-Loubiere, Charlotte, et al. "Sporothriolide-related compounds from the fungus *Hypoxyylon monticulosum* CLL-205 isolated from a *Sphaerocladina* sponge from the Tahiti coast." *Journal of natural products* 80.10 (2017): 2850-2854.
- [67] Dawson, Michael J., et al. "The squalastatins, novel inhibitors of squalene synthase produced by a species of *Phoma*." *The Journal of antibiotics* 45.5 (1992): 639-647.
- [68] Nicolaou, Kyriacos C., et al. "Total synthesis of zaragozic acid A/squalastatin S1." *Angewandte Chemie International Edition in English* 33.21 (1994): 2190-2191.
- [69] Bonsch, B., et al. "Identification of genes encoding squalastatin S1 biosynthesis and in vitro production of new squalastatin analogues." *Chemical Communications* 52.41 (2016): 6777-6780.
- [70] El-Hasan, Abbas, et al. "Detection of viridiofungin A and other antifungal metabolites excreted by *Trichoderma harzianum* active against different plant pathogens." *European Journal of Plant Pathology* 124.3 (2009): 457-470.
- [71] Esumi, Tomoyuki, et al. "Synthesis of viridiofungin A trimethyl ester and determination of the absolute structure of viridiofungin A." *Tetrahedron letters* 39.8 (1998): 877-880.
- [72] Onishi, J. C., et al. "Antimicrobial activity of viridiofungins." *The Journal of antibiotics* 50.4 (1997): 334-338.
- [73] Moppett, C. E., and J. K. Sutherland. "The biosynthesis of glauconic acid: C 9 precursors." *Chemical Communications (London)* 21 (1966): 772-773.
- [74] Barton, D. H. R., and J. K. Sutherland. "329. The nonadrides. Part I. Introduction and general survey." *Journal of the Chemical Society (Resumed)* (1965): 1769-1772.
- [75] Bloomer, J. L., C. E. Moppett, and J. K. Sutherland. "The biosynthesis of glauconic acid." *Chemical Communications (London)* 24 (1965): 619-621.
- [76] Bloomer, J. L., C. E. Moppett, and J. K. Sutherland. "The nonadrides. Part V. Biosynthesis of glauconic acid." *Journal of the Chemical Society C: Organic* (1968): 588-591.
- [77] Cox, Robert E., and John SE Holker. "Biosynthesis of glauconic acid from [2, 3-13 C] succinate." *Journal of the Chemical Society, Chemical Communications* 15 (1976): 583-584.
- [78] Spencer, Paul, et al. "Biosynthetic studies on the fungal secondary metabolites CP-225,917 and CP-263,114." *Journal of the American Chemical Society* 122.2 (2000): 420-421.

- [79] NAKAJIMA, MUTSUO, et al. "Cornexistin: a new fungal metabolite with herbicidal activity." *The Journal of antibiotics* 44.10 (1991): 1065-1072.
- [80] Bai, Jian, et al. "A Cascade of Redox Reactions Generates Complexity in the Biosynthesis of the Protein Phosphatase-2 Inhibitor Rubratoxin A." *Angewandte Chemie International Edition* 56.17 (2017): 4782-4786.
- [81] Kozich, James J., et al. "Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform." *Applied and environmental microbiology* 79.17 (2013): 5112-5120.
- [82] Branton, Daniel, et al. "The potential and challenges of nanopore sequencing." *Nanoscience and technology: A collection of reviews from Nature Journals*. 2010. 261-268.
- [83] Hamer, Lisbeth, et al. "Gene discovery and gene function assignment in filamentous fungi." *Proceedings of the National Academy of Sciences* 98.9 (2001): 5110-5115.
- [84] Wang, Hong-Liang, Bradley L. Postier, and Robert L. Burnap. "Optimization of fusion PCR for in vitro construction of gene knockout fragments." *Biotechniques* 33.1 (2002): 26-32.
- [85] Hwang, Woong Y., et al. "Efficient genome editing in zebrafish using a CRISPR-Cas system." *Nature biotechnology* 31.3 (2013): 227-229.
- [86] Ran, F. Ann, et al. "Genome engineering using the CRISPR-Cas9 system." *Nature protocols* 8.11 (2013): 2281-2308.
- [87] Liu, Rui, et al. "Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system." *Cell Discovery* 1.1 (2015): 1-11.
- [88] Fuller, Kevin K., et al. "Development of the CRISPR/Cas9 system for targeted gene disruption in *Aspergillus fumigatus*." *Eukaryotic Cell* 14.11 (2015): 1073-1080.
- [89] Nødvig, Christina S., et al. "A CRISPR-Cas9 system for genetic engineering of filamentous fungi." *PloS one* 10.7 (2015): e0133085.
- [90] Fire, Andrew. "RNA-triggered gene silencing." *Trends in Genetics* 15.9 (1999): 358-363.
- [91] Meister, Gunter, and Thomas Tuschl. "Mechanisms of gene silencing by double-stranded RNA." *Nature* 431.7006 (2004): 343-349.
- [92] Matzke, Marjori, Antonius JM Matzke, and Jan M. Kooter. "RNA: guiding gene silencing." *Science* 293.5532 (2001): 1080-1083.
- [93] Hu, Yang, et al. "A PKS gene, pks-1, is involved in chaetoglobosin biosynthesis, pigmentation and sporulation in *Chaetomium globosum*." *Science China Life Sciences* 55.12 (2012): 1100-1108.
- [94] Deshmukh, Radhika, and Hemant J. Purohit. "siRNA mediated gene silencing in *Fusarium* sp. HKF15 for overproduction of bikaverin." *Bioresource technology* 157 (2014): 368-371.
- [95] Fire, Andrew, et al. "Genetic inhibition by double-stranded RNA." U.S. Patent No. 6,506,559. 14 Jan. 2003.
- [96] Morris, Kevin V., et al. "Small interfering RNA-induced transcriptional gene silencing in human cells." *Science* 305.5688 (2004): 1289-1292.
- [97] Pratt, Ashley J., and Ian J. MacRae. "The RNA-induced silencing complex: a versatile gene-silencing machine." *Journal of Biological Chemistry* 284.27 (2009): 17897-17901.
- [98] Desai, Ruchir P., and Eleftherios T. Papoutsakis. "Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*." *Applied and Environmental Microbiology* 65.3 (1999): 936-945.
- [99] Taxman, Debra J., et al. "Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown." *RNA therapeutics*. Humana Press, 2010. 139-156.

- [100] Heneghan, Mary N., et al. "First heterologous reconstruction of a complete functional fungal biosynthetic multigene cluster." *ChemBioChem* 11.11 (2010): 1508-1512.
- [101] Lebe, Karen E., and Russell J. Cox. "Oxidative steps during the biosynthesis of squalestatin S1." *Chemical science* 10.4 (2019): 1227-1231.
- [102] Jin, Feng Jie, et al. "Development of a novel quadruple auxotrophic host transformation system by argB gene disruption using adeA gene and exploiting adenine auxotrophy in *Aspergillus oryzae*." *FEMS microbiology letters* 239.1 (2004): 79-85.
- [103] Palys, Sylvester, Thi Thanh My Pham, and Adrian Tsang. "Biosynthesis of alkylcitric acids in *Aspergillus niger* involves both co-localized and unlinked genes." *bioRxiv* (2019): 714071.
- [104] Kwon, Min Jin, et al. "Beyond the biosynthetic gene cluster paradigm: Genome-wide co-expression networks connect clustered and unclustered transcription factors to secondary metabolic pathways." *bioRxiv* (2020).
- [105] Wasil, Zahida, et al. "Oryzines A & B, maleidride congeners from *Aspergillus oryzae* and their putative biosynthesis." *Journal of Fungi* 4.3 (2018): 96.
- [106] Chesters, Nicola CJE, and David O'Hagan. "Biosynthesis of the fungal metabolite, piliformic acid (2-hexylidene-3-methylsuccinic acid)." *Journal of the Chemical Society, Perkin Transactions 1* 6 (1997): 827-834.
- [107] Jenni, Simon, et al. "Structure of fungal fatty acid synthase and implications for iterative substrate shuttling." *Science* 316.5822 (2007): 254-261.
- [108] Bergmann, Sebastian, et al. "Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*." *Nature chemical biology* 3.4 (2007): 213-217.
- [109] Boettger, Daniela, and Christian Hertweck. "Molecular diversity sculpted by fungal PKS–NRPS hybrids." *ChemBioChem* 14.1 (2013): 28-42.
- [110] Du, L., and B. Shen. "Biosynthesis of hybrid peptide-polyketide natural products." *Current opinion in drug discovery & development* 4.2 (2001): 215-228.
- [111] Gondry, Muriel, et al. "Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes." *Nature chemical biology* 5.6 (2009): 414-420.
- [112] Lautru, Sylvie, et al. "The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases." *Chemistry & biology* 9.12 (2002): 1355-1364.
- [113] Belin, Pascal, et al. "The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways." *Natural product reports* 29.9 (2012): 961-979.
- [114] MacDonald, J. C. "Biosynthesis of pulcherriminic acid." *Biochemical Journal* 96.2 (1965): 533-538.
- [115] MacDonald, J. C. "The structure of pulcherriminic acid." *Canadian Journal of Chemistry* 41.1 (1963): 165-172.
- [116] Yu, Xia, et al. "Biosynthesis of strained piperazine alkaloids: uncovering the concise pathway of herquiline A." *Journal of the American Chemical Society* 138.41 (2016): 13529-13532.
- [117] Schoch, Conrad L., et al. "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi." *Proceedings of the National Academy of Sciences* 109.16 (2012): 6241-6246.
- [118] Blouin, Michael S. "Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer." *International journal for parasitology* 32.5 (2002): 527-531.
- [119] Eren, A. Murat, et al. "A filtering method to generate high quality short reads using Illumina paired-end technology." *PloS one* 8.6 (2013): e66643.

- [120] Wibberg, Daniel, et al. "Establishment and interpretation of the genome sequence of the phytopathogenic fungus *Rhizoctonia solani* AG1-IB isolate 7/3/14." *Journal of biotechnology* 167.2 (2013): 142-155.
- [121] Cuomo, Christina A., et al. "Genome sequence of the pathogenic fungus *Sporothrix schenckii* (ATCC 58251)." *Genome announcements* 2.3 (2014).
- [122] Teixeira, Marcus M., et al. "Comparative genomics of the major fungal agents of human and animal Sporotrichosis: *Sporothrix schenckii* and *Sporothrix brasiliensis*." *BMC genomics* 15.1 (2014): 1-22.
- [123] DiGuistini, Scott, et al. "Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen." *Proceedings of the National Academy of Sciences* 108.6 (2011): 2504-2509.
- [124] Banani, Houda, et al. "Genome sequencing and secondary metabolism of the postharvest pathogen *Penicillium griseofulvum*." *BMC genomics* 17.1 (2016): 1-14.
- [125] Laver, Thomas, et al. "Assessing the performance of the oxford nanopore technologies minion." *Biomolecular detection and quantification* 3 (2015): 1-8.
- [126] Weber, Tilmann, et al. "antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters." *Nucleic acids research* 43.W1 (2015): W237-W243.
- [127] Mulder, Nicola, and Rolf Apweiler. "Interpro and interproscan." *Comparative genomics*. Humana Press, 2007. 59-70.
- [128] Zdobnov, Evgeni M., and Rolf Apweiler. "InterProScan—an integration platform for the signature-recognition methods in InterPro." *Bioinformatics* 17.9 (2001): 847-848.
- [129] Karpusas, Mihail, Bruce Branchaud, and S. James Remington. "Proposed mechanism for the condensation reaction of citrate synthase: 1.9-. ANG. structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A." *Biochemistry* 29.9 (1990): 2213-2219.
- [130] Waldron, C., et al. "Resistance to hygromycin B." *Plant Molecular Biology* 5.2 (1985): 103-108.
- [131] Larkin, P. J. "Purification and viability determinations of plant protoplasts." *Planta* 128.3 (1976): 213-216.
- [132] Blochlinger, Karen, and Heidi Diggelmann. "Hygromycin B phosphotransferase as a selectable marker for DNA transfer experiments with higher eucaryotic cells." *Molecular and cellular biology* 4.12 (1984): 2929-2931.
- [133] Osumi, Takashi, and Takashi Hashimoto. "Acyl-CoA oxidase of rat liver: a new enzyme for fatty acid oxidation." *Biochemical and biophysical research communications* 83.2 (1978): 479-485.
- [134] Osumi, Takashi, Takashi Hashimoto, and Nobuo Ui. "Purification and properties of acyl-CoA oxidase from rat liver." *The Journal of Biochemistry* 87.6 (1980): 1735-1746.
- [135] Reddy, Janardan K., et al. "Transcription regulation of peroxisomal fatty acyl-CoA oxidase and enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase in rat liver by peroxisome proliferators." *Proceedings of the National Academy of Sciences* 83.6 (1986): 1747-1751.
- [136] Ghisla, Sandro, and Colin Thorpe. "Acyl-CoA dehydrogenases." *European Journal of Biochemistry* 271.3 (2004): 494-508.
- [137] Kim, J. J., Ming Wang, and Rosemary Paschke. "Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate." *Proceedings of the National Academy of Sciences* 90.16 (1993): 7523-7527.
- [138] Ansari, Mohd Zeeshan, et al. "NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases." *Nucleic acids research* 32.suppl_2 (2004): W405-W413.
- [139] Yun, Choong-Soo, Takayuki Motoyama, and Hiroyuki Osada. "Biosynthesis of the mycotoxin tenuazonic acid by a fungal NRPS-PKS hybrid enzyme." *Nature communications* 6.1 (2015): 1-9.

- [140] Moutiez, Mireille, et al. "Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases." *Nature Communications* 5.1 (2014): 1-7.
- [141] Sauguet, Ludovic, et al. "Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis." *Nucleic acids research* 39.10 (2011): 4475-4489.
- [142] Belin, Pascal, et al. "The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways." *Natural product reports* 29.9 (2012): 961-979.
- [143] Bonnefond, Luc, et al. "Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog." *Proceedings of the National Academy of Sciences* 108.10 (2011): 3912-3917.
- [144] Moutiez, Mireille, Pascal Belin, and Muriel Gondry. "Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis." *Chemical reviews* 117.8 (2017): 5578-5618.
- [145] Cusack, Stephen. "Aminoacyl-tRNA synthetases." *Current opinion in structural biology* 7.6 (1997): 881-889.
- [146] Delarue, Marc. "Aminoacyl-tRNA synthetases." *Current opinion in structural biology* 5.1 (1995): 48-55.
- [147] Shepard, Alyssa, Kiyotaka Shiba, and Paul Schimmel. "RNA binding determinant in some class I tRNA synthetases identified by alignment-guided mutagenesis." *Proceedings of the National Academy of Sciences* 89.20 (1992): 9964-9968.
- [148] Sylvers, Lee A., et al. "A 2-thiouridine derivative in tRNAGlu is a positive determinant for aminoacylation by Escherichia coli glutamyl-tRNA synthetase." *Biochemistry* 32.15 (1993): 3836-3841.
- [149] Lapointe, Jacqueslouis Duplain, and M. A. R. I. O. Proulx. "A single glutamyl-tRNA synthetase aminoacylates tRNAGlu and tRNAGln in Bacillus subtilis and efficiently misacylates Escherichia coli tRNAGln1 in vitro." *Journal of Bacteriology* 165.1 (1986): 88-93.
- [150] Fried, Stephen D., Sayan Bagchi, and Steven G. Boxer. "Extreme electric fields power catalysis in the active site of ketosteroid isomerase." *Science* 346.6216 (2014): 1510-1514.
- [151] Pollack, Ralph M. "Enzymatic mechanisms for catalysis of enolization: ketosteroid isomerase." *Bioorganic chemistry* 32.5 (2004): 341-353.
- [152] Batzold, F. H., et al. "The Δ^5 -3-ketosteroid isomerase reaction: catalytic mechanism, specificity and inhibition." *Advances in enzyme regulation* 14 (1976): 243-267.
- [153] Sulikowski, Gary A., et al. "Studies on the biosynthesis of phomoidride B (CP-263,114): Evidence for a decarboxylative homodimerization pathway." *Organic letters* 4.9 (2002): 1447-1450.
- [154] Huff, Rkce Moppett, and J. K. Sutherland. "A novel synthesis of a nine-membered ring." *Chemical Communications (London)* 19 (1968): 1192-1193.
- [155] Sulikowski, Gary A., Fabio Agnelli, and R. Michael Corbett. "Investigations into a Biomimetic Approach toward CP-225,917 and CP-263,114." *The Journal of organic chemistry* 65.2 (2000): 337-342.
- [156] Baldwin, Jack E., et al. "Studies towards the biomimetic synthesis of the nonadrides CP-225,917 and CP-263,114." *Tetrahedron* 57.34 (2001): 7409-7416.
- [157] Artimo, Panu, et al. "ExpASy: SIB bioinformatics resource portal." *Nucleic acids research* 40.W1 (2012): W597-W603.
- [158] Möller, Steffen, Michael DR Croning, and Rolf Apweiler. "Evaluation of methods for the prediction of membrane spanning regions." *Bioinformatics* 17.7 (2001): 646-653.
- [159] Nielsen, Henrik. "Predicting secretory proteins with SignalP." *Protein function prediction*. Humana Press, New York, NY, 2017. 59-73.

- [160] Nielsen, Henrik, et al. "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." *Protein engineering* 10.1 (1997): 1-6.
- [161] Song, Jianxing. "Insight into "insoluble proteins" with pure water." *FEBS letters* 583.6 (2009): 953-959.
- [162] Fink, Anthony L. "Chaperone-mediated protein folding." *Physiological reviews* 79.2 (1999): 425-449.
- [163] Dice, J. Fred. "Chaperone-mediated autophagy." *Autophagy* 3.4 (2007): 295-299.
- [164] Shuo-shuo, Cui, Lin Xue-zheng, and Shen Ji-hong. "Effects of co-expression of molecular chaperones on heterologous soluble expression of the cold-active lipase Lip-948." *Protein expression and purification* 77.2 (2011): 166-172.
- [165] Foury, Françoise, and Ornella Cazzalini. "Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria." *FEBS letters* 411.2-3 (1997): 373-377.
- [166] Kleff, Susanne, et al. "Identification of a gene encoding a yeast histone H4 acetyltransferase." *Journal of Biological Chemistry* 270.42 (1995): 24674-24677.
- [167] Asadollahi, Mohammad A., et al. "Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering." *Metabolic engineering* 11.6 (2009): 328-334.
- [168] Zhu, Baogong, et al. "In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations." *Biotechniques* 43.3 (2007): 354-359.
- [169] Sharma, Kailash Gulshan, Rupinder Kaur, and Anand K. Bachhawat. "The glutathione-mediated detoxification pathway in yeast: an analysis using the red pigment that accumulates in certain adenine biosynthetic mutants of yeasts reveals the involvement of novel genes." *Archives of microbiology* 180.2 (2003): 108-117.
- [170] Dolle, Roland E., et al. "Preparation of (±)-(erythro)- and (±)-(threo)-2-vinyl citric acids as potential mechanism-based inhibitors of ATP-citrate lyase." *Tetrahedron letters* 32.35 (1991): 4587-4590.
- [171] Reddick, Jason J., et al. "First biochemical characterization of a methylcitric acid cycle from *Bacillus subtilis* strain 168." *Biochemistry* 56.42 (2017): 5698-5711.
- [172] Krawczyk, Hanna, and Tomasz Martyniuk. "Characterisation of the 1H and 13C NMR spectra of methylcitric acid." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 67.2 (2007): 298-305.
- [173] Wiegand, Georg, and Stephen J. Remington. "Citrate synthase: structure, control, and mechanism." *Annual review of biophysics and biophysical chemistry* 15.1 (1986): 97-117.
- [174] Weitzman, P. D. J., and Michael J. Danson. "Citrate synthase." *Current topics in cellular regulation*. Vol. 10. Academic Press, 1976. 161-204.
- [175] Russell, Rupert JM, et al. "The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution." *Biochemistry* 36.33 (1997): 9983-9994.
- [176] Man, Wai-Jin, et al. "The binding of propionyl-CoA and carboxymethyl-CoA to *Escherichia coli* citrate synthase." *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1250.1 (1995): 69-75.
- [177] Guex, Nicolas, and Manuel C. Peitsch. "SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling." *electrophoresis* 18.15 (1997): 2714-2723.
- [178] Ferraris, Davide M., et al. "Structures of citrate synthase and malate dehydrogenase of *Mycobacterium tuberculosis*." *Proteins: Structure, Function, and Bioinformatics* 83.2 (2015): 389-394.
- [179] DeLano, Warren L. "Pymol: An open-source molecular graphics tool." *CCP4 Newsletter on protein crystallography* 40.1 (2002): 82-92.

- [180] von Heijne, Gunnar. "The signal peptide." *The Journal of membrane biology* 115.3 (1990): 195-201.
- [181] Janda, Claudia Y., et al. "Recognition of a signal peptide by the signal recognition particle." *Nature* 465.7297 (2010): 507-510.
- [182] Liu, Nicholas, et al. "Identification and heterologous production of a benzoyl-primed tricarboxylic acid polyketide intermediate from the zaragozic acid A biosynthetic pathway." *Organic letters* 19.13 (2017): 3560-3563.
- [183] Acharya, Rejwi. *Overexpression, purification, and characterization of MmgD from Bacillus subtilis strain 168*. The University of North Carolina at Greensboro, 2009.
- [184] Brock, Matthias, et al. "Oxidation of propionate to pyruvate in Escherichia coli: Involvement of methylcitrate dehydratase and aconitase." *European journal of biochemistry* 269.24 (2002): 6184-6194.
- [185] Jia, Zhi Sheng, et al. "Borderline between E1cB and E2 mechanisms. Chlorine isotope effects in base-promoted elimination reactions." *The Journal of organic chemistry* 67.1 (2002): 177-181.
- [186] Seddiqi, N., et al. "Characterization and localization of the rat, mouse and human testicular phosphatidylethanolamine binding protein." *Experientia* 52.2 (1996): 101-110.
- [187] Rautureau, Gilles JP, et al. "NMR structure of a Phosphatidyl-ethanolamine Binding Protein from drosophila." *Proteins-Structure Function and Bioinformatics* 78.6 (2010): 1606.
- [188] Fisch, Katja Maria. "Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS." *RSC advances* 3.40 (2013): 18228-18247.
- [189] Cacho, Ralph A., et al. "Understanding programming of fungal iterative polyketide synthases: the biochemical basis for regioselectivity by the methyltransferase domain in the lovastatin megasynthase." *Journal of the American Chemical Society* 137.50 (2015): 15688-15691.
- [190] Anderson, Vernon E., and Gordon G. Hammes. "Stereochemistry of the reactions catalyzed by chicken liver fatty acid synthase." *Biochemistry* 23.9 (1984): 2088-2094.
- [191] Piech, Oliver, and Russell J. Cox. "Reengineering the programming of a functional domain of an iterative highly reducing polyketide synthase." *RSC Advances* 10.31 (2020): 18469-18476.

Curriculum vitae

Sen Yin was born in 25.10.1990, Chongqing, China.

Oct 2016 - Jul 2020	PhD student, Organic Chemistry , Gottfried Wilhelm Leibniz Universität Hannover, Hannover, Germany
Sep 2013 - Jul 2016	Master, Pharmacy , Peking Union Medical College, Beijing, China
Sep 2009 - Jun 2013	Bachelor, Traditional Chinese Pharmacy , China Pharmaceutical University, Nanjing, China

List of publications

1. Yang, X. L., Friedrich, S., **Yin, S.**, Piech, O., Williams, K., Simpson, T. J., & Cox, R. J. (2019). Molecular basis of methylation and chain-length programming in a fungal iterative highly reducing polyketide synthase. *Chemical science*, 10(36), 8478-8489.
2. **Yin, S.**, & Kong, J. Q. (2016). Transcriptome-guided gene isolation and functional characterization of UDP-xylose synthase and UDP-d-apiose/UDP-d-xylose synthase families from *Ornithogalum caudatum* Ait. *Plant cell reports*, 35(11), 2403-2421.
3. **Yin, S.**, & Kong, J. Q. (2016). Transcriptome-guided discovery and functional characterization of two UDP-sugar 4-epimerase families involved in the biosynthesis of anti-tumor polysaccharides in *Ornithogalum caudatum*. *RSC advances*, 6(44), 37370-37384.
4. **Yin, S.**, Liu, M., & Kong, J. Q. (2016). Functional analyses of OcRhS1 and OcUER1 involved in UDP-L-rhamnose biosynthesis in *Ornithogalum caudatum*. *Plant Physiology and Biochemistry*, 109, 536-548.
5. **Yin, S.**, Sun, Y. J., Liu, M., Li, L. N., & Kong, J. Q. (2016). cDNA isolation and functional characterization of UDP-d-glucuronic acid 4-epimerase family from *Ornithogalum caudatum*. *Molecules*, 21(11), 1505.
6. Yuan, S., **Yin, S.**, Liu, M., He, J. M., & Kong, J. Q. (2018). OcUGT1-catalyzed glycosylation of testosterone with alternative donor substrates. *Process Biochemistry*, 73, 82-85.
7. Yuan, S., **Yin, S.**, Liu, M., & Kong, J. Q. (2018). Isolation and characterization of a multifunctional flavonoid glycosyltransferase from *Ornithogalum caudatum* with glycosidase activity. *Scientific reports*, 8(1), 1-13.
8. Chen, D., Fan, S., Chen, R., Xie, K., **Yin, S.**, Sun, L., Liu, J., Yang, L., Kong, J., Yang, Z. & Dai, J. (2018). Probing and engineering key residues for bis-C-glycosylation and promiscuity of a C-glycosyltransferase. *ACS Catalysis*, 8(6), 4917-4927.