Rational Engineering of
an iterative Polyketide Synthase

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Katharina Schmidt, M. Sc.

2024

Abstract

Keywords: natural products, biosynthesis, tenellin, protein engineering, tropolones, alkyl citrates, *Hypoxylon lienhwacheense*

The focus of the presented work is the engineering of polyketide-derived natural products in different stages of their biosynthesis. Within this context, one project centres on the engineering of a polyketide synthase (PKS) to enable the production of unnatural products, while another project focusses on unravelling intricacies of biosynthetic pathways and the understanding of tailoring enzymes.

The iterative PKS TenS is known to produce the pentaketide-tyrosine hybrid tenellin **30**. Products of iterative PKS can differ in reduction level, methylation pattern and chain length. Thus, programming must take place. In TenS, the programming of the chain length is known to be determined by the KR domain. Structure prediction of the KR domain enabled the identification of a relevant substrate binding helix, which was the focus of *in vivo* swap experiments. Swap experiments were conducted with analogous substrate binding helices from related PKS, which produce similar hexa- and heptaketides (**33**, **34**). The hybrid PKS products were analysed and indicated an influence of the helix swap on the programming. Further, the simultaneous swap of only four rationally selected amino acids showed a small effect on the chain length, as well as mutation of single isolated amino acids during a following alanine scan of the helix.

The fungus *Hypoxylon lienhwacheense* is a producer of the tropolone alkyl citrate conjugates lienhwalides A-C (**130-132**) and cordyanhydride B **61**, consisting of three alkyl citrate units linked linearly. Genome analysis revealed two distinct gene clusters proposed to be responsible for the biosynthesis of tropolones and alkyl citrates, respectively. Both clusters were analysed bioinformatically, followed by the elucidation of both pathways by a combination of heterologous expression in the host *Aspergillus oryzae* NSAR1 and protein isolation with following *in vitro* assays. For alkyl citrates, the biosynthesis was revealed to be analogous to other known fungal alkyl citrate biosynthetic pathways. Furthermore, the discovery of a related alkylcitrate **143** of unknown origin hints on the biosynthetic pathway of cordyanhydride B **61**. The formation of pentaketide tropolones was revealed to be catalysed by a set of four key enzymes. Lastly, the coupling of the two natural product classes was investigated, but remains mysterious and requires investigations in the future.

Zusammenfassung

Schlagwörter: Naturstoffe, Biosynthese, Tenellin, Protein Engineering, Tropolone, Alkylcitrate, *Hypoxylon lienhwacheense*

Der Schwerpunkt der vorgestellten Arbeiten liegt auf dem Engineering von Polyketid-Naturstoffen in verschiedenen Stadien ihrer Biosynthese. In diesem Zusammenhang konzentriert sich ein Projekt auf das Engineering einer Polyketid-Synthase (PKS), um die Herstellung unnatürlicher Produkte zu ermöglichen, während sich ein weiteres Projekt auf die Aufklärung von Biosynthesewegen und das Verständnis von Tailoring-Enzymen konzentriert.

Die iterative PKS TenS ist für die Herstellung des Pentaketid-Tyrosin-Hybrids Tenellin **30** bekannt. Die Produkte von iterativen PKS können sich im Reduktionsgrad, im Methylierungsmuster und in der Kettenlänge unterscheiden. Daher muss während der Polyketidbiosynthese eine Programmierung stattfinden. Es ist bekannt, dass die Programmierung der Kettenlänge in TenS durch die KR-Domäne bestimmt wird. Die Strukturvorhersage der KR-Domäne ermöglichte die Identifizierung einer relevanten substratbindenden Helix, die im Mittelpunkt von *in vivo*-swap-Experimenten stand. Die Swap-Experimente wurden mit analogen Substratbindungshelices aus verwandten PKS durchgeführt, die ähnliche Hexa- und Heptaketide (**33**, **34**) produzieren. Die Hybrid-PKS-Produkte wurden analysiert und wiesen auf einen Einfluss der Helix auf die Programmierung hin. Außerdem zeigte der gleichzeitige Austausch von nur vier rational ausgewählten Aminosäuren eine geringe Auswirkung auf die Kettenlänge, ebenso wie die Mutation einzelner isolierter Aminosäuren bei einem Alanin-Scan der Helix.

Der Pilz *Hypoxylon lienhwacheense* produziert die Tropolon-Alkylcitrat-Konjugate Lienhwalide A-C (**130-132**) und Cordyanhydrid B **61**, das aus drei linear verknüpften Alkylcitrateinheiten besteht. Eine Genomanalyse ergab zwei separate Gencluster, die für die Biosynthese von Tropolonen bzw. Alkylcitraten verantwortlich sind. Beide Cluster wurden bioinformatisch analysiert, gefolgt von der Untersuchung beider Stoffwechselwege durch eine Kombination aus heterologer Expression im Wirtsorganismus *Aspergillus oryzae* NSAR1 und Proteinisolierung mit anschließenden *in-vitro*-Tests. Für die Alkylcitrate zeigte sich, dass die Biosynthese analog zu anderen bekannten Pilz-Alkylcitrat-Biosynthesewegen verläuft. Darüber hinaus gibt die Entdeckung eines verwandten Alkylcitrats **143** unbekannter Herkunft Hinweise auf den Biosyntheseweg von Cordyanhydrid B **61**. Es wurde festgestellt, dass die Bildung von Pentaketid-Tropolonen durch eine Gruppe von vier Schlüsselenzymen katalysiert wird. Schließlich wurde die Kopplung der beiden Naturstoffklassen untersucht, die jedoch nach wie vor rätselhaft ist und künftige Untersuchungen erfordert.

Acknowledgement

First, I would like to thank my supervisor Prof. Russell J. Cox for giving me the opportunity to work under his guidance on captivating and challenging projects. Thank you for all the valuable advice and encouragement and for being a patient and encouraging discussion partner, whose door is (literally) always open.

Additionally, I would like to thank Prof. Dr. rer. nat. Jakob Franke, Prof. Dr. rer. nat. Thomas Brüser, and PD Dr. rer. nat. habil. Carsten Zeilinger for dedicating their time as co-referees of my PhD thesis and the chair of the examination.

Collaborations are the basis of scientific work. I would like to thank the collaborators from CebiTec (Bielefeld) and HZI (Braunschweig).

Furthermore, my gratitude goes to the media kitchen and especially to (große) Katja, who lays the foundation for our endeavours and making our research possible. My appreciation also extends to the analytical department of OCI, particularly the NMR and mass spectrometry facilities.

I had a wonderful time with past and present members of the Cox group. I have to thank especially Henrike, who has been my partner through thick and thin (and in sickness and in health during conferences). Thanks to Maurice, Jenny, Dongsong, Yunlong, Mary, Ahmed, Carsten, Lukas, Slawik, Olli, Kevin and many others whose contributions have been invaluable. Thanks to Leon, who generated great results for me as my Master's student. A special acknowledgement goes to Dr. Eric Kuhnert, who discussed, supervised, and advised me on my interesting and also often mysterious project.

The unwavering support of my parents, family, and Family Fitz deserves my deepest gratitude. You are the biggest support anybody could ask for. Thanks to my amazing friends who have enriched the last years immeasurable. I appreciate all the support 'behind the scenes', without which all of this would not be possible. Lastly, the biggest thank you and sorry (?) goes to David. Thanks for always being there for me when I just had to complain about something. I can always count on all of you and appreciate all your love and support.

List of Abbreviations and Units

μ   micro-

A   adenine

A   adenylation domain

ACL   ATP-dependent acyl-CoA ligase

ACP   acyl carrier protein

AT   acetyltransferase

BGC   biosynthetic gene cluster

BLAST   basic local alignment search tool

bp   basepairs

C   condensation domain

C   cytosine

Carb   carbenicillin

cDNA   complementary DNA

C-MeT C-methyltransferase

CoA   coenzyme A

COSY   correlation spectroscopy

CS   alkylcitrate synthase

d   days

Da   Dalton

DAD   diode array detector

ddH$_2$O   double distilled H$_2$O

DH   dehydratase

DHMBA   2,4-dihydroxy-3-methyl-6-(2-oxopropyl)-benzaldehyde

DmbS   desmethylbassianin synthase

DNA   deoxyribonucleic acid

EDTA   ethylenediaminetetraacetic acid

EIC   Extracted ion chromatogram

ELSD   evaporative light scattering detector

ER   enoyl reductase

ESI   electro-spray ionisation

EtOAc   ethyl acetate

FAD   flavin adenine dinucleotide

(v)FAS   (vertebrate) fatty acid synthase

FMO   flavin-dependent monooxygenase

G   guanine

gDNA   genomic DNA

GOI   gene of interest

GRAS   generally recognized as safe

h   hour(s)

H   hydrolase

HF   high fidelity

HMBC   heteronuclear multiple bond correlation

HPLC   High performance liquid chromatography

hr   highly reducing

HRMS   high resolution mass spectrometry

HSQC   heteronuclear single quantum correlation

Hz   Hertz

IMAC   immobilized metal-ion affinity chromatography

IPTG   isopropyl-β-D-1-thiogalactopyranoside

Kan   kanamycin

kb   kilobases

KR   β-ketoreductase

KS   β-ketosynthase

l   litre

LCMS   liquid chromatography – mass spectrometry

M   molar

m   milli-

m/z mass to charge ratio

MDC   maleidride dimerising cyclase

MDR   medium chain reductase/dehydrogenase

mg   milligram

MilS   premilitarinone C Synthase

min   minute(s)

ml   millilitre

mM   milli molar

mRNA   messenger RNA

MS   mass spectrometry

n   nano

NAD(P)H   nicotinamid adenine dinucleotide (phosphate)

NHI   non-heme iron dioxygenase

NMR   nuclear magnetic resonance

nr   non-reducing

NRPS   non-ribosomal peptide synthtase

OD   optical densitiy

OR   oxidoreductase

P450   cytochrome P450 oxidase

PAGE   polyacrylamide gel electrophoresis

PCR   polymerase chain reaction

PEG   polyethylene glycol

pH   *potential hydrogenii*

PKS   polyketide synthase

ppm   parts per million

pr   partially reducing

R   reductive release domain

RNA   ribonucleic acid

rpm   rotations per minute

RT   room temperature

s   seconds

SAM S-adenosyl methionine

SAT   starter unit acyl carrier protein

SDR   short-chain dehydrogenase/reductase

SDS   sodium dodecyl sulfate

SN   supernatant

SNAC N-acetylcystamine thioester

SOC   super optimal broth with catabolite repression

sp.   species

T   thiolation domain

T   thymine

TAE   tris-acetate-EDTA

TE   thioesterase

TEMED N, N, N', N'-tetramethylethylenediamine

TenS   pretenellin A synthtase

TES N-Tris(hydroxymethyl) methyl-2-amino-ethanesulfonic acid

TIC   total ion current

t_R retention time

Tris   tris(hydroxymethyl)aminomethane

U   uracil

U   Units (enzyme activity U = 1 μmol/min)

UV   ultraviolet

vs.   *versus*

v/v   volume per volume

w/v   weight per volume

WT   wild type

2MDH   2-methylcitrate dehydratase

Table of Contents

1   Introduction

Natural products are indispensable in both traditional and modern medicine. At its peak in 1990, 80 % of drugs in use were either natural products or natural product derived structures.[1] Many natural products have beneficial bioactivities like antibiotic, anticancer, antifungal or immunosuppressive properties.[2] Hence natural products, which are produced in great variety in nature, are the key source for the development of drugs.[3]

Every living organism synthesises natural products. These molecules fulfil specific roles and functions within cellular processes, resulting in a great diversity of carbon scaffolds, molecular structures and distinctive properties.[4] Certain natural products contribute to the primary metabolic pathways, essential for cellular growth, development, and reproduction, thereby ensuring the organism's survival. Examples are amino acids, sugars and carbohydrates, vitamins, nucleotides or fatty acids.[5]

In contrast, secondary metabolites are not directly indispensable for cellular survival. However, they add specialized abilities to an organism, thus leading to evolutionary advantages for survival in its environment.[6] For example, plants synthesize vivid pigments like flavonoids to attract insects for reproductive purposes,[7] and bacteria use secondary metabolites to facilitate intercellular communication.[8] The use of secondary metabolites is a common defence strategy to compete with other organisms in one ecosystem.[9]

Secondary metabolites are classified based on their biosynthesis and thereby their precursor molecules. Usually, the main precursors for the biosynthesis of secondary metabolites are derived directly from primary metabolism.[10,11] The four main categories are: terpenes; alkaloids; peptides; and polyketides.[12] Terpenes are composed of isoprene units, which are linked together and modified in various ways. Non-ribosomal-peptides (NRPs) are built of proteinogenic and non-proteinogenic amino acids by very large multimodular proteins (Non-ribosomal-peptide synthetases [NRPS]). Alkaloids include a basic nitrogen atom in their structures, which are derived from amino acids. Polyketides, which will be the focus in this thesis, are derived from

small carboxylic acids (*e.g.* acetate, propionate, butyrate *etc.*) that are linked to each other by multidomain enzymes, called polyketide synthases (PKS).[10]

All four classes of secondary metabolites can be biosynthesised in fungi, bacteria, plants and animals. Fungi, in particular, produce a very wide range of unique natural products, due to their presence in nearly every habitat on earth, which requires the specialisation of molecules to enable their survival.[13]

1.1 Secondary Metabolites from Fungi

Fungi are an underestimated life form when it comes to their survival and diversity in nature, whose potential has not yet been exhausted. The oldest finds of fungi in fossils are around one billion years old.[14] Approximately 140,000 species of fungi are known, but predictions for the total number are estimated from at least 2.2 million up to 5.1 million species.[15,16] They play important roles for the balance in their ecological niches as symbionts, pathogens or decomposers.[17] In 2021 the majority of reported microbial natural compounds in the Natural Product Atlas[18] (a database for microbial-derived natural products) are derived from fungi (about 61 %).[19]

Different bioactive compounds from fungi or other organisms have been used in medicinal or agricultural approaches over the last hundreds of years. Classic examples for secondary metabolites from fungi (Figure 1.1) are penicillin-G **1** (peptide) and 6-methysalicylic acid **2** (6-MSA, polyketide) with antibiotic properties. Others are orsellinic acid **3** or lovastatin **4** (polyketides), which is as an anti-cholesterol drug (inhibiting the hydroxymethylglutaryl-coenzyme reductase).[20] Equisetin **5** is an example of a polyketide, which has antiviral properties as an HIV integrase inhibitor. Furthermore, toxins such as fumonisin B1 **6** (polyketide), vomitoxin **7** (terpene) and α-amanitin **8** (peptide) are known. The alkaloids muscimol **9** and psilocybin **10** are representatives of psychoactive metabolites from fungi.[13,21] The examples (Figure 1.1) show the structural complexity and diversity of fungal metabolites. Among all classes of secondary metabolites, polyketides are the most abundant in fungi.[10] Despite a great range of possible structures, they all have the same carbon backbone consisting of head-to-tail assembled acetate-units.
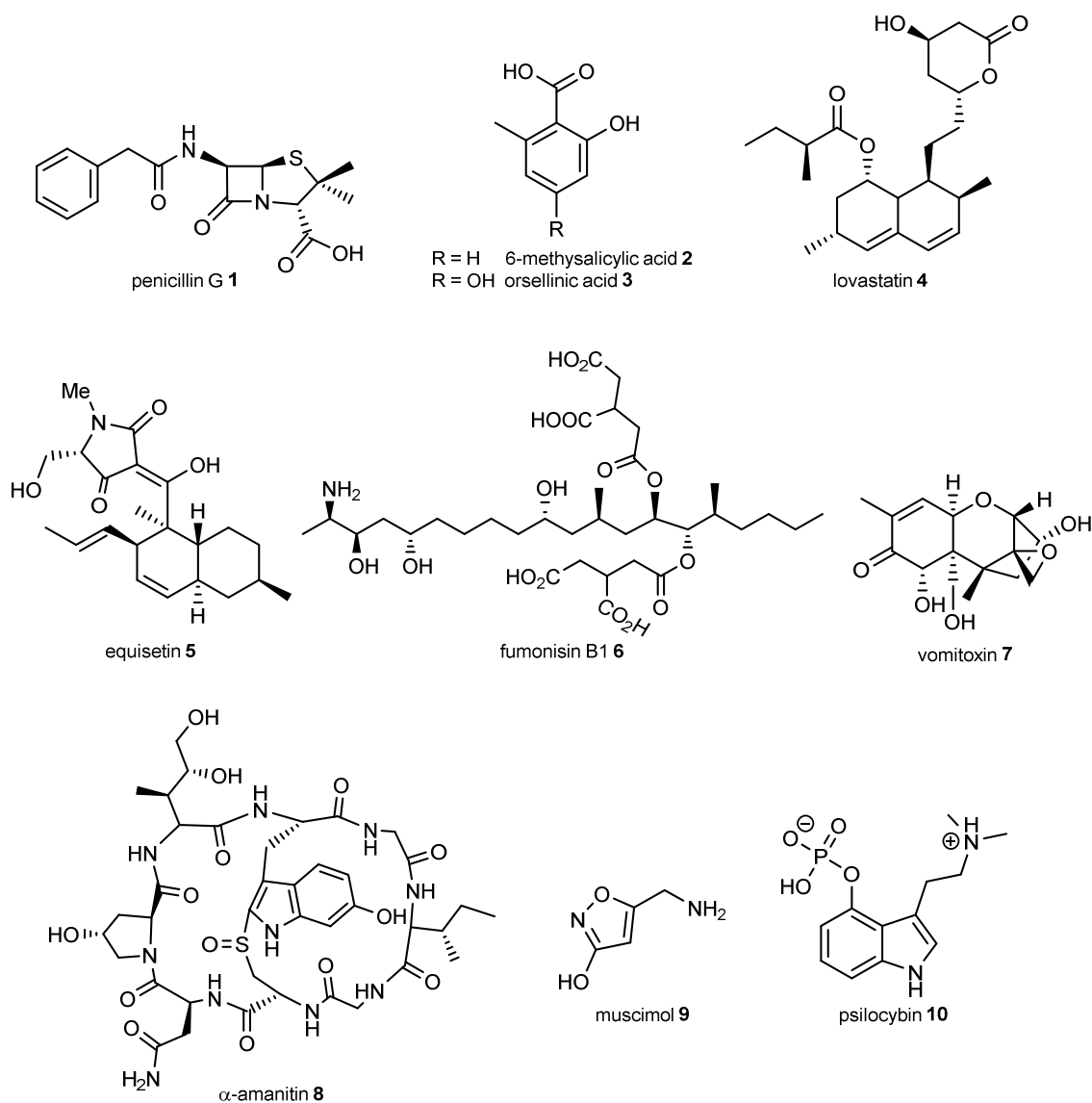
Figure 1.1 Examples of secondary metabolites from fungi.

6-MSA **2** was the first polyketide-derived structure, whose origin from acetate precursors was proven by isotope labelling in 1955.[22] The carbon-carbon coupling reactions and further modifications of the backbone are catalysed by large multidomain proteins known as polyketide synthases (PKS). Polyketides differ in chain length, oxidation-level, coupling to other classes of metabolites or other tailoring steps, resulting in a large variety of complex structural properties, functions and bioactivty.[2]

1.2 Polyketide Biosynthesis

The biosynthesis of polyketides is closely related to the fatty acid biosynthesis, which is part of primary metabolism.[2] Both, polyketides and fatty acids derive from acetate units joined head-to-tail by a Claisen-reaction after prior activation with Coenzyme A (CoA). Furthermore, PKS and fatty acid synthases (FAS) have a similar structural composition, which can be divided into similar functional domains, which catalyse distinctive reaction steps on the growing polyketide chain (Scheme 1.1 B).

As the first preliminary step, the acyl carrier protein (*apo*-ACP) domain is post-translationally modified to enable the first reaction in the biosynthesis: Phosphopantetheine (PP) **11** is covalently bound to a conserved serine of the ACP protein chain as a cofactor by the enzyme phosphopantetheinyl transferase (PPTase) (Scheme 1.1 A). The thiol group carries the growing polyketide chain, while the phosphopantetheine **11** group acts as a flexible arm, shuttling the intermediates of the biosynthesis to the active sites of other catalytic domains.[13]

The acyl transferase domain (AT) is responsible for selection and loading of the *starter units* (acetyl-CoA **12**) for the polyketide or fatty acid, respectively. Therefore, the AT domain binds the acyl molecule to a serine residue and transfers the starter unit to a cysteine residue of the keto synthase domain (KS, Scheme 1.1 B). Next, malonyl-CoA **13** is loaded on the AT domain as the *extender unit* and is transferred to the ACP domain.[13]

The carbon-carbon linkage is built up by decarboxylative Claisen reactions, catalysed by the KS domain (Scheme 1.1 B). The β-carbon of the elongated chain **14** undergoes further modifications in FAS, which are optional for polyketide intermediates. The β-ketoacyl reductase (KR) domain catalyses the NADPH-dependent reduction of the β-ketoacyl group to a secondary alcohol **15**. Then a dehydratase (DH) domain is required to catalyse dehydration to an unsaturated thiolester **16**. The enoyl reductase domain (ER) is responsible for the following last NADPH-dependent reduction to the fully saturated thiolester **17** at the ACP. The resulting intermediate is transferred to the KS domain and either undergoes further chain elongation steps or is passed to the thiolesterase (TE) domain from which it can be released (often as a free acid **18**).[13]
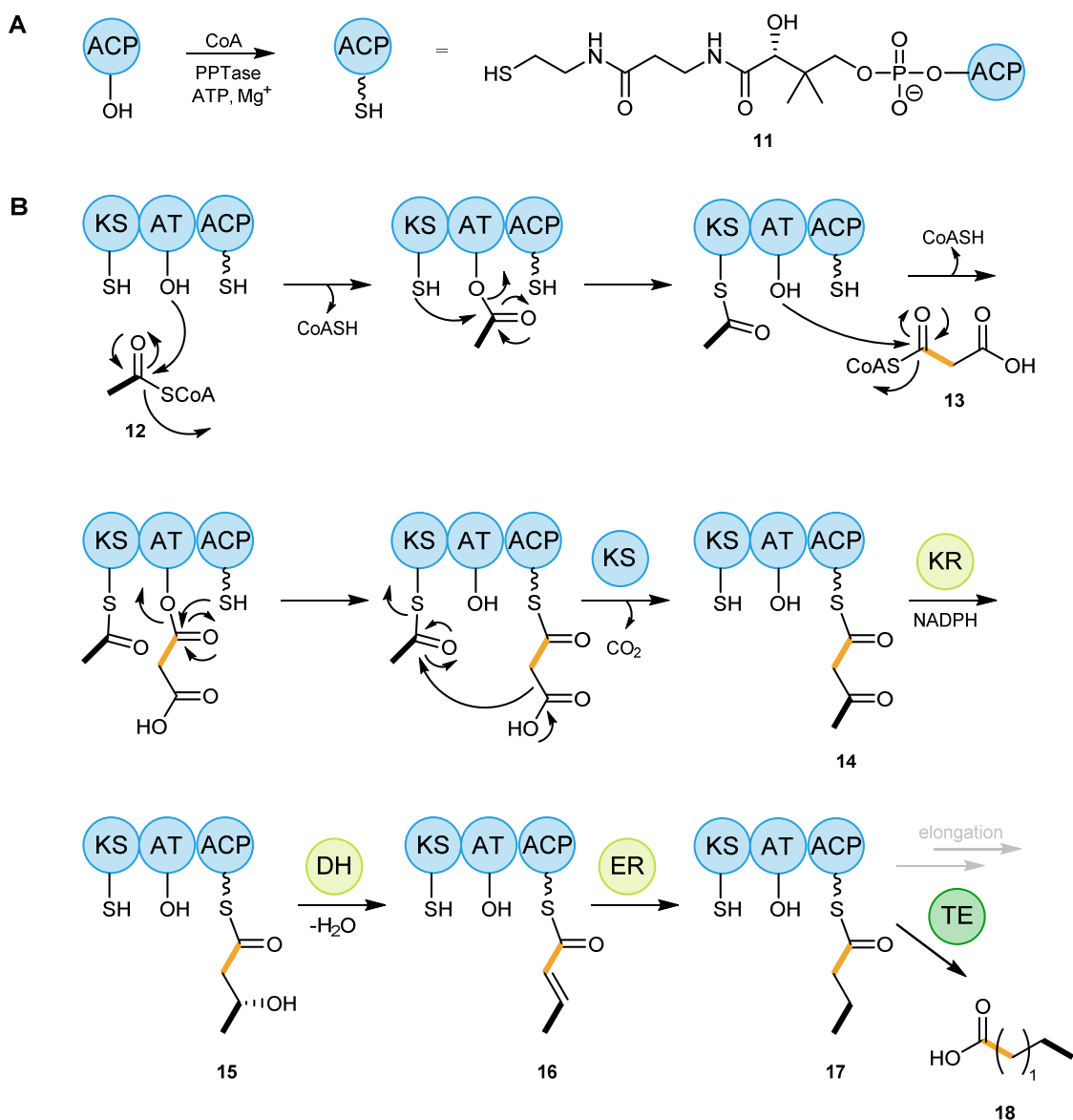
Scheme 1.1 Biosynthesis of fatty acids and polyketides: **A**, conversion of *apo*-APC to *holo*-ACP by PPTase; **B**, chain elongation and β-processing.

The combination of modifications and number of elongation cycles, in which these steps (Scheme 1.1 B) are carried out by the catalytic enzyme domains is the fundamental difference between the PKS and FAS, resulting in different classes of metabolites as products. During fatty acid biosynthesis usually seven to nine acetate units are used resulting in a chain length of 14, 16 or 18 carbons.[2] In contrast, the number of chain elongations steps during polyketide biosynthesis is very variable. Moreover, in FAS all β-processing steps take place in each elongation step to result in fully saturated fatty acids. During polyketide biosynthesis, the modification of the intermediate β-keto thiolester is optional for each catalytic step, resulting in a ketone **14**, secondary alcohol **15**, unsaturated **16** or saturated β-carbon **17**. Further complexity is

added by the possibility of the PKS being selective for different starter and extender units. In bacteria methylmalonyl-CoA is often used for elongation, for example in the erythromycin A biosynthesis.[23]

The main PKS gene responsible for a specific secondary metabolite is surrounded by other co-regulated genes in fungal genomes. Adjacent genes often encode proteins for catalysis of later tailoring steps in the biosynthetic pathway of a compound. The DNA region responsible for the biosynthesis of one specific natural product is called biosynthetic gene cluster (BGC). The gene cluster can encode tailoring enzymes like mono-, and dioxygenases, group transferases (like acyltransferases or glycosyltransferases) or halogenases, leading to a further increase of complexity of polyketide structures. Furthermore, transporter proteins, self-resistance enzymes and transcriptional regulators are often encoded by additional genes in a BGC.[19]

1.2.1 Classification of FAS and Fungal PKS

Three types of PKS are known. Type I PKS are found in bacteria and fungi. They are large multifunctional proteins, which derive from a single gene. A well-known example of a type I fungal PKS is the lovastatin PKS, which synthesises the intermediate dehydromonacolin L **19** for lovastatin **20** biosynthesis (*vide infra*, Scheme 1.2), an early successful anti-cholesterol drug (section 1.2.4). Type I PKS (which are the most common in fungi) can be further divided into modular PKS and iterative PKS (iPKS). All known type II and III PKS and FAS work iteratively (*vide infra*).[21,24] Type II PKS systems only occur in bacteria and consist of dissociated domains working together. Type III PKS are present in plants, fungi, and bacteria. They are isolated dimeric KS enzymes working independent of an ACP or other modifying domains. Typical examples of type III PKS are chalcone and stilbene synthases.[13,23] In contrast, FAS can be categorised into two classes: Type I FAS also consist of one multidomain enzyme, analogous to type I PKS and are found in mammals and fungi. Type II FAS are found in plants and bacteria and consist of multiple dissociated enzymes.[13]

In modular PKS systems, the elongation of the polyketide or further modifications takes place in modules consisting of the catalytic domain sets, where each module acts only once on the intermediate. After the reactions of one module, the intermediate is transferred to the next module where the biosynthesis of the compound

continues. Consequently, the minimal module contains the AT, ACP and KS domain for elongation while the presence of domains for β-processing are optional. The components of the single modules as well as the linear composition of multiple modules together is predictable by analysing the protein sequence of the PKS. A well-known example is the erythromycin polyketide synthase (DEBS1-3), consisting of three separate proteins. DEBS1 includes the loading module and two additional extender modules. DEBS2 and DEBS 3 contain two additional modules each and the TE domain. Thus it can be predicted that the product, 6-deoxyerythronolide B is a heptaketide.[2]

For iterative PKS, the structure of the final product cannot be predicted only by sequence information. The enzymes work iteratively which means that one polyketide is elongated and modified multiple times by the same set of enzymatic domains without a transfer to other modules. The catalysis of β-processing reactions are optional after each elongation cycle. This suggests that iterative PKS must be programmed to determine the chain length and modification of the polyketide. Thus, in iterative systems, the programming is cryptic.

1.2.2 Iterative Fungal Polyketide Synthases

Iterative PKS are classified into highly reducing (hr)-, partially reducing (pr)-, and non-reducing (nr)-PKS.[24] In nr-PKS the growing polyketide chain undergoes no reductive reactions resulting in a poly-β-keto intermediate which cyclize to aromatic products. Pr-PKS only modify the β-carbonyl. In hr-PKS, all shown reductive steps can take place. Hence, every possible reductive state in one molecule is feasible.[25]

Not only the reductive states of the resulting polyketides are different, but also the composition of the catalytic domains and general structure of the PKS (Figure 1.2). Nr-PKS do not include any of the enzymatic domains for β-carbon modification. However, they usually include a unique starter unit ACP transacylase (SAT) domain that loads the starter units, which can be derived from either a FAS, another PKS or acetyl CoA **12**. The extension unit is located after this loading unit, consisting of a β-ketoacylsynthase (KS), acyl transferase (AT), product template (PT) and ACP domain (Figure 1.2 A). The domains work together for the extension of the polyketide with malonyl-CoA **13**. The PT domain is suggested to be evolved from a DH domain that is involved in programmed cyclisation.[26] A domain for termination is present at the C-

terminus of the nr-PKS or other processing domains, like thiolesterases (TE), Claisen cyclases (CLC), or reductases (R).[13,24] Examples of products of nr-PKS are orsellinic acid **3**, aflatoxin **21** and zearalenone **22** (Figure 1.3).
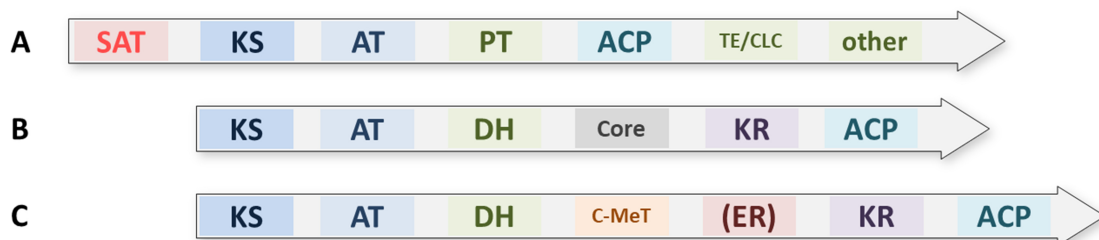
Figure 1.2 Structure of: **A**, nr-PKS; **B**, pr-PKS; **C**, hr-PKR in fungi (SAT = starter unit ACP transayclase, KAS = β-ketoacylsynthase, PT = product template, CLC = Claisen cyclase).

The domain structure of pr-PKS does not contain a starter SAT, PT or TE/CLC domain (Figure 1.2 B). For the reduction of the β-carbon KR and thiolhydrolase (TH) domains are present. The off-loading mechanism is unclear. A core region with no catalytic function is located in the middle of the protein sequences, which is known to play a role in the assembly of the complete synthase by mediating subunit-subunit interaction. Overall much less is known about the relatively rare pr-PKS than about nr-PKS. Typical products derive from 6-methylsalicylic acids, like patulin **23** and *R*-mellein **24** (Figure 1.3).[13]
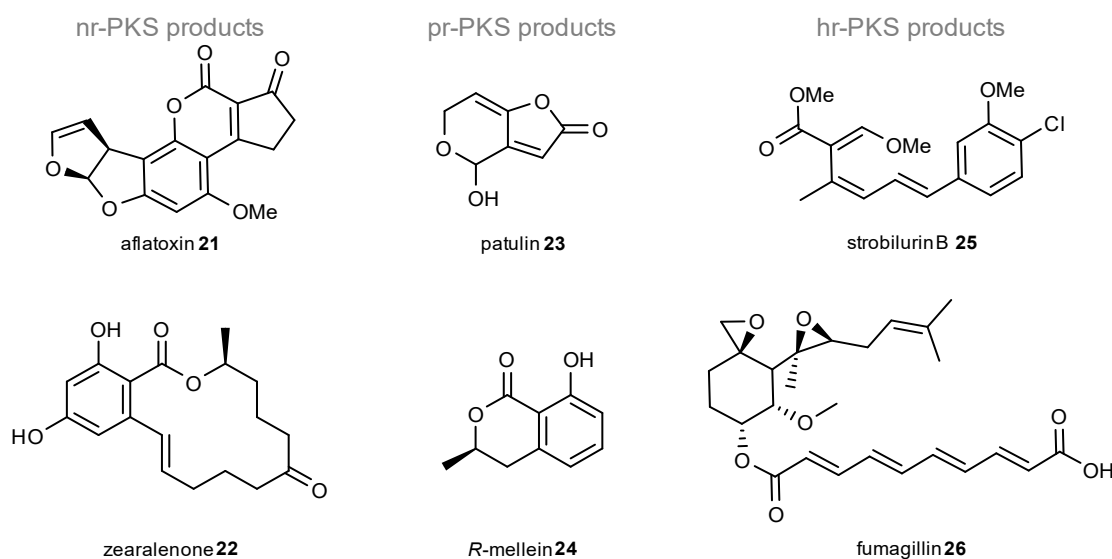
Figure 1.3 Examples of products from nr-/pr-/hr-PKS.

The hr-PKS produces the most complex polyketides, such as strobilurin **25** and fumagillin **26** (Figure 1.3). Some hr-PKS are known to supply other nr-PKS or NRPS their product as a starter molecule for further elongation in the following module.[21] In contrast to the other classes of PKS, the hr-PKS are able to catalyse all β-processing steps that are present in FAS. Therefore, a KR, DH and in some cases an ER domain are included in the sequence (Figure 1.2 C). In addition, programmed *S*-adenosylmethionine (SAM) dependent α-carbon methylation by a *C*-methyltransferase domain (*C*-MeT) takes place optionally. A previous β-carbonyl reduction by the KR is required for the *C*-methylation. However, the *C*-MeT is not active in every hr-PKS. Inactive *C*-methylation domains are called Ψ*C*-MeT.[13,21] The programming of the hr-PKS is not controlled by a distinct region like in nr-PKS, where the PT domain is responsible for chain length and cyclisation regioselectivity. The protein structure is an important factor to investigate the programming of hr-PKS.

1.2.3 Structure of the Vertebrate FAS

The structure of the vertebrate fatty acid synthase (vFAS) is an important basis to investigate the function and programming of iPKS systems. While it was only recently in 2021 possible to generate cryo-electron microscopy (cryo-EM) structures of a hr-PKS[20] (section 1.2.4), for the longest time the vertebrate FAS (vFAS) was used as a template for structural analysis.

Iterative PKS have an overall sequence identity of approx. 20 % to vFAS.[21] However, at a structural level vFAS and hr-PKS are very similar. They have the same catalytic domains, organized in the same order with a high conservation in active and cofactor-binding sites (Figure 1.4 C).[27,28] vFAS also include a Ψ*C*-MeT domain, which is supportive for this model.[21]

The crystal structure of the vFAS shows an overall homodimeric "X" shape (Figure 1.4 A, B). The two polypeptides are mainly in contact through hydrophobic interactions within the KS and ER domains to dimerize (Figure 1.4 A, B). The architecture exposes all other enzymatic domains outwards, forming two lateral clefts, each surrounded by one set of each enzymatic domain. The condensation domains (KS, AT) are located in the lower part of the structure. The β-carbon modification domains KR, DH, ER and *C*-MeT domains are present in the upper region. The modifying region

and the condensing region of the vFAS are connected very loosely by covalent peptide linkage (Figure 1.4 B). This narrow linking region suggests flexibility between the modifying and condensing regions towards each other, allowing a twisting motion.[21]

The dimeric KS domains are the centre of the vFAS architecture. Each AT domain is located at the side of a dimeric KS domain, connected through a linker domain (LD, Figure 1.4). The KS domains are connected to the DH domains of the upper modifying region, which are arranged in a "V"-shape with little contact to each other. The dimeric ER structure is found above the DH domains. The KR domains are linked to the ER dimer structure.
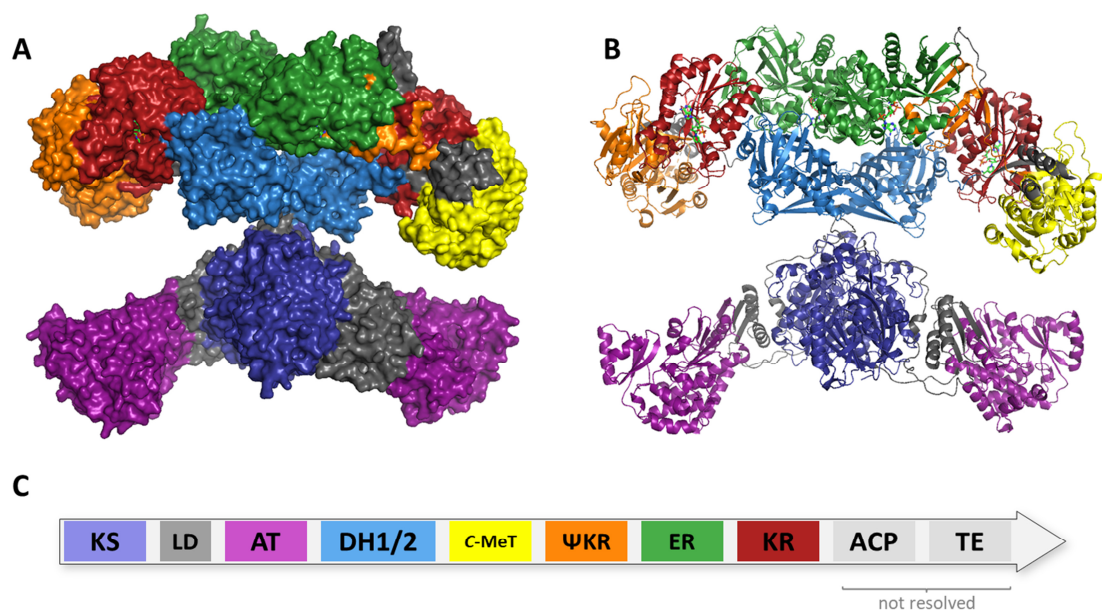
Figure 1.4 Architecture of mFAS: **A**, protein surface depiction; **B**, protein structure; **C**,schematic illustration of domain sequence.[21,28]

The non-catalytic Ψ-*C*MeT domain is structurally related to *S*-adenosylmethionine (SAM)–dependent methyltransferases of PKS but is inactive in the case of the vFAS, since it lacks the conserved active site residues. Additionally, a ΨKR structure is located between KR and Ψ-*C*MeT domains (Figure 1.4 C) with no catalytic, but structural function. The TE and ACP domains are not visualized in the measured crystal structure due to their loose connection to the main vFAS structure. They are connected with a flexible linker of 20 or more residues.[21] However, the anchor point for the ACP is located at the upper site of the lateral clefts, respectively.

Overall, the structural domains of vFAS are more related to type I PKS domains than to the type II FAS found in bacteria. It is hypothesized that vFAS and PKS systems both derived from a common precursor with active *C*-MeT domains, leading to a close evolutionary relationship.[28] The architecture of vFAS and modular PKS is also similar, which shows that the iterative and non-iterative systems are closely related.[21,29]

1.2.4 Structure of the hr-PKS LovB/LovC

The first structural analysis of an hr-PKS was recently reported by Wang *et al.* (2021). The study revealed the cryo-electron microscopy (cryo-EM) structure of the LovB-LovC complex in 3.6 Å resolution.[20] LovB is, together with the *trans*-acting ER LovC, responsible for the synthesis of the dehydromonacolin L **19**, the nonaketide precursor of lovastatin **20** (Scheme 1.2). The structure of the LovB (PKS)-LovC (*trans*-ER) complex was also reported. **19** is later oxidized by LovA (P450 monoxygenase) to **27** and connected to the diketide product **28** synthesized by LovF to result in lovastatin acid **29** (Scheme 1.2) catalysed by the transferase LovD.[30]
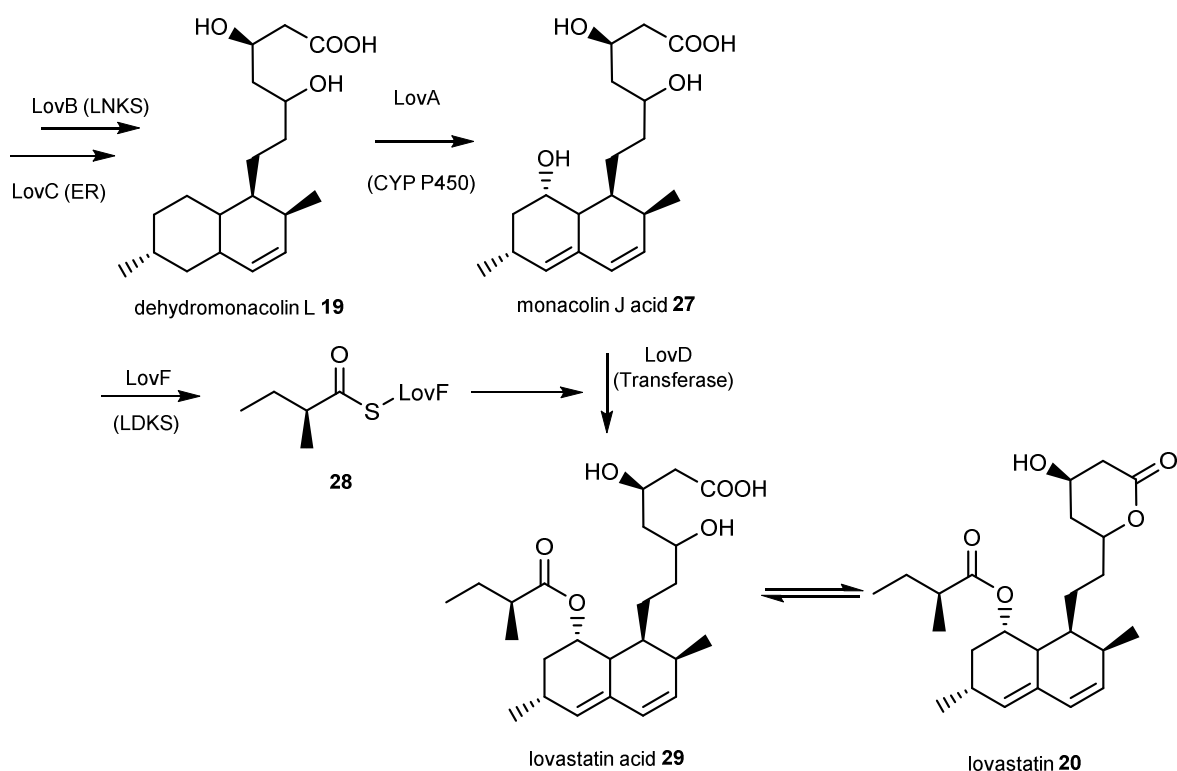
Scheme 1.2 Biosynthetic pathway of lovastatin.

The structural analysis confirmed the similarities to vFAS (overall X-shape in a face-to-face dimer) with few differences.[20] The binding of the *trans*-ER LovC at the AT-domain was described, leading to an L-shaped catalytic chamber, surrounded by the six active domains (Figure 1.5 A, B). Additionally, the loop of the *trans*-ER, which is essential for the interaction to form the complex was identified and confirmed *via* mutation experiments.

A key difference between the vFAS and LovB structures is the contact between the KS and DH domains in the centre of the X-shape. LovB shows only minimal flexibility (~0.4°) between condensing and modifying domains, while a high flexibility was reported vFAS.[21] However, the protein structure of LovB shows high similarities to the vFAS structure, confirming their evolutionary relationship.
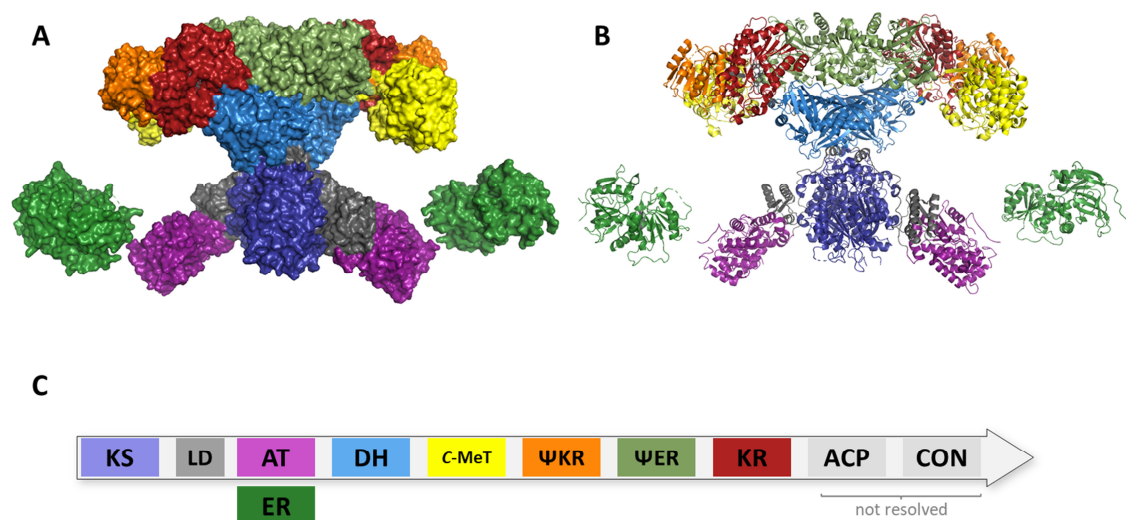
Figure 1.5 Architecture of LovA-LovC complex: **A**, protein surface depiction; **B**, protein structure; **C**,schematic illustration of domain sequence (CON = condensation domain of LovB).[20]

Another recent milestone in the analysis of protein structures was the development of the machine learning based AlphaFold2 algorithm (developed by DeepMind, Inc.), which enables a computational structure prediction *ab initio* from the peptide sequence.[31] It uses neuronal-network based methods to approximate protein model structures.[31] In less than a year after release, over 250 citations referred to the AlphaFold database, portraying the impact on protein structure predictions through all life science fields.[32] Advantages and limitations will be discussed in section 2.3.1.

1.3 Investigation of Natural Product Biosynthesis

The development of molecular biology methods such as genome sequencing and recombinant DNA technology improved the ability to elucidate biosynthetic pathways, enabled metabolic engineering to obtain new compounds, and improved titres of promising structures.[1] The two main methods used in this work are heterologous expression in a fungal host (section 1.3.1) and recombinant protein expression (section 1.3.2) in *Escherichia coli* (*E. coli*) followed by *in vitro* assay experiments.

1.3.1 Fungal Heterologous Expression

A traditional method for the investigation of biosynthetic pathways is the knock-out of putative responsible genes in the wild type fungus and the detection of changes in the resulting product. However, there are multiple reasons for the use of heterologous expression methods instead of the classical analysis of a biosynthetic pathway in the wild type fungus. The cultivation of a wild type fungus under laboratory conditions is often not successful. Furthermore, it is not always possible to genetically manipulate every fungal species with current methods. Many BGCs are silent or barely expressed during the cultivation, which also hinders the easy extraction and identification of natural products. A selected heterologous host is usually easy to cultivate. Furthermore, heterologous expression has potential to increase titres of the target compound in comparison to the wild type producer, due to the easy selection of effective and controllable promoters.[19]

One example of a successful heterologous expression was the 6-methylsalicylic acid synthase (6-MSAS) expression, producing the polyketide 6-MSA **2** in *S. coelicolor* CH999 in 1995.[33,34] However, a prokaryotic host is not always suitable for the expression of fungal genes due to differences in RNA processing and the functional folding of enzymes.[19] In the case of 6-MSA **2**, the titres were remarkably increased by using the fungal host *A. nidulans* as a platform for the biosynthesis.[1,35]

Filamentous fungi are capable of a more similar mRNA processing to the wild type organism as hosts, which are more related to target species, enabling a transformation of DNA without excluding introns in the sequence first. Additionally, heterologous transformation allows the co-expression of genes from different BGCs and species, opening up a nearly endless source of new combinations to build pathways for

the discovery of new hybrid natural products with potential useful bioactivities. This method was further improved by the availability of bioinformatic tools, algorithms and databases, which have become increasingly important in recent years.[36] Furthermore, some transformable filamentous fungi, like *Trichoderma reesii* use waste material as a carbon source, which enables a cheap, sustainable growth.[37]

Commonly used filamentous fungal hosts are *Aspergillus oryzae*, *Aspergillus nidulans* and *Aspergillus niger*.[19] These fungi are also indicated as GRAS organisms (Generally Recognized As Safe), which is emphasized by the traditional use of *A. oryzae* in food industry as a producer of enzymes and for traditional food fermentation products like sake, miso and soy sauce.[19,38–40] In 1983 the first successful transformation of *Aspergillus* was reported.[41] The first reconstruction of a complete fungal BGC was published in 2010 by Heneghan *et al.*, who expressed four genes of the tenellin BGC from *Beauveria bassiana* successfully in the auxotrophic strain *A. oryzae* M2-3 (more details section 1.4.1).[42]

The species *A. oryzae* is additionally characterized by a comparatively clean background production of compounds, making it an excellent host for natural product biosynthesis analysis.[19] The development of a quadruple auxotrophic stain NSAR1 (deficiencies in nitrate reduction (*niaD*-gene*)*, sulfur metabolism (*sC*-gene), arginine biosynthesis (*argB*-gene), adenine biosynthesis (*adeA*-gene)) furthermore facilitates the use for transformation by establishing a platform for selection markers.[40,43] Additionally, two dominant selection marker (against bleomycin (*ble*) and glufosinate ammonium (*bar*)) are available.[39] The strain can be transformed with a combination of the corresponding expression system with pTY-GS vectors (Figure 1.6), including a selective marker and four promoter/terminator pairs (P/T$_{amyB}$, P/T$_{gpdA}$, P/T$_{eno}$, P/T$_{adh}$).[39,43]

The background compound production of *A. oryzae* NSAR1 was recently decreased by the deletion of the *mcrA* gene to obtain a kojic acid deficient strain, which is usually a common by-product in the extraction.[44]

The successful transformation of fungal DNA with self-replicating vectors in the host is rare. In most cases the recombinant sequence is integrated ectopically into the genomic DNA of the host fungus.[45] The recombination and integration into the genomic DNA is enabled through a multi-protein complex (including a DNA-dependent protein kinase, DNA ligase, exonuclease, and Ku70/Ku80 heterodimer) driven by non-

homologous end joining (NHEJ, Figure 1.6), which occurs with predominant higher frequency than site-specific recombination (< 1%).[45] The NHEJ mechanism decreases the efficiency of knock-out experiments in wild type fungi, as the DNA modification has a small probability of targeting the gene of interest (GOI) successfully, making another strategy more attractive, which is the "knock-out by expression" method. Here, the putative genes for a biosynthetic pathway are transformed into the host fungus. After successful reconstruction of a biosynthetic pathway of a compound, the putative genes are omitted individually to mimic the traditional knock-out experiment, leading to the analysis of intermediates and the elucidation of the pathway.
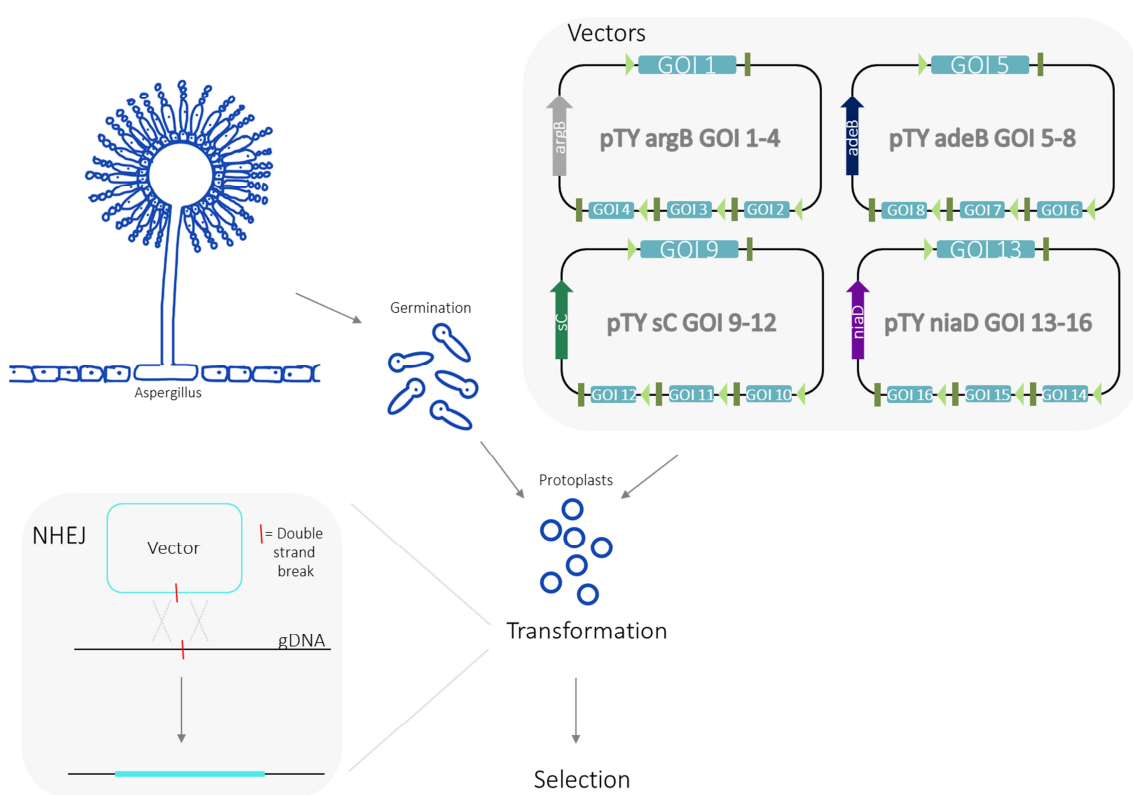
Figure 1.6 General workflow for *Aspergillus* transformation.

The first step for a fungal transformation is to obtain fungal protoplasts from the host strain by the cultivation of *e.g. A. oryzae* NSAR1 spores, followed by enzymatic digestion of the fungal cell wall by an appropriate lysing enzyme (Figure 1.6).[41] Next, polyethylenglycol (PEG)-mediated transformation protocols are carried out. The selection of positive transformants takes place in two rounds on appropriate selective agar.[40,42,43]

1.3.2 *In vitro* Studies with Purified Proteins

Heterologous expression is not always optimal for an investigation of biosynthetic pathways. A main hurdle for working in heterologous organisms is the possible modification of products and intermediates of a pathway by native host enzymes, leading to shunt metabolites.[46,47] Furthermore, products or intermediates are sometimes chemically unstable or toxic to the host. Therefore, the cultivation and extraction of the producing heterologous host does not always lead to a successful detection of compounds. An alternative is the investigation of catalytic enzymes and their products *in vitro*. Furthermore, the investigation of enzymes *in vitro* allows enzymatic studies, for example kinetics or substrate specificity/selectivity tests.[48]

Often the well-studied organism *E. coli* is used as suitable heterologous host for recombinant proteins as the cultivation and expression is cheap, easy and fast in most cases.[49,50] Hurdles like the presupposed post transitional modification of the wild type DNA can be circumvented by rational optimisation of the donor DNA, for example manually excluding known introns or codon optimization.[50] After successful protein expression, soluble proteins can be obtained usually through chromatographic methods.

1.4 Objectives of Research

The main focus of this work are polyketide-derived molecules from fungi. The biosynthesis of the compound families of 2-pyridones like tenellin **30** (section 1.4.1), maleidrides (section 1.4.2), and tropolones (section 1.4.3) will be described in the following sections.

1.4.1 Tenellin and other 2-Pyridone Compounds

Tenellin **30** is a yellow pigment produced by the insect pathogen *Beauveria bassiana.* Initially it was assumed that tenellin **30** might be involved in the pathogenicity of the fungus, but knockout experiments disproved this theory.[13,51] The structure of tenellin **30** was elucidated in 1974 by McInnes *et al.*[51–53] **30** is a representative of an interesting group of natural products: It consists of a pentaketide linked to the amino acid tyrosine **37** *via* an amide linkage (Scheme 1.8), which is enabled through a large protein (4239 amino acids), in which an hr-PKS is combined

with an NRPS module (Figure 1.7).[51,54] The resulting metabolite (after various tailoring steps) belongs to the family of 2-pyridones, which also includes the antifungal antibiotic ilicocolin H **31** from *Cylindrocladium ilicicola*, bassianin **32** isolated from *Beauveria tenella* and desmethylbassianin (DMB) **33** from *Beauveria bassiana* 992.05 (Figure 1.7).[53,55,56] Further examples are the neuroregulator militarione A **34** extracted from *Cordyceps militaris* (Figure 1.7) or farinosone A **35** from *Paecilomyces farinosus*.[53,55,56] **30**, **32**, **33** and **35** only differ in their polyketide precursor in terms of chain length and methylation pattern.
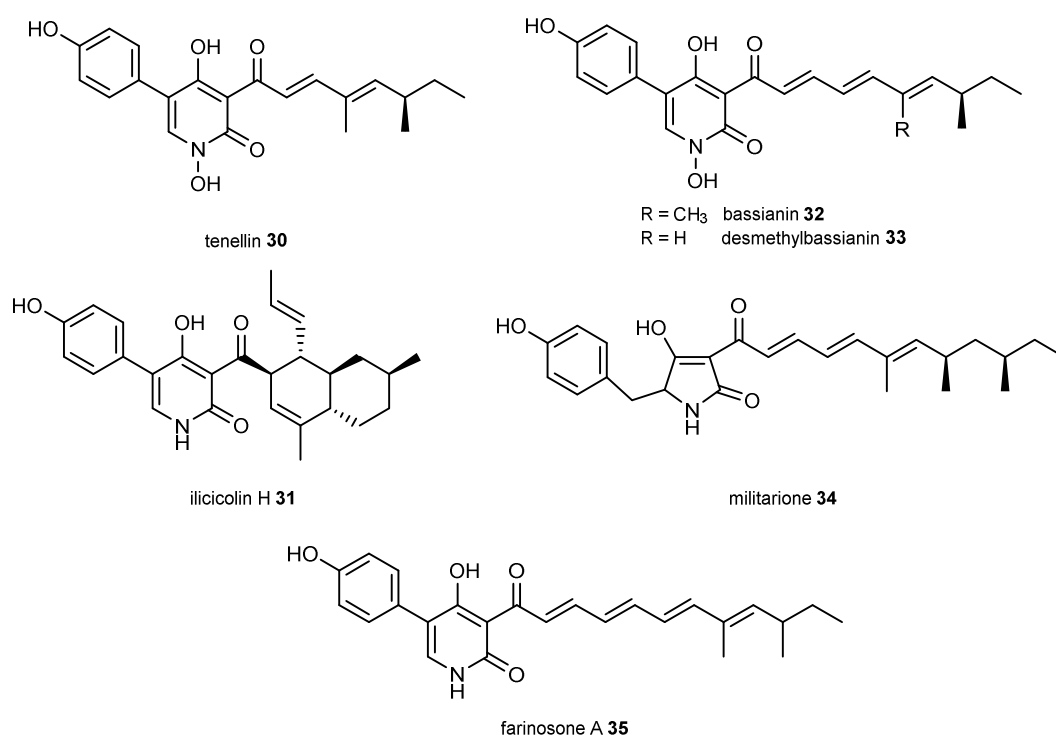
Figure 1.7 Structures of 2-pyridone natural products.

1.4.1.1 Tenellin Biosynthesis

The BGC of tenellin **30** was identified in 2007 by Simpson and co-workers and the biosynthesis was fully solved in 2008 by heterologous expression and gene knock-out/RNAi silencing experiments.[51,54,57] Four genes are present in the cluster (Figure 1.8 A). The longest gene (approx. 12 kb) encodes the hr-PKS-NRPS called TenS, including the KS, AT, DH, *C*-MeT, ER^0, KR and ACP domains, followed by a condensation- (C), adenylation- (A), thiolation- (T), and Dieckmann cyclase- (DKC) domains (Figure 1.8 A). The ER domain (ER^0) within the sequence is not active in the

resulting protein. However, the BGC includes the *trans*-ER domain encoding gene *tenC*. Additionally, two P450 oxidase genes, *tenA* and *tenB* are present in the BGC, acting as tailoring enzymes in later steps of the tenellin **30** biosynthesis (Figure 1.8 B).[54,57] In 2010 the tenellin biosynthetic pathway was the first full pathway reconstructed in a fungal host by Lazarus and coworkers.[42]

As other iterative hr-PKS, TenS is programmed. After the first elongation, the β-carbon is fully reduced to the saturated thiolester. In the following two elongation cycles the reductive circle stops after the dehydration to the unsaturated intermediate. In the last two elongation cycles the β-carbons remains unmodified. After the first and second extension the *C*-MeT domain methylates the α-carbon of the chain, resulting in the ACP-bound double methylated polyketide intermediate **36** (Figure 1.8 B).[58]

The polyketide-intermediate **36** undergoes a condensation reaction with tyrosine **37**, which is selected by the C domain of the NRPS and bound to the T domain. Pretenellin A **39** results from the Dieckmann cyclisation of the molecule **38** as first enzyme free intermediate of the biosynthesis. The cytochrome P450 oxidase encoded by *tenA* catalyses an oxidative ring expansion of the tetramate compound **39** to the 2-pyridone pretenellin B **40**. In the last step of the biosynthesis the second cytochrome P450 oxidase TenB catalyses *N*-hydroxylation, resulting in the final product tenellin **30**.[42]
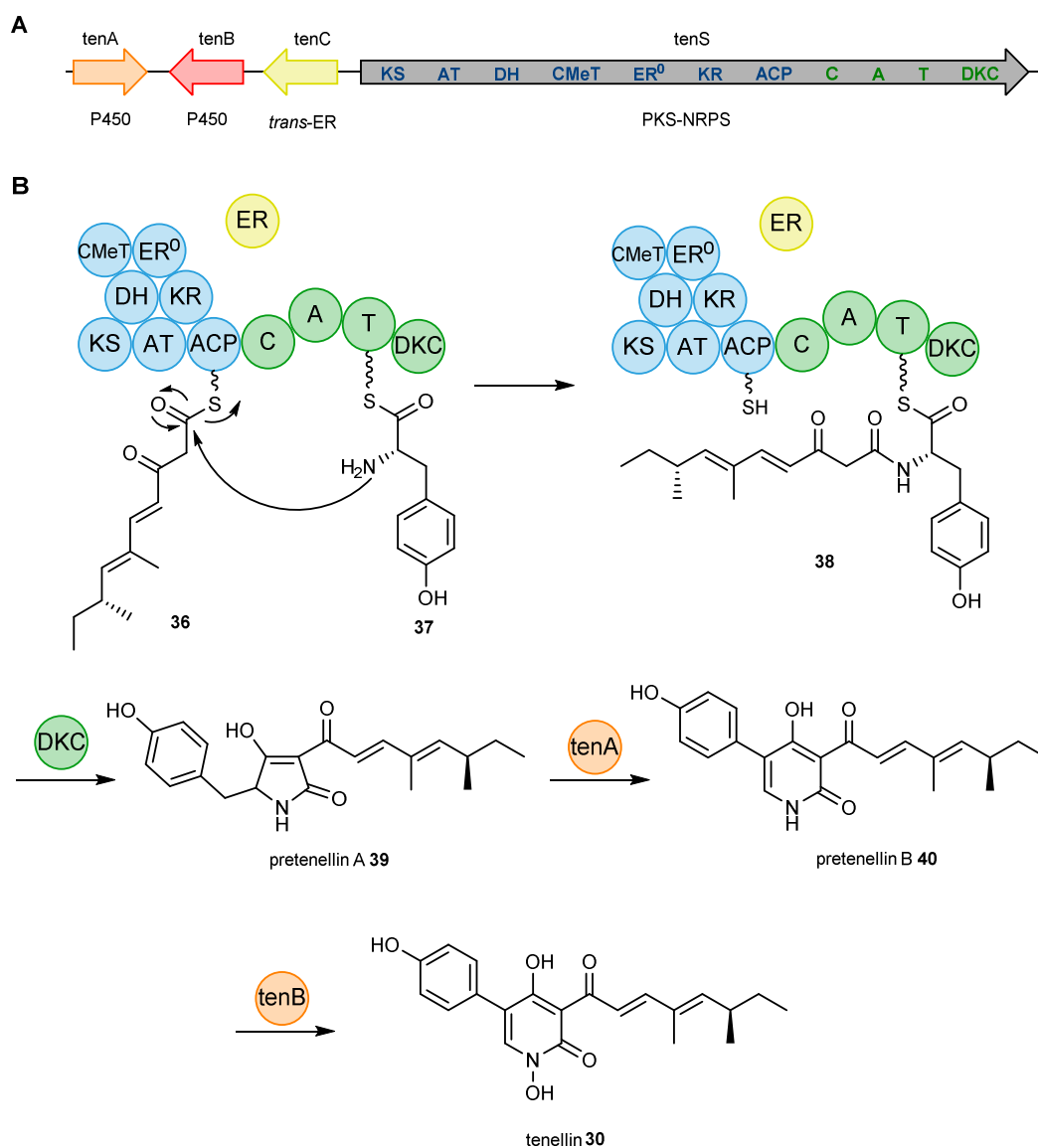
Figure 1.8 Overview of tenellin biosynthesis: **A**, biosynthetic gene cluster of the tenellin biosynthesis; **B**, main reactions of the tenellin biosynthesis.

1.4.1.2 Understanding the Programming of iterative hr-PKS

The programming of iterative PKS remains cryptic. Understanding the programming of different iterative PKS β-processing was studied intensively over past years by a combination of different methods. One strategy were intense *in vitro* studies with *N*-acetylcysteamine thiolesters (SNACs) or pantetheines as analogs of the *holo*-ACP arm.[27] For example, Tang *et al.* showed in 2009 that LovB, expressed in yeast cells, retains correct intermediates until the desired compound is formed. Incorrect intermediates (like pyrones or hydrolytic products) were off-loaded by the PKS.[59] The same group investigated the programming of LovB further, reported in 2015. The DH

domain was deactivated to feed different substrates to the hr-PKS to test the KR and *C*-MeT domain programming. The study showed that the *C*-MeT is selective for unsaturated tetraketide SNACs, which represents the natural substrate in the biosynthetic pathway. However, the KR domain showed no significant selectivity between different methylation pattern and chain lengths.[60] The domains are in competition for the reaction on the polyketide. The feeding of different substrates showed the difference in the effect of both domains: if a triketide was fed, the KR reaction was significantly faster than the methylation. In the case of tetraketides, the methylation was faster.[27,60]

For the programming of DH domains, much less is known. While they show a high stereo *E*-selectivity of the products, the substrate selectivity has not been studied.[27] The ER domain of other hr-PKS was shown to be selective for a broad range of substrates: for example, the SQTKS *cis*-ER catalyses even reactions with unnatural intermediates like *Z*-alkenes.[27,61] Some *trans*-ER domains showed no selectivity, like in the case of TenS/DmbS (*vide infra*, section 1.4.1.3). In contrast, LovC is able to catalyse reactions of non-intermediate like structures, but no diketide or triketide reactions.[27,62]

The combination of *intrinsic* programming (substrate selectivity of each domain) and *extrinsic* programming (competition between domains) determines the overall programming and thereby the resulting polyketide structure.

1.4.1.3 Early Investigation of TenS Programming

In the past, TenS was used as a model-PKS to investigate programming. The basis for investigations of the program of TenS was the comparison of the iPKS with the close relative desmethylbassianin synthase (DmbS) from *Beauveria bassiana* 992.05. DmbS is responsible for the biosynthesis of the 2-pyridone desmethylbassianin **33** (DMB, Figure 1.7), which contains a hexaketide chain with differing methylation pattern from the pentaketide tenellin **30** (Figure 1.7). Despite these differences in the programming of the PKS, the main biosynthesis of the two molecules is the same. This is also reflected by the biosynthetic gene cluster sequence of the two molecules, which share a 90 % sequence identity at peptide level.[56] The main PKS-NRPS hybrid protein has an overall identity of 86 %, where the single domains vary in identity between 80 % and 97 %.[56]

The overall high sequence similarity despite different programming provided the opportunity to perform domain-swap experiments.

Early studies tested the effect of the NRPS on the programming. Chimeric PKS-NRPS were constructed *via* homologous recombination in yeast and heterologous expression of the TenS PKS with DmbS NRPS (but lacking the *trans*-ER).[56] The exchange of the NRPS module did not change the resulting products, indicating that the NRPS plays no role in programming (Figure 1.9 A-B). This finding additionally shows that the C-domain of the NRPS is not selective for the polyketide structure, which is linked to tyrosine.[56]

The effect of the *trans*-acting ER was studied with similar methods. The swap of the *trans*-ER between the two systems did not show differences in the programming, indicating that it is determined in the PKS itself (Figure 1.9 C-D). However, the *trans*-ER has influence on the fidelity of the program. In the absence of the *trans*-ER (Figure 1.9 A-B) the programming fidelity decreases. Different by-products with shorter polyketide chains, namely prototenellin A-C **41-43**, were found. In contrast, the presence of the *trans*-ER leads to the biosynthesis of a single main product (pretenellin A **39** or preDMB A **44**, Figure 1.9 C-D). Lazarus and co-workers suggested that the *trans*-ER has influence on the PKS by binding to it during the polyketide synthesis. In absence of the *trans*-ER the PKS becomes relatively slow, which results in shorter chains because the C domain of the NRPS picks up shorter chains from the PKS earlier.[56]

The finding of TenS being capable of producing a group of polyketides with varying chain length and methylation pattern gave the first evidence that TenS can be reprogrammed.[25] These findings were confirmed later by gene silencing experiments of *tenC* in the wild type fungus by Yakasai *et al.*[25]

After the successful analysis of domain swap experiments regarding the NRPS and *trans*-ER, more sophisticated approaches were also performed by homologous recombination in yeast to construct chimeric iPKS genes, followed by heterologous expression. Through various swapping experiments, the less crucial domains responsible for the programming were revealed stepwise (Figure 1.9 E-L). The AT, KS, DH and ACP domains appeared to play no role in programming (Figure 1.9 F, H-I) as the exchange from the donor PKS had no effect on the resulting product titres or

structures. This suggests that the extender machinery of the hr-PKS works indiscriminately.
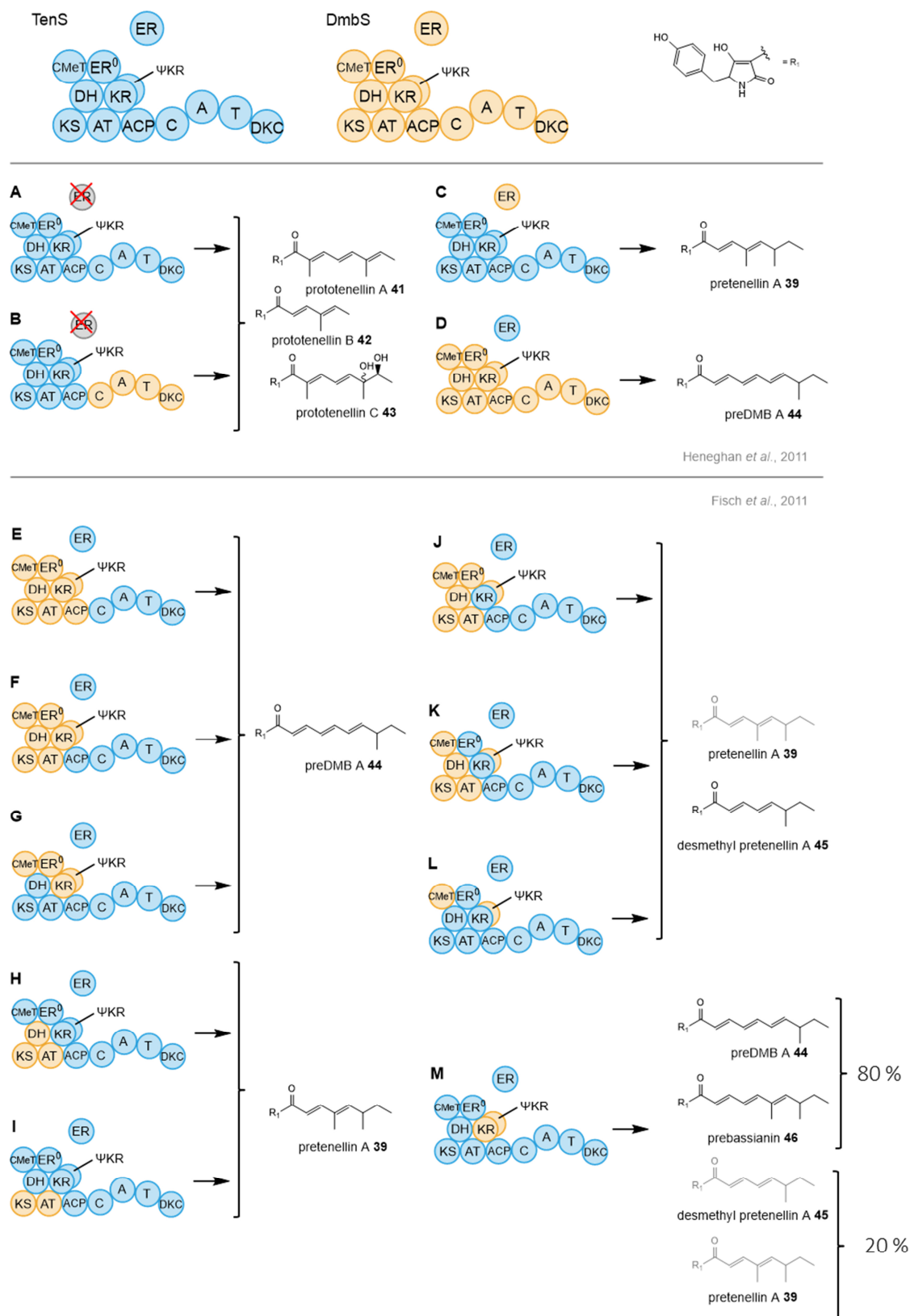
Figure 1.9 Summary of domain swap experiments (minor compounds are shown in grey).

In swaps with the *C*-MeT-ΨKR domain involved, but not the KR domain (Figure 1.9 J-L), the products do not differ in chain length, but in the methylation pattern. A minor amount of the dimethylated pretenellin A **39** was observed, while the main product was desmethylpretenellin A **45**, which is mono-methylated, analogous to preDMB **44**. This indicated that the *C*-MeT plays a significant role in the control of the methylation pattern.

Only when swapping KR, ΨKR, and *C*-MeT domains between TenS and DmbS changes in the resulting product was observed (Figure 1.9 G, J-M).[63] The swap of only the KR domain (Figure 1.9 M) finally lead to a mixture of products, with a majority being hexaketides preDMB A **44** and prebassianin **46** (80%) over pentaketides desmethylpretenellin A **45** and pretenellin A **39** (20%). Furthermore, dimethylated products **46** and **39** (66 %) dominated over mono-methylated products **45** and **44** (34%). Therefore, it was suggested that chain length is mainly controlled by the KR domain, and it has an effect on the methylation pattern, but is less influential than the *C*-MeT domain.[63]

Simpson and co-workers concluded that the modifying domains have intrinsic selectivity for their substrates and rationalised parts of the extrinsic programming by the physical contact of the *C*-MeT-ΨKR-KR domains to each other in the protein structure (Figure 1.5, section 1.2.3). Parts of the KR domain were especially responsible for determination of the chain length, when swapping smaller sub fragments, which will be discussed in more detail in section 2.1.[58]

1.4.2 Maleic Acid Anhydride Compounds: Maleidrides

A large group of natural products, which derive from maleic acid anhydrides (Figure 1.10) are maleidrides. The first reported representative, glauconic acid **47** was isolated from the fungus *Penicillium glaucum/purpurogenum* and is known since 1931, followed by byssochlamic acid **48** (*Byssochlamus fulva*) in 1933.[64,65] However, the elucidation of the structure was only possible 30 year later by chemical degradation and X-ray crystallography studies, revealing a central alicyclic nine-membered ring system, which is the result of two maleic acid anhydrides linked together in head-to-tail orientation (Figure 1.10).[66–68]

The characteristic composition of two C_9 monomer units is the origin of the name nonadrides, which describes this family of compounds since 1965.[69] Various

similar structures were found in different fungal species since then, like the herbicide cornexistin **49** and 1'-hydrocornexistin **50**, produced by *Paecilomyces variotii* or antifungal heveadride **51**, isolated from *Bipolaris heveae.*[70–72]

R = OH glauconic acid **47**
R = H   glaucanic acid **47a**

R1 = O, R2 = H   byssochlamic acid **48**
R1 = H, R2 = OH 1-hydroxy byssochlamic acid **48a**
R1 = OH, R2 = H 10-dihydro byssochlamic acid **48b**

R = H   cornexistin **49**
R = OH 1'-hydroxy cornexistin **50**

heveadride **51**   zopfiellin **52**   viburspiran **53**   agnestadride A **54**

agnestadride B **55**   rubratoxin B **56**   phomoidride D **57**
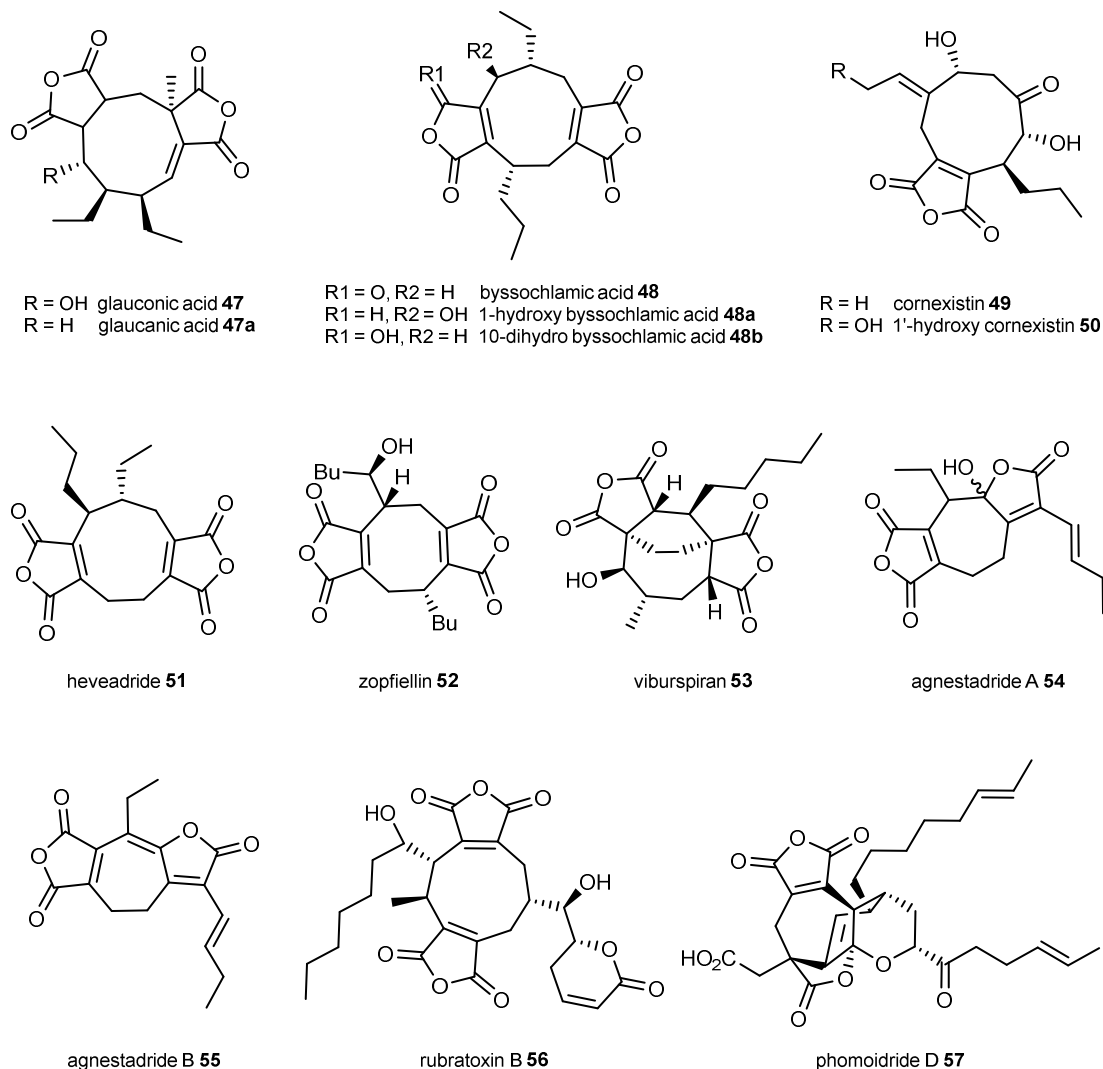
Figure 1.10 Examples of maleic anhydride units as part of dimerized carboxylic rings.

Over the years, the name nonadrides was more associated with the conspicuous central nine-membred ring structure, which established the name octadrides for related structures with an eight-membered ring like zopfiellin **52** or viburspiran **53** (both antifungal).[72,70] Moreover, heptadrides were discovered, which include a central seven-membered ring. Examples for these compounds are agnestadride A **54** and B **55** which are also produced by *Byssochlamus fulva*.[73] To summarize all three terms, the umbrella term "maleidrides" was introduced.[74] Furthermore, more complex structures like rubratoxin B **56** (protein phosphatase 2A [PP2A] inhibitor)[72,75] and phomoidride D **57**

(anticholesterol/ anticancer properties)[70,76] are known, with differing precursor molecules and therefore longer alkyl side chains.

1.4.2.1 Other Alkyl Citrate Compounds

In addition to the maleidride structures, also linear chains of two or three maleic acid anhydride units are known. The examples are cordyanhydride A **60** and B **61**, isolated from *Cordyceps pseudomilitaris* BCC 1620 in 1999 (Figure 1.11).[77] The biosynthesis of these linear compounds remains unknown to date.

Sporothriolide **62** is an alkyl citrate, first reported in 1994 from *Sporothrix sp.*, *Discosia sp.*, and *Pezicula livida* (Figure 1.11).[78,79] Another example for a natural product incorporating oxaloacetic acid is the squalene synthase inhibitor squalestatin S1 **63**, isolated from various fungi *e.g. Phoma sp.*, which contains citrate in its final structure. It is famously known for its cholesterol lowering properties.[80,81]
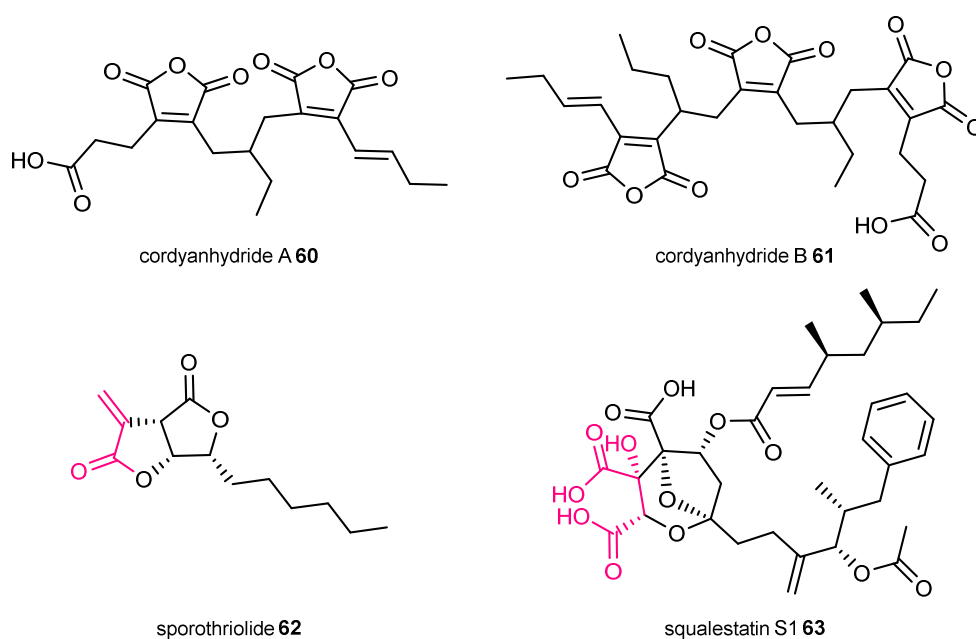
Figure 1.11 Examples of alkyl citrate compounds (pink = origin from oxaloacetic acid).

Carboxylic maleic acid anhydride compounds are the product from the reaction of a citrate synthase like enzyme, which combines oxaloacetic acid **64** and polyketide or fatty acid precursors to alkyl citrate compounds. This was first reported in 1968, when feeding studies with labelled acetate units proved the origin as a polyketide intermediated linked to the remaining C_3-chain, which was identified as oxaloacetic acid

64.[68] The citrate synthase reaction is not only known in secondary metabolism, but also in primary metabolism, where a citrate synthase catalyses the condensation of oxaloactetic acid **64** and acetyl-CoA **12** during the first key enzymatic step of the citric acid cycle, resulting in citrate **65** and CoA (Scheme 1.3).[82]
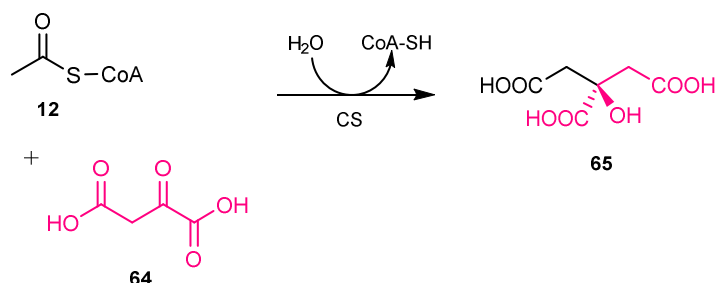
Scheme 1.3 Reaction of the citrate synthase (CS) during citric acid cycle.

1.4.3 Tropolones and Troponoids

Tropolones, characterised by a 7-membered aromatic ring (Figure 1.12, red), are found in plants, bacteria and fungi. In fungi they are a rather rare class of compounds with only ~30 structures known.[83] The tropolone moiety is often part of more complex structures (Figure 1.12), like cytotoxic gukulenin A **66** from the marine sponge *Phorbas gukulensis* or nemanolone F **67** from the fungus *Nemania sp*. that has antimalarial and cytotoxic bioactivity. Additional examples are rubrolone B **68** (*Streptomyces echinoruber*) and (-)-colchicine **69** (*Colchicum autumnale*, plant).[83,84]
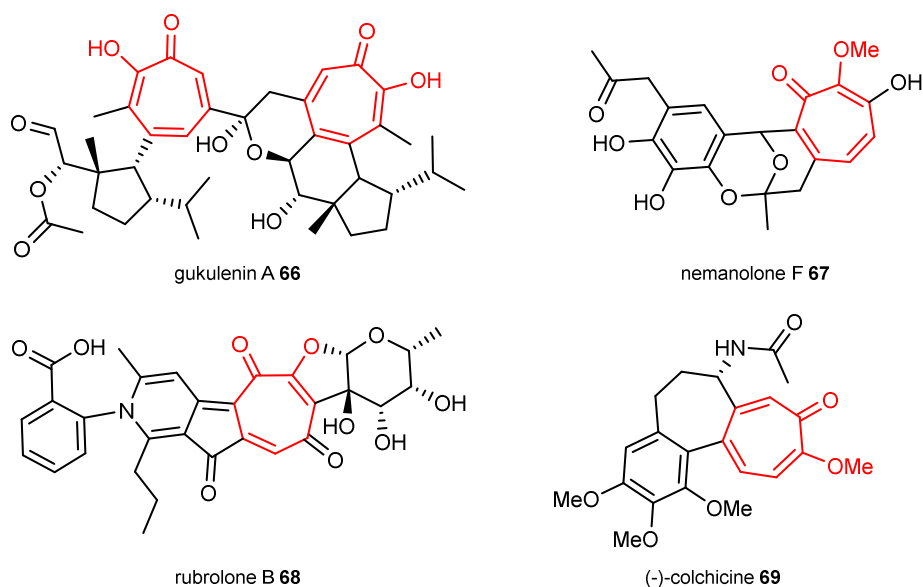
Figure 1.12 Examples of tropolone compounds.

Natural products including tropolone moieties often show antibacterial properties, but also other bioactivities are known like antiviral, antitumor, antioxidant, anti-inflammatory, insecticidal activities or they act as inhibitors of enzymes by chelating metal ions.[85] Therefore, tropolone structures had high relevance in the past decades for drug research.

1.4.3.1 Tropolone Biosynthesis in Bacteria and Plants

The biosynthetic pathway for tropolones has evolved based on different routes fungi, plants and bacteria (section 1.4.3.2, Scheme 1.4, Scheme 1.5).[83]

In bacteria the biosynthetic pathway was elucidated by isotopic labelling experiments, determining the origin of tropolone structures in phenylalanine **70** *via* phenylacetic acid (PAA) **71** based on the shikimate pathway.[83] Later knock-out studies in bacteria like *Phaeobacter inhibiens* and *Burkholderia plantarii* identified the proteins for the later biosynthetic steps. The detailed biosynthetic pathway was reported in 2018 for *Streptomyces cyaneofuscatus* and *S. luteogriseus* (Scheme 1.4)*. PAA **71** biosynthesis is followed by CoA-activation to **72** and the formation of a 7-membered heterocycles **74** *via* the epoxide **73** by the enzymes PaaA-E and PaaG, which is then oxidised by TrlA to form the aldehyde **75**.[86] After Knoevenagel condensation to **76** and various oxidative steps the tropone **77** and tropolone **78** core structure is formed, which

is then modified to several tropolone-containing compounds, *e.g.* 3,7-hydroxytropolone **79**.[86]
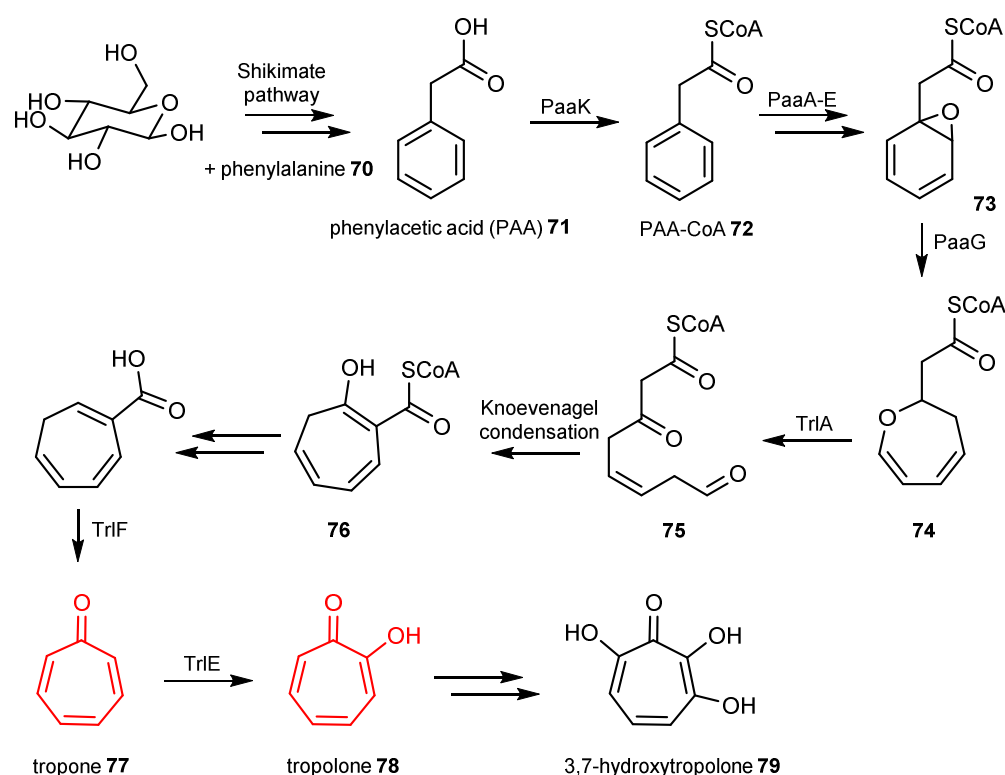
Scheme 1.4 Biosynthetic pathways of tropolones in bacteria to 3,7-hydroxytropolone **79**.

An alternative origin of tropolones in bacteria is the biosynthesis by Type II PKS. Here, the tropolone is often part of a molecule with higher complexity. A complex oxidative rearrangement of the poly-β-ketoacyl intermediate leads to the later compound. It is yet unclear, if the cyclisation or oxidation takes place first.[83,84]

In plants, the biosynthetic origin is not anchored in primary metabolism. Isotopic labelling proved the terpene and/or alkaloid origin of many tropolone structures. One example is the biosynthetic pathway of β-thujaplicin **80** *via* geranyl diphosphate **81** (GDP, Scheme 1.5). After cyclisation to form a terpinolene structure **82**, selective oxidative reactions take place, followed by a rearrangement, in which the methyl-group of the epoxide **83** is incorporated into the final tropolone ring system of **80**.[83,87,88]
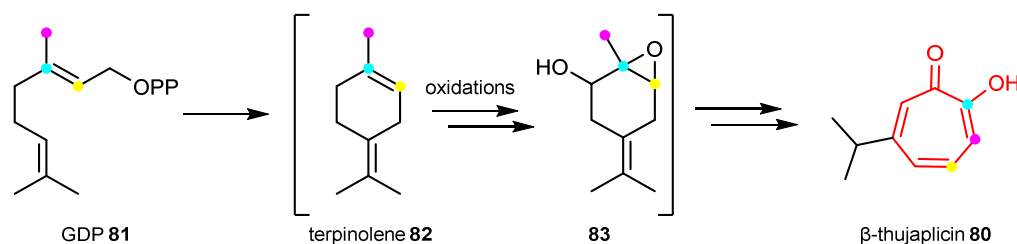
Scheme 1.5 Biosynthetic pathways of tropolones in plants to β-thujaplicin **80**.

However, other examples like colchicine **69** are alkaloid derived, where an oxidative rearrangement is also the key step to form the tropolone.[84] Thus, just like in bacteria, different tropolone moieties are built based on different metabolic pathways.

1.4.3.2 Tropolone Biosynthesis in Fungi

The earliest found examples of tropolones from fungi derived from polyketides, the tetraketides puberulic acid **84** (reported in 1932) from *Penicillium puberulum*, and stipitatic acid **85** from *Talomyces stipitatus* (1942), were both described by Raistrick and coworkers.[85,89–91] The tropolone structure was elucidated in 1945 by investigating colchicine **69**, which uncovered the characteristic 7-membered, aromatic ring as the key structure (Figure 1.13 A).[92]

For fungi, the biosynthetic core steps were elucidated 2012 by Davison *et al.* in case of tetraketides (Figure 1.13 B): An nr-PKS releases 3-methylorcinaldehyde **86**, which undergoes oxidative dearomatisation by a FAD-dependent monooxygenase (FMO) resulting in **87**. The non-heme iron-dependent dioxygenase (NHI) catalyses the next oxidation to **88** enabling a pinacol-type re-arrangement of the six-membered ring, resulting in the seven-membered ring core structure **89**.[93] The responsible core enzymes TropB (FMO) and TropC (NHI) are related to enzymes with similar functions in the well-studied pathways like citrinin (CitB), azaphilone (MrPigN) and sorbicillinoid (SorbC) biosynthetic pathways.[74]

A

R = OH puberulic acid **84**
R = H   stipitatic acid **85**

sepedonin **90**

antibiotic C **91**

spiromaterpene A **94**

malettinin A **92**

xenovulene B **93**

B
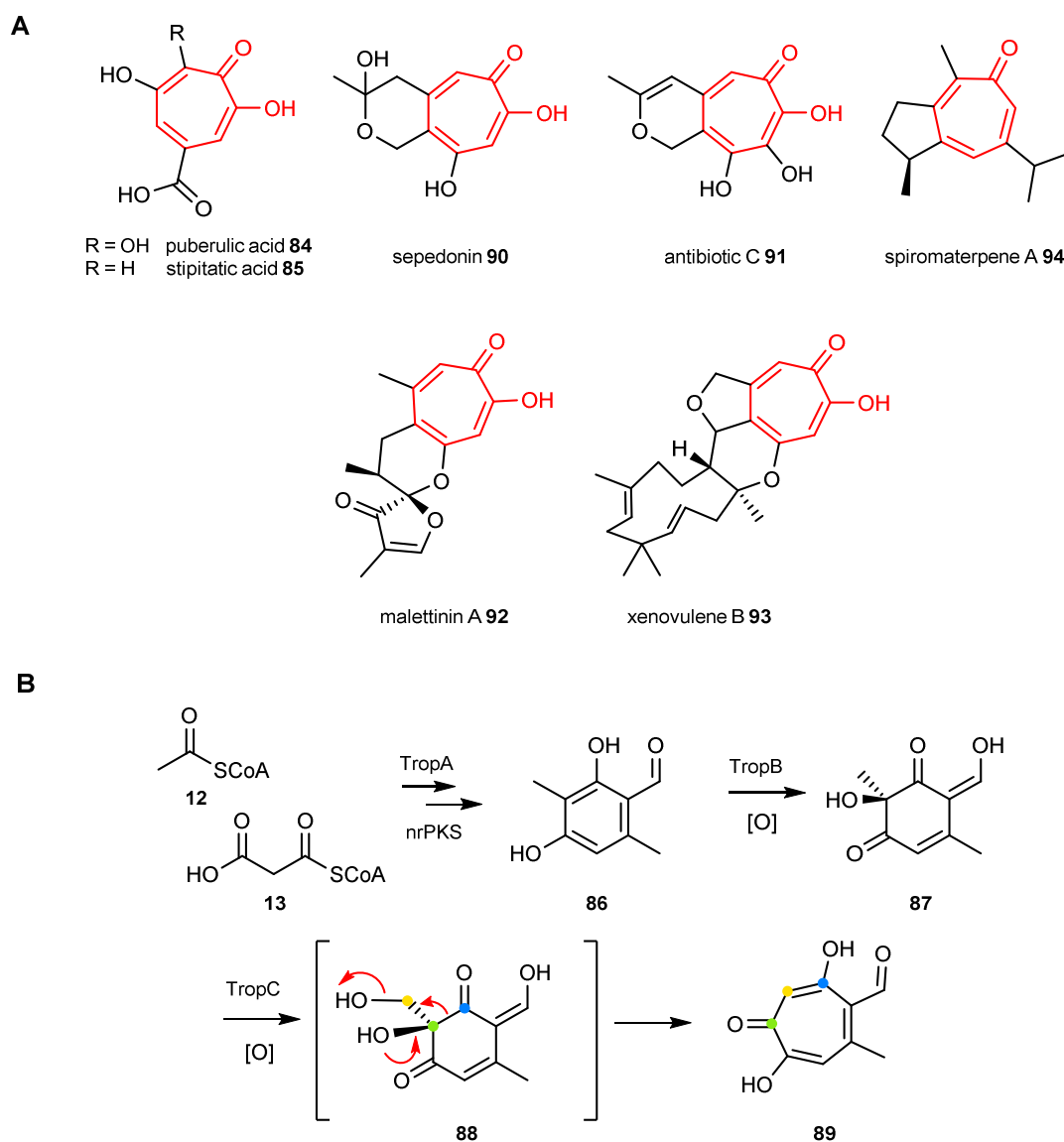
Figure 1.13 Fungal tropolones: **A,** structures of fungal tropolone compounds; **B**, key steps of tropolone biosynthesis in stipitatic acid pathway.

More complex PKS-derived structures like bicyclic sepedonin **90** and antibiotic C **91**, malettinins **92** or the xenovulene B **93** (a tropolone-sesquiterpenoid) are known.[94,95] Examples like sepedonin **90** have been investigated since 1965, when it was first found in cultures from *Sepedonium chrysospermum*.[96] The structure was confirmed only a few years later, but the biosynthesis remained unknown.[97]

In 2022, fungal tropolone structures derived from a terpene biosynthetic origin were reported.[88] Spiromaterpene A **94** from *Spiromstix* sp. is derived from a sesquiterpene, undergoing a terpene cyclisation and reaction by a multifunctional cytochrome P450 enzyme results in the core tropolone structure.[88]

1.5 Overall Aims

The overall aim of this work is the understanding of the biosynthesis of polyketide-derived compounds from fungi. The content of this thesis is separated into two different projects including subprojects. While the tenellin-project will focus on the *programming* of iterative hr-PKSs itself, the lienhwalide project focuses on *biosynthetic pathways*, including numerous tailoring and modification steps, leading to the final compounds.

The understanding of the programming in fungal iterative hr-PKS systems by rational engineering TenS is the focus of the first project (chapter 2). The intrinsic chain-length programming will be investigated. Previous work has already shown that a specific region of the KR domain of TenS is particularly crucial for the chain length determination. Rational amino acid swaps will be developed. TenS will be compared to related fungal iterative hr-PKS in structural protein models and sequence comparison. Therefore, an AlphaFold structural model for TenS will be obtained, analysed and used to interpret and rationalize the results of the effect of amino acid swaps. The investigation will be performed by heterologous expression experiments with the different hybrid variants of TenS and analysis of the obtained extracts by LCMS. Additionally, an alanine scan will be performed to delineate relevant amino acid positions of the TenS KR structure.

Secondly, the lienhwalide project (chapter 3 and 4) focusses on the elucidation of the biosynthetic pathway of lienhwalides **130-132** and related compounds, produced by the fungus *Hypoxylon lienhwacheense*. As these natural products include two classical fungal compound structures (alkyl citrates and tropolones), the project is also divided into two subprojects. The investigation for the elucidation of both pathways starts with genome sequencing and bioinformatic gene prediction methods, which will lead to the discovery of two gene cluster candidates.

The first focus is the biosynthesis of maleic acid anhydride compounds. The biosynthesis is speculated to be closely related to the formation of maleidrides in the early steps based on gene homology. The biosynthetic steps of the pathway will be studied by heterologous co-expression of the putative core enzymes of the pathway. Additionally, an unexpected compound **143** will be found during these experiments, whose biosynthetic origin will be further investigated by *in vitro* methods with the

purified alkylcitrate synthase of the gene cluster. While the formation of the maleic acid anhydride monomeric units was studied before, the formation of the trimeric, linear compound cordyanhydride B is unknown and will also be investigated.

The second subproject is about the biosynthesis of sepedonin **90**-related pentaketide tropolone compounds. The first step is the bioinformatic analysis of the putative BGC. After the analysis of the BGC, a synteny analysis of 35 putative tropolone BGCs from newly sequenced species will be carried out. A high similarity between the BGCs between nearly all investigated species suggests an important ecological role of the BGC. The putative proteins for the formation of the core carbon skeleton will be determined by heterologous expression experiments. Isolation of the core proteins and subsequent reaction *in vitro* will be carried out to elucidate the pathway in more detail.

With the two decoded biosynthetic pathways on hand, the last step of the biosynthesis of lienhwalides will be studied. The final project aim is to elucidate the coupling reaction of maleic acid anhydride monomers to sepedonin. The investigations will be carried out by heterologous co-expression of two full BGCs, which is an interesting starting point for further investigations of the final biosynthetic steps in *H. lienhwacheense*.

2   Rational Engineering of the hr-PKS TenS

2.1 Introduction

Previously described experiments showed that domain swaps in TenS changed its programme. The programme is mainly influenced by the *trans*-ER, *C*-MeT, KR and ΨKR domains of the hr-PKS (section 1.4.1).[54,61] While the *C*-MeT seems to mainly control the methylation pattern, the KR domain influences the chain length. The presence or absence of a *trans*-ER is involved in the control of the fidelity of the programme. The chain length (and therefore the KR domain) will be the focus of investigation in this work.

2.1.1 Reaction of KR Domains: Stereochemistry and Structure

The KR domain is an essential domain of FAS systems and is very frequently found in iterative and modular type I PKS. Since vFAS and modular PKS KR domains have been well studied, these form useful model systems for the study of the much-less studied iterative type I PKS KR domains of fungi, as found in TenS.

2.1.1.1 KR Domains of FAS

The crystal structure analysis (section 1.2.3) of vFAS shed light to the structural organisation of the KR domain within the vFAS. The KR is monomeric and was distinguished into an *N*-terminal structural subdomain (ΨKR), which stabilizes a *C*-terminal catalytic subdomain (KR, Figure 2.1 A). The ΨKR is suggested to be a truncated KR domain, as it is approx. half of the size of the KR and has lost the ability to bind NADP(H).[29] The KR and ΨKR subdomains interact as a pseudo-didomain.[98] The KR/ΨKR didomain is the central connector of the modifying domains of vFAS, interacting additionally with the DH, ER and Ψ*C*-MeT (Figure 2.1 A). Although KR and ΨKR are in close physical contact in the vFAS structure (section 1.2.3, Figure 2.1 A, B), at sequence level the ER domain is inserted between the KR subdomains (Figure 2.1 B, C).
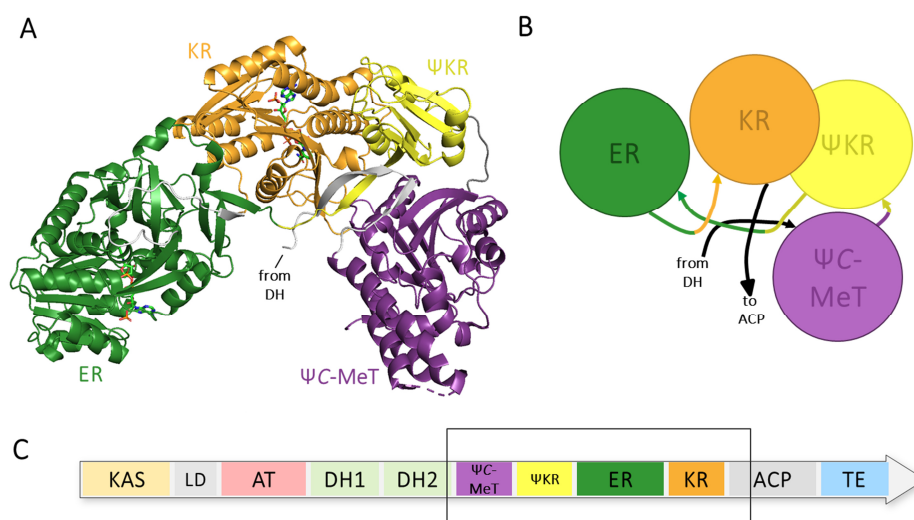
Figure 2.1 Protein architecture of vFAS modifying domains: **A**, crystal structure (PDB 2vz9); **B**, schematic overview; **C**, FAS domains sequence; diagrams based on Maier *et al.*[28]

The stereochemistry of reactions catalysed by the KR domain of vFAS was studied *in vitro* extensively in the 1980s by Hammes and Cornforth. They reported the KR reduction from the *si*-face of the 3-keto moiety of fatty acid intermediates to give the *R*-alcohol by transfer of the 4'-*pro-S* hydride from NADPH (Scheme 2.2, Scheme 2.3).[98,99] The results were in agreement with investigations of many other FAS from different organisms from eukaryotes (yeast,[100] pigeon/rat liver,[101] castor bean seeds,[102] *Chlorella vulgaris*)[102] and prokaryotes (*Brevibacterium ammoniagenes*)[103] which all reported the reduction by the transfer of the 4'-*pro-S* hydride of NADPH.

2.1.1.2 KR Domains of PKS

In contrast to vFAS KR domains, later investigations surprisingly showed that the KR domains of modular PKS *e.g.* DEBS are able to produce hydroxyls of both *R*- and *S*-configurations, depending on the module. In module 1 of DEBS, reduction results in the 3-*R* configuration while the 3-*S*-hydroxyl is the product of modules 2, 5 and 6.[98]

Swapping experiments of the modular engineered PKS DEBS1 KR, including two modules fused to the TE domain of DEBS3 (Scheme 2.1 A), showed that the KR domain sets the orientation for the chiral centre of the 3-hydroxyl group.[104–106] The natural KR domain EryKR2 was swapped with the RapKR2 (and the inactive RapDH2[0], Scheme 2.1 B). The EryKR2 domain is known to produce polyketides with 3-*S*-orientation (like **99**, Scheme 2.1 A) while RapKR2 is expected to produce 3-*R*-

stereocentres (**100**, Scheme 2.1 C). The swap experiment proved a change in the orientation of the resulting polyketide **101** (Scheme 2.1 B).[105–107]
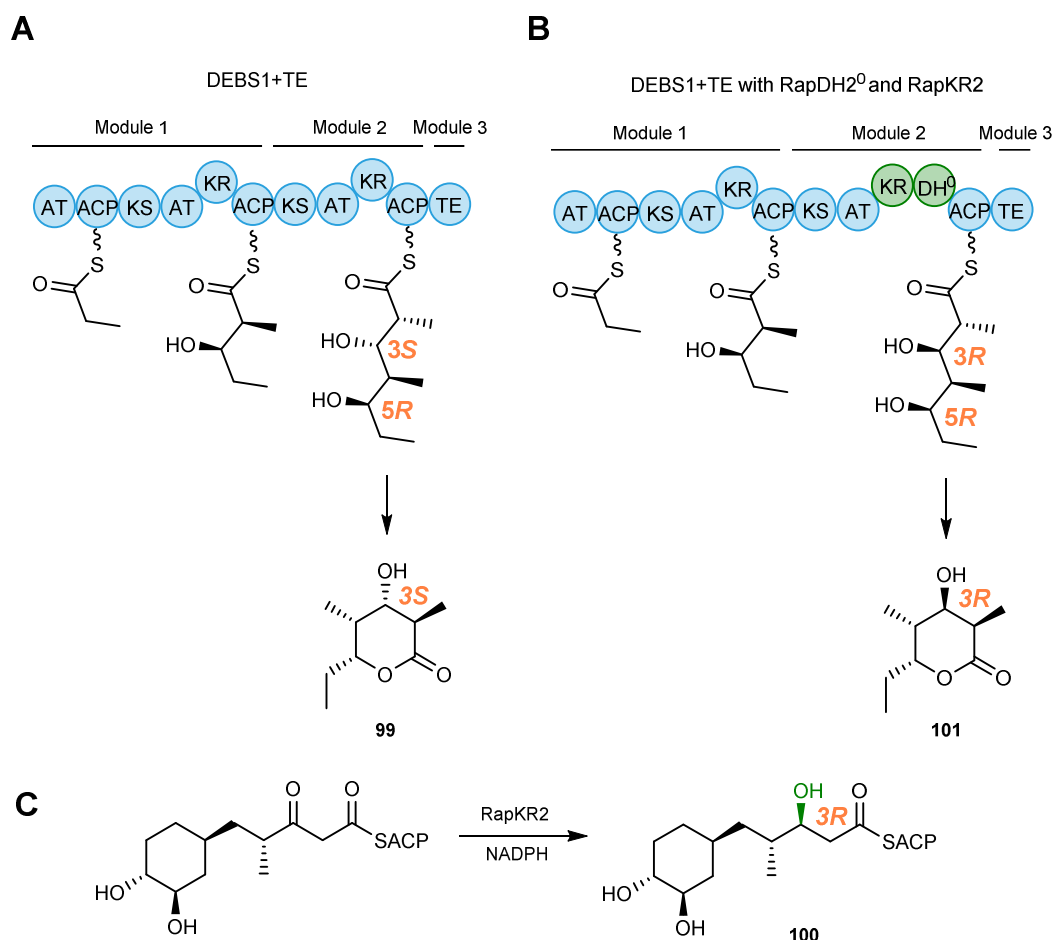
Scheme 2.1 Reaction of engineered DEBS1+TE: **A**, reaction of DEBS1+TE; **B**, reaction of DEBS+TE with swap of EryKR2 with RapDH2[0] and RapKR2; **C**, native reaction of RapKR2.[105,107]

KR domains from vFAS and PKS belong to the family of short-chain dehydrogenase/reductases (SDR). SDR enzymes are NAD(P)H-dependent oxidoreductases, sharing similar sequence motifs and mechanisms across the SDR family. The characteristic core Rossmann fold binds the cofactor in the same location and conformation for different types of SDRs (Figure 2.2 B) by a conserved NADPH binding motif (often TGGTGxLG).[28,108] All SDR enzymes include a highly conserved, catalytic tyrosine and a conserved asparagine, serine and lysine residues (Scheme 2.2, Figure 2.2 C). FAS and PKS KR domains contain the lysine and asparagine residue at swapped positions in comparison to other known SDR enzymes.[28] The conserved catalytic tyrosine and a serine residue are positioned similarly in all SDRs and bind and

polarise the 3-keto group of a polyketide intermediate.[98] The NADPH 4'-*pro-S* hydride attacks the β-carbon **102** as the tyrosine residue donates its proton to the carbonyl oxygen (Scheme 2.2), resulting in the alcohol **103a**.
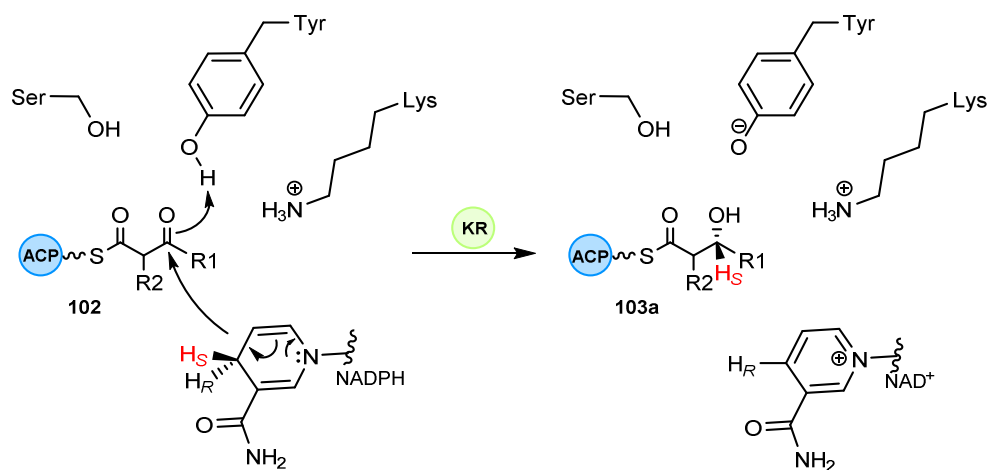
Scheme 2.2 Catalytic site and reduction of KR domains with cofactor NADPH.

The conserved lysine residue is thought to decrease the pK_a of the tyrosine residue and it orients the tyrosine hydroxyl proton towards the substrate.[109] The conserved tyrosine and lysine are in contact to the nicotinamide ribose in addition to the contacts of the NADPH binding motif (Figure 2.2 C).[107] Furthermore, a lid helix was observed in KR domains of modular PKS, which was speculated to form a flexible clamp over the bound polyketide chain (Figure 2.2, dark blue).[107,109]
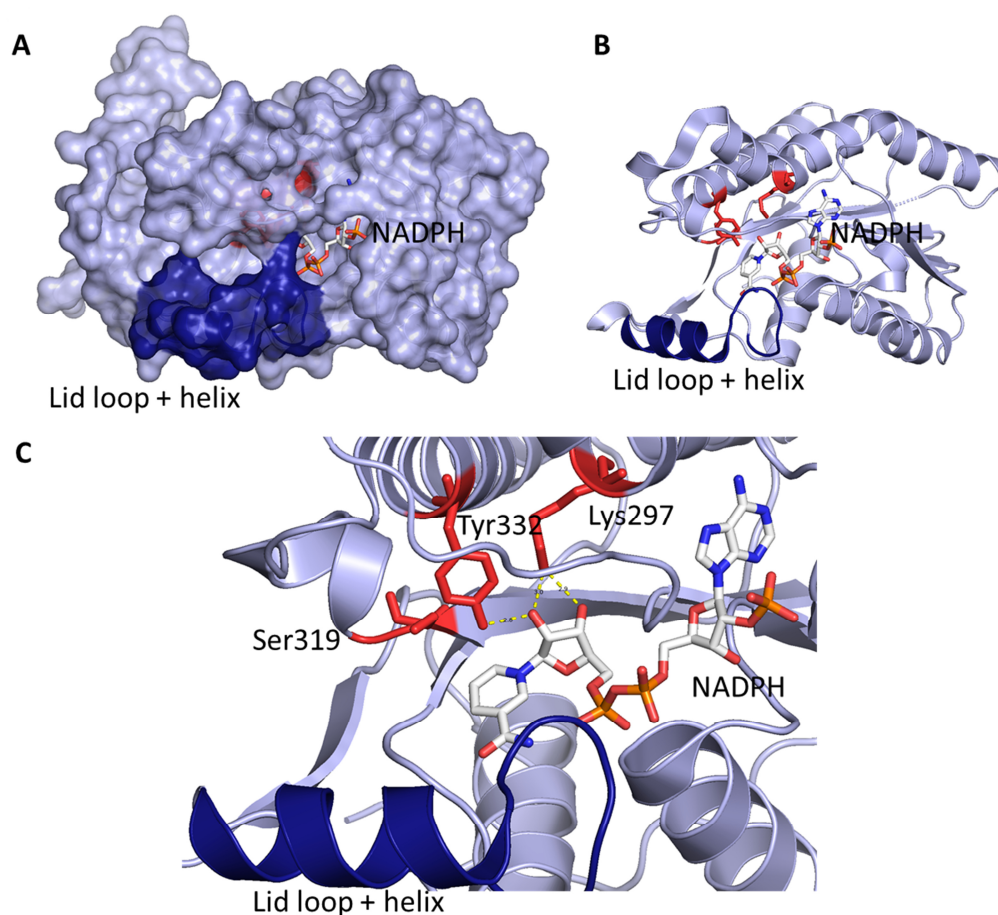
Figure 2.2 Domain architecture of PlmKR1 (PDB 4HXY): **A**, visualization of the protein surface; **B**, overall protein architecture; **C**, active site with NADPH and conserved amino acid residues; diagrams based on Bonnett *et al.*[109]

Based on the resulting β-hydroxyl-stereocentres, KR domains of modular PKS can be classified into different types A-C, which correlate with differing conserved residues of the peptide sequence (Table 2-1). The absolute configuration description after Cahn-Ingold-Prelog (*R/S*) is used for stereocentres with the generally accepted agreement that the head of the polyketide takes precedence over the tail.

Since in PKS and FAS the 4'-*pro-S* hydride of NADPH is transferred in the same orientation relative to the position of the conserved tyrosine, the different possible *R/S* stereospecific outcomes are rationalized by the direction in which the ACP-bound substrate enters the active site pocket (Scheme 2.3).[99,110]

The interaction between KR domains with the ACP domain, which carries the intermediate, was investigated for modular PKS. The interface between ACP and KR domains is located adjacent to the lid helix, near the NADPH binding site of the KR domain or directly at the lid helix.[111–113] This position enables the phosphopantetheine

arm to enter from either of two directions, depending on the KR domain type (Scheme 2.3).[112]
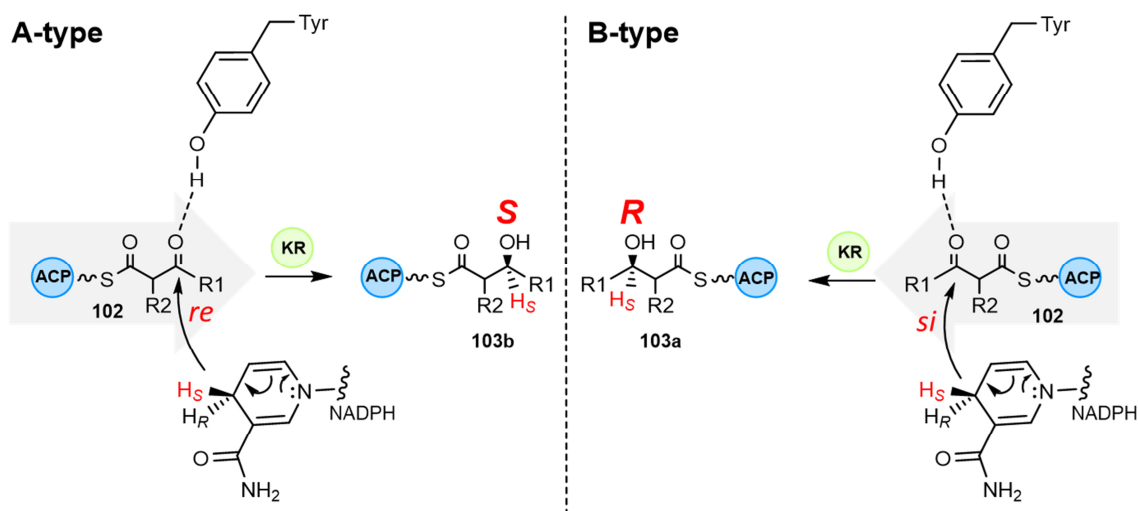
Scheme 2.3 Substrate orientation in A-type and B-type KR domains.
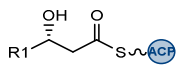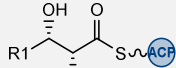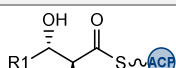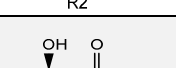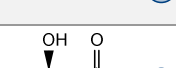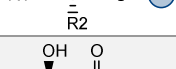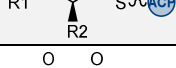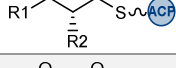
All A-type KRs produce *S*-β-hydroxyls **103b** through the attack of the *re* face of the β-keto group by the NADPH 4'-*pro-S* hydride (Scheme 2.3). For A-type KRs it was shown that a closing of the lid loop, covering over the cofactor and substrate binding site, prevents access of the substrate from an incompatible direction to result in the experimentally observed stereocentre.[109] A conserved tryptophan (W) is present in A-type KRs, located at a loop adjacent to the active site.[107,114]

In contrast, in B-type KRs (like TenS) the *si* face of the β-keto intermediate **102** is attacked by the NADPH hydride to result in *R*-β-hydroxyl groups (**103a**, Scheme 2.3). The conserved LDD/VDD-motif of B-type KRs is located on the opposite side of the active site in comparison to the conserved tryptophan of A-type KR domains. In the LDD/VDD motif the leucine/valine residue is occasionally replaced by an isoleucine residue motif; the second position can be found replaced by several residues, while the third aspartate is strictly conserved.[114] The β-keto substrates of A- and B-type KRs are bound from different sides in the substrate-binding groove to result in opposite hydroxyl group orientation (Scheme 2.3). The characteristic conserved amino acids (Table 2-1) are hypothesized to take part in the control of the entry direction of the polyketide intermediate by an unknown mechanism.[109] According to this classification, vFAS KR domains have similarity to B-type KR-domains, as the reduction results also in *R*-hydroxy groups.[114]

Moreover, KR domains also control the orientation of alkyl substituents at the α-position by epimer-specific reductase activity. This was demonstrated by feeding racemic mixtures of α-substituted β-ketoacyl-*N*-acetylcystamine thiolester substrate analogues to KR domains *in vitro*. *R*-α-substituents are either reduced by the KR without inversion of the stereocentre or are epimerized to α-*S*-orientation in the specific KRs before reduction.[107,114] The α-substituent orientation defines the subtypes of the KR classes. KRs reducing α-unsubstituted intermediates belong to A0 or B0-type KRs. If an *R*-α-substituent is present in the resulting intermediate, the responsible KRs belong to A1 or B1 subtypes. Accordingly, KRs with reactions resulting in *S*-α-substituents belong to A2 or B2 subtypes.

C-type KRs have no β-reductive activity. C1-type KRs are catalytically non-functional in contrast to C2-type KRs, which have an epimerase activity, converting *S*-α-substituents to *R*-α-orientation. The mechanism of the epimerase activity is unknown to date. Based on the presence of the four conserved catalytic residues (tyrosine, serine, asparagine and lysine) in C2-type KRs it is likely that they play a role in epimerisation.

Table 2-1 Overview of KR domain types and chiral centres of the resulting products.

KR type	orientation in product		structure of the product	conserved residues (Y = catalytic tyrosine)
	β-hydroxyl	α-substituent		
A0	*S*	-		W
A1	*S*	*R*		W
A2	*S*	*S*		W and HXX**Y**
B0	*R*	-		LDD or VDD
B1	*R*	*R*		LDD or VDD
B2	*R*	*S*		LDD and **Y**XP
C1	-	-		-
C2	-	*S*		**Y** present, but no NADPH binding motif

2.1.2 Sub-domain Swap Experiments of TenS

Detailed sub-domain swaps experiments (Figure 2.3), were reported in 2019 by Cox and co-workers between the sequences of the related hr-PKS DmbS (mainly producing hexaketide **33**) and the militarinone C synthase (MilS), producing doubly-reduced trimethylated heptaketide **34** (Figure 1.7).[58] The KR and ΨKR domain sequences were divided into sub-fragments, which were swapped to generate chimeric ΨKR/KR domains in the TenS backbone.

Each hybrid PKS was expressed in the host *A. oryzae* NSAR1 together with the *trans*-acting ER domain TenC. The resulting products of the hybrid PKS were analysed in terms of chain length and methylation pattern (Figure 2.3). Swaps within the sequence of the ΨKR domain did not lead to a change in chain length of the polyketide products, leading the focus on to KR domain itself.

Cox and co-workers observed that the control of the programming is not solely based on the intrinsic substrate-selectivity of each catalytic domain, but is most likely a combination of intrinsic (section 2.1.2.2) and extrinsic (section 2.1.2) factors, which are in a complex balance of relative selectivity to result in the programme of TenS.

2.1.2.1 Extrinsic Programming and Competition between the Catalytic Domains

The conclusions on the extrinsic programming of TenS from Cox and co-workers were based on results from domain-swap experiments. All sub-fragment swap experiments (Figure 2.3 B-J) usually did not lead to the production of exclusively one product, but rather to a mixture of products with differences in more than one structural feature of the polyketide. For example, a chimeric KR domain consisting of approx. 50 % DmbS sequence (Figure 2.3 F) gave pentaketides **39**, **104**, **41** and **45** and low titres of the hexaketide **46** that not only differ in the chain length, but also in the methylation pattern. The selectivity of both, the KR *and* the *C*-MeT domains were changed by this swap that only included KR domain sequence areas. This suggests a competition between the KR and *C*-MeT domain at the β-ketothiolester stage. Furthermore, the production of **41** is increased in comparison to the wild type TenS, giving evidence for a change in *trans*-ER selectivity.

Another example, which portrays the extrinsic programming is the complete swap of the KR domain to the sequence of DmbS (hexaketide) or MilS (heptaketide, Figure 2.3 E, I), which did not abolish the production of the native pentaketide product **39**. Furthermore, a swap of the KR domain sequence of the last quarter of the sequence lead to a completely inactive synthase (Figure 2.3 D).
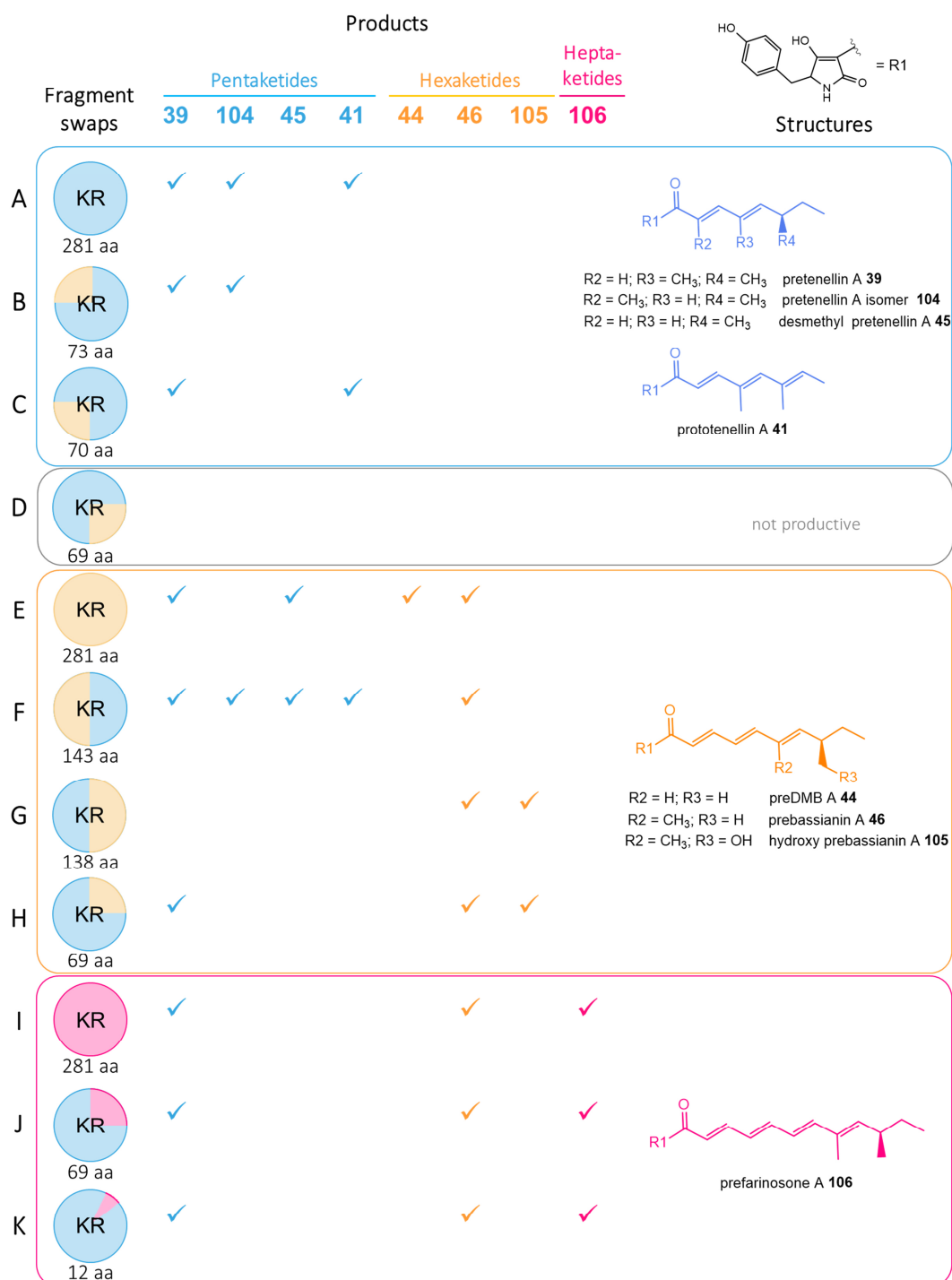
Figure 2.3 Overview of fragment swap experiments (blue = TenS, orange = DmbS, pink = MilS).[58]

The swap experiments (Figure 2.3) gave evidence that the catalytic domains compete for the ACP-bound intermediates and the rates for this competition are influenced by extrinsic factors that do not involve changes in the active sites. Consequently, the overall selectivity cannot be changed completely by swaps focusing on only one domain due to likely protein-protein or domain-domain interactions.[58] The finding is in agreement with previously described results from Simpson and co-workers, where the cleanest change in selectivity was observed by a swap of the complete *C*-MeT-ΨKR-ER^0-KR tetra-domain that maintains these intra-domain interactions.[63]

A delicate balance of each competition situation explains the programming during polyketide biosynthesis (Scheme 2.4). The intrinsic and extrinsic components, the kinetic parameters for each domain and for a specific ACP-bound intermediate affect the outcome of each step individually. Cox and co-workers suggested that the DH domain and the KS/AT domain reaction is controlled chemically with low selectivity – the DH always acts if an appropriate β-alcohol is present after the previous biosynthetic step (Scheme 2.4, black arrows) and the AT domain is continuously suppling additional extender units, while the KS domain acts only on non-β-keto intermediates.

The other domains (*C*-MeT, KR, ER, and C domain) are in competition for the ACP-bound substrate (Scheme 2.4, red arrows), as the domains are all catalytically competent to react on the growing ACP-bound intermediate chain. After the first extension cycle of the KS domain, acetoacetyl-ACP **14** is methylated to **108** due to the selectivity of the *C*-MeT domain, which reacts the fastest for diketides in TenS, DmbS and MilS (Scheme 2.4 A), instead of reduction to **107**. The α-acetyl-diketide **108** can still be a substrate for the KR, which consequentially catalyses the reduction, followed by the DH dehydration and saturation by the ER domain to fully saturated methyl diketide **109**. Extension by the KS domain then results in **110**. After the second extension, the next branching point is reached (Scheme 2.4 B). In DmbS the KR reacts fastest (with following reaction of DH domain to **111**), preventing methylation. In the other PKS, again the *C*-MeT reacts fastest, resulting in a suitable substrate **112** for the KR and DH leading to **114**. This competition process repeats after each extension reaction. After the third extension (Scheme 2.4 C), the KR domain reduces **115** to **116** in TenS and DmbS, while for MilS, a methylation of **115** to **117** takes place before the KR reaction.
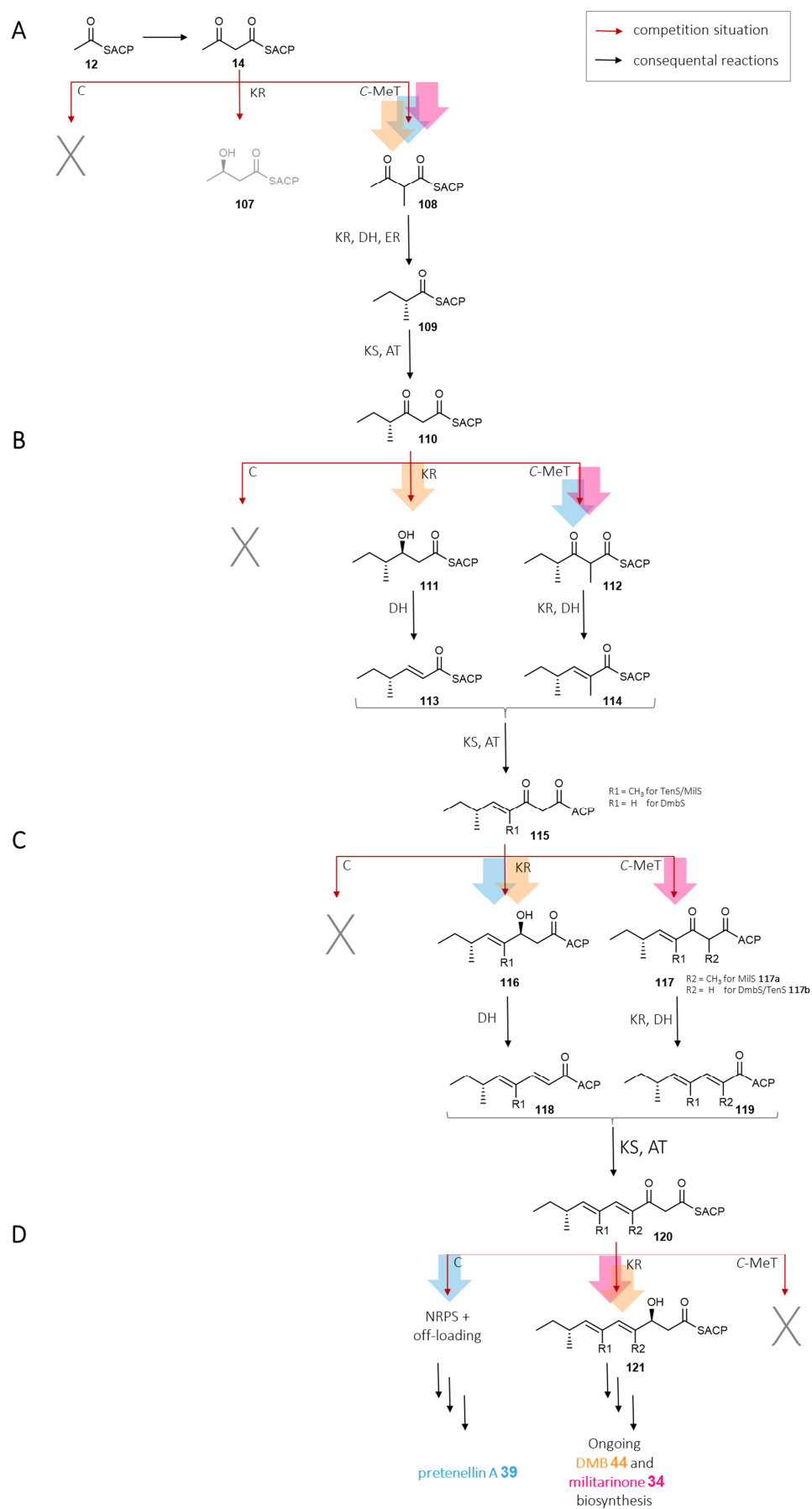
Scheme 2.4 Competition of domain reactions in TenS/DmbS/MilS; **A-D**, respective step of elongation.

During tenellin **30** biosynthesis, the fourth condensation resulting in β-keto-pentaketide **120** is the final extension step. Thus, chain-release by the C domain outcompetes reduction, methylation or further chain extension. In contrast, in MilS and DmbS the reduction by the KR domain outcompetes the chain release (Scheme 2.4 D) and so the chain is reduced, dehydrated to **121** and extended again. Hence, failure of the KR domain at a particular chain-length terminates chain growth. At this point the swapping experiments (Figure 2.3 F, I) can lead to hexaketide production, suggesting that the KR reaction speed is increased, reacting faster than the chain release, as in case of the wild type TenS.

The competition between the domains continues for longer intermediates in DmbS and MilS until the β-ketothiolester intermediates reach the expected chain length. When no other domain catalyses a reaction, the NRPS C domain removes the ACP-bound polyketide, followed by Dieckmann cyclisation by the DKC domain to release the first enzyme free intermediate. The off-loading process is determined by two main factors: it acts selectively on β-ketothiolesters and the chain release is the slowest step in most cases.[58] As triketides like **110** are never observed during swapping experiments, the C domain only acts, if the reaction speed of other domains decreases, as in case of TenS with structure **120**. Overall, the structure of the resulting polyketide gives information of the relative reaction kinetics of the separate domains during the domain-domain competition. For example, structure **30**, **32**, **34**, **35**, and **121** show that methylation is less likely to occur with an increase of the chain length.

2.1.2.2 Intrinsic Programming of TenS

In addition to the extrinsic protein-protein interaction, intrinsic factors also play a role in chain length programming. This is illustrated by an observation by Cox and co-workers that involved a swap experiment including the second half of the KR domain sequence (length of 138 amino acids), which resulted in exclusively hexaketides **46** and **105** (Figure 2.3 G). Furthermore, a fragment with the length of 69 amino acids within the second half of the TenS sequence (positions 2343-2411) also led predominately to the production of hexaketides **46** and **105** (Figure 2.3 H). In contrast, the swaps of the sequence fragments with similar length, but not in this region, led either to only pentaketides **39**, **104**, **45** and **41** (Figure 2.3 B, C) or no production of polyketides (Figure 2.3 D). This shows that the main impact on the intrinsic programming for chain

length is located within the 69 amino acid long fragment of the KR. This finding was confirmed by swapping experiments with MilS as donor in the same region (Figure 2.3 J), leading to **46** and the heptaketide **106**.[58,63]

To get an insight into the intrinsic control on the programming of this 69 amino acid long fragment, a "threaded" protein model of TenS was constructed (Figure 2.4).[58] The complete TenS protein structure was mapped on the architecture of vFAS due to the close relationship between vFAS and hr-PKS systems (section 1.2.3).[2] Separate domains were modelled based on modular PKS domains to get structural information of higher quality and later fused to the threaded TenS complete structure. Therefore, the amphotericin-B producing PKS AmphB2 was used as a template for the TenS KR domain with sequence identity of 27 %.[58] While vFAS and TenS both include B-type KR domains, the modular AmphB2 KR domain is an A-type KR. Nevertheless the AmphB2 KR domain had most sequence identity to TenS with satisfactory resolution and template quality, compared to all available X-Ray structures of KR domains at that time. The resulting KR domain model was fused to the vFAS overall structure for a refined chimeric model *in silico*.[58]

The 69 amino acid sequence region (TenS positions 2343-2411) was structurally evaluated in the protein structure model of the KR domain. Within the 69 amino acid sequence, an α-helix was found, which neighbours the active site groove of the KR domain. This sequence corresponds with the previously described *lid-helix* (section 2.1.1.2). The α-helix between TenS positions 2398 to 2409 was speculated to contact the substrate and was therefore further called the substrate binding helix (sbh, Figure 2.4). In the final swapping experiment, a 12 amino acid long swap of the substrate binding helix sequence between TenS and MilS (Figure 2.3 K) led to the same result, as previously found with the 69 amino acid fragment sequence (Figure 2.3 J) *i.e.* production of a mix of chain-lengths in which the heptaketide **106** was a major component. This confirmed that the substrate binding helix sequence of the KR domain is involved in the determination of the chain length.[58,63]

substate-binding helix
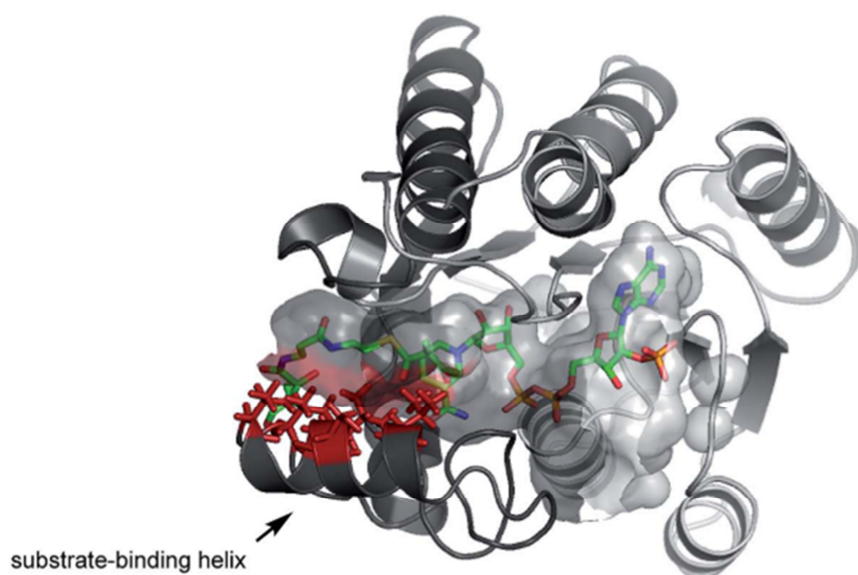
Figure 2.4 Model structure of TenS KR domain (with NADPH and 2'-methylacetoacetyl pantetheine, red = residues V2397, S2400, L2401, T2404, and V2406); diagram from Yang *et al.*[58]

2.2 Project Aims

The identification of the substrate binding helix as a significant influencing component on the intrinsic programming for TenS is the starting point for the experiments of this section. The aim of the project is to rationally engineer the KR domain of TenS to investigate its intrinsic programming in more detail. Because the previous structural model was based on an A-type KR that does not match the TenS KR domain, a new structural model of the TenS KR domain will be developed using the structure prediction tool AlphaFold, which is the basis to discuss structural characteristics in comparison to the previously reported model.

The complete substrate binding helix sequence will be exchanged with the corresponding sequences of DmbS and MilS to confirm the important role of the substrate binding helix. The investigations of chimeric TenS variants will be carried out by heterologous expression in the host *A. oryzae* NSAR1 and following standardized cultivation, extraction and LCMS analysis.

Next, the orientation of single amino acids of the substrate binding helix will be evaluated in terms of direction to the substrate binding site and will serve as the basis to rationally engineer the TenS KR. A set of only four simultaneous amino acid mutations will be carried out to re-programme TenS and observe the effect on chain length control.

In the last section, the aim will be to minimize the number of amino acid mutations, which lead to a change in the produced polyketide chain length. To further refine the identification of crucial amino acid positions, the alanine scan method will be used. Every position of the substrate binding helix will be mutated to alanine to understand the role of every amino acid position separately for the intrinsic programming of TenS.

2.3 Results

2.3.1 Analysis of TenS Structural Models

The previously reported strategy to obtain a threaded structural model with vFAS as the overall template and by using known modular PKS domains as a templates for separate domains (section 2.1.2.2) was the best methodology available at the time.[58] However, it suffered from problems associated with the choice of template for the KR domain. While sequence homology for AmphB2 KR is high (27 %), this KR domain is an A-type, which is stereochemically incorrect.

Dramatically new developments in the field of *ab initio* protein structure prediction enable the rapid construction of reliable protein models based on sequence information alone. One method is the structural prediction by using the artificial intelligence-based tool AlphaFold, which has revolutionized the field of protein modelling.[31] For example, Adams and co-workers evaluated predictions with experimental crystal structures and found that C_α atoms from AlphaFold predicted proteins differ only by a median of 0.6 Å from real structures in areas with predicted high confidence.[115]

Nevertheless, the protein structure prediction tool does not completely replace experimental structural data. Proteins are dynamic and also include flexible areas, which show different, interconverting conformations based of conditions like temperature, solution, pH or the binding of a ligand and other proteins or domains. This is a limitation for protein prediction tools. Overall, experimental data is irreplaceable to-date for ligand docking, covalent modifications or unknown interactions.[115] Another aspect, which needs to be considered, is that the AlphaFold algorithm learned and trained the protein structure predictions based on the Protein Data Bank (PDB),[116] mostly representing well-folded, packed and stable proteins, which is not always necessarily the case for every protein in every naturally occurring conformation. In particular, active sites do not always follow the rules of folding, which can lead to challenges for the prediction.[117] Nevertheless, Adams and co-workers come to the conclusion that AlphaFold predictions are mostly accurate and suggest that predictions can provide hypotheses for mechanisms and are an appropriate starting point for experiment design.[115]

The AlphaFold structure prediction was run with the TenS KR amino acid sequence using ChimeraX software on default settings.[31,118,119] The resulting structure was compared to the previously obtained model (section 2.1.2.2) by alignment analysis and visualised (Figure 2.5) using PyMOL (The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC.). The binding of the cofactor in the KR model was extracted from the AmphB KR domain template as in previous work.[58] The root-mean-square deviation of atomic positions (RMSD) value of 0.88 Å between the two structures indicates a high structural similarity between the models (Figure 2.5 A-C).
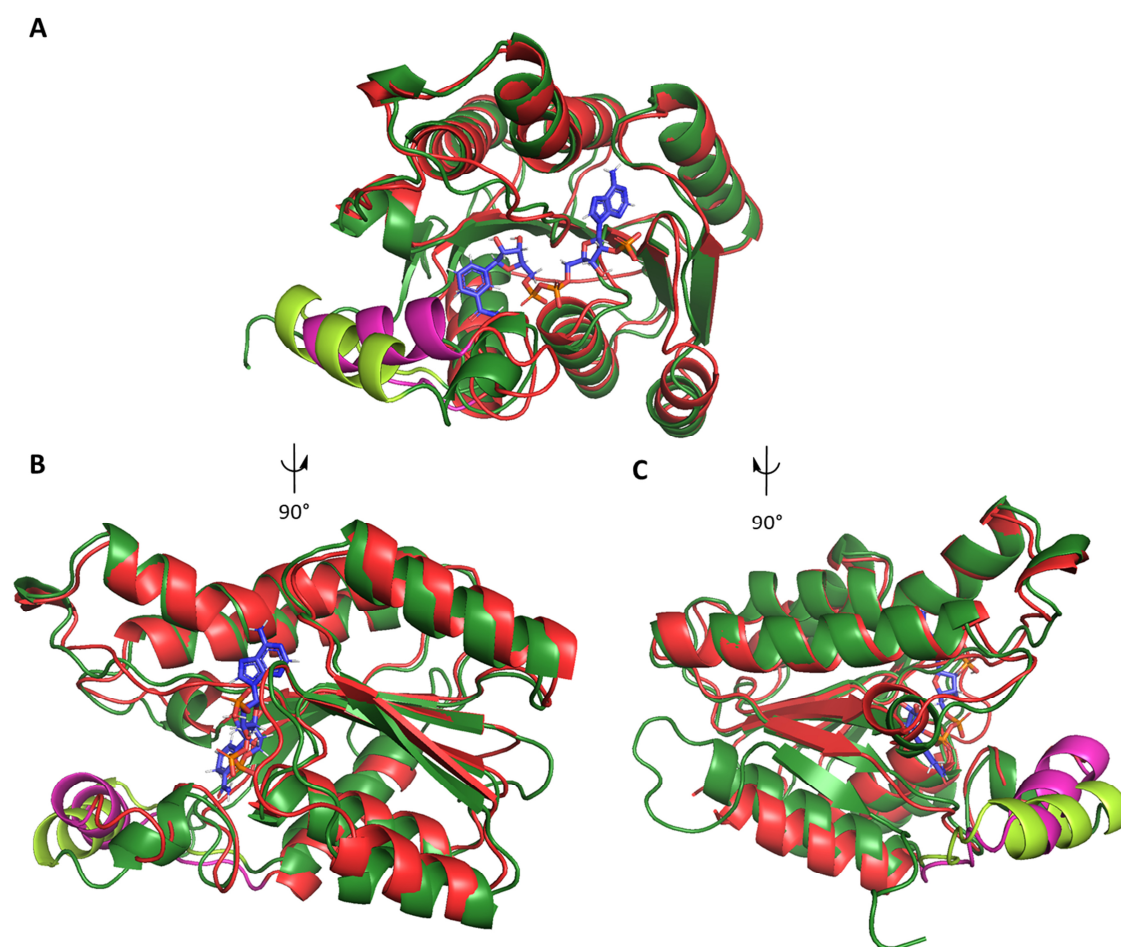
Figure 2.5 Alignment of protein structure from KR domain of TenS with previously obtained threaded model (red, substrate binding helix magenta) *vs.* AlphaFold model (green, substrate binding helix light green): **A**, front view; **B**, side view; **C**, side view; blue = NAD (docking according to previously obtained model).

The substrate binding helix (between T2395 and V2409) is located at the same position in both models, but differs in the angle relative to the active site (Figure 2.5, magenta and light green). For the previously obtained threaded model, the helix is clamped over

the active site (Figure 2.6 A, B). For the AlphaFold model, the active site is more open at the front (Figure 2.6 C). This observation is in agreement with the fact that the template-based model was developed on the basis of an A-type KR (AmphB2), which is known to show the closed conformation of the KR domain in experimental data.[109] However, the TenS KR domain belongs to the group of B-type KR domains. The previous selection of an A-type KR domain template thus inherently leads to a bias in the resulting model. Experimental data of B-type KR show a more open active site, consistent with the AlphaFold model.[109] Therefore, the AlphaFold model may be a better representation for the active site organisation of B-type KR domains.
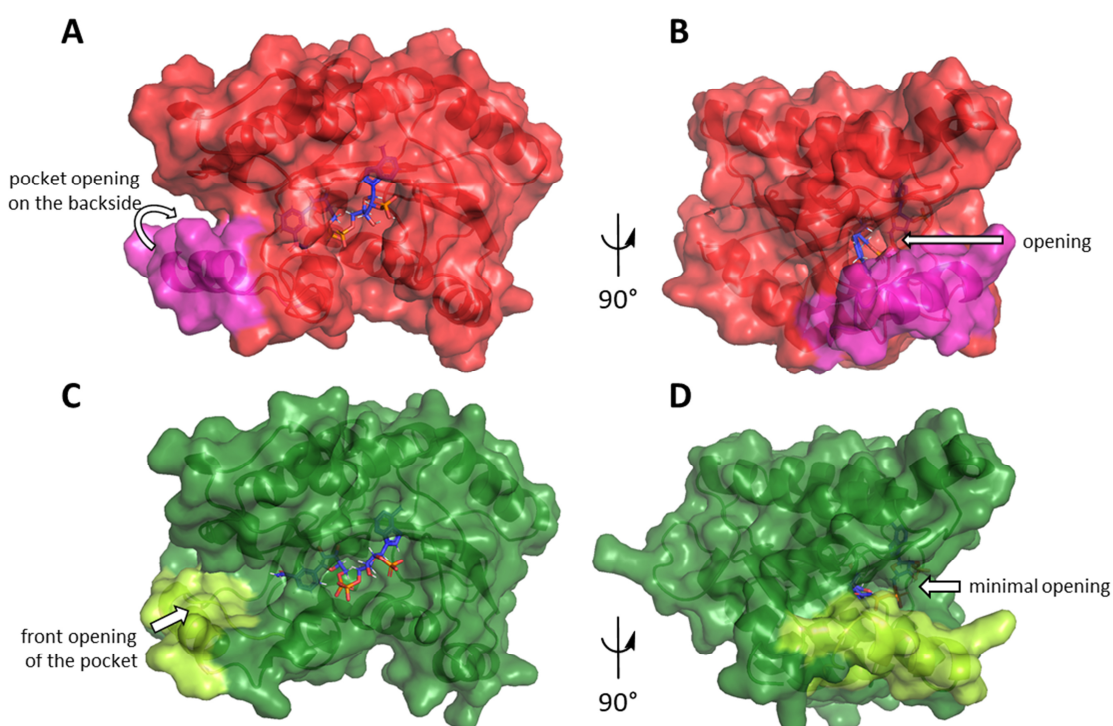
Figure 2.6 Substrate and cofactor binding pocket of TenS KR (green = NAD, yellow = 2'-methylacetoacetyl pantetheine, docking according to previously obtained model): **A**, front view of threaded model (blue, substrate binding helix cyan); **B**, left side view of template-based model; **C**, front view of AlphaFold model (red, substrate binding helix magenta) **D**, left side view of AlphaFold model.

2.3.1.1 Evaluation of Protein Structure Models

The structural prediction of TenS was evaluated using the recently published cryo-EM structure of LovB by Wang and co-workers (described in section 1.2.4).[20] The LovB cryo-EM structure is the most precise model of an iterative hr-PKS KR domain
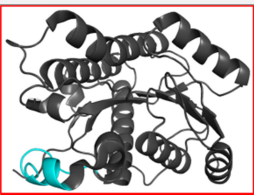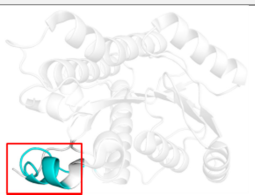
available and is therefore a suitable structure for evaluation. Like TenS, it includes the conserved LXD sequence motif for B-type KR domains. The complete LovB structure was obtained from the SwissProt[120] database (identifier 7cpx) and all protein domains outside the KR domain (preserving the bound cofactor) were manually removed. The AlphaFold model, the threaded KR model and the experimentally obtained LovB structure were visualized using PyMOL and aligned against LovB for comparison (Table 2-2). The low RMSD value of 0.90 Å (Figure 2.7 A, B) for the alignment of the AlphaFold model against the LovB KR structure indicates a more correct approximation for the protein model rather than the previously obtained template-based model (RMSD of 1.25 Å for the alignment of the template-based model against the LovB KR domain, Figure 2.7 C, D).
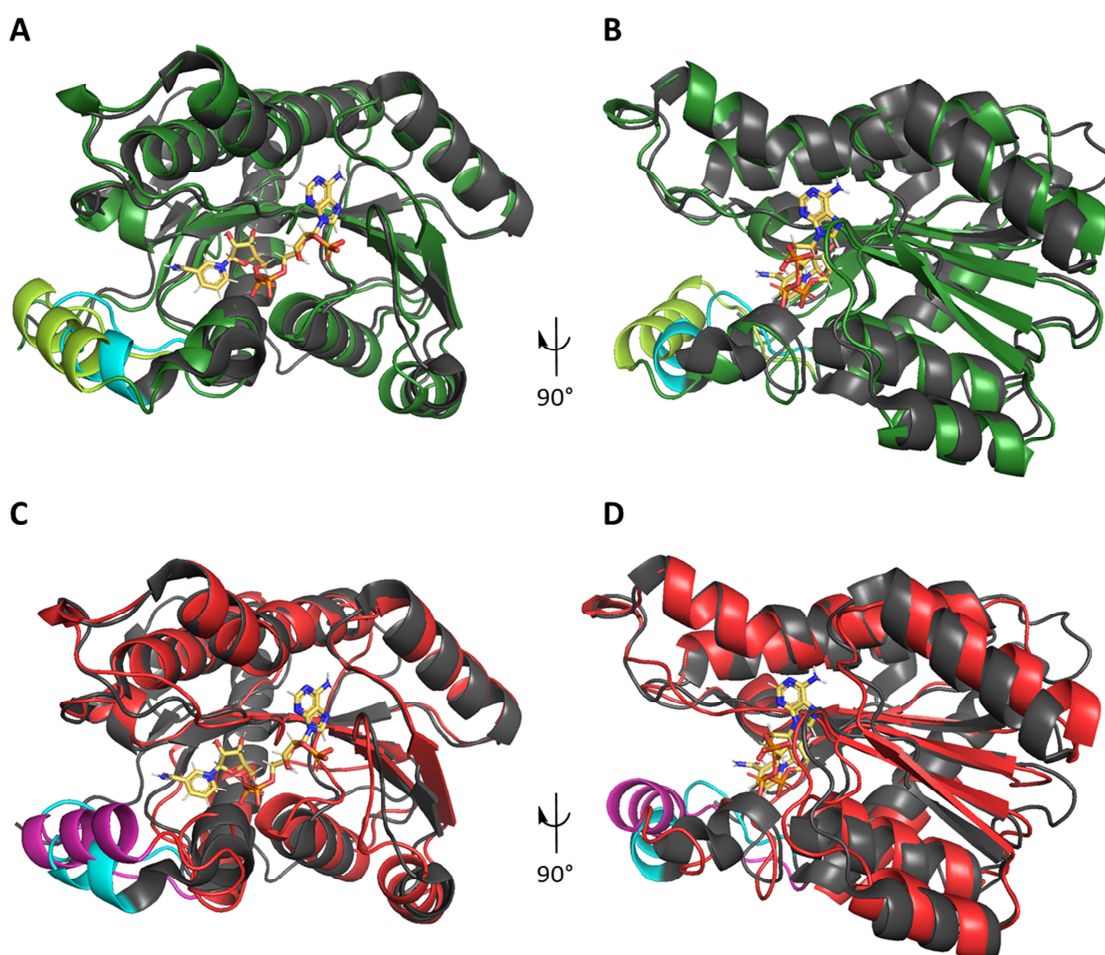
Figure 2.7 Alignment of TenS KR models to LovB cryo-EM KR structure: **A**, TenS AlphaFold model *vs.* LovB in frontal view; **B**, TenS AlphaFold model *vs.* LovB in side view; **C**, TenS template-based model *vs.* LovB in frontal view; **D**, TenS template-based model *vs.* LovB in side view (grey = LovB, green = AlphaFold, red = threaded model, yellow = cofactor from experimental LovB data).

The substrate binding helix area of the LovB sequence was determined by the structural alignment against the TenS KR domain and confirmed by sequence alignment against the TenS KR domain using BLASTp.[121] The substrate binding helix of the three structures (Figure 2.7, cyan/magenta/light green) is the area with most visible differences over all three structures. When considering only the areas outside the substrate binding helix (for TenS between T2395 and V2409), the RMSD values decrease from 1.25 Å to 1.18 Å for the alignment of the template based model to the LovB KR (Table 2-2). Accordingly, the same procedure results in a minimal lower RMSD for the AlphaFold model alignment with the LovB KR structure from 0.90 Å to 0.87 Å. Additionally, a structural alignment of only the substrate binding helix of LovB leads to overall higher values of 3.03 Å (alignment to the template-based model) and 2.81 Å (alignment to the AlphaFold model, Table 2-2). The comparison of RMSD values to the LovB structure demonstrate that the majority of the structural differences in the proteins were focussed in the substrate binding helix, illustrating a possible high flexibility in this structural component.

Table 2-2 Structural comparison and resulting RMSD of LovB with threaded and AlphaFold model.

Model	RMSD to LovB [Å]		
compared area	complete domain	excluding sbh	only sbh
Threaded model	1.25	1.18	3.03
AlphaFold model	0.90	0.87	2.81

In all three comparisons, the AlphaFold model shows a lower RMSD to the experimental LovB structure than the threaded model, indicating a higher accuracy and therefore this model will be used for the further structural analysis of the TenS KR.

2.3.1.2 Substrate and Cofactor Docking for TenS

The cofactor NADP(H) and the substrate analogue 2-methylacetoacetyl pantetheine **125** (Figure 2.8) were introduced into the AlphaFold structure. The NADP(H) was extracted from the AmpKR2 structure (identifier 5xwv) and inserted into the TenS structure.[122] The structure was refined using the energy minimizing YASARA server.[123–125]

Two approaches were used to add a phosphopantethein-bound substrate **125** to the model. First Autodock vina was used. This software is a popular and effective protein-ligand docking programme.[126] Second, manual placement was also attempted.

For Autodock vina docking, the substrate was built in Chem3D (PerkinElmer) and positioned adjacent to the catalytic residues (Y2353) and the NADP(H) attacking hydrogen in PyMOL. The substrate was placed in the known direction for B-type KR domains (section 2.1.1.2), along the LRD-motif. Docking of the polyketide intermediate 2-methylacetoacetyl panthetheine **125** was carried out by AutoDock vina (with grid-box parameters centre: x = 195.889, y = 225.294, z = 145.298; size: x = 18, y = 8, z = 6).[125,127,128] Since no experimental structural data for the TenS KR or other B-type KR domains including the substrate are available, a direct validation of the molecular docking was not possible. The quality of the docking result is suggested to be low (best energy value -1.8 kcal/mol). Therefore, a visual validation of the docking was performed based on criteria like direction and distance to the cofactor and known conserved catalytic tyrosine residue. Docked poses with a distance of > 10 Å for the C_β-ketone to the catalytic residue and NADP(H) were obtained by Autodock vina. Therefore, AutoDock vina results were not used further in the work.

Manual docking was then attempted for the substrate and used for visualization (Figure 2.8, magenta). Again, the substrate **125** was built in Chem3D (PerkinElmer) and inserted into the KR domain structure in PyMOL, while entering the active site from the correct direction according to B-type KR models (Scheme 2.3) for the attack of the *si* face of the 3-keto group. While the positions of the protein residues and the cofactor were unchanged, the substrate was fitted into the active site pocket, and manually adjusted with the "3 button editing" mode of PyMOL. The fitted substrate and the protein structure were refined with YASARA. [123–125] The resulting position of the substrate is only to be regarded as a representation of a possible pose, which can be used for orientation. Conclusions on the final substrate position will require experimental confirmation.

red: Y2353, yellow: sbh, blue: NADPH, orange: LRD motif (L2296-D2298), pink: 2'-methylacetoacetyl pantetheine, dark red: NADPH binding motif (S2239, A2266, R2240, G2214-G2220)
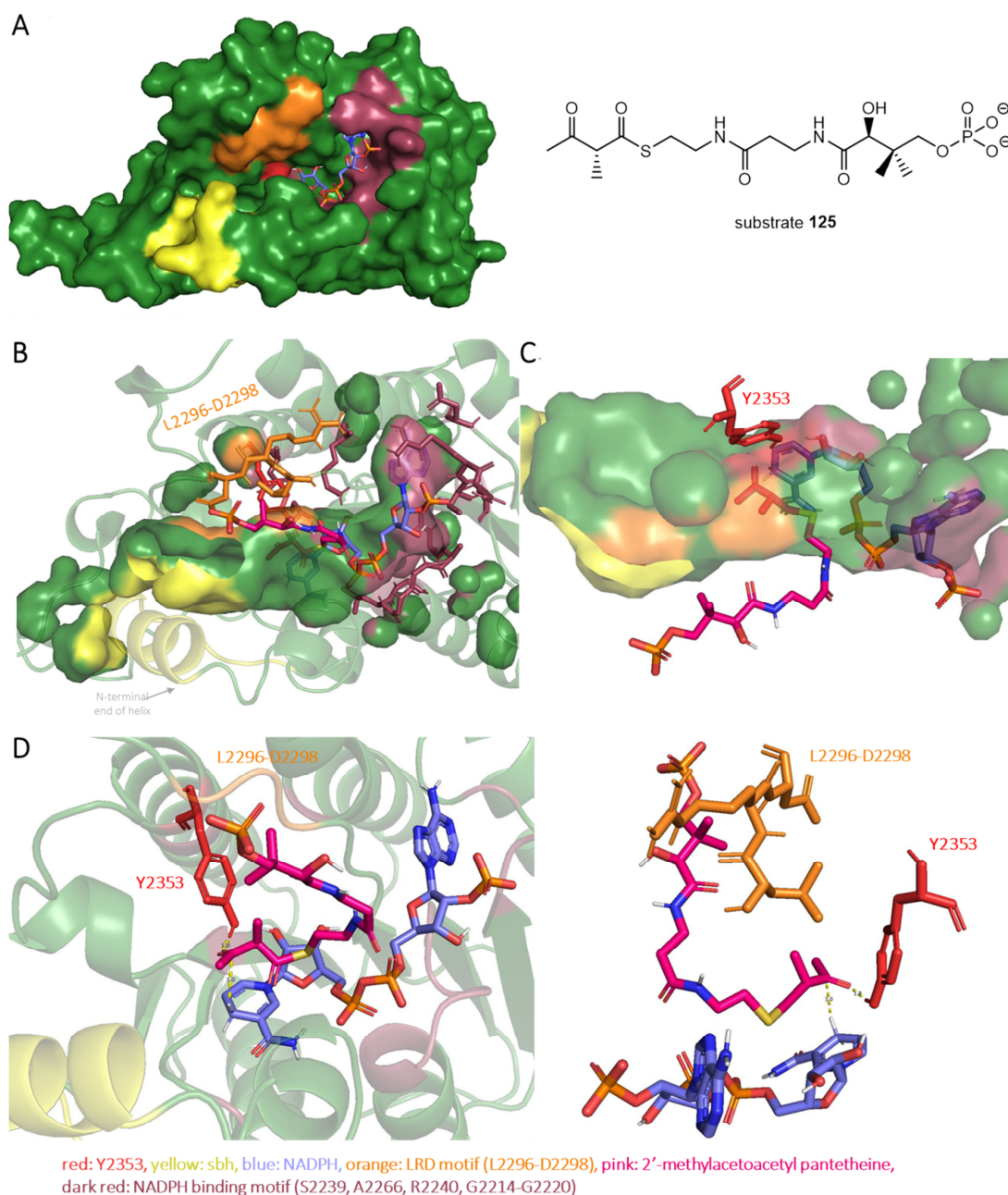
Figure 2.8 Representation of TenS active site including substrate and cofactor **A**, complete TenS KR domain with visualized surface; **B**, front view of the active site pocket; **C**, top view of the active site pocket; **D**, interactions in the substrate binding pocket from the frontal side; **E**, interactions in the substrate binding pocket from the back side.

For the manual placement (Figure 2.8 B-E) reasonable distances were achieved: the β-carbon is located 3.0 Å from the 4'-*pro-S* NADPH hydride and the carbonyl oxygen of the β-position is located 1.8 Å from the catalytic tyrosine (Y2353). The orientation of the substrate would lead to the formation of the observed 3-*R*-hydroxy stereocentre, in agreement with the known stereochemical outcome of B-type KR domains (section 2.1.1.2) The cofactor is bound by the cofactor binding motif (GAAGGLG, position

54

2214-2220) and additionally to S2239, R2240 and A2266 (Figure 2.8 A-C, dark red). The LRD-motif, which is typically found in B-type KR domains, is placed at the expected position adjacent to the active site and the modelled substrate (section 2.1.1.2).

As described previously (section 2.1.1.2), the substrate binding helix of the TenS KR domain corresponds to the *lid-helix* in the known structures of modular PKS KR domains. This helix is speculated to clamp over bound polyketide intermediates during reaction. The AlphaFold model appears to shows an open form. Although precise docking of the substrate within the active site was not possible, the contact of the substrate binding helix to the substrate appears to be very likely for longer polyketide chains (like tetraketide **126**), which reach deeper into the active site binding pocket (Figure 2.9) when closed. Thus, the substrate binding helix is part of the active site architecture, which makes it a reasonable target for mutations, as already observed by previous work of our group.[58,129]

red: Y2353, yellow: sbh, blue: NADPH, orange: LRD motif (L2296-D2298), dark red: NADPH binding motif (S2239, A2266, R2240, G2214-G2220)
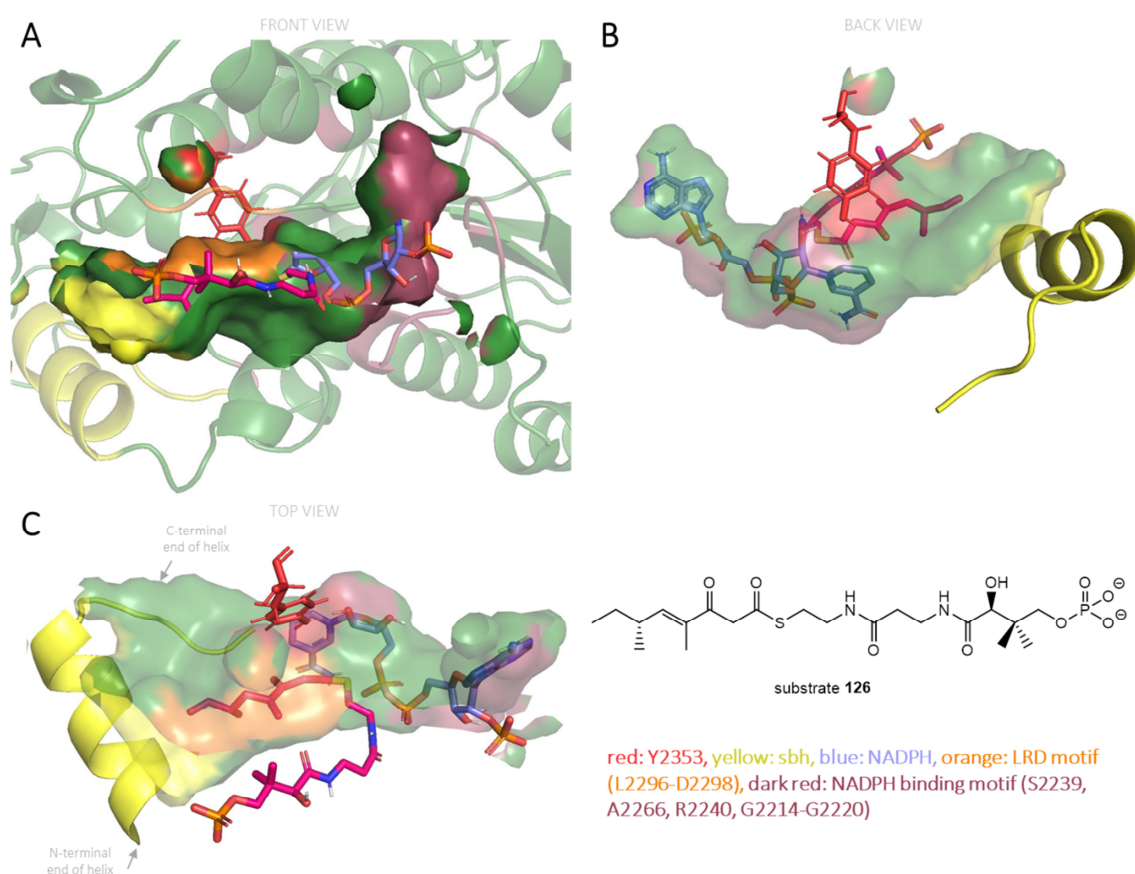
Figure 2.9 Model of the TenS KR active site including tetraketide **126** and NADPH **A**, front view of active site binding pocket; **B**, back view of active site binding pocket, **C**, top view of active site binding pocket.

It was hypothesized that amino acid residues of the substrate binding helix may control the length of the polyketide allowed to enter the active site and bind correctly for reduction. The KS domain appears to be unable to extend β-ketones. Hence, when the chain is no longer a substrate for the KR, then the PKS ceases to extend. For example, incorrect longer substrate chains could collide with the closed helix. Thus, a substrate with an unusual length cannot be placed with the β-carbon at the correct position between the conserved tyrosine and the 4'-*pro-S* proton of NADPH. A change of the amino acid residues of the helix could affect the pocket volume available *e.g.* to fit longer substrates. Another feasible influencing factor is the interaction of longer polyketide carbon chains with the helix residues in the closed state of the active site, which could stabilize the binding of correct substrates. Furthermore, substrate selectivity could be influenced by dynamic factors of the substrate binding helix like structural integrity or flexibility of the substrate binding helix, determined by the included amino acid residues.

2.3.1.3 Comparison between TenS, DmbS, MilS and mFAS Model Structures

The substrate binding helix was analysed by the structure comparison between TenS (producing pentaketides), DmbS (producing hexaketides), MilS (producing heptaketides) and vFAS (producing fatty acids) KR domains. The substrate binding helix has a high structural integrity and a similar position for each hr-PKS (Figure 2.10 A-C), which were modelled with AlphaFold with previously described methods.

The vFAS KR domain was also predicted with the same methods. A multiple alignment identified the sequence corresponding to substrate binding helix in vFAS (section 7.4.3). In contrast to the hr-PKS structures, the vFAS does not form an α-helical structure for the lid helix, but an unstructured loop instead (Figure 2.10 D). The absence of the substrate binding helix correlates with the lack of programming and longer intermediates for vFAS and underlines an important role of the helix on the programming. Fatty acid intermediates of all lengths enter the active site *via* the same opening as NADPH and exit on the other side with variable lengths.
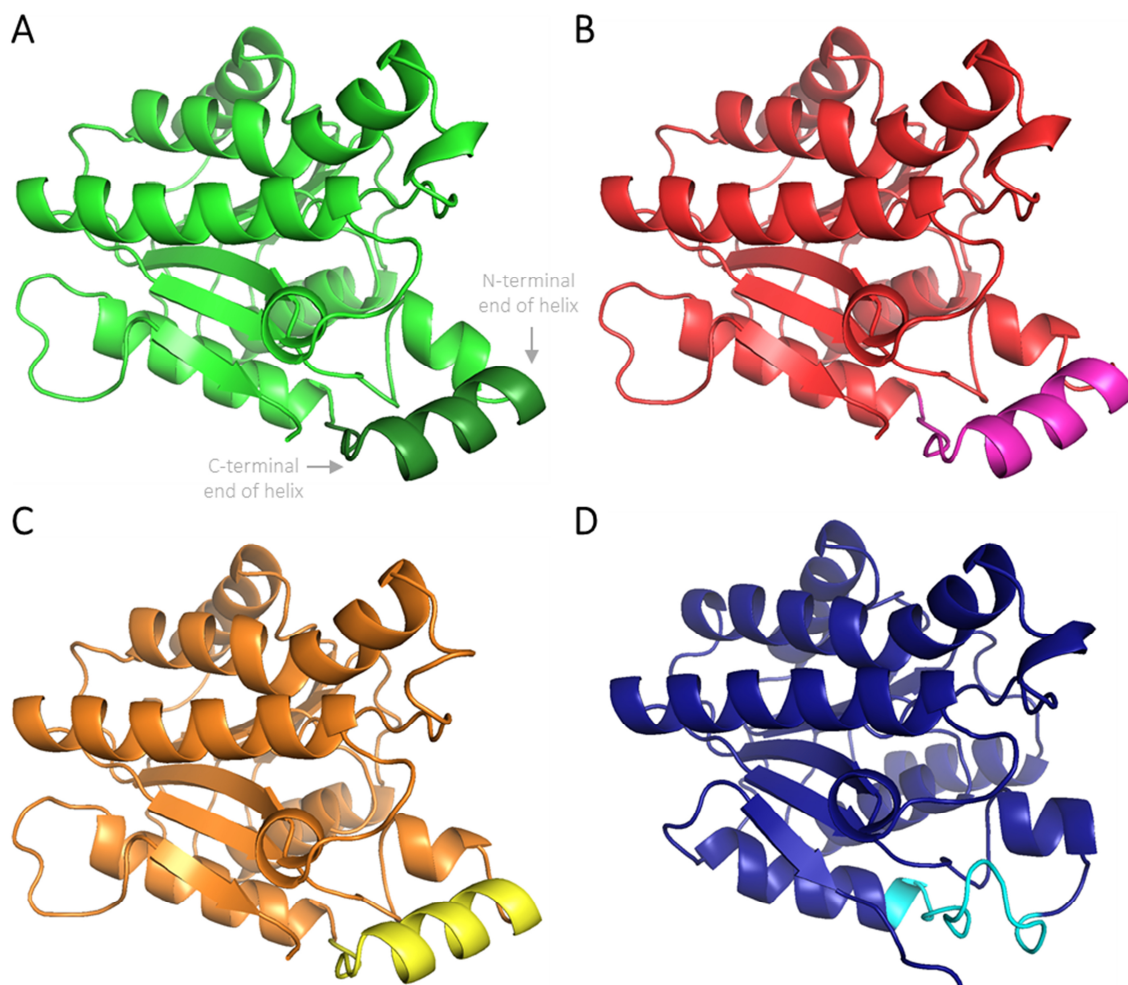
Figure 2.10 Structural protein models based on AlphaFold: **A**, TenS (sbh dark green); **B**, DmbS (sbh magenta); **C**, MilS (sbh yellow); **D**, mFAS (pig, sbh cyan).

When analysing the active site pocket architecture of the four different synthases (TenS, DmbS, MilS, vFAS), the shape of the NADPH binding site does not differ (Figure 2.11 A-D). In contrast, the site of the substrate shows small differences in the overall shape for the hr-PKS. While TenS and DmbS have two potential openings for the substrate binding tunnel directing to the same side (Figure 2.11 A, B), the MilS model contains only one opening for substrates (Figure 2.11 D). Since the tunnel positioned further in the back is significantly narrower than the front tunnel, access to it is probably more difficult for an ACP-bound intermediate. The more accessible front tunnel is more likely to be the real binding position of the substrate. However, no conclusion can be made on the role of the two potential tunnels, because the AlphaFold models most likely represent open forms of the active sites. The open form does not necessarily correspond with the pocket organisation during the reaction, as the helix is hypothesized to clamp over the bound substrate (section 2.3.1). Therefore, the domain structure comparison is

not sufficiently informative to reach a final conclusion on differences between the PKS models. The KR domain with a closed substrate binding helix would presumably be more informative for comparing the different PKS, but could not be generated. A clearer picture on the substrate position will be achieved when co-crystallographic data elucidates the substrate docking in future. Nevertheless, the protein sequence can be used to allow comparisons and for the design of physical experiments.
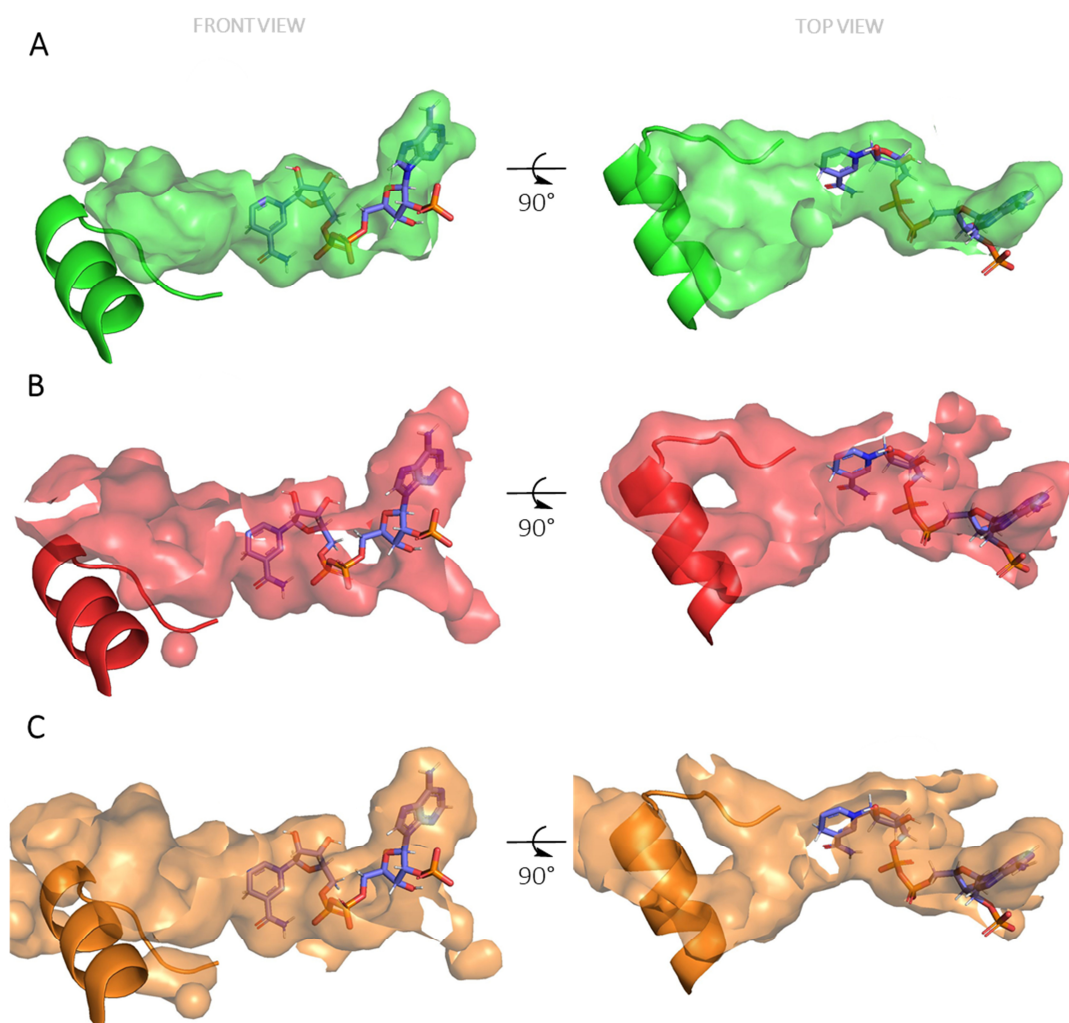
Figure 2.11 Substrate and cofactor binding pocket based on AlphaFold: **A**, TenS; **B**, DmbS; **C**, MilS.

2.3.2 Vector Design and Construction for Hybrid KR Domains

To confirm the important role of the substrate binding helix, two exchange experiments of the complete substrate binding helix to the related DmbS and MilS sequences were designed. Both PKS were used in previous swapping experiments by Cox and co-workers. The DmbS KR sequence was previously used for fragment swaps with a

length of 69 amino acids (approx. 25 % of the KR domain sequence), leading to the formation of hexaketides (section 2.1.2.2, Figure 2.12). The swap design of this work is 15 amino acid long and covers the substrate binding helix. It was hypothesized to lead to similar results as the 69 amino acid swap, producing hexaketides. Furthermore, a 12 amino acid swap (Figure 2.12) experiment between MilS and TenS led to the production of heptaketides in previous work.[58] These findings were hypothesized to be confirmed by analogous MilS substrate binding helix swap of 15 amino acids, covering the same area as the DmbS swap.

2.3.2.1 Selection of TenS Fragment Swap Sequence

The first aim is to swap the substrate-binding helix sequence of TenS between T2395 and V2409 (inclusive, Figure 2.8, yellow) to the corresponding sequence of DmbS (*tenS*(Δsbh:*DmbS*-sbh) and MilS (*tenS*(Δsbh:*MilS*-sbh). The peptide multiple alignment was obtained by previous group members with Geneious (Version 7.1.9). The substrate binding helix shows high sequence variability (Figure 2.12) in the three hr-PKS compared to the overall sequence (see alignment section 7.4.2), which shows the significance of this protein region.[58] In case of the substrate binding helix of DmbS, the swap leads to 11/15 amino acid exchanges and in case of MilS to 12/15 amino acid swaps. The only conserved amino acid in all three PKS is R2405, but no structural role of this amino acid is known for the KR active site.

Figure 2.12 Sequence of substrate binding helix (red = changed amino acids compared to TenS).

2.3.2.2 Recombination Strategy for Chimeric TenS Sequences

The following workflow was used to obtain expression vectors containing the engineered *tenS* gene for heterologous expression in *A. oryzae* NSAR1 for every TenS variant.

The entry vector pEYA-*tenS* (Figure 2.13 A) was provided by previous work from Dr. Oliver Piech and Dr. Sen Yin.[58] The pEYA vector is an *E. coli – S. cerevisiae* shuttle vector, including a *2μ* origin of replication (*ori*) and *ura3* as selection marker (encoding orotidine 5'-phopsphate decarboxylase) for uracil auxotroph *S. cerevisiae*. The *ura3* site enables yeast selection on uracil and uridine free medium. Additionally, pEYA contains genes for replication in *E. coli*: ori (*pUC*) as well as a *kan^R* resistance gene as a selection marker against the antibiotic kanamycin.

To introduce amino acid swaps, pEYA-*tenS* was linearized with *FseI* and *EcoRI* that cut within the *tenS* gene, resulting into two fragments (~ 6 kb and ~ 11 kb, Figure 2.13 A, B). The 6 kb fragment, which contained the sequence of the substrate binding helix, was replaced in the further cloning procedure.

The yeast recombination method requires an identical DNA region of approx. 30 bp at the end of linear fragments. The *tenS* gene was re-constructed with three overlapping DNA fragments (F1 - 3) corresponding to unchanged *tenS* sequence. The fragments (between 1.8 kb and 2.2 kb) were obtained by PCR amplification of the original *tenS* gene of the vector (Figure 2.13 C) with primers A1-A6. The primers were chosen to obtain fragments with a length of approx. 2 kb to ensure an easy and successful PCR amplification. They work as patches to connect the unchanged *tenS* sequence with a fragments containing the hybrid DNA. A 120 bp long synthetic DNA fragment (red fragment in Figure 2.13 A) was used to introduce mutations between F2 and F3.
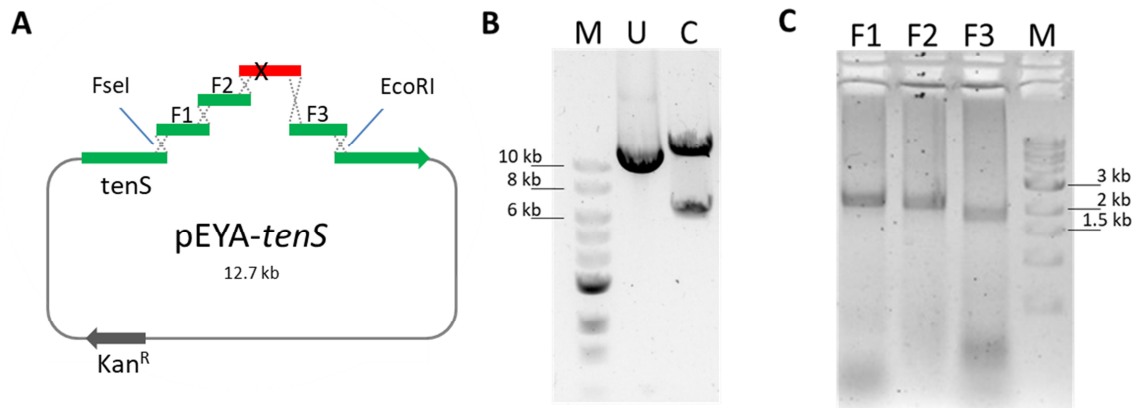
Figure 2.13 Construction of the vector containing *tenS* gene: **A**, scheme of vector construction (green = *tenS* gene, grey = kanamycin ,red = synthetic DNA containing mutations, which are symbolised with x), resistance; **B**, agarose gel of uncut plasmid (U), cut plasmid (C) and marker (M); **C**, agarose gel of PCR fragments 1 - 3 for vector construction and marker (M).

The linearized vector, the PCR fragments and the synthetic DNA were transformed into *S. cerevisiae* with the help of the LiAc/SS carrier DNA/PEG method to execute homologous recombination.[130,131] The constructed vector candidates were purified from yeast and passaged through *E. coli* Top10 selected on kanamycin. Single colonies of *E. coli* were picked, grown over night, and the plasmid was purified.

2.3.2.3 Vector Screening

A strategy to enable fast screening for *E. coli* colonies containing the vector with the desired mutations was developed. The synthetic DNA fragment was designed to contain not only the mutations for amino acid swaps, but also a silent *BstEII* restriction site (Figure 2.14 B). The silent restriction site is located 93 bases upstream of the first mutation site on the synthetic fragment (Figure 2.14 A). The plasmid candidates were digested with *BstEII*, which cuts at two sites for positive vector candidates, resulting in a 3.3 kb DNA fragment (Figure 2.14 A, C). This method allows rapid screening of plasmid candidates.
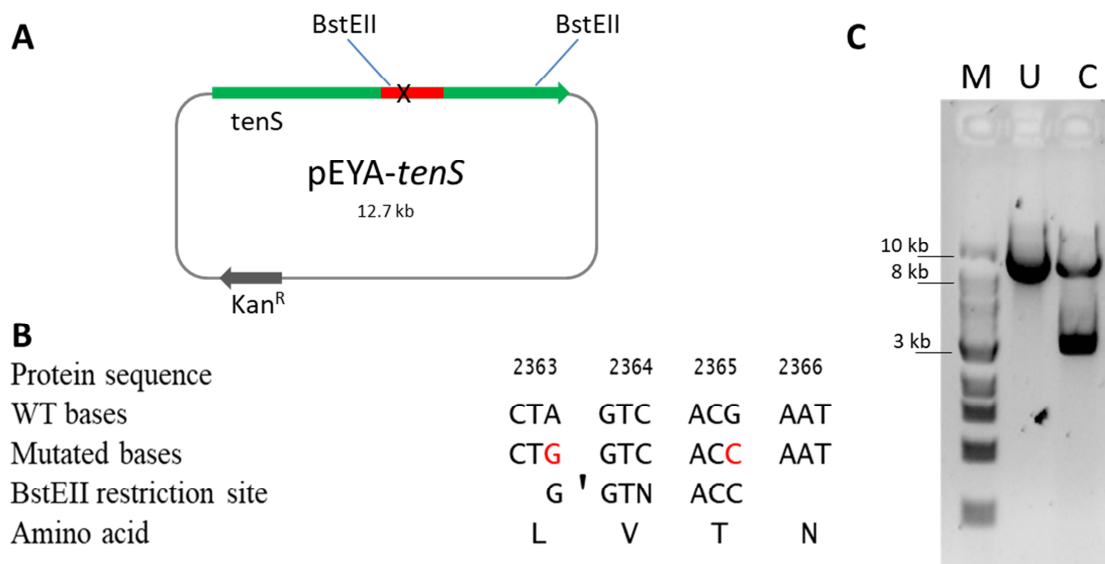
Figure 2.14 Restriction digestion of pEYA-*tenS* with BstEII as a control: **A**, overview of DNA and protein sequence; **B**, scheme of digestion (red fragment = previously built-in synthetic fragment, dark red = *att*-sites for further recombination); **C**, example of agarose gel with uncut plasmid (U) and cut plasmid (C) containing the synthetic fragment and marker (M).

Positive vector candidates were additionally sequenced using an appropriate primer (SQ_KR) to confirm the presence of the desired mutations. If the expected sequence including the desired mutations was observed, the GOI was inserted into the expression vector.

2.3.2.4 Recombination into a Fungal Expression Vector

In vitro LR ("Gateway")-recombination[132] was performed between pEYA-*tenS* and the fungal expression vector (pTY-GS-*argB-tenC*, section 1.3.1), which contains the *tenC* gene encoding the *trans*-acting ER domain, provided by previous work. The Invitrogen LR recombinase kit was used, which contains a mixture of bacteriophage λ integrase proteins (integrase, integration host factor, excisionase).

The entry vector pEYA-*tenS* and the destination vector contain corresponding *att*-sites (Figure 2.15 A, *att*L1/*att*L2 and *att*R1/*att*R2), which enable the LR-recombinase enzymes to interchange the DNA fragment between them. The *att*-sites flank a Gateway cassette in the fungal expression vector, which includes a chloramphenicol resistance gene and a *ccdB* gene, coding for a CcdB killer protein. The CcdB killer protein leads to the death of non-competent *E. coli* and is required for the selection strategy of the recombination method (Figure 2.15 A).
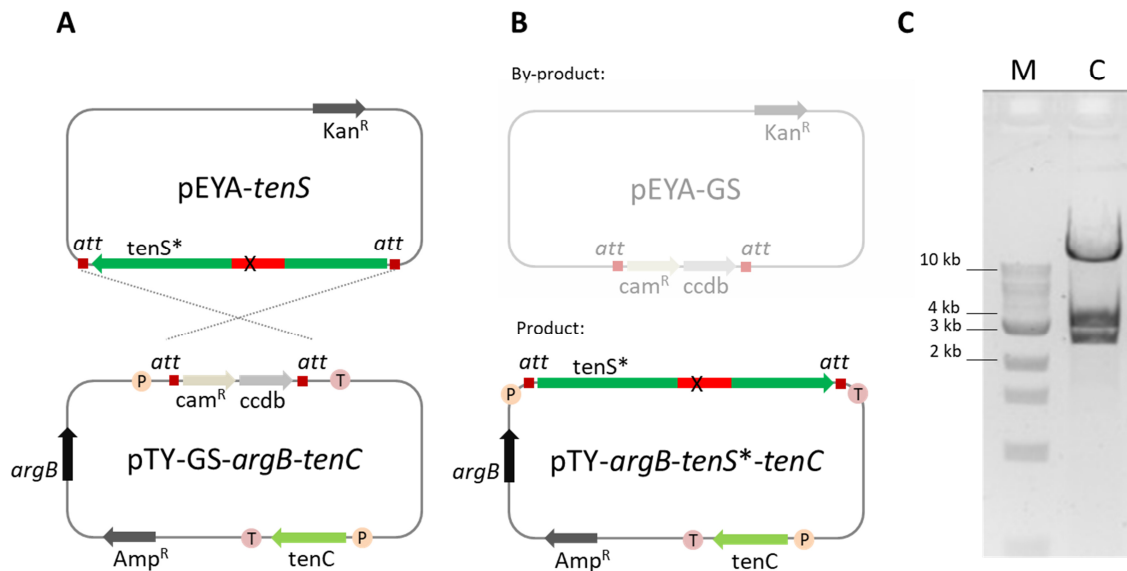
Figure 2.15 LR recombination of pEYA-tenS and pTY-GS-*argB-tenC*: **A**, scheme of pEYA-*tenS* and pTY-GS-*argB-tenC* (*AmpR* = carbenicillin resistance, *argB* = selection marker for fungal transformation experiments, grey = gene for ccdB toxin, P = promoter, T = terminator; **B**, scheme of the resulting vectors with final expression vector and by-product, GC: Gateway casette; **C**, agarose gel of control of the final expressionvector.

Four different vectors are present after *in vitro* recombination: the two original plasmids (Figure 2.15 A), the target vector (pTY-*argB-tenS-tenC*) and the pEYA vector after the recombination, which includes the *ccdB* gene (Figure 2.15 B). The LR-recombinase vector mix was transformed into *ccdB*-incompetent *E. coli* Top 10 cells. The expected vector pTY-*argB-tenS-tenC* (Figure 2.15 B) is selected on agar containing the appropriate antibiotic and due to the *ccdB* gene affected toxicity. An additional control of the putative expression vector took place with restriction digestion by *BstEII*. The expected vector contains three *BstEII* restriction sites, resulting in three fragments (Figure 2.15 C, 2.5 kb, 3.4 kb and 18 kb).

The final expression vector contains the genes encoding the hr-PKS TenS between the starch inducible *amyB*-promoter (P_{amyB}) and terminator (T_{amyB}) and the *trans*-acting ER TenC from the tenellin pathway, located between *adh*-promoter (P_{adh}) and *eno*-terminator (T_{eno}). The expected product of the wild type (WT) system is pretenellin A **39**.

2.3.3 Analysis of TenS WT and TenS Variant Extracts

2.3.3.1 Transformation into the Fungal Expression Host

The fungal heterologous host *A. oryzae* NSAR1 was grown on DPY plates for approx. 7 days. A 50 ml liquid culture with GN media was inoculated with the mycelia and grown overnight to obtain fresh fungal biomaterial for the transformation. The fungus was incubated with *Trichoderma* lysing enzyme/VinoTaste® the next day for 1 - 3 h to remove the cell wall and form protoplasts. After PEG-mediated transformation (section 2.3.2), the protoplasts were plated on minimal media agar lacking arginine. After approx. 5 days, single transformants were picked onto fresh selection plates. The selection process was repeated after three to five days. After the selection process, the transformants were transferred to DPY plates to grow for approx. 5 days at 28 °C. Then, 100 ml CMP liquid cultures were inoculated with mycelia of each transformant. CMP contains maltose that induces the P_{amyB} promotor (the P_{adh} promotor is constitutive) leading to the gene expression of *tenS* and *tenC*. Transformants were cultivated for 7 days at 30 °C at 180 rpm.

The integration of the *tenS* gene into the *A. oryzae* NSAR1 genome of each transformant was monitored by extraction gDNA and PCR amplification (Figure 2.16 A) with the appropriate primers (KR_amp_fw/KR_amp_rev). The resulting 546 bp fragment (Figure 2.16 B) was sequenced.
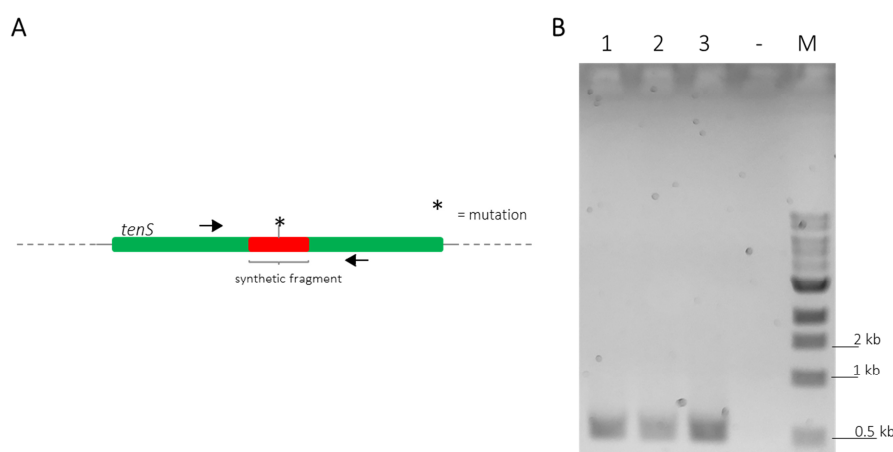
Figure 2.16 Analysis of gDNA by PCR: **A**, amplification of a 0.5 kb fragment including the KR substrate binding helix with mutations; **B**, agarose gel run (1-3 = different transformants as examples, - = negative control, M = marker).

2.3.3.2 Extract Analysis of TenS Wild Type Protein

The unchanged *tenS* gene was used for transformation as a positive control. The CMP liquid cultures were extracted twice with equal amounts of ethyl acetate (EtOAc) after prior acidification with HCl to a pH of 2. The solvent was dried with $MgSO_4$ and then removed under reduced pressure. Extracts were dissolved in 1 ml methanol. The extracts were then used for LCMS analysis.

The production of pretenellin A **39** as the main compound was observed by LCMS for the positive control as expected (Figure 2.17 A). Pretenellin A **39** is characterised by a peak at 8.7 min in the DAD chromatogram. The UV absorbance of pretenellin A has characteristic maxima at 225 nm and 358 nm (Figure 2.17 B). The total ion count mass spectra shows the expected fragments of 354 *m/z* in ES⁻ ([M - H]⁻) and 356 *m/z* ([M + H]⁺) in ES⁺ mode, according to the molecular weight of 355 Da of pretenellin A **39** (Figure 2.17 B).
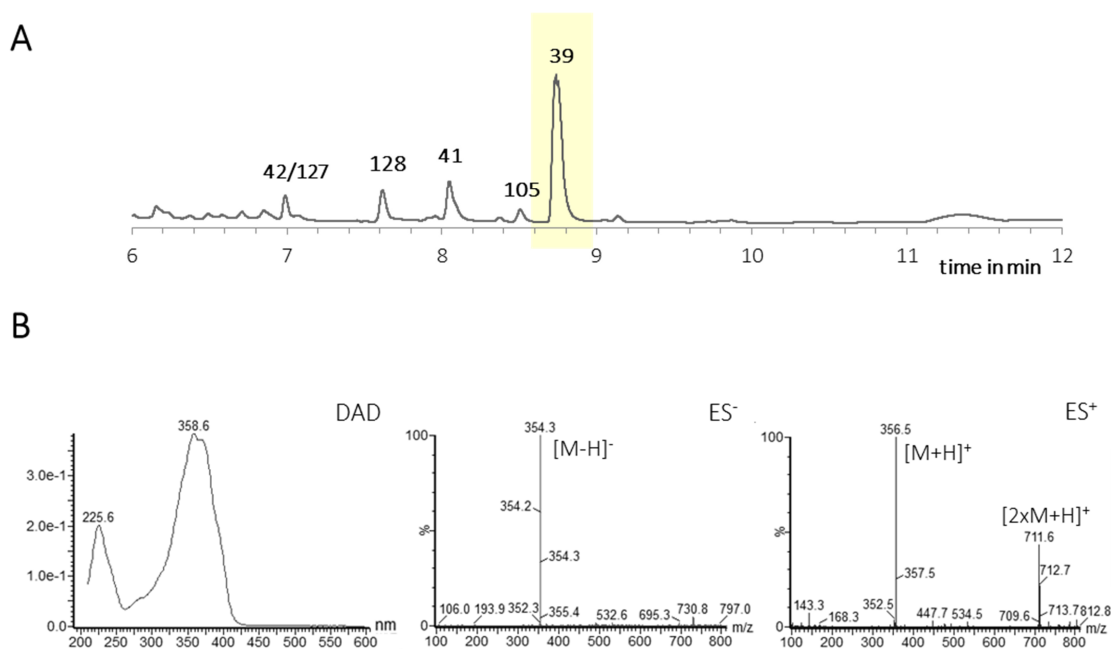
Figure 2.17 LCMS Analysis of extract from positive control containing wild type pTY-*argB-tenS-tenC*: **A**, DAD BPI chromatogram (210 - 600 nm); **B**, characteristic spectra of the peak at 8.7 min.

Additionally to the main pretenellin A peak, minor related pentaketide or tetraketide compounds were observed at 7.0 min, 7.6 min, 8.1 min, 8.6 min and 9.1 min (Figure 2.17 A). They were assigned by mass and UV spectra to the known compounds **41**, **42**,

127, **104** and **128** by comparison to previously obtained characteristic data from the Cox group (Figure 2.18).[58]

anhydro pretenellin A **128**

prototenellin B **42**

prototenellin B isomer **127**

= R

prototenellin A **41**

pretenellin A isomer **104**
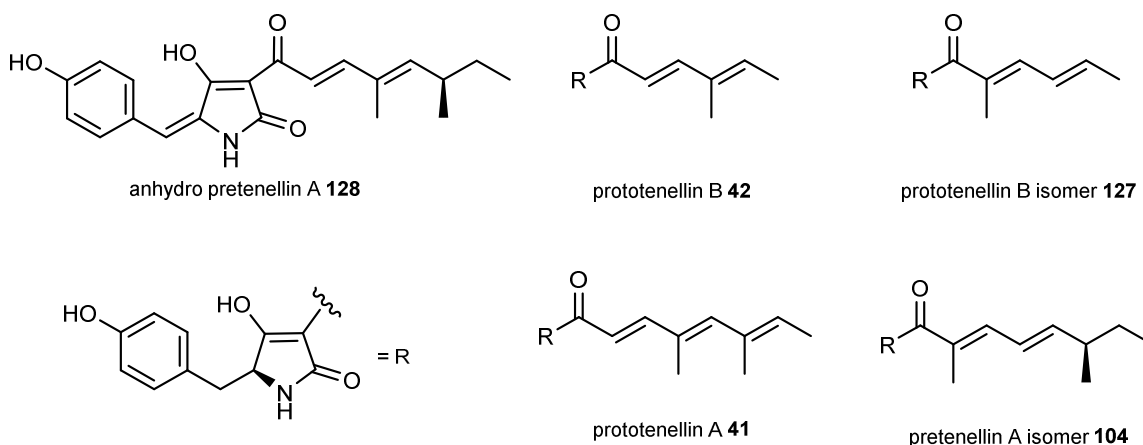
Figure 2.18 Minor compounds in TenS (WT) extract.

2.3.3.3 Substrate Binding Helix (T2395 to V2409) Swap with DmbS

The transformation, selection and cultivation process for hybrid TenS variants was carried out as described previously. Seven transformants containing *tenS*(Δsbh:*DmbS*-sbh) were obtained and all of these produced pretenellin A **39** (Figure 2.19 A), corresponding to the wild type TenS (Figure 2.17). Additionally, two related new peaks were observed at 6.6 min and 9.4 min (Figure 2.19 A) in each transformant.

The MS fragmentation fragmentation of the molecule **46** in the ES⁻ and ES⁺ mode suggests a mass of 381 Da at 9.4 min (Figure 2.19 C), with the *m/z* 382 ([M + H]⁺) and 380 ([M - H]⁻). Additionally, UV spectra with maxima at 224 nm and 398 nm identified the compound as the doubly methylated hexaketide prebassianin **46** by comparing the data to previous obtained spectra from our group. The compound eluting at 6.4 min was identified similarly by the characteristic data (Figure 2.19 B) as the known hydroxylated shunt compound hydroxy prebassianin **105**, most likely derived from oxygenation by *A. oryzae*.[58]
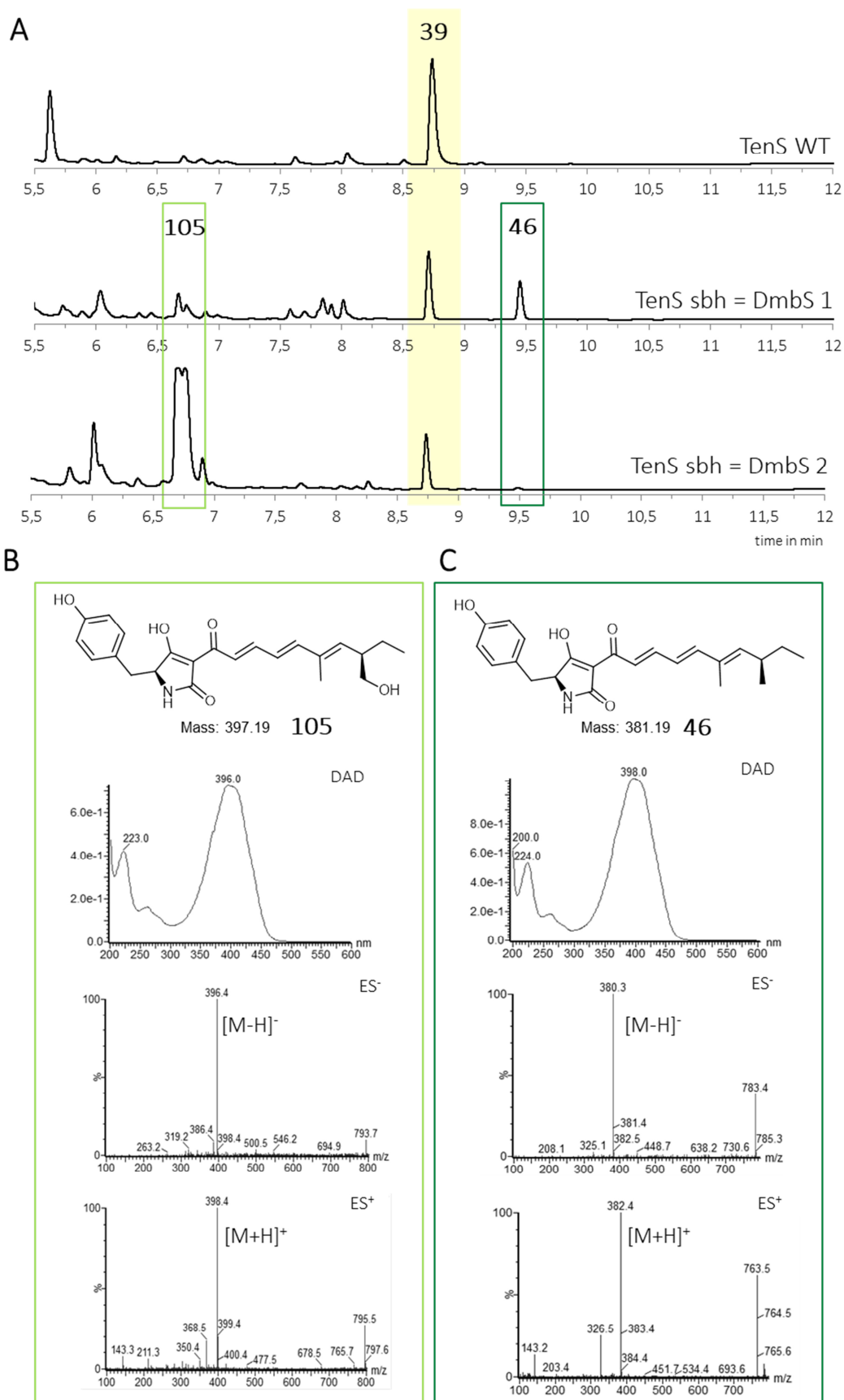
Figure 2.19 LCMS Analysis of extract from pTY-*argB-tenS-tenC* (*tenS*(Δsbh:*DmbS*-sbh): **A**, DAD (210 – 600 nm) BPI chromatogram of WT TenS and two examples of chimeric TenS (Δsbh:*DmbS*-sbh); **B**, characteristic spectra of peak **105** at 6.4 min; **C**, characteristic spectra of peak **46** at 9.4 min.

The ratio between prebassianin **46** and hydroxy prebassianin **105** differs in the different transformants (Figure 2.19 A, example 1 and 2). The ratio between pentaketide **39** and hexaketides (**46** and **105** summarized) was roughly determined in the ELSD chromatogram by peak integration. Considering all producing transformants, peak integration shows that hexaketides are at least 50 % of the total (Figure 2.19 A, example 1), and up to 97 % hydroxy prebassianin **105** in the highest case (Figure 2.19 A, example 2). On average approx. 12 % pentaketide pretenellin A **39** is produced in comparison to approx. 88 % hexaketides.

All producing transformants were scanned for known related polyketide molecules with differing chain length and methylation patterns (Table 2-3), found in previous work (section 2.1).[58] Therefore, a scan based on the molecular masses and UV maxima of known related compounds was carried out. Traces of different pentaketides were found, including prototenellin A **41** and the pretenellin A isomer **104** as well as the described hexaketide structures **46** and **105**. No evidence for heptaketide structures was observed.

Table 2-3 Scan of known compounds in transformants with sbh = DmbS extracts (red = main product, "-" = not detected).

	Structure	WT	Transformant						
			1	2	3	4	5	6	7
Pentaketides	prototenellin A **41**	✓	✓	✓	✓	-	✓	✓	✓
	desmethylpretenellin A **45**	-	-	-	-	-	-	-	-
	pretenellin A isomer **104**	✓	✓	✓	✓	✓	✓	✓	✓
	pretenellin A **39**	✓	✓	✓	✓	✓	✓	✓	✓
Hexaketides	prebassianin A **46**	-	✓	✓	✓	✓	✓	✓	✓
	hydroxy prebassianin **105**	-	✓	✓	✓	✓	✓	✓	✓
Heptaketides	prefarinosone **106**	-	-	-	-	-	-	-	-
	militarinone C **34**	-	-	-	-	-	-	-	-

In summary, the chain length was reprogrammed by the swap of the substrate binding helix with preservation of the methylation pattern of **39** for the main compounds.

2.3.3.4 Substrate Binding Helix Swap (T2395 to V2409) with MilS

The same sequence area (T2395 to V2409, 15 residues) was swapped to the corresponding MilS sequence (*tenS*(Δsbh:*MilS*-sbh)). According to previous results, a swap of 12 residues led to the production of heptaketides when swapping Q2398 - V2409. Therefore, a similar result was expected.

The four mutants containing the substrate binding helix of MilS produced mostly the pentaketide pretenellin A **39** (like the wild type TenS, Figure 2.17) and the hexaketide prebassianin **46**, corresponding to the complete substrate helix swap with DmbS (Figure 2.19). The hydroxylated hexaketide **105** was only observed in traces. Additionally, a minor peak **129** was detected at 10.0 min (Figure 2.20).
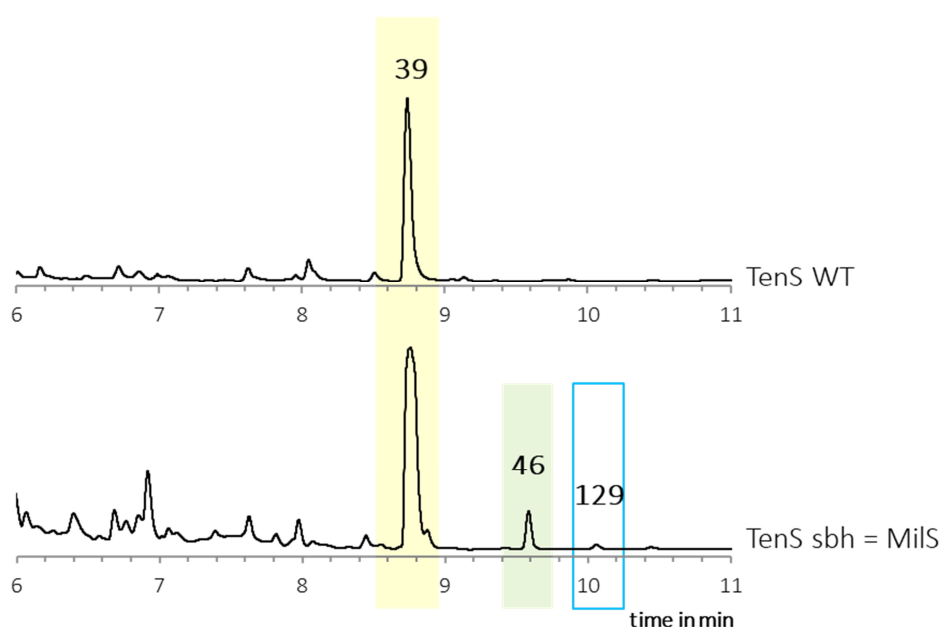
Figure 2.20 DAD (210 – 600 nm) chromatograms of extract from pTY-*argB-tenS-tenC (tenS(Δ*sbh:*MilS*-sbh) and TenS WT.

The peak of compound **129** is characterized by *m/z* values of 394 ([M - H]⁻) and 396 ([M + H]⁺), indicating a mass of 395 Da (Figure 2.21). The compound peak was detectable in two transformants in minor amounts (Table 2-8). The chromatogram shown (Figure 2.20) is the transformant with the most intense peaks achieved for prebassianin **46** and **129**. The UV maximum of 392 nm (Figure 2.21) indicated a trieneone, when comparing to previously obtained data. This underlines the hypothesis that the peak corresponds to compound **129** including a triple-methylated hexaketide

chain with a mass of 395 Da (Figure 2.21). Due to the very low titre of the compound (Figure 2.20) it was not possible to isolate **129** for NMR.
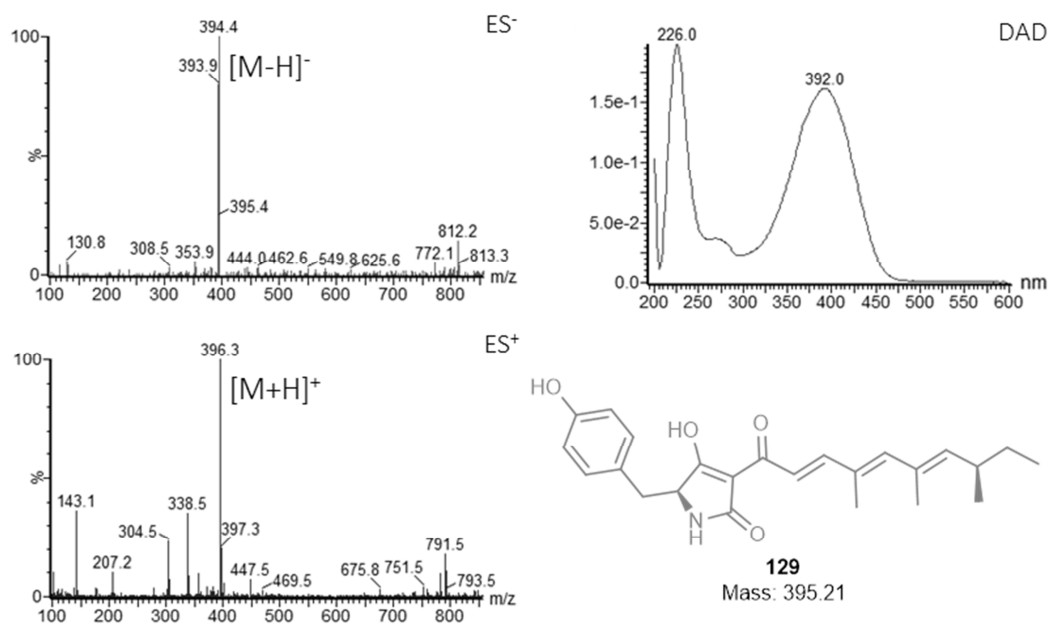
Figure 2.21 Characteristic spectra from extract from pTY-*argB-tenS-tenC (tenS(Δsbh:MilS-sbh)*: **A**, spectra of peak at 10.0 min; **B**, spectra of peak at 10.4 min.

No additional new compounds were found (Table 2-8).

Table 2-4 Scan of known compounds in transformants with sbh = MilS extracts (red = main product, - = not detected).

	Structure	WT	Transformant			
			1	2	3	4
Pentaketides	prototenellin A **41**	✓	✓	-	-	-
	desmethylpretenellin A **45**	-	-	-	-	-
	pretenellin A isomer **104**	✓	✓	-	-	-
	pretenellin A **39**	✓	✓	✓	✓	✓
Hexaketides	prebassianin A **46**	-	✓	✓	-	✓
	hydroxy prebassianin **105**	-	✓	✓	✓	✓
	compound **129**	-	✓	-	-	✓
Heptaketides	prefarinosone **106**	-	-	-	-	-
	militarinone C **34**	-	-	-	-	-

The presence of **46**, **105** and **129** are indications that TenS was reprogrammed to produce hexaketides, but no heptaketides were detectable. The swap to the MilS sequence is less effective in terms of changes in the chain length than the swap to the DmbS sequence, based on the achieved titres of unnatural products. However, **129** indicates an effect on the methylation programming by the T2395 to V2409 swap.

2.3.4 Rationally Selected Amino Acid Swap Experiments

2.3.4.1 Strategy for Rational Selection of Amino Acid Swaps

The sequence alignment of TenS, DmbS and MilS (Figure 2.22) was linked to the structural analysis of the substrate binding helix (Figure 2.23) for the selection of rational amino acid swaps. In the sequence from position 2395 to 2409 in TenS, some amino acids are conserved for two or three hr-PKS (K2396, V2397, Q2398, T2403, R2405). Conserved amino acid positions were excluded for the selection of amino acid swap positions. However, a high occurrence of amino acid changes (Figure 2.22, red) is found for the other positions of the substrate binding helix (complete KR alignment see section 7.4.3).
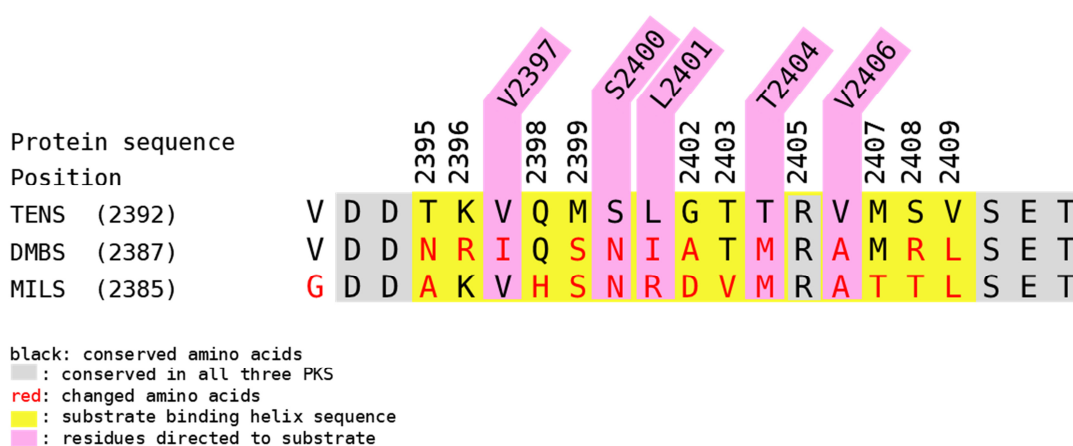
Figure 2.22 Protein sequence of the substrate binding site of the KR of TenS, DmbS and MilS (C: contact to cofactor, S: contact to substrate according to KR model, red: changed amino acids compared to TenS).

With the AlphaFold model, the side chains of amino acids V2397, S2400, L2401, T2404 and V2406 are directed towards the active site of the KR active site pocket (Figure 2.22, Figure 2.23, violet residues), which could indicate possible role for substrate selectivity. Other amino acids of the sequence point mainly to the bottom side

of the pocket (Figure 2.24 D, F, H, yellow) or to the opposite side (Figure 2.24 E-H, yellow) and therefore are not in focus for amino acid swaps, although these positions of the helix are also changed between the different hr-PKS (Figure 2.22).
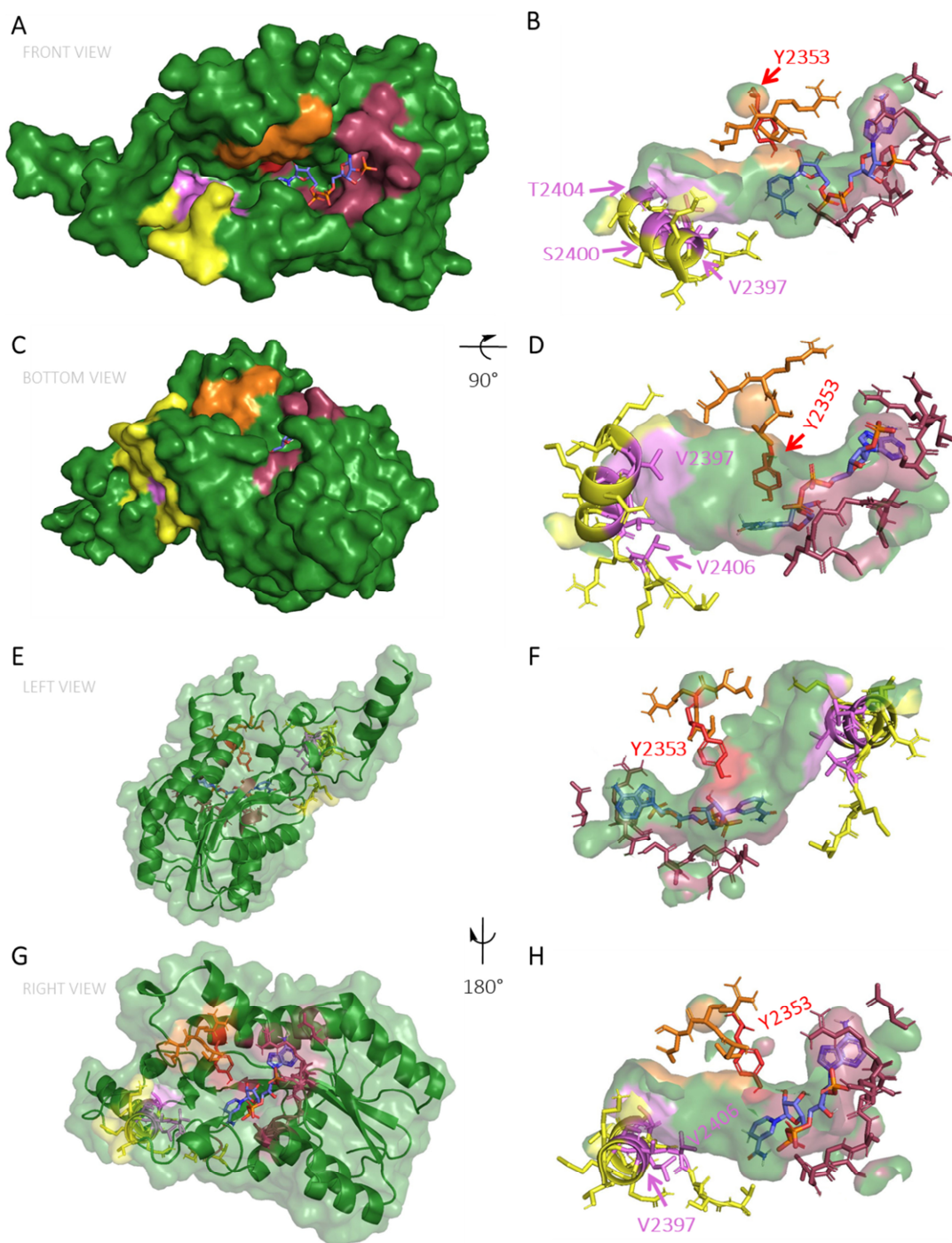
Figure 2.23 Detailed visualisation of the KR active site: **A**, front view of active site with protein surface; **B**, front view of the active site with surface of the active site pocket; **C,** bottom view of the active site with protein surface; **D**, bottom view of the active site with surface of the active site pocket; **E**, left side view of KR domain; **F**, left side view of KR domain active site; **G**, right side view of KR domain; **H**, right side view of KR domain active site.

72

Positions V2397, S2400, L2401, T2404 and V2406 were further evaluated based on sequence alignment to select the appropriate mutations. Residue V2397 is in putative close contact to the substrate. However, position V2397 will not be included in the rational engineering approach, as in two PKS (TenS and MilS) the amino acid is conserved and for DmbS a conservative replacement by isoleucine is present at this position (Table 2-5).

Residue L2401 is in both DmbS and MilS swapped to different amino acids. Leucine (TenS) and isoleucine (DmbS) are structurally similar, since both are aliphatic C_4-residues. Thus, the non-conservative arginine residue mutation was chosen for swapping experiments (Table 2-5). The residues at positions S2400N, T2404M and V2406A are the same in the DmbS and MilS sequence (Table 2-5). Hence, the selection of the rational engineering of the PKS includes the four mutations S2400N, L2401R, T2404M and V2406A.

Table 2-5 Selected amino acid residues for mutation.

Position	Residue in			Selected residue for mutation:
	TenS	DmbS	MilS	
V2397	V	I	V	no mutation
S2400	S	N	N	N
L2401	L	I	R	R
T2404	T	M	M	M
V2406	V	A	A	A

2.3.4.2 LCMS Analysis of S2400N, L2401R, T2404M and V2406A 4-Residue Swap

Five producing transformants containing 4-residue change were analysed as described previously. All five of the transformants showed a clear peak for pretenellin A **39** as the main product (Figure 2.24). Additionally, a minor amount of prebassianin A **46** was identified (Figure 2.24). Prebassianin A **46** was detected in four producing transformants in extracted ion chromatograms (Table 2-6).
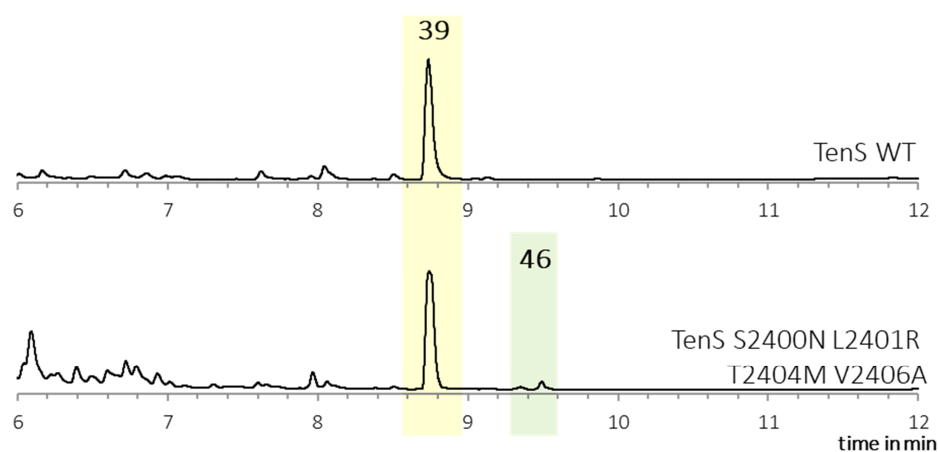
Figure 2.24 LCMS Analysis of extract from pTY-*argB-tenS*-tenC* S2400N, L2401R, T2404M and V2406A.

No additional new compounds were found (Table 2-6).

Table 2-6 Scan of compounds in TenS transformants with mutations S2400N, L2401R, T2404M, V2406A (red = main produc, - = not detected).

	Structure	**WT**	**Transformant**				
			1	2	3	4	5
Pentaketides	prototenellin A **41**	✓	-	-	-	✓	✓
	desmethylpretenellin A **45**	-	-	-	-	-	-
	pretenellin A isomer **104**	✓	✓	✓	✓	✓	✓
	pretenellin A **39**	✓	✓	✓	✓	✓	✓
Hexaketides	prebassianin A **46**	-	✓	✓	-	✓	✓
	hydroxyprebassianin **105**	-	✓	-	-	✓	✓
	compound **129**	-	-	-	-	-	-
Heptaketides	prefarinosone **106**	-	-	-	-	-	-
	militarinone C **34**	-	-	-	-	-	-

The swap of four amino acid positions affected the production of hexaketide **46** and **105**. However, the intensity of the hexaketide peak is decreased in comparison to the swap experiments with the complete substrate binding helix (section 2.3.3.3 and 2.3.3.4).

2.3.5 Comparison of Hybrid TenS Variants Protein Structures

The wildtype TenS KR protein structure was compared to hybrid TenS KR structures in which the performed mutations were introduced by changing the amino acid sequence prior to AlphaFold prediction to analyse possible changes in substrate binding helix and active site pocket architecture due to the swap experiments. An introduction of the mutations into the structural models had no influence on the overall shape and a minimal influence on the position of the substrate binding helix in the predicted structural models (Figure 2.25 A-D).
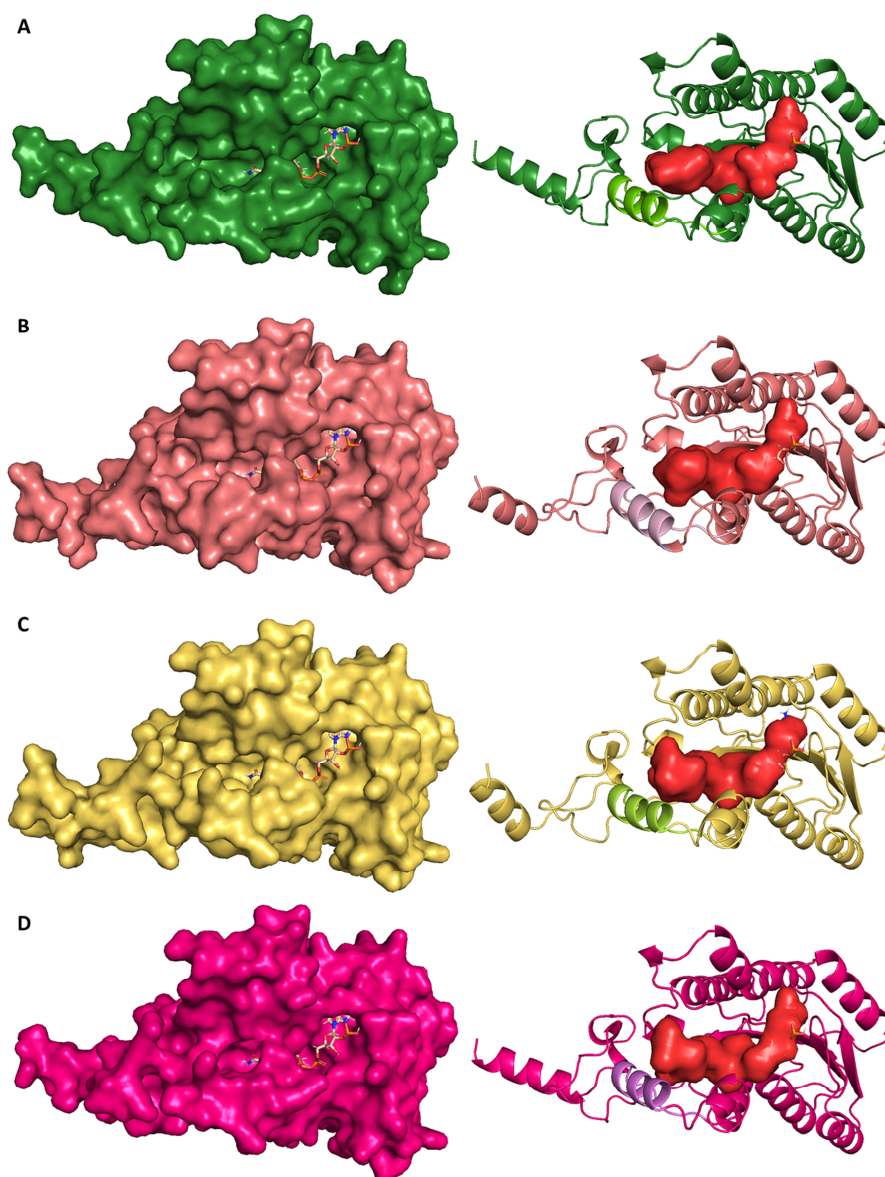
Figure 2.25 Comparison of calculated pocket volume by PyVOL: **A**, TenS WT (sbh light green); **B**, TenS Δsbh:DmbS-sbh (sbh light pink); **C**, TenS Δsbh:MilS-sbh (sbh green); **D**, TenS S2400N, L2401R, T2404M and V2406A (sbh violet).

The characteristics of the active site pocket were analysed for the wild type TenS and each mutant using the PyVOL plugin for PyMOL (Figure 2.25, red).[133] This bioinformatic tool allows the calculation of the volume of the substrate and cofactor binding pocket (used parameters see section 6.3.2). The values of the pocket volume differ with no correlation to the efficiency of the changes in polyketide chain length (Table 2-7) and the visible differences of the calculated pocket are minimal (Figure 2.25). However, as previously described (section 2.3.1) the AlphaFold models most likely represent the open state of the active site pocket. No conclusion is possible for the closed conformation of the active site pocket that is most likely involved in intrinsic substrate selectivity.

Table 2-7 Comparison of calculated pocket volume by PyVOL.

TenS model	pocket volume [Å^3]
TenS WT	1269
TenS Δsbh:*DmbS*-sbh	1114
TenS Δsbh:*MilS*-sbh	1114
TenS S2400N, L2401R, T2404M, V2406A	1205

Overall, no difference in any structural criteria was observed for all hybrid TenS models. Thus, another experimental approach was developed to focus on each amino acid position of the substrate binding helix individually (section 2.3.6).

2.3.6 Alanine Scan of the Substrate Binding Helix

An alanine scan was performed to find single amino acids of the substrate binding helix, which play a crucial role in the KR domain programming. Therefore, each amino acid was exchanged with the amino acid alanine. Alanine scans are a common method to investigate the function of a certain amino acid position, as alanine is an aliphatic amino acid, which includes no charge to enable molecular interactions, leading to comparatively low functionalisation of the residue. By swapping a complex amino acid residue to a simple methyl group, the impact of the certain amino acid on the protein function can be determined.[134]

2.3.6.1 Vector Construction

Overall 15 amino acids of the KR domain from TenS were exchanged to alanine (between D2394 to S2410) in individual experiments, which contain the substrate binding helix by previously described methods (section 2.3.2). In each synthetic fragment, one original codon was exchanged for an alanine codon (GCN) at a different position (Table 6-6). The mutated pEYA-*tenS* vector was obtained through yeast recombination, verified *via* sequence analysis and recombined with the fungal expression vector.

In the sequence of TenS, none of the 15 positions includes an alanine residue. In comparison, the substrate binding helix contains two alanine residues in DmbS and two in MilS at different positions (at position G2402A and V2406A for DmbS and at positions T2395A and V2406A for MilS). The results of the swaps of position S2400A, L2401A, T2404A and V2406A are especially of interest, as they were hypothesized to be important, based on the KR model in the prior analysis (Figure 2.23).

2.3.6.2 LCMS Analysis of Alanine Scan Experiments

All transformants from all 15 alanine scan mutants produced pretenellin A **39** as the main product. However, traces of additional related compounds were detected by searching for extracted ion chromatograms for known compounds (Table 2-8). For example, a run with the TenS variant V2406A (Figure 2.26 E) included traces of the hexaketide structures prebassianin **46** (Figure 2.26 D) and hydroxyl prebassianin **105** (Figure 2.26 A) in addition to pretenellin A as the main product (Figure 2.26 C) and the pentaketide **41** (Figure 2.26 B, C).
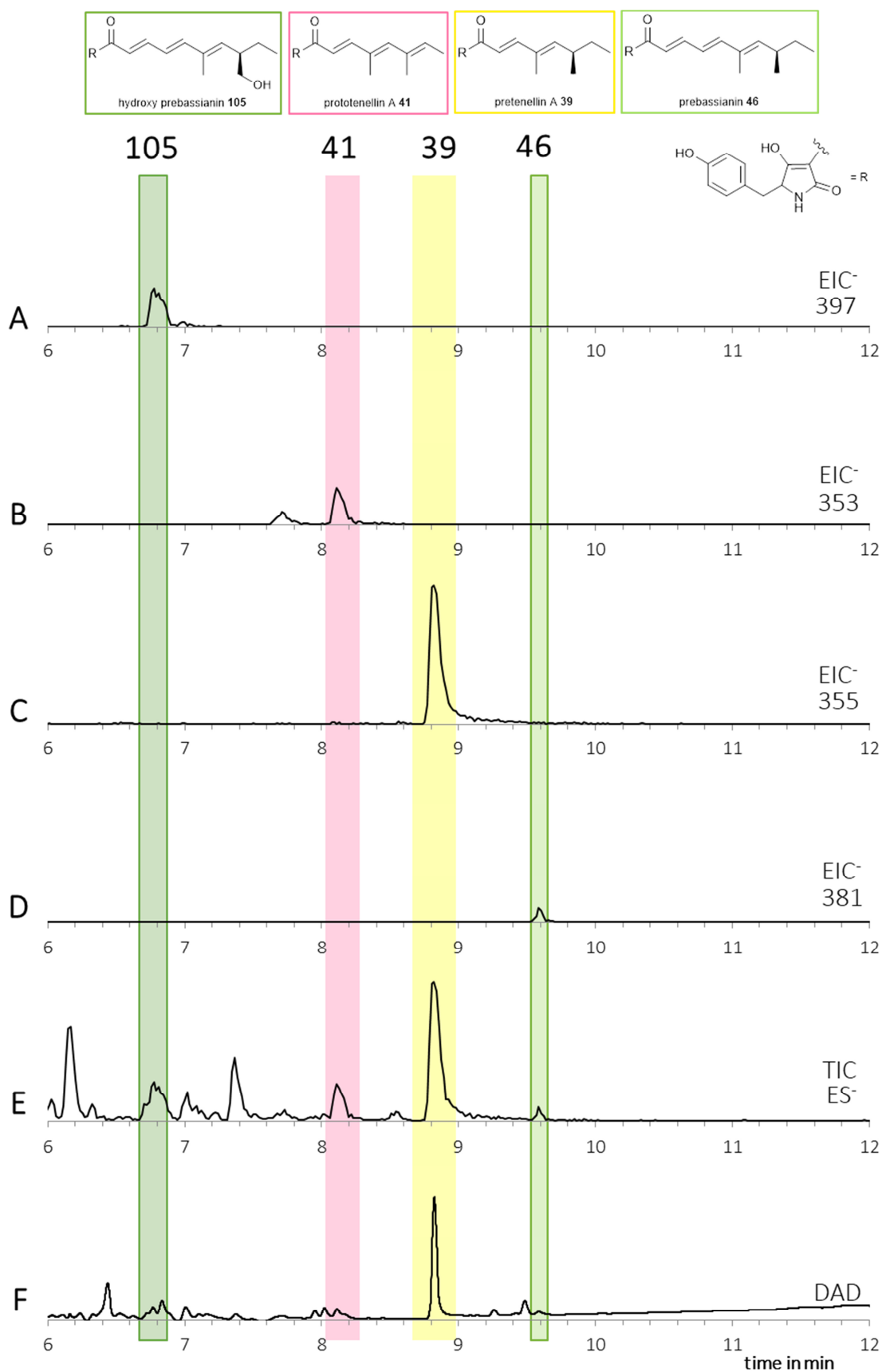
Figure 2.26 Example of TenS variant V2406A chromatogram with: extracted ion chromatograms of *m/z* **A**, 396; **B**, 352; **C**, 354; **D**, 381; **E**, ES⁻ chromatogram.

Hexaketide compounds were found in extracts of transformants including the following mutations (Table 2-8): G2402A (in 5 from 9 extracts), T2404A (in 1 of 7 extracts), V2406A (in 3 from 3 extracts), M2407A (in 1 from 5 extracts).

Table 2-8 Summary of extracted compounds from alanine scan (red = positive for hexaketides).

Transformant	Detected compounds						
	Pentaketides				Hexaketides		
	41 353 Da	**45** 341 Da	**104** 355 Da	**39** 355 Da	**46** 381 Da	**105** 397 Da	**129** 395 Da
WT	++	-	++	++	-	-	-
T2395A	++	+/-	++	++	-	-	-
K2396A	+	+	++	++	-	-	-
V2397A	+/-	-	-	++	-	-	-
Q2398A	++	+/-	+/-	++	-	-	-
M2399A*	++	-	++	++	-	-	-
S2400A	++	+	++	++	-	-	-
L2401A	+/-	+	-	++	-	-	-
G2402A	+	-	++	++	+	-	-
T2403A	++	++	++	++	-	-	-
T2404A	+	-	++	++	-	+/-	-
R2405A	+	++	++	++	-	-	-
V2406A	++	-	++	++	+/-	++	-
M2407A	+/-	+/-	++	++	+/-	-	+/-
S2408A	++	+	++	++	-	-	-
V2409A*	++	++	++	++	-	-	-

* = fewer than 3 transformants analysed   ++ = 100 %, + = 50-99 %, +/- = 1-50 %, - = 0 % of transformants

2.4 Discussion of Rational Engineering of TenS

2.4.1 Swap of Complete Substrate Binding Helix T2395 to V2409

The swap of the complete substrate binding helix to the DmbS and MilS sequences led in both cases to the production of hexaketide compounds **46** and **105**. The main difference between the experiments is the ratio between the produced penta- and hexaketides.

The highest production of hexaketides in all mutation experiments was found for the substrate-binding helix swap with DmbS when taking into consideration that hydroxyl prebassianin **105** derived from adventitious oxygenation of prebassianin **46**

from the host *A. oryzae* is also a hexaketide. The production of hexaketides was possible in a comparatively high amount (up to 97 % of the total). Cox and co-workers reported similar findings when swapping sequence fragments of DmbS of a length of 138 amino acids and 69 amino acids (Figure 2.3 F, G), both including the substrate binding helix.[58] It was possible to achieve similar results with the swap experiment of this work with only 15 amino acid positions from T2395 to V2409. The substrate binding helix is therefore a key region for the control of the substrate binding helix.

For the T2395 to V2406 swap to the MilS sequence, the main product of the transformants remained the pentaketide pretenellin A. However, hexaketides **46**, **105** and **129** were also detected. In a previously reported swap of positions Q2398 - V2409 to MilS amino acids, heptaketides were observed in low titres (but as the dominant overall-product).[58] Only one additional residue was exchanged in case of the experiment of this work. Position K2396 and V2397 are conserved in both KR domains. Thus, the only non-conservative mutation included in position 2305 to 2307 is the mutation T2395A. Position T2395 is therefore an interesting focus for future work, although mutation of this position to alanine did not lead to hexaketides during the alanine scan. Furthermore, the 15-residue MilS swap did show pretenellin A **39** as the main product in opposite to the DmbS swap of the complete substrate binding helix, which could indicate a higher sensitivity of the MilS sequence to extrinsic programming factors. The interaction of the substrate binding helix sequence with other adjacent residues to the KR active site in the closed state could affect the KR domain architecture.

The influencing factors on the intrinsic selectivity are unknown and are probably an interplay of the direct contact to the substrate through amino acid residues, the structural properties of the helix and dynamic factors such as the flexibility and interactions of the helix with other protein regions.

2.4.2 Swap of S2400N, L2401R, T2404M and V2406A

A swap of four rationally chosen amino acids led to the production of hexaketide prebassianin **46**, in addition to the main product pretenellin A **39**. Although the titre of the hexaketide products was lower for four amino acid swaps in comparison to swaps of the complete substrate binding helix, an influence on chain length control with the simultaneous mutation of S2400N, L2401R, T2404M and V2406A was observed.

Hence, the residues at these positions include key positions for affecting intrinsic programming.

In summary, the exchange of residues from the substrate binding helix from different iterative PKS systems (Figure 2.27 A) led to change in the polyketide chain production (Figure 2.27 C). The wild type TenS produced mostly pretenellin A **39** (Figure 2.27 B), in addition to other pentaketide minor products but no longer chains. A swap of the complete substrate binding helix with the corresponding DmbS sequence leads to doubly methylated hexaketides (prebassianin A **46**, **105**). A swap with the sequence of MilS gives the same result. Here, additionally minor amounts of a triple-methylated hexaketide (**129**) were detected, which illustrates that the KR domain is relatively slower than the *C*-MeT domain for early chains.
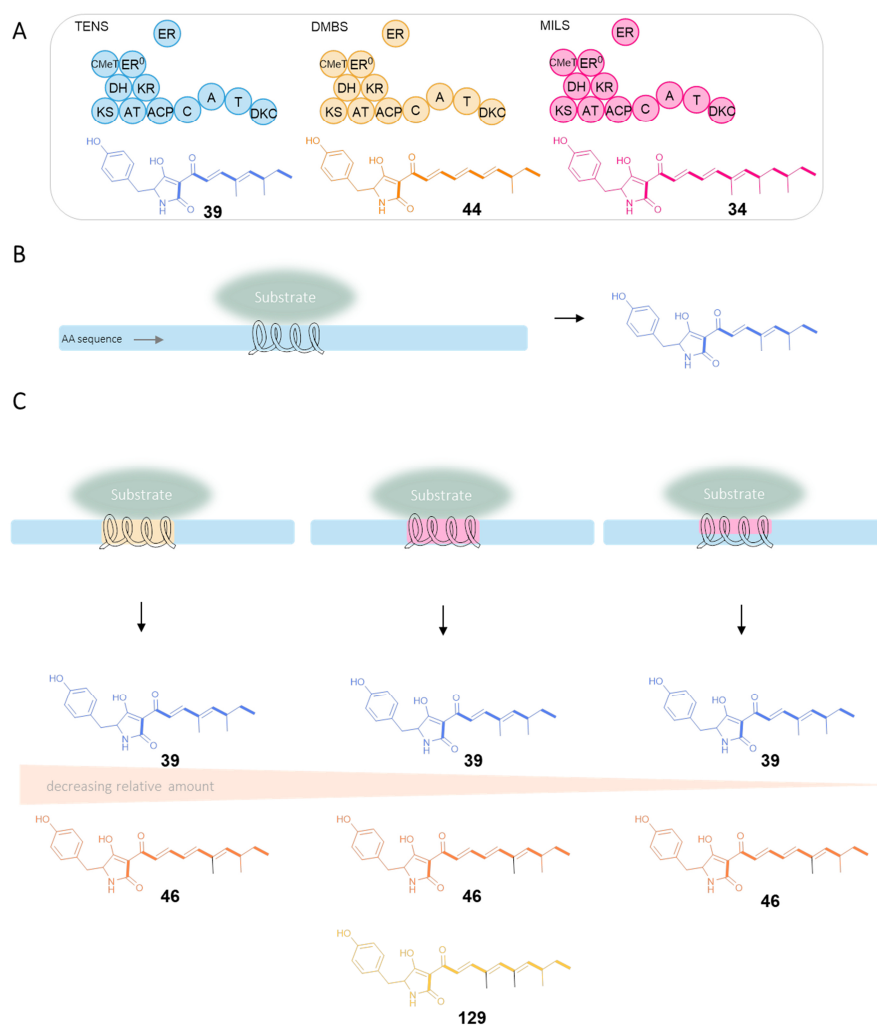
Figure 2.27 Summary of rational engineering experiments of TenS: **A**, overview of iterative hr-PKS and their main products; **B**, scheme of substrate binding helix in the wild type TenS; **C**, scheme of TenS variants and their products.

The lowest hexaketide production was observed with decreasing number of amino acid swaps, which is a result analogous to extrinsic and intrinsic effects on domain level. Cox and co-workers and Simpson and co-workers both described that domain swap experiments are more effective when larger multi-domain swaps are carried out to preserve domain-domain interactions.[58,63] The decrease in re-programmed unnatural products for hybrid TenS with smaller swap-fragments is in agreement with the results of this work.

2.4.3 Protein Structure Analysis of TenS Variants

The predicted protein structure models of the different chimeric TenS variants showed no significant differences in their structural characteristics. Changes in the substrate-binding helix or the resulting binding pocket induced by the mutation of amino acid residues were negligible.

Another aspect, which may play a role in the chain length determination, is the *flexibility* of the substrate binding helix or other structural elements within the protein structure. The AlphaFold model, the threaded model based on the AmphB2 template[58] and the cryo-EM structure of LovB[20] KR domain differ most in the area of the substrate binding helix (section 2.3.1.1, Figure 2.5, Table 2-2), which underlines the aspect of different positioning of the helix. Protein areas with flexibility have a significant influence on the resulting pocket volume of the respective catalytic domain. If the architecture of the catalytic domain is changed, the reaction speed, which is considered a crucial component for the programming in previous publications (Scheme 2.4), also changes. Computational structure analysis methods cannot portray the flexibility with current prediction methods. Therefore, the insights drawn from the structural models can only be interpreted to a very limited extent in relation to programming.

2.4.4 Discussion of Alanine Scan

Of the 15 amino acid positions that were mutated to alanine, 11 mutations did not result in changes in the produced pentaketide products. A change in chain length programming was observed with four positions of alanine scan experiments of the substrate binding helix. However, these results are associated with very low concentration of the hexaketides in each analysis. Not only has the level of production

significance for the reliability of the results, but also the consideration of all producing transformants of one alanine scan position. In the case of the mutations at positions T2404A and M2407A, only one transformant was observed to produce traces of hexaketides, which indicates less reliable results.

At positions V2406A and G2402A, the results were more consistent for the production of hexaketides. In the case of V2406A, hexaketides were detected in 100 % of the cases. For position G2402A, 55 % hexaketide producing transformants were observed. Thus, these positions are suggested to be particularly important for the programming of the PKS.

However, it cannot be ruled out that other amino acid positions may be involved in aspects of the chain length programming. The titre of the products is a crucial factor, as the native pentaketide product was the main product in all obtained transformants and the production of hexaketide compounds is proportionally lower. Therefore, under low overall production levels, the detection of hexaketides was not possible. Additionally, the importance of the three-dimensional protein structure cannot be overstated, as various intermolecular interactions are at play for each position. A simple substitution of an amino acid may not necessarily yield in a pronounced effect, as the interplay with an associated residue may significantly impact the effect of a swap experiment. Extending the analysis from isolated alanine swaps to a combination with the rationally chosen amino acid positions in the future could increase the efficiency of the production of hexaketides.

The hypothesis that the substrate binding helix plays a role in determining the chain length of the resulting polyketide was confirmed. However, the influence of the individual amino acids on the intrinsic selectivity is not clearly defined.

2.5   Conclusion and Outlook

The focus of this project was to determine the influence of the substrate binding helix of the KR domain from TenS on the programming. Previous fragment swaps between TenS and related PKS systems showed that swaps including the substrate binding helix produced hexa- or heptaketides.[58]

For this purpose, TenS structure predictions were successfully obtained using AlphaFold. The protein structure model was evaluated and used for the analysis of the active pocket architecture.

It was possible to confirm the important role of the substrate binding helix by swapping the substrate binding helix sequence of TenS to DmbS. This swap led overall to the most effective change of the observed products with a successful production of mayor amounts of hexaketide **46** and **105** by swapping only 15 amino acids. For a substrate binding helix swap to MilS only hexaketides **46**, **105**, and **129** were observed, but no heptaketides. The amino acid mutation T2395A effected the change in production from heptaketides in previous publications to hexaketides, which is an indication that T2395 could be a decisive position of the helix in the combination with other amino acid positions. An effect of this position on the overall conformation or flexibility of the helix can be speculated, as T2395 is positioned at the N-terminal end of the substrate binding helix (Figure 2.28, green), which clamps over the polyketide intermediate.

Furthermore, the AlphaFold protein model was analysed regarding the orientation of each single amino acid of the substrate binding helix to the active site pocket. Simultaneous mutations of four positions that are likely to contact the substrate (S2400N, L2401R, T2404M and V2406A) achieved the production of hexaketide **46** in heterologous expression experiments.

The results of the alanine scans are in agreement with an involvement in chain length control by amino acids V2406 (Figure 2.28, cyan) and T2404 (Figure 2.28, yellow) on the intrinsic programming of the KR. TenS including mutation G2402A, on the other hand, was not integrated into the simultaneous swap of different positions, but led to the production of hexaketides. While G2402 (Figure 2.28, pink) is not directed to the substrate binding site, the position can still have an impact on structural characteristics of the helix, like the conformation or the interaction with residues outside

of the substrate binding helix. Two follow-up experiments are conceivable to further investigate the individual amino acid positions: (I) A swap of the already simultaneously changed amino acids with an additional swap of position G2402A (or G2402R for a swap to the amino acid of DmbS at this position), which could further enhance the production of hexaketides. (II) Another set of mutations can exclude position S2400N (Figure 2.28, red), L2401R (Figure 2.28, orange) and T2404M (Figure 2.28, yellow) in a simultaneous swap. If this leads to the same results as before, the very important role of position V2406A (and potentially G2402A) is further emphasised.

Additionally, position T2395 (Figure 2.28, green) should be the focus for future work, as mutation of this position leads to the preservation of hexaketide production over heptaketides when swapping the remaining substrate binding helix residues. However, the identification of separate intrinsic and extrinsic factors for chain length control may have reached its limits by the identification of a small number of amino acid residues.

green: T2395, red: S2400, orange: L2401, pink: G2402, yellow: T2404, cyan: V2406
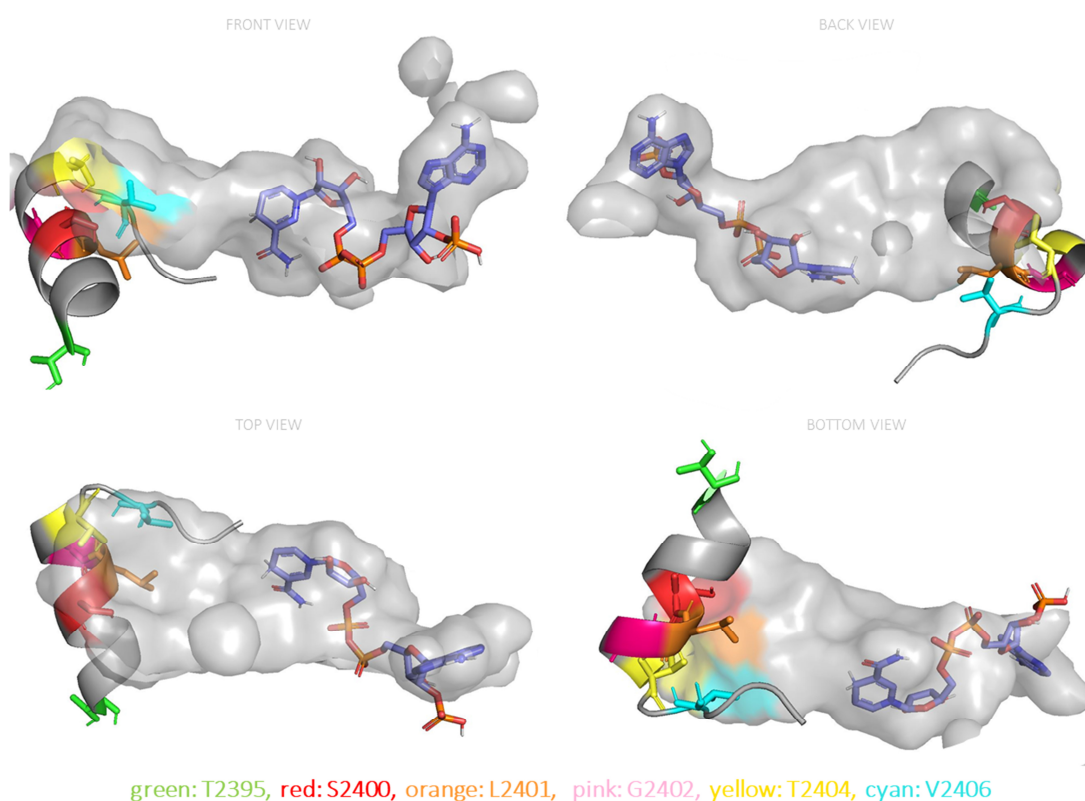
Figure 2.28 TenS substrate binding helix and substrate binding pocket.

The findings reinforce the hypothesis by Cox and co-workers that the change of a small number of amino acids leads to the production of a range of unnatural products.

Furthermore, they speculated that the production of a small library of related compounds could be an evolutionary advantage, where a change of few positions in an hr-PKS sequence leads to fast and easy adaptation to changes in the environmental conditions.[58]

In summary, the substrate binding helix swap with DmbS shows for the first time that production of hexaketides as the main product is possible by changing only 15 amino acid positions. Furthermore, a change of only four amino acid positions, which were rationally chosen, leads to hexaketides in addition to pentaketides as the main product. Finally, the swap of only one amino acid position was able to affect the programming by producing traces of hexaketide compounds. However, this approach has probably reached its limit. In future, better models will be required that take dynamics of the system into account and that are able to sample more closed confirmations of the KR domain.

3   Maleic Acid Anhydride Biosynthesis in *H. lienhwacheense*

3.1 Introduction

The *Xylariaceae* (Xylariales, Ascomycota), including the genera *Xylaria*, *Daldinia* and *Hypoxylon*, are not only one of the largest families among ascomycetes (with over 1300 species in 2013), but also outstanding when it comes to the prolific production of secondary metabolites.[135] For this reason, they have been studied extensively in the last years resulting in the discovery of diverse azaphilones,[136,137] sporothiolides,[78,138] and cytochalasins[139,140] among others.[141]

From the group of *Xylariaceae,* in particular the *Hypoxylaceae,* have been the focus of studies due to their high diversity of over 350 species.[141] Despite the high potential for the discovery of secondary metabolites, many species have only recently been studied on a genomic level. In 2021 a group of 14 selected genomes was sequenced to understand phylogenetic family relationships, including the species *Hypoxylon lienhwacheense*.[141] This wood decaying fungus was found in Northern Thailand and first described in 1996.[142] The species produces high amounts of natural products, which were extracted and analysed in previous work by the collaborating group of Prof. Dr. Stadler and co-workers (Helmholtz Institute for Infection Research, Braunschweig). Due to the interesting structure of the secondary metabolites found, they are the objects of research for this project.

3.1.1 Natural Products from *Hypoxylon lienhwacheense*

The new fungal compounds lienhwalides A-C **130** - **132** and the known compound cordyanhydride B **61** (Figure 3.1) were discovered in the extracts of *H. lienhwacheense* cultures and identified by isolation and NMR analysis by Dr. Frank Surup (Helmholtz Institute for Infection Research, Braunschweig) and co-workers.

The two structural components of the lienhwalides belong to known compound classes: tropolones and maleic acid anhydrides - both derived from PKS-products (section 1.4.2 and 1.4.3). Lienhwalides B **131** and C **132** are diastereomers. In

comparison to lienhwalide A **130**, they are *O*-methylated and oxidised at the aromatic seven-membered ring system.

lienhwalide A **130**   lienhwalide B **131**   lienhwalide C **132**

cordyanhydride B **61**   **133**
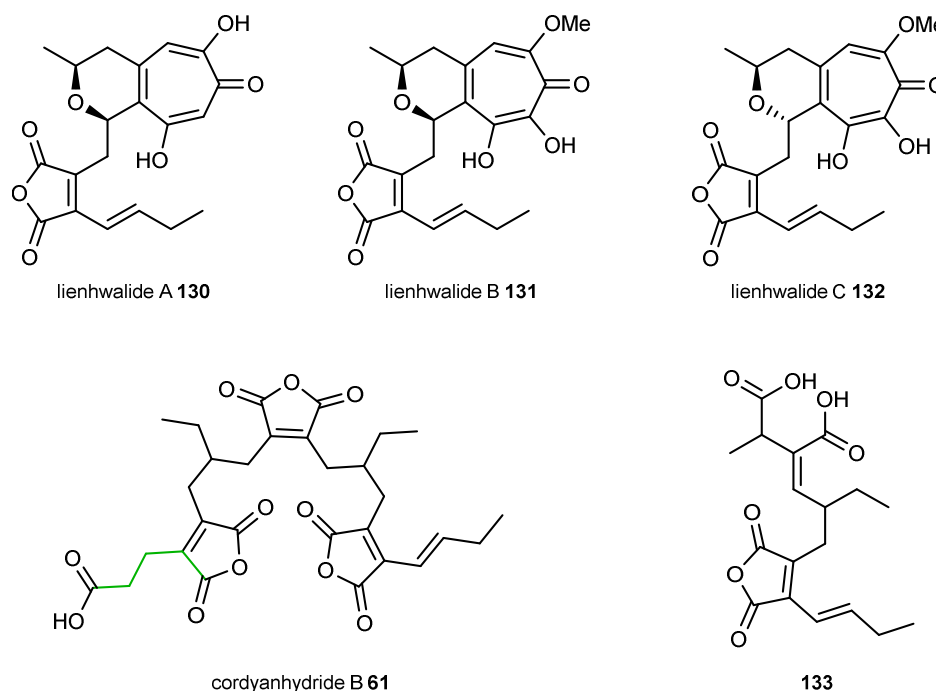
Figure 3.1 Natural products from *H. lienhwacheense*.

Cordyanhydride B **61** does not include a tropolone moiety. It consists of three maleic acid anhydride units linked together linearly. Additionally it contains one maleic acid anhydride unit with a longer carbon-chain compared to the usual maleic acid monomers (Figure 3.1, green). This compound was found previously in the extracts of other fungi (*Cordyceps pseudomilitaris*, section 1.4.2.1), but its biosynthetic pathway has not been studied previously.[77] In *Cordyceps pseudomilitaris*, linear linked maleic acid anhydride compounds containing only two units were also found (cordyanhydride A **60**).[77] Interestingly, cordyanhydride A **60** was not isolated from *H. lienhwacheense* by our collaborators. However, compound **133**, consisting of a conjugate of maleic anhydride and maleic acid was detected. **133** is a possible intermediate for cordyanhydride B production.

3.1.2 Genome Analysis of *H. lienhwacheense*

Genome sequencing was performed by a combination of Illumina and Oxford nanopore technology with the strain *H. lienhwacheense* MFLUCC 14-1231.[141] Gene prediction was performed using Augustus version 3.2,[143,144] GeneMark-ES 4.3.6,[145] GenDB2.0[146] and the databases COG,[147] KEGG[148] and SWISS-PROT.[120,141,149] Sequencing, assembly and annotation were perfomed by Dr. Daniel Wibberg at the Centre for Biotechnology (CeBiTec, Bielefeld). The identification of the BGCs (Table 3-1) was based on the BLASTp Algorithm[150,151] in combination with Geneious v. 9.1.8 searches of the genome against known enzymes for specific biosynthetic pathways, performed by Dr. Eric Kuhnert.[149] The co-localised predicted genes were further analysed by BLASTp homology search against the Swiss-Prot database.[149] The high quality genome sequencing analysis (Table 3-1) revealed a rather small genome size (35.8 Mbp) and total BGC number (25) in comparison to related species.[141]

Table 3-1 Characteristic parameters of genome sequence analysis of *H. lienhwacheense*.[149]

Parameter		Value
Genome size		35.78 Mbp
Contigs		61
N50		1.6 Mbp
Genes		9942
BGCs total		25
BGC types:		
PKS	nr-PKS	2
	hr-PKS	10
	pr-PKS	1
	Type III	1
NRPS (-like)	PKS-NRPS	1
	NRPS Siderophore	2
	Other NRPS	2
	NRPS-like	6
Alkyl-citrates	PKS	1
	FAS	1
	Terpene	3

Lienhwalides **130** - **132** combine both tropolone and maleidride moieties. However, a single BGC containing genes that are involved in biosynthesis of both moieties was not discovered. Instead, two separate putative BGC candidates for the biosynthesis of lienhwalides were identified by Dr. Eric Kuhnert. One BGC encodes tropolone genes; the other encodes maleidride genes. The finding of two separate gene clusters for the biosynthesis of the lienhwalide components allows discussing the pathways separately

in the following sections (chapter 3 for maleic acid anhydrides and chapter 4 for tropolones).

3.1.3 Background of Maleic Acid Anhydride Biosynthesis

3.1.3.1 Biosynthetic Pathway of Maleic Acid Anhydrides in Fungi

The analysis of the biosynthetic origin of maleidrides started in 1965 by Sutherland and co-workers by isotopic labelling and degradation studies. It was shown, that the C_9-monomer of glauconic acid **47** is built from two different carbon chains: oxaloacetic acid **64** and hexanoate that was suggested to be derived from fatty acid biosynthesis.[152,153] The biosynthetic origin of two carbon chains was later confirmed for rubratoxins **56** and phomoidrides **57** (which contain alternative side chains) by similar methods.[154,155]

Only decades later, with the development in molecular biology methods and the analysis of genomes, was it possible to study the biosynthetic pathway in more detail. Oikawa and co-workers reported the BGC responsible for the biosynthesis of the maleidride phomoidride in 2015.[156] A combination of heterologous expression experiments, gene knock-outs and *in vitro* assays revealed the biosynthetic pathway of the maleic acid anhydride precursors.[72,156,157] One of the best studied examples is the biosynthesis of maleic acid anhydride monomers in *Byssochlamus fulva* (Scheme 3.1).[158,159] An hr-PKS synthesizes the tetraketide **134**, which is coupled to oxaloacetic acid **64** by an alkylcitrate synthase. Later work showed that only CoA thiolesters are substrates for the alkylcitrate synthase.[160] This is followed by the dehydration reaction of **135** by an alkylcitrate dehydratase, which enables the formation of the maleic acid anhydride **136**. Spontaneous decarboxylation reaction results in the reactive monomers **138** and/or **139**, presumably *via* anion **137a-b**.[158]
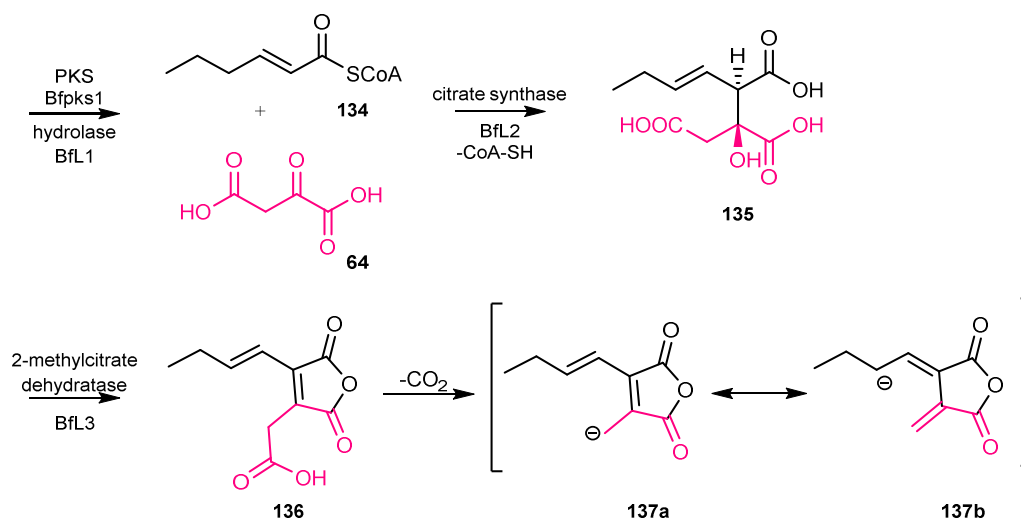
Scheme 3.1 Biosynthetic pathway of byssochlamic acid precursors in *B. fulva*.

3.1.3.2 Coupling Reactions of Maleic Acid Anhydride Molecules

Sutherland and co-workers first hypothesized 1962 the biosynthetic origin of maleidrides *via* a coupling reaction to form homodimers.[66,69] The first evidence was obtained by feeding two isotopically labelled C$_9$-precursors *in vivo* to *Penicillium purpurogenum*.[161]

Decades later, Cox and co-workers reported the dimerization catalysed by two enzymes in the BGC of *B. fulva* in 2016: the ketosteroid isomerase (KI)-like protein and a phosphatidylethanolamine-binding-like protein (PEBP). While the KI-like enzyme appears to catalyse the dimerization, the PEBP seemed to improve the titre of the dimerized products.[158] Similar findings were published in the case of cornexistin **49** and zopfiellin **52** biosynthesis.[71,162] In 2021 Cox and co-workers reported that the dimerization reaction is enabled by the KI-like enzyme alone based on *in vitro* studies with yeast cell-free extracts.[160] Recently, the name maleidride dimerising cyclase (MDC) was introduced for the KI-like enzyme.[72] The PEBP protein was speculated to chaperone the reaction, *e.g.* by binding the unstable anionic intermediates **137a-b** as related PEBP are known to have a high affinity for anions.[70,163]

The size, substitution pattern and stereochemistry of the resulting cyclic dimers are determined by the orientation of the monomers to each other (Figure 3.2).[158,164] A two-step reaction was suggested, where one acceptor monomer (**138** or **139**) undergoes the condensation reaction with the donor **136**, which is previously decarboxylated.[158,164] The dimerization reaction of the monomers results in byssochlamic acid **48** like

nonadrides, when a head-to-tail orientation of the monomers is present (Figure 3.2 B). In contrast an orientation of head-to-side leads to the heptadride agnestadride A **54** (Figure 3.2 C), found in the same fungus as **48**. Monomers **139** (Figure 3.2 E) or **138** (Figure 3.2 D) are also involved in head-to-head dimerization, leading to glauconic acid **47** type or heveadride **51** type maleidrides, respectively. Two already decarboxylated monomers do not dimerize in the *in vitro* studies by Yin *et al.*[160] In contrast, phomoidride **57** type products are reported to dimerize by a reaction between two monomers without the previous decarboxylation of the donor (Figure 3.2 F).[164]
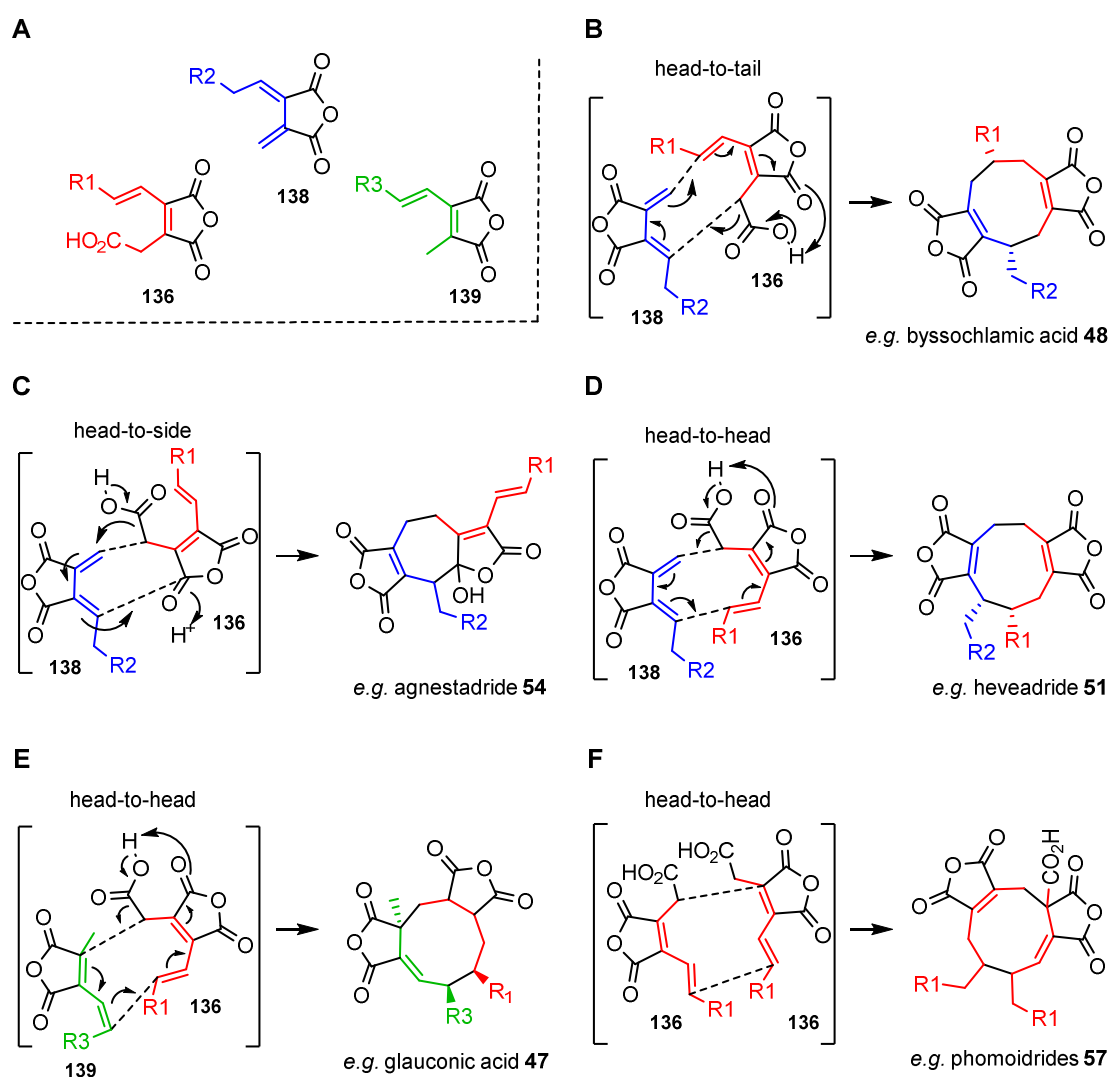
Figure 3.2 Proposed dimerization mechanism of maleic acid anhydrides: **A,** maleic acid anhydride monomer precursors; **B-F**, dimerization modes and resulting molecule structure.

3.2 Project Aims

The overall aim of the project is the understanding of the biosynthesis of secondary metabolites of *H. lienhwacheense*. This chapter focuses on the biosynthesis of maleic acid anhydride **136** metabolites.

The genome sequencing, analysis and gene prediction was carried out by previous work.[141,149] The starting point for the investigation is the analysis of the putative BGC for maleic acid anhydride biosynthesis. Putative genes for the biosynthesis will be identified and used in heterologous expression experiments in the host *A. oryzae* NSAR1.

After the determination of the key enzymes for the biosynthesis, the aim will be to increase the maleic acid anhydride production in the host by introducing a CoA ligase enzyme of the BGC for activation of the biosynthetic precursors for maleic acid anhydride synthesis. During the studies, an unexpected maleic acid anhydride congener will be detected, purified and structurally analysed. The biosynthetic origin will be investigated by *in vitro* studies with the *H. lienhwacheense* alkylcitrate synthase.

Finally, an attempt to analyse the biosynthetic origin of cordyanhydride B **61** will be carried out, which focusses on the MDC enzyme of the BGC by *in vivo* investigations.

3.3 Results

3.3.1 Cultivation of *H. lienhwacheense*

H. lienhwacheense was cultivated for 13 days at 25 °C and 150 rpm with 200 ml YMG liquid media in 500 ml shake flasks, followed by extraction (EtOAc) and analysis of the organic extracts by LCMS (Figure 3.3). The cultivation under laboratory conditions shows the production of the tropolone maleic acid anhydride conjugates **130**, **131**, and **132** and other related maleic anhydride compounds **61** and **133**.
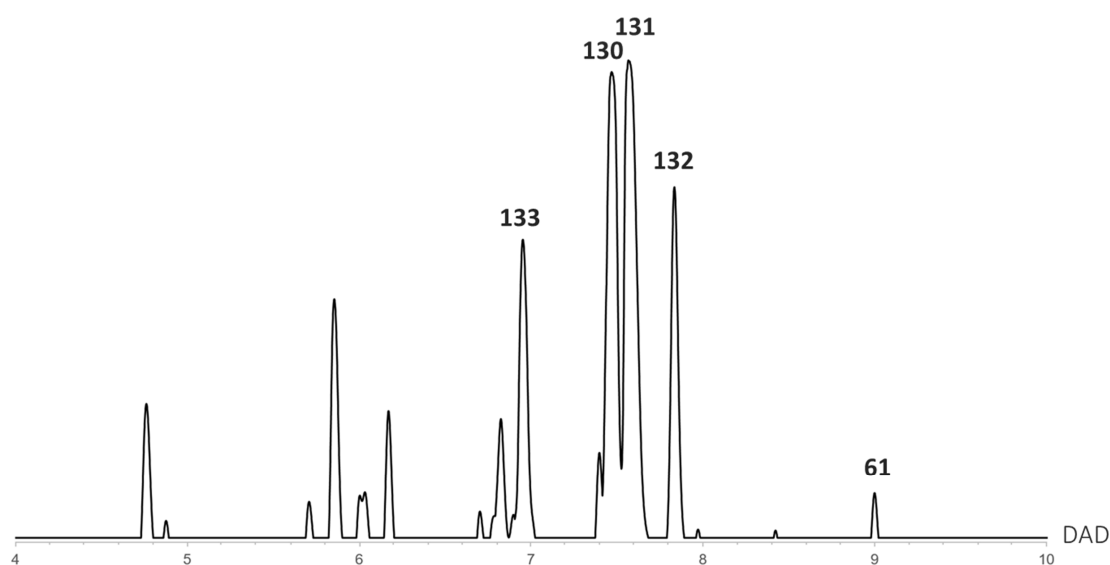
Figure 3.3 DAD chromatogram (210 – 600 nm) of extract from *Hypoxylon lienhwacheense* in 200 ml YMG culture at 25 °C for 13 days at 150 rpm.

The analysis of the extract revealed the same secondary metabolites as previously found by our collaborators (Helmholtz Institute for Infection Research) in high amounts, indicated by intense peaks in the LCMS chromatograms and the characteristic UV and MS data (Figure 3.4).
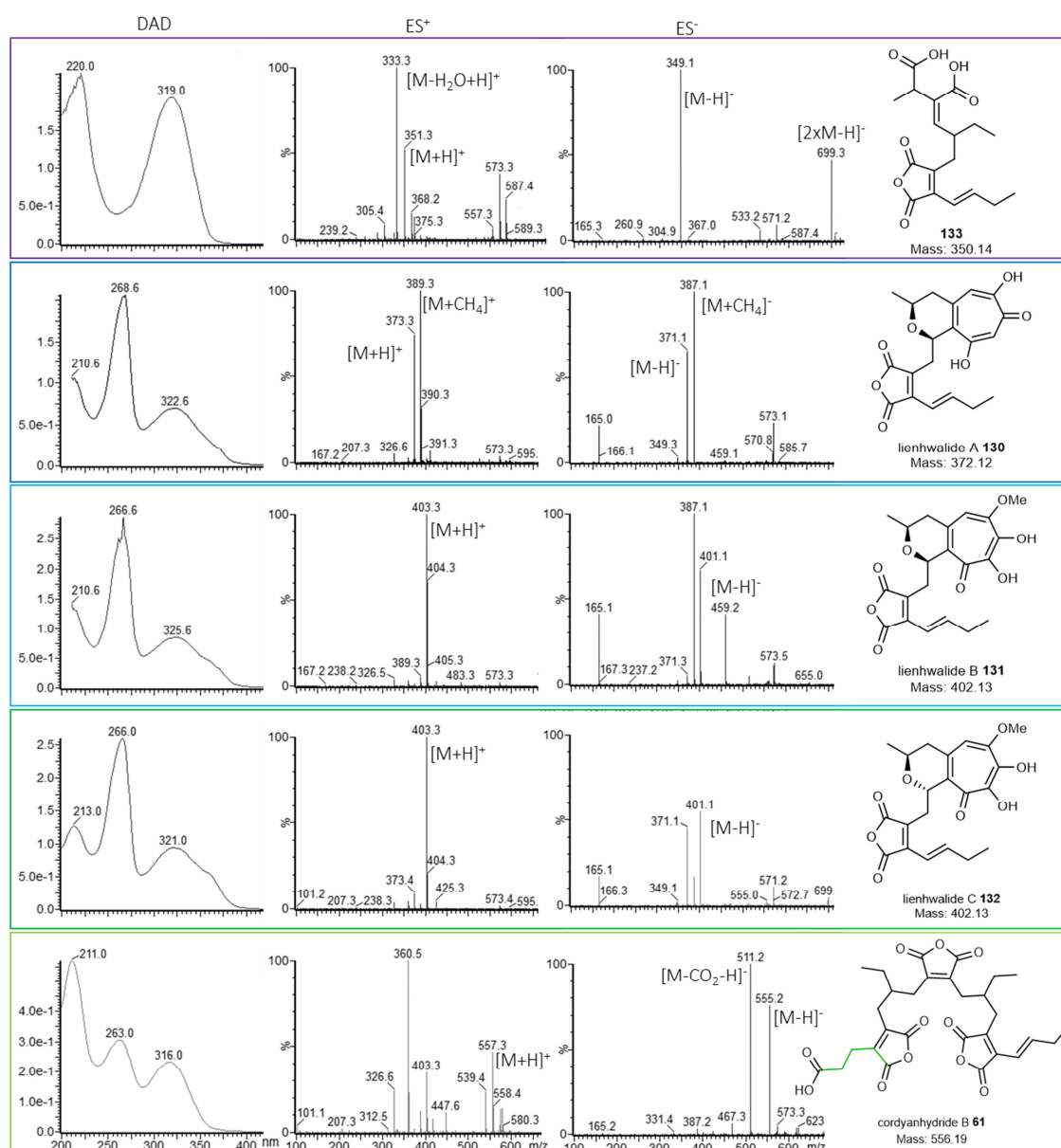
Figure 3.4 Summary of characteristic data from compounds of *H. lienhwacheense*.

Lienhwalides A-C **130**-**132** have similar UV spectra with three maxima at around 210 nm, 266 nm and 322 nm (Figure 3.4). Lienhwalides B **131** and C **132** are detected by the same *m/z* in ES⁻ and ES⁺ spectra and can therefore only be differentiated by the retention time (with B: 7.6 min and C: 7.9 min, Figure 3.3). Lienhwalide A **130** can be identified by its mass of 372 Da and a retention time of 7.5 min (Figure 3.3, Figure 3.4).

Cordyanhydride B **61** was identified by the retention time of 9 min and a mass of 556 Da, in addition to the characteristic UV spectra with similar values of 211 nm, 263 nm and 316 nm. Additionally, compound **133** with a retention time of 7 min was detected with a mass of 350 Da.

Different liquid media were tested to try to find producing and non-producing conditions to enable transcriptome analysis. In all tested media (YMG, PDB, DPY, MMK2) it was not possible to observe non-producing conditions. Nonetheless, the mRNA was extracted from the mycelia of an *H. lienhwacheense* liquid cultivation (RNA Clean and Concentrator^TM Kit, *Zymo Research*, Figure 3.5). The transcriptome analysis was carried by Dr. Daniel Wibberg (CeBiTec, Bielefeld). The analysis of the transcriptome was used to confirm previously predicted genes and introns by bioinformatic methods described before.
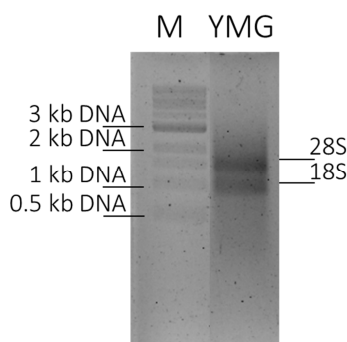
Figure 3.5 RNA gel electrophoresis (run with 5 µl RNA at 70 V for 45 min).

The RNA was reverse transcribed to cDNA (High Capacity RNA-to-cDNA Kit, *Thermo Fisher Scientific*) to give intron-free sequence. The obtained cDNA was used for the expression analysis of the biosynthetic pathways of maleic acid anhydride and tropolone monomers in the following studies.

3.3.2 Analysis of Biosynthetic Gene Cluster and related Pathways

Maleic acid anhydrides and alkyl citrates are known from a broad range of fungi (section 1.4.2). The gene clusters of related compounds from related fungi were used as a basis for the analysis of the pathway. Such pathways include either a PKS or FAS that builds the main carbon skeleton with a following acyl CoA activation, and an alkylcitrate synthase that adds the key oxaloacetic acid **64**. BGCs that encode maleidride biosynthesis enzymes are therefore easily detected by searching for the key alkylcitrate synthase-encoding genes.

The search in the *H. lienhwacheense* genome revealed two alkylcitrate synthase genes within putative BGCs (Figure 3.6). The *bf*-BGC from *Byssochlamus fulva* is

known to produce maleidrides like byssochlamic acid **48** and agnestadrides **54-55** (section 1.4.2) and is well understood.[43] The bioinformatic alignment tool clinker was used to show a relationship between the *bf*-BCG with the putative maleic anhydride producing BGCs from *H. lienhwacheense* (Figure 3.6).[165]
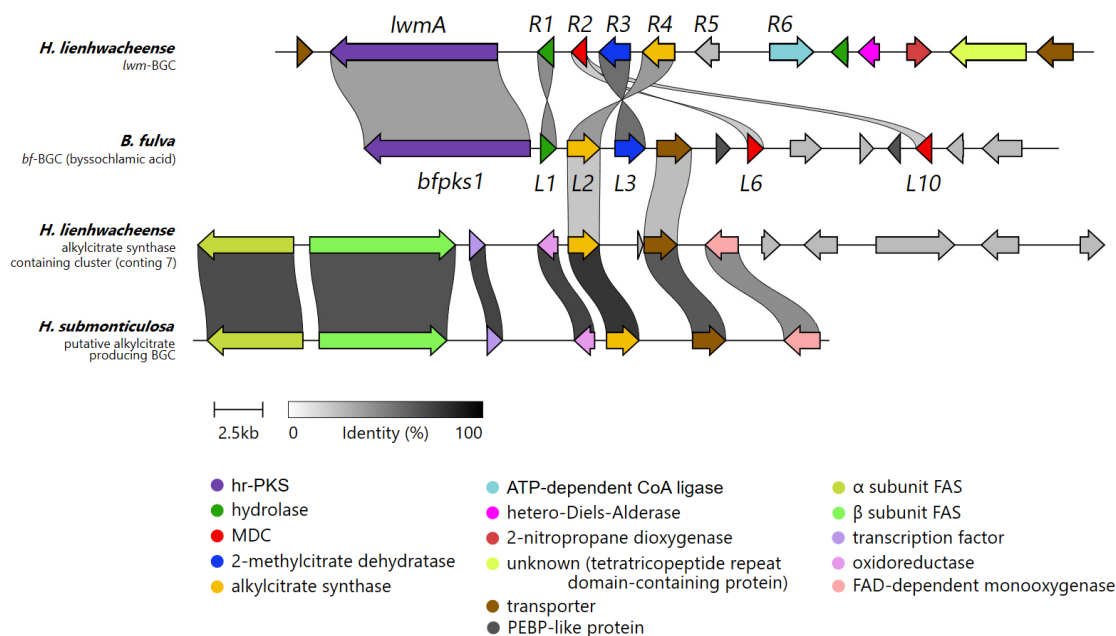
Figure 3.6 Gene cluster comparison of alkylcitrate synthase containing BGC from *H. lienhwacheense*, *B. fulva* and *H. submonticulosa.*

One of the putative BGCs, which will further be called the *lwm*-BGC, includes five genes, each with high similarity to corresponding *bf*-BGC genes. In contrast, the second cluster containing an alkylcitrate synthase includes only two genes with a high similarity to *bf*-BGC genes, from which one is a non-catalytic putative transporter protein encoding gene. Thus, the *lwm*-BGC was used for further investigations of the maleic acid anhydride pathway. The second BGC on contig 7 (Figure 3.6) includes putative genes encoding a FAS and five additional genes, which all show homology to a known putative alkylcitrate producing BGC (*e.g.* from *H. submonticulosa*), which is being investigated in context of other projects from our group.

The similarity of the *lwm*- and *bf*-BGC (Figure 3.6) is an indication that early steps of the byssochlamic acid **48** and putative maleidride pathways are analogous in the case of *H. lienhwacheense*. The core proteins (hr-PKS, hydrolase, MDC, 2-methycitrate dehydratase, citrate synthase) of both clusters show a high level of similarity. The putative genes (Table 3-2) for the *lwm*-BGC were additionally reviewed using the BLASTx algorithm.[121,150,151] In *H. lienhwacheense* remaining putative genes, which

were not similar to *B. fulva* genes are: a hypothetical gene with unknown function (*lwmR5*), a putative acyl-CoA ligase encoding gene (*lwmR6*), and a second putative hydrolase gene, which shows low similarity with the hydrolase from *B. fulva*. Downstream of the putative hydrolase encoding gene, a hetero-Diels-Alderase is present, next to a putative 2-nitropropane dioxygenase-like protein encoding gene. The borders of the BGC were defined according to the homology to the *bf*-BGC. In contrast to other known maleidride producing BGCs, no putative PEBP encoding gene is included in the *lwm*-BGC.

Table 3-2 Annotation of maleic acid anhydride gene cluster from *H. lienhwacheense with b*lastx search against SwissProt database.

putative Gene	length [bp] / [aa]	putative function	blastx hit, identity	identity *bf*-BGC protein
lwmA	7830 / 2610	hr-PKS	Type I Iterative PKS, *Monascus purpureus*, 54 %	47 %
lwmR1	666 / 222	hydrolase	esterase alnB, *Aspergillus nidulans*, 40 %	54 %
lwmR2	684 / 228	MDC	-	31 % (each)
lwmR3	1482 / 494	alkylcitrate dehydratase	2-methylcitrate dehydratase-like protein oryR, *A.oryzae*, 51 %	65 %
lwmR4	1326 / 442	alkylcitrate synthase	citrate synthase-like protein oryE, *A.oryzae*, 43 %	50 %
lwmR5	1215 / 405	unknown	-	-
lwmR6	1758 / 586	acyl-CoA ligase	acyl-CoA ligase easD, *Aspergillus nidulans*, 41 %	-
-	753 / 251	hydrolase	hydrolase pyvD, *Aspergillus violaceofuscus*, 28 %	-
-	1056 / 352	unknown	hetero-Diels-Alderase	-
-	1002 / 334	unknown	2-nitropropane dioxygenase, *Pleomassaria siparia*, 79 %	-

The known key enzyme homologues for the biosynthesis of byssochlamic acid **48** precursors were selected for the heterologous expression in *A. oryzae* NSAR1.

3.3.3 Heterologous Expression of Proposed Core Enzymes

3.3.3.1 Vector Construction

Vectors for heterologous expression were constructed by yeast recombination and LR-recombination methods, similarly as described in section 2.3.2. The template for PCR amplification (Figure 3.7 A and D) was cDNA (section 3.3.1) with appropriate primer

pairs (H1/H2, 2MDH1/2MDH2, CS1/CS2, Table 6-7). The primer included overhangs of 30 bp homologous to regions to the sequence of the insertion site of the vector to result in appropriate fragments for yeast recombination (Figure 3.7 A, D). The genes *lwmR1*, *lwmR3* and *lwmR4* were inserted into the fungal expression vector pTY-GS-*ade* after restriction digestions with the enzyme *AscI* (Figure 3.7 E).
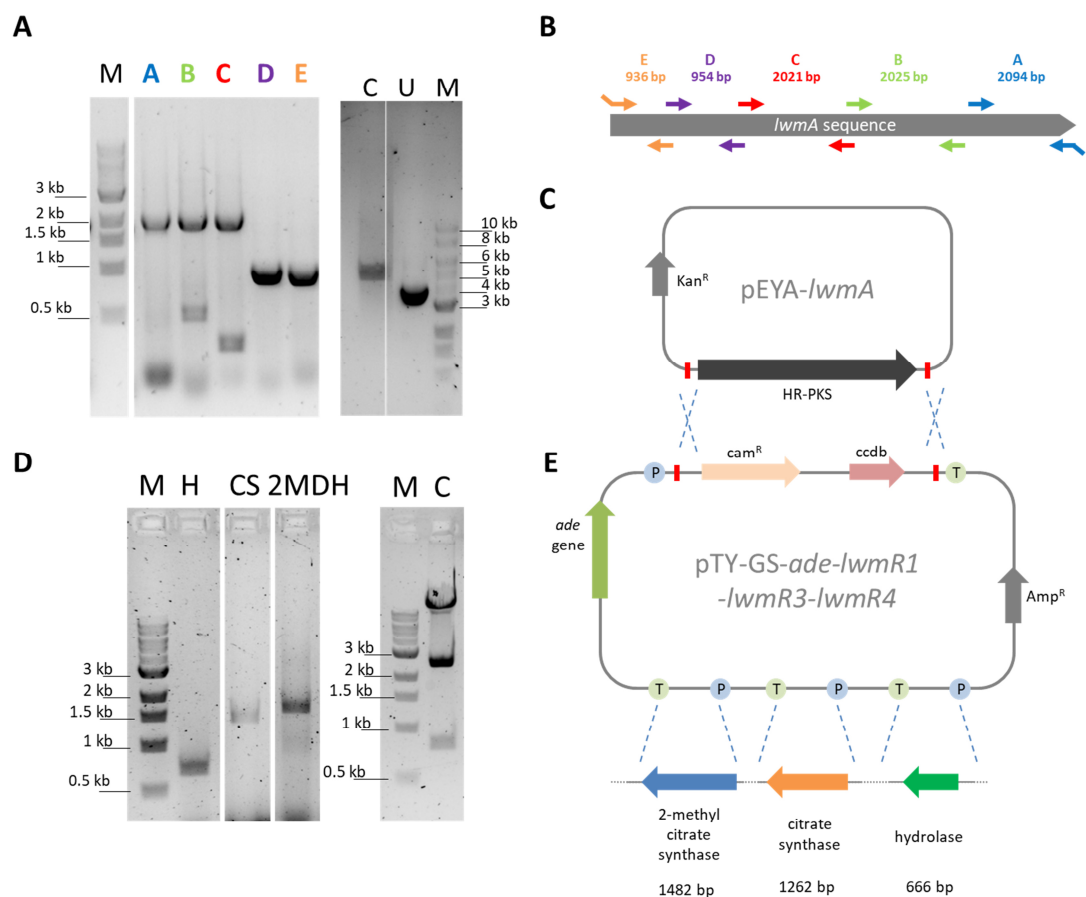
Figure 3.7 Vector construction of fungal expression vector including 4 putative genes for maleic anhydride biosynthesis: **A**, fragments used for pEYA-*lwmA* construction (C = cut vector, U = uncut vector); **B**, scheme of *lwmA* fragments for yeast recombination; **C**, resulting pEYA-*lwmA* vector after yeast recombination; **D**, fragments used for construction of the expression vector (C = cut vector); **E**, resulting pTY-GS-*ade-lwmR1-lwmR3-lwmR4* vector.

The putative PKS-gene *lwmA* was first reconstructed in the entry vector pEYA. The vector was previously digested using the restriction enzymes *NotI* and *AscI*. Due to the length of approx. 7800 bp, yeast recombination was performed using five overlapping fragments (Figure 3.7 A-B, fragments A-E) of the PKS sequence using primers pairs A1-E2 (Table 6-7). After recombination, the expected resulting vector (Figure 3.7 C)

was verified with PCR amplification and following sequencing methods as described previously. Subsequent performance of *in vitro* LR recombination of vectors pEYA-*lwmA* and pTY-GS-*ade-lwmR1-lwmR3-lwmR4* completed the construction of the final expression vector pTY-*ade-lwmA -lwmR1-lwmR3-lwmR4*.

3.3.3.2 Transformation of Proposed Core Genes

The four core genes (encoding PKS, hydrolase, alkylcitrate synthase, 2-methycitrate dehydratase) of the proposed pathway (Scheme 3.1) were all shown to be present in the vector pTY-*ade-lwmA-lwmR1-lwmR3-lwmR4* by PCR. This vector was transformed into the heterologous host *A. oryzae* NSAR1. The transformants undergo a selection process, cultivation in liquid media and an extraction followed by LCMS analysis (see section 2.3.3.1).

Overall, 16 transformants were obtained. Four transformants showed production of maleic acid monomer **136** (Figure 3.8), characterized by a new peak at 6 min with the typical UV maximum of 314 nm and a characteristic *m/z* of 165 ([M - CO_2]$^-$) in ES$^-$ spectra. These characteristic data and the retention time indicate the maleic acid anhydride compound **136**, identical to previous studies by Dr. Sen Yin.[160]
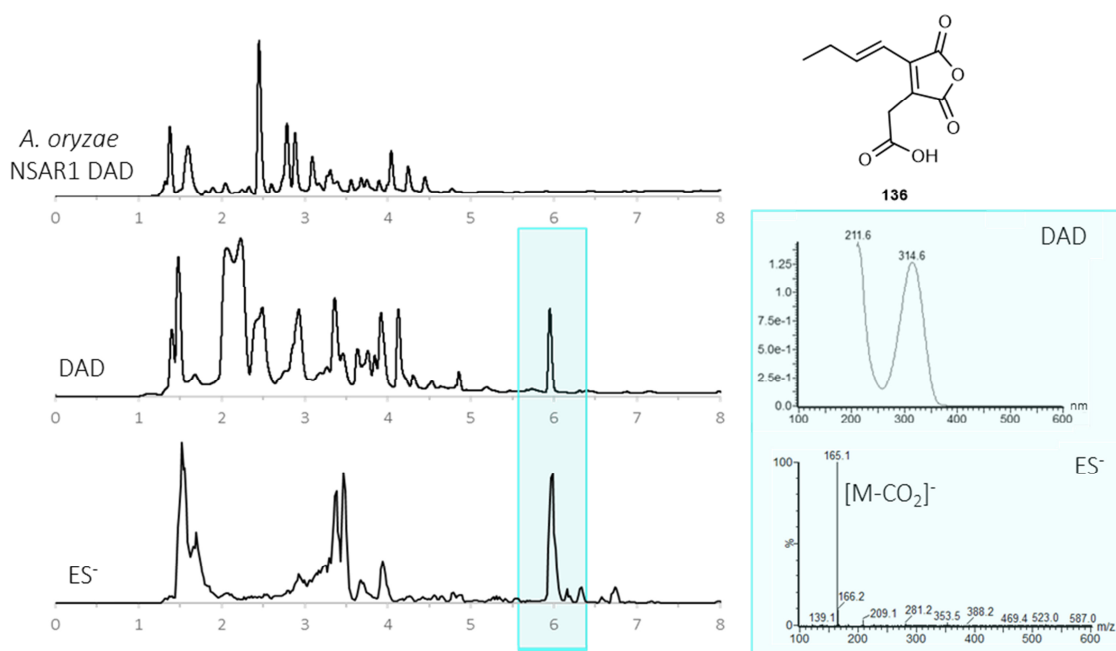
Figure 3.8 LCMS analysis of transfomant including *lwmA, lwmR1, lwmR3, lwmR4*: DAD chromatograms (210 – 600 nm) of WT and transformant, ES$^-$ and ES$^+$ chromatograms of transformant and corresponding spectra at 6 min.

One transformant (from the overall obtained 16 transformants) showed an intense peak for the production of **136** (shown in Figure 3.8). The three additional positive transformants (data not shown) showed minor amounts of **136**, only visible when searching for the expected fragment ion in the extracted ion chromatogram. For the other twelve transformants, no related compounds were detectable.

3.3.3.3 Co-expression of LwmR6 with LwmA + R1 + R3 + R4

Previous *in vitro* experiments of proteins from the *B. fulva* byssochlamic acid **48** BGC from Cox and co-workers in 2021 showed that the citrate synthase BfL2 only accepts acyl CoA-substrates.[160] ATP-dependent acyl-CoA ligases catalyse the reaction of the polyketide **140** with CoA (Scheme 3.2) to form **134.** This leads to the hypothesis that addition of the putative acyl-CoA ligase gene *lwmR6*, which is present in the BCG (Figure 3.6), may increase the efficiency of the pathway and therefore its inclusion may lead to higher yields.
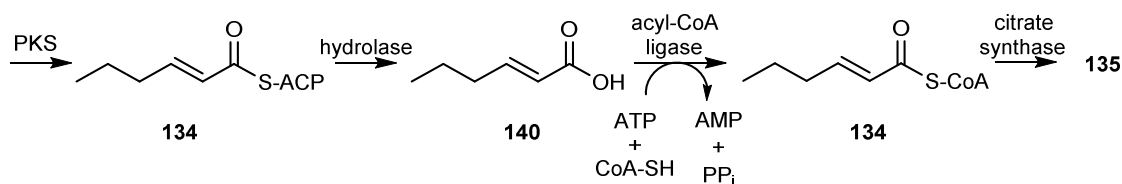
Scheme 3.2 Reaction of acyl-CoA ligase.

The putative ligase encoded in the BGC (*lwmR6*) was added to the heterologous expression system with the vector pTY-GS-*argB-lwmR6* (constructed with previously described methods with primers ACL1/ACL2), which was simultaneously transformed with the previously obtained vector pTY-*ade-lwmA-lwmR1-lwmR3-lwmR4* followed by the usual selection, cultivation and LCMS analysis process.

As a result, 13 out of 17 of the selected transformants produced maleic acid monomers **136** (Figure 3.9) in high amounts. The yield of **136** increased approx. 4-fold for the comparison of highest producing transformants (with and without *lwmR6* respectively, Figure 3.8 and Figure 3.9) as indicated by peak areas of the DAD chromatogram , but titres were not quantified during the studies. An improvement of the biosynthetic pathway efficiency, according to Scheme 3.2 was confirmed.
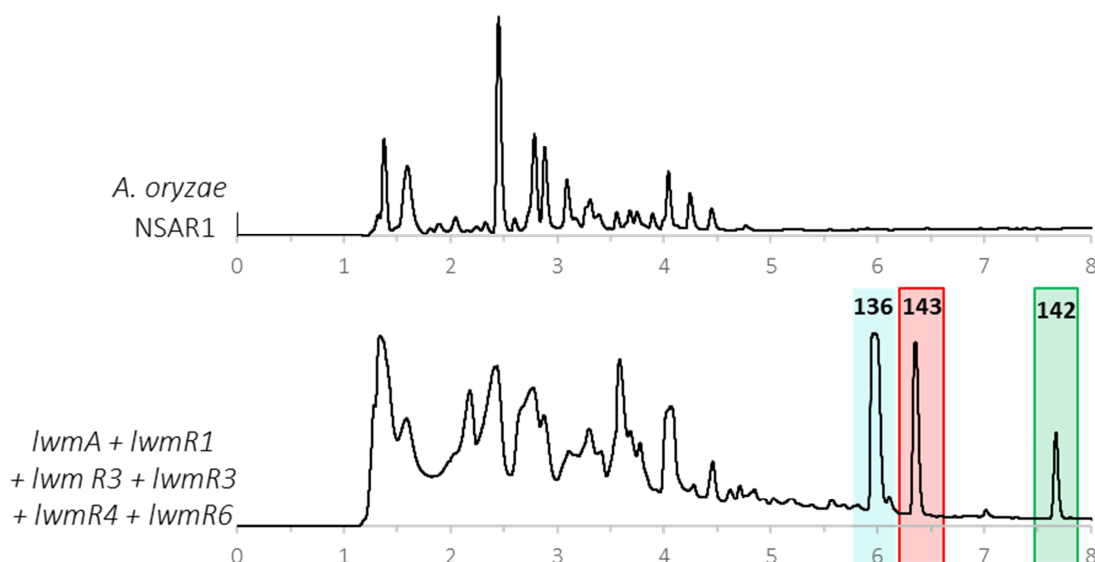
Figure 3.9 DAD chromatogram of transformant containing 4 core enyzmes and the acyl CoA ligase.

In comparison to the previous expression of transformations including the core 4-gene set of enzymes for the pathway, two additional peaks were observed (Figure 3.9, red and green). The peak at 7.5 min (green) was identified as a known compound from previous studies from Dr. Sen Yin.[160] The characteristic data (Figure 3.10 A) and retention time is consistent with structure **142**, which is the product of a spontaneous decarboxylation of **136** (Figure 3.10 B).
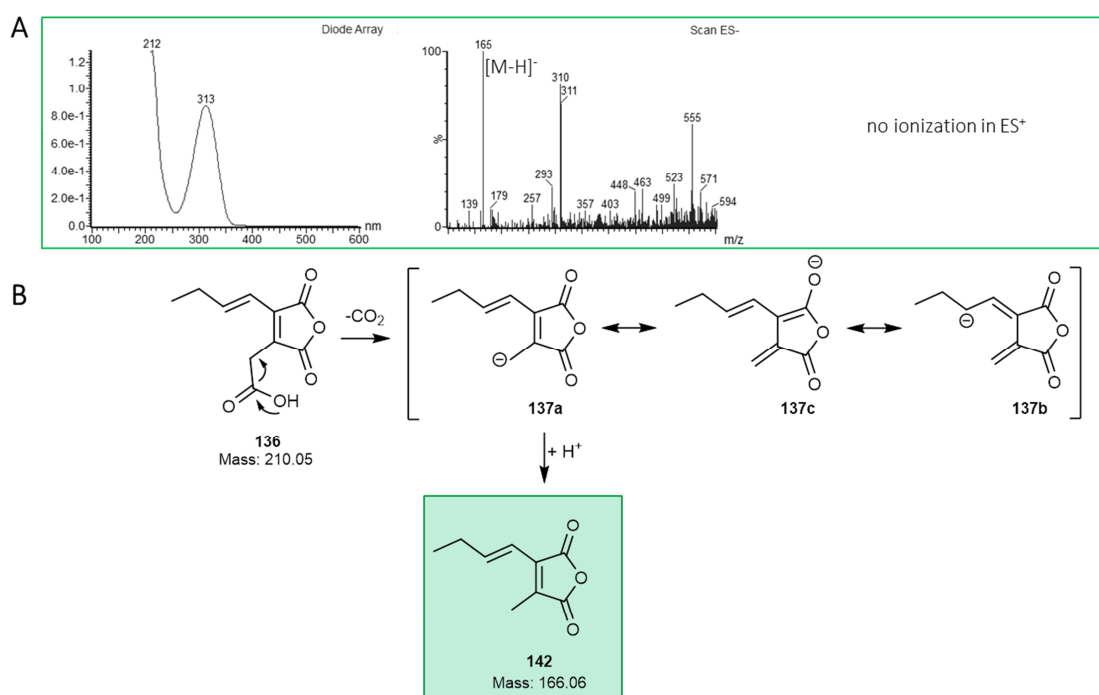
Figure 3.10 Characterisation of the peak at 7.7 min: **A**, LCMS spectra; **B**, proposed mechanism.

Furthermore, 12 of the 17 transformants produced not only the expected compounds **136** and **142**, but also showed another new compound **143** peak at 6.5 min. The intensity of peaks for compound **143** was only a fraction of the maleic acid anhydride **136** peak. The characteristic UV at 311 nm (Figure 3.11) indicated a relationship to the maleic acid anhydride **136** and **142**. The mass was speculated to be 224 Da by ES$^+$/ES$^-$ spectra with peaks at a *m/z* of 447 ([2 M - H]$^-$), 179 ([M - CO_2 - H]$^-$) and 225 ([M + H]$^+$).
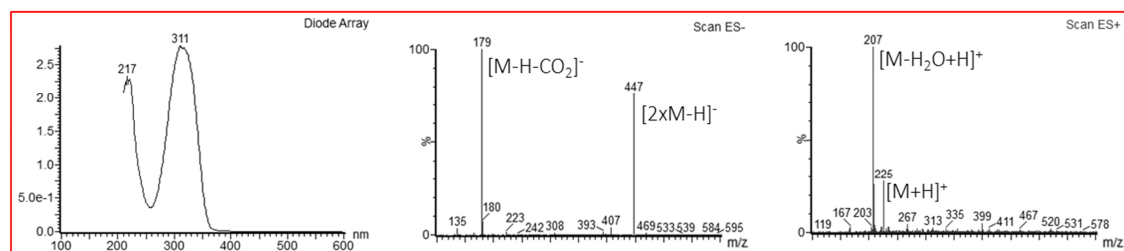
Figure 3.11 Characterisation of the peak **143** at 6.5 min.

For the identification of compound **143** the transformant was cultivated at higher scale (1.9 litre) and **143** was isolated by preparative LCMS. Overall, approx. 1.4 mg/L were obtained. The structure of **143** was elucidated by NMR analysis (Figure 3.12, Figure 3.13, Table 3-3) and assigned to the new peak (details section 7.4.1.1). Additional 2D-NMR data is summarized in section 7.4.1.1.
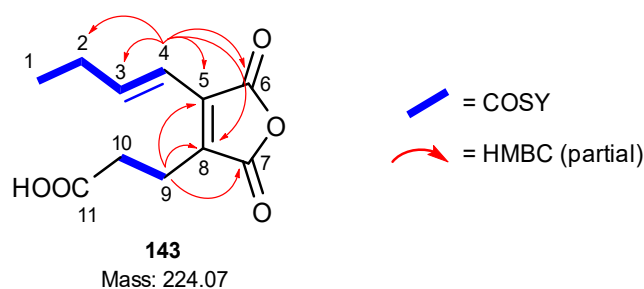
Figure 3.12 Identified structure **143**.

Table 3-3 Chemical shifts of compound **143** in CDCl₃ (600 MHz).

Position	C-type	δ_C/ppm	δ_H/ppm (mult, J in Hz)	COSY	HMBC	literature[166]	
						δ_C/ppm	δ_H/ppm
1	CH₃	12.6	1.14 (t, 7.42)	2	2, 3	12.5	1.11
2	CH₂	27.6	2.35 (p, 7.36)	1, 3, 4	1, 3, 4	19.2	2.35
3	CH	150.5	7.31 (m)*	2, 4	1, 2, 5	116.4	7.27
4	CH	116.3	6.35 (d, 15.89)	2, 3	2, 3, 5, 6, 8	150.2	6.31
5	C	138.5				165.9 or 164.3 or 138.4 or 136.3	
6	C	165.9					
7	C	164.2					
8	C	136.1					
9	CH₂	19.3	2.85 (t, 7.25)	10	5, 7, 8, 10, 11	31.3 or 27.5	2.79
10	CH₂	31.0	2.76 (t, 7.25)	9	8, 9, 11		2.79
11	C	176.4				-	

* overlapped by solvent peak
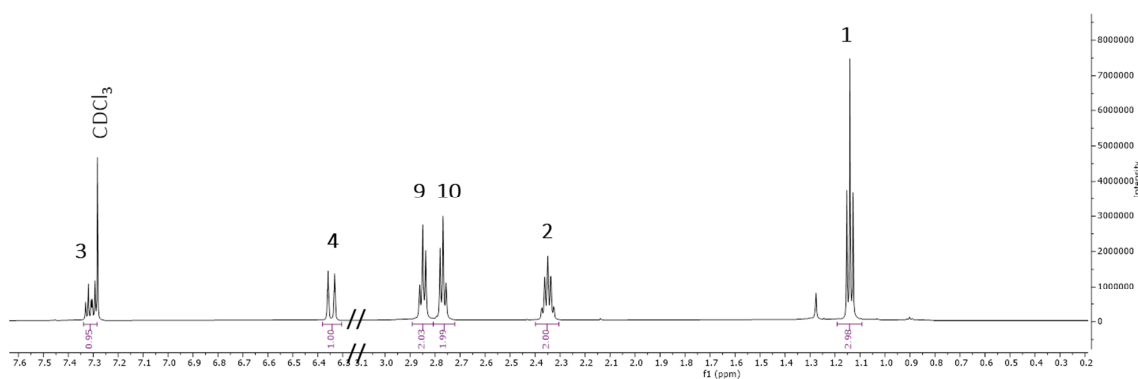
Figure 3.13 ¹H NMR (600 MHz) of compound **143** in CDCl₃.

In comparison to the usual maleic acid anhydride **136** this structure **143** appears to contain a non-polyketide carbon chain extended by one carbon. Cordyanhydride B **61** contains a maleic acid anhydride monomer containing a longer carbon chain (Scheme 3.3), in addition to two maleic acid anhydride monomers. Therefore, structure **143** is a conceivable precursor for the biosynthesis of cordyanhydride B **61**.
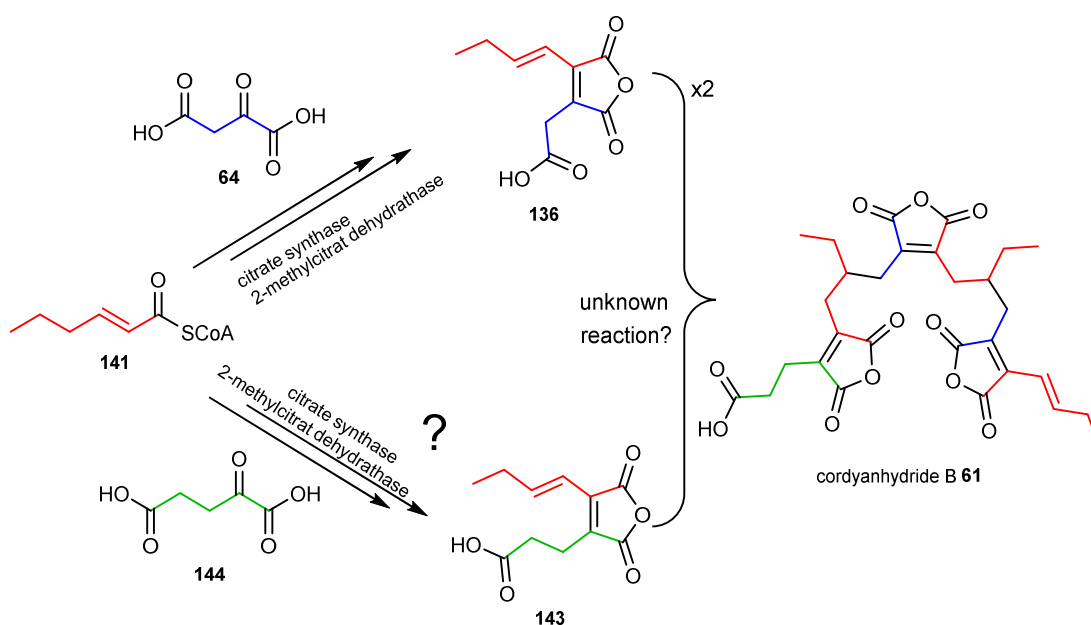
Scheme 3.3 Hypothetical biosynthesis of cordyanhydride B **61**.

The biosynthetic origin of **143** is unknown. It was hypothesized that the acylcitrate synthase LwmR4 might accept both oxaloacetate **64** and α-ketoglutaric acid **144** as substrates (Scheme 3.3). Both are primary metabolites and available in *A. oryzae*. The ring closing reaction is then enabled by the 2-methycitrate dehydratase to form compound **143**. This hypothesis will be investigated in section 3.3.4 by *in vitro* studies.

3.3.4 Alkylcitrate Synthase *in vitro* Studies

3.3.4.1 Recombinant Protein Expression Vector Construction

The substrate selectivity of the *H. lienhwacheense* alkylcitrate synthase was tested *in vitro*. Before the construction of a protein expression vector appropriate for heterologous expression in *E. coli*, the citrate synthase sequence was characterized using bioinformatic tools. Analysis with TMHMM-2.0[167] predicted no transmembrane regions. Analysis with the SignalP 5.0 Server[168] indicated no signal peptide in the protein sequence. The Expasy ProtParam analysis tool[169] was used to predict the expected theoretical weight (49250.53 Da), the *pI* (9.01), aliphatic index (88.46) and instability index (42.47). Overall, bioinformatic analysis was promising for successful isolation of the soluble protein after expression in *E. coli*.

The DNA sequence of the *H. lienhwacheense* alkylcitrate synthase was codon optimized for *E. coli* (Thermo Fisher Scientific) and purchased as a synthetic fragment (Twist Bioscience). The amplification took place with primer overhangs appropriate for the restriction enzymes used for vector digestion (*NdeI* and *NotI*). The PCR product was digested with these restriction enzymes and inserted into the pET28a(+) vector (Figure 3.14 A) by ligation according to the manufacturer's instruction (section 6.1.4.3).
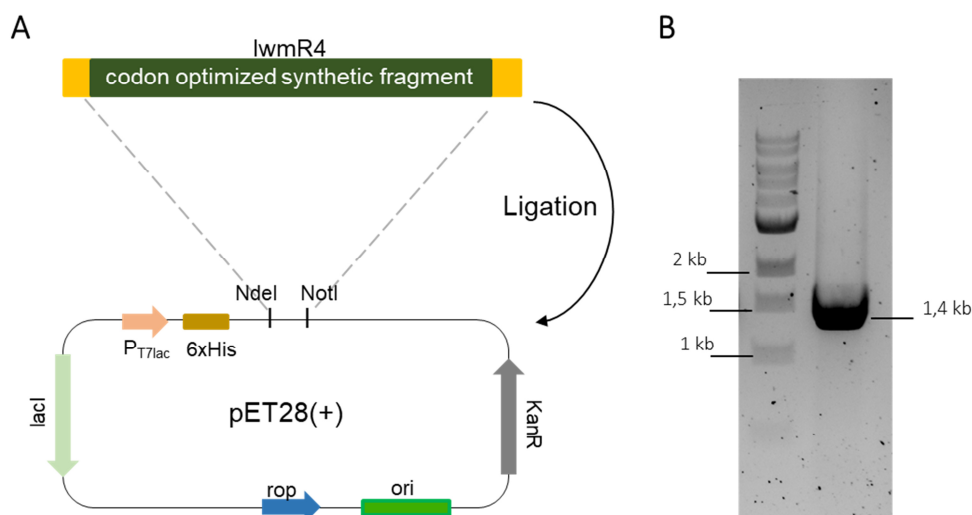
Figure 3.14 Vector construction of pET28(+)-lwmR4: **A**, schmatic overview of ligation strategy; **B**, PCR amplification of lwmR4.

The correct integration of the gene was verified by PCR amplification (Figure 3.14 B) and sequence analysis. The vector was then transformed into *E.coli* BL21.

3.3.4.2 Recombinant Protein Isolation

E. coli BL21 was grown in a 10 ml preculture over night at 37 °C and used to inoculate 1 L LB-media, followed by induction with IPTG at an OD_{600} of 0.6. After induction, the temperature was set to 16 °C and the culture was incubated for approx. 20 h. The cells were harvested and disrupted by sonication.

The recombinant protein contains a histidine-tag, which enables the purification *via* immobilized metal ion affinity chromatography (Ni-NTA). Therefore, the supernatant was mixed with the Ni-NTA resin. The resin was placed in an empty column with inserted PE filter element. The proteins were washed and eluted in multiple steps with appropriate phosphate buffers containing increasing imidazole

concentrations (0 mM – 500 mM, section 6.1.4.3). The resulting fractions were collected and analysed. The SDS PAGE of the protein purification process (Figure 3.15) shows the protein in the elution buffer 2 fraction with a mass of approx. 49 kDa. This fraction was used to concentrate the enzyme to a concentration of approx. 20 µg/ml.
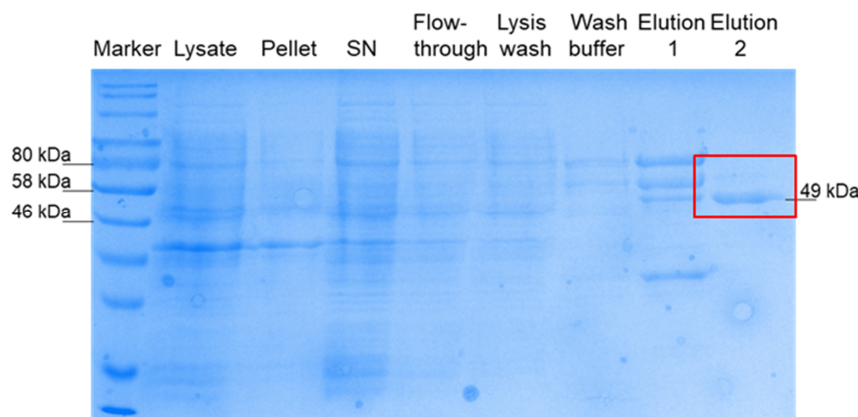
Figure 3.15 SDS PAGE of the purification process of *H. liehwacheense* alkylcitrate synthase from *E.coli* BL21 cells (SN = supernatent).

3.3.4.3 Alkylcitrate Synthase Assay with Hexanoyl-CoA

The first substrate tested for the reaction of the *H. lienwacheense* alkylcitrate synthase was hexanoyl-CoA **145** (see section 3.3.3), which mimics the natural tetraketide product, obtained by previous group members. The expected products after the coupling reaction are the corresponding alkylcitrates **146** and **147** (Figure 3.16 A) depending on the second substrate oxaloacetic acid **64** or α-ketoglutaric acid **144**.

The assay was performed with 1 mM hexanoyl-CoA **145** and 1 mM oxaloacetic acid **64** or α-ketoglutaric acid **144**, with 4 µM *H. lienwacheense* alkylcitrate synthase in 100 µl 50 mM sodium phosphate buffer. In a control reaction mixture, no alkylcitrate synthase was added to the assay (Figure 3.16 B). The assays were incubated at 30 °C for 2 h and stopped by adding 100 µl CH₃CN. The reaction mixtures were mixed thoroughly and centrifuged at maximum speed to pellet the protein. Afterwards, the supernatant of each reaction mixture was analysed directly by LCMS.
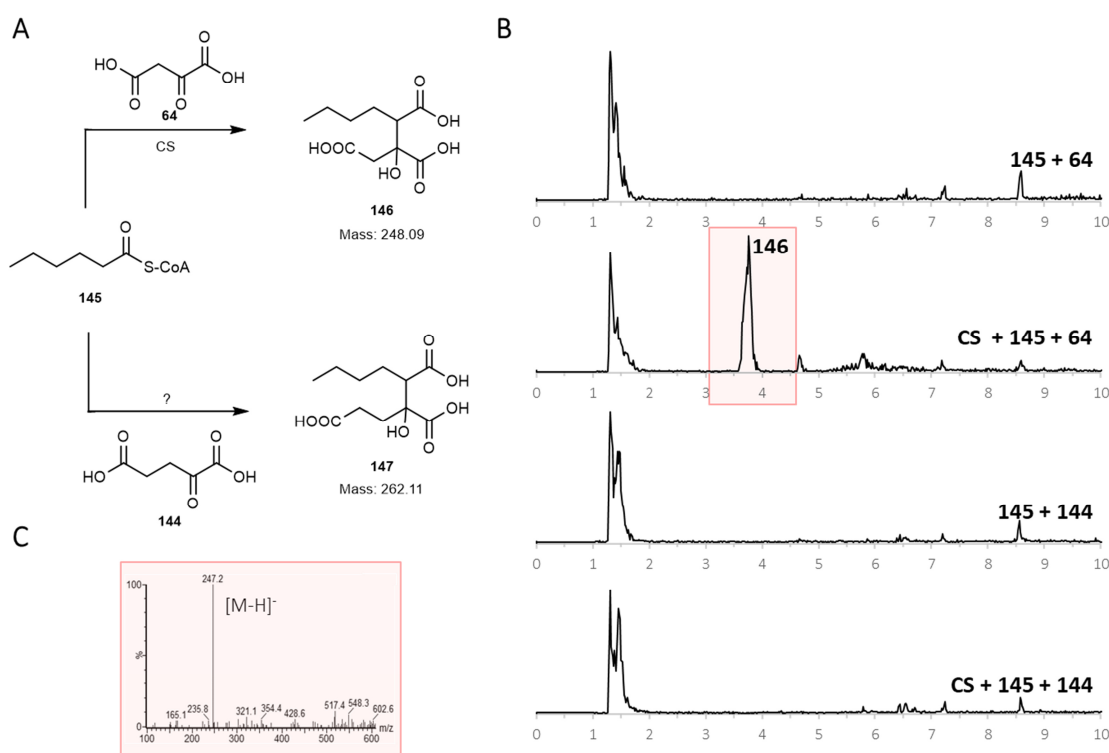
Figure 3.16 *In vitro* assay with *H. lienhwacheense* alkylcitrate synthase and hexanoyl-CoA: **A**, expected reaction of citrate synthase with hanoyl-CoA; **B**, ES⁻ chromatogram of *in vitro* assay with citrate synthase; **C**, ES⁻ spectra of the peak at 3.7 min.

The alkylcitrate synthase catalysed a reaction between oxaloacetic acid **64** and hexanoyl-CoA **145**, which is indicated by the peak at 3.7 min (Figure 3.16 B) with an associated *m/z* of 247 in ES⁻ spectra (Figure 3.16 C), as expected. For α-ketoglutaric acid **144** no reaction took place, as no new peaks were observed. This finding contradicts the hypothesis that LwmR4 is able to use α-ketoglutaric acid **144** and hexanoyl-CoA **145** for the formation of alkylcitrates like **147**.

3.3.4.4 Alkylcitrate Synthase Assay with *E*-hex-2-enoyl-CoA

Hexanoyl-CoA **145** is not the natural polyketide substrate of the citrate synthase according to previous work (Scheme 3.1). Therefore, the unsaturated natural substrate, *E*-hex-2-enoyl CoA **134** (Figure 3.17 A) was tested as a substrate for the citrate synthase assays under the same conditions. It was synthesized from *E*-hex-2-enoic acid **140** *via* **148** using methods developed by Dr. Steffen Friedrich (Scheme 3.4).[160,170]

Scheme 3.4 Synthesis of hexenoyl-CoA as a substrate for citrate synthase assays.
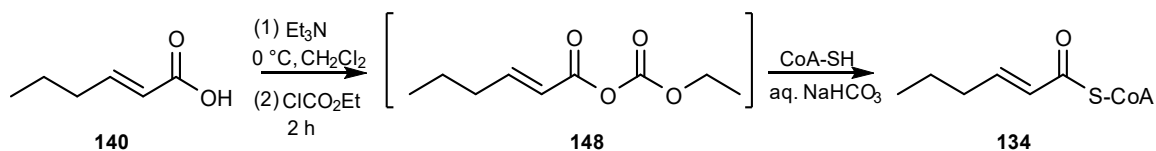
Oxaloacetic acid **64** and α-ketoglutaric acid **144** were tested for the reaction with **134** with previously described methods. The expected products are the unsaturated alkylcitrates **135** and **149** (Figure 3.17 A) for a reaction with oxaloacetic acid **64** and α-ketoglutaric acid **144**, respectively.

The formation of **135** was detected by the extracted ion chromatogram for the search of the fragment ion with an *m/z* of 245, according to the mass of **135** (Figure 3.17 B). With the search for the fragment ion with an *m/z* of 259 for **149** no peak was observed. The assay with hexenoyl-CoA confirmed the previous results (section 3.3.4.3), as no reaction took place to form the expected product **149** (Figure 3.17 B) with α-ketoglutaric acid **144** as a substrate, but only with oxaloacetic acid **64** (Figure 3.17 B, C). The same result was confirmed with a reaction with an incubation time of > 16 h.
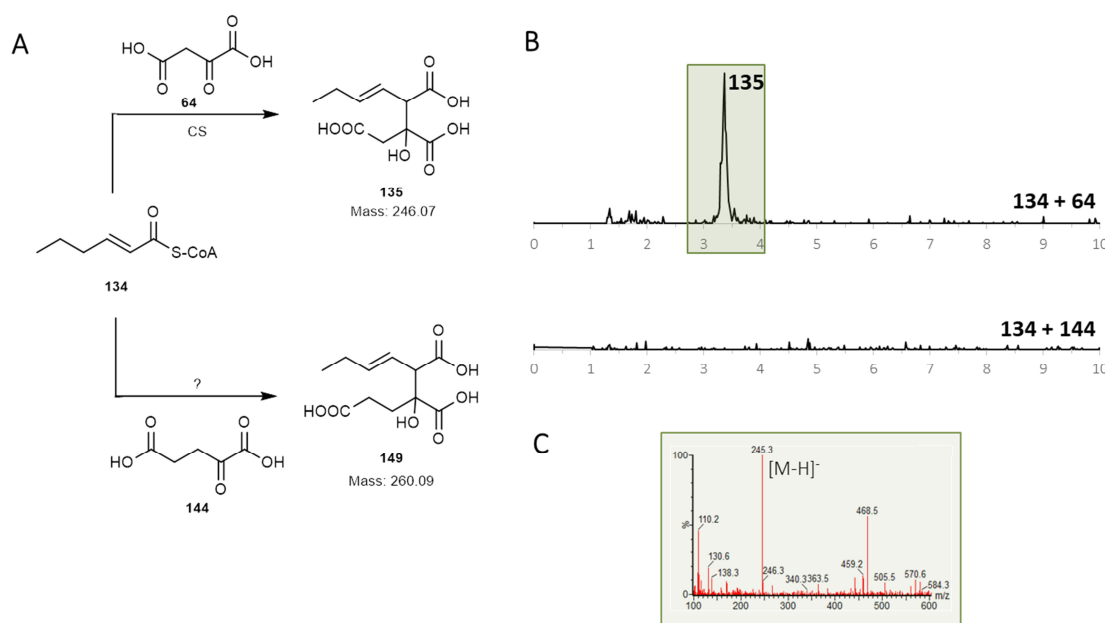
Figure 3.17 *In vitro* assay results with *H. lienhwacheense* alkylcitrate synthase and hexenoyl-CoA: **A**, Expected reaction of citrate synthase with henoyl-CoA: **B**, ES⁻ chromatogram of *in vitro* assay with citrate synthase with EIC for expected product; **C**, ES⁻ spectra of the peak at 3.2 min.

In conclusion, α-ketoglutaric acid **144** is not a substrate for the *H. lienhwacheense* alkylcitrate synthase, but the enzyme accepts different acyl CoAs **134** and **145**. This leaves the question of how compound **143** (section 3.3.3) is formed unanswered.

3.3.5 Studies on the Biosynthesis of Cordyanhydride B

Maleic acid anhydrides (like **136**) are not only incorporated into lienhwalide structures found in *H. lienhwacheense*, but also in the trimeric cordyanhydride B **61**. Cordyanhydride B **61** consists of three maleidride monomers linked linearly. The monomers are apparently derived from two equivalents of **136** and one equivalent of **143**. It is not known how the monomers are linked together. However, in the case of the maleidrides like *e.g.* byssochlamic acid **48**, two monomers **136** are cyclised to form the distinctive 9-membered ring (section 3.1.3.2). Previous work in the Cox group has shown that this is probably achieved by the MDC enzyme, although the precise mechanism is still unknown. Early *in vivo* work showed that the often-associated PEBP protein may be required, but later *in vitro* work showed this is probably not the case and PEBP proteins were speculated to represent a chaperone-like function *in vivo* (section 3.1.3.2).[71,158,160,162] The *lwm*-BGC of *H. lienhwacheense* encodes an MDC, but no PEBP (Figure 3.6). Therefore, it was hypothesised that the putative MDC LwmR2 may be responsible for the linking reactions to form cordyanhydride B **61**.

3.3.5.1 Heterologous Co-Expression of Core Genes + LwmR2 (+R5)

The gene *lwmR2* was added to the heterologous expression system (primer KI1/KI2) using vector pTY-*argB-lwmR2-lwmR6*, obtained by previously described methods. The pTY-*argB-lwmR2-lwmR6* vector was co-transformed with the previously achieved vector pTY-*ade-lwmA-lwmR1-lwmR3-lwmR4,* leading to the potential co-expression of the PKS LwmA, hydrolase LwmR1, acyl-CoA-ligase LwmR6, alkylcitrate synthase LwmR4, 2-methyl alkylcitrate synthase LwmR3 and the putative MDC LwmR2.

After the analysis of > 20 maleic acid anhydride **136** producing transformants including the appropriate genes, no new compound was observed (Figure 3.18), suggesting that the MDC is not (or not alone) responsible for the coupling reactions to form cordyanhydride B. However, other obstacles to an active enzyme cannot be ruled

out, such as an incorrect intron/exon prediction or a problem with compartmentation *in vivo*, as the exact organelle target and reaction mechanism of the MDC is unknown.
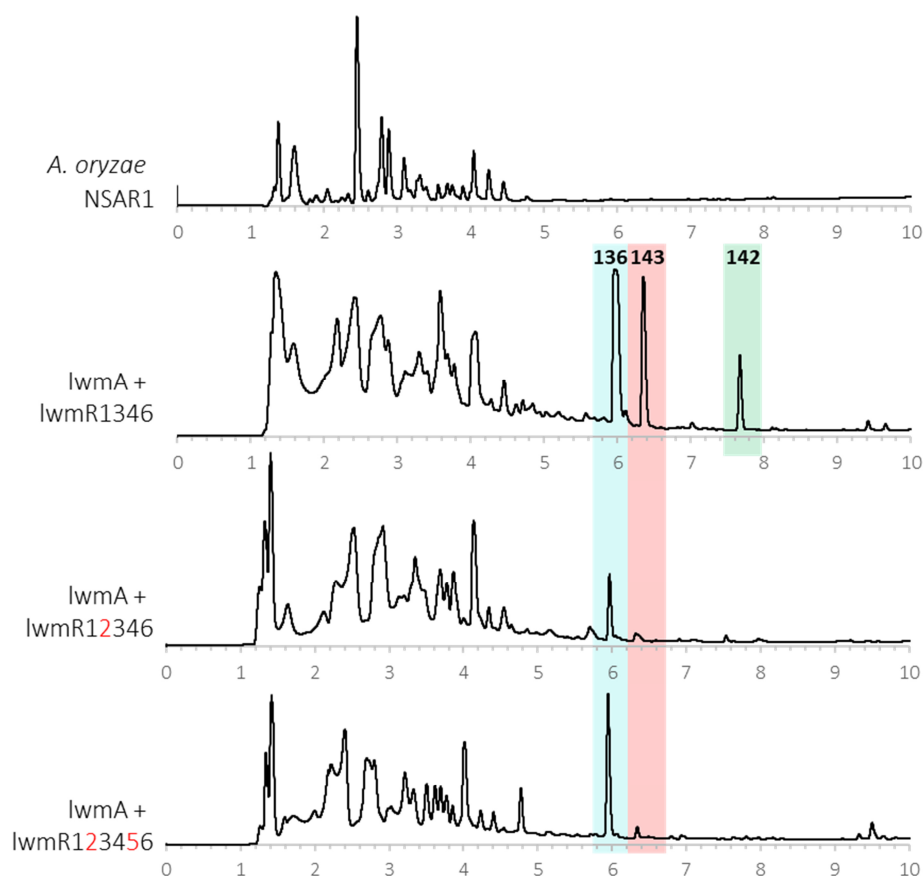
Figure 3.18 DAD chromatograms of heterologous co-expression of core enzymes + *lwmR2* (+ *lwmR5*).

Since a second enzyme, the mysterious PEBP protein, is known to be involved in the biosynthesis of maleidride dimers in related systems (section 3.1.3.2), it was speculated that the MDC LwmR2 may be effective in co-expression with a helper protein. The role of a helper protein could either involve a chaperon-like function by working together with the MDC or be part of a resistance mechanism to circumvent cell-death induced by resulting toxic metabolites of the pathway. However, no analogue to PEBP was found in the BGC (Figure 3.6) or in other loci of the complete *H. lienhwacheense* genome. When considering the BGC, only two putative genes (*lwmR5* and *lwmR7*) were not used in previous heterologous expression experiments. The gene *lwmR7* was hypothesized to encode a hydrolase. The function of the putative gene *lwmR5* is unknown and therefore might be involved in the dimerization.

The gene *lwmR5* was added to the expression system (primer R5gpdA_fw/ R5eno_rev), resulting in vector pTY-*argB-lwmR2-lwmR6-lwmR5*. The vector pTY-*argB-lwmR2-lwmR6-lwmR5* was again co-transformed with the previously obtained vector pTY-*ade-lwmA-lwmR1-lwmR3-lwmR*. The co-expression of the two vectors includes the complete BGC expect for the putative hydrolase LwmR7. The introduction of LwmR2 and LwmR5 in co-expression with the known genes for maleic acid anhydride production did not lead to the biosynthesis of new metabolites (Figure 3.18) for all five analysed transformants. As a result, no catalytic role was assigned the putative MDC encoding gene in the *lwm*-BGC or the unknown putative protein LwmR5 by heterologous expression.

3.3.5.2 Recombinant Expression of LwmR2

Apparent lack of activity of LwmR2 (encoding the putative MDC) *in vivo* led to considering activity assays *in vitro*. Therefore, pET28a(+)-*lwmR2* was prepared for protein expression in *E. coli* BL21 to test the reaction of the putative MDC *in vitro*. The protein sequence was analysed using the SignalP 5.0[168] website for analysis of signal peptides and cleavage sites in eukaryotes. An *N*-terminal signal peptide of 19 amino acids was predicted (Figure 3.19).

Figure 3.19 Prediction of LwmR2 protein sequence for signal peptides (SP = secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I, CS = cleavage site, other = the probability that the sequence does not have any kind of signal peptide).

The protein expression vector pET28(+)-*lwmR2* was designed with appropriate primer (KI-fw-pET28a/KI-rev-pET28a) to exclude the predicted signal peptide from the sequence and to insert the restriction sites for the digestion of the fragment after PCR (Figure 3.20 A). After successful construction, the transformation into *E. coli* BL21 was carried out. The cells were cultivated as described before. However, it was not possible to obtain soluble protein after successful induction of the expression (Figure 3.20 B).
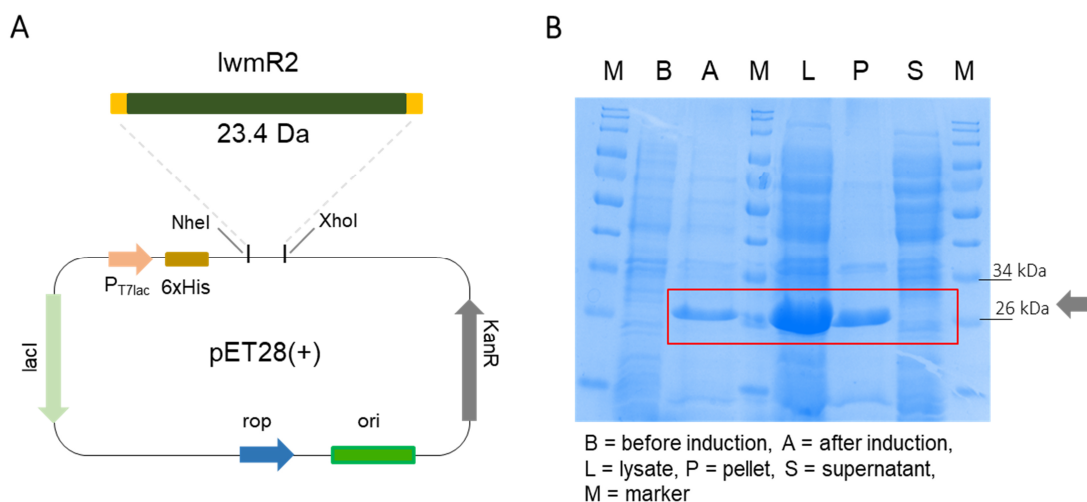
Figure 3.20 Preparation of LwmR2: **A**, vector construction of pET28(+)a-*lwmA*; **B**, SDS-PAGE of the purification process of LwmR2 from *E.coli* BL21 cells.

Consequently, it was not possible to set up *in vitro* experiments to investigate the putative reaction of LwmR2 with compounds **136** and/or **143**. It was not possible to investigate the biosynthesis of cordyanhydride further.

3.4 Discussion: Maleic Anhydride Pathway

In summary, **136**, **142** and **143** were the only maleic acid anhydride related compounds found in the *in vivo* studies of the *lwm* pathway (Table 3-4 A-D). No other coupled or related compounds were found. For *in vitro* studies, the intermediate **135** was detected (Table 3-4 E) as a product of LwmR4.

Table 3-4 Summary of *lwm* –BGC pathway *in vivo* and *in vitro* experiments.

	lwm genes							detected compounds		
	A	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**136**	**142**	**143**
A	✓	✓	-	✓	✓	-	-	✓	-	-
B	✓	✓	-	✓	✓	-	✓	✓	✓	✓
C	✓	✓	✓	✓	✓	-	✓	✓	-	✓
D	✓	✓	✓	✓	✓	✓	✓	✓	-	✓
E (*in vitro*)				✓					**135**	

3.4.1 Early Steps of Maleic Acid Anhydride Biosynthesis

The early steps of the biosynthetic pathway (Scheme 3.1) were confirmed by the heterologous expression of the putative genes *lwmA*, *lwmR1*, *lwmR3* and *lwmR4* in *A. oryzae* NSAR1 by the formation of **136** and **142**. The formation of **135** in *in vitro* assays of the alkylcitrate synthase are in agreement with these results. The PKS, hydrolase, citrate synthase and 2-methycitrate dehydratase are the core enzymes for the biosynthesis of maleic acid anhydrides **136** and are highly homologous to enzymes of related maleidride pathways (Figure 3.21).

The yields of maleic acid anhydride monomers were improved by adding the acyl-CoA ligase LwmR6 to the heterologous expression system. In comparison to other maleidride producing clusters, only the rubratoxin cluster in *Talaromyces stipitatus* contains a homologue of the gene.[171] An improvement of the production of maleidrides by the acyl-CoA ligase was not investigated in this species. However, in case of the previous *bf*-BGC heterologous expression was successful despite the lack of an analogous activating enzyme.[158] Likewise, the successful production of **136** in experiments without co-expression of the *lwmR6* gene is in agreement with this finding. This is explained by the presence of native fatty acid CoA ligases in the host fungus or other analogous enzymes. For example, a blastx search of *lwmR6* against the *A. oryzae*

genome reveals eight hits, the highest score indicated by 94 % query cover and 50 % identity. The natural activity of acyl-CoA ligases of primary metabolism from the heterologous host are therefore likely to be sufficient to form **134** for further reactions. Nevertheless, the activity of LwmR6 is speculated to be higher than native background activities, reasoned by the up-regulated expression of the heterologous genes by the selected promotors.

Six other related BGCs are known to be directly responsible for the production of maleidrides in other organisms. These compounds are byssochlamic acid **48**/ agnestadrides **54-55**,[158] rubratoxins **56**,[171] cornexistin **49**,[71] zopfiellin **52**[157,162] (two homolog clusters) and scytalidin **150**.[70,157] The gene clusters for the biosynthesis of these compounds (except rubratoxin BGC that is not available) were compared to the *lwm*-BGC (Figure 3.21). All clusters share a high homology between the core enzyme set while differing in other additional putative genes, indicating that the first biosynthetic steps for maleic acid anhydride formation are similar in all species.
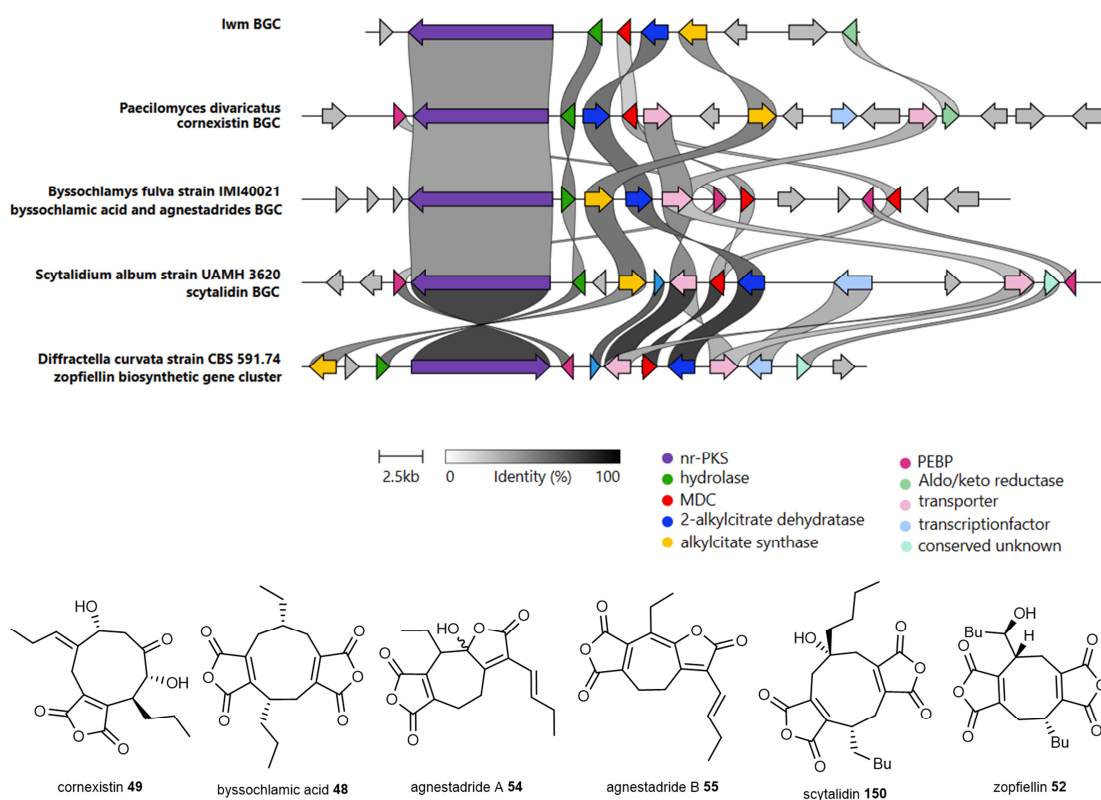
Figure 3.21 Gene Cluster comparison between maleidride producing clusters and their products.

3.4.2 Biosynthesis of Compound 143

The co-expression of the catalytic core genes together with the acyl-CoA ligase LwmR6 showed that the same set of enzymes synthetizes not only the maleic acid anhydride **136**, but also compound **143** *in vivo*. Compound **143** was first discovered in 1980 in extracts of *Paecilomyces variotii*.[166] While the biosynthetic origin is unknown, it was excluded that the citrate synthase is responsible for a coupling reaction to form the alkylcitrates with α-ketoglutaric acid **144** under the tested conditions. The *in vitro* assays showed that the alkylcitrate synthase LwmR4 is able to accept hexanoyl-CoA **145** or *E*-hex-2-enoyl-CoA **134** together with oxaloacetic acid **64**, but not α-ketoglutaric **144**. As compound **143** was only detected *in vivo*, it is possible that in *H. lienhwacheense* as well as in *A. oryzae* an unknown enzyme or mechanism plays a role in the formation of **143**.

3.4.3 Biosynthesis of Cordyanhydride B

It was speculated that compound **143** is a precursor for the biosynthesis of cordyanhydride B **61**. Cordyanhydride B **61** and/or A **60** was reported in extracts of *Cordyceps pseudomilitaris BCC 1620*,[77,172] *Paecilomyces tenuipes*[173] and *Taleromyces stipitatus*.[174] In all cases, the biosynthetic pathway of the precursor as well as the linear linked compounds is unknown. This also applies to *H. lienhwacheense*, where the biosynthetic origin of **61** remains a mystery. Co-expression of the putative MDC LwmR2 did not (alone or in co-expression with the adjacent putative protein LwmR5) synthetize any new coupled compounds *in vivo*, therefore leaving the question of the coupling reaction to form cordyanhydride B **61** open. The investigation of the MDC LwmR2 reaction was not possible *in vitro* due to unsuccessful soluble protein production in *E. coli,* which is in agreement with previous *E.coli* MDC expression experiments. Cox and co-workers attempted to express the MDCs BfL6 and BfL10 in *E. coli*, but were unable to obtain active protein. However, the use of yest cell-free extracts were used for *in vitro* experiments with MDCs and PEBP-like proteins.[160]

Furthermore, co-expression of LwmR2 (MDC) with the core enzymes *in vivo* did not show the production of any related maleidride-like cyclic dimers (*e.g.* byssochlamic acid **48** or agnestadrides **54-55**), despite an identity value of 36 % - 38% to other known MDCs (ScyR6, PvL3, BfL10, EpiR6) based on blastx comparison.[121,151]

Therefore, it can be speculated that the MDC of the *lwm*-BGC is either inactive or less active, or would need an assistant protein analogous to PEBP. It was excluded, that the putative gene product LwmR5 in combination with the MDC in heterologous expression experiments leads to any new, coupled compounds.

In comparison to other maleidride forming clusters, a PEBP protein homologue is missing from the *lwm*-BGC, and no homolog was found in the complete *H. lienhwacheense* genome. Taking the absence of the PEBP and the unsuccessful synthesis of cordyanhydride B into account, it indicates that the linear coupling of maleic acid anhydrides is not closely related to dimerization in the cases of maleidrides.

A spontaneous coupling reaction of maleic acid anhydride monomers was never observed during handling of the pure compounds or liquid culture extracts, which is in agreement with previous attempts of achieving the dimerized products by synthetic methods. Sutherland and co-workers aimed for cyclic maleidride products. It was possible to generate epimers of byssochlamic acid **48**, but not the natural compounds by a coupling reaction of **136**.[175,176] The hurdle in biomimetic studies was speculated to be caused by the instability of monomer compounds, which could also be the case in the studies of this work. The coupling reaction remains a mystery.

3.5 Conclusion

The *lwm*-BGC was discovered by the genome analysis of *H. lienhwacheense*. Homologs for the known core genes for maleidride biosynthesis were found. The investigation of the *lwm*-BGC confirmed that the biosynthetic origin of the maleic acid anhydride compounds is analogous to other related fungal maleidrides (Scheme 3.5 A). For further studies (section 5), the core set used was defined with overall five enzymes: LwmA (PKS), LwmR1 (hydrolase), LwmR3 (dehydratase), LwmR4 (alkylcitrate synthase) as the catalytic core-enzymes for the formation of **136**, and LwmR6 (acyl-CoA ligase) to increase the titre of the metabolites.
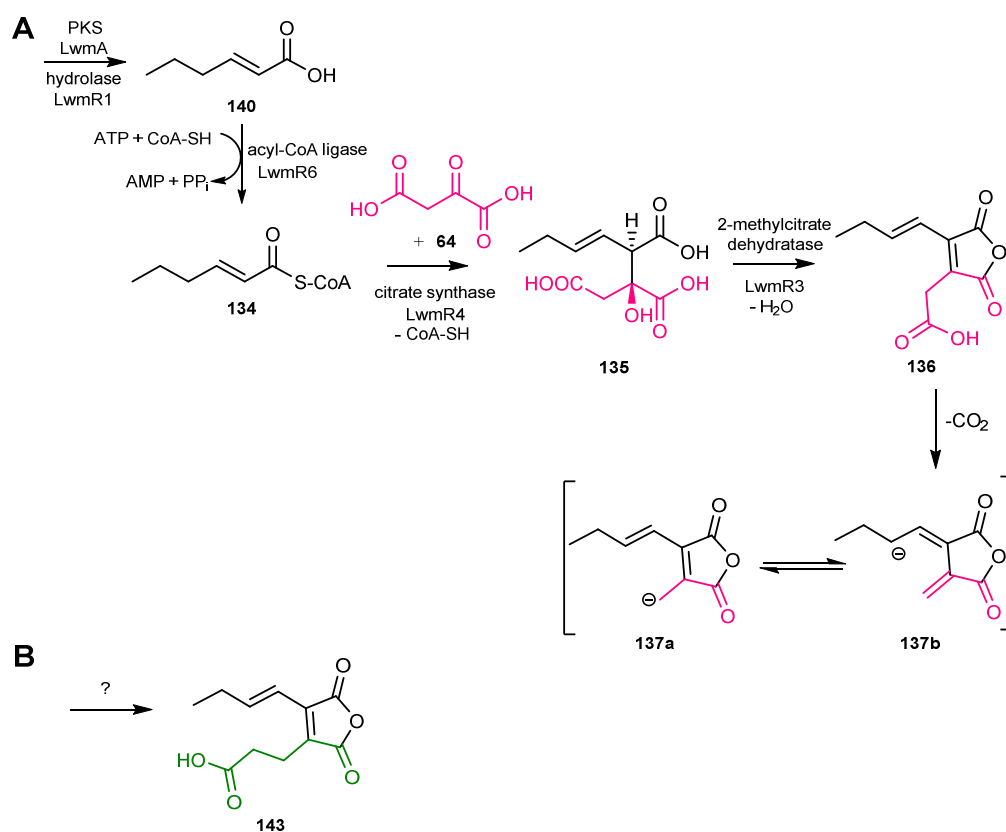
Scheme 3.5 Biosynthetic pathway of maleic acid anhydrides in *H. lienhwacheense*: **A**, proposed biosynthethic pathway; **B**, unknown origin of **143**.

The putative intermediate **143** (Scheme 3.5 B) for cordyanhydride B **61** biosynthesis was detected, isolated, and structurally analysed. It includes two additional carbons in the non-polyketide chain of maleic acid anhydride compounds in comparison to the known maleic acid anhydride **136**. However, the origin of **143** remains a mystery as *in vitro* assays with the alkylcitrate synthase did not identify the substrate for the formation

of **143**. The investigations of the biosynthetic pathway of compound **143** will possibly reveal a new and unexpected origin of the compound in the future. Isotopic labelling studies are an appropriate follow-up experiment to investigate the unknown pathway.

The discovery of **143** *in vivo* represents the first step towards the elucidation of the mysterious origin of cordyanhydride B **61**. The heterologous biosynthesis of cordyanhydride B **61** was not possible. The inclusion of the MDC LwmR2, which was speculated to be responsible for the coupling reaction, in heterologous expression experiments did not lead to coupled products. For the elucidation of the biosynthesis of cordyanhydride B **61**, extensive analysis of the genome of *H. lienhwacheense* will be required.

4 Tropolone Biosynthesis in *H. lienhwacheense*

4.1 Introduction

Lienhwalides **130-132** consist of linked alkylcitrate and tropolone moieties. Chapter 3 focussed on the alkyl citrates; this chapter will focus on the tropolone moiety.

4.1.1 Pentaketide Tropolone Pathways in Fungi

The most well-studied fungal tropolones are tetraketides that include *e.g.* stipitatic acid **85** and related compounds (section 1.4.3.2).[84,93] Pentaketide tropolones such as sepedonin **90** are much less studied. However, Gerke *et al.* have recently reported the biosynthesis of sepedonin in *Aspergillus nidulans* using knock-out methodology.[177]

The *dba*-BGC in *A. nidulans* contains genes encoding proteins homologous to TropA, TropB and TropC (Figure 1.13), indicating a biosynthetic pathway similar to tetraketide tropolone biosynthesis.[178] The dba pathway was analysed by knock-out experiments and genes were assigned to the responsible enzymes for the core steps (Scheme 4.1).[177] The first enzyme-free intermediate is the methylated pentaketide aldehyde 2,4-dihydroxy-3-methyl-6-(2-oxopropyl)-benzaldehyde (DHMBA) **156**, which is oxidized by the FMO DbaH to **157**, analogous to the reaction catalysed by TropB from **86** to **87** (Figure 1.13). Gerke *et al.* suggest that the pentaketide undergoes a spontaneous ring-forming reaction, resulting in the hemiacetal **158**. **158** dehydrates to azanidulone **159** as a shunt metabolite. While it was shown that the reaction of **158** to sepedonin **90** requires the catalyst TroA (or AN7893, NHI oxidase), no other enzyme was identified. However, the reaction of **158** to sepedonin **90** must be a redox-neutral reaction. Therefore, a reductive reaction remains unknown in the pathway. **90** dehydrates to anhydrosepedonin **160** which is then oxidized by another FMO (DbaB) to antibiotic C **91**, the final product of the pathway.[177]

The products of the pathway, anhydrosepedonin **160** and antibiotic C **91**, inhibit bacterial growth.[177] Additionally, anhydrosepedonin **160** is known to have antifungal properties.[179]
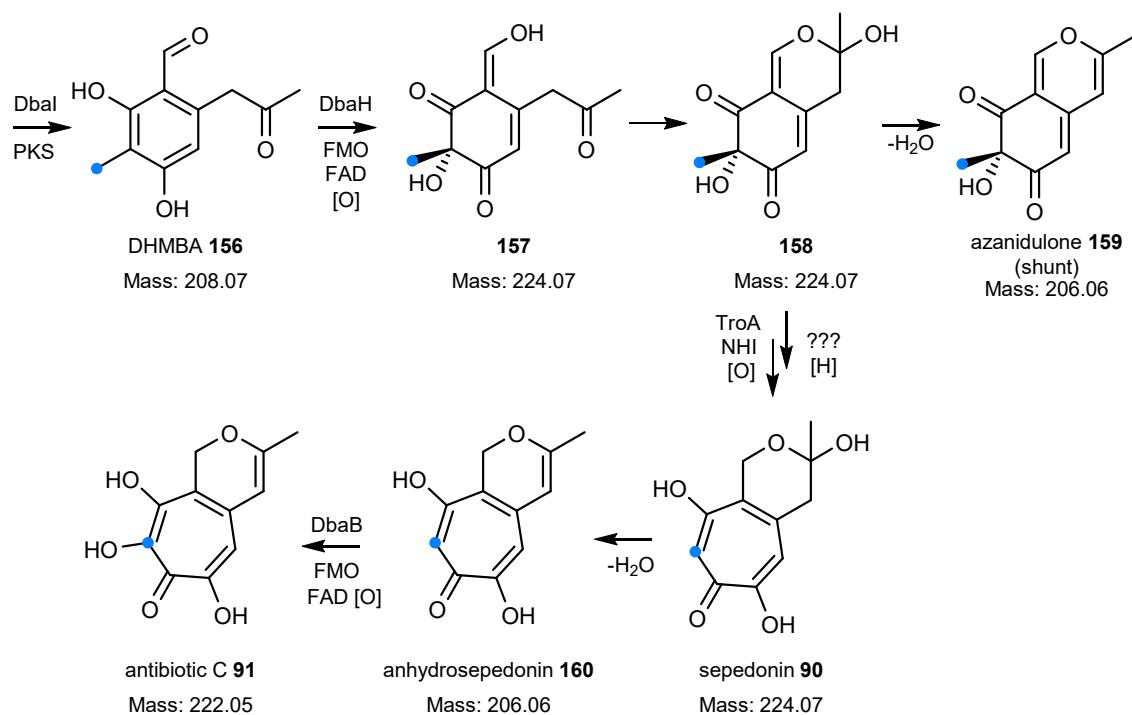
Scheme 4.1 Key steps of antibiotic C biosynthesis in *A. nidulans*.[177]

4.1.2 Synteny Analysis of Pentaketide Tropolone BGCs in previous Projects

Former studies by Kuhnert *et al*. revealed that out of 14 genomes from *Hypoxylaceae* species and the related *Xylaria hypoxylon*, 13 genomes include a putative tropolone BGC, including the genes coding for key enzymes for polyketide synthesis (TropA), oxidation (TropB) and ring expansion (TropC) that form the tropolone core. *H. lienhwacheense* was one of the 14 investigated species. Not only were the core genes highly conserved, but all additional genes encoding putative tailoring proteins, transport proteins and transcription factors were also conserved. Furthermore, all analysed BGCs were highly homologous in regards of the order of the genes and the position of the BGC in the genomes.[149] However, no tropolone-containing compounds were reported from any of the 14 studied species, except *H. lienhwacheense*.[141,149]

4.2 Project Aims

The aim of this section is to understand the biosynthetic origin of the tropolone moiety in the lienhwalides **130-132**. The underlying synteny of the putative tropolone BGCs in the 12 genomes from *Hypoxylaceae* species and the related *Xylaria hypoxylon* suggests an important ecological role of the cluster and the resulting compounds for this group of fungi. Therefore, the *H. lienhwacheense* pathway serves as a model pathway for the analysed species. As described previously (section 3.1.2) extensive genome analysis of *H. lienhwacheense* was carried out by the co-worker Dr. Eric Kuhnert, whose work is the starting point for the investigation of the tropolone pathway.

The previously identified tropolone BGC will be analysed for the putative function of each gene by comparison with related gene clusters. In addition, the BGC will be used as a reference to search bioinformatically for potential tropolon-forming clusters and synteny analysis in 35 additional newly sequenced genomes of *Hypoxylacae* or related species.

For the elucidation of the tropolone pathway, preliminary heterologous expression experiments were carried out by Dr. Jing Feng. The results provided the basis for further investigations, which were achieved by systematic heterologous expression. The investigations here will also include the structural elucidation of shunt metabolites and intermediates of the pathway. Additionally the aim is to investigate the unknown reductive step for the tropolone moiety formation. Heterologous expression will also be used to investigate the late stage modifications of the sepedonin-core scaffold, which could lead to intermediates for lienhwalide B **131** and C **132** formation.

Furthermore, *in vitro* studies will be performed in order to better understand the individual biosynthetic steps of the tropolone pathway and to circumvent the presence of shunt intermediates observed *in vivo*. The results of the *in vivo* and *in vitro* experiments will be summarized to develop a mechanistic pathway for tropolone formation.

4.3 Results

4.3.1 Gene Cluster Analysis

The analysis of the *H. lienhwacheense* genome by Dr. Eric Kuhnert (section 3.1.2) revealed a single putative tropolone gene cluster in *H. lienhwacheense* (*lwt*), which includes putative homologs of the previously described core genes for the tetraketide stipitatic acid **85** BGC (*tropA-C*, section 1.4.3.2). The known core genes for the formation of tropolones (section 1.4.3.2) encode the nr-PKS (*tropA/lwtS,* Figure 4.1, violet), the FAD-dependent monooxygenase (*tropB/lwtR1,* Figure 4.1, pink) and the α-ketoglutaric acid-dependent dioxygenase (*tropC/lwtR7,* Figure 4.1, blue).[93] This set of genes is present in every known tropolone-forming cluster illustrated by the clinker analysis, which was carried out for the previously studied *trop*-BGC (stipitatic acid **85**, *T. stipitatus*),[93] *as*-BGC (xenovulene B **93**, *A. strictum*),[95,180] *dba*-BGC (antibiotic C **91**, *A. nidlans*),[177,178] and *lwt*-BGC (Figure 4.1).
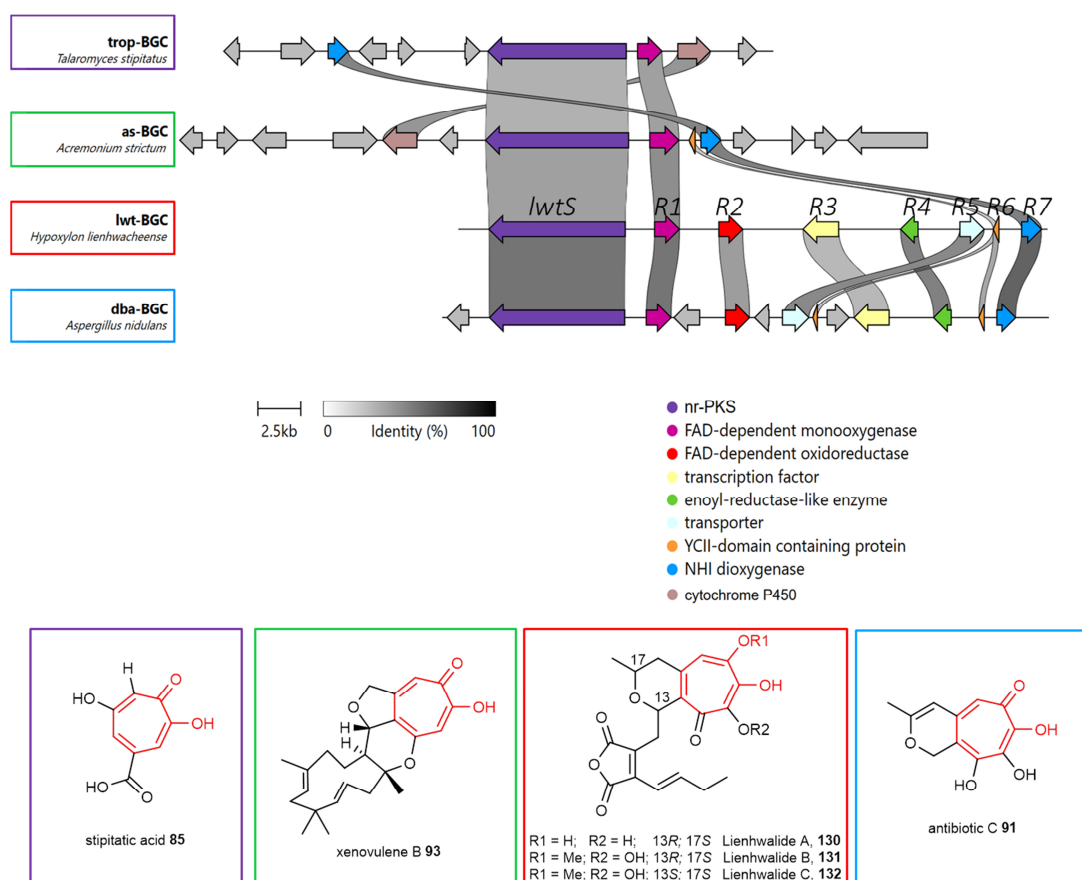
Figure 4.1 Gene cluster analysis of putative tropolone gene cluster in *H. lienhwacheense* in comparison to *dba*-BGC (*A.nidulans*), *as*-BGC cluster (*A. strictum*) and *trop*-BGC (*Talaromyces stipitatus*).[165]

123

Additional to the known genes encoding the three core enzymes, the *dba-*, *lwt-* and *as-*BGCs (Figure 4.1) also include a small putative gene with unknown function encoding a YCII-like domain of only 110 residues (*e.g. lwtR6*, Figure 4.1, orange). In the *dba-*BGC two putative YCII domain encoding genes are present (*AN7894* and *dbaC*). The function of this putative protein is unknown.

The *dba-*BGC was previously investigated by knock-out experiments to elucidate the function of most enzymes of the pathway (section 4.1.1). The *dba-*BGC includes a gene homolog for every putative gene of the *lwt-*BGC (Figure 4.1).[177,178] A putative FAD-dependent oxidoreductase (*lwtR2,* Figure 4.1, red), an enoyl-reductase-like medium chain reductase/dehydrogenase (MDR) enzyme (*lwtR4,* Figure 4.1, green), a transcription factor (*lwtR3*, Figure 4.1, yellow) and a transporter protein (*lwtR5*, Figure 4.1, light blue) are also encoded in the *lwt-*BGC, according to clinker homology analysis. These homologies were confirmed by blastx analysis (Table 4-1).

In comparison to the *dba* cluster, three genes are missing from in the *lwt-*BGC. *DbaG* (encoding a transcription factor), *dbaE* (encoding a serine hydrolase) and *dbaB* (encoding an FAD-dependent monooxygenase) homologs are absent from the *lwt-*BGC. Previous work suggests the function of these proteins has no correlation with the formation of the sepedonin carbon skeleton, as the knock-out of these genes did not lead to any impact on the biosynthesis in *A. nidulans* (Scheme 4.1). DbaB is known to be responsible for the conversion of anhydrosepedonin **160** to antibiotic C **91** (Scheme 4.1).[177]

Table 4-1 Annotation of tropolone gene cluster from *H. lienhwacheense*.

putative gene	length [bp] / [aa]	putative function (blastx)	identity to corresponding Dba protein
lwtS	7833 / 2611	nr-PKS	62 %
lwtR1	1389 / 463	FAD-dependent monooxygenase	62 %
lwtR2	1395 / 465	FAD-dependent oxidoreductase	48 %
lwtR3	1623 / 541	transcription factor	45 %
lwtR4	1023 / 341	enoyl-reductase-like medium chain reductase/dehydrogenase (MDR) enzyme	58 %
lwtR5	1416 / 472	transporter	58 %
lwtR6	330 / 110	YCII-like domain containing protein	47 % (to AN7894 and DbaC each)
lwtR7	1002 / 334	non-heme iron-dependent dioxygenase	69 %

4.3.2 Synteny Analysis of Tropolone BGCs in *Hypoxylacae* Genomes

The analysis of putative sepedonin-forming BGCs was part of a genome-sequencing project, in cooperation with the Stadler group (Helmholtz Institute for Infection Research, Braunschweig) and the Centre for Biotechnology (CeBiTec, Prof. Dr. Jörn Kalinowski, University of Bielefeld). In this new project 35 genomes of *Hypoxylacae* or related species were sequenced and assembled by Dr. Bart Verwaaijen (CeBiTec, University of Bielefeld, detailed list section 7.4.2). The genome sequencing was performed by a combination of Illumina and Oxford nanopore technology. Methods for genome prediction and analysis were similar to those described in section 3.1.2. Gene prediction was performed with GeneMark.[181]

The identification of putative tropolone-forming BGCs was based on a cblaster search[182] of the obtained genomes (section 6.3.3) in combination with the blastx algorithm[151] and Geneious v. 9.1.8 blast searches. The resulting amino acid sequences for all translated proteins of the *lwt*-BGC, excluding the transporter and transcription factor encoding genes (*lwtR3*, *lwtR5*), were used as the references for the search.

The analysis showed that the 35 species can be categorized into three groups. 30 species include one putative tropolone BGC each, with near identical gene composition and order as the *H. lienhwacheense* BGC (Figure 4.2). In all 30 species, only two putative genes were found additionally to the putative genes of the *lwt*-BGC in the corresponding analysed BGCs. In the tropolone BGC of *H. lechatii* a putative transposase gene was identified by blastx (Figure 4.2, grey). In the tropolone BGC of *H. griseobrunneum* an additional hypothetical gene was found, with no putative function assigned (Figure 4.2, grey). Thus, no additional putative genes with catalytic function for the pathway were identified for any of these 30 BGCs.
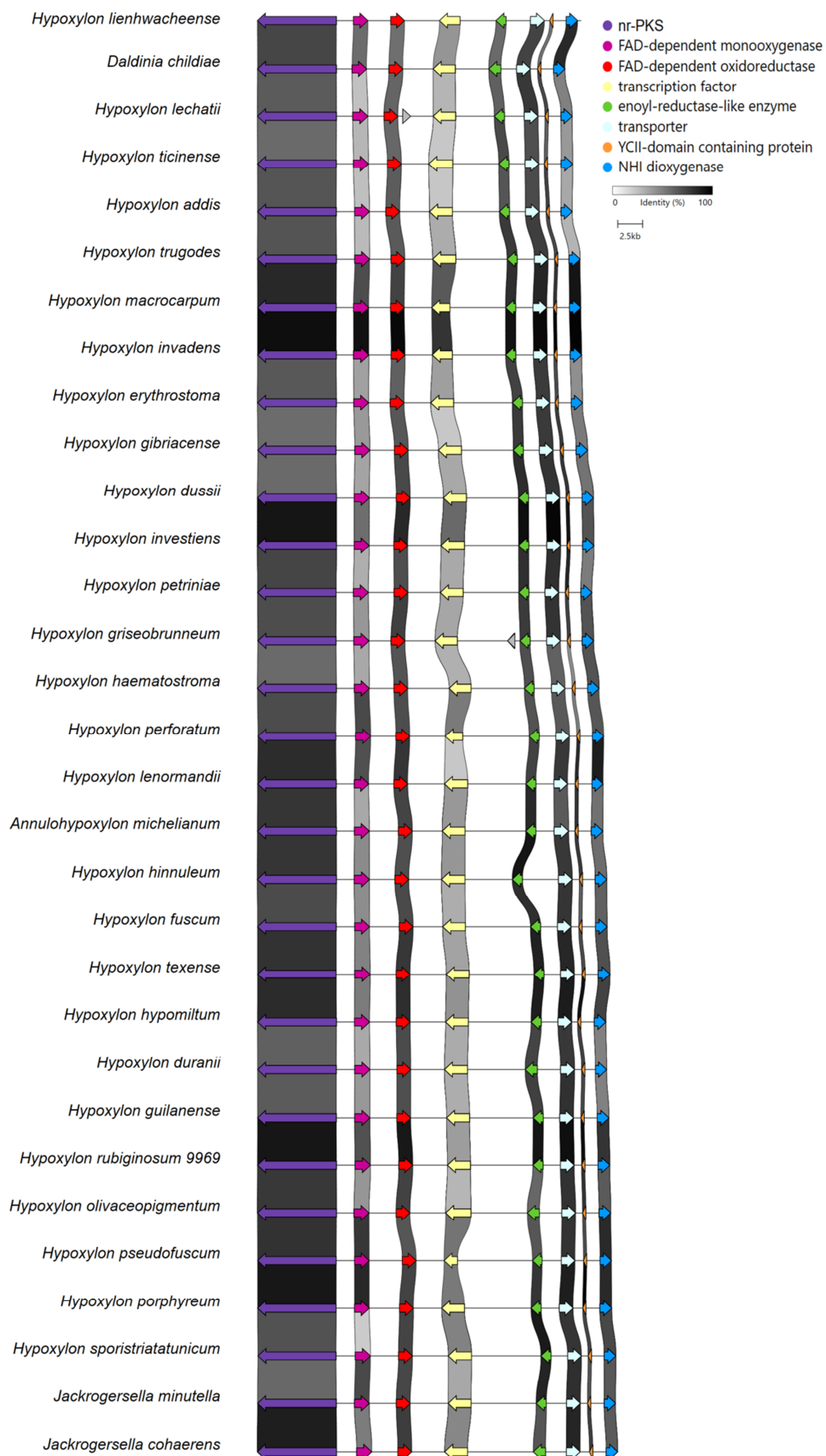
Figure 4.2 Alignment of putative sepedonin BGC of 30 species in comparison to *H. lienhwacheense*.

In case of *H. lateripigmentum* all putative genes with homology to *lwt* genes are localized in the same order, but are contained on two different contigs (Figure 4.3). The *lwtS* and *lwtR1* homologues are located on contig 5, adjacent to the same homologous putative genes as in *H. lienhwacheense* (microtubuli-binding protein encoding gene, a gene with no assigned function, and a cytochrome P450 encoding gene). The gene homologs of *lwtR2* - *lwtR7* are located on contig 2. The putative genes are not at the ends of the contigs, which excludes an interruption of the gene cluster reasoned by the sequencing methods. A transposition event was speculated for this BGC.
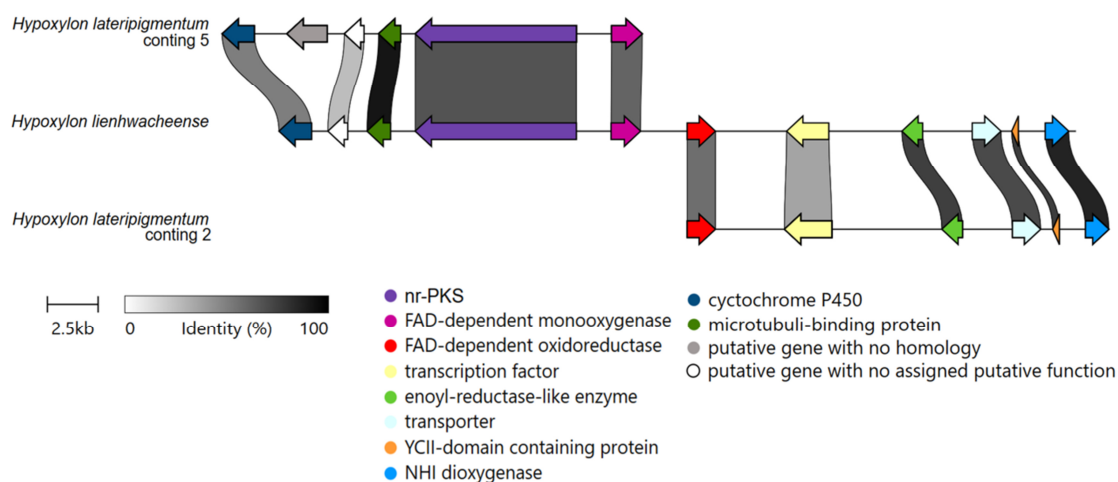
Figure 4.3 Alignment of putative sepedonin BGCs of *Hypoxylon lateripigmentum* comparison to *H. lienhwacheense*.

Three species of the 35 *Hypoxylacae* or related fungi contain putative tropolone-forming core genes (nr-PKS, FMO, NHI) in addition to a gene with a homology to the α-humulene cyclase (Figure 4.4, pink). Thus, the species *Parahypoxylon papillatum*, *Parahypoxylon ruwenziorense*, and *Creophaeria sassafras* form the second category of putative tropolone BGCs. A hetero-Diels-Alderase (Figure 4.4, light green) is present in two species (*Parahypoxylon papillatum*, *Creophaeria sassafras*) shown by the homology to the noreupenifeldin **161** BGC from *Phaeosphaeriaceae sp. CF-150626* and the *Phoma sp.* eupenifeldin **162** BGC.[95] The formation of meroterpenoids was speculated for these two BGCs, but further investigations are required for determination of the pathway products. Nonetheless, the presence of the well-known tropolone-forming core enzymes suggests the inclusion of a tropolone moiety as product or intermediate of the resulting biosynthetic pathway in *P. papillatum*, *P. ruwenziorense*, and *C. sassafras*.
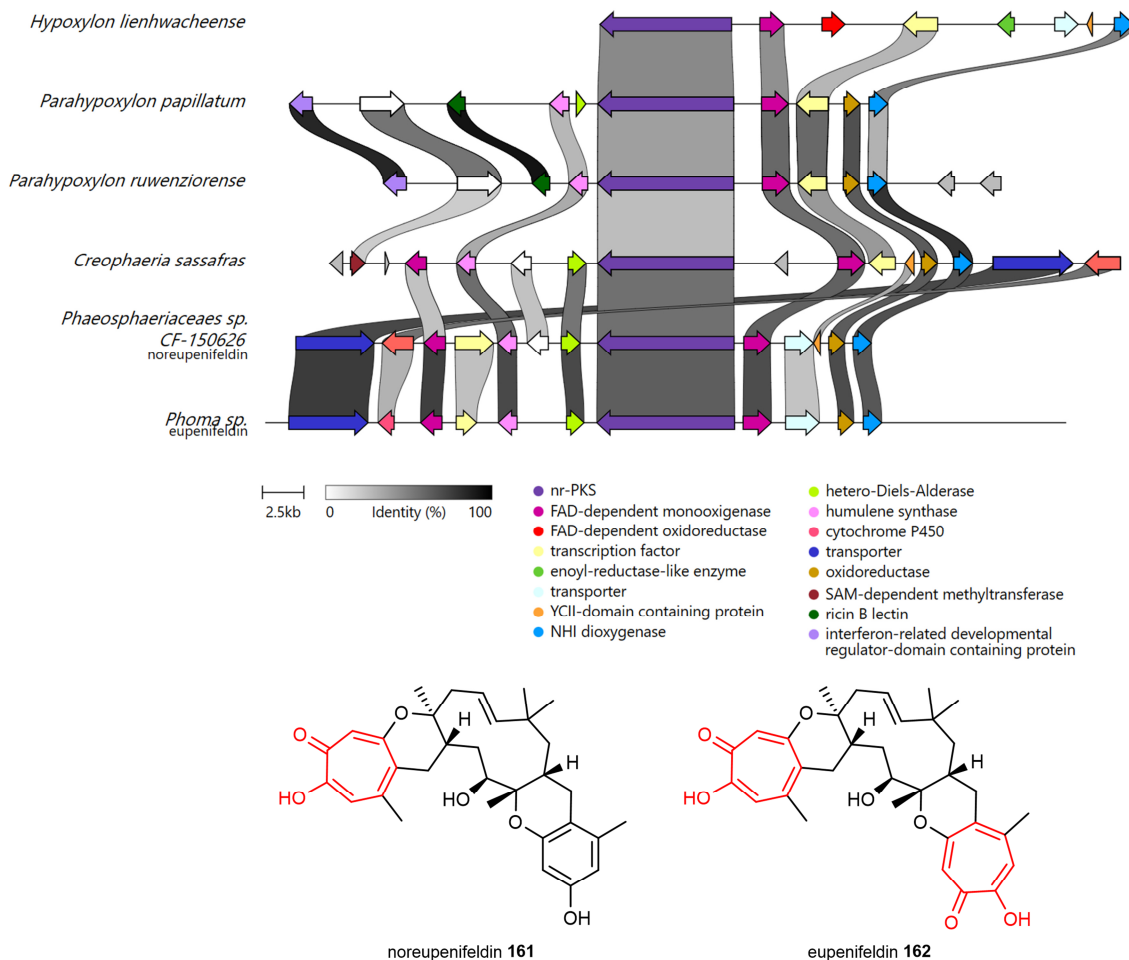
Figure 4.4 Alignment of putative tropolone BGC of 3 species in comparison to the known BGCs *H. lienhwacheense*, *Phaeosphaeriaceae sp. CF-150626* (noreupenifeldin), and *Phoma sp.* (eupenifeldin; white = hypothetical genes, grey = no homology to other genes).[95,180]

The final category includes only the species *Hypoxylon laschii* (Figure 4.5). The gene cluster with most related genes to the *lwt*-BGC contains only a putative PKS and an FMO encoding gene. Accordingly, a functional tropolone-forming biosynthetic pathway seems to be absent from *H. laschii*.
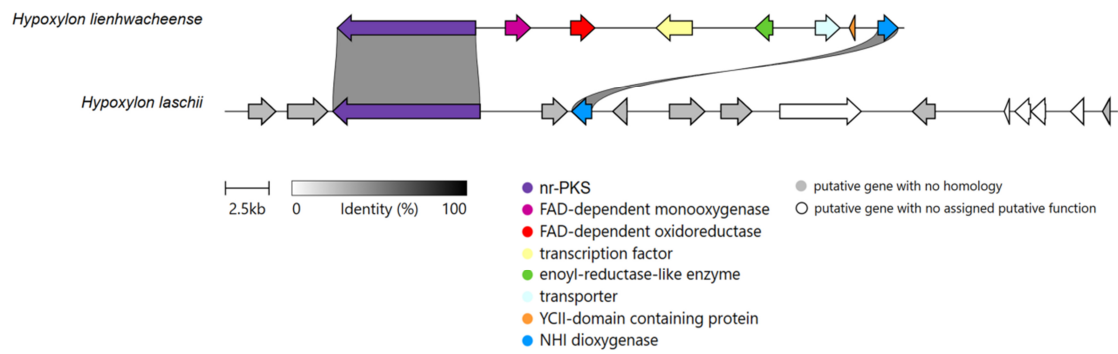
Figure 4.5 Alignment of putative BGC of *Hypoxylon laschii* in comparison to *H. lienhwacheense*.

Geneious v. 9.1.8 was used to build a phylogenetic tree (Figure 4.6), based on multiple alignment of all putative tropolone-BGCs. The DNA sequence of each BGC was used for the alignment to obtain distance values, which were the basis for the visualisation of the relation of the BGCs. The same three groups were observed (Figure 4.6, blue, grey, and black). *H. laschii* and the species *C. sassafras*, *P. papilatum* and *P. ruwenzoriense* tropolone BGCs are less close related to the tropolone BGCs in the remaining species.
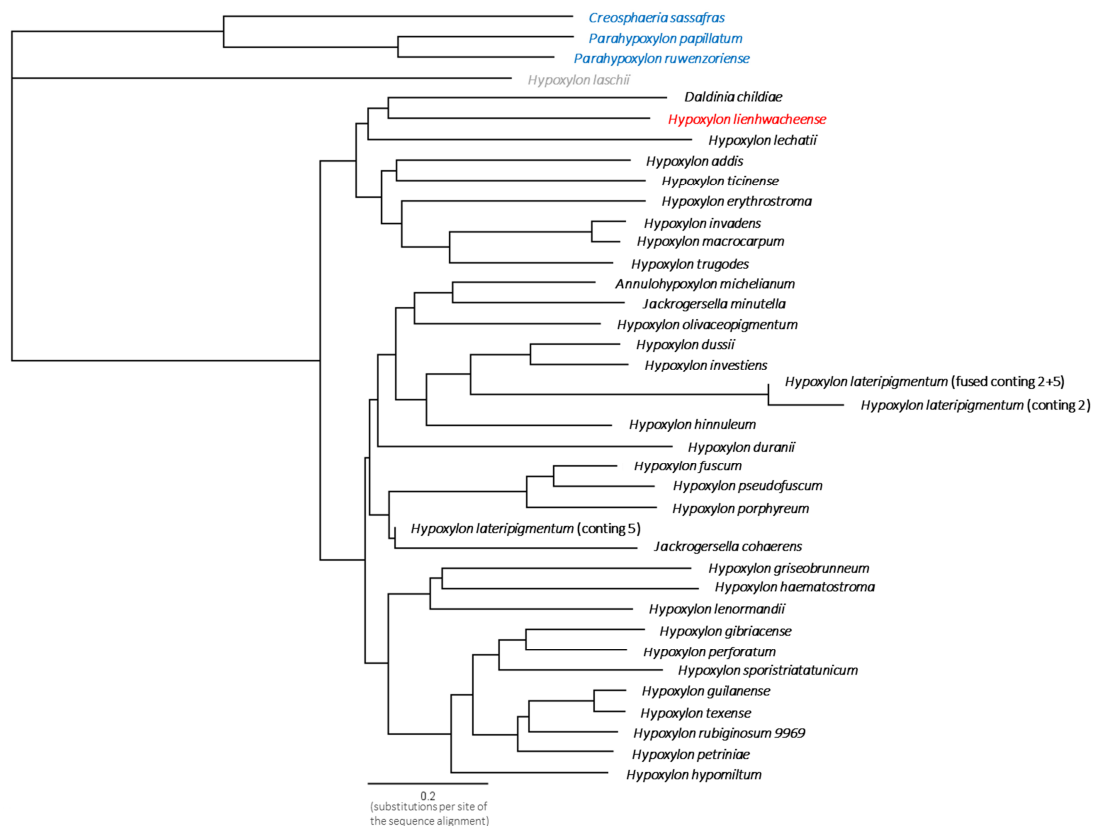
Figure 4.6 Phylogenetic tree of analysed species based on distance matrix.

129

Out of the 35 genomes of species that were analysed, 34 genomes contain a putative tropolone-forming BGC. This observation may suggest a significant ecological role of the proposed tropolone metabolite, as it was highly conserved during the evolutionary process.

4.3.3 First Approaches on Heterologous Expression – Summary of Previous Work

Previous heterologous expression work had only investigated the biosynthesis of tetraketide tropolones (section 1.4.3.2). In the case of sepedonin and related compounds in *A. nidulans*, investigations had only been done using gene knockouts (section 4.1.1). The aim here, therefore, was to use heterologous expression as a synthesis platform to investigate the production of pentaketide tropolones.

In previous unpublished work by Dr. Jin Feng, attempts were made to understand the *H. lienhwacheense* tropolone pathway by heterologous expression of varied gene sets in *A. oryzae* NSAR1 (Table 4-3). The biosynthetic genes were inserted into fungal expression vectors (Table 4-2), combined and transformed into *A. oryzae* NSAR1 as previously described. The extracts of the transformants were analysed by LCMS, but the analysis was unfinished at the start of this project.

Table 4-2 Summary of vectors used in first approaches of heterologous expression of the *lwt*-BGC.

Vector name	Putative enzymes	Selection marker
pTY-*argB-lwtS*	PKS	*argB*
pTY-*argB-lwtS-R1*	PKS, FMO	*argB*
pTY-*argB-lwtS-R1-R7*	PKS, FMO, NHI	*argB*
pTY-*argB-lwtS-R1-R7-R2*	PKS, FMO, NHI, OR	*argB*
pTY-*argB-lwtS-R1-R7-R4*	PKS, FMO, NHI, MDR	*argB*
pTY-GS-*sC-lwtR6*	YCII	*sC*
pTY-GS-*sC-lwtR4-R6*	YCII, MDR	*sC*
pTY-GS-*sC-lwtR2*	OR	*sC*
pTY-GS-*sC-lwtR4*	MDR	*sC*

4.3.3.1 Heterologous Biosynthesis of DHMBA

Heterologous expression experiments by Dr. Jin Feng first introduced the putative nr-PKS gene (*lwtS*) into *A. oryzae* NSAR1 (Figure 4.7 A), expecting the transformant to

produce the first enzyme-free pentaketide intermediate, DHMBA **156**. Transformation, selection, and liquid cultivation took place as described before. The resulting eight DPY liquid cultures were extracted twice with equal amounts of ethyl acetate (EtOAc), but without previous acidification of the culture. The solution was dried (MgSO$_4$) and then the solvent was removed under reduced pressure. Surprisingly, the expression of the PKS alone did not lead to DHMBA **156**. Overall eight transformants were screened, of which two produced no new compounds. In the other six transformants, a new peak was found (Figure 4.7 A) at 5.5 min.
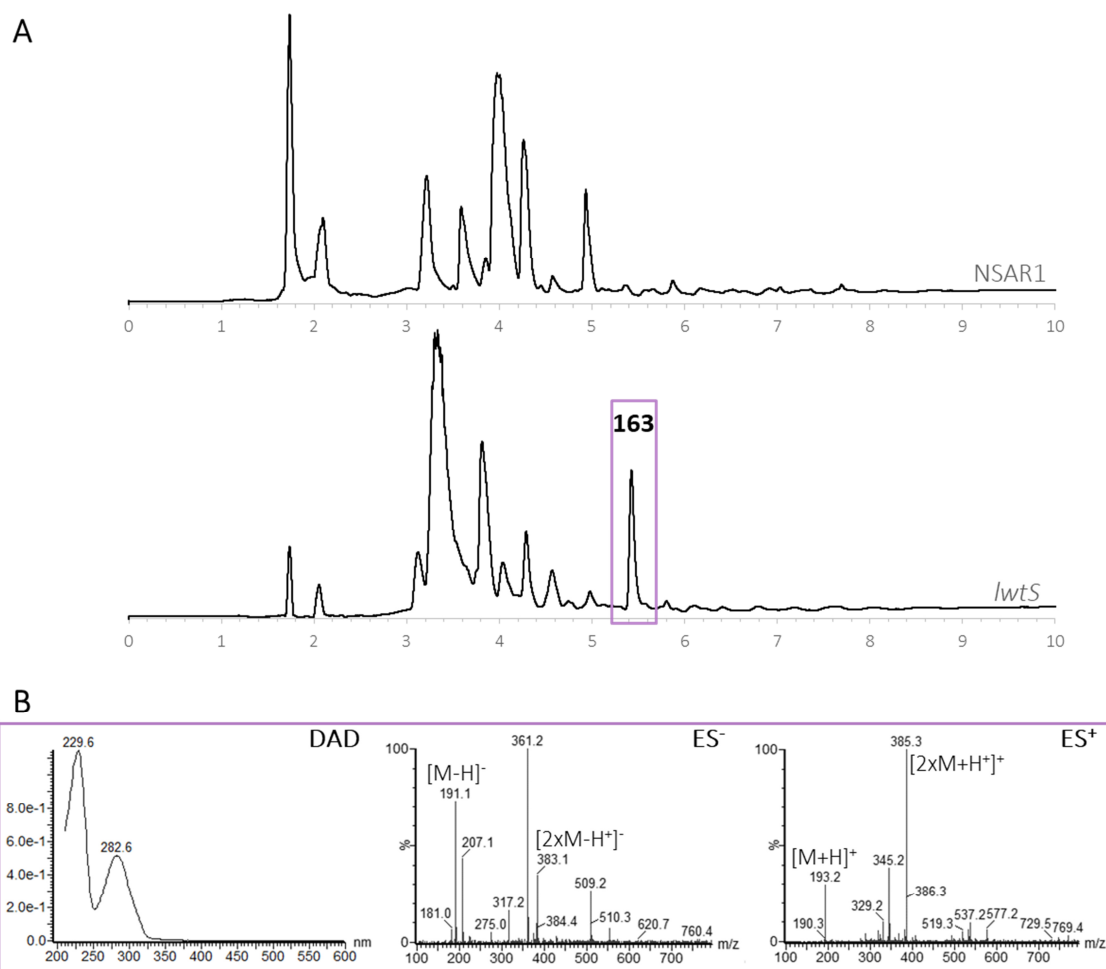
Figure 4.7 Heterologous expression in *A. oryzae* NSAR1 with LwtS: **A**, DAD chromatograms (210 - 600 nm); **B**, characteristic UV and mass spectra of **163**.

The new peak of **163** was characterized by UV maxima at 229/282 nm (Figure 4.7 B), which does not correspond to the characteristic data of DHMBA **156** (220/295 nm).[177] Although a fragmentation ion of a *m/z* of 207 was found in the ES$^-$ spectrum (Figure 4.7

B), which would be in agreement to the mass of DHMBA (208 Da), the main peaks of ES⁻/ES⁺ spectra indicate a different putative mass of the compound **163** (192 Da). Structural analysis was not carried out with the unknown compound **163** at that time (*vide infra* section 4.3.4). It was not possible to produce the expected product DHMBA **156**. The mass difference between DHMBA **156** (208 Da) and **163** (192 Da) of 16 Da (Scheme 4.2) may indicate a reduction (+ 2 Da) and the loss of water (- 18 Da). Therefore, **163** was suggested to be a shunt metabolite, resulting from DHMBA **156**.

4.3.3.2 Heterologous Co-Expression Including Putative Genes for Tailoring Steps

Different combinations of the *lwt* genes were assembled in expression vectors (Table 4-3) as a follow-up experiment by Dr. Jin Feng. These were transformed into *A. oryzae* NSAR1 and analysed as before. In most extracts, the unknown shunt compound **163** remained the main product (Table 4-3 A-D). For the co-expression of other gene combinations, DHMBA **156** was detected (Table 4-3 E-J, Figure 4.8 A). DHMBA **156** is characterized by a mass of 208 Da (indicated by ES⁻/ES⁺ detection mode, *e.g.* Figure 4.8 B), and a UV maximum at 220 nm and 295 nm in the DAD chromatogram, (Figure 4.8 B). The structure of DHMBA **156** was confirmed by isolation and NMR (Dr. Jin Feng, section 7.4.1.2).
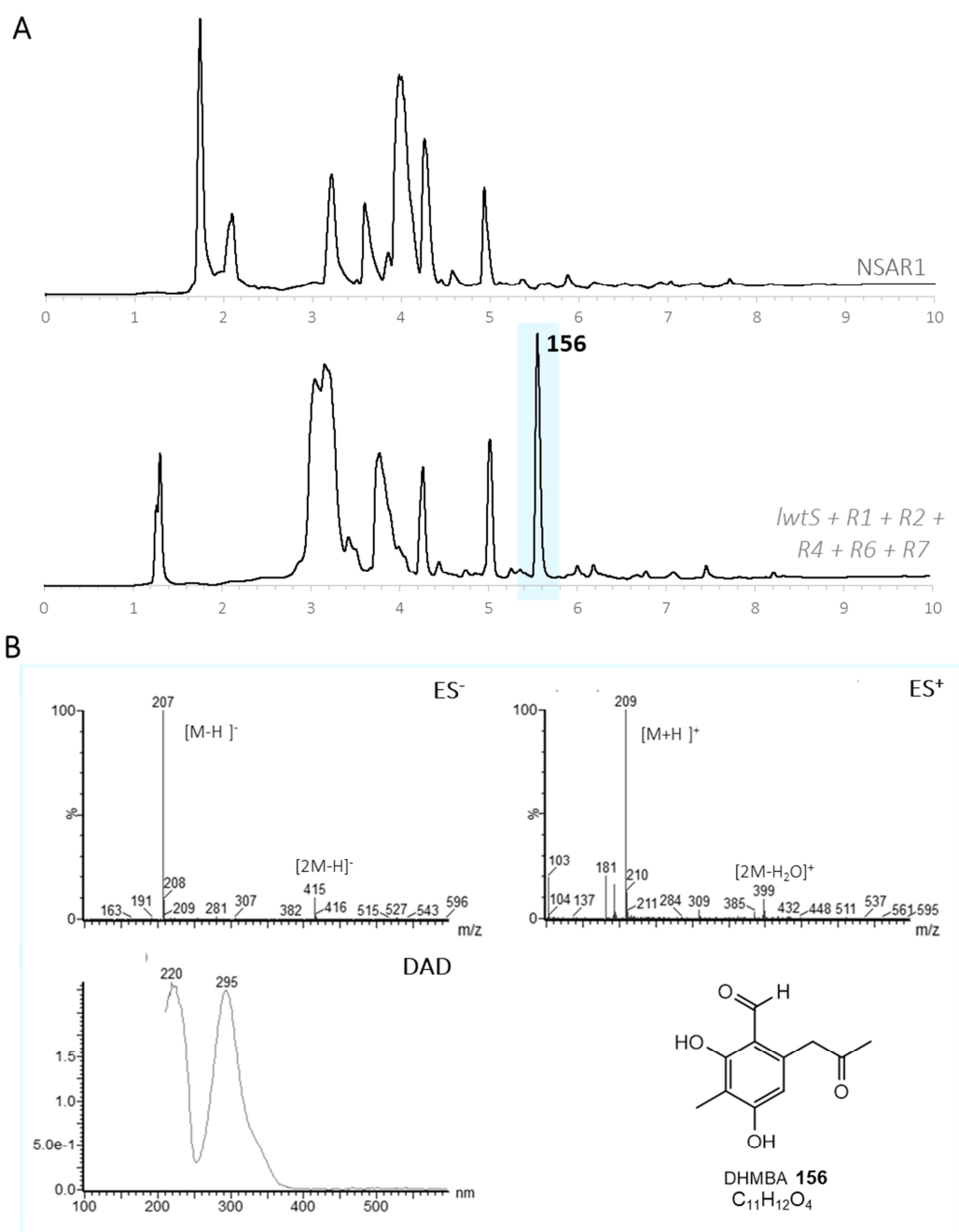
Figure 4.8   LCMS data of a transformant including genes *lwtS+R1+R2+R4+R6+R7*: **A**, DAD chromatogram; **B**, characteristic DAD, ES[+] and ES[-] spectra for DHMBA **156**.

In the overall picture, the transformants mainly either produced the unknown shunt compound **163**, or DHMBA **156** (Table 4-3, experiments A-J), but no other later intermediate of the pathway, although putative genes encoding enzymes for tailoring of DHMBA were introduced (Figure 4.8 A, Table 4-3). Late intermediates of the pathway were also not detected, when the full cluster (Table 4-3, experiment I) was transformed. The proposed pathway for pentaketide tropolones according to previous research by Gerke *et al*. (Scheme 4.1) suggests the formation of DHMBA **156**, which reacts to

133

sepedonin **90** or antibiotic C **91** by the inclusion of the full BGC (Scheme 4.1). Therefore, the expected products of the transformants of experiments B-E and H-I (Table 4-3, Figure 4.8) are, in addition to DHMBA, also **91** and sepedonin **90**, as well as anhydrosepedonin **160** or even other unknown tropolone-related compounds. No intermediates of the later proposed biosynthetic pathway were detected.

Table 4-3 Summary of heterologous expression experiments by Dr. Jin Feng.

| Exp. | putative *lwt* genes | | | | | | number of transformants | Number producing | |
	S	R1	R2	R4	R6	R7		DHMBA 156	163
A	✓	-	-	-	-	-	8	0	6
B	✓	✓	-	-	-	✓	7	0	5
C	✓	✓	✓	-	-	✓	4	0	3
D	✓	✓	-	✓	-	✓	10	0	5
E	✓	✓	✓	✓	-	✓	11	6	4
F	✓	-	-	-	✓	-	4	1	2
G	✓	✓	-	-	✓	-	3	1	2
H	✓	✓	-	-	✓	✓	1	1	0
I	✓	✓	✓	✓	✓	✓	10	6	3

These previous findings by Dr. Jin Feng raise two questions as a starting point for further analysis: (I) the structure of the shunt compound **163** remains unclear and should be investigated further (section 4.3.4); and (II) an investigation of the absence of later intermediates (section 4.3.5) and late steps of the biosynthetic pathway is required (section 4.3.6).

4.3.4 Analysis of the Formation Shunt Intermediates

Additional heterologous expression experiments were carried out by previously described methods to elucidate the early steps of the pathway in more detail and investigate the formation of **163**. Thus, additional transformation experiments were achieved involving only the *lwtS* gene by using the vector pTY-*argB-lwtS*, kindly provided by Dr. Jin Feng.

4.3.4.1 Extraction of LwtS Transformant

After transformation, the transformants were selected on appropriate selection agar and cultivated with previously described methods. During the extraction, each culture was separated into two parts (Figure 4.9). One half was extracted without prior acidification (Figure 4.10) similarly to the experiments by Dr. Jin Feng (section 4.3.3). The extraction of the second half was carried out with prior acidification of the media (Figure 4.10), according to the standardised protocol (section 6.2.3.1). For each extraction, equal amounts of ethyl acetate (EtOAc) were used. The solvent was dried with MgSO$_4$ and then removed under reduced pressure.
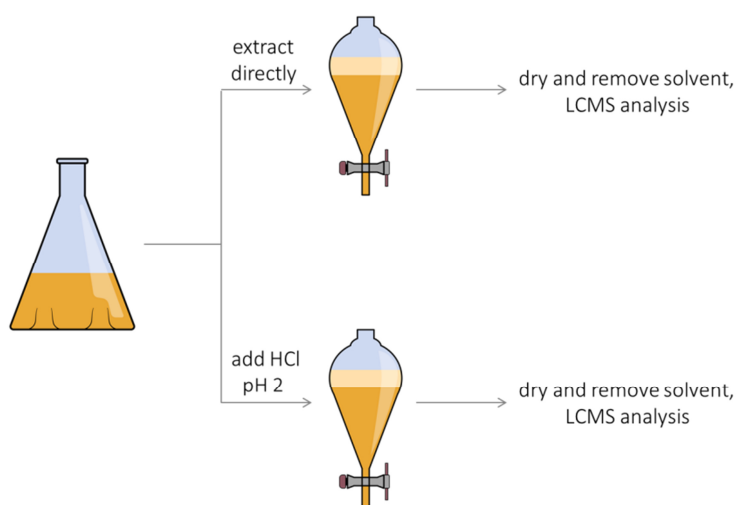
Figure 4.9 Workflow for the extraction of *lwtS* transformants.

From eight transformants including *lwtS*, four produced the unknown compound **163**, which was characterized by the same characteristic data like previously detected by Dr. Jin Feng (Figure 4.10 D). When no acid was added during the extraction process (Figure 4.10 B), one peak **163** at 5.6 min was observed when analysing the extracts of transformants. DHMBA **156** was not detectable for any transformant.
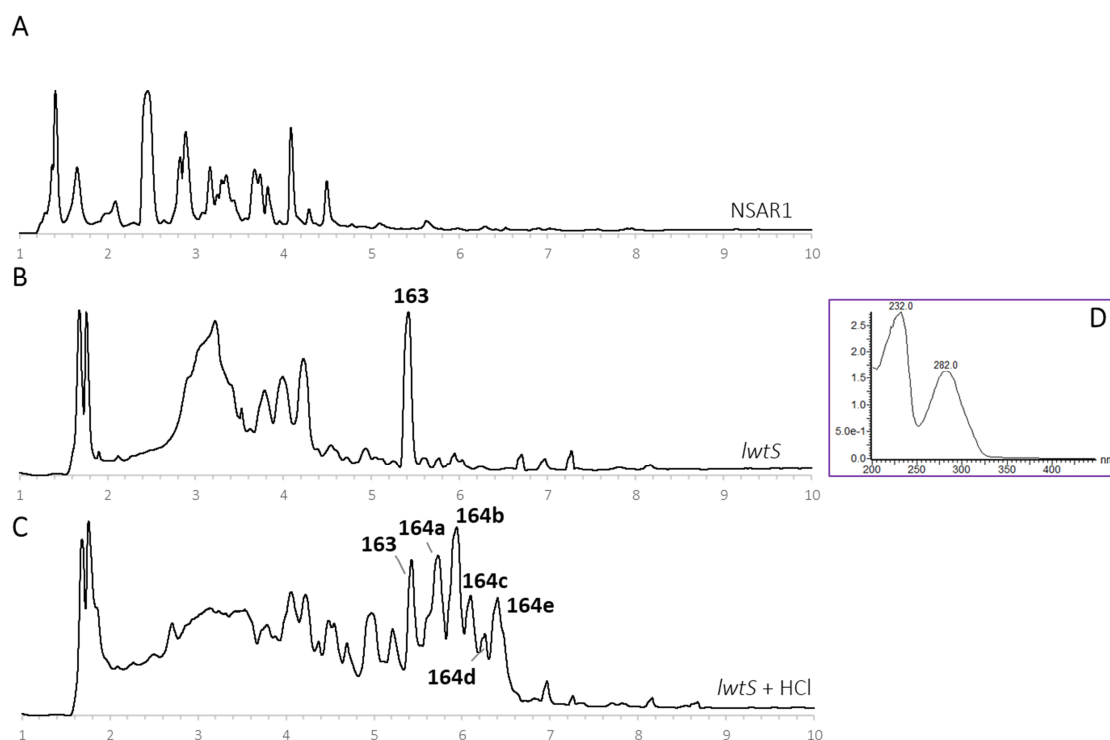
Figure 4.10 Comparison of transformants with *lwtS* with and without prior acidification before extraction: **A**, DAD chromatogram of *A. oryzae* NSAR1; **B**, DAD chromatogram of transformant without acidification during the extraction process; **C**, DAD chromatogram of transformant with acidification during the extraction process; **D**, UV spectrum of **163**.

The analysis of the same transformant was carried out with prior acidification before extraction. The transformant that previously produced only **163** (Figure 4.10 B), showed five new overlapping peaks **164a-e** between 5.6 min – 6.5 min (Figure 4.10 C) by the addition of HCl during the extraction process.

All of the new appearing compound peaks **164a-e** show a similar fragmentation pattern in their ES^+/ES^- spectra, giving the mass values of 192 Da, 384 Da, 576 Da, 768 Da and 960 Da (Figure 4.11). The fragment ions show a consistent *m/z* difference of 192 Da between each peak (Figure 4.11). The UV absorbance maxima shift from 236 nm/280 nm increased with later retention time to 226 nm/331 nm.
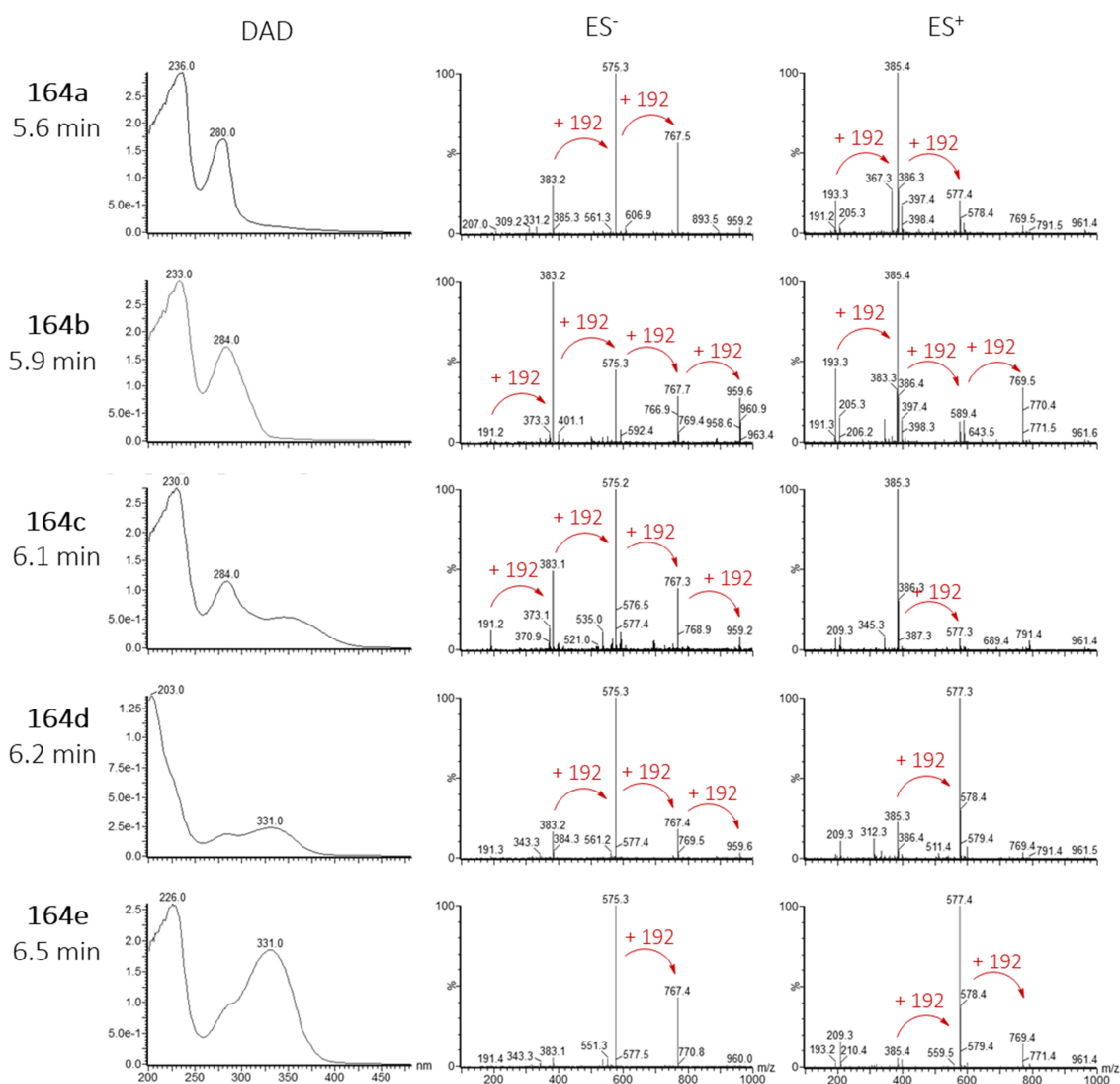
Figure 4.11 Overview of UV spectra and ES$^+$/ES$^-$ spectra of unknown compound peaks between 5.6 min - 6.5 min.

Based on the MS data it was hypothesized, that **163** is a reduction shunt of **156**. The two obvious positions for reduction are the C-1 aldehyde and C-9 ketone of DHMBA **156** (Scheme 4.2, red). Furthermore, it was hypothesised that shunt compound **163** forms oligomers **164a-e** under acidic conditions, adding the mass of 192 Da for each structure (Scheme 4.2).
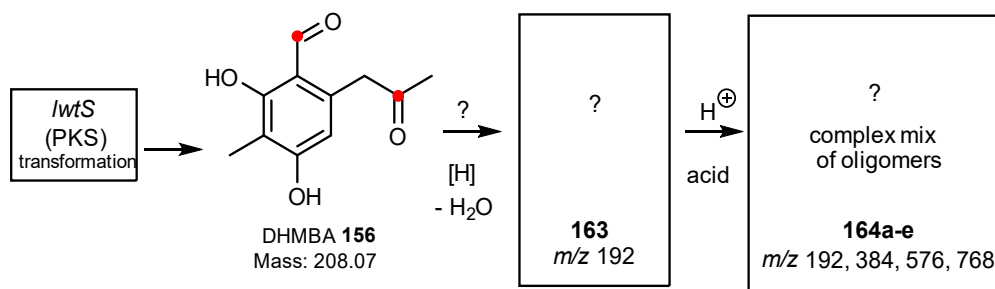
Scheme 4.2 Overview of hypothesized reactions leading to shunt compounds (red = reducible positions).

4.3.4.2 Compound Isolation

The isolation of **163** was attempted by preparative LCMS (Figure 4.12). A titre of 16 mg/l was obtained. However, after purification the fraction appeared as a mixture. Other minor related peaks (**165** - **167**) were also observed. The main component in the mixture was peak **164f** at 5.6 min. **164f** has a different fragmentation pattern than **163** in ES$^-$/$^+$ mode, characterized by the same peak pattern in ES$^-$/$^+$ mode as before **164a-e** with masses of 192 Da, 384 Da, 576 Da, 768 Da and 960 Da. Furthermore, **164f** has similar UV maximum absorbance values as **163**, indicating a relationship to **163**.
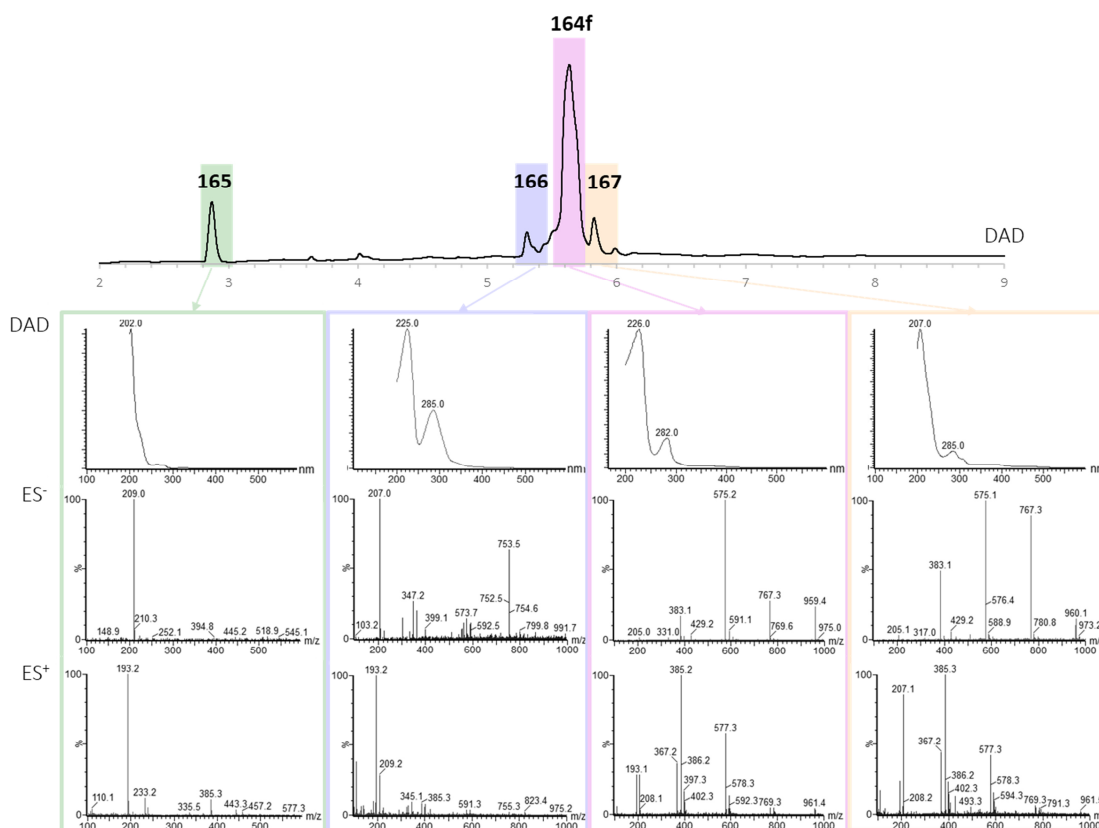
Figure 4.12 DAD chromatogram of isolated fraction and characteristic spectra of each peak.

4.3.4.3 UPLC-HR-MS Analysis of the Shunt Compound Mixture

The unknown compound mixture was analysed with Ultra Performance Liquid Chromatography high resolution- (UPLC-HR)-MS to obtain the chemical formula of the unknown fragments (Table 4-4). For the mass of 192 Da, a chemical formula of $C_{11}H_{12}O_3$ was calculated. In comparison to DHMBA **156** (chemical formula $C_{11}H_{12}O_4$) the chemical formula $C_{11}H_{12}O_3$ is in agreement with the prior hypothesis of the formation of shunt metabolites by reduction of DHMBA **156** and the loss of water (Scheme 4.2).

Table 4-4 Summary UPLC-HRMS Analysis.

m/z in ES$^+$ [M+H]$^+$	calculated mass	calculated chemical formula
193.0870	193.0865	$C_{11}H_{12}O_3$
385.1640	385.1651	$C_{22}H_{24}O_6$
577.2446	577.2438	$C_{33}H_{36}O_9$

The chemical formula for the *m/z* of 385 and 577 is in agreement with the oligomerization of an unknown compound by adding $C_{11}H_{12}O_3$ (192 Da) to each additional fragmentation ion *m/z* in the ES$^+$ mode.

4.3.4.4 NMR Analysis of the Shunt Compound Mixture

The ^{13}C NMR analysis (Figure 4.13 A) of the purified fraction in wet CH_3CN confirmed the presence of a complex compound mixture. MS analysis showed that there is a probable reduction of DHMBA **156**. The two most likely positions for reduction are the C-1 aldehyde and C-9 ketone of DHMBA **156** (Scheme 4.2). Ketones and aldehydes are distinguishable in ^{13}C NMR. Aldehyde carbonyls tend to resonate below 200 ppm while ketones resonate above 200 ppm. The presence of three clear peaks at 205.8 ppm, 206.1 ppm and 206.7 ppm (Figure 4.13 B) and no peaks between 190 - 200 ppm suggest that **156** has been reduced at the C-1 aldehyde position. As a first hypothesis, it was assumed that the produced primary alcohol is the starting material for further oligomerisation reactions. The ^{13}C and ^1H NMR data was then examined to find additional evidence to support this idea.
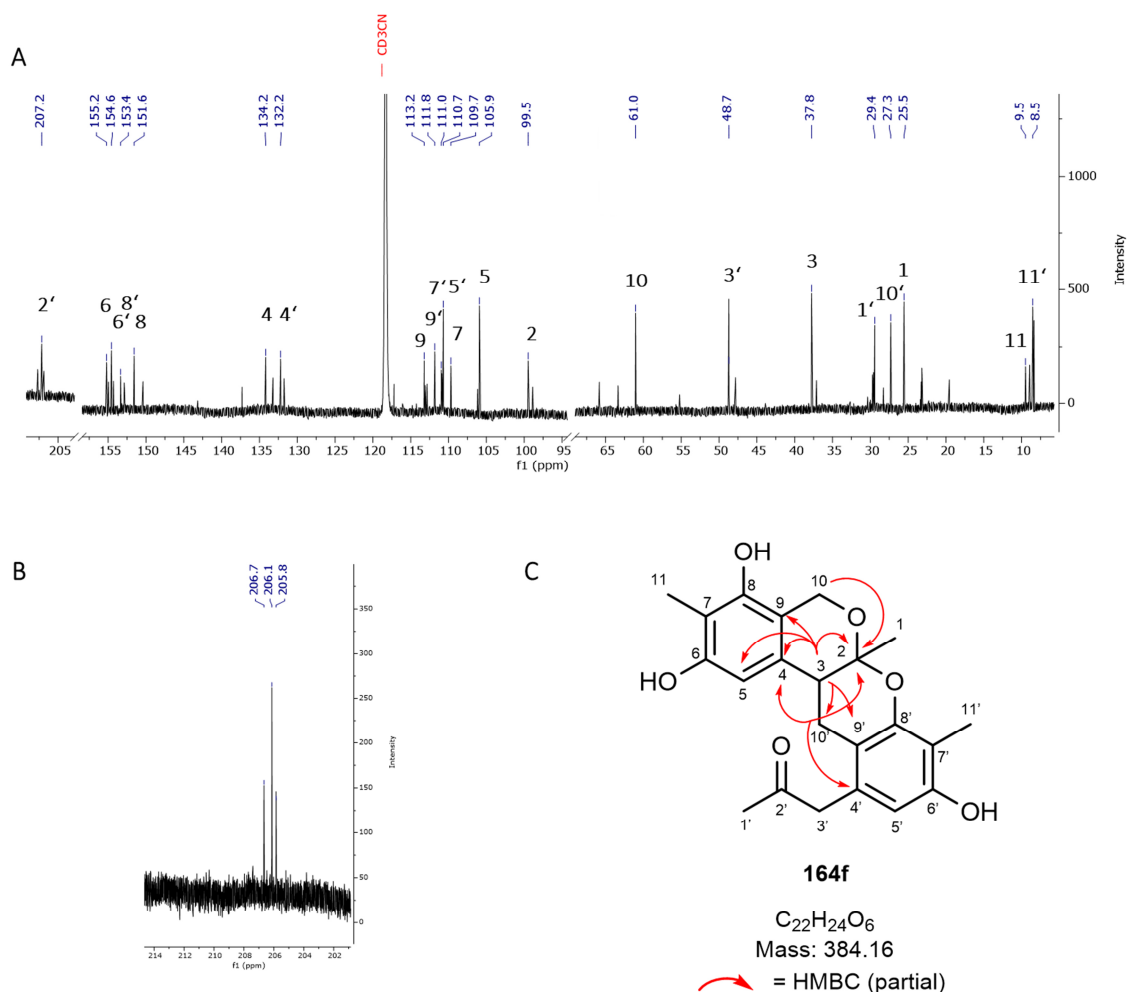
Figure 4.13 ^{13}C NMR data for shunt compound **164f**: **A**, ^{1}H NMR (126 Hz) of **164f** in CD$_3$CN; **B**, ^{1}H NMR (600 Hz) of **164f** in CD$_3$CN with zoom between 202 – 214 ppm; **C**, identified structure **164f**.

The NMR analysis is consistent with the structure of **164f** (Figure 4.13 C, Table 4-5) for the main component of the mixture. The HSQC shows signals for the positions 11/11', 10/10', 5/5', 3/3' and 1/1' (Figure 4.14). Furthermore, a ^{13}C NMR signal at 99.2 ppm is indicative for the presence of a hemiacetal or acetal, supporting the hypothesis of this linkage for oligomerisation. As the HSQC detects no correlation to position 2, an acetal structure is most likely present at position 2 (Figure 4.14, green). Furthermore, the diastereotopic 10/10'-hydrogens show a correlation to the 2-carbon, connecting two monomers to each other (Figure 4.15 C, green, Figure 4.13 C). A further key signal which correlates with structure **164f** is the 3-hydrogen signal, appearing as a doublet of doublets (Figure 4.15 B), coupling with the diastereotopic 10'-hydrogens (Figure 4.15 A) and additionally correlating to 2, 4, 5, 9, 9' in the HMBC spectrum (Figure 4.15 C, blue). Furthermore, the 10' hydrogens correlates with both, 4-C and 4'C in the HMBC spectrum (Figure 4.15 C, yellow).

140
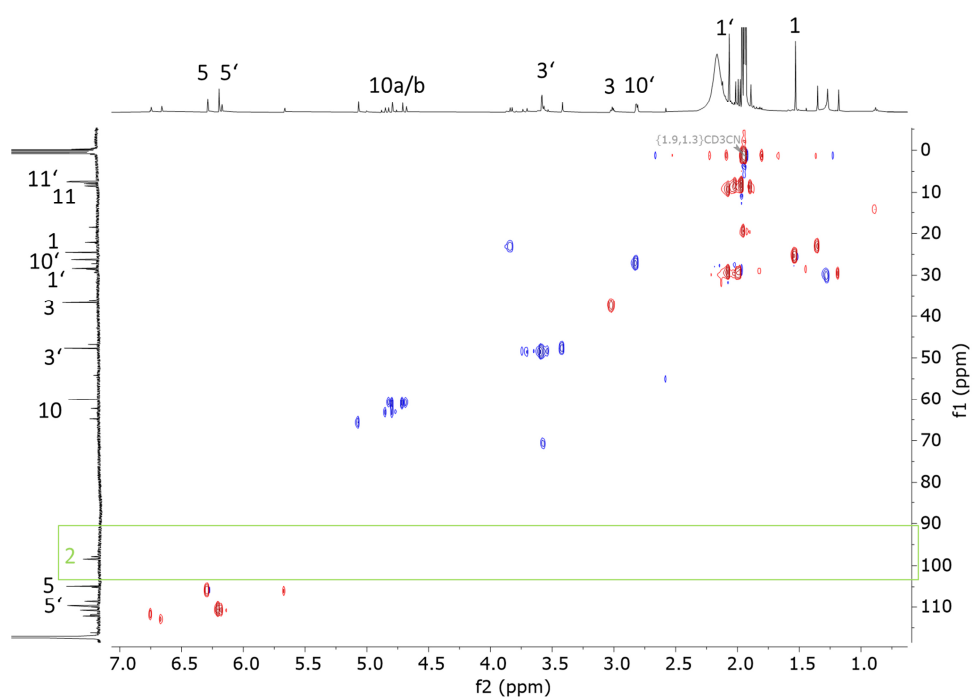
Figure 4.14 HSQC of **164f** in CD$_3$CN.
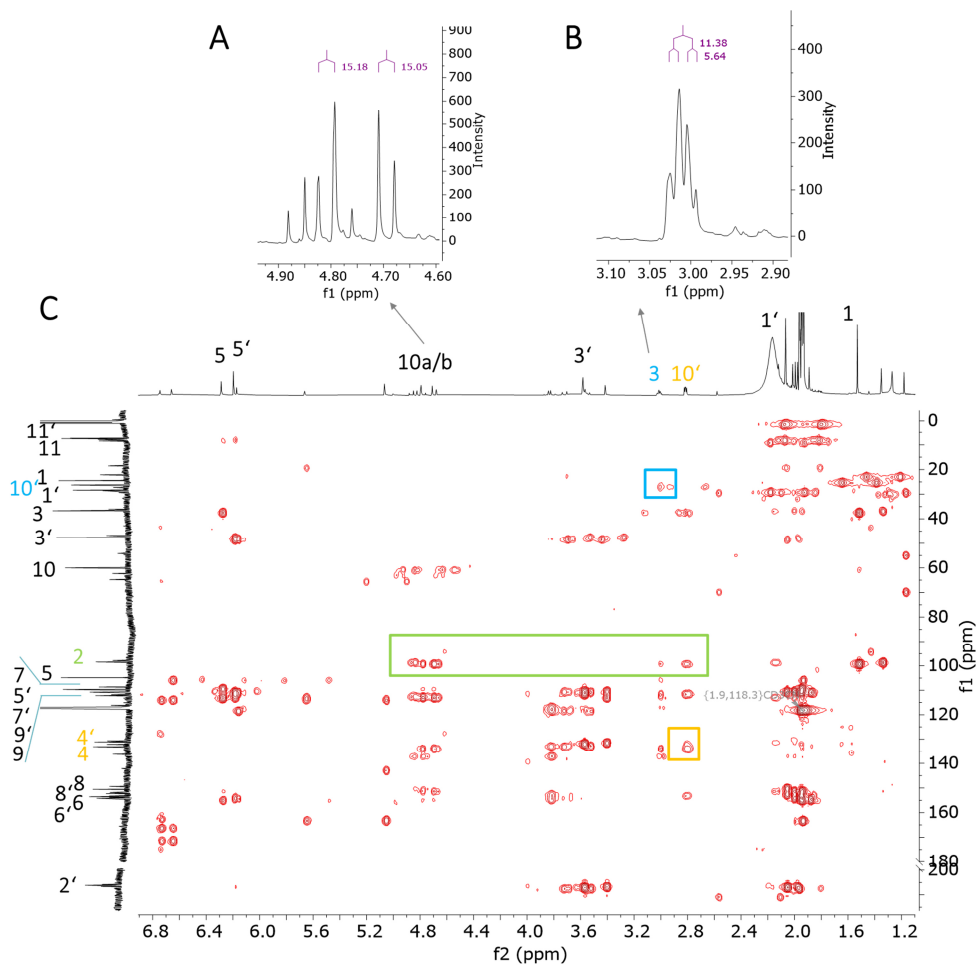
Figure 4.15 HMBC of 1**64f** in CD$_3$CN.

The structure of **164f** correlates with chemical shifts according to literature (Table 4-5).[183] The structure of **164f** was concluded to be the main compound for the analysed compound mixture. It presumably arises by dimerization of **163** related compounds during its attempted purification.

Table 4-5 Chemical shifts of **164f** in DMSO (600 MHz).

Atom	C-type	δ_C/ppm	δ_H/ppm (mult, J in Hz)	HMBC	Literature[183] in CDCl$_3$ (10% CD$_3$OD) δ_C/ppm	δ_H/ppm
1	CH$_3$	25.6	1.54 (s)	2, 3	25.3	1.60 (s)
2	C	99.5			98.5	
3	CH	37.8	3.01 (dd, 5.4, 11.4 Hz)	2, 4, 5, 9,9', 10'	36.8	3.02 (dd, 5.1, 4.4 Hz)
4	C	134.2			132.3	
5	CH	105.9	6.29 (s)	3, 7, 6, 9	104.3	6.16 (s)
6	C	155.2			154.1	
7	C	109.7			108.3	
8	C	151.6			150.2	
9	C	113.2			112.0	
10	CH$_2$	61.0	4.69 (d, 15.1 Hz) 4.81 (d, 15.2 Hz)	2, 4, 8, 9	60.2	4.93 (d, 14.7 Hz) 4.81 (d, 14.7 Hz)
11	CH$_3$	9.5	2.08 (s)	*	7.6	2.03 (s)
1'	CH$_3$	29.3	2.02 (s)		28.6	2.02 (s)
2'	C	207.2			210.4	
3'	CH$_2$	48.7	3.56 (m*)	4', 2', 5'	49.4	3.60 (d, 15.1 Hz) 3.46 (d, 15.1 Hz)
4'	C	132.2			129.5	
5'	CH	110.7	6.20 (s)	3', 6', 7'	109.9	6.25 (s)
6'	C	154.6			153.5	
7'	C	111.0			111.2	
8'	C	153.4			152.7	
9'	C	111.8			110.5	
10'	CH$_2$	27.2	2.83 (dd, 5.4, 3.1 Hz)	2, 3, 4, 4', 8', 9'	26.1	2.98 (dd, 16.1, 4.4 Hz) 2.86 (dd, 16.1, 5.1 Hz)
11'	CH$_3$	8.5	1.98*	*	8.0	2.06 (s)

* overlapped by solvent peak/impurity

4.3.4.5 Proposed Reaction Sequence leading to Shunt Metabolites in *A. oryzae* NSAR1

The ^{13}C NMR gave evidence for a DHMBA structure, which was reduced at the C-1 position, preserving the C-9 ketone group. The reductive reaction leads to compound **163**, which is the intermediate for further reactions. **163** with a mass of 210 Da reacts to the ortho-quinomethide **165** (Scheme 4.3) by acidification or in the MS source. **165** is reactive and forms, for example, the hemiacetal **166**. **165** and **166** have the same chemical formula of $C_{11}H_{12}O_3$ and mass of 192 Da, which is in agreement with the obtained MS data.
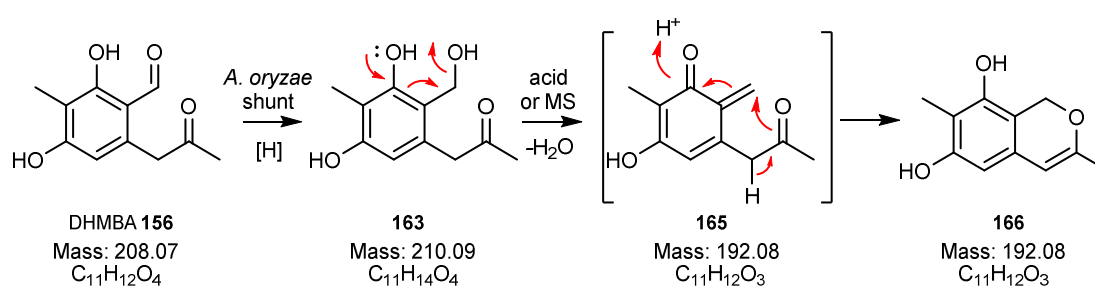
Scheme 4.3 Proposed formation of shunt metabolites form DHMBA **156**.

The formation of **164f** indicates that **165** is reactive and leads to different oligomers (Scheme 4.4). The formation of **164f** is based on a dimerization of **165** by a hetero Diels Alder reaction (Scheme 4.4). Multiple other dimeric forms of **165** are feasible, for example **167a-c**. The formation of **167a-c** is probably indicated by the minor components in the compounds mixture, but **164f** is the dominant component, leading to the hypothesis that **164f** is the most stable shunt metabolite.
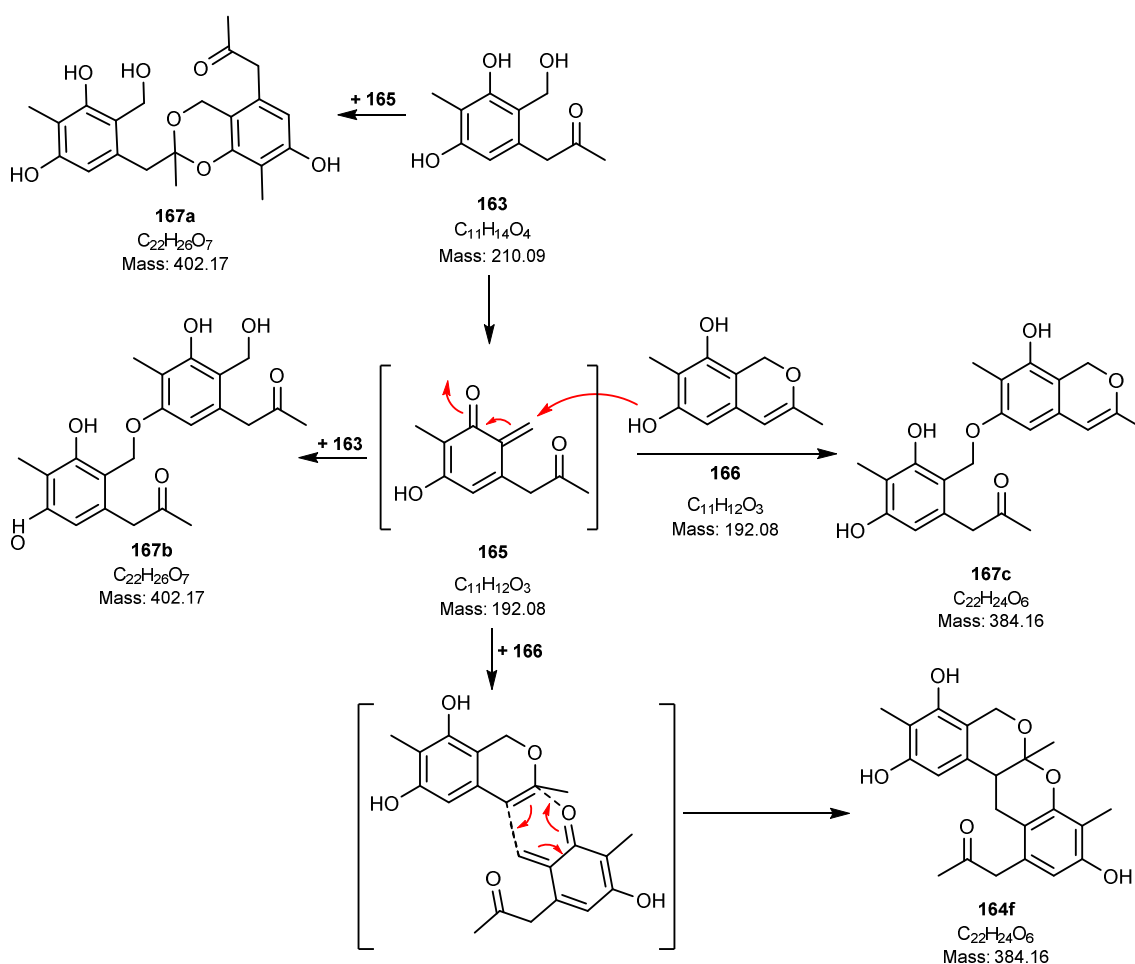
Scheme 4.4 Proposed formation of shunt metabolites forming dimers.

Reactions forming longer oligomers are also possible (Scheme 4.4).The observed mass of 576 Da in the LCMS spectra (Figure 4.11) corresponds to the presence of trimers ($C_{33}H_{36}O_9$). Accordingly, tetramers would be indicated by the mass of 768 Da (Table 4-4).

Since the peaks of the compounds **164a-e** (Figure 4.11) all show a similar fragmentation of the peaks in $ES^{+/-}$ detection mode, it is not possible to clearly assign the structures **164f** to one of the peaks **164a-e**. It is likely that significant fragmentation and recombination of all species occurs in the high energy source of the MS. Nevertheless, it is most likely that peaks **164a-e** all indicate mixtures of various ortho-quinomethide **165** derived structures.

163 - 167 are shunt metabolites arising from DHMBA **156**. Thus, the presence of **163 - 167** were considered as indirect evidence of prior formation for DHMBA **156** in the following work. The peaks of the shunt metabolites **164a-e**, which were observed repeatedly in the following chromatograms, are marked with an asterisk (*).

4.3.5 Analysis of the Absence of late-stage Intermediates

An additional inconsistency of the previous transformation experiments was the observation of only either DHMBA **156** or the reduction shunt **163** in the extracts (section 4.3.3.2), instead of the expected later intermediates when genes encoding later steps from the *lwt* pathway are present (Scheme 4.3).[177] As the pathway never gives any de-aromatised products, it was hypothesized that the FMO encoded by *lwtR1* may be broken or inactive. Therefore, a series of experiments to probe the presence of *lwtR1* in the heterologous expression were devised.

4.3.5.1 Transformant gDNA Screening for *lwtR1*

Transformants obtained by Dr. Jin Feng (Table 4-3) were analysed by the extraction of gDNA and subsequent PCR amplification of the expected genes with appropriate primers (Table 6-7). The PCR analysis (Figure 4.16) indicated that only one of the analysed transformants contained an intact *lwtR1* gene (Figure 4.16 A-B, transformants from Table 4-3, experiment G and I). Hence, the absence of later pathway products beyond DHMBA **156** is reasoned by the absence of the *lwtR1* gene, encoding the FMO TropB-homolog, which is necessary for tropolone formation according to the proposed biosynthetic pathway (Scheme 4.3).
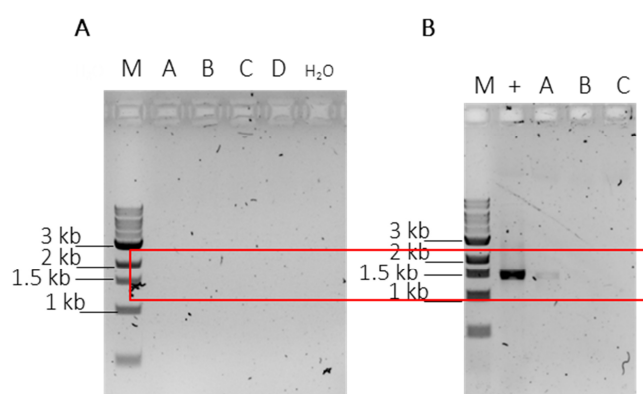
Figure 4.16 PCR amplification control of transformed genes with gDNA template: **A**, agarose gel of transfomants expected to contain *lwtS+R1+R2+R4+R6+R7* genes (experiment G); **B**, agarose gel of transfomants expected to contain *lwtS+R1+R6* genes (experiment I).

The analysis of the gDNA suggests that the integration of *lwtR1* into the host genome failed during the transformation process, and this likely explains why no late-stage

intermediates were observed. A possible rationale was speculated in which the *lwtR1* gene includes an unknown sequence pattern that may be particularly susceptible to digestion by host restriction enzymes to integrate the vector DNA into the genomic DNA. A detailed explanation of the failed integration requires further investigation. However, the main focus of the elucidation of later steps of the biosynthetic pathway was achieved by a different approach. A homolog of the FMO encoding gene *lwtR1*, *dbaH* from the antibiotic C pathway from *A. nidulans* (Figure 4.1, identity of 62 %) was chosen to complete the pathway for all following heterologous expression experiments.

4.3.5.2 Exchange of FMO LwtR1 by the Analogue DbaH

A. nidulans strain AGB 552, kindly provided by Dr. Jennifer Gerke (Braus group, Georg-August-University, Göttingen), was cultivated (details section 6.1.3.1) and mRNA was obtained by previously described methods. cDNA was synthesized and used for PCR amplification (primer DbaH_Peno_fw/DbaH_Teno_rev) to construct vectors for enzyme expression in the heterologous host (Figure 4.17). Yeast recombination was carried out in context of the Master's project of Leon Buttchereit. The three vectors pTY-*ade-lwtS-lwtR6-lwtR7-dbaH* (Figure 4.17 A), pTY-*ade-lwtS-lwtR6-dbaH,* and pTY-GS-*argB-lwtR2-lwtR4* (Figure 4.17 B) were constructed. The successful construction of the expression vectors was confirmed by PCR (Figure 4.17 C) and sequence analysis.
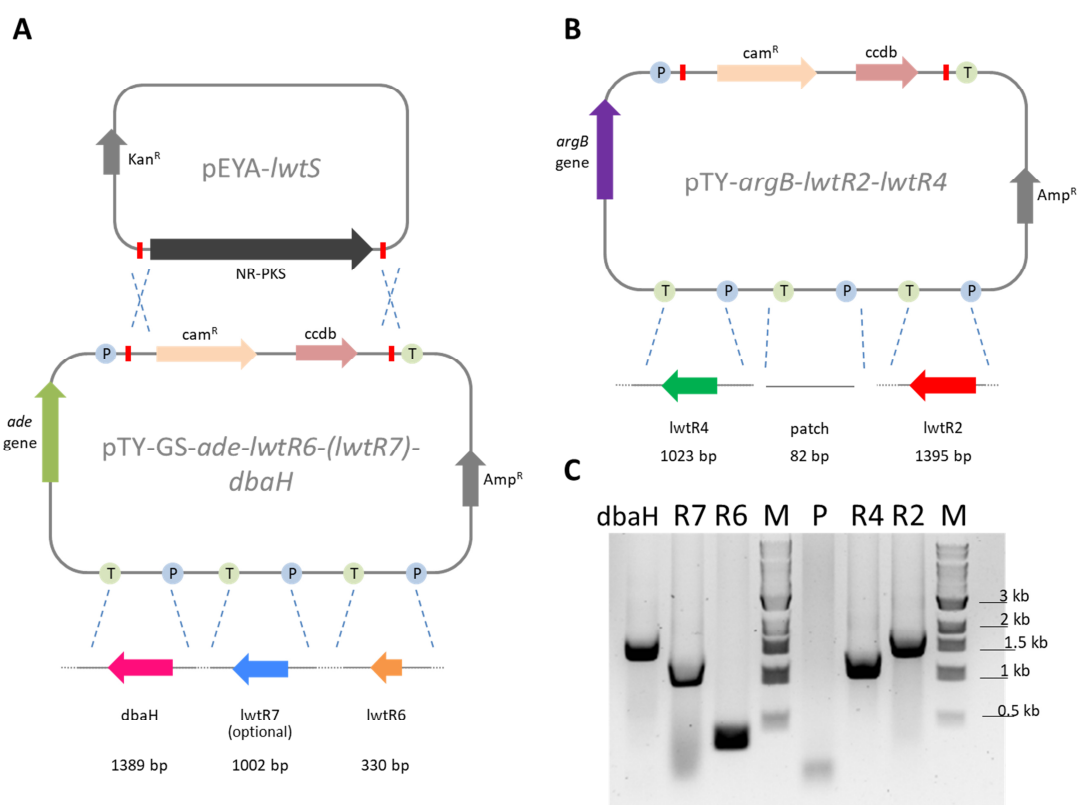
Figure 4.17 Vector construction for tropolone pathway: **A**, scheme of vector pTY-*ade-lwtS-lwtR6-lwtR7-dbaH*; **B**, scheme of vector pTY-GS-*argB-lwtR2-lwtR4*; **C**, PCR fragments for yeast recombination (P = patch).

4.3.6 Investigation of the *lwt*-BGC by Heterologous Expression

A systematic heterologous co-expression of the catalytic enzymes of the *lwt*-BGC was carried out to build up the pathway step by step, but using *dbaH* instead of *lwtR1*. The putative gene *lwtR6* encoding a putative YCII domain containing protein with unknown function was present in all heterologous expression experiments by default.

4.3.6.1 Co-Expression of LwtS + LwtR6 + DbaH

A combination of *lwtS* (encoding the nr-PKS), the *lwtR1* analogue *dbaH* (encoding the FMO), and *lwtR6* (encoding the YCII domain containing protein) were introduced into *A. oryzae* NSAR1 with the vector pTY-*argB-lwtS-lwtR6-dbaH*. The transformation was carried out in the context of Leon Buttchereit's master project. The transformation, selection and cultivation took place as previously described, including standardized acidification of the cultures before extraction. The analysis led to 15 transformants

producing pathway related compounds (Table 4-6) from a total of 24 selected transformants.

The early reduction shunt metabolites **164a-f** were detected in all producing transformants (Table 4-6, Figure 4.18 A, *), indicating the prior presence of **163**, as a reduction product of DHMBA **156** (Scheme 4.3). Additionally, DHMBA **156** was directly observed in two of the transformants.

Table 4-6 Summary of observed compounds of transformants containing *lwtS + lwtR6 + dbaH*.

Structure	Transformant														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
reduction shunt **164a-f**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DHMBA **156**	-	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-
azanidulone **159**	-	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	-	✓	-
sepedonin **90**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
anhydro-sepedonin **160**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
antibiotic C **91**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Ten transformants produced azanidulone **159**, which is a known shunt of the pathway (Figure 4.18 B-D, Table 4-6). **159** co-elutes with an unrelated compound. However, extracted ion chromatograms (Figure 4.23 B) enabled identification of traces of azanidulone **159** at 3.1 min with the help of the ES$^+$ *m/z* of 207 and a UV maxima at 219/337 nm (Figure 4.23 D), which corresponds to literature data by Gerke *et al.*[177] The presence of azanidulone **159** indicates indirectly the successful reaction of the FMO DbaH to **157** according to literature.[177] **157** leads to the formation of the hemiacetal **158** that dehydrates to azanidulone **159** (Figure 4.23 C). No other related compound was identified (Table 4-6, Figure 4.23 B).
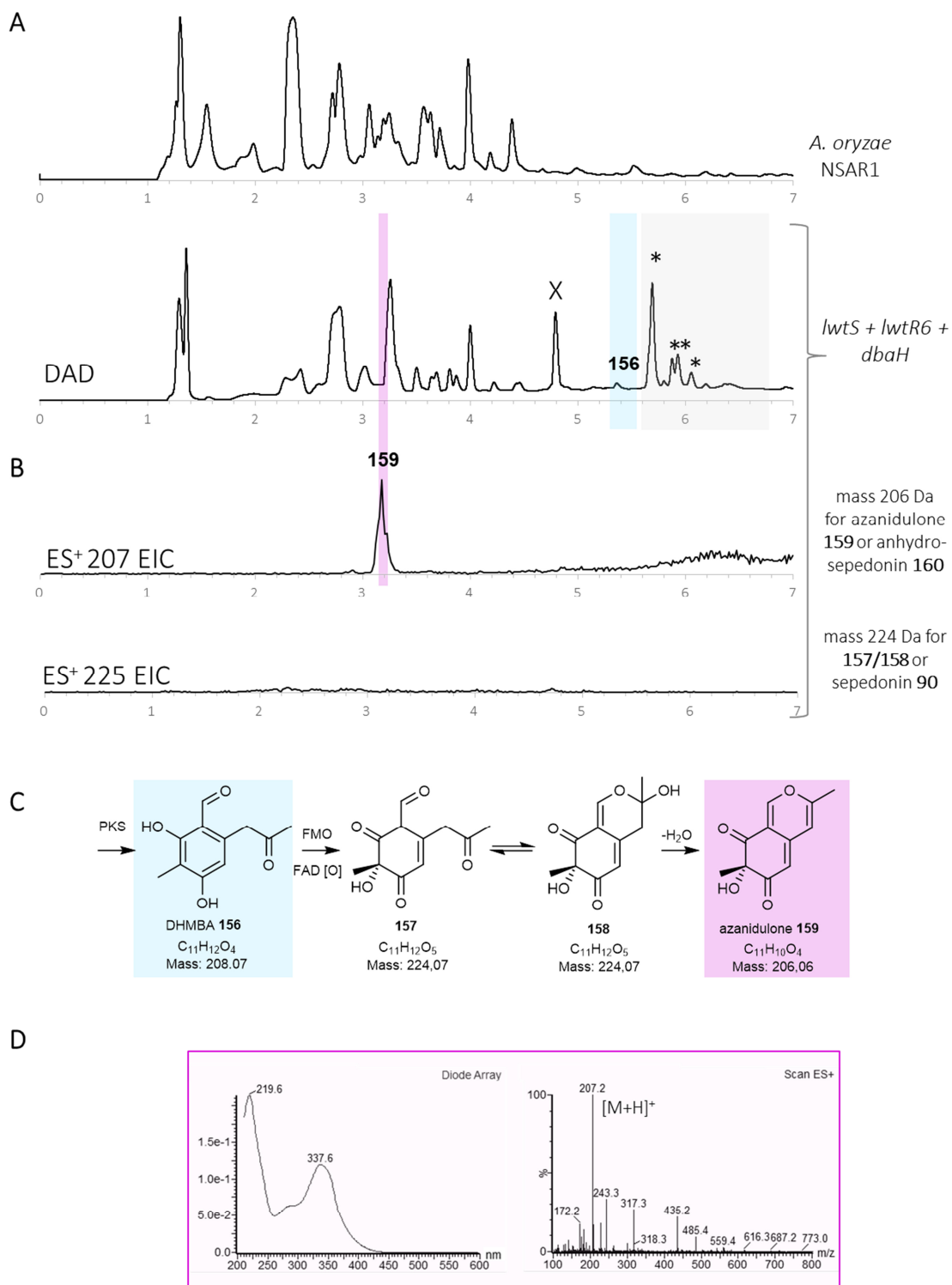
Figure 4.18 Heterologous expression of LwtS (PKS), DbaH (FMO) and LwtR6 (YCII): **A**, DAD chromatogram of *A.oryzae* NSAR1 and transformant (* = **164a-f**, X = unrelated); **B**, EIC for *m/z* 207 in ES⁺ mode **C**, proposed reaction of **156** to **159**. **D**, DAD and ES⁺ spectra of azanidulone **159**.

4.3.6.2 Co-Expression LwtS + LwtR6 + DbaH + LwtR7

Next, the *lwtS*, *lwtR6*, *dbaH* and *lwtR7* (encoding the NHI) genes were co-expressed (Figure 4.19) using vector pTY-*argB-lwtS-lwtR6-lwtR7-dbaH*. Twenty two transformants were obtained, of which 13 produced pathway intermediates. Similar to previous findings (with only LwtS, LwtR6, and DbaH), only early intermediates of the pathway were found (Figure 4.19 A, Table 4-7). No tropolones were observed. Early shunt peaks **164a-f** were observed in all 13 producing transformants. For four transformants, additionally DHMBA **156** was detected. Azanidulone **159** was the only additional compound found, detected in nine transformants (Table 4-7).
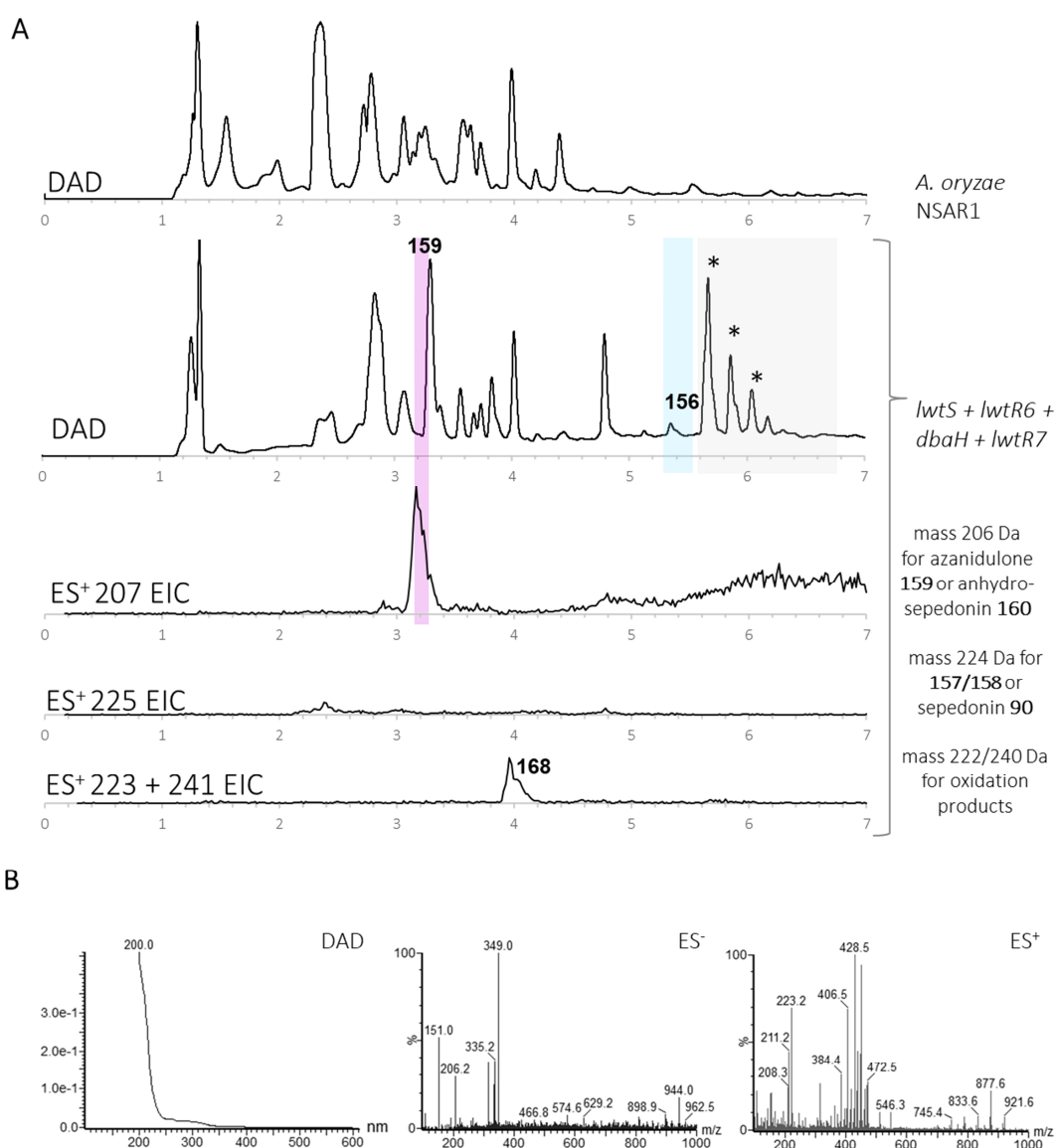

Figure 4.19 Heterologous Expression of LwtS (PKS), DbaH (FMO), LwtR6 (YCII) and LwtR7 (NHI): **A**, DAD chromatograms of *A. oryzae* NSAR1 and transformant with selected EIC of ES$^+$ (* = **164a-f**); **B**, characteristic data for peak **168**.

The α-ketoglutarate dependent dioxygenase encoded by *lwtR7* is hypothesized to catalyse the ring-expansion similar to TropC in the related tetraketide pathways like stipitatic acid **85** (Figure 1.13 B, section 1.4.3.2). Thus, additionally to the known intermediates of the pathway, all ES^+/ES^- chromatograms were screened for the hypothesized mass for the oxidation product of intermediate **157**. The oxidation of **157** (mass 224 Da) would lead to the mass of 240 Da (+16 Da) or 222 Da (-2 Da). A search in the EIC chromatogram results in peak **168** corresponding to the mass of 222 Da at 4 min (Figure 4.19 A). The UV absorbance maxima are not similar to tropolone-related compounds (Figure 4.19 B). The MS analysis does not give evidence for a relation of **168** to a pathway intermediate with a mass of 222 Da (Figure 4.19 B). Furthermore, also following results (*vide infra*, section 4.3.7.2) are in agreement with **168** not being an intermediate of the tropolone pathway. In conclusion, peak **168** is unrelated to the pathway. Overall, no tropolone-formation was detected in any of the transformants including the *lwtS*, *lwtR6*, *dbaH,* and *lwtR7*.

Table 4-7 Summary of observed compounds in transformants containing *lwtS + lwtR6 + dbaH + lwtR7*.

Structure	Transformant												
	1	2	3	4	5	6	7	8	9	10	11	12	13
reduction shunt **163** or **164a-f**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DHMBA **156**	✓	-	-	✓	-	-	-	-	-	-	✓	✓	-
azanidulone **159**	✓	-	✓	✓	-	✓	✓	✓	✓	-	-	✓	✓
sepedonin **90**	-	-	-	-	-	-	-	-	-	-	-	-	-
anhydrosepedonin **160**	-	-	-	-	-	-	-	-	-	-	-	-	-
antibiotic C **91**	-	-	-	-	-	-	-	-	-	-	-	-	-

4.3.6.3 Co-Expression LwtS + LwtR6 + DbaH + LwtR7 + LwtR2

Additional vectors were constructed similar to previously described methods (summarized in Table 6-5). The vectors pTY-*argB-lwtS-lwtR6-lwtR7-dbaH* and pTY-GS-*ade-lwtR2* were transformed together, followed by selection, cultivation and LCMS analysis as described previously. Only the early intermediates **156** (one transformant), the shunt peaks **164a-f** (eleven transformants), and **159** (seven transformants) were observed for eleven selected transformants (Figure 4.20, Table 4-8).

Table 4-8 Summary of observed compounds from transformants containing *lwtS + lwtR6 + dbaH + lwtR7+ lwtR2*.

Structure	Transformant										
	1	2	3	4	5	6	7	8	9	10	11
reduction shunt **163** or **164a-f**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DHMBA **156**	-	-	-	-	-	✓	-	-	-	-	-
azanidulone **159**	✓	✓	✓	-	✓	-	-	✓	✓	✓	-
sepedonin **90**	-	-	-	-	-	-	-	-	-	-	-
anhydrosepedonin **160**	-	-	-	-	-	-	-	-	-	-	-
antibiotic C **91**	-	-	-	-	-	-	-	-	-	-	-
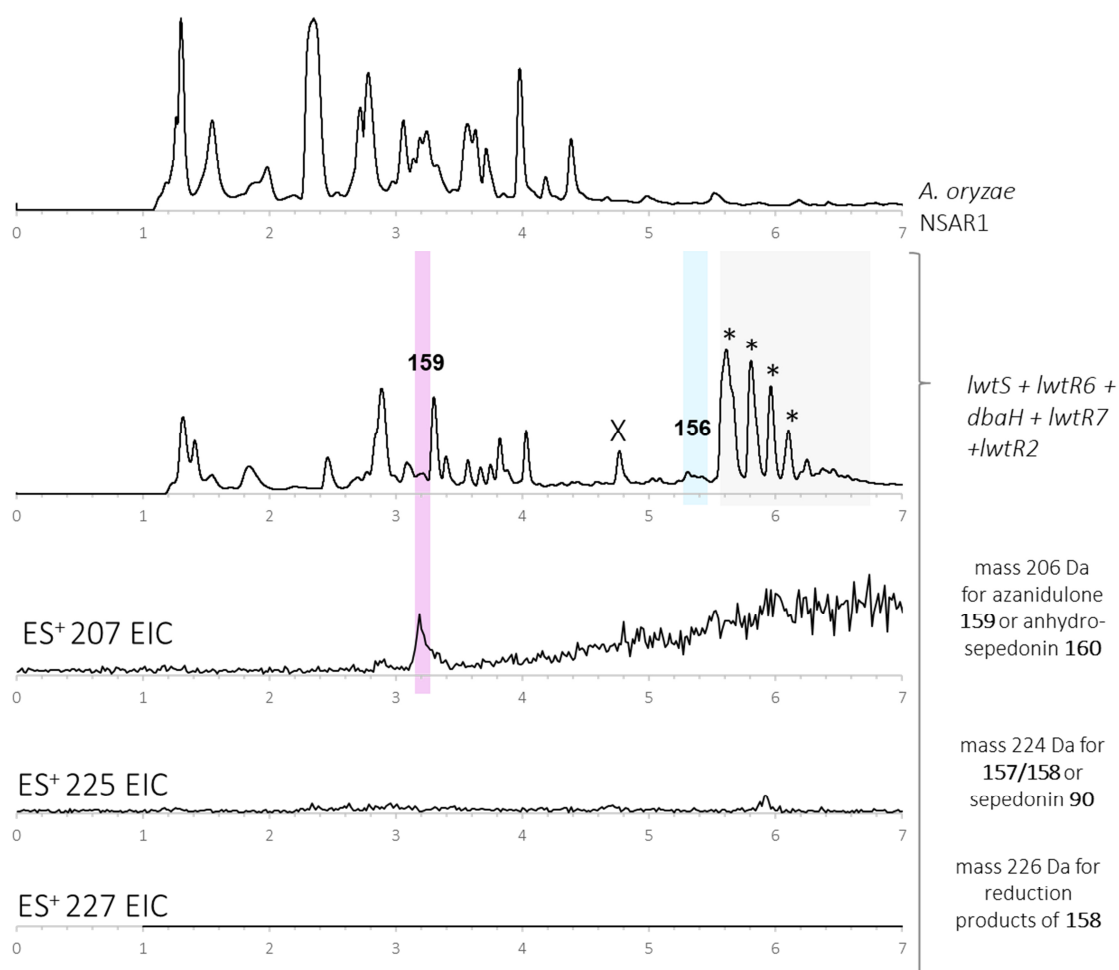
Figure 4.20 Heterologous Expression of LwtS (PKS), DbaH (FMO), LwtR6 (YCII), LwtR7 (NHI) and LwtR2 (putative oxidoreductase; * = **164a-f**, X = unrelated).

4.3.6.4 Co-Expression LwtS + LwtR6 + DbaH + LwtR7 + LwtR4

The heterologous co-expression of the genes *lwtS*, *lwtR6*, *dbaH*, *lwtR7* and *lwtR4* (encoding the MDR) led for the first time to the detection of tropolones (Figure 4.21).
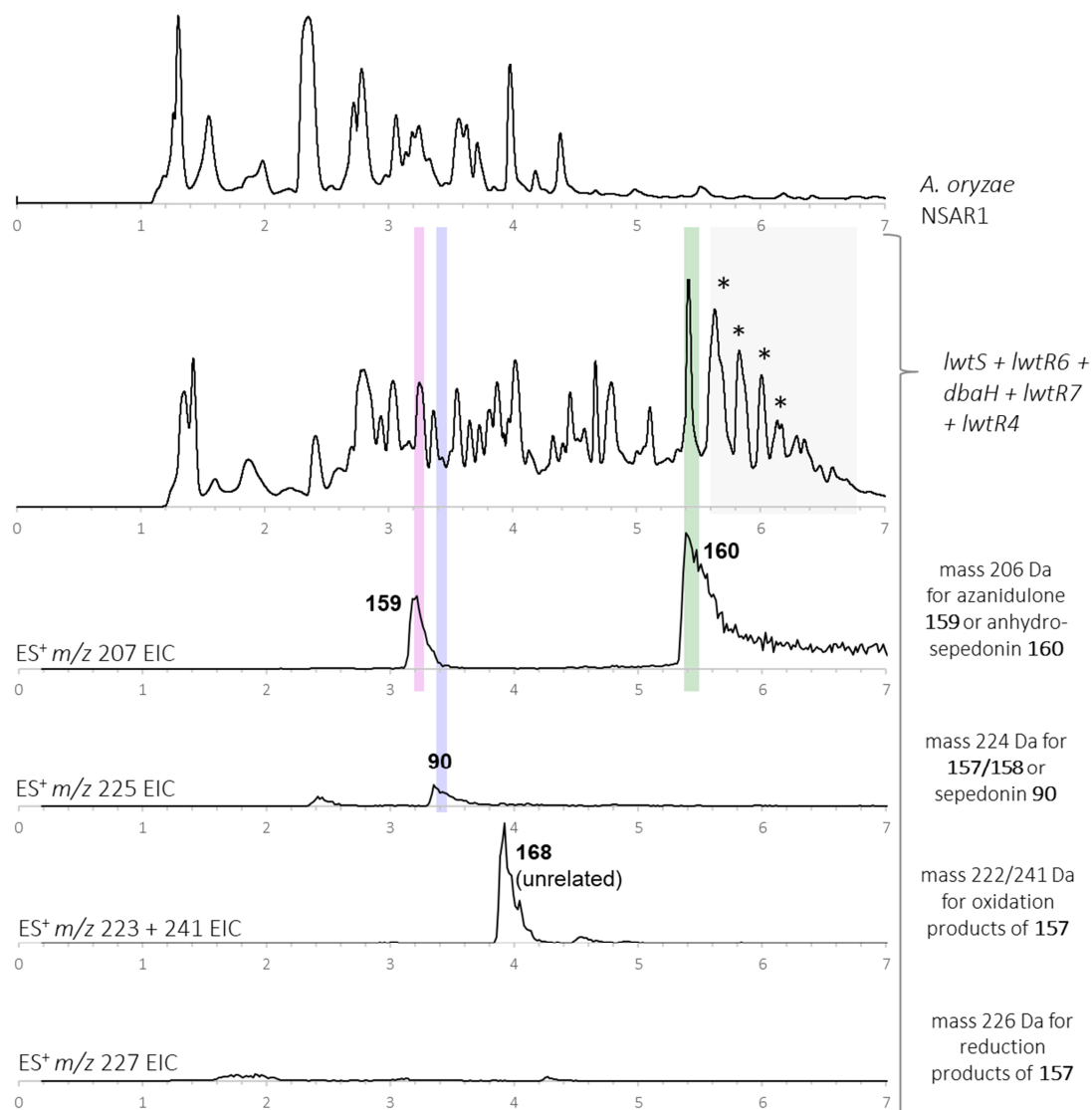
Figure 4.21 DAD chromatogram of LwtS, DbaH, LwtR6, LwtR7, and LwtR4 and EIC for masses 206, 224, 222, 240, 226 Da (* = **164a-e**).

The production of anhydrosepedonin **160** (Figure 4.21) was indicated by a peak at 5.4 min with a characteristic UV absorbance of 248/286/369 nm (Figure 4.22 A) and a mass of 206 Da with corresponding *m/z* of 205 ([M-H]$^-$) and 207 ([M+H]$^+$). Additionally, sepedonin **90** was discovered as a minor peak at 3.4 min with a similar UV spectrum to anhydrosepedonin **160** (210 nm/253 nm/327 nm) and a mass of 224 Da

(Figure 4.22 A). The identification of the compounds **160** and **90** was later confirmed by isolation and NMR analysis (section 4.3.6.6). Sepedonin **90** and anhydrosepedonin **160** were identified in the extract of two of three producing transformants (Figure 4.21). In one additional transformant, only **164a-f** and azanidulone **159** were found.
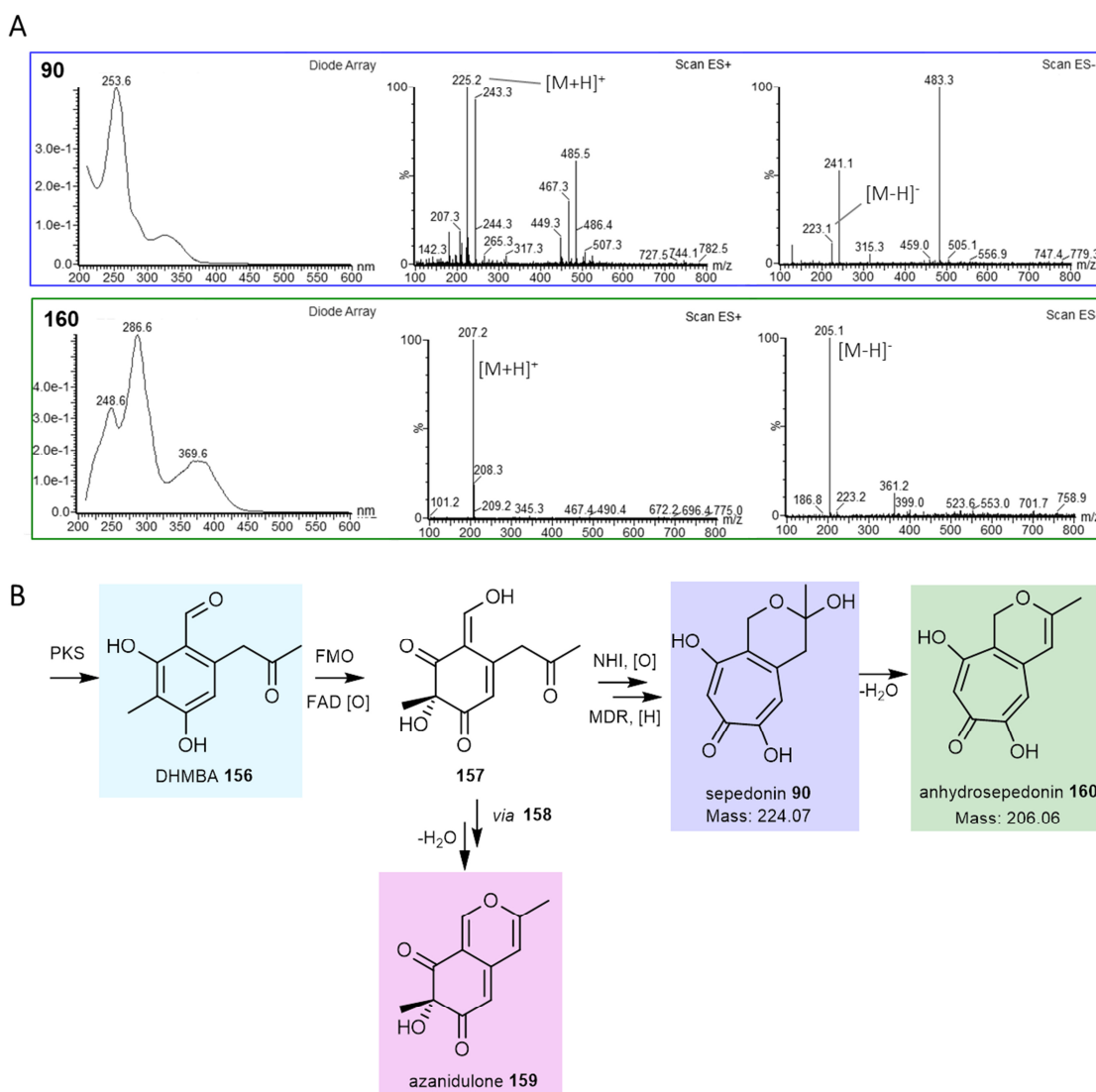
Figure 4.22 Formation of anhydrosepedonin **160** and sepedonin **90**: **A**, characteristic DAD, ES⁻, and ES⁺ spectra for sepedonin **90** and anhydrosepedonin **160**; **B**, proposed pathway.

Anhydrosepedonin **160** and sepedonin **90** were identified using the UV spectroscopic data and mass of known compounds from literature (Figure 4.22). However, possible other related intermediates were not observed. The successful high production of compounds related to sepedonin **90** appears to be associated with an increased number of peaks found in the extracts of transformants. This increase in peaks challenges the

identification (or later isolation) of compounds related to the sepedonin-pathway, due to overlapping peaks. The newly observed peaks could be linked to pathway intermediates or may potentially arise from host compounds that may also cross-react with pathway intermediates. The identification of potential intermediates of the sepedonin pathway will be investigated by *in vitro* studies in the following section 4.3.7.

4.3.6.5 Heterologous Expression of the Full Gene Cluster

All putative catalytic enzymes of the *lwt*-BGC (Figure 4.1) were introduced into *A. oryzae* NSAR1 by transformation with pTY-*argB-lwtS-lwtR6-lwtR7-dbaH* and pTY-GS-*ade-lwtR2-lwtR4,* carried out as part of the Master's project of Leon Buttchereit. The obtained transformants (Figure 4.23 A) included all putative catalytic genes of the cluster (Table 4-1), encoding the nr-PKS (LwtS), YCII-domain containing protein (LwtR6), FMO (DbaH), NHI (LwtR7), FAD-dependent oxidoreductase (OR, LwtR2) and MDR (ER-like, LwtR4). Overall, 13 transformants were obtained, of which nine produced compounds related to the pathway (Table 4-9, Figure 4.23 C).

Table 4-9 Summary of observed compounds in transformants including all putative *lwt*-BGC genes encoding catalytic enzymes.

Structure	Transformant								
	1	2	3	4	5	6	7	8	9
reduction shunt **163** or **164a-f**	-	✓	-	✓	✓	✓	✓	✓	✓
DHMBA **156**	-	-	✓	-	-	✓	✓	-	-
azanidulone **159**	-	✓	-	✓	✓	✓	✓	✓	✓
sepedonin **90**	-	-	-	-	-	-	-	✓	✓
anhydrosepedonin **160**	✓	✓	-	-	✓	✓	✓	✓	✓
antibiotic C **91**	-	-	-	-	-	-	-	-	-

Two transformants produced only DHMBA **156**, azanidulone **159** or **164a-f**, which suggests an incomplete gene manifest for these transformants (Table 4-9, transformant 3, 4). For all other transformants (Table 4-9), anhydrosepedonin **160** was the main pathway product (Figure 4.23 A, B). For two transformants, sepedonin **90** was additionally detected (Figure 4.23 A, B). No antibiotic C **91** was detected in the extracts of the transformants, indicating the set of enzymes is not sufficient for its biosynthesis (Figure 4.23 A, B). Although a peak **168** at 4 min has a mass of 222 Da (identical to the

mass of antibiotic C **91**), its retention time and UV absorbance are not in agreement with known data for antibiotic C **91**.[177] However, antibiotic C **91** is suggested to be an intermediate of the lienhwalide pathway, as *O*-methylated antibiotic C **91** forms part of the lienhwalide B **131** and C **132** structures.
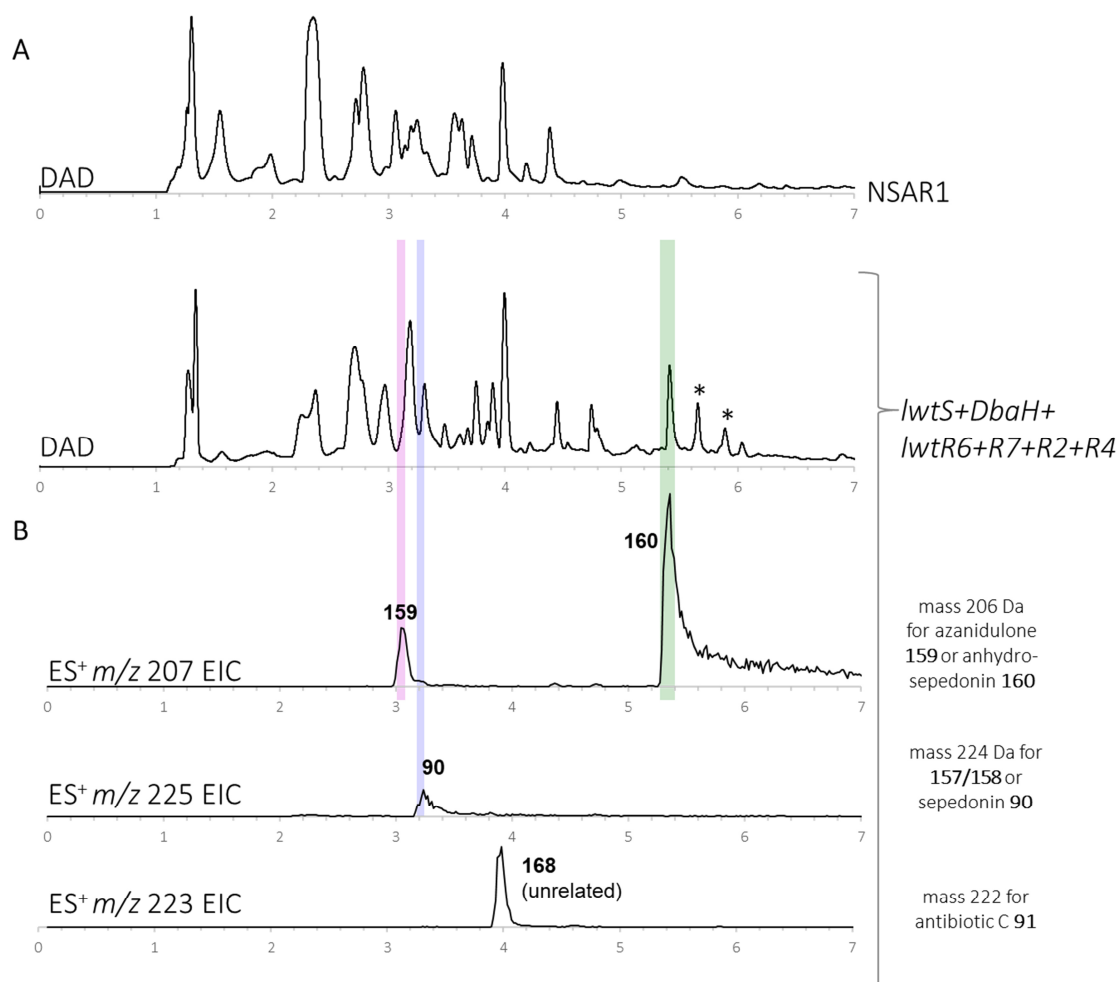
Figure 4.23 Heterologous expression of LwtS (PKS), DbaH (FMO), LwtR7 (NHI), LwtR6 (YCII), LwtR2 (OR) and LwtR4 (MDR): **A**, DAD chromatogram of *A. oryzae* NSAR1 and transformant; **B**, extracted ion chromatogram for *m/z* 207, 225, and 223 in ES$^+$ mode (* = **164a-e**).

4.3.6.6 Compound Isolation and NMR Analysis of Pathway Products

To fully confirm and characterize the structures of anhydrosepedonin **160**, the transformant shown in Figure 4.21 was fermented at greater scale (2.4 L) and **160** was obtained by preparative LCMS to give 26.6 mg (11.1 mg/l) of a yellow powder eluting at 5.4 min. The LCMS analysis of the purified compound showed it to be an equilibrating mixture (Figure 4.24) of **90** and **160**.
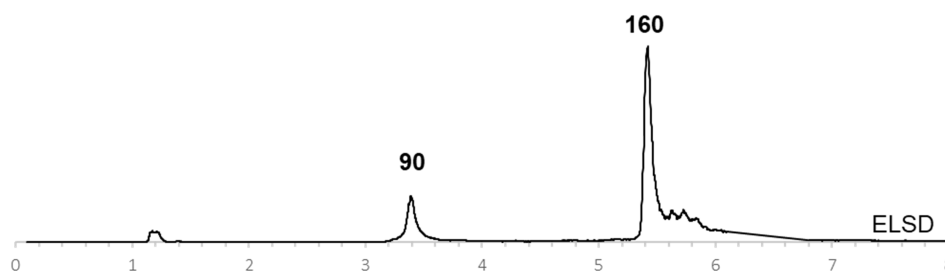
Figure 4.24 LCMS chromatogram of the obtained putative anhydrosepedonin fraction with ELSD detector.

The characteristic UV/ES-/ES+ data indicated a mixture of anhydrosepedonin **160** as the main compound, accompanied by minor amounts of sepedonin **90**. The mixture was analysed by NMR and this confirmed anhydrosepedonin **160** (Figure 4.25, Table 4-10) as the main compound. Sepedonin **90** was the minor compound in the mixture (Figure 4.25, Table 4-11).

90 and **160** were identified from the extract of the transformant, therefore verifying the proposed biosynthetic pathway. The presence of sepedonin **90** and anhydrosepedonin **160** in the purified fraction suggest a spontaneous inter-conversion of the two compounds.

Table 4-10 Chemical shifts of anhydrosepedonin **160** in DMSO (600 MHz).

Atom	C-type	δ_C/ppm	δ_H/ppm (mult)	HMBC	Literature[177] in DMSO-d_6	
					δ_C/ppm	δ_H/ppm
1	CH$_2$	64.6	5.00 (s)	2, 3, 8, 10	64.5	5
2	C	112.8			112.7	
3	C	162.1			161.6	
4	CH	111.3	6.77 (s)	1, 2, 3, 5/6	111.2	6.73
5	C	169.7			169.9	
6	C	165.9			165.8	
7	CH	112.4	6.58 (s)	2, 5/6, 9	112.4	6.59
8	C	140.9			140.9	
9	CH	105.3	5.72 (s)	2, 10, 11	105.2	5.73
10	C	161.1			161.2	
11	CH$_3$	18.6	1.92 (s)	9, 10	19.1	1.93

A

red: anhydrosepedonin **160** blue: sepedonin **90**

B

= HMBC (H to C)

anhydrosepedonin **160**
Chemical Formula: $C_{11}H_{10}O_4$
Mass: 206.06

sepedonin **90**
Chemical Formula: $C_{11}H_{12}O_5$
Mass: 224.07
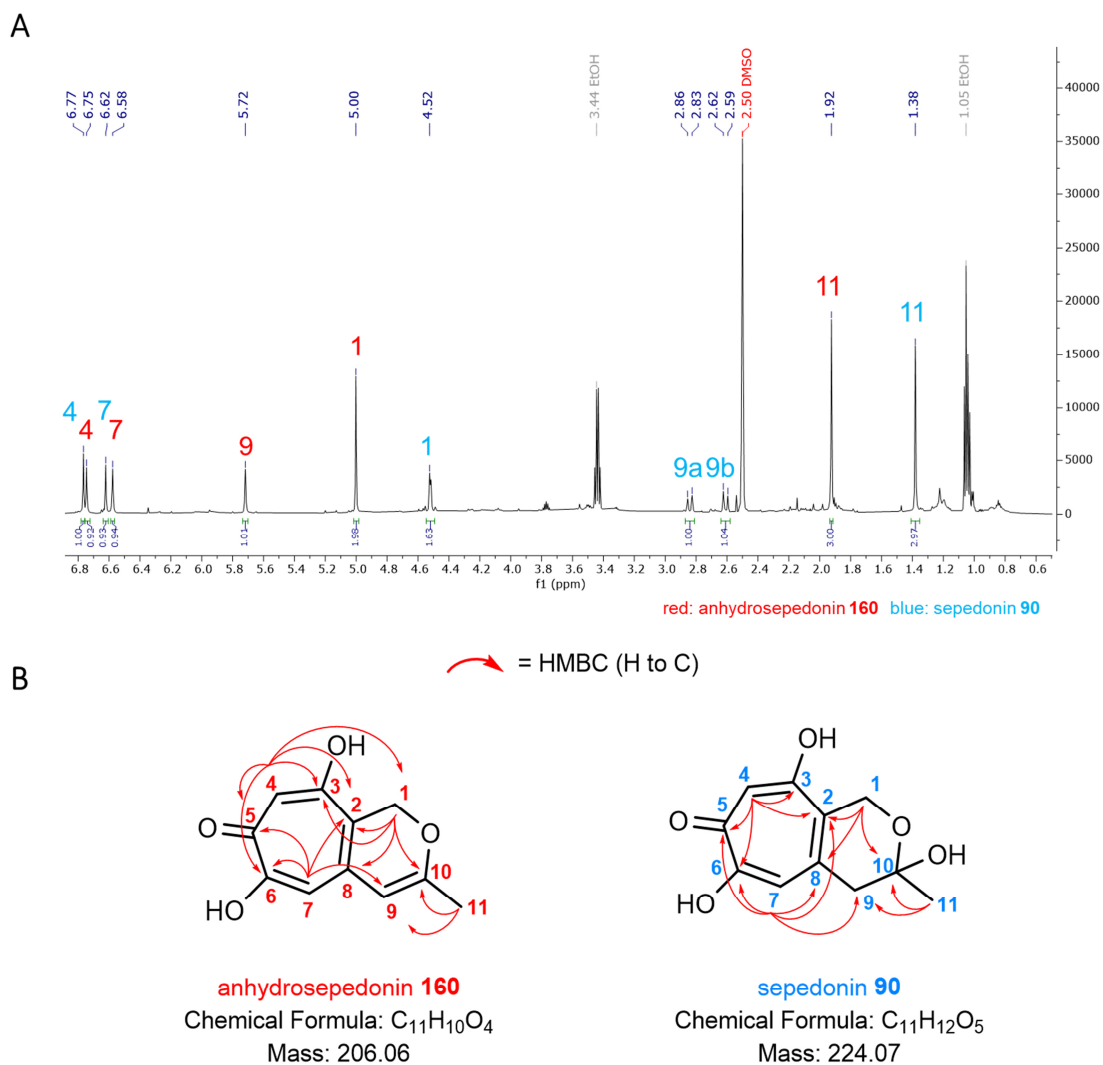
Figure 4.25 NMR analysis of **160** and **90**: **A**, ^{1}H NMR (600 MHz) of anhydrosepedonin **160** and sepedonin **90** in DMSO; **B**, identified structure **160** and **90**.

Table 4-11 Chemical shifts of sepedonin in DMSO (600 MHz).

| Atom	C-type	δ_C/ppm	δ_H/ppm (mult)	HMBC	Literature[97,184,185] in pyridine-*d*5	
					δ_C/ppm	δ_H/ppm
1	CH$_2$	59.4	4.52 (s)	2, 8, 10	60.6	5.09
2	C	126.9			-	
3	C	163.1			166.0	
4	CH	112.9	6.75 (s)	2, 3, 5, 6	113.5	7.15
5	C	173.1			174.3	
6	C	160.7			161.6	
7	CH	115.2	6.62 (s)	2, 5, 6, 8, 9	115.6	6.86
8	C	140.0			-	
9	CH$_2$	43.1	a: 2.61 (d, J = 17.4) b: 2.84 (d, J = 17.5)	2, 7, 8, 10	44.0	2.92
10	C	92.9			93.6	
11	CH$_3$	28.3	1.38 (s)	9, 10	29.0	1.72

4.3.6.7 Investigation of Late Tailoring Steps and Modifications

Lienhwalide B **131** and C **132** contain *O*-methylated forms of antibiotic C **91**. Therefore, the anhydrosepedonin-core structure must undergo oxidation and methylation for the formation of the lienhwalides.

Formation of Antibiotic C

While heterologous biosynthesis of anhydrosepedonin **160** was detected in the previous transformation experiments (section 4.3.6.5), the questions of how antibiotic C **91** is formed as a possible intermediate for lienhwalides B **131** and C **132** remains open. Antibiotic C **91** is presumably an intermediate of lienhwalide biosynthesis, which is suggested to be the product of the oxygenation of anhydrosepedonin **160**, according to the proposed pathway by Gerke *et al.* (Scheme 4.1). In the dba pathway from *A. nidulans* anhydrosepedonin **160** is oxidized to antibiotic C **91** by the FMO DbaB.[177] The *lwt*-BGC does not include a homologue to DbaB (Figure 4.1). A putative cytochrome P450 oxygenase encoding gene was found approx. 5 kbp upstream of the nr-PKS gene *lwtS* (Figure 4.26) when searching for genes that encode oxygenases.
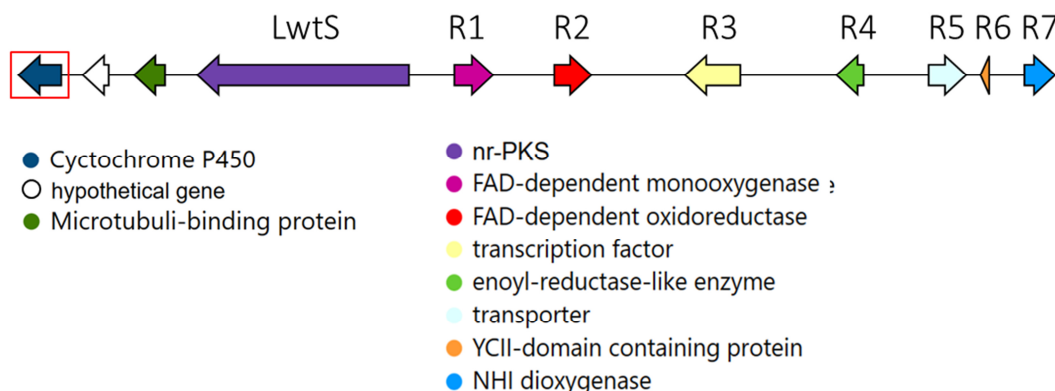
Figure 4.26 *lwt*-BGC of *H. lienhwacheense*.

The putative cytochrome P450 oxygenase gene was transformed together with the enzymes for formation of the tropolone core. The vector pTY-*ade-lwtR4-P450* was constructed by similar methods as already described. Appropriate primer pairs (P450HL_fw/P450HL_rev) were used for cDNA amplification to exclude introns. The putative gene encoding the cytochrome P450 was inserted into the pEYA vector and LR-recombination into the already obtained pTY-GS-*ade-lwtR4* vector (Table 6-5) was carried out. The vector pTY-*argB-lwtS-lwtR6-lwtR7-dbaH* was used additionally to the newly constructed vector pTY-*ade-lwtR4-P450*.

After transformation, selection and cultivation by usual methods, no formation of antibiotic C **91** was observed for the six producing (from overall eleven) obtained transformants (Table 4-12, Figure 4.27). Mostly early pathway products **156** and **163**/**164a-f** were detected. The late pathway product anhydrosepedonin **160** was only observed in traces for one of the six transformants (Table 4-12). The peak of **160** overlaps with one of the typical shunt peaks **164g**, which characterized with similar *m/z* values of 192 and a UV absorbance maximum of 237/282 nm. **164g** is most likely related to the identified shunt compounds **163 – 166** (section 4.3.4.5), but co-elutes with anhydrosepedonin **160** in previously shown chromatograms.
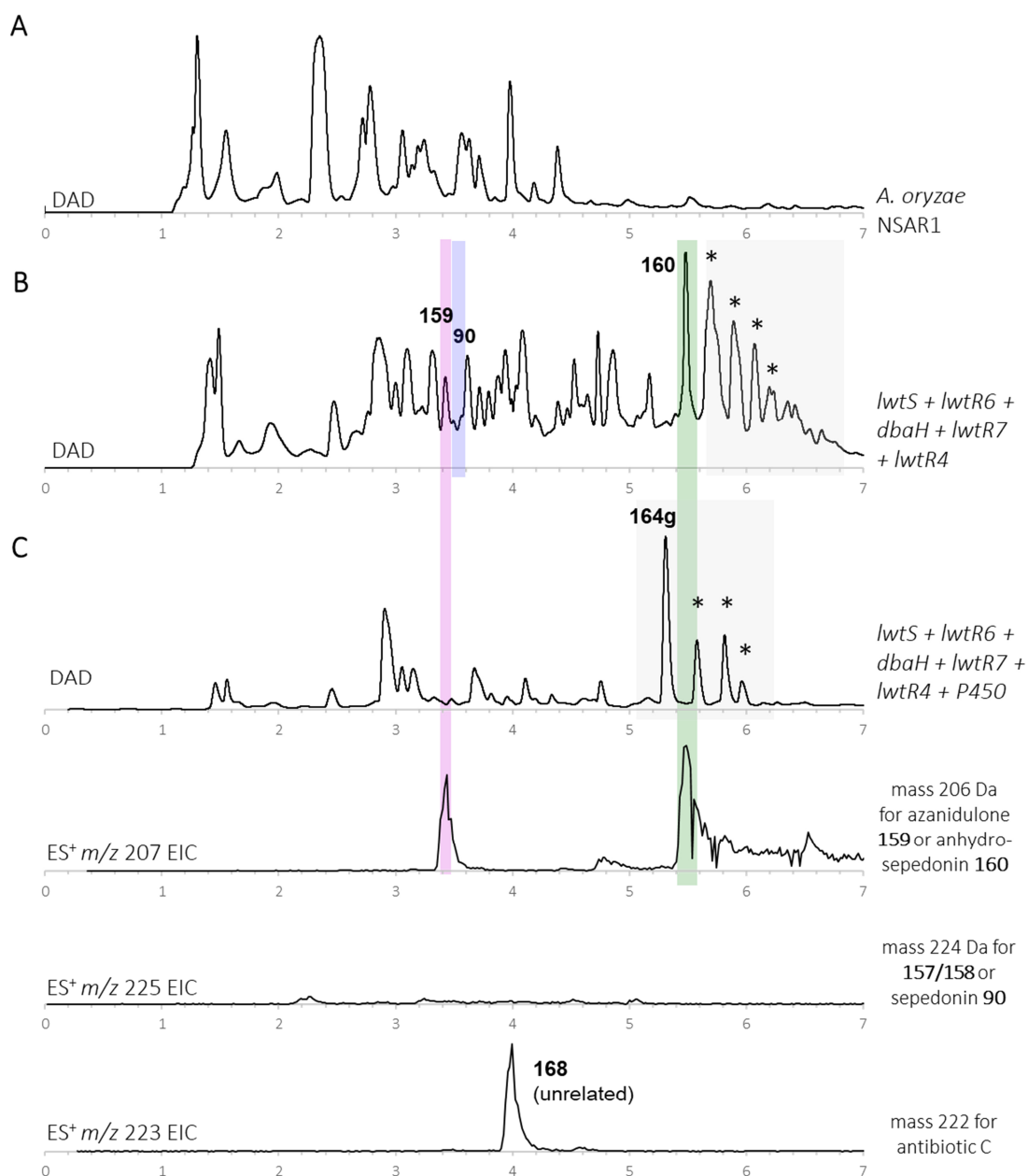
Figure 4.27 Chromatograms of heterologous expression with LwtS (PKS), DbaH (FMO), LwtR6 (YCII), LwtR7 (NHI), LwtR4 (MDR), and P450: **A**, DAD chromatogram of *A.oryzae* NSAR1; **B**, DAD chromatogram of transformant including *lwtS + lwtR6 + dbaH + lwtR7 + lwtR4*; **C** ES⁺ as TIC and EIC for transformant including *lwtS + lwtR6 + dbaH + lwtR7 + lwtR4 + P450* (* = **164a-e**).

Sepedonin derived compounds such as antibiotic C **91** have antifungal properties.[179] It was hypothesized that the transformants do not survive the presence of the toxic antibiotic C. Additionally, it was speculated that antibiotic C **91** production might be improved by transporting it outside the cells after biosynthesis. Thus, the transport protein LwtR5 could serve as a resistance mechanism against **91**. Therefore, the transporter gene *lwtR5* (Figure 4.28) was added to the transformation system (using

primers lwtR5_Pgpda/lwtR5_Teno). The vector pTY-*argB-lwtS-lwtR6-lwtR7-dbaH* was used in combination with the newly constructed vector pTY-*ade-lwtR4-P450-lwtR5* for transformation.

After transformation, selection, cultivation and extraction as described previously, the five obtained producing transformants (from overall seven transformants) were further analysed. The inclusion of *lwtR5* into the transformation experiment did not lead to the detection of antibiotic C **91** (Figure 4.28, Table 4-12).

Table 4-12 Summary of observed compounds for co-transformation of *lwtS + lwtR6 + dbaH + lwtR7+ lwtR4 + P450 (+lwtR5)*.

Structure	Transformant with: *lwtS + lwtR4 + R6 + R7 + dbaH*										
	+ P450						+lwtR5 + P450				
	1	2	3	4	5	6	1	2	3	4	5
reduction shunt **163/164a-f**	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-
DHMBA **156**	-	-	-	-	✓	-	-	-	-	-	-
azanidulone **159**	✓	-	-	-	-	-	✓	-	✓	-	-
sepedonin **90**	-	-	-	-	-	-	✓	-	✓	-	-
anhydrosepedonin **160**	✓	-	-	-	-	-	✓	-	✓	✓	✓
antibiotic C **91**	-	-	-	-	-	-	-	-	-	-	-

The titres of sepedonin **90** were increased by the introduction of the transporter gene *lwtR5* **91**. The LCMS sample was diluted by the factor 8 in comparison to all samples measured in this work. Nevertheless, the DAD chromatogram is overloaded (Figure 4.28). A comparison of the anhydrosepedonin **160** production of a transformant including the transporter encoding gene *lwmR5* (Figure 4.28) with the highest producing previously obtained transformant (including *lwtS, lwtR6, dbaH, lwtR7, lwtR4*; section 4.3.6.4, Figure 4.21) was carried out. The **160** production was increased by the factor 5 in DAD chromatogram peak integral comparison. Taking into consideration that the sample is diluted, the anhydrosepedonin **160** production is at least approx. 40-fold increased. The increase of anhydrosepedonin **160** production is most likely higher than 40-fold, as the DAD chromatogram (Figure 4.28) is overloaded.
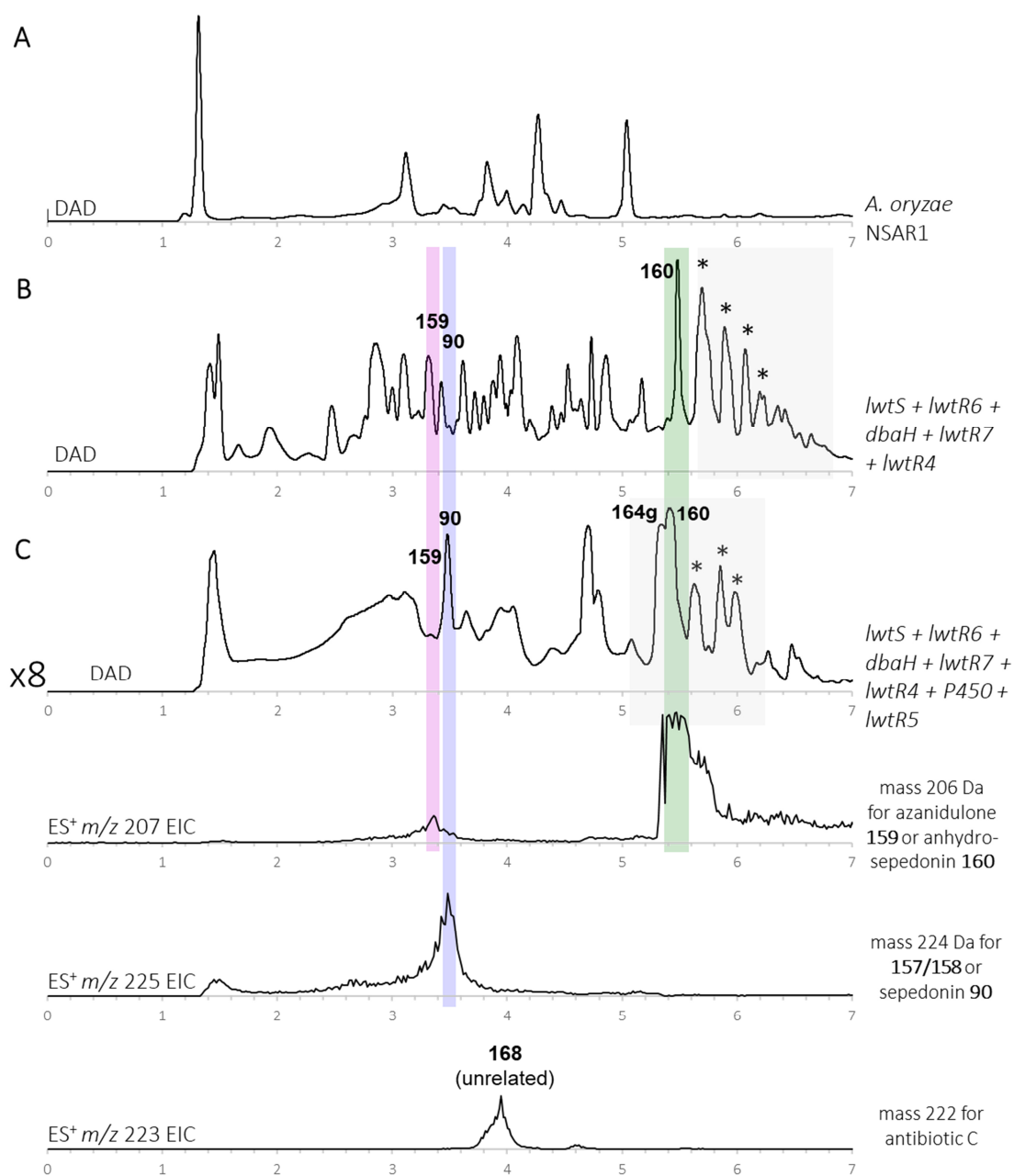
Figure 4.28 Chromatograms of heterologous expression with LwtS, DbaH, LwtR6, LwtR7, LwtR4, P450, and LwtR5 (transporter): **A**, DAD chromatogram of *A.oryzae* NSAR1; **B**, DAD chromatogram of transformant including *lwtS + lwtR6 + dbaH + lwtR7 + lwtR4*; **C** transformant including *lwtS + lwtR6 + dbaH + lwtR7 + lwtR4 + P450 + lwtR5* (* = **164a-f**).

O-Methylation of Tropolones

In addition to late pathway oxidation that forms antibiotic C, *O*-methylation must occur for lienhwalides B **131** and C **132**. No gene encoding a putative methyltransferase is located near the *lwt*-BGC. It was hypothesized that tropolone structures may undergo spontaneous *O*-methylation by the treatment of the compounds with the solvent methanol during extraction or compound isolation process.

A mixture of anhydrosepedonin **160** and sepedonin **90** was dissolved in methanol and in parallel in acetonitrile (ACN), to test if the tropolones **160** and **90** are able to be methylated by the solvent methanol. Tropolone intermediates sepedonin **90** and anhydrosepedonin **160** are stable against methylation by the solvent methanol (Figure 4.29). Therefore, it was excluded that methylation arises from treatment of compounds with methanol.
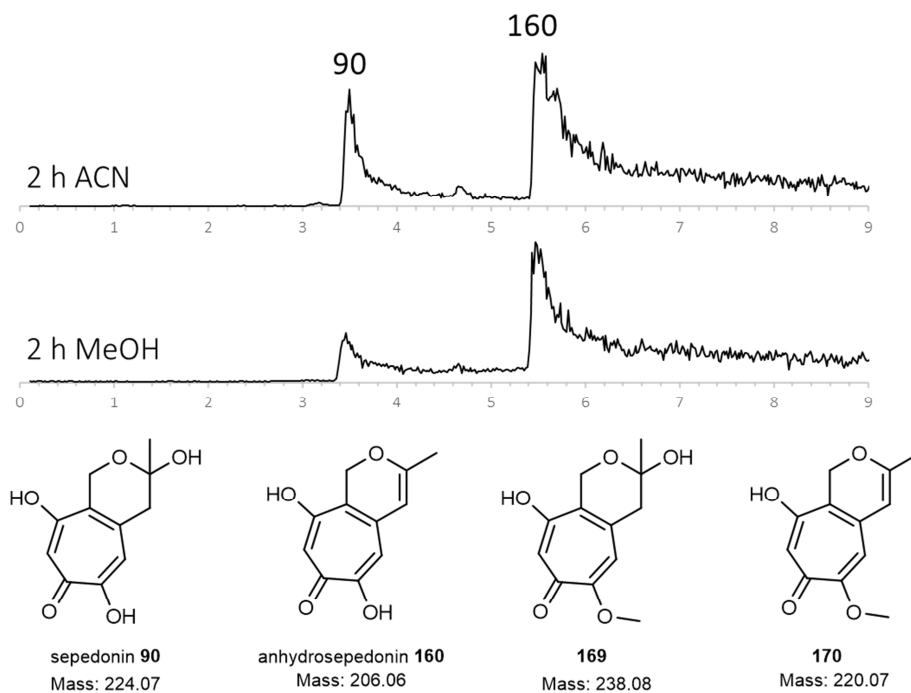
Figure 4.29 Comparison of tropolone stability in acetonitrile and methanol for 2 h; ES$^+$ EIC for *m/z* 207 (**90**) + 225 (**160**)+ 221 (**170**) + 239 (**169**).

4.3.6.8 Summary of Heterologous Expression Experiments

The results of the above experiments (summary Table 4-13) confirmed that genes *lwtS*, *lwtR6*, *dbaH*, *lwtR7*, and *lwtR4* are sufficient to encode the biosynthesis of the tropolone compounds **90** and **160**. The role of the oxidoreductase encoded by *lwtR2* remains unknown. No other intermediates of the pathway were found.

Table 4-13 Summary of heterologous expression experiments.

	putative *lwt* genes								163 reduced shunt	156 DHMBA	159 aza-nidulone	90 sepe-donin	160 anhydro-sepe-donin	91 anti-biotic C
	S	R1	R2	R4	R5	R6	R7	P-450						
A	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-
B	✓	✓	-	-	-	✓	-	-	✓	✓	✓	-	-	-
C	✓	✓	-	-	-	✓	✓	-	✓	✓	✓	-	-	-
D	✓	✓	✓	-	-	✓	✓	-	✓	✓	✓	-	-	-
E	✓	✓	-	✓	-	✓	✓	-	✓	✓	✓	✓	✓	-
F	✓	✓	✓	✓	-	✓	✓	-	✓	✓	✓	✓	✓	-
G	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-

## 4.3.7 Investigation of the *lwt*-BGC by *in vitro* Assays

The heterologous expression approach led to a high number of shunt metabolites in the LCMS chromatograms (section 4.3.4, **163**, **164a-g**, **166**) that potentially mask the presence of putative intermediates. Furthermore, the ortho-quinomethide shunt intermediate **165** is reactive and may not only form dimers (like **164f**), but also potentially react with host metabolites. This also impedes easy identification of important sepedonin pathway intermediates. While the core genes of the sepedonin formation were found by the heterologous expression approach, the intermediates of the pathway remain unclear. For example, the results show that both, an oxidation and reduction are required for sepedonin formation, but the order and intermediates are unknown.

Additionally, it was not possible to express *lwtR1* from the *lwt*-BGC of *H. lienhwacheense* in the heterologous host *A. oryzae* NSAR1. To confirm the obtained results (with LwtR1 instead of DbaH) and to further investigate the order of the pathway reactions and the intermediates, *in vitro* assays were carried out.

4.3.7.1 Recombinant Protein Isolation and *in vitro* Assay Setup

The genes *lwtR1* (FMO), *lwtR7* (NHI) and *lwtR4* (reductase) were inserted into pET28a(+) expression vectors in previous work by Dr. Jin Feng. The proteins were expressed and purified from *E. coli* BL21 cultures using Tris-HCl buffers, with methods described previously (Table 6-2, Figure 4.30, section 6.1.4.3).
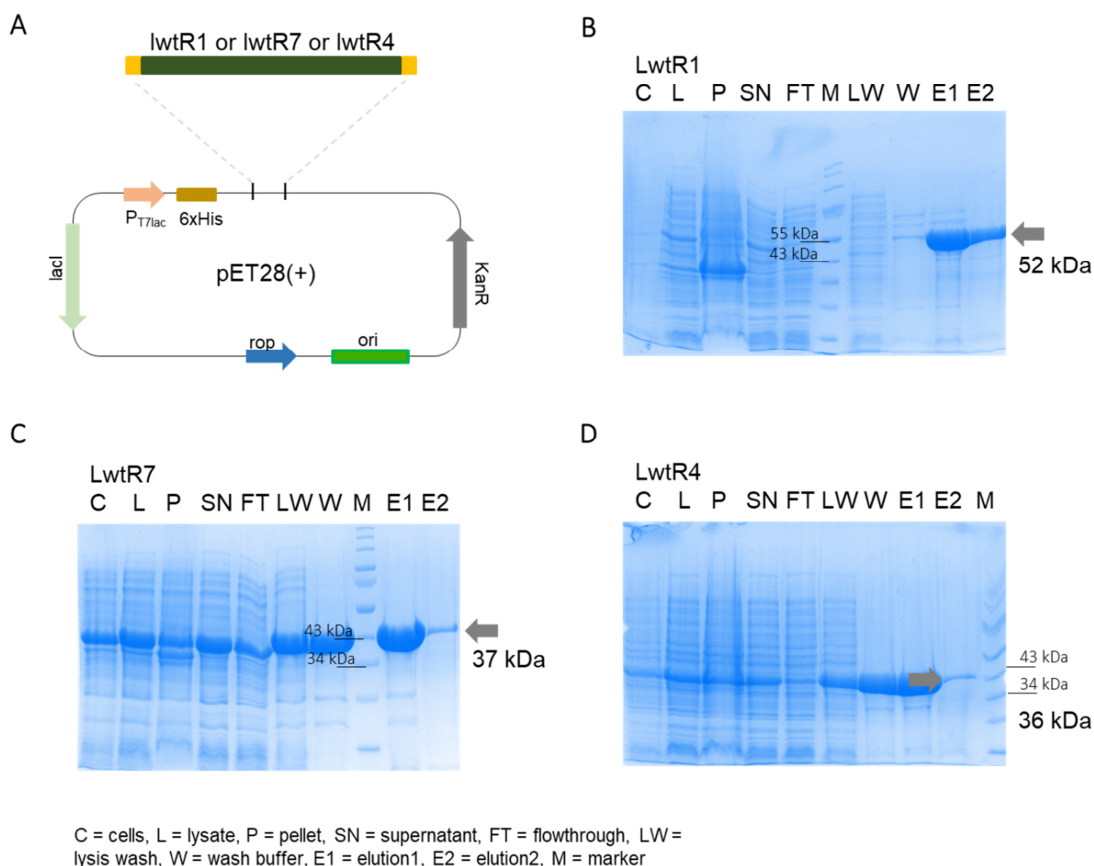
Figure 4.30 Protein expression and purification of LwtR1, LwtR7 and LwtR4: **A**, vector construction scheme; **B**, purification of LwtR1; **C**, purification of LwtR7; **D**, purification of LwtR4.

DHMBA **156** was purified from producing *A. oryzae* NSAR1 transformants (Figure 4.8) by Dr. Jin Feng as the substrate for *in vitro* investigations. DHMBA was incubated with appropriate cofactors for four differing combinations of proteins (LwtR1 [experiment I], LwtR1 + LwtR7 [experiment II], LwtR1 + LwtR4 [experiment III], LwtR1 + LwtR4 + LwtR7 [experiment IV], Table 4-14) to enable a sequential reaction of the intermediates without purification steps in between the reactions.

Table 4-14 Composition of Lwt-enzymes *in vitro* assays.

final concen-tration	component	experiment			
		I LwtR1	**II** LwtR1+R7	**III** LwtR1+R4	**IV** LwtR1+R4+R7
2.5 mM	DHMBA **156**	✓	✓	✓	✓
5 mM	α-ketoglutaric acid	-	✓	✓	✓
8 mM	ascorbate	-	✓	✓	✓
0.1 mM	FeSO$_4$	-	✓	✓	✓
1 mM	NADPH	✓	✓	✓	✓
5 mM	glucose-6-phosphate	✓	✓	✓	✓
1 U/ml	glucose-6-phosphate dehydrogenase	✓	✓	✓	✓
5 mM	TES buffer (1 M, pH 7.5)	✓	✓	✓	✓
10 µM	LwtR1	✓	✓	✓	✓
10 µM	LwtR4	-	-	✓	✓
10 µM	LwtR7	-	✓	-	✓
ad 50 µl	H$_2$O	✓	✓	✓	✓

In addition to the Lwt-proteins and appropriate cofactors, an NAD(P)/NAD(P)H recycling system was introduced to the *in vitro* assay composition, to improve the turnover of the reaction.[186] Therefore, glucose-6-phosphase (G6P) dehydrogenase and G6P were added into the reaction mixture (Scheme 4.5), analogous to similar literature-known approaches that provide a steady supply of NAD(P)H.[187,188]
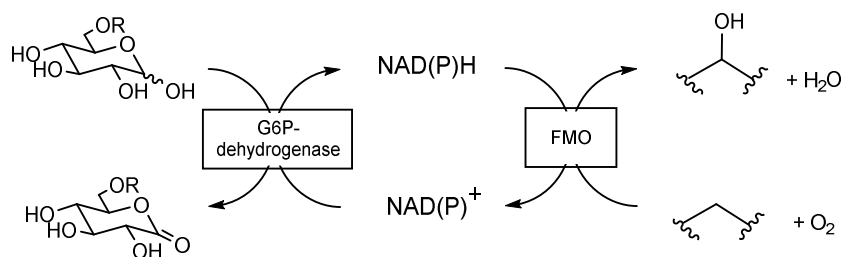
Scheme 4.5 Concept of NAD(P)H recycling system.

After 2 h incubation, 100 µl CH₃CN was added to the reaction mixture, the protein was precipitated by centrifugation, and the supernatant was examined directly by LCMS.

4.3.7.2 Results of *in vitro* Assays

All *in vitro* assays were analysed by LCMS by EIC search in ES$^+$ mode for the *m/z* 209 (mass of 208 Da for DHMBA **156**), 207 (mass of 206 Da for anhydrosepedonin **160** and azanidulone **159**), 225 (mass of 224 Da for sepedonin **90**, **157**, **158**), 223, 241 and 227 (mass of 222 Da, 240 Da, and 226 Da for potential oxidized or reduced intermediates).

Reaction of LwtR1 – Experiment I

The *in vitro* experiments confirmed the turnover of DHMBA **156** by the FAD-dependent monooxygenase LwtR1 (Figure 4.31 A). Adding LwtR1 leads to two new peaks in the LCMS chromatograms at 2.5 min and 3.4 min.

For the peak at 2.5 min (Figure 4.31 A), the *m/z* data indicates a mass of 224 Da, which is in agreement with the formation of the oxygenated and de-aromatized benzaldehyde structure **157**. Due to the lack of literature data for the compounds **157** and **158**, it is unknown if this peak indicates the ring-opened structure **157** or the hemiacetal form **158** as they are characterized by the same mass of 224 Da. The compound eluting at 3.4 min (Figure 4.31 A) with a characteristic UV maxima at 337 nm and a mass of 206 Da is consistent with literature data for azanidulone **159**.
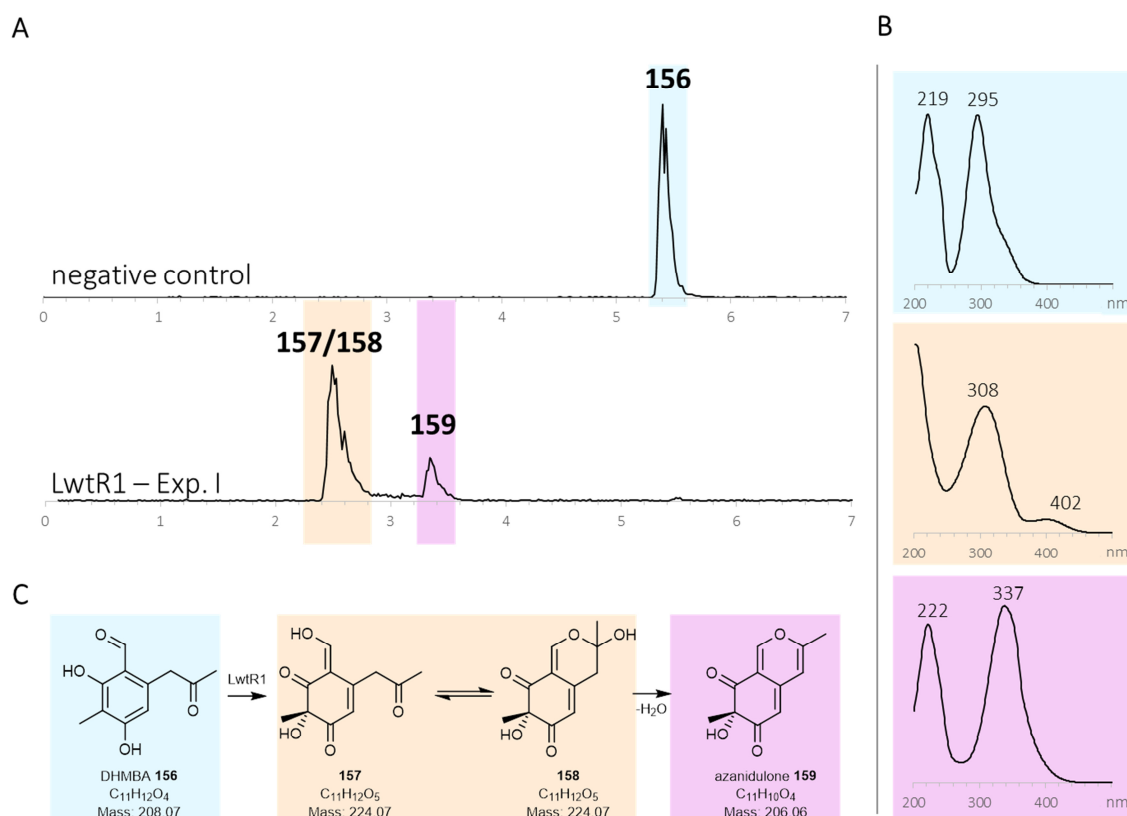
Figure 4.31 Results for experiment I of *in vitro* enzyme reactions: **A**, ES$^+$ EIC chromatograms for *m/z* of 209 + 225 + 207 + 223 + 227 + 241 for the negative control and experment I; **B**, DAD spectra of according peaks; **C**, proposed pathway.

Reaction of LwtR1 + LwtR7 – Experiment II

The key step for the formation of the 7-membered aromatic ring system in tetraketide tropolones is catalysed by the non-heme iron dioxygenase, which oxidizes the 3-methy-group, enabling the pinacol-like ring-expansion rearrangement (section 1.4.3.2, Figure 1.13). As in the *in vivo* studies, the *in vitro* assay showed that LwtR7 alone is not sufficient for the formation of the tropolone skeleton (Figure 4.32 A).
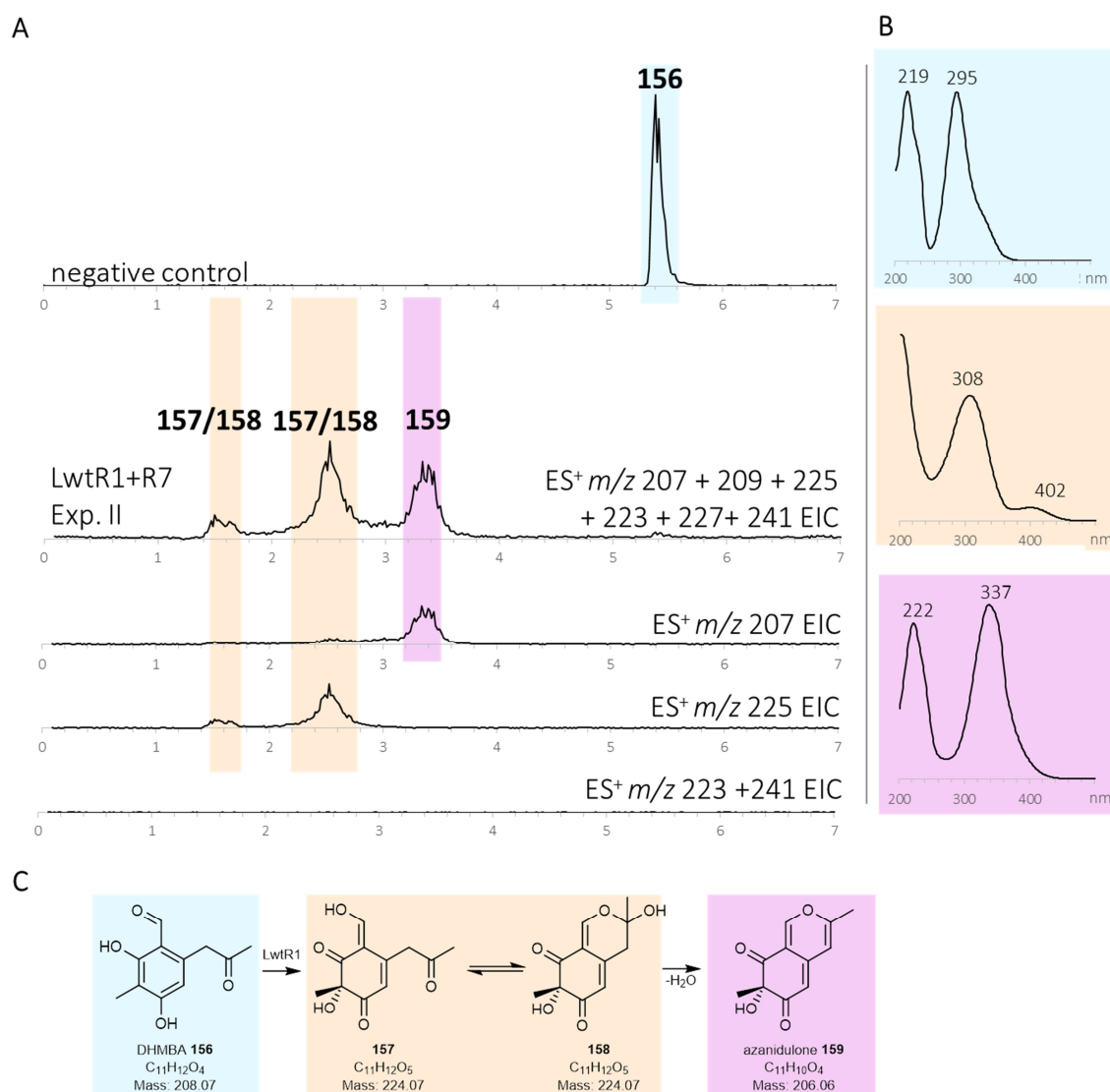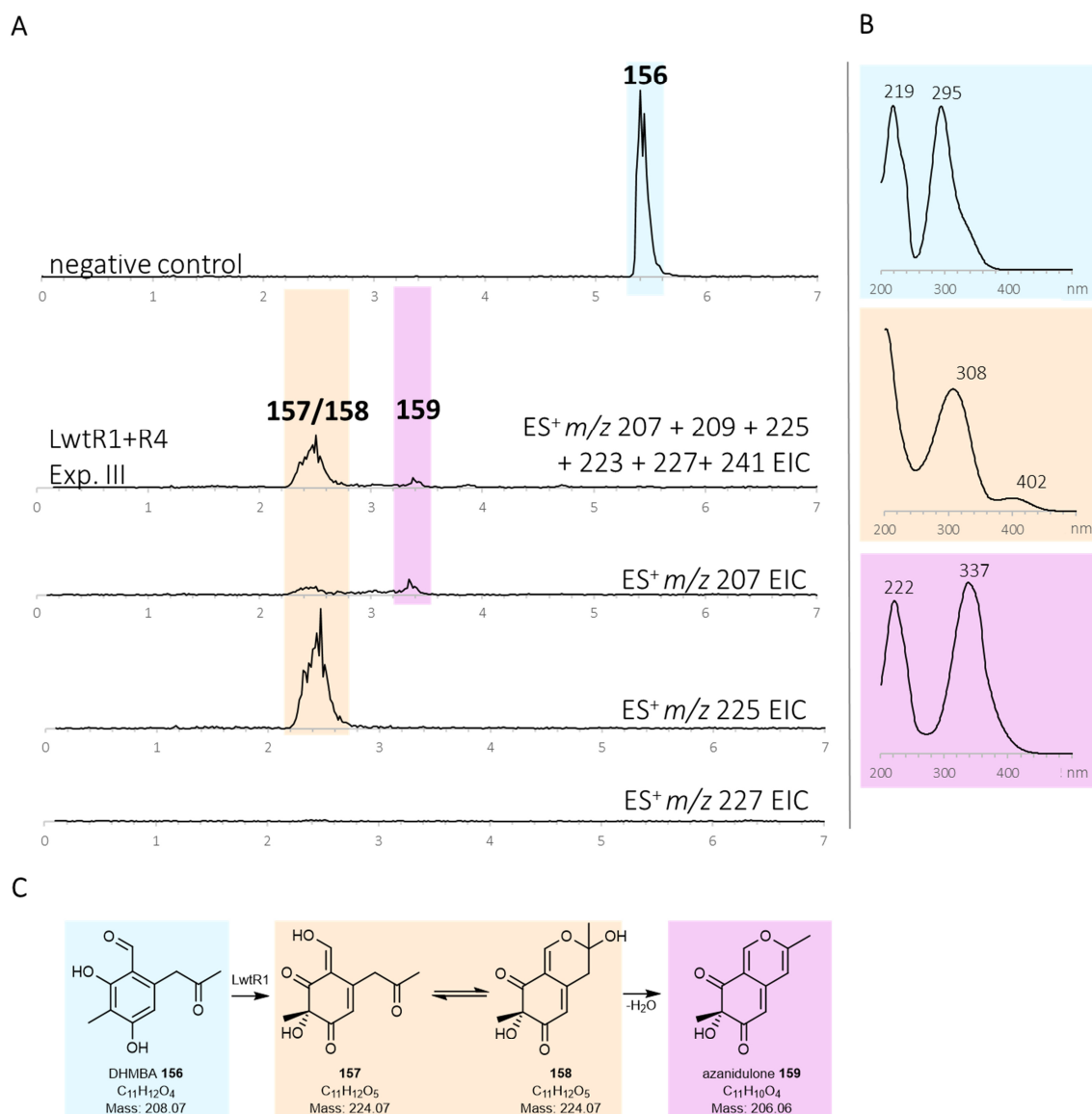
Figure 4.32 Results for experiment II of *in vitro* enzyme assays: **A**, ES$^+$ EIC chromatograms for *m/z* of 209 + 225 + 207 + 223 + 227 + 241 for the negative control and experment II; **B**, DAD spectra of according peaks; **C**, proposed pathway.

No new tropolone-containing compound was found with the combined assay of LwtR1 together with LwtR7, besides the already described compounds **159** and **157** or **158**. A peak at 1.6 min with an *m/z* data by ES$^{+/-}$ spectra indicates again a mass of 224 Da, which could again confirm the presence of the de-aromatized benzaldehyde structure **157** or the hemiacetal **158**. Both, the peak at 1.6 min and at 2.5 min with a mass of 224 Da are characterized by a similar UV maximum at 308 nm and 311 nm (Figure 4.31 B), confirming the structural relationship between **157** and **158**. Remarkably, no potential oxidized product with a mass of 240 Da or 222 Da (+ O, - H$_2$O) was observed (Figure 4.31 A). The absence of **168** *in vitro* confirms that **168** is unrelated to the pathway *in vivo* (section 4.3.6).

Reaction of LwtR1 + LwtR4 – Experiment III

Adding LwtR1 in combination with LwtR4 to the assay did not lead to any new compounds (Figure 4.33). Instead, the already observed compounds **157/158** and **159** were detected. No potentially reduced compounds (mass of 226 Da) was detected (Figure 4.33 A).

Figure 4.33 Results for experiment III of *in vitro* enzyme assays: **A**, ES[+] EIC chromatograms for *m/z* of 209 + 225 + 207 + 223 + 227 + 241 for the negative control and experment III; **B**, DAD spectra of according peaks; **C**, proposed pathway.

Reaction of LwtR1 + LwtR4 + LwtR7 – Experiment IV

When LwtR1, LwtR4 and LwtR7 were added together (Figure 4.34), sepedonin **90** was detected at 3.3 min with ES$^{+/-}$ spectra (*m/z* of 223 and 225) and UV spectrum (maxima at 219/252/326 nm), additionally to **157/158**.
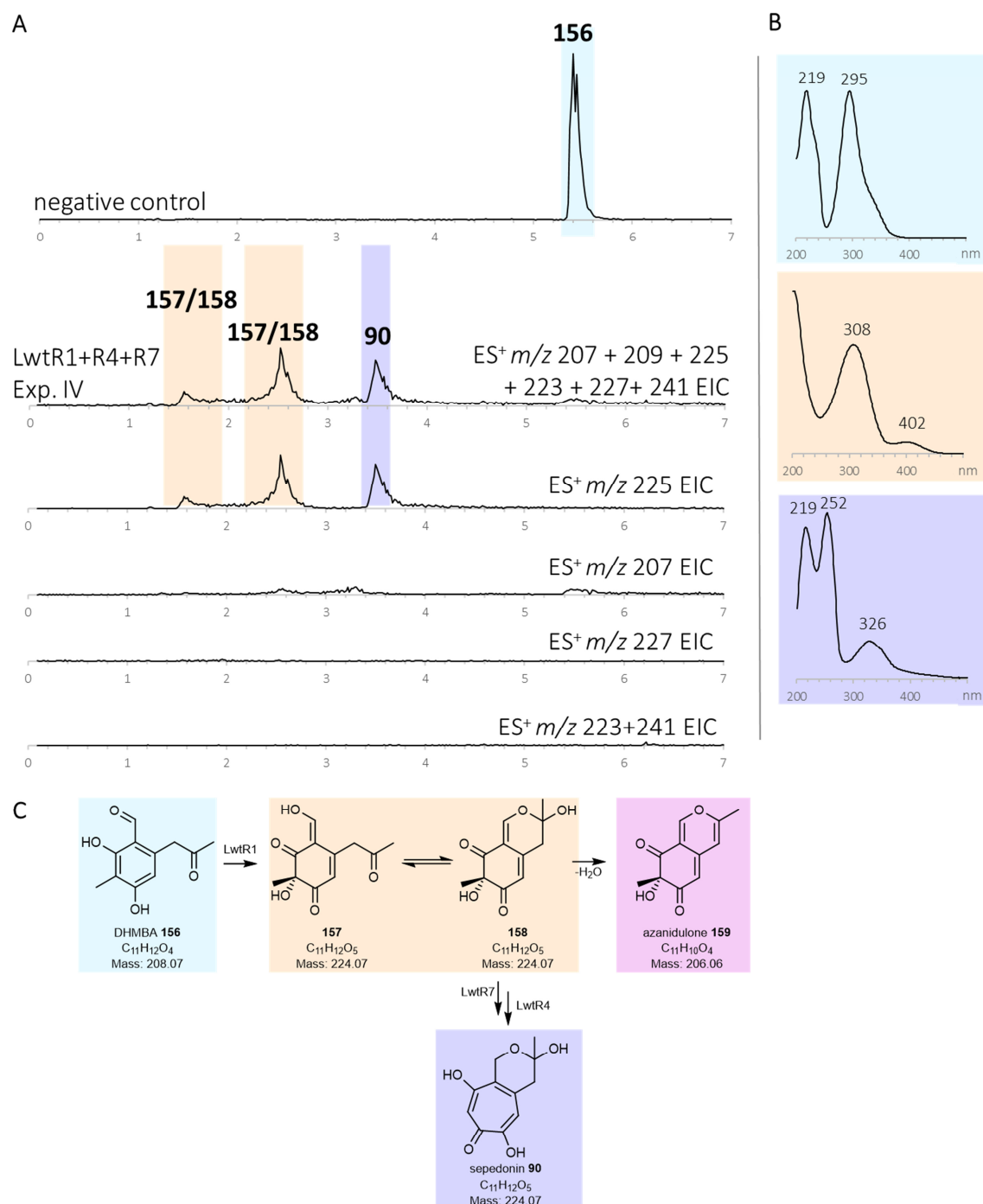
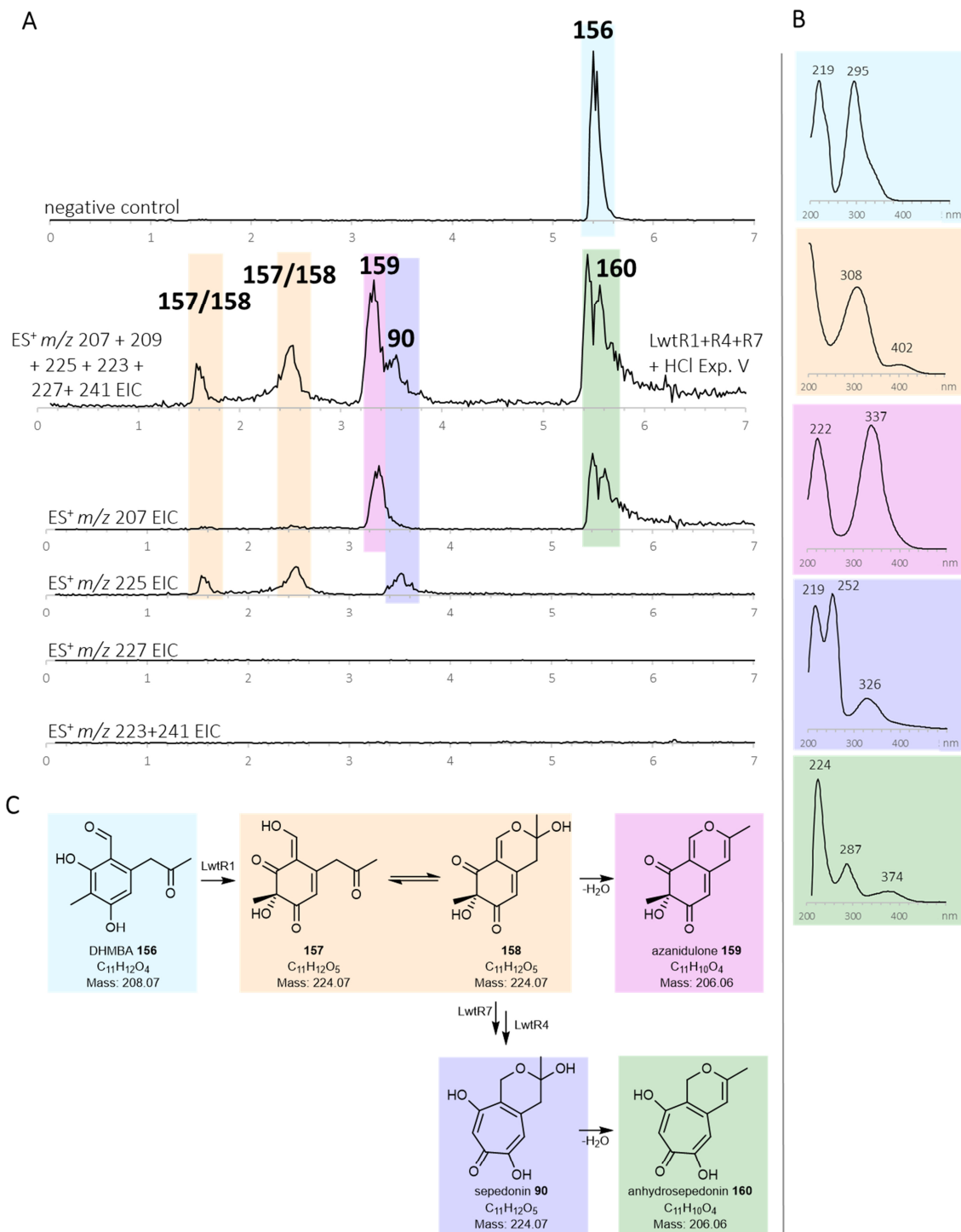**Figure 4.34** Results for experiment IV of *in vitro* enzyme assays: **A**, ES$^+$ EIC chromatograms for *m/z* of 209 + 225 + 207 + 223 + 227 + 241 for the negative control and experment IV; **B**, DAD spectra of according peaks; **C**, proposed pathway.

172

The turnover of the reaction is not at 100 %, as the peak for the FMO product **157** or **158** is still present in the mixture. Changing the conditions of the reaction by varying incubation time or concentration of the substrate/proteins did not lead to a higher efficiency of the reaction. An intermediate for the sepedonin **90** formation was not detected in any of the assays (Figure 4.34).

Finally, the extract of the reaction mixture was acidified with 3 µl of 2 M HCl (experiment V) before LCMS analysis. The acidification converted sepedonin to anhydrosepedonin **160** (Figure 4.35), suggesting that the dehydration reaction to anhydrosepedonin **160** is spontaneous under acidic conditions.

**Figure 4.35** Results for experiment IV + HCl of *in vitro* enzyme assays: **A**, ES$^+$ EIC chromatograms for *m/z* of 209 + 225 + 207 + 223 + 227 + 241 for the negative control and experment V; **B**, DAD spectra of according peaks; **C**, proposed pathway.

Formation of compound **163** was not observed in *in vitro* approaches, which confirms that **163** most likely is a shunt product of *A. oryzae* NSAR1.

4.4 Discussion: Tropolone Biosynthetic Pathway

4.4.1 BGC Synteny Analysis

Synteny analysis showed highly homologous, conserved putative tropolone-producing BGCs in nearly all of the analysed 35 species, which correlates to former work by Kuhnert *et al.*, where 13 from 14 species included an analogous cluster.[149] Combining these new results with those previously reported by Kuhnert *et al.*[149] (Figure 4.36), a total of 49 mostly *Hypoxylacae* genomes were analysed for tropolone BGCs. All except two (*H. laschii* and *H. fragiforme,* Figure 4.36) were found to contain tropolone BGCs. Of the remaining 47 species, 44 contain BGCs that are highly homologous to the *lwt*-BGC in terms of gene content, gene order and gene direction (section 4.1.2). This may be an indication of an important ecological role of metabolites made by the clusters, as the BGC has been highly conserved during evolution of these fungi. The final three tropolone BGCs (*P. papilatum*, *P. ruwenziorense*, *C. sassafras*) are more related to the tropolone meroterpenoid BGCs previously studied in the Cox group.[95,180] Two of these species, *P. papilatum*, and *P. ruwenziorense*, were recently assigned to the new genus *Parahypoxylon*, indicating a less near relationship to other investigated species, while *C. sassafras* does not belong to *Hypoxylacae*.[189]
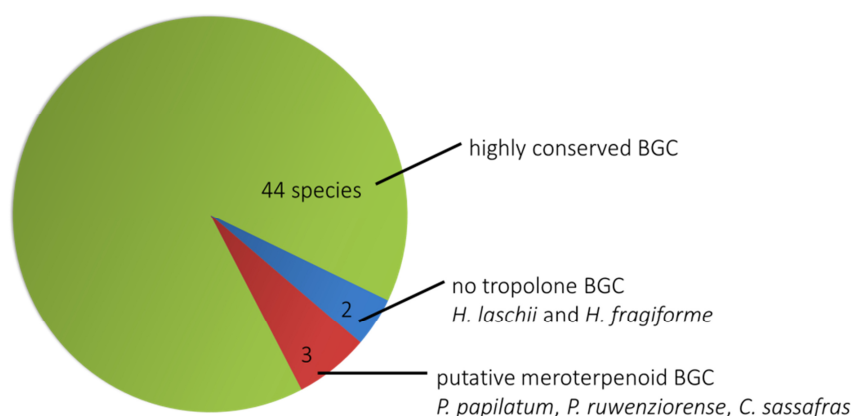
Figure 4.36 Distribution of putative tropolone producing BGCs in the analysed *Hypoxylacae* species.

Interestingly, *H. lienhwacheense* is the only investigated species, where tropolone-containing structures were found in extracts of the fungus to date. The formation of the specialised secondary metabolites often requires an activation of gene expression by

certain triggers and conditions. In the case of *A. nidulans*, activation of the *dba*-BGC by co-cultivation with *Streptomyces* species was observed.[177,178] In further studies Gerke *et al.* showed that the antifungal metabolite phleomycin is the key factor for fungal tropolone gene cluster activation in *A. nidulans*.[177] Accordingly, certain unknown triggers and conditions could lead to the formation of tropolone metabolites in the fungi studied, which can be identified in the future.

Furthermore, the putative *lwtR6* gene is mysterious as it is highly conserved in all putative sepedonin-related gene clusters (Figure 4.2), suggesting an important role despite the very unusual length of 330 bp. No function was assigned to the YCII domain containing putative protein encoded by *lwtR6* by heterologous expression experiments. In order to obtain an indication of the function of *lwtR6* and the resulting YCII domain containing protein, a blastx search against the Swiss-Prot database[151] was carried out. In bacteria only two catalytic enzymes including a YCII domain were assigned to a catalytic function. In *Burkholderia phenoliruptrix* AC1100 and in *Rhodococcus opacus* YCII-domain containing enzymes catalyse dechloronations.[190,191] A functional relationship to LwtR6 is excluded, as no halogenated compound was detected in any investigation with sepedonin forming BGCs. Additionally, a structural alignment analysis was carried out. Therefore, the LwtR6 structure was predicted using AlphaFold.[31] The resulting protein structure was used as a template for a search on the Dali server,[192] which did not lead to a consistent hypothesis for the function of LwtR6. In contrast to the blastx search, Dali server hits indicate non-catalytic functions of proteins including a YCII domain. Examples of hits are the ATX1 copper chaperone (*S. cerevisiae*),[193] nickel responsive regulator (*E. coli*)[194] or transcriptional regulators (*Neisseria meningitidis*).[195]

For the related biosynthesis of antibiotic C **91** in *A. nidulans* (section 1.4.3.2) also no function was assigned to the two present putative YCII-domain including enzymes of the *dba*-BGC (*AN7894* and *dbaC*). The separate knock-out of *AN7894* and *dbaC* maintained tropolone production. A simultaneous knock-out of both genes was not carried out. However, Gerke *et al.* showed by transcriptome analysis that activation of the BGC by phleomycin also increased gene expression of the two putative YCII domain-containing proteins AN7894 and DbaC. They speculated on a regulative role of the two genes *AN7894* and *dbaC*.[177,178] Consequently, a clear conclusion on a putative

function of the putative LwtR6 protein requires more investigations on this mysterious protein.

4.4.2 Biosynthetic Steps for Sepedonin Formation

A combination of heterologous expression experiments and *in vitro* protein assays leads to the proposed biosynthetic pathway of the sepedonin-related secondary metabolites (Table 4-15, Scheme 4.6).

Table 4-15 Summary of heterologous expression and *in vitro* experiments (green = catalytic core enzymes, * = LwtR1 *in vivo* exchanged by DbaH).

	putative lwt genes								163 reduced shunt	156 DHM BA	157/158 de-aromatized	159 azani-dulone	90 sepe-donin	160 an-hydro-sepedo-nin	91 anti-biotic C
	S	R1*	R2	R4	R5	R6	R7	P-450							
						heterologous expression									
A	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-
B	✓	✓	-	-	-	✓	-	-	✓	✓	-	✓	-	-	-
C	✓	✓	-	-	-	✓	✓	-	✓	✓	-	✓	-	-	-
D	✓	✓	✓	-	-	✓	✓	-	✓	✓	-	✓	-	-	-
E	✓	✓	-	✓	-	✓	✓	-	✓	✓	-	✓	✓	✓	-
F	✓	✓	✓	✓	-	✓	✓	-	✓	✓	-	✓	✓	✓	-
G	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	-
						in vitro assays									
I	-	✓	-	-	-	-	-	-	-	-	✓	✓	-	-	-
II	-	✓	-	-	-	-	✓	-	-	-	✓	✓	-	-	-
III	-	✓	-	✓	-	-	-	-	-	-	✓	✓	-	-	-
IV	-	✓	-	✓	-	-	✓	-	-	-	✓	-	✓	-	-
V	-	✓	-	✓	-	-	✓	-	-	-	✓	✓	✓	✓	-

The proposed biosynthetic pathway (Scheme 4.6) will be discussed step-by-step in the following sections 4.4.2.1 - 4.4.2.4 .
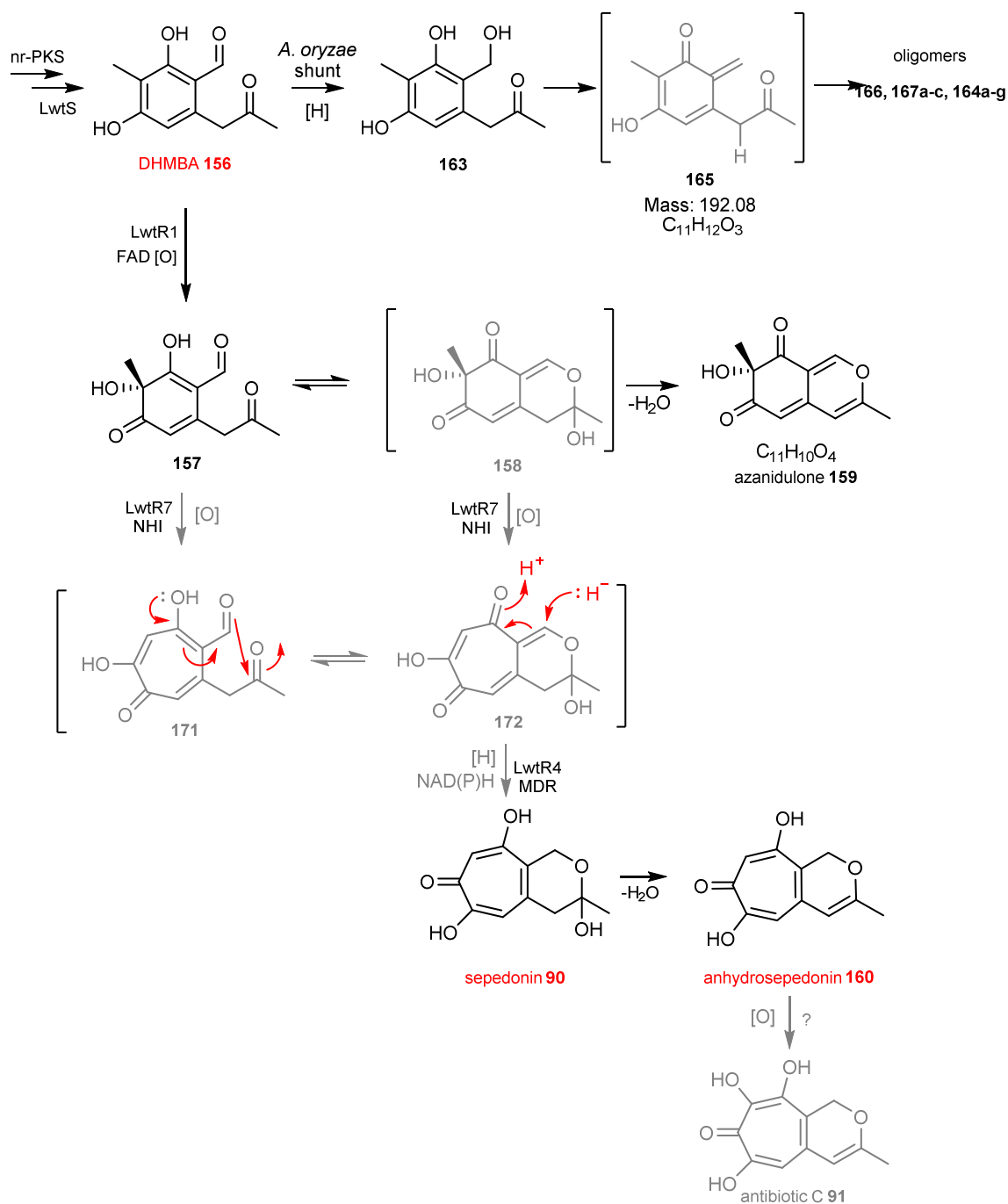
Scheme 4.6 Proposed biosynthesis of anhydrosepedonin **160** (red = identified with NMR analysis; black = identified by retention time, UV maxima and fragmentation in ES$^+$/$^-$ in LCMS analysis; grey = not detected).

4.4.2.1 First Enzyme-free Intermediates of the Tropolone Pathway

The nr-PKS releases the first enzyme free tetraketide intermediate DHMBA **156**. The structure was verified by NMR. However, the production of DHMBA led to the unpredictable formation of shunt metabolite **163** during expression in *A. oryzae* NSAR1

(section 4.3.4, Table 4-15 experiment A). The mass of **163** was not found in assays with isolated proteins, which indicates that DHMBA is stable, but is reduced in *A. oryzae* NSAR1 at the C-1 aldehyde to the corresponding primary alcohol **163**, which loses water to form the key structure **165**. The ortho-quinomethide **165** is most likely the key intermediate for the formation of various shunt metabolites **167a-c** and **164a-g** (Scheme 4.4). The dimeric structure of **164f** was confirmed by NMR.

The formation of reduced shunt metabolites by *A. oryzae* is not surprising. Similar reductions were found when investigating the citrinin pathway BGC of *Monascus ruber* M7 by He and Cox. Heterologous expression in *A. oryzae* confirmed a similar reduced aldehyde compound, leading to a cyclic hemiacetal.[46] Other examples are known, where the intracellular conditions of *A. oryzae* reduce aldehyde intermediates to form shunts. Examples of common shunt reactions appear in the context of the heterologous expression in *A. oryzae* of cytochanasans BGCs,[196,197] xenovulenes,[180] sporothriolides,[78] brasilanes,[198] and azaphilones[199] to name just a few examples.

Although the formation of shunt metabolites does not directly affect the pathway, it does reduce the throughput of the intermediates available to downstream enzymes and thus the pathway is less efficient.

4.4.2.2 Reaction of LwtR1

The PKS releases DHMBA **156**. The oxidative de-aromatisation reaction from DHMBA **156** to **157** (Scheme 4.6) is catalysed by the FMO LwtR1 (for *in vitro* experiments, DbaH for *in vivo* experiments), in agreement with known pathways to the tetraketide tropolone stipitatic acid (section 1.4.3.2, Table 4-15 experiment I). Although the intermediate **157** and **158** were not detectable in heterologous expression experiments, *in vitro* protein assays detect the conversion of DHMBA to two compounds **157** and **158** (Table 4-15 experiment I). After spontaneous cyclisation, which forms the second heterocycle, a dehydration reaction follows to form azanidulone **159** *in vivo* and *in vitro* (Table 4-15 experiment B, I). Azanidulone **159** is a well-known shunt product in related studies from Gerke *et al*. The compound was also observed in studies by Pyser *et al*. from a reaction with DHMBA **156** and the TropC homologous monooxygenases AfoD and AzaH (Scheme 4.8).[177,187]
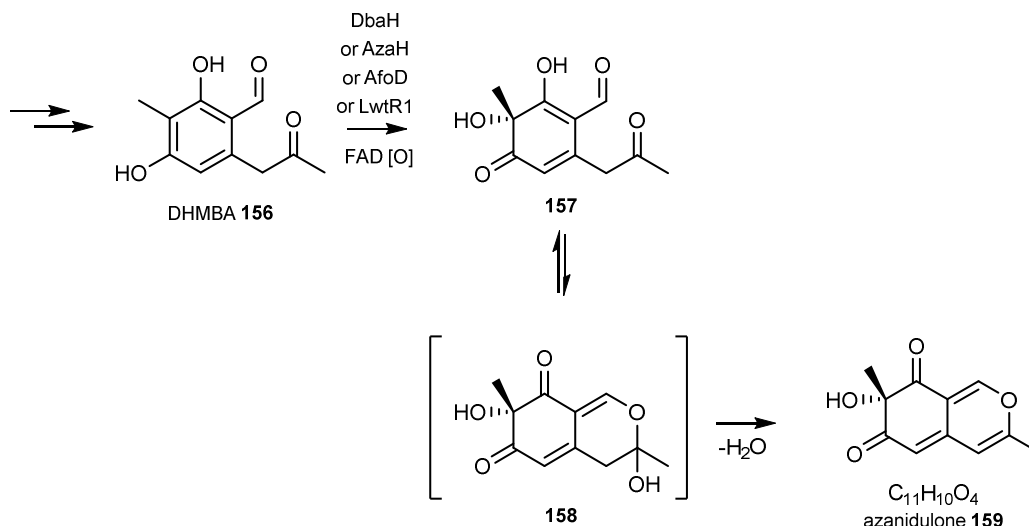
Scheme 4.7 Known reactions leading to shunt product azanidulone **159**.

4.4.2.3 Reaction of LwtR4 and LwtR7

The formation of sepedonin **90** from **157** was only detectable, when LwtR7 (NHI) *and* LwtR4 (ER-like enzyme) were both expressed in heterologous expression experiments, as well as *in vitro* protein assays (Table 4-15, experiments E, IV). Excluding one of the two redox enzymes abolished sepedonin production (Table 4-15, experiments D, II, III). No intermediate was detected during *in vivo* or *in vitro* experiments when adding only one of the redox enzymes. Compound **171** and **172** (Scheme 4.6, Scheme 4.8) are both feasible as intermediates of this reaction. The absence of a detectable intermediate similar to **171** or **172** suggests that the intermediate either is unstable and needs to react immediately with the next enzyme, or the two enzymes work together for the formation of sepedonin **90**. The substrate **157/158** for the reaction to **90** seems not to be consumed in any of the reaction mixtures including only one of the enzymes. Hence, a cooperative reaction of the two proteins LwtR7 and LwtR4 is most likely.

In related pathways (section 1.4.3.2) the ring expansion reaction takes place immediately after the oxidative de-aromatisation (Scheme 4.8 A). Analogously this would lead to compound **171**, which reacts to the hemiacetal **172** and is followed by the reduction (Scheme 4.8 B, route B). The sequence of reactions is also possible, when the hemiacetal **158** is formed first and then followed by the ring expansion reaction to **172** (Scheme 4.8 B, route A). As **158** was detectible in *in vitro* assays (section 4.3.7.2) route A is the more likely mechanism, which leads to **90**.

180
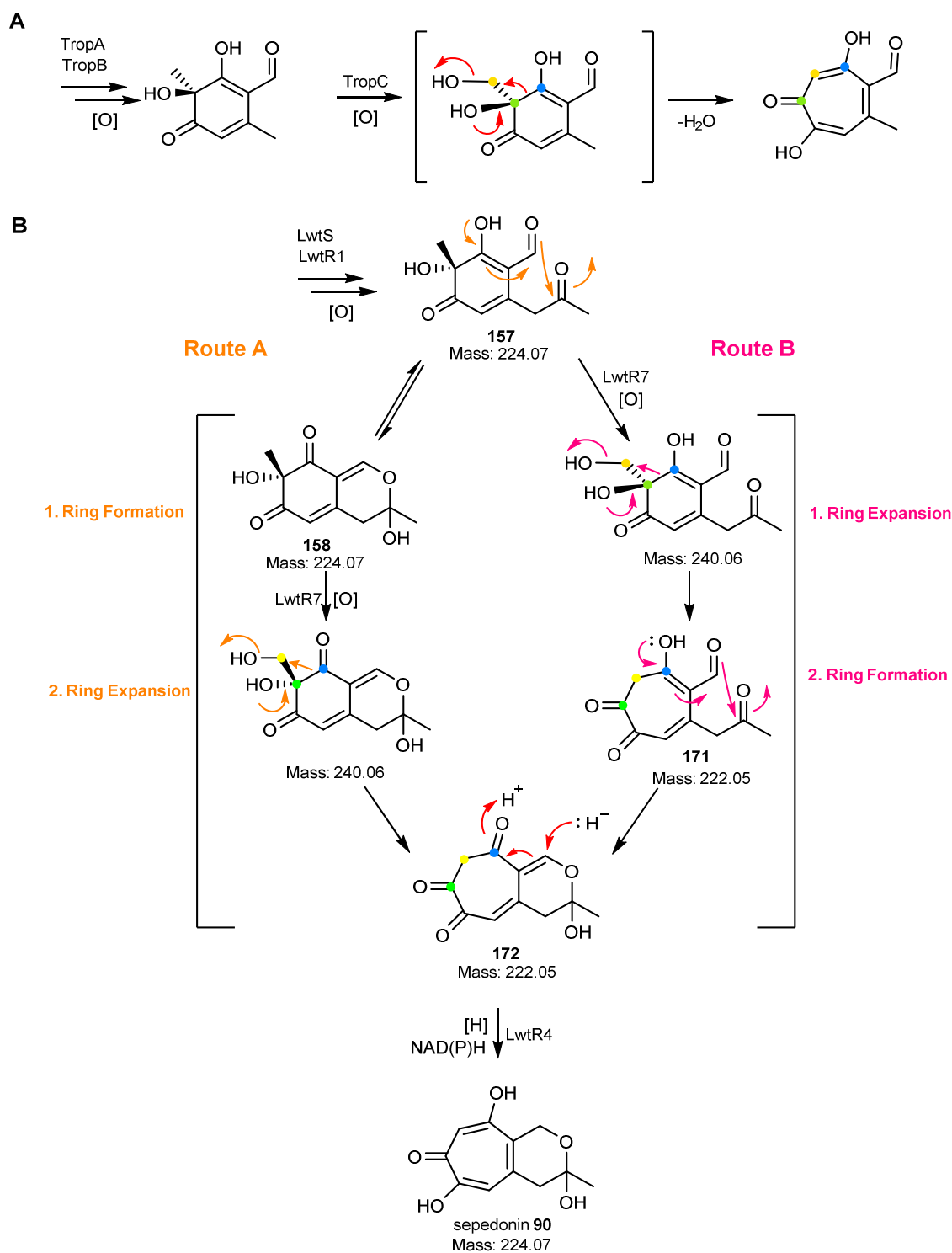
Scheme 4.8 Mechanistic steps of tropolone formation: **A**, mechanism of TropB and Trop C; **B**, proposed mechanism of LwtR1, LwtR7 and LwtR4.

In the case of previous publications on the pentaketide tropolone antibiotic C **91** in *A. nidulans* by Gerke *et al.* the ring-expansion reaction to sepedonin **90** was not traced back to the MDR enzyme AN7895, which is present in the cluster (Figure 4.1).[177] The

knock-out strains produced anhydrosepedonin despite the lack of the MDR (AN7895), while knock-out of AN7893 (*troA,* NHI) led to the cessation in the production of sepedonin-like structures. Therefore, the close cooperation between the NHI TropC analogue AN7893 (TroA) and the MDR enzyme AN7895, analogous to the cooperation of LwtR7 (NHI) and LwtR4 (ER-like enzyme), was not found and requires confirmation by further studies.[177] However, knock-out experiments can lead to inaccuracy, because the background activity of other enzymes cannot be excluded. Thus, it is possible that reductive activity was performed by a different enzyme of the fungus. In support of this idea, a blastp search with the *dba* MDR protein sequence (encoded by AN7895) against the *Aspergillus nidulans* genome results in a hit with 41.3 % identity. The hit (XP_658634) is a hypothetical protein. Thus, the function is unknown. XP_658634 shows also an identity of 39.53 % with LwtR4 and may catalyse the reaction in the knock-out strain.

Furthermore, LwtR4 shows homology to enzymes from other biosynthetic pathways. For example, the clinker synteny analysis of the *lwt*-BGC with the related azaphilone BGC from *Monascus ruber* (section 4.4.3, Figure 4.37) illustrates the identity of 45 % to MrPigH. The function of the LwtR4 homologue MrPigH was studied before (Scheme 4.9). It was suggested that MrPigH of *Monascus* fungi contributes to the saturation of azaphilone intermediate **173** (but is not a requirement) by reducing the C5(2') double bond to **174**.[199–202]
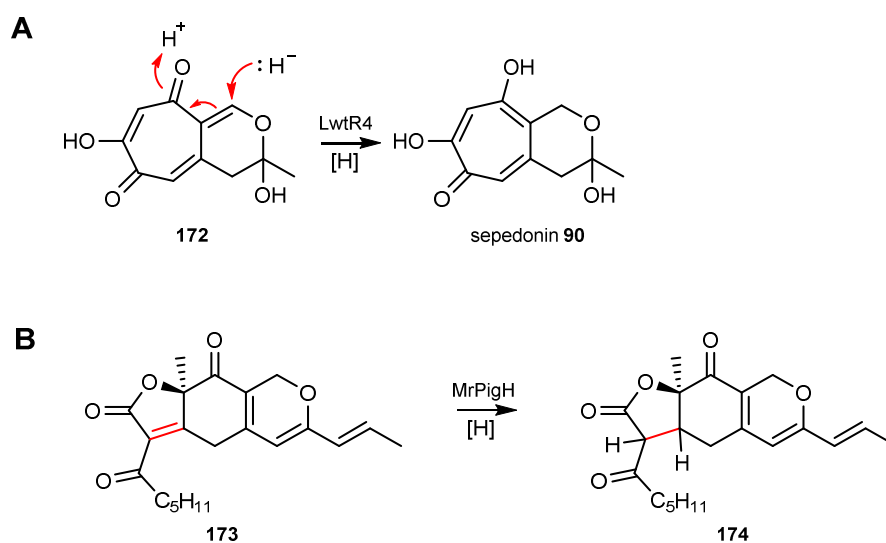
Scheme 4.9 Proposed reaction of enoyl-reductase like enzymes: **A,** LwtR4 in *H. lienhwacheense*; **B,** MrPigH in *Monascus* fungi.[199–202]

4.4.2.4 Late Modification Steps of the Sepedonin Skeleton

Finally, the sepedonin derivative anhydrosepedonin **160** is formed by dehydration of sepedonin **90**, most likely occurring spontaneously, as acidification of the reaction mixture after *in vitro* assays led to the loss of water (Table 4-15, experiment V).

It is an open question how antibiotic C **91** is formed in the biosynthetic pathway. Antibiotic C **91** is a putative intermediate during lienhwalide formation. It was not possible to identify a putative redox enzyme to catalyse the hydroxylation of **160**. Not only in the *lwt*-BGC but also in the complete *H. lienhwacheense* genome no DbaB analogue was found.[177] The heterologous co-expression of the core genes with the transport protein encoding gene *lwtR5* led to increased production of antifungal tropolone compounds, most likely achieved by the de-toxification of the fungal cells by the transport protein LwtR5.

Furthermore, *O*-methylated antibiotic C derivatives are included in lienhwalide B and C. The experiments (section 4.3.6.7) showed that spontaneous formation of methyl ethers by the incubation in methanol does not occur. An unknown enzyme most likely catalyses the introduction of the methyl group.
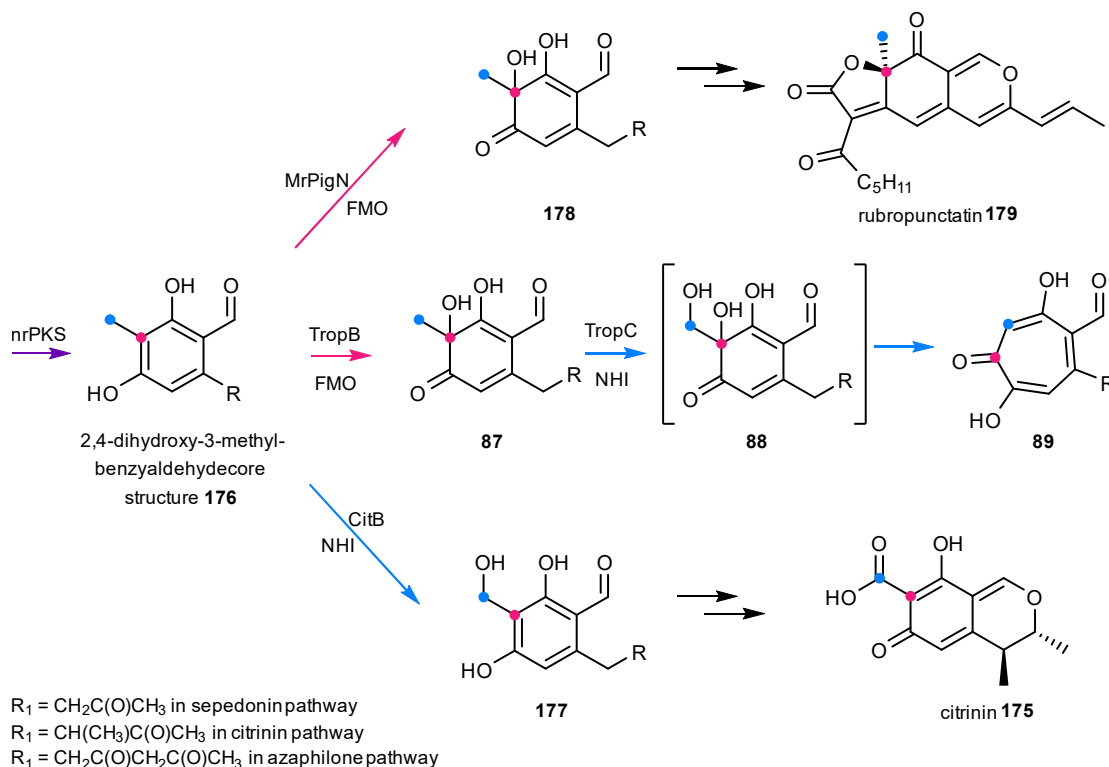
The responsible enzymes for the formation of the final intermediates (for final aromatic hydroxylation and *O*-methylation) for the lienhwalide formation are not clustered alongside the other catalytic genes of the BGC. It is unknown, if the final modification steps take place before or after the formation of the lienhwalide **130**-**132** scaffold.

4.4.3 Related Secondary Metabolite Pathways

Taking the complete tropolone-forming pathway into consideration, a (evolutionary) relationship to related pathways of other fungi and the included core enzymes is proposed (Figure 4.37). For example, the NHI enzyme CitB (TropC homolog, Figure 4.37) of the citrinin **175** biosynthetic pathway catalyses the oxidation of the 3-methylgroup of **176** to **177**, without previous de-aromatisation (Scheme 4.10) due to the absence of a LwtR1 FMO homologue in the cluster (Figure 4.37).[46]
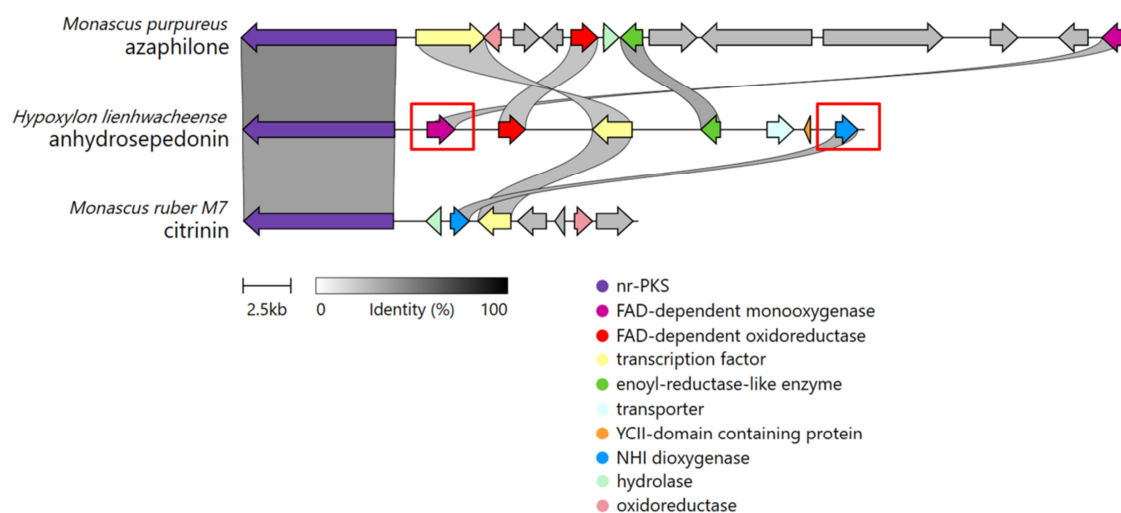
Figure 4.37 Comparison of BGCs for sepedonin, citrinin **175** and azaphilone **179** biosynthesis.

R₁ = CH₂C(O)CH₃ in sepedonin pathway
R₁ = CH(CH₃)C(O)CH₃ in citrinin pathway
R₁ = CH₂C(O)CH₂C(O)CH₃ in azaphilone pathway

Scheme 4.10 Pathway relationships between tropolones, citrinin and azaphilones.

On the other hand, during azaphilone **179** biosynthesis the core 2,4-dihydroxy-3-methyl-benzaldehyde **176** structure undergoes only the FAD-dependent oxidative de-aromatisation, catalysed be MrPigN (TropB-homolog, Figure 4.37) to **178** (Scheme 4.11). While other modifications take place, no re-aromatisation is part of the later azaphilone biosynthesis, which is the most important difference to tropolone pathways as no NHI enzyme is present in the BGC (Figure 4.37) for the early biosynthesis.[199] The

combination of both enzymes, the FMO together with the NHI leads to tropolones, like **89** (Scheme 4.11).

Furthermore, azaphilone BGCs include two additional genes homologous to *lwt*-genes coding for catalytic tailoring enzymes. The LwtR4 homolog MrPigH (section 4.4.2.3) and the homologue of the FAD-dependent oxidoreductase LwtR2, MrPigF are present in azaphilone BGCs (Figure 4.37). While the function of LwtR2 is unknown for the sepedonin pathway, MrPigF was proposed to be responsible for the reaction of **178** to **179** (Scheme 4.12).
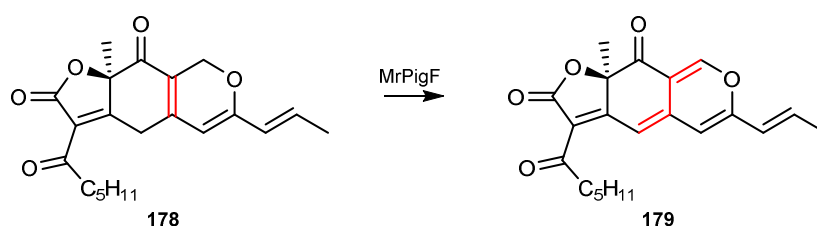
Scheme 4.11 Reaction of MrPigF.[199–202]

The three BGCs (for tropolones, azaphilones and citrinin, Figure 4.37) and the associated biosynthetic pathways are an example for the evolutionary connection between fungal BGCs, and the emerge of diverse secondary metabolites through the combination of different tailoring enzymes to develop diverse pathways. Homologs for all genes encoding catalytic enzymes of the *lwt*-BGC were identified in BGCs for the citrinin **175** and the azaphilone pathway. Therefore, the combination of gene homologs of the citrinin pathway and the azaphilone pathway, leads to the production of tropolones.

4.5 Conclusion and Outlook

The biosynthetic pathway of tropolone formation in *H. lienhwacheense* was elucidated, starting with the identification of the *lwt*-BGC as the responsible BGC for tropolone formation by heterologous expression. The *lwt*-BGC includes the genes encoding for core enzymes known for tropolone biosynthesis.

Heterologous expression experiments were the basis for the identification of all required key enzymes for the biosynthesis. The formation of the tropolone skeleton requires four key enzymes of the cluster: the nr-PKS LwtS, the FMO LwR1, the NHI enzyme LwtR7 and the MDR LwtR4. The key role of LwtR4 as an essential reductive enzyme has not been demonstrated before. Hence, an important aspect of pentaketide tropolone formation was elucidated by heterologous expression methods and was further confirmed by *in vitro* assays with isolated recombinant proteins. However, the order and the potential cooperation of the enzymatic steps for the tropolone formation are unknown. During the studies, the structures of anhydrosepedonin, sepedonin, and an often occurring *A.oryzae* shunt metabolite were elucidated.

While the formation of the core structure is thus solved, further tailoring steps leading to the formation of intermediates incorporated into lienhwalides remain unknown. The oxidation and methylation of the tropolone ring was not carried out by enzymes located in the *lwt*-BGC and requires further investigation.

The synteny of the tropolone-producing BGCs in 44 genomes of analysed related fungal species suggests that tropolone including compounds (likely to be closely related to sepedonin) may play an important ecological role in the entire family of *Hypoxylacae* and relatives. *Gerke et al.* showed that sepedonin is produced in response to bacterial production of phleomycin, and thus sepedonin might have an antibacterial role in these fungi. The number and order of genes is preserved in each of the 44 genomes. Therefore, similar biosynthetic steps can be speculated for similar pathways for the 44 species. However, the true identity of the final products of the pathways remains unknown for other species.

Successful heterologous production of sepedonin and anhydrosepedonin is a prerequisite for further analysis of lienhwalide formation (chapter 5).

5 Studies on the Lienhwalide Formation

5.1 Introduction

Lienhwalides consist of an alkyl citrate and tropolone moieties linked together by a carbon-carbon bond. The core steps of the biosynthesis from the maleic acid anhydride compounds like **136** and **138/139** by the *lwm*-BGC (chapter 3) and tropolone metabolites like **90** and **160** by the *lwt*-BGC (chapter 4) were elucidated previously. The coupling of the two compound classes will be investigated in the following chapter.

5.1.1 Hypothesis of Lienhwalide Coupling by MDC-related Mechanism

It was hypothesized that maleidrides such as the nonadrides byssochlamic acid **48** arise by dimerization of maleic acid anhydrides like **136**, which undergo decarboxylation, resulting in **138/139** (section 3.1.3.2, Scheme 5.1 A).[70–72] **136** is the donor in a two-step coupling reaction, where one acceptor monomer **138/139** is attacked by **136** in a condensation reaction (Figure 3.2, Scheme 5.1 B).[158,164] On this basis, it was speculated that the decarboxylated nucleophilic intermediate of **136** attacks electrophilic Michael-acceptor-like positions of tropolone intermediates like **172**, leading to lienhwalide-like compound **180** (Scheme 5.1 B, pink). The decarboxylation takes place spontaneously, leading to appropriate nucleophiles for the coupling reaction. Tropolone intermediate **172** is a putative product of oxidation by LwtR7, which undergoes reduction by LwtR4 for sepedonin biosynthesis (section 4.4.2.3, Scheme 5.1 B, green).
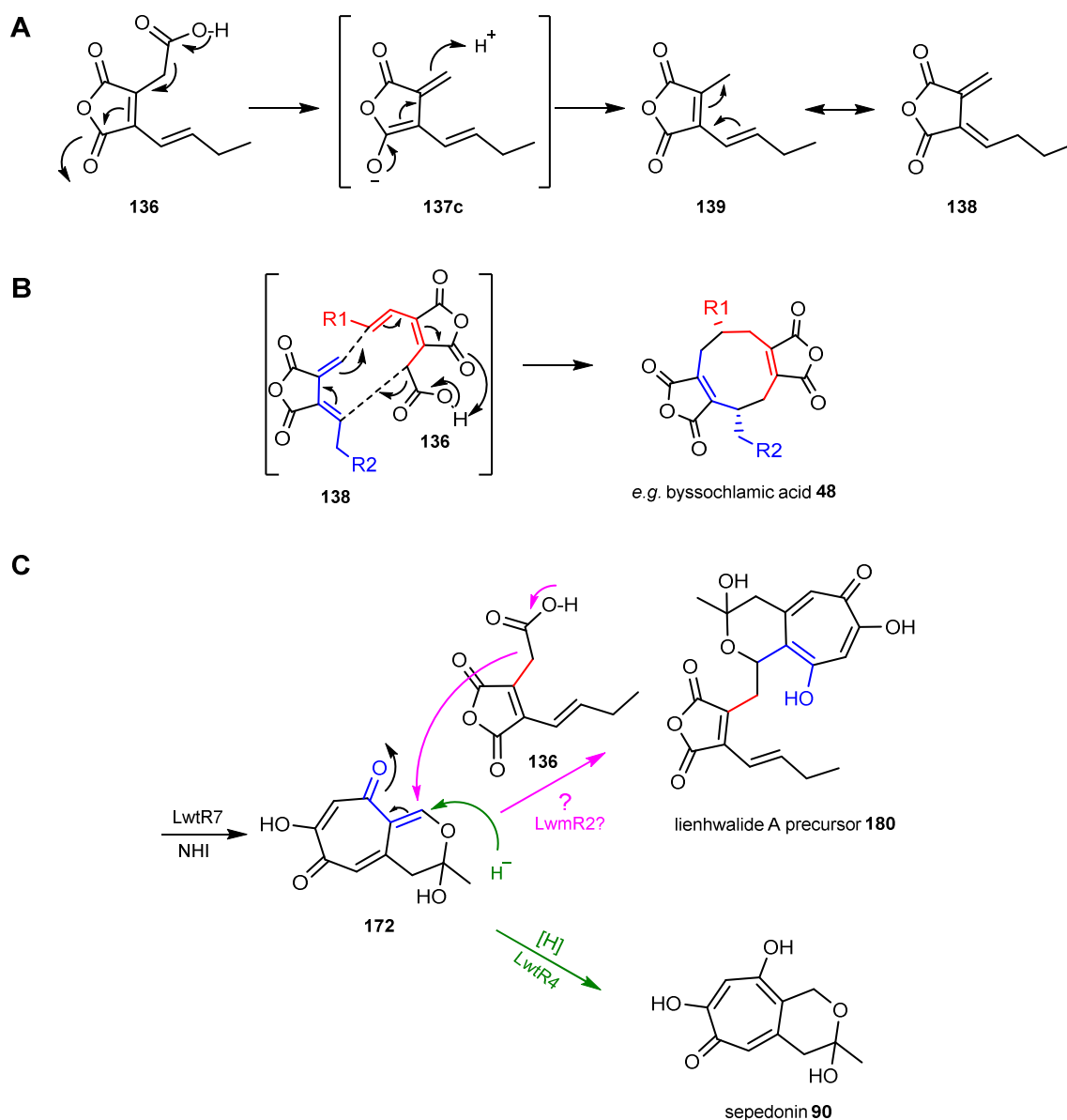
Scheme 5.1 Proposed mechanism of dimerisation: **A**, decarboxylation reaction of **136** to **138**; **B**, example of maleidride dimerization; **C**, proposed mechanism for lienhwalide coupling.

Two hypotheses for the coupling reaction were suggested, involving the same reactive steps. Firstly, the reaction could take place spontaneously perhaps encouraged by high concentrations during extraction or purification. Secondly, the coupling reaction could be catalysed by enzymes *e.g.* with chaperone-like functions. Previous studies have shown that MDCs catalyse the maleidride formation, including two monomeric maleic acid anhydride molecules (section 3.1.3.2). The *lwm*-BGC includes the putative MDC LwmR2, which has not been assigned a function by previous experiments (section 3.3.5). Therefore, LwmR2 could be a candidate to catalyse a coupling reaction.

5.2 Project Aims

First, the possibility of spontaneous coupling of tropolones and maleic acid anhydrides will be tested by co-cultivating maleic acid anhydride **136** and sepedonin **90** producing transformants. The assumption here is that products or intermediates of the separate pathways may meet in the fermentation media, or during extraction and concentration. If a spontaneous reaction is possible, heterodimers will be detected.

Second, the possibility of an enzyme-assisted coupling reaction will be investigated. Since previous studies have shown that MDCs catalyse homodimer formation during maleidride biosynthesis (section 3.1.3.2), the enzyme LwmR2 of the *lwm*-BGC will be focus of heterologous expression experiments. LwmR2 was not assigned to a biosynthetic function in *H. lienhwacheense* yet (section 3.4.3). The previously identified key enzymes for the formation of both components of the lienhwalides will be transformed together. Here, the assumption is that pathway products or intermediates will meet *in vivo*, with additional possibility of catalysis by LwtR2. Thus, all relevant enzymes from two different clusters (*lwm* and *lwt*) will be co-expressed with the MDC LwmR2.

5.3 Results

5.3.1 Co-Cultivation of Tropolone and Alkyl Citrate Producers

To investigate if spontaneous coupling of the compounds is possible in the media or during extraction, anhydrosepedonin **160** and maleic acid anhydride **136** producing transformants were incubated together (culture A and B, Scheme 5.2). The spontaneous coupling of **172** and **136** leads to **180**. Lienhwalide precursor **180** undergoes the loss of water to **181** and reduction to lead to lienhwalide A **130** (Scheme 5.2).
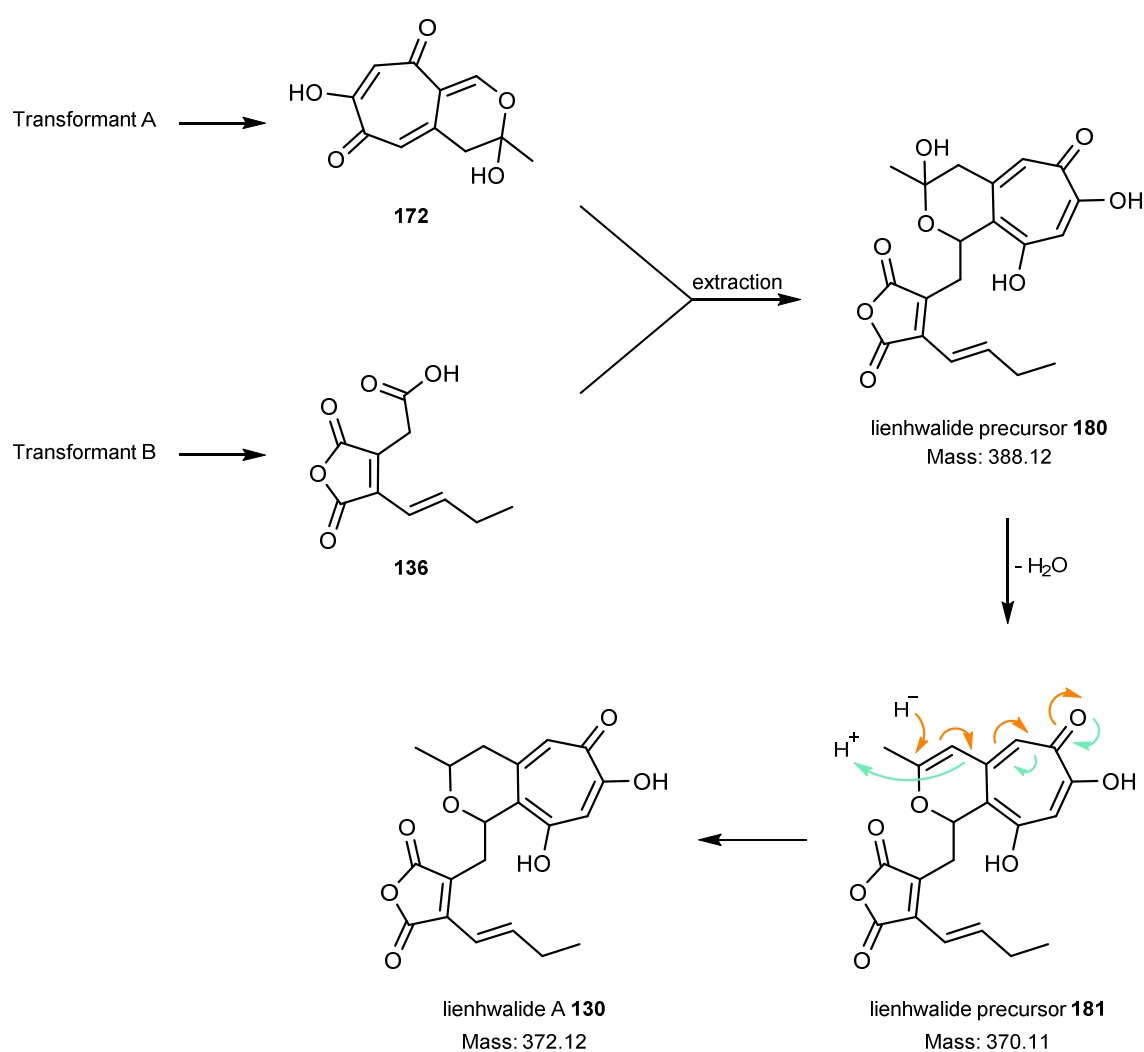

Scheme 5.2 Experimental setup and spontanous coupling reaction resulting in lienwalide A **130** *via* **180** and **181**.

Three tropolone producing transformants and three maleic acid anhydride producing transformants were chosen from previous experiments (section 3, section 4), giving overall nine different co-cultivations. The 100 ml DPY liquid cultures were inoculated with approx. the same amount of mycelia of both transformants in one 500 ml baffled flask and cultivated for 5 days at 28 °C and 110 rpm. The cultures were extracted as previously described (Figure 5.1).

All co-cultivations were screened for the potential coupled products **180**, **181** and lienhwalide A **130** (Scheme 5.2) with the expected masses (Figure 5.1 B). Additionally a screening for the expected UV absorption maxima was carried out (Figure 5.1 C), expecting maxima at approx. 210 nm, 266 nm, and 322 nm (Figure 3.4) for coupled products, which are expected to have a retention time of approx. 7 – 10 min.

No new compounds were found in the analysis of any of the co-culture extracts (Figure 5.1). Only the known compounds anhydrosepedonin **160,** the tropolone pathway shunt compounds **164a-f** and the monomeric maleic acid anhydrides **136**, **142**, and **143** were detected (Figure 5.1 B).
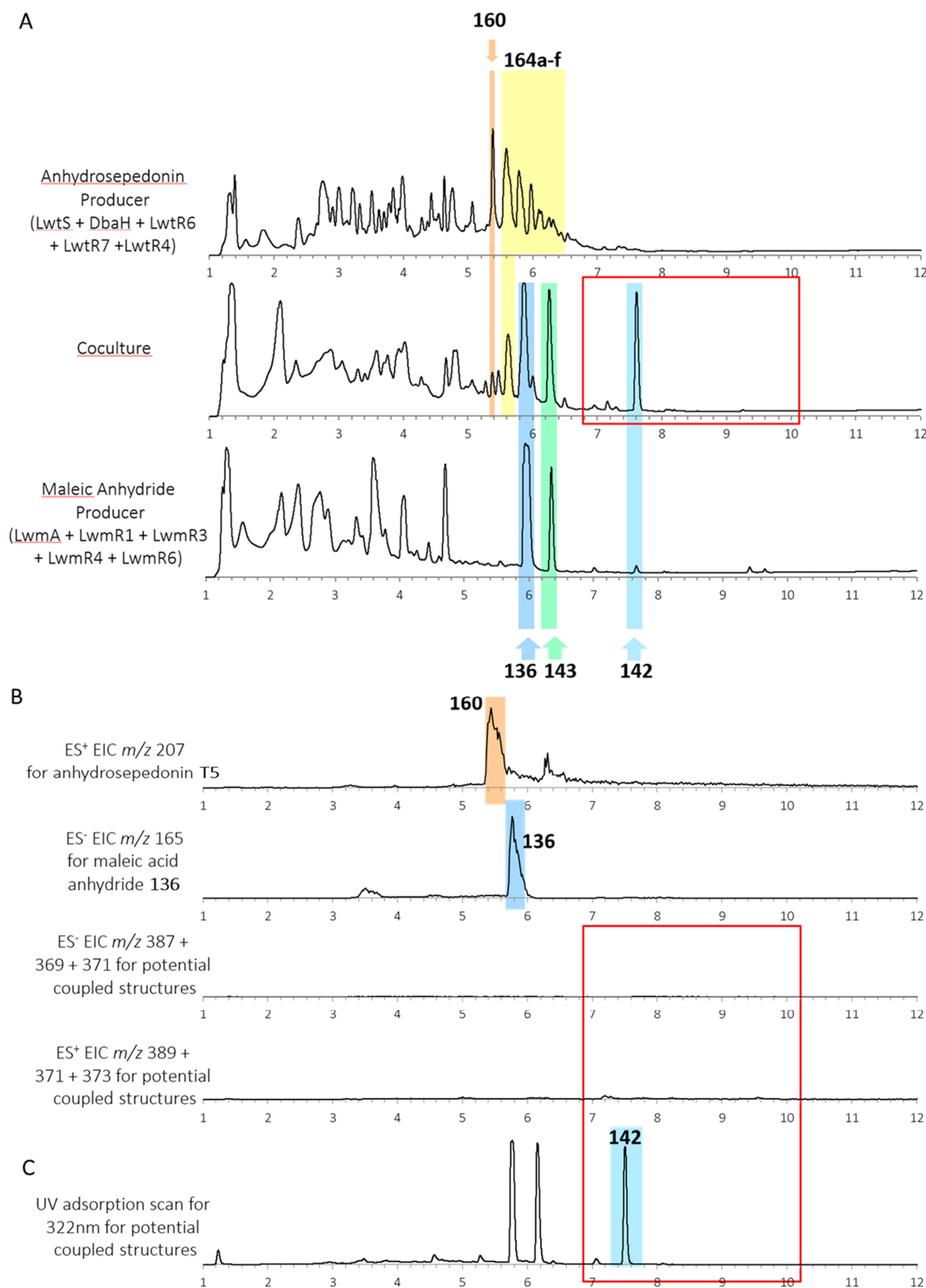
Figure 5.1 Cocultivation in comparison to maleic anhydride and tropolone producer (red box: expected retentiontimes for conjugates): **A**, DAD chromatograms; **B**, selected EIC chromatograms; **C**, UV absorption scan for 322 nm.

The spontaneous coupling of maleic acid anhydride **136** with sepedonin **90** or anhydrosepedonin **160** during the extraction and concentration was not observed. It is unknown if intermediates of the pathways are exported from the transformed cells into the media or are only present inside the cells. No transporting proteins were additionally transformed in transformants for either pathway. Accordingly, spontaneous coupling of tropolones or maleidride or their precursors may be possible, but not under the conditions investigated.

5.3.2 Heterologous Co-Expression of two Fungal BGCs

A second approach was carried out to investigate the possible coupling reaction. The two gene clusters were co-transformed into the heterologous host. Therefore, the vectors pTY-*ade-lwmA-lwmR1-lwmR3-lwmR4*, pTY-*argB-lwtS-lwtR6-dbaH-lwtR7* and pTY-*sC-lwmR2-lwmR6-lwtR2-lwtR4* (Figure 5.2), obtained by previously described methods, were transformed into the host *A. oryzae* NSAR1 by usual methods. It was speculated, that the MDC might be responsible for coupling reactions, analogous to other MDCs catalysing dimerizing reactions during maleidride biosynthesis (section 3.1.3.2). Thus, the gene *lwmR2* was added into the expression system. The MDC gene was amplified excluding the previously predicted signal peptide sequence (section 3.3.5).
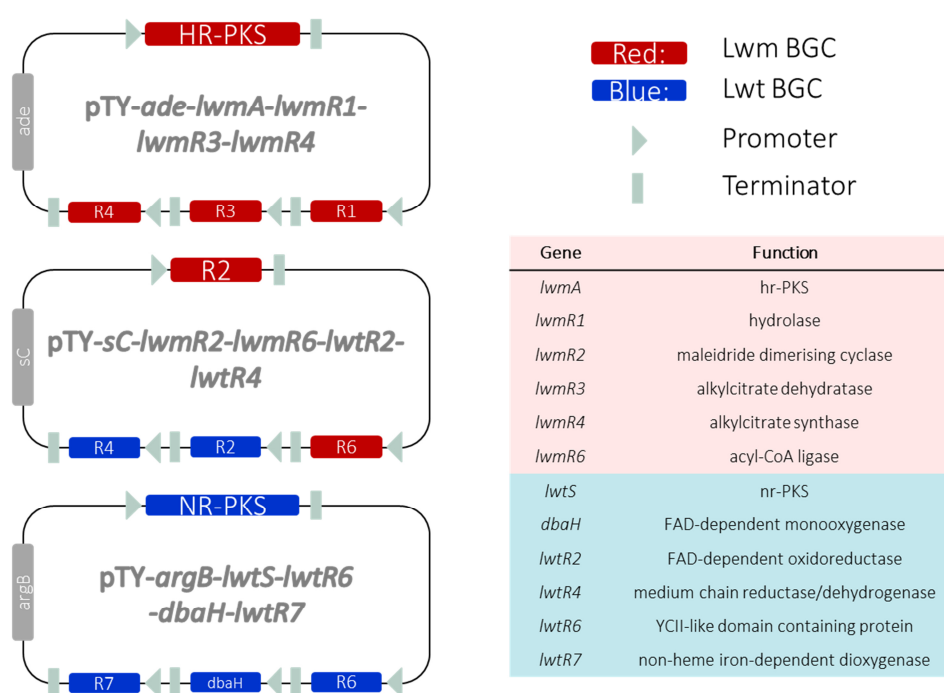

Gene	Function
lwmA	hr-PKS
lwmR1	hydrolase
lwmR2	maleidride dimerising cyclase
lwmR3	alkylcitrate dehydratase
lwmR4	alkylcitrate synthase
lwmR6	acyl-CoA ligase
lwtS	nr-PKS
dbaH	FAD-dependent monooxygenase
lwtR2	FAD-dependent oxidoreductase
lwtR4	medium chain reductase/dehydrogenase
lwtR6	YCII-like domain containing protein
lwtR7	non-heme iron-dependent dioxygenase

Figure 5.2 Overview of the vectors used for transformation of both BGCs.

5.3.2.1 LCMS Analysis of Transformants Including all Catalytic *lwm-* and *lwt*-BGC Genes

Transformation, selection, cultivation, extraction, and LCMS analysis were carried out as described previously. A screening for coupled products was carried out searching for the masses of potential coupled products **180** (388 Da), **181** (370 Da) and lienhwalide A **130** (372 Da, Figure 5.1 D) and the expected UV absorption maxima (search for 322 nm, Figure 5.1 E). No new coupled compounds were found in transformants including the shown vectors (Figure 5.3).
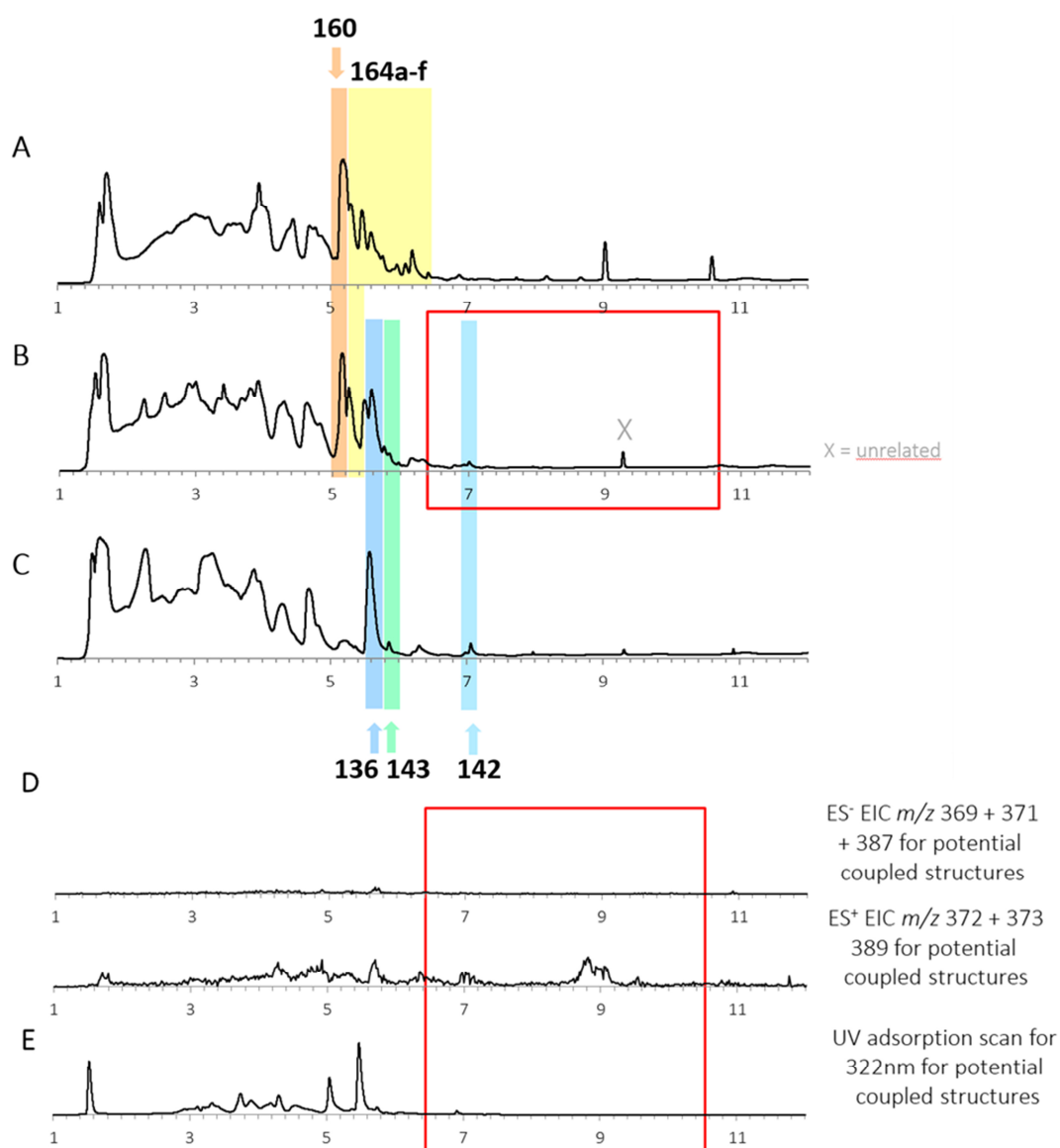
Figure 5.3 DAD chromatograms of host transfomed with two BGC core enzymes: **A**, example of tropolone producing transformant; **B**, example of producer of both compound classes; **C**, example of maleic anhydride producing transformant **D**, selected EIC chromatograms for transformant **B**; **E**, UV absorption scan for 322 nm for transformant B.

For nine transformants (from overall 9 producing transformants) tropolone-related compounds like **156** or **160** (Figure 5.3 A, Table 5-1) were observed. Maleic anhydrides like **136**, **142**, or **143** were only found in three of the producers (Figure 5.3 C, Table 5-1 A, F, G). Only for one transformant both classes of secondary metabolites were detected (Figure 5.3 B, Table 5-1 A).

Table 5-1 Scan of transformants for the expression of the two BGCs *lwt* and *lwm* together.

Structure	Maleic acid Anhydride Pathway Compounds			Tropolone Pathway Compounds					
	136	142	143	156	164 a-g	157/158	159	90	160
A	✓	✓	✓	-	-	-	✓	✓	✓
B	-	-	-	-	✓	-	✓	-	✓
C	-	-	-	-	✓	-	✓	✓	✓
D	-	-	-	-	✓	-	✓	✓	✓
E	-	-	-	-	✓	-	✓	✓	✓
F	✓	✓	✓	-	-	-	-	-	-
G	✓	✓	✓	-	-	-	-	-	-
H	-	-	-	-	-	-	✓	✓	✓
I	-	-	-	-	✓	-	✓	✓	✓
J	-	-	-	-	✓	-	✓	-	✓
K	-	-	-	-	✓	-	✓	-	✓

5.3.2.2 Analysis of gDNA for Transformants Combining two BGCs

Extraction of the gDNA from the transformants and PCR amplification of selected genes was carried out. The successful transformation of *pTY-ade-lwmA-lwmR1-lwmR3-lwmR4* was confirmed by PCR amplification of the gene *lwmR4* (Figure 5.4 B). For *pTY-argB-lwtS-lwtR6-dbaH-lwtR7* the confirmation took place by PCR amplification of the *lwtR7* gene (Figure 5.4 B). The transformation of the last vector *pTY-sC-lwmR2-lwmR6-lwtR2-lwtR4* was tested by PCR amplification of the gDNA for *lwtR4* gene (Figure 5.4 B). Additionally, the presence of the MDC encoding gene *lwmR2* was used for gDNA control.
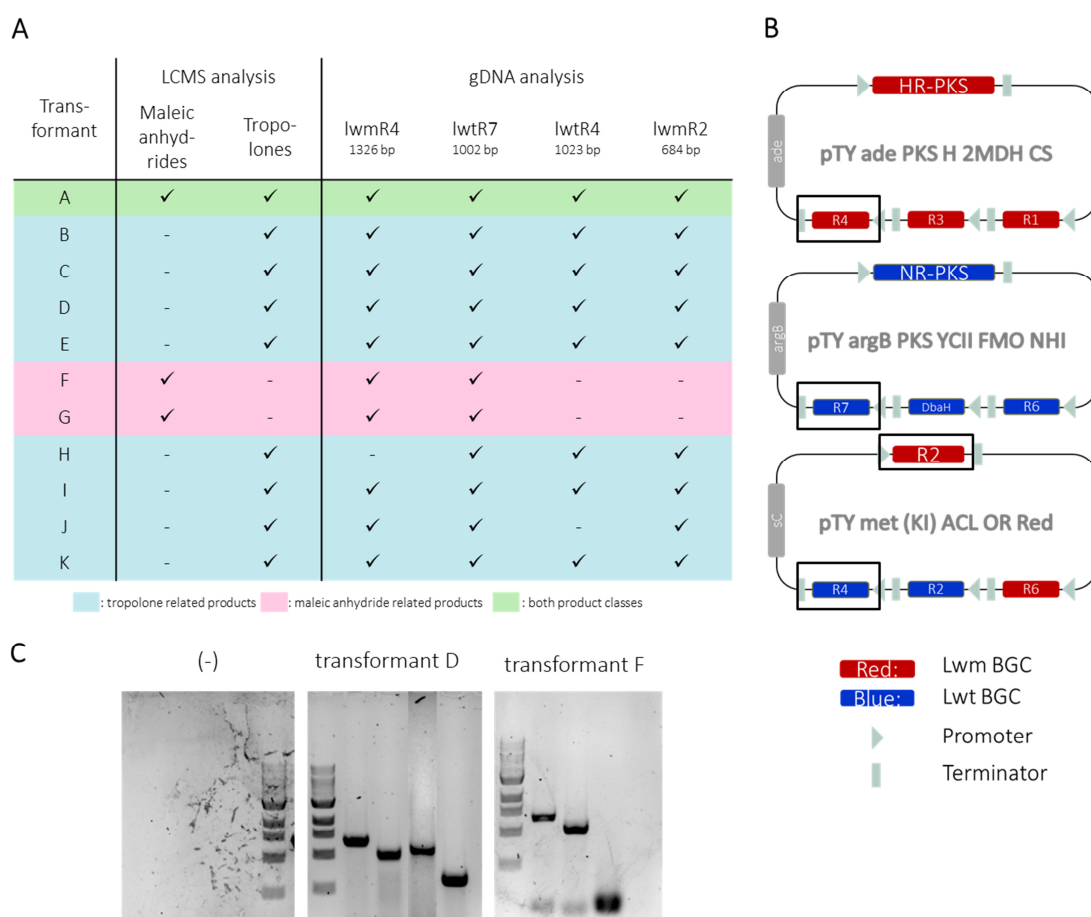
Figure 5.4 Summary of transformants with two transformed BGCs: **A**, table of transformants, products and gDNA analysis; **B**, schematic overview of transformed vectors; C, example of gDNA PCR amplification.

The transformation showed to be incomplete in transformants, which only produced maleic acid anhydrides as the main products (Figure 5.4 A, C; transformant F, G). For these transformants, PCR amplification of *lwtR4* and *lwmR2* failed. Thus, the vector pTY-*sC-lwmR2-lwmR6-lwtR2-lwtR4* was most likely not successfully integrated in the gDNA during transformation (Figure 5.4 B). While all catalytically necessary enzymes for maleic acid production are present in the other two included vectors (*lwmA, lwmR1, lwmR3, lwmR4*), the production of tropolones is not possible, when the *lwtR4* gene is absent, as it was shown to be required for the catalytic step of the sepedonin formation (section 4.3.7).

For most remaining transformants (Table 5-1 B-E, H-K), the tropolone related compounds are produced in excess in comparison to maleic acid anhydrides. This leads to overlaps in the peaks and therefore the maleic anhydrides were not detectable.

Only one transformant produced similar amounts of both classes of secondary metabolites (Figure 5.3 B, Figure 5.4 A, green). This transformant was also used in a time course, where no new compounds were detected in a range of 3-12 days. This finding confirms that spontaneous or MDC-catalysed coupling of the tropolone and maleidride moieties was not possible without biochemical catalysis during the co-expression of two BGCs.

5.4 Discussion: Lienhwalide Formation

5.4.1 Lienhwalide Coupling by MDC-related Mechanism

The co-fermentation experiment, where it is expected that intermediates of the two pathways cannot meet, only allows completed tropolones and maleic acid anhydrides to meet. This could be either in the media, or during extraction and concentration prior to LCMS analysis. No linkage here means that the two moieties cannot react spontaneously. In the co-expression intermediates could meet, but there is still no reaction. This shows that a biochemical catalyst is required, or the correct intermediate(s) are not formed.

One aspect of the unsuccessful coupling is speculated to be related to unstable intermediates. The proposed mechanism of the coupling reaction (Scheme 5.1) is based on **136** and **172** as intermediates. For the reaction, compound **137**, which results from the decarboxylation of maleic acid anhydride **136**, attacks the intermediate **T16**, leading to the lienhwalide structure. If the coupling reaction is related to the proposed mechanism, both structures are required in reaction mixtures in sufficient amounts to enable the reaction. **172** of the sepedonin pathway was never observed during experiments of the tropolone pathway (section 4). Furthermore, the coupling reaction is in competition to the reduction of **172**, leading to sepedonin (Scheme 5.1 B). *In vitro* protein assays (section 4.3.7) showed clearly that the reaction of LwtR7 and LwtR4 work closely together, which raises the question how the intermediate **172** is present in the reaction mixture of the wild type fungus. A chaperone like enzyme might be required for a successful coupling reaction between sepedonin- and maleic acid anhydride intermediates.

The presence of the putative MDC *in vivo*, which homologs are known to catalyse coupling reactions in other related fungal pathways, did not catalyse dimer formation. Similar to the biosynthetic pathway of cordyanhydride B (section 3.3.5), the MDC is not (alone) able to form any new compounds.

To overcome the hurdles related to the unstable intermediates, further experiments are conceivable. The introduction of PEBP proteins from other related pathways could improve the stability of **137** (section 3.1.3.2). However, this is speculative, as the function of PEBP has not been fully elucidated in literature. To increase the stability of the intermediate **172**, it would be possible to exclude the later pathway steps, for example by excluding the expression of LwtR4, which could lead to accumulation of **172**. However, this approach is also speculative, since the intermediate **172** has never been detected before in *in vivo* and *in vitro* experiments. Nevertheless, these two approaches could potentially increase the stability of the intermediates to enable spontaneous coupling of the intermediates.

5.4.2 Heterologous Co-Expression of two Complete BGCs

Interestingly, the transformants including two BGCs were either able to produce tropolone-related compounds in high amounts with no maleic acid anhydride detectable or only maleic acid anhydrides were detected. A similar amount of both compound families was only found in one of eleven transformants. The gDNA analysis by PCR amplification showed that in cases where maleic acid anhydride production is dominant, the tropolone pathway is incomplete due to the absence of *lwtR4*, which is located on the pTY-*sC-lwmR2-lwmR6-lwtR2-lwtR4* vector. Accordingly, the biosynthesis of tropolone-related products was observed to be dominant in nearly all transformants, when all selected genes were successfully transformed. An imbalanced uptake of acetyl-CoA units for the polyketide production (section 1.2) is speculated to occur, and leads to different production levels of the compound families.

H. lienhwacheense produces lienhwalides in high amounts while the precursors of maleic acid anhydride related compounds were found as minor components in the extracts. In contrast, sepedonin related structures and other free tropolones are not found in extracts of *H. lienhwacheense* outside of lienhwalide structures (section 3.3.1). Complex systems of many enzymes and intermediates working together to form one

secondary metabolite are usually highly regulated in fungi. In the WT fungus, the balance between the two compound classes is most likely regulated at the gene expression level (promotors, transcription factors) and/or by compartmentation of enzymes and reactive steps in different organelles with different, appropriate conditions for the potentially unstable intermediates. While the regulation mechanisms are unknown for *H. lienhwacheense*, no regulation at all takes place in the heterologous expression approach in *A. oryzae* NSAR1. If certain conditions or concentrations of intermediates or proteins are necessary for the successful coupling reaction, the lack of regulation in *A. oryzae* could rationalise the failed coupling of maleic acid anhydrides and   tropolones.

5.4.3 Alternative Mechanism for Lienhwalide Formation

The analysis of the *H. lienhwacheense* genome revealed that a putative hetero-Diels-Alderase (hDA) encoding gene is located near the *lwm*-BGC (Figure 5.5, pink). The putative gene showed homology to hetero-Diels-Alderases by using the blastp[150,151] algorithm against the SwissProt[120] database. The search resulted in two hits (Figure 5.5) for the putative hetero-Diels-Alderase from *Phoma sp.* and *Sarocladium sp.* for the biosynthesis of eupenifeldin **162** (query cover 88 %, approx. 30 % identity) and xenovulenes (query cover 91 %, approx. 31 % identity). The putative hetero-Diels-Alderase encoding gene will further be referred to as *lwmR8*.

| Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Putative hetero-Diels-Alderase; AltName: Full=Eupenifeldin biosynthesis cluster protein F; Flag...	Phoma sp.	122	122	88%	8e-31	29.91%	352	A0A4P8GEA3.1
RecName: Full=Putative hetero-Diels-Alderase asR5; AltName: Full=Xenovulene A biosynthesis cluster protein ...	Sarocladium sp. ...	116	116	91%	4e-28	31.46%	401	A0A2U8U2M1.1

Figure 5.5 *lwm*-BGC (blue box) expanded with neighboring putative genes (red box = blastp hits for *lwmR8*)

The putative protein sequence of LwmR8 was used to predict the protein structure using AlphaFold.[31] The resulting protein structure was used as a template for structural search of the DALI server database.[192] The two hits with most similarity include the hDAs PycR1 (32 % identity) and EupR1 (34 % identity).

PycR1, EupF and AsR5 are known to catalyse similar hDA reactions of tropolone quinomethides such as **185** and **191** (Scheme 5.3). After reduction of stipitaldehyde **183** to intermediate **184**, PycR1 and EupF catalyse the formation of an ortho-quinomethide **185**. Then, the hetero-Diels-Alder reaction between reduced stipitaldehyde derivative **185** and the sesquiterpenes **186** (EupF) or **188** (PycR1) to the meroterpenoids **187** and **189** takes place respectively. For AsR5, intermediate **190** is the result of prior oxidation of **183**. Then, **190** is converted to the ortho-quinomethide **191** that reacts with **192** to give the meroterpenoid **193**.[95,180,203,204]

Scheme 5.3 Reactions of EupF, AsR5 and PycR1.[95,180,203–205]
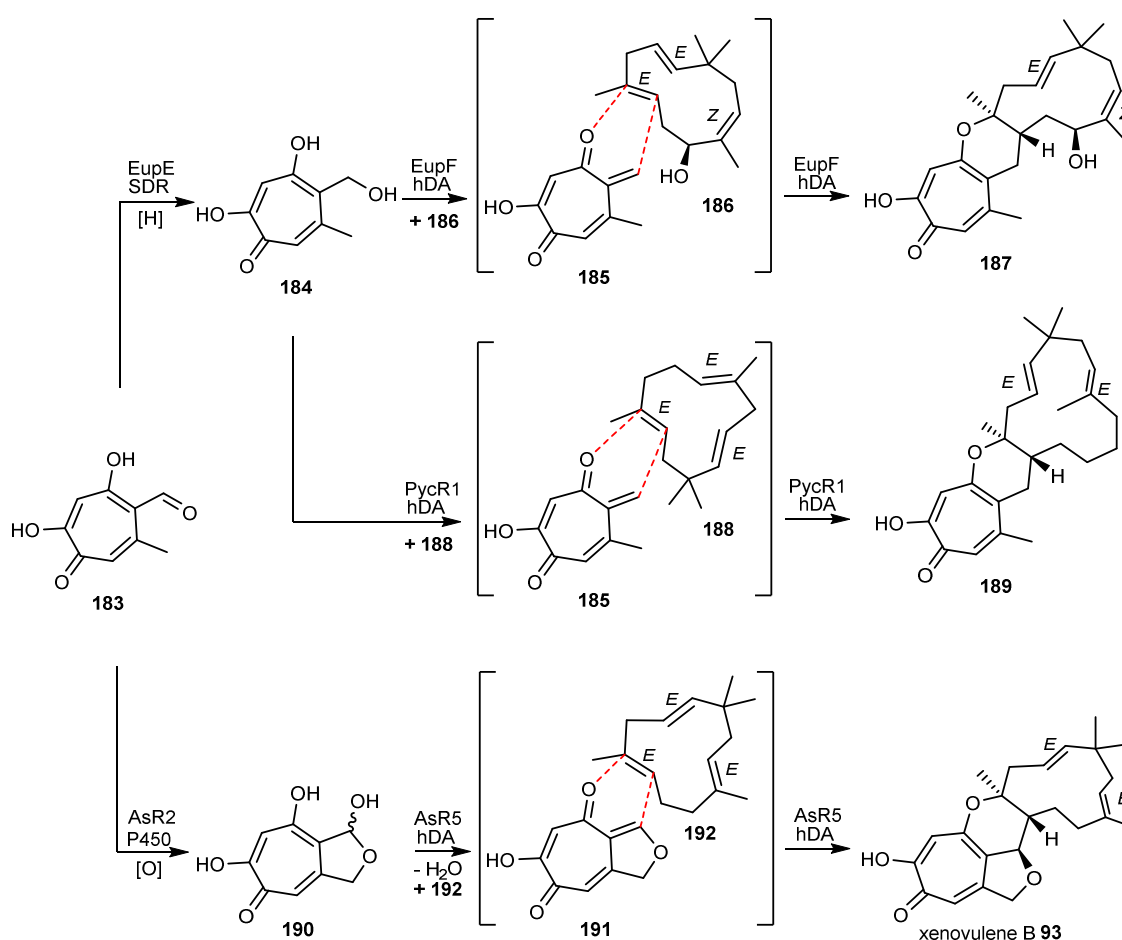
Remarkably, all three enzymes related to LwmR8 react with a tropolone intermediate **184/185** or **190/191** (Scheme 5.3). This is good circumstantial evidence that LwmR8 reacts with tropolones. The close location of *lwmR8* to the *lwm*-BGC (Figure 5.6) could indicate that LwmR8 is part of the *lwm*-BGC. Thus, a connection between maleic acid anhydrides and tropolones correlates with the presence of an hDA enzyme (Figure 5.6).
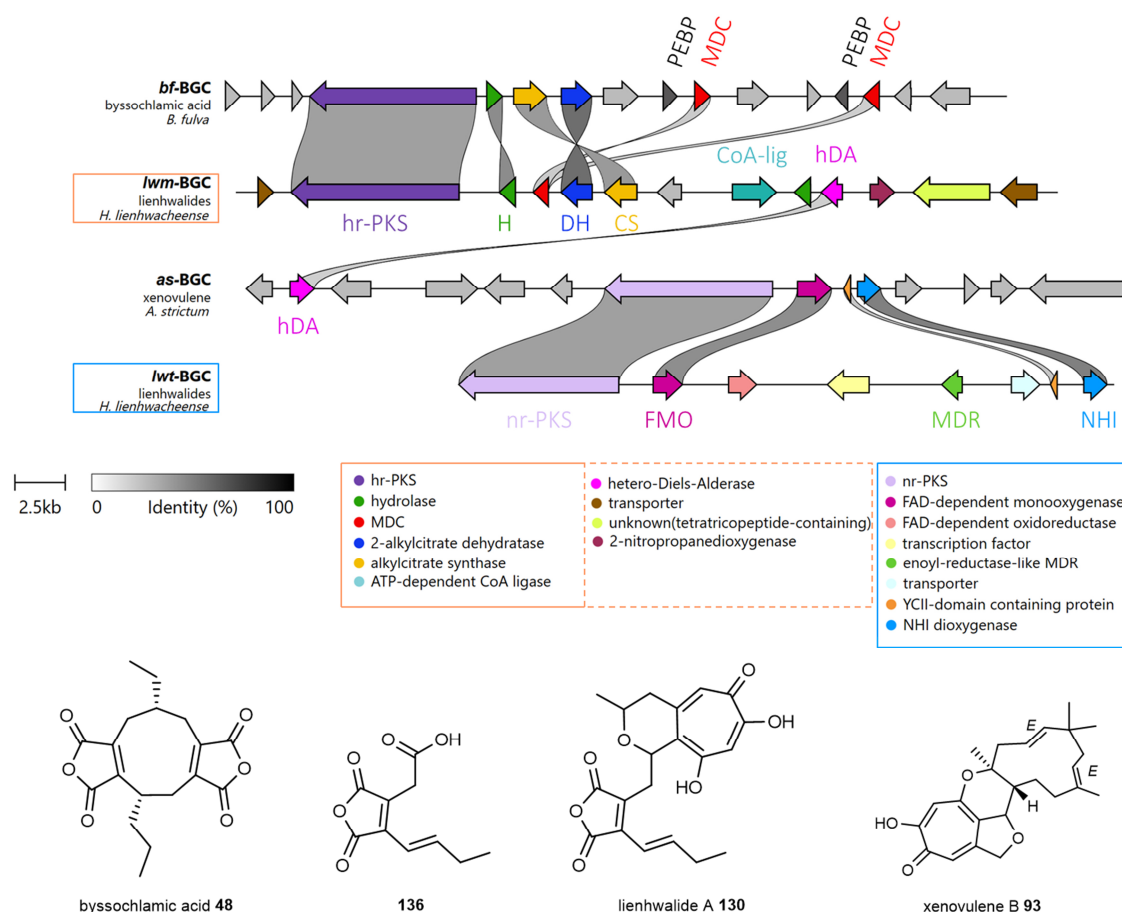
Figure 5.6 Gene cluster analysis *lwt-* and *lwm*-BGC from *H. lienhwacheense* in comparison to *as*-BGC (*A. strictum*) and *bf*-BGC (*B. fulva*); clinker minimal alignment sequence identity was adjusted to 0.25.[95,158,160,203]

The putative Diels-Alderase LwmR8 is therefore a candidate to catalyse a coupling reaction to form lienhwalides, according to Scheme 5.4. Maleic acid anhydride **136** decarboxylates to **138**. In maleidride pathways, **136** and **138** are coupled by the MDC (and PEBP). For lienhwalides, the decarboxylated species **138** is speculated to act as the dienophile to the α,β-unsaturated ketone motif of the oxidised tropolone intermediate **172**. Structure **194** is the result of the Diels-Alder reaction, which ring-opens to **180**, the

precursor to lienhwalide A **130**. The hypothesis requires further investigations to give evidence for the involvement of LwmR8 in the coupling reaction.
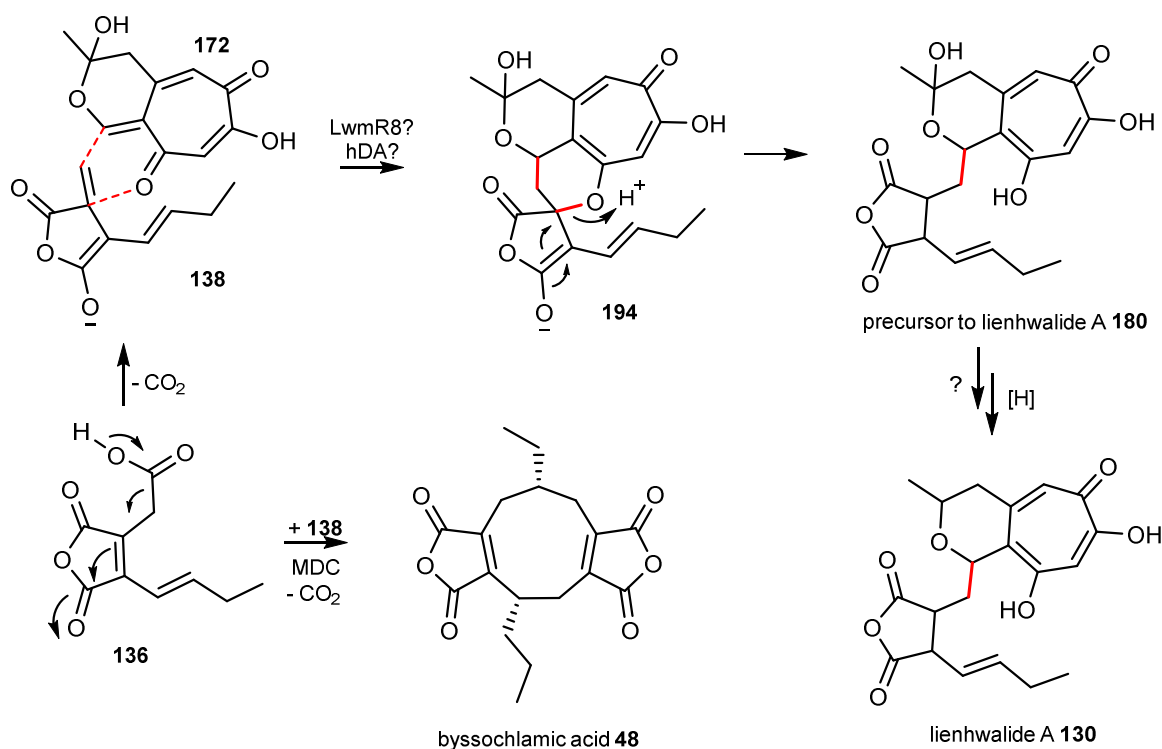
Scheme 5.4 Proposed mechanism of dimerisation by hetero Diels Alder reaction.

Interestingly, a putative hDA encoding gene was found in *A. nidulans* near the *dba*-BGC (Figure 5.7, red box). The presence of a putative hDA encoding gene near the antibiotic C **91** producing cluster may be an indication for the involvement of the hDA in the pathway. This hypothesis also requires further investigations to give evidence for a possible function of the hDA.
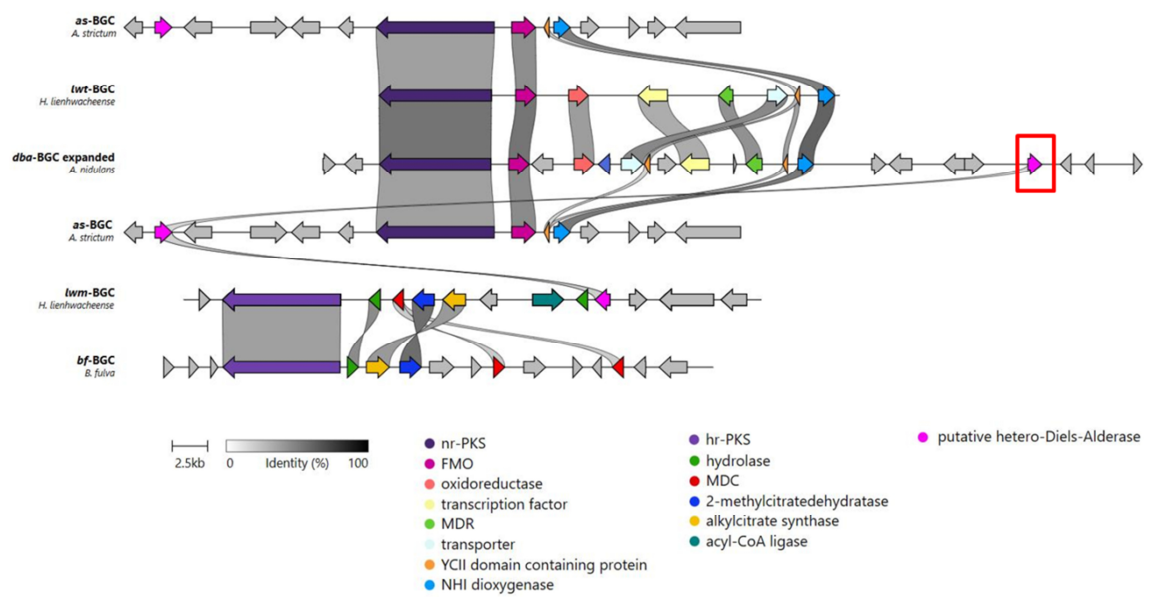
Figure 5.7 Gene cluster analysis of *dba*-BGC from *A. nidulans* with *lwt*- and *lwm*-BGC from *H. lienhwacheense* in comparison to *as*-BGC (*A. strictum*) and *bf*-BGC (*B. fulva*); clinker minimal alignment sequence identity was adjusted to 0.25.[95,158,160,177,178,203]

5.5 Conclusion and Prospect of Lienhwalide Biosynthesis

The overall aim of the elucidation of the biosynthetic pathway of lienhwalides (and related compounds like cordyanhydride B) was successful, but not completed. The biosynthetic pathway of each individual compound class was elucidated and the core enzymes identified.

For the formation of maleic acid anhydrides the requirement of the core-enzymes LwmA, LwmR1, LwmR3 and LwmR4 is in agreement to the pathway of other related maleidrides (Scheme 5.5). This was confirmed by the formation of **136** and **142** *in vivo* and **146** in *in vitro* studies. Additionally, the efficiency of the biosynthesis was improved by the acyl-CoA ligase LwmR6. In addition to these compounds, the biosynthesis of the putative intermediate **143** for the cordyanhydride B **61** formation was also enabled *in vivo*. However, the coupling reaction to form cordyanhydride B **61** is still unknown. Future studies of the mysterious biosynthesis of **143** and the subsequent coupling reaction could determine a new mechanism of coupling in comparison to known dimerization reactions of MDCs.

The sepedonin pathway of the *lwt*-BGC was elucidated, which is catalysed by LwtS, LwtR1, LwtR7 and LwtR4 (Scheme 5.5). The requirement of the reductive enzyme LwmR4 for the formation of the tropolone core structure is new in comparison to known tetraketide tropolone pathways, where the reductive enzyme is not required by the pathway. However, for this pathway some biosynthetic steps also remain a mystery. The formation of antibiotic C **91** and *O*-methylated intermediates for lienhwalide B **131** and C **132** production is unknown. The responsible enzymes for these reactions are not clustered adjacent the other core enzymes.

The biosynthesis of both, maleic acid anhydrides and tropolones, enables formation of lienhwalides (Scheme 5.5). However, the mechanism and the required enzymes for the final lienhwalide formation remains unknown. The approach to simultaneously transform all known catalytic enzymes from two BGCs in one transformant was carried out, but did not lead to lienhwalide related compounds. A catalytic function of the MDC LwmR2 was not found during heterologous expression experiments. Nevertheless, the path for future investigations in the field of coupling reactions has been laid. The final formation of the heterodimer will be the focus of

future studies, since the reaction could require a previously unknown type of reaction and/or enzyme.
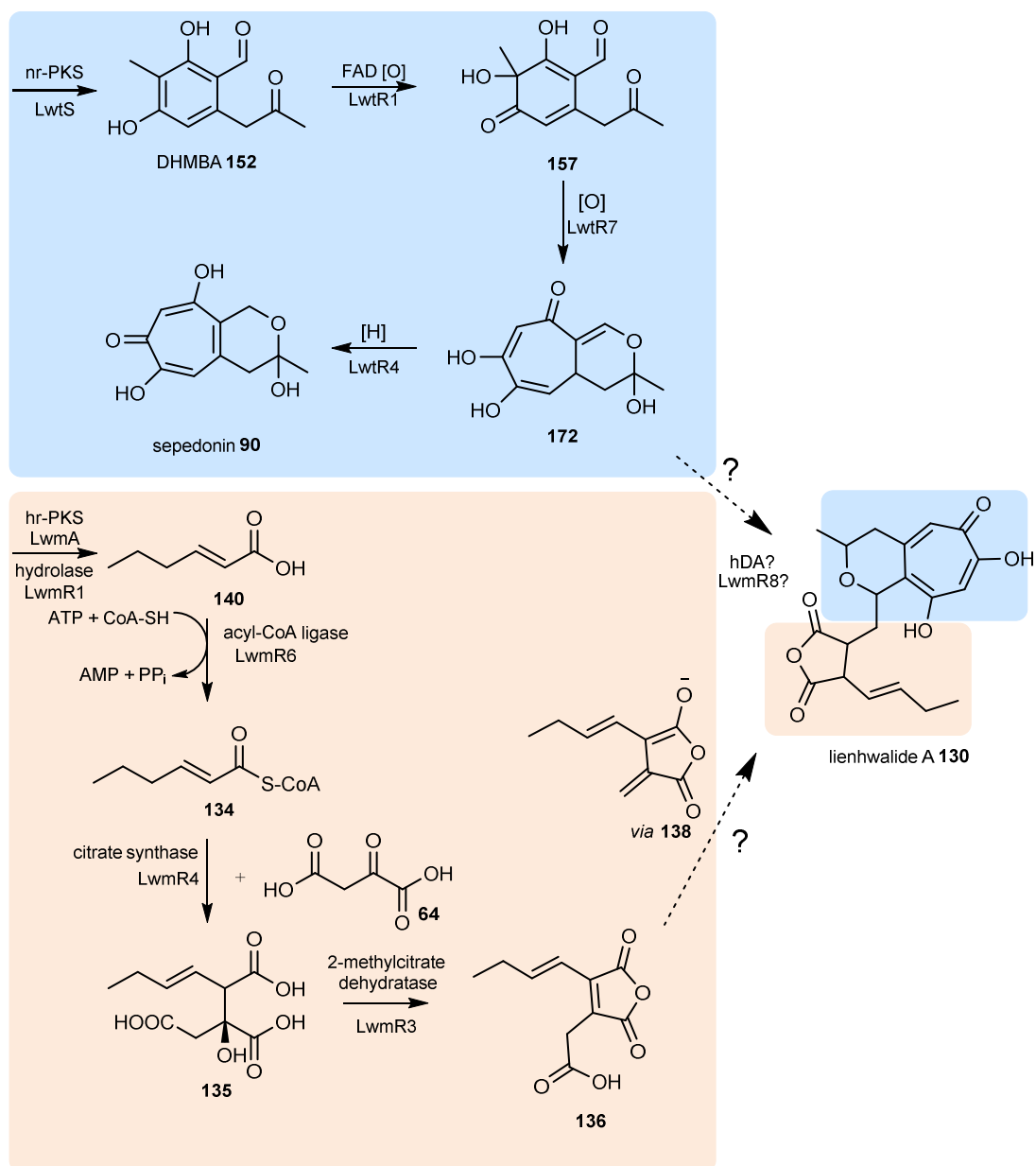
Scheme 5.5 Overall illustration of lienhwalide pathway (orange = maleic acid anhydride pathway, blue = tropolone pathway).

6 Experimental

All mentioned chemicals were purchased in analytical grade (or higher) from one of the following companies: AMRESCO (Solon, OH, USA), Acros Organics (part of Thermo Fisher Scientific, Geel, Belgium), Applichem (Darmstadt, Germany), Bio-Rad (Hercules, CA, USA), Honeywell (Charlotte, NC, USA), Carl Roth (Karlsruhe, Germany), New England Biolabs (Beverly, MA, USA), Sigma Aldrich (Steinheim, Germany) or Thermo Fisher Scientific (Waltham, MA, USA).

6.1 Biological Methods

6.1.1 Media, Buffers and Antibiotics

All media (Table 6-1), buffers and solutions (Table 6-2) used in this work were prepared with Millipore water (GenPure Pro UV/UF millipore device, *Thermo Fisher Scientific*) and sterilised by autoclaving 15 min at 121 °C (Autoclave 2100 Classic, *Prestige Medical*) or sterilised by disposable syringe filters (pore size 0.2 – 0.4 µm, *Carl Roth*). The pH was adjusted with 2 M HCl or 2 M NaOH by using a FiveEasy Standard pH Meter Line (*Mettler Toledo*).

Table 6-1 Media used in this work.

Media	Composition [% (w/v)]
CMP media	3.5 % Czapek Dox broth (Duchefa Biochemie), 2 % D(+)-Maltose monohydrate (Duchefa Biochemie), 1 % Polypeptone (Roth), 3.5 % Czapek Dox broth (Duchefa Biochemie)
CZD/S agar	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.05 % Adenine (Roth), 0.15 % L-Methionine (Roth), 1.5 % Agar (Duchefa Biochemie)
CZD/S agar w/o methionin	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.05 % Adenine (Roth), 1.5 % Agar (Duchefa Biochemie)
CZD/S softagar	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.05 % Adenine (Roth), 0.15 % L-Methionine (Roth), 0.8 % Agar (Duchefa Biochemie)
CZD/S softagar w/o methionin	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.8 % Agar (Duchefa Biochemie)

CZD/S1 agar	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.15 % L-Methionine (Roth), 1.5 % Agar (Duchefa Biochemie)
CZD/S1 agar w/o methionin	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 1.5 % Agar (Duchefa Biochemie)
CZD/S1 softagar	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 0.15 % L-Methionine (Roth), 0.8 % Agar (Duchefa Biochemie)
CZD/S1 agar w/o methionin	3.5 % Czapek Dox broth (Duchefa Biochemie), 18.22 % D-Sorbitol (=1 M) (Roth), 0.1 % Ammonium sulfate (Roth), 1.5 % Agar (Duchefa Biochemie)
DPY agar	2 % Dextrin from potato starch (Sigma Aldrich), 1 % Polypeptone (Roth), 0.5 % Yeast extract (Duchefa Biochemie), 0.5 % Monopotassium phosphate (Roth), 0.05 % Magnesium sulfate hexahydrate (Sigma Aldrich), 2.5 % Agar (Duchefa Biochemie)
DPY media	2 % Dextrin from potato starch (Sigma Aldrich), 1 % Polypeptone (Roth), 0.5 % Yeast extract (Duchefa Biochemie), 0.5 % Monopotassium phosphate (Roth), 0.5 % Magnesium sulfate hexahydrate (Sigma Aldrich)
GN media	2 % D(+)-Glucose Monohydrate (Roth), 1 % Nutrient broth Nr. 2 from Oxoid (Fisher Scientific)
Hutner's Trace Element Solution	5,00 % EDTA disodium salt (Roth), 2,20 % (w/v) Zinc sulfate heptahydrate (Acros), 1,14 % (w/v) Boric acid (Roth), 0,506 % (w/v) Manganese-(II)-chloride tetrahydrate (Roth), 0,161 % (w/v) Cobalt-(II)-chloride hexahydrate (Roth), 0,157 % (w/v) Copper-(II)-sulfate heptahydrate (Roth), 0,11 % (w/v) Ammonium molybdate tetrahydrate (Roth), 0,499 % (w/v) Iron-(II)-sulfate heptahydrate (Sigma Aldrich), pH 6.5-6.8; Mix all except EDTA, Bring to boil, then add EDTA (as solution)
LB agar	0.5 % Yeast extract (Duchefa Biochemie), 1 % Tryptone (Duchefa Biochemie), 0.5 % Sodium chloride (Roth or VWR), 1.5 % Agar (Duchefa Biochemie)
LB media	0.5 % Yeast extract (Duchefa Biochemie), 1 % Tryptone (Duchefa Biochemie), 0.5 % Sodium chloride (Roth or VWR)
London agar	1 % ´´-Glucose Monohydrate (Roth), 2 % Aspergillus Salt Solution (2.6 % Potassium chloride (Roth), 2.6 % Magnesium sulfate hexahydrate (Sigma Aldrich), 7.6 % Monopotassium phosphate (Roth), 5 % Hunter's Trace Element solution), 1 % nitrogen solution (10 mM Sodium nitrate (Roth)), 0.1 % PABA, 2 % Agar (Duchefa Biochemie)
London medium	1 % D(+)-Glucose Monohydrate (Roth), 2 % Aspergillus Salt Solution (2.6 % Potassium chloride (Roth), 2.6 % Magnesium sulfate hexahydrate (Sigma Aldrich), 7.6 % Monopotassium phosphate (Roth), 5 % Hunter's Trace Element solution), 1 % nitrogen solution (10 mM Sodium nitrate (Roth)), pH 6.5; 0.1 % PABA; Induction with 1µg/ml Zeocin (InvivoGen) after 16 h
MMK2 media	4 % D(-)-Mannitol (Roth), 0,5 % Yeast extract (Duchefa Biochemie), 0,43 % Murashige & Skoog salt (Phygenera)
PDB media	2.4 % (w/v) Potato Dextrose broth (Formedium)
SM-URA agar	0.17 % Yeast nitrogen base (Sigma Aldrich), 0.5 % Ammonium sulfate (Roth), 2 % D(+)-Glucose monohydrate (Roth), 0.077 % Complete supplement mixture minus Uracil (Sigma Aldrich), 2.5 % Agar (Duchefa Biochemie)

SOC media		0.5 % Yeast extract (Duchefa Biochemie), 2 % Tryptone (Duchefa Biochemie), 0.06 % Sodium chloride (Roth or VWR), 0.02 % Potassium chloride (Roth), 25 mM Magnesium chloride hexahydrate (Roth), 1 % D(+)-Glucose monohydrate (Roth)
YMG media		0.4 % D(+)-Glucose Monohydrate (Roth), 0.4 % Yeast extract (Duchefa Biochemie), 1 % (w/v) Malt extract (Roth)
YPAD agar		1 % Yeast extract (Duchefa Biochemie), 2 % Tryptone (Duchefa Biochemie), 2 % D(+)-Glucose monohydrate (Roth), 0.03 % Adenine (Roth), 1.5 % Agar (Duchefa Biochemie)
YPAD media		1 % Yeast extract (Duchefa Biochemie), 2 % Tryptone (Duchefa Biochemie), 2 % D(+)-Glucose monohydrate (Roth), 0.03 % Adenine (Roth)

Table 6-2 Buffers and solutions used in this work.

Application	Buffer/Solution	Composition
A.oryzae transformation	Solution 1	0.8 M Sodium chloride, 10 mM Calcium chloride, 50 mM Tris-HCl, pH 7.5
	Solution 2	60 %(w/v) PEG 3350, 0.8 M Sodium chloride, 10 mM Calcium chloride, 50 mM Tris-HCl, pH 7.5
Agarose gel electrophoresis	50x TAE buffer	2 M Tris acetate, 0.05 M EDTA, pH 8.3
Protein purification Phosphate buffers (LwmR4)	Lysis buffer	50 mM Sodium phosphate, 10 mM Imidazole, 300 mM NaCl, 10 % Glycerol, pH 7.5
	Wash Buffer	9.5 ml Lysis buffer, 0.5 ml Elution buffer 2
	Elution buffer 1	3.5 ml Lysis buffer, 1.5 ml Elution buffer 2
	Elution buffer 2	50 mM Sodium phosphate, 500 mM Imidazole, 150 mM NaCl, 10 % Glycerol, pH 7.5
	Storage buffer	50mM Sodiumphosphate, 20 mM NaCl, 10 % Glycerol, pH 7.5
	Assay buffer	0.5 M Na_2HPO_4, NaH_2PO_4
Protein purification Tris-HCl buffers (LwtR1, LwtR4, LwtR7)	Lysis buffer	50 mM Tris/HCl, 150 mM NaCL, 10% (v/v) Glycerol, 20 mM Imidazole, pH 8
	Wash Buffer	9.5 ml Lysis buffer, 0.5 ml Elution buffer 2
	Elution buffer 1	3.5 ml Lysis buffer, 1.5 ml Elution buffer 2
	Elution buffer 2	50 mM Tris/HCl, 150 mM NaCL, 10% (v/v) Glycerol, 500 mM Imidazole, pH 8
	Storage buffer	50 mM Tris/HCl, 20 % (v/v) Glycerol, add ddH2O, pH 7.5
	Assay buffer	1 M TES (N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid)
SDS PAGE	10x SDS buffer	25 mM Tris-HCl, 192 mM Glycine, 0.1 % (w/v) SDS, pH 8.3
	Coomassie dye	25 % (v/v) acetic acid, 10 % (v/v) isopropanol, 0.1 % (w/v) coomassie
	Coomassie bleach	25 % (v/v) acetic acid, 10 % (v/v) isopropanol
	4x Lämmli	10 % (v/v) β-mercaptoethanol, 0.25 % (w/v) bromophenol blue, 30 % glycerol, 0.25 % xylene cyanol

Antibiotics (Carbenicillin, Kanamycin) were prepared in 1000x concentrated stock solution in distilled water in a concentration of 50 µM. Stocks were sterilized through 0.45 µm syringe filter and stored at -20 °C. Antibiotics were diluted for a final concentration of 50 µg/ml by adding them to the final media.

6.1.2 Strains, Vectors and Oligonucleotides

Information about the used strains in this work are summarised in Table 6-3.

Table 6-3 Summary of used strains.

Organism	Strain	Genotype	Reference
E. coli	OneShot® Top10	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG*	Thermo Fisher Scientific
	OneShot® ccdB survival 2T1ᴿ	*F⁻mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (Strᴿ) endA1 nupGfhuA::IS2*	Thermo Fisher Scientific
	OneShot® BL21 (DE3)	*F⁻ ompT hsdSв (rв⁻mв⁻) gal dcm (DE3)*	Thermo Fisher Scientific
S. cerevisiae	CEN.PK2	*MATa/a ura3-52/ura3-52trp1-289/trp1-289 leu2-3_112/leu2-3_112 his3D1/his3 D1MAL2-8C/MAL2-8C SUC2/SUC2*	Euroscarf
A. oryzae	NSAR1	*ΔargB sC⁻ ΔadeA niaD⁻*	Lazarus group, Bristol
H. lienhwacheense	MFLUCC 14-1231	wild-type	Stadler group, Braunschweig[141]
A. nidulans	AGB 552	*pabaA1; ΔnkuA::argB*	Jennifer Gerke[177,178]
	AGB 1418	*ΔtroA(AN7893)::β-six-site; pabaA1; ΔnkuA::argB*	Jennifer Gerke[177,178]

Information about the used vectors in this work are summarised in Table 6-4 for section 2 and in Table 6-5 for section 3, 4, and 5.

Table 6-4 Summary of used vectors for tenellin-project.

Name	Description	Selection	Reference
pEY *tenS*	pEYA shuttle vector including *tenS* gene	Kan^R	Dr. Oliver Piech
pTY-*argB-tenC*	fungal expression vector with trans-ER *tenC* gene	*argB*, arb^R, *URA3*	Dr. Oliver Piech
pTY-*argB-tenS-tenC*	pTY-*argB-tenC* including *tenS* gene	*argB,* arb^R, *URA3*	this work
pEYA-*tenS** sbh:DmbS	T2395 to V2409 swap to DmbS sequence	Kan^R	this work
pEYA-*tenS** sbh:MilS	T2395 to V2409 swap to MilS sequence	Kan^R	this work
pEYA-*tenS** 2400N, L2401R, 2404M, V2406A	Mutations 2400N, L2401R, 2404M, V2406A	Kan^R	this work
pTY-*argB-tenS*-tenC* sbh:DmbS	fungal expression vector with trans-ER *tenC* gene, T2395 to V2409 swap to DmbS sequence	*argB,* arb^R, *URA3*	this work
pTY-*argB-tenS*-tenC* sbh:MilS	fungal expression vector with trans-ER *tenC* gene, T2395 to V2409 swap to MilS sequence	*argB,* arb^R, *URA3*	this work
pTY-*argB-tenS*-tenC* 2400N, L2401R, 2404M, V2406A	fungal expression vector with trans-ER *tenC* gene, mutations 2400N, L2401R, 2404M, V2406A	*argB,* Carb^R, U*RA3*	this work
pEYA-*tenS** AlaN	pEYA-tenS with alanine muatations at position N (N between D2394 to S2410)	Kan^R	this work
pTY-*argB-tenS*AlaN-tenC*	pTY-*argB-tenC* including *tenS* gene alanine muatations at position N (N between D2394 to S2410)	*argB,* Carb^R, *URA3*	this work

Table 6-5 Summary of used vectors for lienhwahlide-project (section 3, 4, and 5).

Name	Selection	Reference
pEYA-*lwmA*	Kan^R	this work
pEYA-*lwm2*	Kan^R	this work
pTY-GS-*ade-lwmR1-R3-R4*	*ade*, Carb^R, *URA3, ccdB*	this work
pTY-*ade-lwmA -lwmR1-R3-R4*	*ade*, Carb^R, *URA3*	this work
pTY-GS-*argB-lwmR6*	*argB*, Carb^R, *URA3, ccdB*	this work
pTY-*argB-lwmR2-lwmR6*	*argB*, Carb^R, *URA3*	this work
pTY-*argb-lwmR2-lwmR6-lwmR5*	*argB*, Carb^R, *URA3*	this work
pEYA-*lwtS*	Kan^R	this work, Dr. Jin Feng
pTY-*argB-lwtS*	*argB*, Carb^R, *URA3*	this work
pTY-GS-*sC-lwtR6*	*sC*, Carb^R, *URA3, ccdB*	this work, Dr. Jin Feng
pTY-*argB-lwtS -lwtR6-R7-dbaH*	*argB*, Carb^R, *URA3*	this work
pTY-*argB-lwtS-lwtR6-dbaH*	*argB*, Carb^R, *URA3*	this work
pTY-GS-*ade-lwtR2-R4*	*ade*, Carb^R, *URA3, ccdB*	this work
pTY-GS-*ade-lwtR4*	*ade*, Carb^R, *URA, ccdB*	this work
pTY-GS-*ade-lwtR2*	*ade*, Carb^R, *URA, ccdB*	this work
pTY-*ade-P450-lwtR4*	*ade*, Carb^R, *URA, ccdB*	this work
pTY-*ade-P450-lwtR4-lwtR5*	*ade*, Carb^R, *URA, ccdB*	this work
pTY-GS-*sC-lwmR6-lwtR2-lwtR4*	*sC*, Carb^R, *URA3, ccdB*	this work
pTY-*sC-lwmR2-lwmR6-lwtR2-lwtR4*	*sC*, Carb^R, *URA3*	this work

Information about the used oligonucleotides in this work are summarised in Table 6-6 (tenellin project) and Table 6-7 (lienhwalide project). They were synthesized by Sigma Aldrich and supplied lyophilised. Oligonucleotides were dissolved according to the manufacturer's instructions.

Table 6-6 Summary of used oligonucleotides for tenellin experiments.

Description	Name	Sequence
Fragment 1	A1	CGCATTCTCCACGGCTATTGGAAAC
	A2	CCTGCTGATCTTCCTGAACGTCG
Fragment 2	A3	GATGCCCAGCTCCAAAAGGCC
	A4	CTGTATTATTCAGAATGGCAGCGCTCGAGCTTAGCAAGACAAAAAAGTC
Fragment 3	A5	CGGCTCCCACAACATCATAATGG
	A6	GCTTTGGACGATGCGGCGCGG
Fragment 2/3 for alanine scan	A4_alascan	GTCGTCAACCAAGCGGGCAAC
	A5_alascan	GTCGTCAACCAAGCGGGCAAC

synthetic fragments including mutations	sbh = DmbS	GCTCGAGCGCTGCCATTCTGAATAATACAGGCCAGTCAAACTACCACTG CGCAAATCTCTACATGGACAGCCTGGTCACCAATCGGCGCTCGAGAGGA CTCGCAGCTTCCATTATCCATATCGGTCATGTCTGCGACACGGGATACGT TGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAACCTAGGTACCATG CGAGCCATGAGTGTCTCTGAGACGGATGTGCATCATGCCTTTGCTGAGG CGGTCCGCGGGGGGCAGCCAGACAGCCGGAGCGGCTCCCACAACATCAT AATGGGTATTGA
synthetic fragments including mutations	sbh = MilS	GCTCGAGCGCTGCCATTGCGAATAATACAGGCCAGTCAAACTACCACTG CGCAAATCTCTACATGGACAGCCTGGTCACCAATCGGCGCTCGAGAGGA CTCGCAGCTTCCATTATCCATATCGGTCATGTCTGCGACACGGGATACGT TGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAACCGAGGTACCATG CGAGCCATGAGTGTCTCTGAGACGGATGTGCATCATGCCTTTGCTGAGG CGGTCCGCGGGGGGCAGCCAGACAGCCGGAGCGGCTCCCACAACATCAT AATGGGTATTGA
synthetic fragments including alanine-swaps for alanine scan	Ala1	ACGGGATACGTTGCCCGCTTGGTTGACGACGCCAAGGTGCAGATGAGCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala2	ACGGGATACGTTGCCCGCTTGGTTGACGACACCGCGGTGCAGATGAGCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala3	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGCGCAGATGAGCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala4	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGGCGATGAGCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala5	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGGCGAGCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala6	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGGCCC TAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala7	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCG CAGGTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala8	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGCTACCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala9	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTGCCACGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala10	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCGCGCGAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala11	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCACGGCAGTCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala12	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCACGCGAGCCATGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala13	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCACGCGAGTCGCGAGTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala14	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCACGCGAGTCATGGCTGTCTCTGAGACGGATGTGCATCATGC CTTTGCT
	Ala15	ACGGGATACGTTGCCCGCTTGGTTGACGACACCAAGGTGCAGATGAGCC TAGGTACCACGCGAGTCATGAGTGCCTCTGAGACGGATGTGCATCATGC CTTTGCT
primer for control and sequencing	KR_amp_fw	ATGGTCTTGCGTGACAAGCTT
	KR_amp_rev	CATCTGATTCTCCAGGGTGCTAAA
	SQ_KR	GCAAGGTACGGAGCATCTGGACTCG

Table 6-7 Summary of used oligonucleotides for lienhwalide experiments.

Description	Name	Sequence
lwmA	A1	TGCCAACTTTGTACAAGAAAGCTGGGTCGGCTAGTCCCCCTTTTCTGCCT
	A2	GAAGGTTTCATCAGCATCGAGGTC
	B1	GTCTTTGCGGGTGAGGCCTGA
	B2	CCATGTTCGTTGCGCTGGGGT
	C1	CAGCTTTGCGCCTGACCCCA
	C2	AAGCCAAACCGATACGTACCC
	D1	GGACTGGCTTCTTCTTAAGGG
	D2	TGGATAGTGTCATGGTCTCTCAAGG
	E2	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGACACCAATTTCCCTCGAAG
	E1	TGGATAGTGTCATGGTCTCTCAAGG
	LWPKS_FR1_rev	CACGAACCGAGAATGGAGCGTT
	LWPKS_FR2_fw+30	CTAGCCTGAACGCTCCATTCTCGGTTCGTGAGGAAATCACGCGGAACCCTG
	LWPKS_FR3_rev	CCTAAGAACCATCGCTGCTTGGA
	LWPKS_FR4_fw+30	GGAGTTATCCAAGCAGCGATGGTTCTTAGGGATAACTATATCGAGAAGATGACCGTGG
lwmR1	H1	TTCTATGCGTTATGAACATGTTCCCTGGCGCTACCATACATGCAAAACTTCCAT
	H2	TTTCAACACAAGATCCCAAAGTCAAAGGCGATGCCTTTAAGATTACTTTGCCTTC
lwmR3	2MDH1	CAATGTCCATATCATCAATCATGACCGGCGTCATAGCTTCGGACCAGAGCC
	2MDH2	GCTACCCCGCTTGAGCAGACATCACCGGCGATGACTCCTACCCCTAAATCTAGC
lwmR4	CS1	GGCTGGTAGACGTCATATAATCATACGGCGCTACAGCTTTGACGAGACATCCCTG
	CS2	ACTGACCAATTCCGCAGCTCGTCAAAGGCGATGTCGGAAGGGACTCTTCACGTG
	CS_pET28_NdeI_fw	CGGCAGCCATATGATGAGCGAAGGCACCCTGCA
	CS_pET28_NotI_re	GCTCGAGTGCGGCCGCTTACAGTTTGCTGCTAACATCAC
lwmR2	KI1	ACTTTAAGAAGGAGCCCTTCACCAAGGGTGATGAAGATCTCTTTTACGATTCCCGC
	KI2	AATGCCAACTTTGTACAAGAAAGCTGGGTCTTAAGGAGCAAGAGAATTGGGATCG
	KI_exSP_fw	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGGCGCATCGGGGATATGCCAAC
	KI-fw-pET28a	CATATGGCTAGCATGGCGCATCGGGGATATGC
	KI-rev-pET28a	GTGGTGCTCGAGTTAAGGAGCAAGAGAATTGGGA
lwmR6	ACL1	TTCTTTCAACACAAGATCCCAAAGTCAAAGATGGTCTTCGTATCTTTCCCTAAAG
	ACL2	TTTCATTCTATGCGTTATGAACATGTTCCCCTAAAGTCTCGCTGCTGTGGC
lwmR5	R5gpdA_fw	TGACCCACTGGGGTTTTAGGAGGTCAATTGATGCTAGAGTCTTTCTTCGGT
	R5eno_rev	CAGGTTGGCTGGTAGACGTCATATAATCATACGCTAACTATAATTAATGTTACTT
lwtS	Lwt5F	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGATCAGCATGGGGGAGCG
	Lwt3F	AGAGGCGAAGGAGTAGCTGC
	Lwt3R	AGAGCTTCTTCAAGAACACC

213

	Lwt18F	CCAGCTGATAAGATGAATATCCCA
	Lwt15F	CGCTATCCAGTATGCCTGTGCGCG
	Lwt15R	CCGCTACAGGTTTGACACCAGACTCG
	Lwt19F	TGAACACATGCGGAATCCCGT
	Lwt14F	CTCCGCTGCTCCTGCCACCCTACCAG
	Lwt2F	TGAACTTCCTCCAGCCCAGA
	Lwt1R	TAGCTTCTCTGACTCAATCT
	Lwt13R	GGGCATATGGGTGCGGTCTGCGCAAC
	Lwt20F	GACGAAGTCTTCTCGAACAGG
	Lwt5F	AGCTTGGGGGAGCG
	Lwt16F	CGGCTGCGAAACCATCCGAGCCTG
	Lwt16R	TGGGACAATCTCCGTGGGCGCAAC
	Lwt4F	ACGGTATCAAGTTGATTTTC
	Lwt4R	TCGCGCCCCTCAGGACATCC
	Lwt14R	CGAGCCACTTGGTCGTGCCTCCAG
	Lwt5R	AACTTTGTACAAGAAAGCTGGGTCGGCGCGTCATTGGCTTAAGAATCCCA
lwtR1	Lwt30F (pET28a-LwtR1)	ACTGGTGGACAGCAAATGGGTCGCGGATCCATGCCTGAAATGATCAATGAAAACC
	Lwt30R (pET28a-LwtR1)	ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTCATTCATATGCGACAGCACCGTTC
	Lwt8F	TCTTTCAACACAAGATCCCAAAGTCAAAGGATGCCTGAAATGATCAATGA
	Lwt8R	CTATGCGTTATGAACATGTTCCCTGGCGCGTCATTCATATGCGACAGCAC
lwtR2	Lwt26F	GTCGACTGACCAATTCCGCAGCTCGTCAAAATGAAGTTAGCCGCGTTGTG
	Lwt26R	GGTTGGCTGGTAGACGTCATATAATCATACTCACAGCTCAGAAGCTGCGA
	R2_fw_Pgpda	TAACAGCTACCCCGCTTGAGCAGACATCAC ATGAAGTTAGCCGCGTTGTGC
lwtR4	LwtR4_Padh_fw	TTTCTTTCAACACAAGATCCCAAAGTCAAAATGGCTACTAACACTGCTGCT
	LwtR4_Padh_fw	TTCATTCTATGCGTTATGAACATGTTCCCTTTATTCCAAAGTAACAACCAACT
	Lwt27F	GTCGACTGACCAATTCCGCAGCTCGTCAAAATGGCTACTAACACTGCTGC
	Lwt27R	GGTTGGCTGGTAGACGTCATATAATCATACTTATTCCAAAGTAACAACCA
	Lwt33F (pET28a-LwtR4)	ACTGGTGGACAGCAAATGGGTCGCGGATCCATGGCTACTAACACTGCTGCTTGGA
	Lwt33R (pET28a-LwtR4)	ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTATTCCAAAGTAACAACCAACTTA
lwtR6	Lwt37F	TTTCTTTCAACACAAGATCCCAAAGTCAAAATGTCGCAGCTGAATGAATT
	Lwt37R	TTCATTCTATGCGTTATGAACATGTTCCCTTCAGAGAGGCTTCCTCACTG
lwtR7	Lwt11F	CAGCTACCCCGCTTGAGCAGACATCACCGGATGGGTAGCATCGAGGATTC
	Lwt11R	ATGTCCATATCATCAATCATGACCGGCGCGTTACTTCAGGGCCTTGCCGG
	Lwt31F (pET28a-LwtR7)	ACTGGTGGACAGCAAATGGGTCGCGGATCCATGGGTAGCATCGAGGATTCTCGAG

	Lwt31R (pET28a-LwtR7)	ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTACTTCAGGGCCTTGCCGGCCTCG
dbaH	DbaH_Peno_fw	CGACTGACCAATTCCGCAGCTCGTCAAAGGATGCCAGGTACCGTGCGCC
	DbaH_Teno_rev	CAGGTTGGCTGGTAGACGTCATATAATCATCTATTTTTCAACCGACCCCTCTGAC
cytochrome P450	P450HL_fw	GCCAACTTTGTACAAAAAAGCAGGCTCCGCATGGCGGGCTTCCTACACC
	P450HL_rev	TGCCAACTTTGTACAAGAAAGCTGGGTCGGTCAGAATTTCCGTTGCTTGACAAG
lwtR5	lwtR5_Pgpda	TAACAGCTACCCCGCTTGAGCAGACATCACATGGCTCCAGACGCCGATA
	lwtR5_Teno	AGGTTGGCTGGTAGACGTCATATAATCATATTACGCCTTGACGTTGAGCG

6.1.3 Microbiological Methods

6.1.3.1 Growth and Maintenance

E.coli

E. coli cells were grown on LB-agar or in liquid LB-medium with corresponding antibiotics. The cells were cultivated at 37 °C for approx. 12 h. If grown in liquid media, the culture was shaking at 200 rpm. For long-term storage, the liquid culture was mixed with 50 % glycerol to a final concentration of 25 % glycerol and stored at -80 °C.

S. cerevisiae

S. cerevisiae cells were grown on solid YPAD agar at 30 °C for 3 – 5 days. One single colony was used for inoculation for 10 ml liquid YPAD medium. The culture was grown overnight at 30 °C and 200 rpm.

After transformation of cells with vectors containing *ura3* selection marker the cultivation took place with selective SM-URA-Agar and at 30°C for 3-5 days.

Fungal strains

A. oryzae strain NSAR1 was grown on DPY-agar plates for 3-7 days at 28 °C. For long-term storage, the liquid culture was stored at - 80 °C in 25 % glycerol-stocks.

H. lienhwacheense was grown on DPY-agar plates for 7-14 days at 28 °C. For liquid cultures, it was cultivated for 13 days at 25 °C and 150 rpm with 200 ml YMG liquid media in 500 ml shake flasks.

A. nidulans strains were cultivated for 1-3 days at 30 °C on London-Agar. Liquid cultures of 125 ml London-media were grown for 16 h after inoculation at 37 ° C and 150 rpm. Induction took place with 1µg/l Zeocin.

6.1.3.2 Transformations

Heat-shock Transformation of *E.coli* Strains

Competent *E.coli* strains were thawed on ice after - 80 °C storage. 60 – 100 ng purified plasmid was added to 50 µl *E.coli* cells and placed on ice for 30 min, followed by a heat shock at 42 °C for 30 s and cooling on ice for 2 min. Finally, 250 µl SOC-medium was added to the cells and the mixture was incubated at 37°C and 300 rpm for 1 h. The transformed cells were spread out on LB-agar plates containing appropriate antibiotics and incubated at 37 °C overnight.

Transformation *of S. cerevisiae* and Yeast Recombination

The transformation was done using the LiOAc/SS carrier DNA/PEG protocol developed by Gietz and Woods.[206,207] Therefore, a single yeast colony was picked to inoculate 10 ml YPAD medium and grown overnight at 30 °C while shaking at 200 rpm. 40 ml fresh YPAD medium was added to the pre-culture the next day and incubated under the same conditions for 4-5 h. After harvesting the cells by centrifugation (all following centrifugations: Pico™ 17 or Fresco™ 21 Microcentrifuge, Thermo Scientific) at 3000 x g for 5 min at 4 °C they were washed with 25 ml water. The cell pellet was then re-suspended with 1 ml water and transferred to a 1.5 ml tube. The mixture was centrifuged again for 30 s at 11000 x g. The supernatant was discarded. A 1 ml suspension was obtained from the pellet with water, which was divided into 100 µl aliquots. For the preparation of the transformation mixture the following ingredients (Table 6-8) were mixed together on ice.

Table 6-8 Transformation mixture for yeast recombination.

Volume	Ingredient	Comment
50 µL	ssDNA (single-Strand Carrier DNA or salmon sperm DNA, 2 mg/ml)	prepared by boiling at 95 °C for 5 min
240 µl	PEG 3350	50 % w/v
36 µl	LiAc	1 M
34 µl	DNA in water (cut plasmid + fragments, equimolar)	up to 5 µg

The cells were incubated with the transformation mix for 42 °C for 50 min. Cells were pelleted at 11000 x g for 30 s and re-suspended with 500 µl water. 250 µl of the cell mixture was spread on selective SM-URA plates and incubated for 3-5 days at 30 °C.

PEG-mediated Transformation *of A. oryzae* NSAR1

A. oryzae NSAR1 was grown on DPY-plates at 28 °C for 5-7 days. The mycelium was used to inoculate 50 ml GN-medium in 250 ml shake flask. The culture was incubated at 28 °C and 110 rpm for approx. 18h.

The biomass was separated from the media by filtration with a miracloth filter. Mycelia was incubated with 10 ml Trichoderma lysing enzyme (Sigma Aldrich) solution or VinoTaste® Pro (Novozymes) solution (10 mg/ml enzyme, previously sterilised by disposable sterile filter with 0.45 µm pore size [Roth]) while shaking at room temperature and 2 rpm for 3-5 hours.

After incubation, the biomass was gently pipetted up and down to release the protoplasts from hyphal strands. The protoplasts were obtained by filtration with a miracloth filter and centrifuged at 3000 x g for 5 min. The resulting pellet was re-suspended with transformation solution 1 (100 µl per transformation). 500 – 4000 ng of prepared DNA (depending on the number of vectors used) was added to 100 µl protoplast suspension and incubated on ice for 2 min. Then 1 ml of transformation solution 2 was added to the mixture and incubated at room temperature for 20 min. After incubation, 5 ml of appropriate selective softagar was added to the mixture and overlaid over prepared plates with corresponding agar. Plates were incubated at 28 °C for 4-6 days.

When mycelia was visible on the plate, the transformants undergo two further selection steps to avoid false-positive transformants. The colonies are picked from the agar, placed on fresh selection agar plates, and grown for 3-5 days.

For the preparation of liquid cultures, the transformants were grown on DPY agar plates for 5 days. The spores were used to inoculate 100 ml of DPY or CMP liquid medium.

6.1.4 Molecular Biology Methods

6.1.4.1 DNA and RNA Extraction

The different methods and kits used to purify and extract DNA/RNA are summarised in Table 6-9. The manufacturer's instructions and buffers were used.

Table 6-9 List of kits used for DNA/RNA extraction.

Source	DNA/RNA	Kit	Comment
E. coli	vector DNA	Nucleospin$^®$ Plasmid kit (*Machery-Nagel*)	from overnight culture
S. cerevisiae	vector DNA	Zymoprep Yeast Plasmid Miniprep II kit (*Zymo Research*)	after growing on SM-URA plates
A. oryzae NSAR1 transformants	genomic DNA	GeneElute^{TM} Plant Genomic DNA Miniprep Kit (*SIGMA Life Science*)	from mycelia from appropriate plates
H. lienhwacheense, A. nidulans	mRNA	RNA Clean and Concentrator^{TM}-5 (*Zymo Research*)	from liquid cultures under producing conditions
	mRNA to cDNA	High Capacity RNA-to-cDNA Kit (*Thermo Fisher Scientific*)	
PCR	PCR fragments	Nucleospin$^®$ Gel an PCR Clean-up kit (*Machery-Nagel*)	
agarose gel	linearized vectors		

The concentration of DNA and RNA samples was determined with a DeNovix$^®$ DS-11+ Spectrophotometer.

6.1.4.2 Cloning Procedure

Polymerase Chain Reaction (PCR)

PCR was used to amplify DNA Fragments from genomic DNA, cDNA or vectors. The proofreading Q5$^®$ 2x Master Mix (New England Biolabs) was used to obtain DNA fragments needed for further cloning procedure. The information provided by the manufacturer served as a template.

The OneTaq$^®$ 2X Master Mix (New England Biolabs) was used to control if the GOI is present in a vector or gDNA. The information provided by the manufacturer served as a template.

For colony PCR one single colony was picked and prepared with OneTaq$^®$ 2X Master Mix (New England Biolabs). The initial heating temperature time was increased from 30 s to 7 min.

Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to visualize DNA or RNA. Therefore, the gel was prepared with 0.5-2 % agarose in TAE-Buffer. 1 µl Roti$^®$-Safe GelStain (Roth) was added to 25 ml agarose.

The volume of 5 µl of DNA was mixed with 6 x loading buffer. When using OneTaq$^®$ 2X Master Mix no additional loading buffer was necessary. As a marker 2 µl of the 1 kb DNA Ladder (New England Biolabs) was used. The gel was run at 110 V and 400 mA for 25 min in a Bio-Rad gel chamber containing 0.5 % TAE-buffer.

In case of large DNA fragments (> 5000 bp), for example after restriction digestion, 0.5 % Agarose gel was used and the gel was run at 90 V for 40 min.

The DNA was visualized with the Molecular Imager Gel doc XR+ (Bio-Rad) system under UV-light (312 nm).

Restriction Enzyme Digestion

All enzymes (Table 6-10) used in this work were purchased from New England Biolabs (Beverly, MA, USA) and used according to the manufacturer's instructions with appropriate buffers.

Table 6-10 List of enzymes.

Enzyme	Vector/Fragment	Digestion site
EcoRI	pEYA-*tenS*	G'AATTC
FseI	pEYA-*tenS*	GGCCGG'CC
BstEII	pEYA-*tenS*	G'GTNACC
AscI	pTY GS *argB/ade/sC*	GG'CGCGCC
NotI-HF	pTY GS *argB/ade/sC* pET-28a + lwmR4	GC'GGCCGC
NdeI	pET-28a + lwmR4	CA'TATG
NheI-HF	pET-28a + lwmR2	G'CTAGC
XhoI	pET-28a + lwmR2	C'TCGAG

Gateway Cloning

Gateway^{TM} LR Clonase^{TM} II Enzyme mix kit (Invitrogen)[132] was used to transfer genes from the entry vector to the destination vector. The manufacturer's instructions were followed. For *E.coli* Top10 transformation, the vector mixture (10 µl) was added to 50 µl competent cells.

DNA Sequencing

DNA samples were sequenced by *Eurofins Genomics* (Mix2Seq OVERNIGHT, Ebersberg).

6.1.4.3 Protein Purification Procedure

Vector Construction

The T4 DNA Ligase (New England Biolabs) was used for ligation of DNA fragments into the destination vector. The manufacturer's instructions were followed with appropriate buffers.

Recombinant Protein Production

E. coli BL21 (DE3) were used as expression host for the expression of recombinant proteins. A 10 ml LB-media seed-culture with the appropriate antibiotic (kanamycin) was inoculated with transformed cells. The culture was incubated over night at 37 °C and 200 rpm for 12-16 h. The main culture of 1 l LB media in a 2 l shake flask with antibiotic was inoculated using the seed culture and incubated at 37 °C and 220 rpm. At an OD_{600} of 0.25 the temperature was decreased to 16 °C. At OD_{600} of 0.6 the culture was inoculated with 100-500 µM IPTG (isopropyl β-*D*-1-thiogalactopyranoside) and incubated for another 16 h at 170 rpm and 16 °C. The cells were harvested by centrifugation at 6000 x g for 15 min (Thermo Scientific™ Sorvall LYNX 6000 Superspeed Centrifuge).

Cell Lysis and Protein Purification with Immobilized Metal Affinity Chromatography

The obtained cell pellet was homogenized in 20 ml of appropriate lysis buffer (phosphate buffer or Tris-HCl buffer). The cell disruption was achieved by sonication (SONOPULS KE76, Bandelin) for 7 min with 10 s pulse intervals and 34 % amplitude on ice. Cell debris was separated from the lysate by centrifugation for 40 min at 20000 x g at 4 °C (Thermo Scientific™ Sorvall LYNX 6000 Superspeed Centrifuge). The supernatant was obtained by carefully decanting.

All vector constructs encoded an *N*- or *C*- terminal polyhistidine tag (His_6). Ni^{2+} nitrilotriacetic (NTA)-affinity chromatography resin (PureCube) was used for the gravity-flow chromatography of the recombinant expressed proteins. 2 ml resin was prepared according to manufacturer's instructions, added to the lysate and incubated at 4 °C for 1 h under light rotation.

The mixture was carefully pipetted into CHROMABOND® empty column with PE filter element (15 mL; Macherey-Nagel) and the flow through was stored on ice for analysis with SDS-PAGE. The resin was washed with buffers including an increasing concentration of imidazole (from 20 mM up to 500 mM). Each fraction was collected and stored on ice for the analysis by SDS PAGE.

The determination of fractions including the protein was enabled by SDS-PAGE analysis. The appropriate fractions were collected and concentrated using Amicon® Ultra centricons (Merck Millipore) with an appropriate cut-off for each protein. The

obtained protein was stored at 20 mg/ml at most at -20/-80 °C or used immediately for assay reactions.

The concentration of protein solutions was determined with a DeNovix® DS-11+ Spectrophotometer. The ProtParam tool (https://web.expasy.org/protparam/) was used for determination of the protein molecular weight and extinction coefficients.

SDS-Polyacrylamide Gel Electrophoresis

Samples were boiled at 95 °C for 5 min after preparing them by adding 10 µl 4x Lämmli buffer to 30 µl of protein solution.

SDS polyacrylamide gel electrophoreses were performed with 12 % polyacrylamide gels (Table 6-11) in combination with the gel casting system and electrophoresis (Bio-Rad).

Table 6-11 Composition of 12 % SDS-polyacrylamide gel electrophoresis.

Composition	Volume [ml]	
	separating gel	stacking gel
30 % acrylamide/bisacrylamide (Rotiphorese® Gel 30 [37,5:1])	3	0.54
ddH$_2$O	2.45	1.7
1.5 M Tris-HCl, pH 8.8	1.9	-
0.5 M Tris-HCl, pH 6.8	-	0.25
10 % (w/v) SDS	0.075	0.002
10 % (w/v) APS	0.075	0.002
TEMED	0.0003	0.0002

The Colour Prestained Protein Standard (4µl; Broad Range, 11 - 245 kDa; New England Biolabs) was used as a reference marker and 15 µl of the prepared sampled were carefully loaded on the gel. The gel was run at 75 mA for 45 – 60 min.

Coomassie staining solution (approx. 20 ml) was used to stain the bands of the gel for approx. 1 h. Afterwards, Coomassie bleach (20 ml) was used to de-stain the gel and therefore visualize the bands. The gel and incubated for 20 – 30 min repeatedly, until the protein bands were clearly visible. The gels were scanned with the Molecular Imager Gel Doc XR+ system (Bio-Rad).

6.1.4.4 *In vitro* Assay Composition

Citrate Synthase Assays (Maleic Acid Anhydride Pathway)

The reaction mix (Table 6-12) was incubated at 37 °C for 2 h.

Table 6-12 Composition of reaction mix for citrate synthase assays.

Concentration	Composition
1 mM	hexanoyl-CoA or hexenoyl-CoA
1 mM	oxaloacetic acid or α-ketoglutaric acid
4 µM	protein
ad 100 µl	phosphate assay buffer

The volume of 100 µl acetonitrile was added and mixed to stop the reaction. The protein was precipitated by centrifugation for 10 min at 21.1 x g (maximum speed). The supernatant was used for LCMS analysis.

Tropolone Pathway Assays

The reaction mix (Table 6-13) was incubated at 37 °C for 2-4 h.

Table 6-13 Composition of reaction mix for Lwt protein assays.

Final Concentration	Composition
2.5 mM	DHMBA **156**
5 mM	α-ketoglutaric acid
8 mM	ascorbate
0.1 mM	$FeSO_4$
1 mM	NADPH
5 mM	glucose-6-phosphate
1 U/ml	glucose-6-phosphate dehydrogenase
5 mM	TES buffer (1 M, pH 7.5)
10 µM	LwtR1
10 µM	LwtR4 (optional)
10 µM	LwtR7 (optional)
ad 50 µl	H_2O

The double volume of acetonitrile was added and mixed to stop the reaction. The protein was precipitated by centrifugation for 10 min at 21.1 x g (maximum speed). The supernatant was used for LCMS analysis.

6.2 Chemical Methods

6.2.1 Chemical Synthesis of *E*-Hex-2-enoyl CoA [160,170]

For the synthesis of the *E*-hex-2-enoyl CoA 6 mg of *E*-hex-2-enoic acid (0.052 mmol) was dissolved in 1.5 ml anhydrous CH_2Cl. After cooling to 0 °C, 5.5 µl (0,052 mmol) ethylchloroformate was added and incubated for 2 h. The CH_2Cl was removed and 1.5 ml DMF was added. Coenzyme A trilithium salt (21.4 mg, 0.026 mmol) dissolved in 1.5 ml aqueous $NaHCO_3$ was added to the reaction mixture at room temperature. After 10 minutes 100 µl formic acid was added for acidification. The reaction mixture was then added to 50 ml water and frozen for the following lyophilisation (Alpha 1-4 LDplus, Martin Christ). The product (71.9 mg) was purified by preparative LCMS.

6.2.3 Analysis and Isolation of Compounds

6.2.3.1 Extraction of Fungal Liquid Cultures

After cultivation of liquid cultures, the biomass of fungal cultures was homogenized with a blender and flirted by vacuum filtration in most cases (*H. lienhwacheense*, *A. oryzae* NSAR1 transformants of lienhwalide project). Alternatively, the media was filtered directly without homogenisation (*A. nidulans*, *A. oryzae* NSAR1 transformants of tenellin project).

If not stated differently, the supernatant was acidified with 2 M HCl to pH 2-4. The media was extracted twice with an equal amount of ethyl acetate. The organic layers were dried with anhydrous magnesium sulfate ($MgSO_4$). The organic phase was removed with a rotary evaporator. The extract was then dissolved in an appropriate amount of solvent (methanol, acetonitrile or dichloromethane) filtered with glass wool and analysed by LCMS.

6.2.3.2 Analytical Liquid Chromatography Mass Spectrometry (LCMS)

Analytical LCMS was run to analyse the extracts from fugal cultures. The Waters LCMS system containing a Waters 2767 autosampler, Waters 2545 pump, a

Phenomenex Kinetex column (2.6 μm, C18, 100 Å, 4.6 x 100 mm), a Phenomenex Security Guard precolumn (Luna, C5, 300 Å) was used with a flow rate of 1 ml/min. The equipped detectors were a diode array detector (Waters 2998) in the range 210 to 600 nm and an ELSD detector (Waters 2424) together with a mass spectrometer, Waters SQD-2 mass detector (ES$^+$ and ES$^-$, 150 to 1000 *m/z*). For elution, a solvent gradient was run for 15 min starting at 10 % acetonitrile/90 % HPLC grade water (0.05 % formic acid) and ramping to 90 % acetonitrile water (0.045 % formic acid).

6.2.3.3 Preparative LCMS

Isolation of compounds was achieved using a Waters 2767 autosampler, a Waters 2454 pump system using a flowrate of 20 ml/min and a Waters mass-directed autopurification system (equipped with a Phenomenex Kinetex Axia column [5μ; C18; 100 Å; 21.2 x 250 mm] and a Phenomenex Security Guard column [Luna C5; 300 Å]). The flow was split after the column (100:1). The major part of the flow was collected with the fraction system and the other minor part was analysed by the equipped detectors. The minority flow was compensated with a mixture of acetonitrile and water (1:1) at 0.8 ml/min. Then, the minority flow was analysed using the same detectors as described previously. A solvent gradient was starting at 10 % acetonitrile/90 % HPLC grade water (0.05 % formic acid) and ramping to 90 % acetonitrile water (0.045 % formic acid).

The compounds were collected according to the mass analysis in glass tubes. After combining identical fractions from repeating LCMS runs, residual acetonitrile was removed *in vacuo*. The left over aqueous solution was frozen and lyophilised (Alpha 1-4 LDplus, Martin Christ).

6.2.3.4 Ultra Performance Liquid Chromatography-High Resolution Mass Spectrometry (UPLC-HRMS)

The concentration of the compound samples were adjusted to < 1 mg/ml before measurement. High Resolution Mass Spectrometry data was acquired using a Waters QTof premier spectrometer and Acquity UPLC-domain (Waters, incl. TUV detector). The electron spray ionisation took place in positive or negative mode, depending on the measured compound, respectively.

6.2.3.5 Nuclear Magnetic Resonance Spectroscopy (NMR)

Compound samples were measured on one of the following spectrometers: Bruker Ascend 600 MHz, Bruker Ultrashield 500 MHz, Bruker Ascend 400 MHz, Bruker Ultrashield 400 MHz. Data was acquired at 400/500/600 MHz.

Chemical shifts are defined by parts per million (ppm) relative to tetramethysilane standard. Raw data was referenced by the deuterated solvent used. 2D experiment data were obtained for complete structural elucidation including Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Coherence (HSQC), and Heteronuclear Multiple Bond Correlation (HMBC). The data was analysed using the MestReNova 14.2.3 software (MestreLab Research).

6.3 Bioinformatic Methods

6.3.1 AlphaFold and ChimeraX

For AlphaFold prediction, the software UCSF ChimeraX v. 1.4[118,119] was used together with the AlphaFold[31,208] tool v. 4 on default settings.

6.3.2 PyVOL

The PyVOL[133] plugin for PyMOL was used to calculate the pocket size of the substrate and cofactor binding site of TenS models. The minimum radius for pocket calculation was set to 1.8 Å and the maximum radius to 4.8 Å. For other parameters, the default settings were used for the calculation. The adjustment of the minimum/maximum pocket radius were based on the docking of the cofactor.

6.3.3 cblaster and clinker

Default settings were used for the use of cblaster except for the gap size, which was adjusted to 50 000 bp to allow a relatively large window size to analyse potential BGCs manually. The cblaster search was carried out with following sequences (Table 6-14):

Table 6-14 cblaster reference seqeunces.

Enzyme	Amino Acid Sequence
LwtS	MISMGERLRSPVLGPTVFLFGSLALSFDESTFVQIRKTVVENEEFSWVLETVAEFPQIWKSLTKSIPSLQTPTALSQL EDLVNALRTGRPLDTPSPLPNKLLIPLVIISHLTQYATFLRNSAEEQQVDAFSTSELDRETLGLCTGLLSAFAVSSAS NSLQFKKYAAVAVRLAMLVGTVVDAQDESTEPSKSLSAAWNSTEGAEQLQRIVKSFPETYVSVNYDENRATITTKALA ISKLQQQLKNAGLVASEVSLYGRFHTVVNAKILGPLFEYVDARPEFQYMDASSGLVMPTRSNKGGEMITQGPLHHHAL QSILAEQSQWFQTVSAMRDSILKSKDSRILSFGPERSVPPSLLRSLGSQVVYVADLDSPSSRFATFSQKDLHTFSDDD VAVVGMSIKVAGAETLDDFWDLLVAAKSQHKEVPKERFTFETAFRDVDPKRKWFGNFIESYDKFDNKFFKKSPRESST MDPQQRNLYQCAYQAVEQSGYFHLAEKDSRIGCFVGVCSRDYDCNVACHTPNAFTTTGNLQSFIAGKISHYFGWTGPG LVIDTACSSSAVAIHQACRAIITGECTAALAGGTNIFTNPLWFQNLAGASFLSTTGQCKPFDANADGYCRGEGVAAVF LKKLSAAVADGDQILGVVAATGVQQNENCTPIFVPNGPSLADLFRGVTKQARLKPSQISVVEAHGTGTGVGDPAEYDG VRQVLGGPASGREKSLFLSSVKGLVGHTEASSGIVSMIKVMLMIQKGMIPPQASHTSINPAIGATPADKMNIPTNLQA WDADFRAALINNYGASGSNASLIITQSPHREIKDSPVAISGIKYPFWLSGLDGQSLRRYAKALRKFLNNRSYSAADLS LPNISFNMARQSNRSLDKAVLFNVRSVEEFDQKMASLEKGDASITPIDRPKSKPVVLCFGGQVSTFVGLDEQLYKKIA VLRKHLNHVDAVTRSLGAGSIFPKIFETTPMHDTVQLQTILFAIQYACARSWIESGVKPVAVVGHSFGELTALAISQV LSLEDAVKLIINRATLIRDNWGSDKGAMIAVEADLADVEKLLTESNSKIPREKSASIACYNGPRSFTLAGPTAAIDSV AATIPSSLRSKRLNVTNSFHCALVDPLVDALDQVGKMLTFRKPVIHVERASEFPIQEKHTAKFVAEHMRNPVYFNHAI QRLAKQHPSAVFLEVGSNSTITSMASRALGNPSSSHFQAMNITNNDKAWDSLTDATMNLWKAGVATQHWAHQAAQTRE HSPLLLPPYQFEQTTHWIDFKVPPKVSELPPAQKIESEKLPESLLTFVGYEDIEKRVAKFRINTMIPKYERYLQGHVV AQTAPICPSTLQLDLVIEAVRGIRSDLAEAKFEPQARVIENQSPICINPTRCVWIEAQEEPADNNSSWNFQVFSTETV HSSAKTMHTTGKIVFYSAEDLTPKLEFARLERLTGHRRCVELLQSGDVDEVFSNRNIYKIFSEVVDYGEDYRGVQKLV CRGNFSAGHVVKKYNPETWLDTHLADSFCQVGGIFVNCMTDRSPADMYISNGIEQWTRSPKLRSGDAWPDSFHVLAAH HPSDKQYLTDVFVFHPETGALLEAILGISYVKIPKASMSKLLSRLTAKGATAAKPSEPVKVTLFEVAPTEIVPPRQAP TKAANKTEAPKPKKVKKQSSKSYVTIKVKEILAELSGLDLEEIKDDSQLADLGIDSLMGMEMAHEIEGAFKVMLPESE LMDIVDMPGLIKCVEKAVSGDAVSAGDTSTQSEDDASSDSETGSYSISEASSGHPTGLTTPAVDYDKDPYDSLSNCDG ELELPFATIMEAFNETKALTDDKIAEYGQTKYFDTVMPLQTDMCIAFTLEAFDQLGQRLREAKPGDKFSRISHPKEHS RLVDYLYRMLEKESHIINVQGDTIIRTAVAAPTRSSKEILQDLLAHFPDQTTADKLTFYTGSKLAEVLKGQTDGIKLI FGCPEGRELVSGMYAEWPLNRLLYNQMEDFLARLCTKLDMSEGPLKILEMGAGTGGTTKWLVPLLASMNVPVEYTFTD LAPSFVAAARKKFKEYPFMKFRTHDIEKVPADDLIETQHIVIASNAVHATHSLRESGKNIRAALRPDGFLIMLEMTGI MYWVDMIFGLFDGWWFFDDGRQHAITNELKWKEELQAVGYGHVDWTDGVRPENKCEKLIIAMASGERCEQSPILSSPP VKSLSADLVARQAVVDKYVRGLTEGFSVKLAETPQSASVRTYPDGYTILVTGATGSLGSHLVAHFAELPNVARVICLN RCSRSDPKERQRQALIKKGIFLSEDASKKLCVFETDMSKGFLGLPIEDYEGLVHDVTHIVHNAWLMNAKWPIRNFESQ FQIMKNLLEFCREISIKRAPGTKVTFQFISSIATVGHRPIWSGKPVVPEERMTIESVLPTGYGDAKYVCELMVDETLH KHPDRFRAMVVRPGQIAGSSISGYWNPMEHLSFLIKSSQTLKALPDFDGLLSWTPVDKVAGALADLVLLSEDTTHYPI YHVDNPIRQTWKKMIPVLAGALDISLQDIVPFRDWLQCVRNHPRQVEGPGGENPAFLLIDFLDDNFIRMSCGGLLLDT VKIREHSKTLANAGPVSEEVTRRFIQSWKDMGFLSQ
LwtR1	MPEMINENPVEIAIVGGGIIGLILTAGLIKQDVKVKVYEQAKSFREIGAGMAFSANARQCMDLMNPDILAALRSGGSV ATSMDRNDPNDYQRYVDGYNRRRKDDPLYQKVLYKIDAGYKGFEGTRRDQFLEALVKIIPKDIVELERRLDTIEETGP DGRLRLNFADGTTAEADAIGCDGIKSRVRELMFGQGNPASYPHYTHKVAYRTLVPMEKAVEALGEYKARNQHTHIGPN AHIIHYPVANQTMVNIAAFISDSDEWLDDKLSVMSGQRKDLETAFAGWNPRLTALLQYFPEKLEKWAVFDTWDYPAPY FNKGNICLAGDAAHASSPHHGAGACCGVEDALCLFTLIGLIDTTPREKSVVKTASRKEAFVTAFEVYDKIRRARAQWL VNSSRRMCDFYHQPEWANEQKWIKAETCFEEMKDRSYKIWRFDYHAMMKETVVEYNRKILIKPTNGAVAYE
LwtR7	MGSIEDSREMAPDAIPTVDISPFTNPNASEDAKNGVVDAIRHACTTYGFFYLAGHGVSEEKRSGIVKCAKKFFELPQE QRMEVWMGKSMGKSFRGYEPPGIQTHQEGLLPDIKESFVVGHEVPADDPDAGTFSTGPNMWPKSLKDEEFRTPIMEYQ ATMLALSKVLLQLLARALPKSWGHPPNVMDEFAVNPSMPMRLLHYAPQNVLDERQFGGDHTDFGGVTILLQEMNTKGL EVWYPPTETWIPVPPKEGTYVINMGDMMQKWTAGYYRSARHRVITSGTNHRYSVPWFLNGQLKLKCKALDGSGVETIV GEHIRQRLINTMPEAGKALK
LwtR4	MATNTAAWIKEAKAHPFVVESAPLWTPESGEILIKNHAVAINPVDGSLQAAGWWPLNYPTMLGQDVAGVVTAVGDNAQ GFKVGDRVLGHAVGMATKRDQDNAFQQYTILKTNMAAHLPANIPFEKAVVLPLAMSTAACALYQDTHFKLQFPTVPAR ESTGQTFIVWGGSSAVGSNAIQLLVNSGYEVITTASPKNFEYVKKLGASQVFDYSSPTVQDDLIKAMKGKTSAGALDC VGGAPQGILQQVISTVEGTKAVASTKRGWPEPPAGVTMYSIFGATLKDNNVGAAIYNDFLPAALEKGTYVPAPEPTIV GNGLEKIQDAVDLIKKGVSATKLVVTLE
LwtR6	MSQLNEFLVQIPDKPNVLATRVSNVQTHLARLKGLVESGTIVMSGPTLAAHPKTADESLAVNGSVWLVRAGSVQEVHA LVADDIYAKLGVWDLDNVTVTPYKCAVRKPL

For clinker default settings were used, if not stated otherwise.

7 Appendix

7.1 List of Figures

7.2 List of Schemes

7.3 List of Tables

7.4 Additional Data

7.4.1 NMR Spectra

7.4.1.1 Compound 143
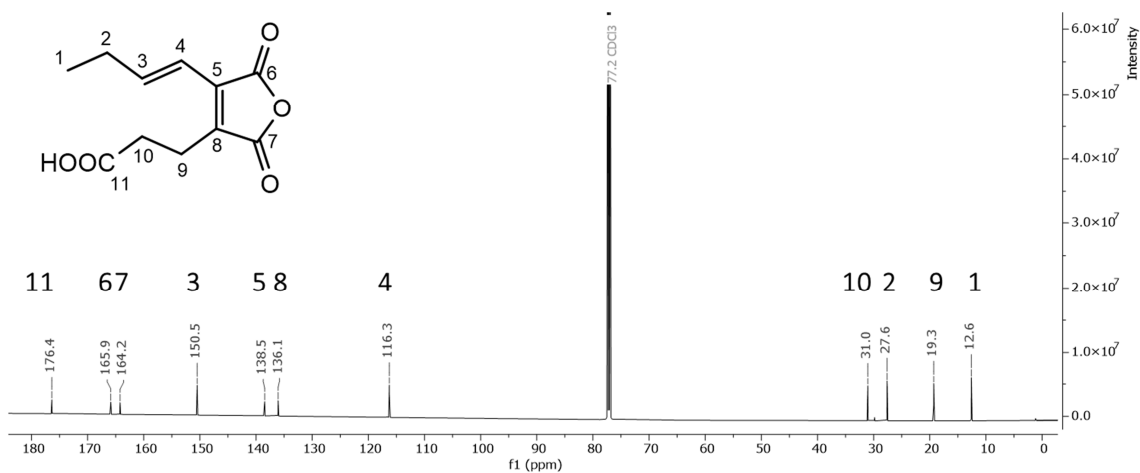
Figure 7.1 ^{13}C NMR (150 Hz) of compound **143** in CDCl$_3$.
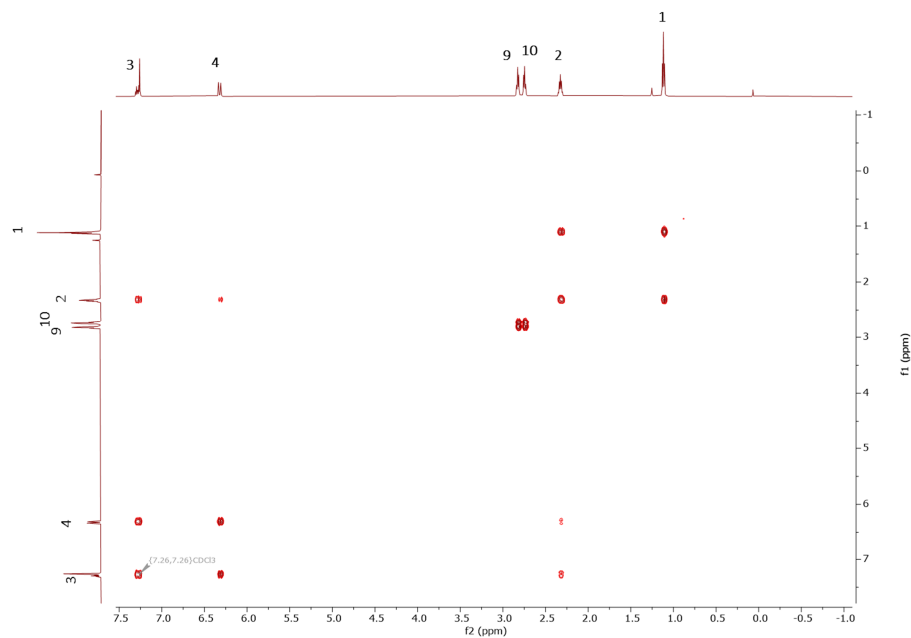
Figure 7.2 COSY (600 Hz) of compound **143** in CDCl$_3$.
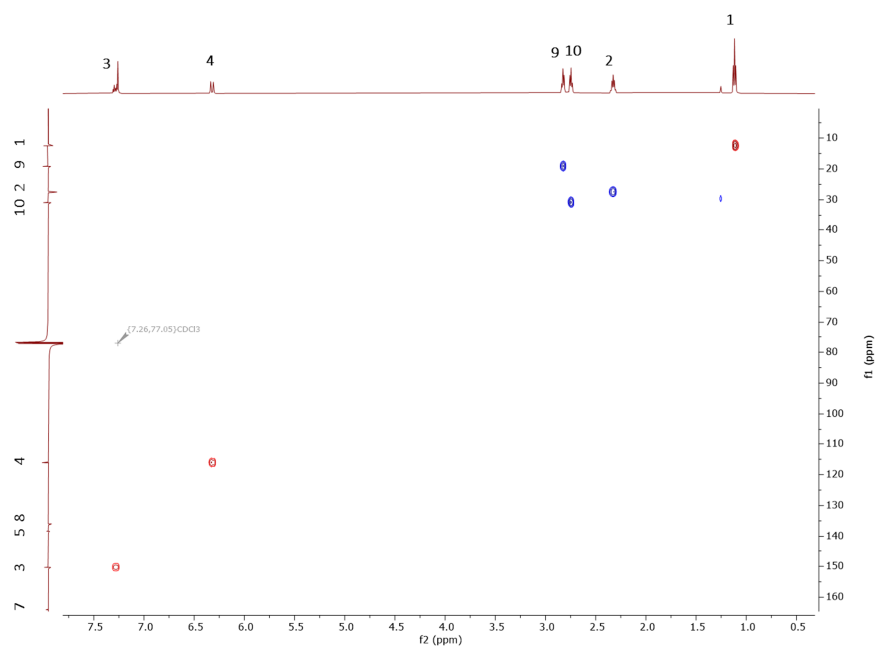
Figure 7.3 HSQC (600 Hz) of compound **143** in CDCl$_3$.
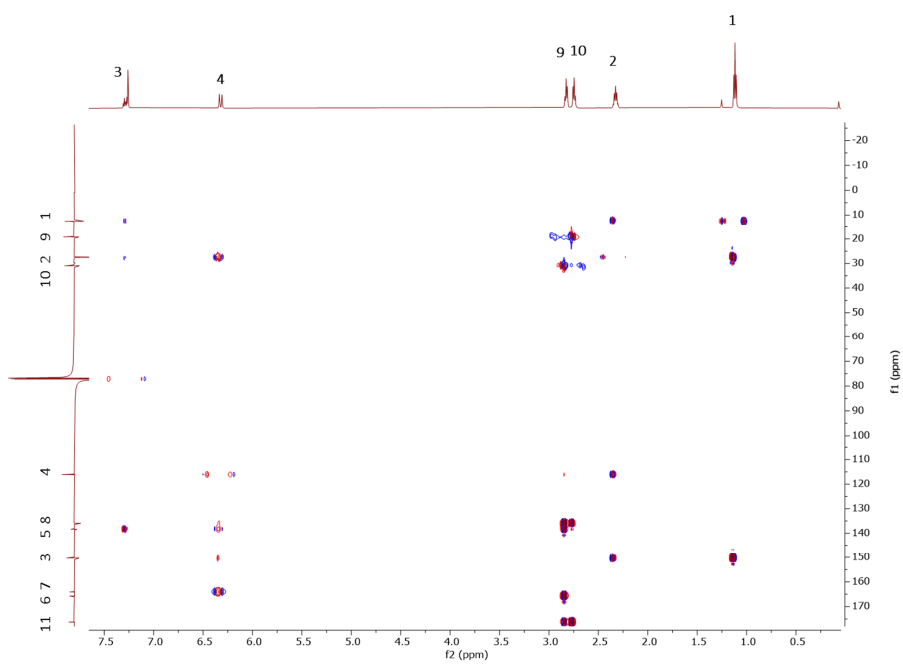
Figure 7.4 HMBC (600 Hz) of compound **143** in CDCl$_3$.

7.4.1.2 DHMBA 156
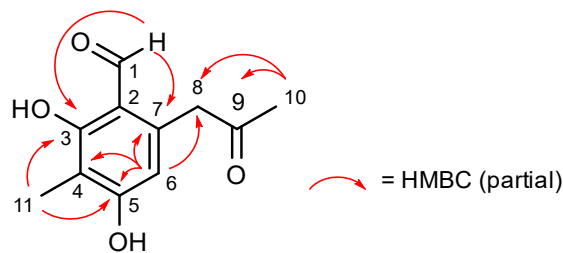
Figure 7.5 Structure of DHMBA **156**.

Table 7-1 Chemical shifts of compound **156** in CDCl$_3$ (600 MHz).

Atom	C-type	δ$_C$/ppm	δ$_H$/ppm (mult, *J* in Hz)	HMBC	literature[209] δ$_C$/ppm	literature[209] δ$_H$/ppm
1	COH	195.0	9.82	3, 7		
2	C	139.3				
3	C	164.8				12.68
4	C	110.8			7.3, 29.5, 46.7, 110.8, 111.6, 113.3, 138.9, 163.5, 164.7, 194.4, 205.6	
5	C	163.5				9.30
6	CH	111.5	6.35	4, 5, 7, 8		6.38
7	C	113.6				
8	CH$_2$	46.7	4.09	2, 6, 7, 9		4.09
9	C	205.3				
10	CH$_3$	29.7*	2.19	8, 9		2.04 or 2.21
11	CH$_3$	7.3	2.01*	3, 5, 6		2.04 or 2.21
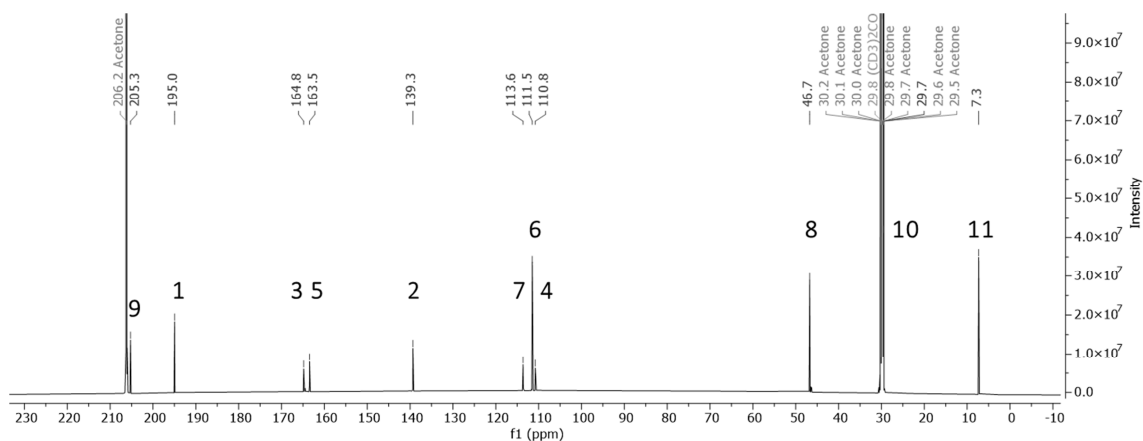
Figure 7.6 ^{13}C NMR (600 Hz) of DHMBA **156** in (CD$_3$)$_2$CO.

238

Figure 7.7 ^{1}H NMR (600 Hz) of DHMBA **156** in $(CD_3)_2CO$.

Figure 7.8 COSY of DHMBA **156** in $(CD_3)_2CO$.

239

Figure 7.9 HSQC of DHMBA **156** in $(CD_3)_2CO$.

Figure 7.10 HMBC of DHMBA **156** in $(CD_3)_2CO$.
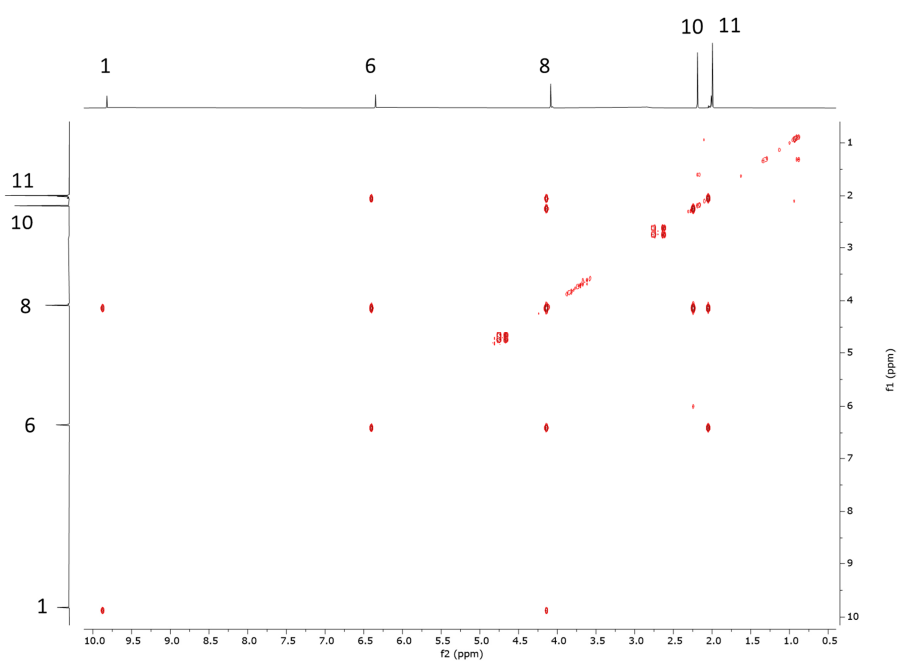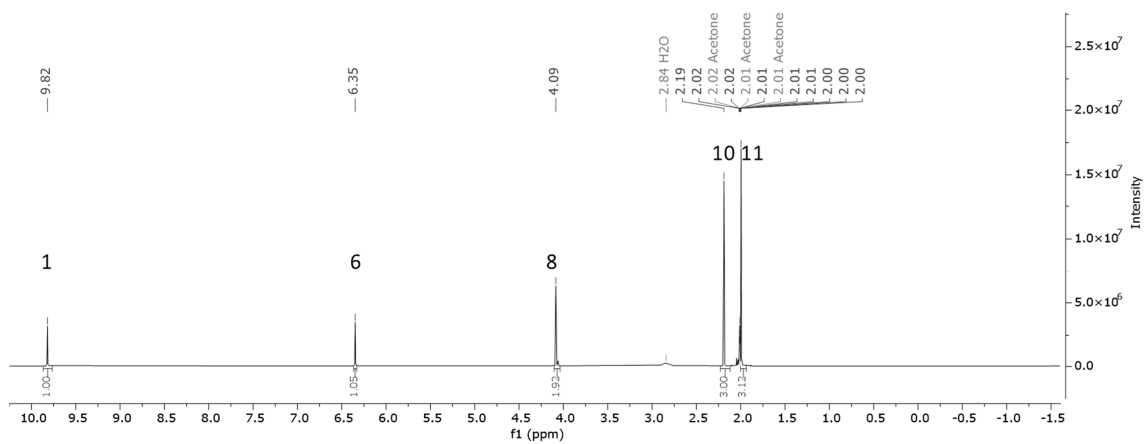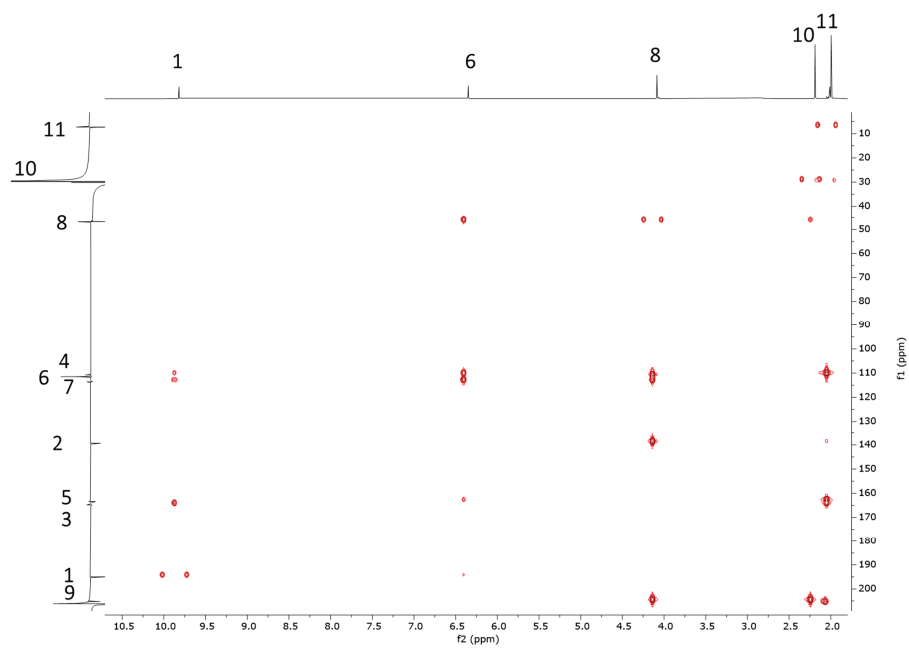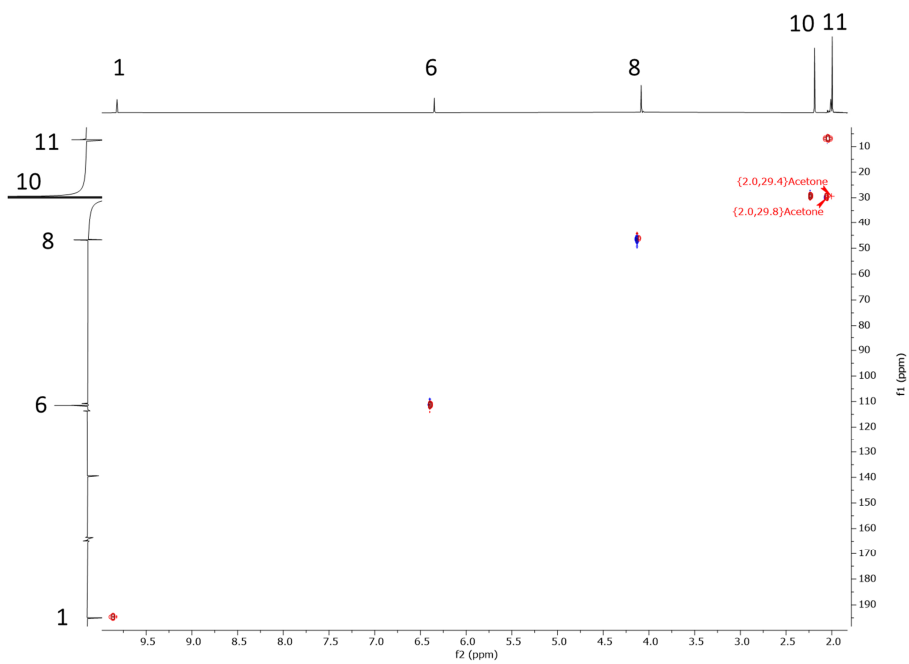
7.4.1.3 Sepedonin 90 and Anhydrosepedonin 160
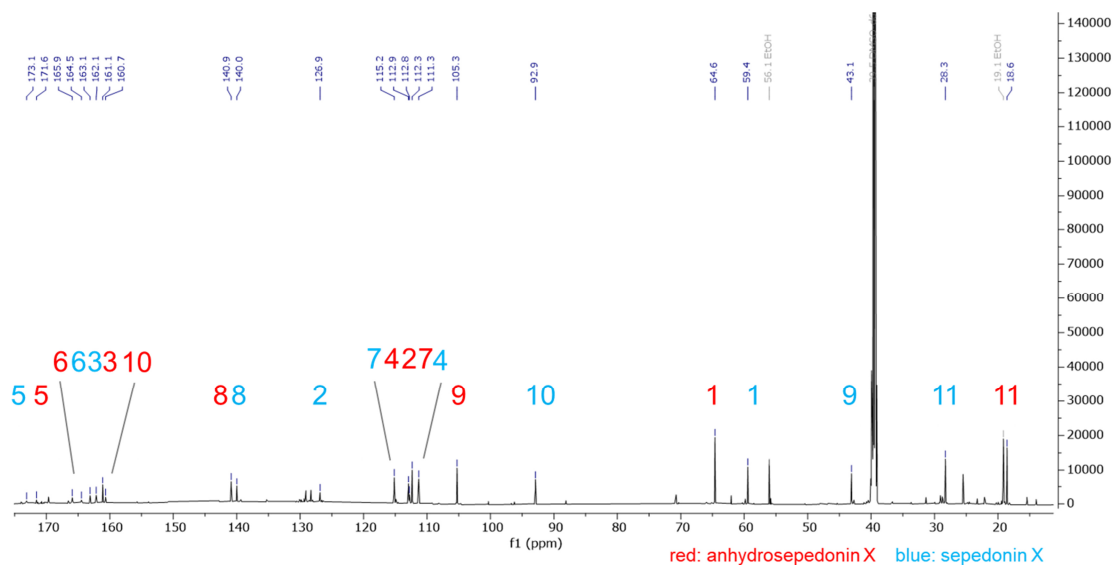
Figure 7.11 ^{13}C-NMR of anhydrosepedonin **160** and sepedonin **90** mixture (600 MHz) in $(CD_3)_2SO$.
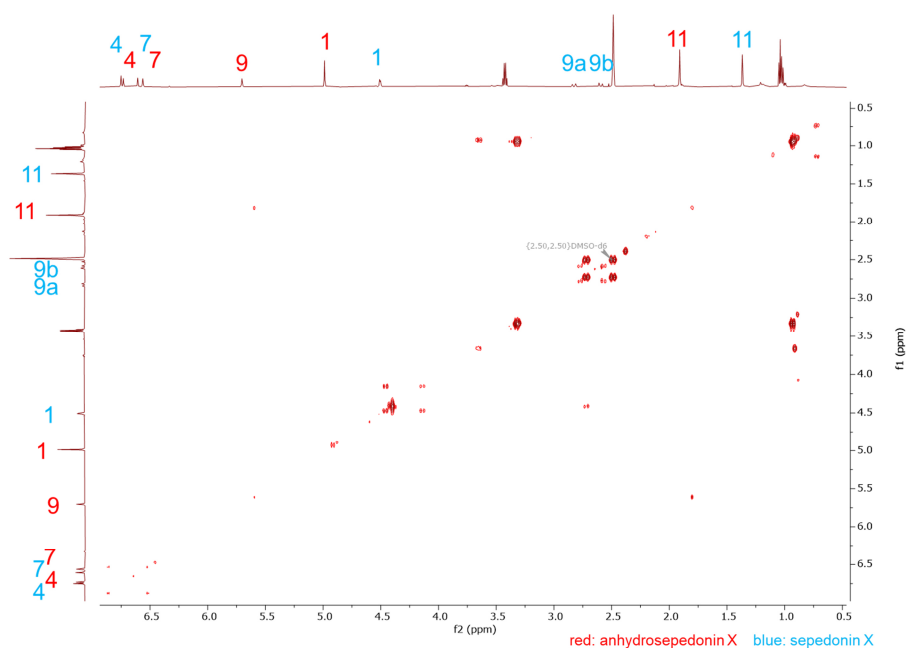
Figure 7.12 COSY of anhydrosepedonin **160** and sepedonin **90** mixture (600 MHz) in $(CD_3)_2SO$.
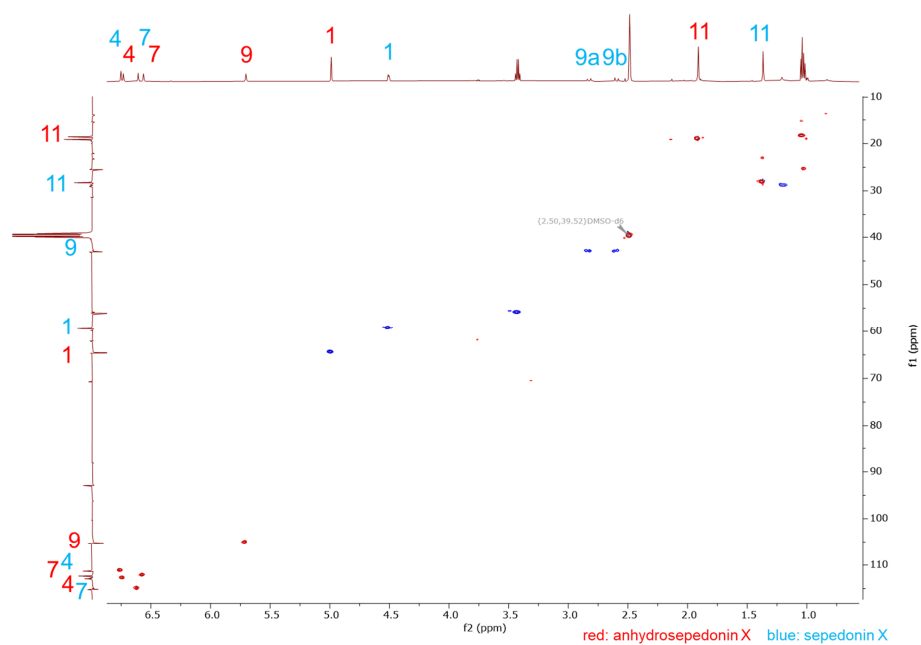
Figure 7.13 HSQC of anhydrosepedonin **160** and sepedonin **90** mixture (600 MHz) in (CD$_3$)$_2$SO.
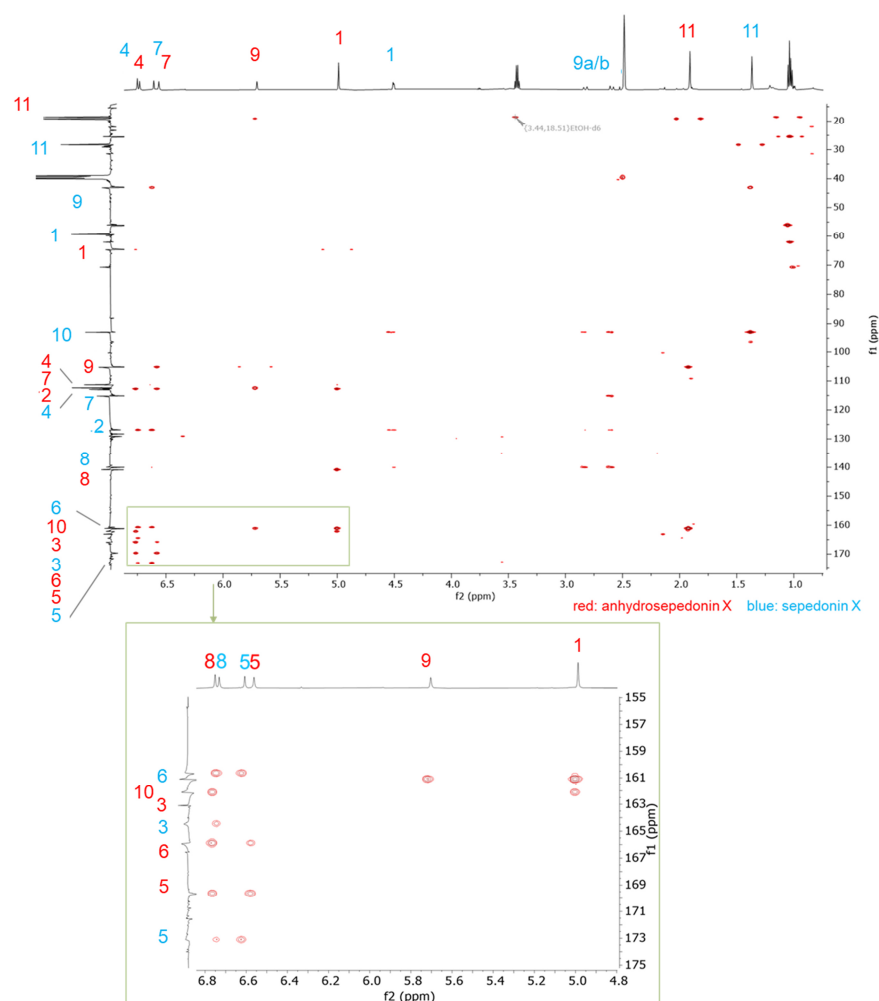
Figure 7.14 HMBC of anhydrosepedonin **160** and sepedonin **90** mixture (600 MHz) in (CD$_3$)$_2$SO.

7.4.2 List of analysed Genomes for Tropolone BGCs

Table 7-2 Overview of species included into the genome-sequencing project.

Species (Strain ID)	abbreviation	related BGC?	Ref.
Annulohypoxylon michelianum (CBS 119993)	Ami	+	this work
Annulohypoxylon truncatum (CBS 140778)	At	+	[149]
Creosphaeria sassafras (CBS 127876)	Cs	+/-	this work
Daldinia concentrica (CBS 113277)	Dc	+	[149]
Daldinia childiae (CBS 122881)	Dch	+	this work
Entonaema liquescens (ATCC 46302)	El	+	[149]
Hypoxylon addis (MUCL 52797)	Had	+	this work
Hypoxylon duranii (ATCC58730)	Hd	+	this work
Hypoxylon dussii (MUCL 53766)	Hdu	+	this work
Hypoxylon erythrostoma (MUCL 53759)	Her	+	this work
Hypoxylon fragiforme (MUCL 51264)	Hf	+	[149]

Species (Strain ID)	abbreviation	related BGC?	Ref.
Hypoxylon fuscum (CBS 113049, STMA 13090)	Hfu	+	this work
Hypoxylon griseobrunneum (CBS 331.73, STMA12142)	Hg	+	this work
Hypoxylon gibriacense (MUCL 52698)	Hgi	+	this work
Hypoxylon guilanense (MUCL 57726)	Hgu	+	this work
Hypoxylon haematostroma (MUCL 53301, STMA 10239)	Hh	+	this work
Hypoxylon hinnuleum (MUCL 3621)	Hhi	+	this work
Hypoxylon hypomiltum (MUCL53306)	Hhm	+	this work
Hypoxylon investiens (MUCL53316)	Hi	+	this work
Hypoxylon invadens (MUCL 51475)	Hin	+	this work
Hypoxylon lienhwacheense (MFLUCC 14-1231)	Hl	+	[149]
Hypoxylon lateripigmentum (CBS 129031, MUCL 53304, STMA 10237)	Hla	+	this work
Hypoxylon laschii (MUCL 52796)	Hlas	-	this work
Hypoxylon lechatii (MUCL 54609)	Hle	+	this work
Hypoxylon lenormandii (STMA 14013, MFLUCC 13-0596)	Hlm	+	this work
Hypomontagnella monticulosa (MUCL 54604)	Hm	+	[149]
Hypoxylon macrocarpum (CBS 119012, STMA 18259)	Hmc2	+	this work
Hypoxylon olivaceopigmentum (DSM 107924)	Hol	+	this work
Hypoxylon pulicicidum (ATCC 74245)	Hp	+	[149]
Hypoxylon perforatum (DSM 107930)	Hpe	+	this work
Hypoxylon pseudofuscum (DSM 112035, STMA 18200)	Hpf2	+	this work
Hypoxylon porphyreum (CBS 119022)	Hpo	+	this work
Hypoxylon petriniae (CBS 114746)	Hpt	+	this work
Hypoxylon rickii (CBS 129345)	Hr	+	[149]
Hypoxylon rubiginosum (MUCL 52887)	Hrub	+	[149]
Hypoxylon rubiginosum (9969, STMA 17058, DSM 106870)	Hrub2	+	this work
Hypomontagnella submonticulosa (DAOMC 242471)	Hsm	+	[149]
Hypomontagnella spongiphila (CLL-205)	Hsp	+	[149]
Hypoxylon sporistriatatunicum (STMA18003)	Hst	+	this work
Hypoxylon trugodes (MUCL 54794, CBS 135444)	Ht	+	this work
Hypoxylon texense (DSM 107933)	Hte	+	this work
Hypoxylon ticinense (CBS 115271)	Hti	+	this work
Jackrogersella cohaerens (CBS 119126)	Jc	+	this work
Jackrogersella multiformis (CBS 119016)	Jm	+	[149]
Jackrogersella minutella (CBS 135445)	Jmi	+	this work
Pyrenopolyporus hunteri (MUCL 49339)	Ph	+	[149]
Parahypoxylon papillatum (ATCC 58729, STMA 12136)	Php	+/-	this work
Parahypoxylon ruwenziorense (MUCL 51392)	Phr	+/-	this work
Xylaria hypoxylon (CBS 122620)	Xh	-	[149]

+ : BGC highly homologous (genes, number/order of genes)

+/- : BGCs with 2-4 homologous genes

- : no putative tropolone-forming cluster

7.4.3 Multiple Alignment TenS, MilS, DmbS KR Sequence

```
                     2201                                              2250
                     ER        KR                                     KR
TENS      (2196)     PPLQTRGLFKSDRTYLMVGAAGGLGTSICRWMVRNGARHVVVTSRN--PK
DMBS      (2191)     PPLQTRGLFKSDRTYLMVGAAGGLGTSLCRWMVRNGARHVVVTSRN--PK
MILS      (2196)          LFQSDRTYLMVGAAGGVGTSLCRWMVRHGARHVIVTSRN--PK
AmphB     (235)      -------RPPVHGSVLVTGGTGGIGGRVARRLAEQGAAHLVLTSRRGAD-
mFAS pig  (1873)     LTGLSKTFCPPHKSYVITGGLGGFGLQLAQWLRLRGAQKLVLTSRSGIRT
mFAS rat  (1867)     ISAISKTFCPEHKSYIITGGLGGFGLELARWLVLRGAQRLVLTSRSGIRT

                     2251                                              2300
                     KR                                               KR
TENS      (2244)     ADPEMLNEAERYGAAVQVVPMDACSKDSVQTVVDMIRATMPPIAGVCNAA
DMBS      (2239)     ADPEMLNEAERYGAIVRVVPMDACNKDSVQTVVDTIRATMPPIAGVCNAA
MILS      (2237)     GDPTMLSEAKQYGATVRVVSMDVCDRRSVEAVVGMIRATMPPIACVCNAA
AmphB     (279)      GAAELRAELEQLGVRVTIAACDAADREALAALLAEL-PEDAPLTAVFHSA
mFAS pig  (1923)     GYQARQVREWRRQGVQVLVSTSNASSLDGARSLITEATQLGPVGGVFNLA
mFAS rat  (1917)     GYQAKHVREWRRQGIHVLVSTSNVSSLEGARALIAEATKLGPVGGVFNLA

                     2301                                              2350
                     KR                                               KR
TENS      (2294)     MVLRDKLFLDMNVDHMKDVLGPKMQGTEHLDSIFAQEP--LDFFVLLSSS
DMBS      (2289)     MVLCDKLFLDMDVDQMNNTLGPKVDGTEYLDSIFAHEP--LDFFILLGSA
MILS      (2287)     MVLCDKLFLDMDVDILNNTLGPKVDGTEILDSIFSEEA--LDFFILLGST
AmphB     (328)      GVAHDD-PVDLTLGQLDALMRAKLTAARHLHELTADL--DLDAFVLFSSG
mFAS pig  (1973)     MVLRDAVLENQTPEFFQDVSKPKYSGTANLDRVTREACPELDYFVIFSSV
mFAS rat  (1967)     MVLRDAMLENQTPELFQDVNKPKYNGTLNLDRATREACPELDYFVAFSSV

                     2351                                              2400
                     KR                                               KR
TENS      (2342)     AAILNNTGQSNYHCANLYMDSLVTNRRSRGLAASIIHVGHVCDTGYVARL
DMBS      (2337)     AAILNNMGQSNYHCANLYMDSLVKHRRSRGLAASIIHIGHVCDTGYVARM
MILS      (2235)     ATIANNIGQSNYHCANLYMDSLVAQRRSRGLAASIIHIGYICDTGYVARL
AmphB               AAVFGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMATDP
mFAS pig  (2023)     SCGRGNAGQANYGFANSAMERICEKRRHDGLPGLAVQWGAIGDVGVVLET
mFAS rat  (2017)     SCGRGNAGQSNYGFANSTMERICEQRRHDGLPGLAVQWGAIGDVGIILEA

                     2401                                              2450
                     KR                                               KR
TENS      (2392)     VDDTKVQMSLGTTRVMSVSETDVHHAFAEAVRGGQPDSRSGSHNIIMGIE
DMBS      (2387)     VDDNRIQSNIATMRAMRLSETDVHHAFAQAVRGGQLDSRSGSYNIIMGIE
MILS      (2385)     GDDAKVHSNRDVMRATTLSETDVHHAFAEAVRGGSPGSPIGSYNIIMGID
AmphB     (337)      EVHDRLVRQGVLAMEPPEHALGALDQMLNDDTAAAPITMDWEMFAPAFTN
mFAS pig  (2073)     MGTNDTVIGGTLPQRIASCLEVLDLFLSQPHPVLS---------------
mFAS rat  (2067)     MGTNDTVVGGTLPQRISSCMEVLDLFLNQPHAVLS---------------
                     2451                                              2500
                     KR                                      KR
TENS      (2442)     PPTKPLDLTKRKPVWISDPRLGPCLPFSTLENQMMASEQAAAASAVDSLA
DMBS      (2437)     PPTKPLDLTRRQAVWLSDPRLGHMLPYSTLENQMIASGQAAA-S-ADSLA
MILS      (2435)     PPTKSLDSSRRKALWLSDPRLGHMVPYSASADQAVTSEQA
AmphB     (478)      RPSALLSTVPEAVSALSDE-------------------
mFAS pig  (2108)     ------------------------SFVLAEKKAAAPRDGSSQK------
mFAS rat  (2102)     ------------------------SFVLVEKKAVAHGDGEAQR------
```

▢ : Identity within PKS-NRPS, ▢ : Identity within mFAS, RED: Identity between PKS-NRPS and mFAS

7.4.4 Summary Table for Tenellin related Compounds

Table 7-3 Overview of all pretenellin A producing transformants.

Experiment	Transformant	Pentaketides				Hexaketides		
		41 353 Da	**45** 341 Da	**104** 355 Da	**39** 355 Da	**46** 381 Da	**105** 397 Da	**129** 395 Da
substrate binding helix = DmbS	KS_T1101_C1002_CD	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_A	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_B	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_C	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_D			X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_E	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_F	X		X	X	X	X	
	KS_T1705_C0707_E1407_Ten-CD_H	X		X	X	X	X	
	KS_T1101_C1002_CD	X		X	X	X	X	
substrate binding helix = MilS	KS_T0311_tenS-CM_C1701_A_1ml	X		X	X	X	X	X
	KS_T2411_tenS-CM_C1701_A_1ml				X	X	X	
	KS_T2411_tenS-CM_C1701_B				X		X	
	KS_T2810_tenS-CM_C1701_A_1ml				X	X	X	X
S2400N, L2401R, T2404M, V2406A	KS_T1705_C1207_E1907_Ten-MiSB_A			X	X	X	X	
	KS_T1705_C1207_E1907_Ten-MiSB_B			X	X			
	KS_T1705_C1207_E1907_Ten-MiSB_C			X	X			
	KS_T1705_C1207_E1907_Ten-MiSB_E	X		X	X	X	X	
	KS_T1705_C1207_E1907_Ten-MiSB_F	X		X	X	X	X	
T2395A	KS T1507 1-5-8 E - ALA 1	X		X	X			
	KS_T1101_C1002_Ala1	X		X	X			
	KS_T1401_Ala1_A_7d_10x	X		X	X			
	KS_T1401_Ala1_B_6d_nhcl_10x	X	X	X	X			
	KS_T1401_Ala1_C_7d_10x	X		X	X			
	KS_T1401_Ala1_D_7d_10x	X		X	X			
	KS_T1401_Ala1_E_6d_nhcl_10x	X	X	X	X			
K2396A	KS T3007 2-10-14 A 1z10 - ALA2	X	X	X	X			
	KS_T1401_Ala2_A_7d_10x	X	X	X	X			
	KS_T1401_Ala2_B_7d_10x	X		X	X			
	KS_T1401_Ala2_C_7d_10x	X	X	X	X			
	KS_T1401_Ala2_D_7d_10x			X	X			
V2397A	KS_T0311_C2911_E0612_Ala3_A	X			X			
	KS_T0311_C2911_E0612_Ala3_B				X			
	KS_T1111_C2911_E0612_Ala3_A				X			
Q2398A	KS_T0311_C2911_E0612_Ala4_A	X			X			
	KS_T1111_C2911_E0612_Ala4_A	X	X		X			
	KS_T1111_C2911_E0612_Ala4_B	X			X			
	KS_T1711_C0812_E1512_Ala4_A	X		X	X			
M2399A	KS_T0209_1-5-8_D_1z10 - ALA5	X		X	X			
S2400A	KS_T1401_Ala6_A_7d_10x	X		X	X			
	KS_T1401_Ala6_B_7d_10x	X		X	X			
	KS_T2611_ala6_B_2x	X	X	X	X			
	KS_T2611_ala6_C_2x	X	X	X	X			
	KS_T2611_ala6_D_2x	X	X	X	X			
L2401A	KS_Ala7_0611_D_10x		X		X			
	KS_T1401_Ala7_A_6d_nhcl_10x	X	X		X			
	KS_T1401_Ala7_C_7d_10x				X			
G2402A	KS T0907 1-5-8 C - ALA 8	X		X	X	X		
	KS T1507 1-5-8 D - ALA 8			X	X			

	KS T3007 1-5-8 A 1z10 - ALA8	X		X	X	X	
	KS T3007 1-5-8 B 1z10 - ALA8	X		X	X	X	
	KS T3007 1-5-8 E 1z10 - ALA8	X		X	X	X	
	KS T3007 1-5-8 F 1z10 - ALA8	X		X	X		
	KS T3007 1-5-8 G 1z10 - ALA8	X		X	X		
	KS T3007 1-5-8 H 1z10 - ALA8	X		X	X	X	
	KS_T2808_1-5-8_A_1z10 - ALA8	X		X	X		
T2403A	KS_Ala9_0611_A_10x	X	X	X	X		
	KS_Ala9_0611_B_10x	X	X	X	X		
	KS_Ala9_0611_C_10x	X	X	X	X		
	KS_Ala9_0611_D_10x	X	X	X	X		
	KS_Ala9_0611_E_10x	X	X	X	X		
	KS_Ala9_0611_F_10x	X	X	X	X		
	KS_Ala9_0611_G_10x	X	X	X	X		
	KS_T0209_7-9-11_D_1z10 - ALA 9	X	X	X	X		
T2404A	KS_2-10-14_T1206_BG - ALA10	X		X	X		
	KS_2-10-14_T1206_BG_z10 - ALA10	X		X	X		
	KS_2-10-14_T1206_CG - ALA10	X		X	X		
	KS_2-10-14_T1206_EG - ALA10			X	X		
	KS_T1101_C1002_Ala10_A			X	X		
	KS_T1101_C1002_Ala10_B						
	KS_T2808_2-10-12_A_1z10 - ALA10	X		X	X		X
R2405A	KS_T1009_7-9-11_A - ALA11		X	X	X		
	KS_T1009_7-9-11_B - ALA11	X	X	X	X		
	KS_T1009_7-9-11_C - ALA11	X	X	X	X		
V2406	KS_T1401_Ala12_B_7d_10x	X		X	X		X
	KS_T1401_Ala12_D_7d_10x	X		X	X		X
	KS_T1401_Ala12_E_7d_10x	X		X	X	X	X
M2407A	KS_T1401_Ala13_A_7d_10x			X	X		
	KS_T1401_Ala13_B_7d_10x			X	X		
	KS_T1401_Ala13_C_7d_10x			X	X		
	KS_T2801_Ala13_B_7d_10x	X		X	X		
	KS_T2801_Ala13_C_7d_10x	X	X	X	X	X	X
S2408A	KS T3007 2-10-14 B 1z10 - ALA14	X		X	X		
	KS T3007 2-10-14 C 1z10 - ALA14	X	X	X	X		
	KS T3007 2-10-14 E 1z10 - ALA14	X		X	X		
	KS T3007 2-10-14 F 1z10 - ALA14	X	X	X	X		
	KS T3007 2-10-14 G 1z10 - ALA14	X	X	X	X		
V2409A	KS_T0209_12-13-15_C_1z10-ALA 15	X	X	X	X		
	KS_T1401_Ala15_D_7d_10x	X	X	X	X		

8 References

1. Pickens, L. B., Tang, Y. & Chooi, Y. H. Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* **2**, 211–236 (2011).

2. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).

3. Harvey, A. L. Natural products in drug discovery. *Drug Discov. Today* **13**, 894–901 (2008).

4. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).

5. Drew, S. W. & Demain, A. L. Effect of primary metabolites on secondary metabolism. *Annu. Rev. Microbiol.* **31**, 343–356 (1977).

6. Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. Why are secondary metabolites (Natural Products) biosynthesized. *J. Nat. Prod.* **52**, 1189–1208 (1989).

7. Ewa, Mł. Survey of plant pigments: Molecular and environmental determinants of plant colors. *Acta Biol. Cracoviensia Ser. Bot.* **51**, 7–16 (2009).

8. Netzker, T. *et al.* Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters. *Front. Microbiol.* **6**, 1–13 (2015).

9. Brakhage, A. A. Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **11**, 21–32 (2013).

10. Keller, N. P., Turner, G. & Bennett, J. W. Fungal secondary metabolism - from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937–947 (2005).

11. Pott, D. M., Osorio, S. & Vallarino, J. G. From central to specialized metabolism: An overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. *Front. Plant Sci.* **10**, (2019).

12. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–837 (2004).

13. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010–2026 (2007).

14. Loron, C. C. *et al.* Early fungi from the Proterozoic era in Arctic Canada. *Nature* **570**, 232–235 (2019).

15. Lücking, R. *et al.* Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* **11**, 14 (2020).

16. Blackwell, M. The fungi: 1, 2, 3 ... 5.1 million species? *Am. J. Bot.* **98**, 426–438 (2011).

17. Caesar, L. K., Kelleher, N. L. & Keller, N. P. In the fungus where it happens: History and future propelling Aspergillus nidulans as the archetype of natural products research. *Fungal Genet. Biol.* **144**, 103477 (2020).

18. Van Santen, J. A. *et al.* The Natural Products Atlas 2.0: A database of microbially-derived natural products. *Nucleic Acids Res.* **50**, D1317–D1323 (2022).

19. Meng, X. *et al.* Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity. *Biotechnol. Adv.* **54**, 107866 (2022).

20. Wang, J. *et al.* Structural basis for the biosynthesis of lovastatin. *Nat. Commun.* **12**, 1–10 (2021).

21. Herbst, D. A., Townsend, C. A. & Maier, T. The architectures of iterative type I PKS and FAS. *Nat. Prod. Rep.* **35**, 1046–1069 (2018).

22. Birch, A. J., Gager, F., Mo, L., Pelter, A. & Wright, J. J. Studies in relation to biosynthesis. 2-hydroxy-6-methylbenzoic acid in Penicillium griseofulvum. *Aust. J. Chem.* **8**, 539–544 (1955).

23. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **26**, 90–114 (2009).

24. Skellam, E. Biosynthesis of fungal polyketides by collaborating and trans-acting enzymes. *Nat. Prod. Rep.* **39**, 754–783 (2021).

25. Yakasai, A. A. *et al.* Nongenetic reprogramming of a fungal highly reducing polyketide synthase. *J. Am. Chem. Soc.* **133**, 10990–10998 (2011).

26. Crawford, J. M. *et al.* Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139–1143 (2009).

27. Cox, R. J. Curiouser and curiouser: progress in understanding the programming of iterative highly-reducing polyketide synthases. *Nat. Prod. Rep.* **40**, 9–27 (2023).

28. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science (80-.).* **321**, 1315–1322 (2008).

29. Paiva, P. *et al.* Animal Fatty Acid Synthase: A Chemical Nanofactory. *Chem. Rev.* **121**, 9502–9553 (2021).

30. Ma, S. M. & Tang, Y. Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB. *FEBS J.* **274**, 2854–2864 (2007).

31. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

32. Varadi, M. & Velankar, S. The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics* 1–10 (2022).

33. Bedford, D. J., Schweizer, E., Hopwood, D. A. & Khosla, C. Expression of a functional fungal polyketide synthase in the bacterium Streptomyces coelicolor A3(2). *J. Bacteriol.* **177**, 4544–4548 (1995).

34. Alberti, F., Foster, G. D. & Bailey, A. M. Natural products from filamentous fungi and production by heterologous expression. *Appl. Microbiol. Biotechnol.* **101**, 493–500 (2017).

35. Fujii, I. *et al.* Cloning of the polyketide synthase gene atX from Aspergillus terreus and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. *Mol. Gen. Genet.* **253**, 1–10 (1996).

36. Medema, M. H. The year 2020 in natural product bioinformatics: An overview of the latest tools and databases. *Nat. Prod. Rep.* **38**, 301–306 (2021).

37. Shenouda, M. L., Ambilika, M., Skellam, E. & Cox, R. J. Heterologous Expression of Secondary Metabolite Genes in Trichoderma reesei for Waste Valorization. *J. Fungi* **8**, (2022).

38. Schneider, P., Misiek, M. & Hoffmeister, D. In vivo and in vitro production options for fungal secondary metabolites. *Mol. Pharm.* **5**, 234–242 (2008).

39. Pahirulzaman, K. A. K., Williams, K. & Lazarus, C. M. *A toolkit for heterologous expression of metabolic pathways in aspergillus oryzae. Methods in Enzymology* vol. 517 (Elsevier Inc., 2012).

40. Jin, F. J., Maruyama, J. I., Juvvadi, P. R., Arioka, M. & Kitamoto, K. Development of a novel quadruple auxotrophic host transformation system by argB gene disruption using adeA gene and exploiting adenine auxotrophy in Aspergillus oryzae. *FEMS Microbiol. Lett.* **239**, 79–85 (2004).

41. Lubertozzi, D. & Keasling, J. D. Developing Aspergillus as a host for heterologous expression. *Biotechnol. Adv.* **27**, 53–75 (2009).

42. Heneghan, M. N. *et al.* First heterologous reconstruction of a complete functional fungal biosynthetic multigene cluster. *ChemBioChem* **11**, 1508–1512 (2010).

43. Lazarus, C. M., Williams, K. & Bailey, A. M. Reconstructing fungal natural product biosynthetic pathways. *Nat. Prod. Rep.* **31**, 1339–1347 (2014).

44. Bailey, A. M. *et al.* Cleaning the Cellular Factory--Deletion of McrA in Aspergillus oryzae NSAR1 and the Generation of a Novel Kojic Acid Deficient Strain for Cleaner Heterologous Production of Secondary Metabolites. *Front. Fungal Biol.* **2**, 1–12 (2021).

45. Kück, U. & Hoff, B. New tools for the genetic manipulation of filamentous fungi. *Appl. Microbiol. Biotechnol.* **86**, 51–62 (2010).

46. He, Y. & Cox, R. J. The molecular steps of citrinin biosynthesis in fungi. *Chem. Sci.* **7**, 2119–2127 (2016).

47. Song, Z. *et al.* Heterologous expression of the avirulence gene ACE1 from the fungal rice pathogen Magnaporthe oryzae. *Chem. Sci.* **6**, 4837–4845 (2015).

48. Lowry, B., Walsh, C. T. & Khosla, C. In Vitro reconstitution of metabolic pathways: Insights into nature's chemical logic. *Synlett* **26**, 1008–1025 (2015).

49. Kaur, J., Kumar, A. & Kaur, J. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. *Int. J. Biol. Macromol.* **106**, 803–822 (2018).

50. Burgess-Brown, N. A. *et al.* Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr. Purif.* **59**, 94–102 (2008).

51. Eley, K. L. *et al.* Biosynthesis of the 2-pyridone tenellin in the insect pathogenic fungus Beauveria bassiana. *ChemBioChem* **8**, 289–297 (2007).

52. Mcinnes, A. G., Smith, D. G., Wat, C.-K., Vining, L. C. & Wright, J. L. C. Tenellin and Bassianin, Metabolites of Beauueria Species. Structure Elucidation with 15N- and Doubly 13C-Enriched Compounds using 13C Nuclear Magnetic Resonance Spectroscopy. *J. Chem. Soc. Chem. Commun.* **8**, 281–282 (1974).

53. Wat, C.-K., Mcinnes, A. G., Smith, D. G., Wright, J. L. C. & Vining, L. C. The yellow pigments of Beauveria species. Structures of tenellin and bassianin . *Can. J. Chem.* **55**, 4090–4098 (1977).

54. Halo, L. M. *et al.* Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires cooexpression with an enoyl reductase. *ChemBioChem* **9**, 585–594 (2008).

55. Matsumoto, M. & Minato, H. Structure of ilicicolin H, an antifungal antibiotic. *Tetrahedron Lett.* 3827–3830 (1976).

56. Heneghan, M. N. *et al.* The programming role of trans-acting enoyl reductases during the biosynthesis of highly reduced fungal polyketides. *Chem. Sci.* **2**, 972–979 (2011).

57. Halo, L. M. *et al.* Late stage oxidations during the biosynthesis of the 2-pyridone tenellin in the entomopathogenic fungus beauveria bassiana. *J. Am. Chem. Soc.* **130**, 17988–17996 (2008).

58. Yang, X.-L. *et al.* Molecular basis of methylation and chain-length programming in a fungal iterative highly reducing polyketide synthase. *Chem. Sci.* **10**, 8478–8489 (2019).

59. Ma, S. M. *et al.* Complete reconstitution of a highly reducing iterative polyketide synthase. *Science (80-.).* **326**, 589–592 (2009).

60. Cacho, R. A. *et al.* Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J. Am. Chem. Soc.* **137**, 15688–15691 (2015).

61.   Roberts, D. M. *et al.* Substrate selectivity of an isolated enoyl reductase catalytic domain from an iterative highly reducing fungal polyketide synthase reveals key components of programming. *Chem. Sci.* **8**, 1116–1126 (2017).

62.   Ames, B. D. *et al.* Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11144–11149 (2012).

63.   Fisch, K. M. *et al.* Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* **133**, 16635–16641 (2011).

64.   Wijkman, N. Über einige neue, durch Schimmelpilze gebildete Substanzen. *Justus Liebigs Ann. Chem.* **485**, 61–73 (1931).

65.   Raistrick, H. & Smith, G. The metabolic products of Byssochlamys fulva Olliver and Smith. *Biochem J.* **27**, 1814–1819 (1933).

66.   Baldwin, J. E. *et al.* The constitutions of glauconic, glaucanic and byssochlamic acids. *Experientia* **18**, 345–352 (1962).

67.   Hamor, T. A., Paul, I. C., Monteath Robertson, J. & Sim, G. A. The structure of byssochlamic acid. *Experientia* **18**, 352–354 (1962).

68.   Bloomer, B. J. L., Moppett, C. E. . & Sutherland, J. K. The Nonadrides. Part V.l Biosynthesis of Glauconic Acid. *J. Chem. Soc. C Org.* 588–591 (1968).

69.   Barton, D. H. R. & Sutherland, J. K. The Nonadrides. Part I. Introduction and General Survey. *J. Chem. Soc.* 1769–1772 (1965).

70.   Williams, K. *et al.* Maleidride biosynthesis - construction of dimeric anhydrides - more than just heads or tails. *Nat. Prod. Rep.* (2022).

71.   Williams, K. *et al.* Genetic and chemical characterisation of the cornexistin pathway provides further insight into maleidride biosynthesis. *Chem. Commun.* **53**, 7965–7968 (2017).

72.   Williams, K., de Mattos-Shipley, K. M. J., Willis, C. L. & Bailey, A. M. In silico analyses of maleidride biosynthetic gene clusters. *Fungal Biol. Biotechnol.* **9**, 1–34 (2022).

73.   Chen, X. X., Zheng, Y. & Shen, Y. Natural products with maleic anhydride structure: Nonadrides, tautomycin, chaetomellic anhydride, and other compounds. *Chem. Rev.* **107**, 1777–1830 (2007).

74.   Schor, R. & Cox, R. Classic fungal natural products in the genomic age: The molecular legacy of Harold Raistrick. *Nat. Prod. Rep.* **35**, 230–256 (2018).

75.   Büchi, G; Snader, K. M.; White, J.D.; Zanos Gougoutas, J. . S. S. Structures of Rubratoxins A and B. *J. Am. Chem. Soc.* **92**, 6638–6641 (1970).

76.   Sulikowski, G. A., Agnelli, F., Spencer, P., Koomen, J. M. & Russell, D. H. Studies on the Biosynthesis of Phomoidride B (CP-263,114): Evidence for a Decarboxylative Homodimerization Pathway. *Org. Lett.* **4**, 1447–1450 (2002).

77. Isaka, M., Tanticharoen, M. & Thebtaranonth, Y. Cordyanhydrides A and B. Two unique anhydrides from the insect pathogenic fungus Cordyceps pseudomilitaris BCC 1620. *Tetrahedron Lett.* **41**, 1657–1660 (2000).

78. Tian, D. S. *et al.* The sporothriolides. A new biosynthetic family of fungal secondary metabolites. *Chem. Sci.* **11**, 12477–12484 (2020).

79. Krohn, K., Ludewig, K., Aust, H. J., Draeger, S. & Schulz, B. Biologically Active Metabolites from fungi. 3. sporothriolide, discosiolide, and 4-epi-ethisolide—new furofurandiones from sporothrix sp., discosia sp., and pezicula livida. *J. Antibiot. (Tokyo).* **47**, 113–118 (1994).

80. Lebe, K. E. & Cox, R. J. Oxidative steps during the biosynthesis of squalestatin S1. *Chem. Sci.* **10**, 1227–1231 (2019).

81. Dawson, M. J. *et al.* The squalestatins, novel inhibitors of squalene synthase produced by a species of phoma. V. minor metabolites. *J. Antibiot. (Tokyo).* **47**, 740–754 (1994).

82. Verschueren, K. H. G. *et al.* Structure of ATP citrate lyase and the origin of citrate synthase in the Krebs cycle. *Nature* **568**, 571–575 (2019).

83. Guo, H., Roman, D. & Beemelmanns, C. Tropolone natural products. *Nat. Prod. Rep.* **36**, 1137–1155 (2019).

84. Cox, R. J. & Al-Fahad, A. Chemical mechanisms involved during the biosynthesis of tropolones. *Curr. Opin. Chem. Biol.* **17**, 532–536 (2013).

85. Zhao, J. & Zhao, J. Plant Troponoids: Chemistry, Biological Activity, and Biosynthesis. *Curr. Med. Chem.* **14**, 2597–2621 (2007).

86. Chen, X. *et al.* Biosynthesis of Tropolones in Streptomyces spp.: Interweaving Biosynthesis and Degradation of Phenylacetic Acid and Hydroxylations on the Tropone Ring. *Appl. Environ. Microbiol.* **84**, (2018).

87. Fujita, K., Yamaguchi, T., Itose, R. & Sakai, K. Biosynthetic Pathway of beta-Thujaplicin in the Cupressus lusitanica Cell Culture. *J. Plant Physiol.* **156**, 462–467 (2000).

88. Liu, J. *et al.* Concise Biosynthesis of Tropone-Containing Spiromaterpenes by a Sesquiterpene Cyclase and a Multifunctional P450 from a Deep-Sea-Derived Spiromastix sp. Fungus. *J. Nat. Prod.* **5**, 2–9 (2022).

89. Acid, S. Structure of Stipitatic Acid Age of the Baker ' s Hole Coombe Rock , Resins and Wood-Resin Composites , and. **155**, 50–51 (1945).

90. Birkinshaw, J. H., Chambers, A. R. & Raistrick, H. Studies in the biochemistry of micro-organisms: Stipitatic acid, $C_8H_6O_5$, a metabolic product of Penicillium stipitatum Thom. *Biochem. J.* **36**, 242–51 (1942).

91. Birkinshaw, J., Journal, H. R.-B. & 1932, U. Studies in the biochemistry of micro-organisms: Puberulic acid $C_8H_6O_6$ and an acid $C_8H_4O_6$, new products of the metabolism of glucose by Penicillium puberulum. *Biochem. J.* **26**, 441–453 (1932).

92.   Deway, M. J. S. Structure of Colchicine. *Nature* **155**, 141–142 (1945).

93.   Davison, J. *et al.* Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7642–7647 (2012).

94.   Bentley, R. A fresh look at natural tropolonoids. *Nat. Prod. Rep.* **25**, 118–138 (2008).

95.   Schotte, C., Li, L., Wibberg, D., Kalinowski, J. & Cox, R. J. Synthetic Biology Driven Biosynthesis of Unnatural Tropolone Sesquiterpenoids. *Angew. Chem. Int. Ed. Engl.* **59**, 23870–23878 (2020).

96.   Oxford, A. E. & Raistrick, H. Studies in the biochemistry of micro-organisms. *Biochem. J.* **42**, 323–329 (1948).

97.   Wright, J. L. C., McInnes, A. G., Smith, D. G. & Vining, L. C. Structure of sepedonin, a tropolone metabolite of Sepedomum chrysospermum Fries . *Can. J. Chem.* **48**, 2702–2708 (1970).

98.   Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: A tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041–1072 (2007).

99.   Anderson, V. E. & Hammes, G. G. Stereochemistry of the Reactions Catalyzed by Chicken Liver Fatty Acid Synthase. *Biochemistry* **23**, 2088–2094 (1984).

100.   Seyama, Y. *et al.* Origin of hydrogen atoms in the fatty acids synthesized with yeast fatty acid synthetase. *J. Biochem.* **82**, 1325–1329 (1977).

101.   Dugan, R. E., Slakey, L. L. & Porter, J. W. Stereospecificity of the transfer of hydrogen from reduced nicotinamide adenine dinucleotide phosphate to the acyl chain in the dehydrogenase-catalyzed reactions of fatty acid synthesis. *J. Biol. Chem.* **245**, 6312–6316 (1970).

102.   Saito, K. *et al.* Stereospecificity of hydrogen transfer by pyridine nucleotide-dependent enoyl reductases in fatty acid synthesis: Studies with enzymes obtained from developing castor bean seeds and Chlorella vulgaris. *Plant Cell Physiol.* **21**, 9–19 (1980).

103.   Seyama, Y., Kasama, T., Yamakawa, T., Kawaguchi, A. & Okuda, S. Stereochemical studies of hydrogen incorporation from nucleotides with fatty acid synthetase from Brevibacterium ammoniagenes. *Adv. Exp. Med. Biol.* **101**, 37–43 (1978).

104.   Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science (80-.).* **268**, 1487–1489 (1995).

105.   Kao, C. M. *et al.* Alcohol Stereochemistry in Polyketide Backbones Is Controlled by the â-Ketoreductase Domains of Modular Polyketide Synthases. *J. Am. Chem. Soc.* **120**, 2478–2479 (1998).

106.   McDaniel, R., Kao, C. M., Hwang, S. J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4**, 667–674 (1997).

107. Zheng, J. & Keatinge-Clay, A. T. The status of type I polyketide synthase ketoreductases. *Medchemcomm* **4**, 34–40 (2013).

108. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. The SDR superfamily: func tional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).

109. Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20**, 772–783 (2013).

110. McPherson, M., Khosla, C., Cane, D. E., V, S. U. & January, R. V. Erythromycin Biosynthesis : The -Ketoreductase Domains Catalyze the Stereospecific Transfer of the 4- pro - S Hydride of NADPH. *J. Am. Chem. Soc.* **7863**, 3267–3268 (1998).

111. Passmore, M., Gallo, A., Lewandowski, J. R. & Jenner, M. Molecular basis for acyl carrier protein-ketoreductase interaction intrans-acyltransferase polyketide synthases. *Chem. Sci.* **12**, 13676–13685 (2021).

112. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).

113. Zhang, Y. M. *et al.* Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on β-Ketoacyl-ACP Synthase III. *J. Biol. Chem.* **276**, 8231–8238 (2001).

114. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **14**, 898–908 (2007).

115. Terwilliger, T. C. *et al.* AlphaFold predictions are valuable hypotheses, and accelerate but do not replace experimental structure determination. *bioRxiv* (2023) doi:https://doi.org/10.1101/2022.11.21.517405.

116. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

117. Nussinov, R., Zhang, M., Liu, Y. & Jang, H. AlphaFold, allosteric, and orthosteric drug discovery: Ways forward. *Drug Discov. Today* **28**, 103551 (2023).

118. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

119. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

120. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).

121. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).

122. Liu, C. *et al.* Substrate-bound structures of a ketoreductase from amphotericin modular polyketide synthase. *J. Struct. Biol.* **203**, 135–141 (2018).

123. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. *Proteins Struct. Funct. Genet.* **47**, 393–402 (2002).

124. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinforma.* **77**, 114–122 (2009).

125. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 255–461 (2010).

126. Gaillard, T. Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark. *J. Chem. Inf. Model.* **58**, 1697–1706 (2018).

127. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).

128. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **9**, 1–5 (1996).

129. Piech, O. Computational and In Vitro Study of Isolated Domains from Fungal Polyketide Synthases. (Gottfried Wilhelm Leibniz Universität Hannover, 2020).

130. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).

131. Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 35–37 (2007).

132. Katzen, F. Gateway® recombinational cloning: A biological operating system. *Expert Opin. Drug Discov.* **2**, 571–589 (2007).

133. Smith, R. H. B., Dar, A. C. & Schlessinger, A. PyVOL: a PyMOL plugin for visualization, comparison, and volume calculation of drug-binding sites. *bioRxiv* **c**, 816702 (2019).

134. Chen, K. H., Le, S. P., Han, X., Frias, J. M. & Nowick, J. S. Alanine scan reveals modifiable residues in teixobactin. *Chem. Commun.* **53**, 11357–11359 (2017).

135. Stadler, M., Kuhnert, E., Peršoh, D. & Fournier, J. The Xylariaceae as model example for a unified nomenclature following the 'One Fungus-One Name' (1F1N) concept. *Mycology* **4**, 5–21 (2013).

136. Kuhnert, E. *et al.* Lenormandins A—G, new azaphilones from Hypoxylon lenormandii and Hypoxylon jaklitschii sp. nov., recognised by chemotaxonomic data. *Fungal Divers.* **71**, 165–184 (2015).

137. Surup, F. *et al.* Identification of fungal fossils and novel azaphilone pigments in ancient carbonised specimens of Hypoxylon fragiforme from forest soils of Châtillon-sur-Seine (Burgundy). *Fungal Divers.* **92**, 345–356 (2018).

138. Surup, F. *et al.* Sporothriolide derivatives as chemotaxonomic markers for Hypoxylon monticulosum. *Mycology* **5**, 110–119 (2014).

139. Yuyama, K. T. *et al.* Cytochalasans act as inhibitors of biofilm formation of staphylococcus aureus. *Biomolecules* **8**, 1–13 (2018).

140. Wang, C. *et al.* Investigating the Function of Cryptic Cytochalasan Cytochrome P450 Monooxygenases Using Combinatorial Biosynthesis. *Org. Lett.* **21**, 8756–8760 (2019).

141. Wibberg, D. *et al.* High quality genome sequences of thirteen Hypoxylaceae (Ascomycota) strengthen the phylogenetic family backbone and enable the discovery of new taxa. *Fungal Divers.* **106**, 7–28 (2021).

142. Vasilyeva, L. N., Stephenson, S. L., Hyde, K. D. & Bahkali, A. H. Some stromatic pyrenomycetous fungi from northern Thailand - 1. Biscogniuxia, Camillea and Hypoxylon (Xylariaceae). *Fungal Divers.* **55**, 65–76 (2012).

143. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

144. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309-12 (2004).

145. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).

146. Meyer, F. *et al.* GenDB - An open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31**, 2187–2195 (2003).

147. Tatusov, R. L. *et al.* The COG database: An updated vesion includes eukaryotes. *BMC Bioinformatics* **4**, 1–14 (2003).

148. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, (2004).

149. Kuhnert, E. *et al.* Secondary metabolite biosynthetic diversity in the fungal family Hypoxylaceae and Xylaria hypoxylon. *Stud. Mycol.* **99**, 1–43 (2021).

150. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

151. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

152. Bloomer, J. L., Moppett, C. E. & Sutherland, J. K. The Biosynthesis of Glauconic Acid. *Chem. Commun.* **24**, 619–621 (1965).

153. Cox, R. E. & Holker, J. S. E. Biosynthesis of glauconic acid from [2,3-^{13}C]succinate. *J. Chem. Soc. Chem. Commun.* 583–584 (1976).

154. Nieminen, S., Payne, T. G., Senn, P. & Tamm, C. Zur Biosynthese der Rubratoxine. *Helv. Chim. Acta* **64**, 2162–2174 (1981).

155. Spencer, P., Agnelli, F., Williams, H. J., Keller, N. P. & Sulikowski, G. A. Biosynthetic Studies on the Fungal Secondary Metabolites CP-225,917 and CP-263,114. *J. Am. Chem. Soc.* **122**, 420–421 (2000).

156. Fujii, R. *et al.* Biosynthetic Study on Antihypercholesterolemic Agent Phomoidride: General Biogenesis of Fungal Dimeric Anhydrides. *Org. Lett.* **17**, 5658–5661 (2015).

157. De Mattos-Shipley, K. M. J. *et al.* Uncovering biosynthetic relationships between antifungal nonadrides and octadrides. *Chem. Sci.* **11**, 11570–11578 (2020).

158. Williams, K. *et al.* Heterologous Production of Fungal Maleidrides Reveals the Cryptic Cyclization Involved in their Biosynthesis Angewandte. *Angew. Chemie* **55**, 6784–6788 (2016).

159. Szwalbe, A. 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. *Chem. Commun.* **51**, 17088–17091 (2015).

160. Yin, S., Friedrich, S., Hrupins, V. & Cox, R. J. In vitro studies of maleidride-forming enzymes. *RSC Adv.* **11**, 14922–14931 (2021).

161. Moppett, C. E. & Sutherland, J. K. The Biosynthesis of Glauconic Acid: C9 Precursors. *Chem. Commun.* 772–773 (1966).

162. Shiina, T. *et al.* Oxidative Ring Contraction by a Multifunctional Dioxygenase Generates the Core Cyclooctadiene in the Biosynthesis of Fungal Dimeric Anhydride Zopfiellin. *Org. Lett.* **22**, acs.orglett.0c00340 (2020).

163. Serre, L., Vallée, B., Bureaud, N., Schoentgen, F. & Zelwer, C. Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: A novel structural class of phospholipid-binding proteins. *Structure* **6**, 1255–1265 (1998).

164. Yamamoto, S. *et al.* Elucidation of Late-Stage Biosynthesis of Phomoidride: Proposal of Cyclization Mechanism Affording Characteristic Nine-Membered Ring of Fungal Dimeric Anhydride. (2022).

165. Gilchrist, C. L. M. M. & Chooi, Y.-H. H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).

166. Aldrige, David C.Carman, R. M. & Moore, R. B. A New Tricarboxylic Acid Anhydride from Paeciomyces variotii By. *J. Chem. Soc. Perkin Trans. 1* 2134–2135 (1980).

167. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. PredictingTransmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

168. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).

169. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the Expasy Server. *John M. Walk. Proteomics Protoc. Handbook, Humana Press* 571–608 (2005).

170. Bond-Watts, B. B., Weeks, A. M. & Chang, M. C. Y. Biochemical and structural characterization of the trans-enoyl-coa reductase from treponema denticola. *Biochemistry* **51**, 6827–6837 (2012).

171. Bai, J. *et al.* A Cascade of Redox Reactions Generates Complexity in the Biosynthesis of the Protein Phosphatase-2 Inhibitor Rubratoxin A. *Angew. Chemie - Int. Ed.* **56**, 4782–4786 (2017).

172. Isaka, M., Kongsaeree, P. & Thebtarannonth, Y. Bioxanthracenes from the insect pathogenic fungus Cordyceps pseudomilitaris BCC 1620 I Taxonomy, Fermentation, Isolation and Antimalarial Activity. *J. Antibiot. (Tokyo).* **54**, 36–43 (2001).

173. Kikuchi, H. *et al.* Novel trichothecanes, paecilomycine A, B, and C, isolated from entomopathogenic fungus, Paecilomyces tenuipes. *Tetrahedron Lett.* **45**, 6225–6228 (2004).

174. Da Silva, P. H. F. *et al.* Antifungal polyketides and other compounds from Amazonian endophytic talaromyces fungi. *J. Braz. Chem. Soc.* **29**, 622–630 (2018).

175. Huff, R. K., Moppett, C. E. & Sutherland, J. K. The Nonadrides. Part VI. Dimerisation of the C9 Unit in vivo and in vitro. *J. Chem. Soc.* 2584–2590 (1972).

176. Huff, B. R. K., Moppett, C. E. & Sutherland, J. K. A Novel Synthesis of a Nine-membered Ring. *Chem. Commun.* 1992–1993 (1968).

177. Gerke, J. *et al.* Biosynthesis of Antibacterial Iron-Chelating Tropolones in Aspergillus nidulans as Response to Glycopeptide-Producing Streptomycetes. *Front. Fungal Biol.* **2**, 1–19 (2022).

178. Gerke, J. *et al.* Breaking the silence: Protein stabilization uncovers silenced biosynthetic gene clusters in the fungus Aspergillus nidulans. *Appl. Environ. Microbiol.* **78**, 8234–8244 (2012).

179. Quang, D. N. *et al.* Ampullosine , a new Isoquinoline Alkaloid from Sepedonium ampullosporum (Ascomycetes). *Nat. Prod. Commun.* **5**, 869–872 (2010).

180. Schor, R., Schotte, C., Wibberg, D., Kalinowski, J. & Cox, R. J. Three previously unrecognised classes of biosynthetic enzymes revealed during the production of xenovulene A. *Nat. Commun.* **9**, 1963 (2018).

181. Borodovsky, M. & McIninch, J. GENMARK: Parallel gene recognition for both DNA strands. *Comput. Chem.* **17**, 123–133 (1993).

182. Gilchrist, C. L. M. *et al.* Cblaster: a Remote Search Tool for Rapid Identification and Visualization of Homologous Gene Clusters. *Bioinforma. Adv.* **1**, 1–19 (2021).

183. Asai, T. *et al.* Use of a biosynthetic intermediate to explore the chemical diversity of pseudo-natural fungal polyketides. *Nat. Chem.* **7**, 737–743 (2015).

184. McInnes, A. G., Smith, D. G., Vining, L. C. & Wright, J. L. C. Use of carbon-13 in biosynthetic studies. Incorporation of isotopically labelled acetate and formate into the fungal tropolone, sepedonin. *Chem. Commun.* 1669–1670 (1968).

185. J. Wright, D. G. Smith, A. G. McInnes & L. C. Vining. Use of 13C in biosynthetic studies. Incorporation of acetate and formate into the fungal tropolone, sepedonin. *Can. J. Biochem.* **47**, 945–949 (1969).

186. Chenault, H. K. & Whitesides, G. M. Regeneration of Nicotinamide Cofactors. *Appl. Biochem. Biotechnol.* **14**, 147–197 (1987).

187. Pyser, J. B. *et al.* Stereodivergent, Chemoenzymatic Synthesis of Azaphilone Natural Products. *J. Am. Chem. Soc.* **141**, 18551–18559 (2019).

188. Dockrey, S. A. B. *et al.* Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **10**, 119–125 (2018).

189. Cedeño-sanchez, M. *et al.* Segregation of the genus Parahypoxylon (Hypoxylaceae , Xylariales) from Hypoxylon by a polyphasic taxonomic approach. **162**, 131–162 (2023).

190. Hayes, R. P., Lewis, K. M., Xun, L. & Kang, C. H. Catalytic mechanism of 5-chlorohydroxyhydroquinone dehydrochlorinase from the YCII superfamily of largely unknown function. *J. Biol. Chem.* **288**, 28447–28456 (2013).

191. Roth, C., Gröning, J. A. D., Kaschabek, S. R., Schlömann, M. & Sträter, N. Crystal structure and catalytic mechanism of chloromuconolactone dehalogenase ClcF from Rhodococcus opacus 1CP. *Mol. Microbiol.* **88**, 254–267 (2013).

192. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).

193. Arnesano, F., Banci, L., Bertini, I., Huffman, D. L. & O'Halloran, T. V. Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1. *Biochemistry* **40**, 1528–1539 (2001).

194. Schreiter, E. R. *et al.* Crystal structure of the nickel-responsive transcription factor NikR. *Nat. Struct. Biol.* **10**, 794–799 (2003).

195. Ren, J. *et al.* The structure and transcriptional analysis of a global regulator from Neisseria meningitidis. *J. Biol. Chem.* **282**, 14655–14664 (2007).

196. Hantke, V., Wang, C., Skellam, E. J. & Cox, R. J. Function of pathway specific regulators in the: ACE1 and pyrichalasin H biosynthetic gene clusters. *RSC Adv.* **9**, 35797–35802 (2019).

197. Zhang, H., Hantke, V., Bruhnke, P., Skellam, E. J. & Cox, R. J. Chemical and Genetic Studies on the Formation of Pyrrolones During the Biosynthesis of Cytochalasans. *Chem. - A Eur. J.* **27**, 3106–3113 (2021).

198. Feng, J. *et al.* Biosynthesis of oxygenated brasilane terpene glycosides involves a promiscuous N-acetylglucosamine transferase. *Chem. Commun. (Camb).* **56**, 12419–12422 (2020).

199. Chen, W. *et al.* Orange, red, yellow: biosynthesis of azaphilone pigments in Monascus fungi. *Chem. Sci.* **8**, 4917–4925 (2017).

200. Li, M. *et al.* Monasone naphthoquinone biosynthesis and resistance in monascus fungi. *MBio* **11**, 1–16 (2020).

201. Qiao, J. *et al.* Transcriptome analysis revealing molecular mechanisms of enhanced pigment yield by succinic acid and fluconazole. *Prep. Biochem. Biotechnol.* **52**, 990–1000 (2022).

202. Chen, W., Feng, Y., Molnár, I. & Chen, F. Nature and nurture: Confluence of pathway determinism with metabolic and chemical serendipity diversifies: Monascus azaphilone pigments. *Nat. Prod. Rep.* **36**, 561–572 (2019).

203. Liu, J. *et al.* Tandem intermolecular [4 + 2] cycloadditions are catalysed by glycosylated enzymes for natural product biosynthesis. *Nat. Chem.* (2023).

204. Chen, Q. *et al.* Enzymatic Intermolecular Hetero-Diels-Alder Reaction in the Biosynthesis of Tropolonic Sesquiterpenes. *J. Am. Chem. Soc.* **141**, 14052–14056 (2019).

205. Li, L. & Cox, R. J. Stereochemical and Biosynthetic Rationalisation of the Tropolone Sesquiterpenoids. *J. Fungi* **8**, (2022).

206. Gietz, R. D. & Schiestl, R. H. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 35–37 (2007).

207. Gietz, R. D. & Woods, R. A. Yeast transformation by the LiAc/SS Carrier DNA/PEG method. *Methods Mol. Biol.* **313**, 107–120 (2006).

208. King, A. D., Pržulj, N. & Jurisica, I. Protein complex prediction with RNSC. *Methods Mol. Biol.* **804**, 297–312 (2012).

209. Suzuki, T. *et al.* Synthesis of 7-Acetyloxy-3,7-dimethyl-7,8-dihydro-6H-isochromene-6,8-dione and its Analogues. *J. Heterocycl. Chem.* **19**, 1409–1418 (2001).

Curriculum Vitae

Name:	Katharina Schmidt
Born:	07.12.1994

Apr. 2019 – Jan. 2024	**PhD student**, Organic Chemistry, Gottfried Wilhelm Leibniz University of Hannover, Germany
Oct. 2016 – Jan. 2019	**Master of Science**, Life Science, Gottfried Wilhelm Leibniz University of Hannover, Germany
Oct. 2013 – Oct. 2016	**Bachelor of Science**, Life Science, Gottfried Wilhelm Leibniz University of Hannover, Germany
Jun. 2013	**Abitur**, Ursulaschule Osnabrück, Germany

List of Publications

1. Schmidt, K. & Cox, R. J. Investigation of chain-length selection by the tenellin iterative highly-reducing polyketide synthase. *RSC Adv.* **14**, 8963–8970 (2024).