Understanding and Exploiting the Fungal Biosynthesis of Bisorbicillinoids and Terrein

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Lukas Kahlert, M. Sc.

2021

Abstract

key words: natural products, biosynthesis, polyketide, sorbicillinoids, terrein

The biosynthetic pathways towards the polyketides terrein **94** and (dimeric) sorbicillinoids, such as the Diels-Alder derived bisorbicillinol **151**, were investigated. Both pathways were reconstituted step by step in *Aspergillus oryzae* and the key biosynthetic transformations were adressed through a combination of *in vivo* and *in vitro* studies.

Previous work showed that terrein **94** is formed *via* an oxidative decarboxylation/ ring contraction sequence of the precursor 6-hydroxymellein **113**. The production of **113** was established in *Aspergillus oryzae* NSAR1 by co-expression of two genes encoding the non-reducing polyketide synthase (PKS) TerA and the PKS-like multidomain protein TerB. Deliberate point mutations of catalytic key residues within both proteins revealed that TerB is a novel type of of *trans*-acting enzyme. The catalytically active ketoreductase domain of TerB reduces the growing polyketide intermediate that is tethered to the PKS TerA; the catalytically inactive dehydratase domain of TerB appears to mediate protein-protein interactions with TerA.

The flavin-dependent monooxygenase (FMO) TerC was demonstrated to open the lactone ring of **113** *via* oxidative decarboxylation using a combination of heterologous expression in *A. oryzae* and *in vitro* assays. The chemically unstable aromatic product is hydroxylated by the second FMO TerD *in vitro*, but further reconstitution of the pathway was not successful. Heterologous expression of a homologous biosynthetic gene cluster (BGC) from a marine *Roussoella* sp. yielded a related aromatic intermediate. Although the ring contraction was not observed, the pivotal lactone opening during the biosynthesis of **94** and related metabolites was experimentally elucidated for the first time.

Co-expression of three genes from the sorbicillinoid BGC of *Trichoderma reesei* reconstituted the biosynthetic pathway up to the highly reactive sorbicillinol **148a** that was effectively reduced in the heterologous host *A. oryzae*. Dimeric sorbicillinoids were exclusively produced upon presence of a fourth gene encoding the FMO TrSorD. Further feeding studies confirmed that TrSorD is an unprecedented FMO that independently catalyzes both epoxidation and *intermolecular* dimerization of **148a** with itself or other suitable molecules, following a Diels-Alder or Michael-like mechanism. The non-homologous FMO PcSorD from *Penicillium chrysogenum* plays the same catalytic role as TrSorD, and furthermore oxidizes **148a** into oxosorbicillinol **173**. These results answer the long-standing question for the involvement of an *in vivo* catalyst during the dimerization of **148a**. Mutational studies indicated that covalent flavinylation is essential for enzyme activity of both TrSorD and PcSorD.

Zusammenfassung

Schlagwörter: Naturstoffe, Biosynthese, Polyketid, Sorbicillinoide, Terrein

Die Biosynthese der Polyketide Terrein **94** und (dimeren) Sorbicillinoide, wie dem Diels-Alder Derivat Bisorbicillinol **151**, wurde untersucht. Beide Pfade wurden schrittweise in *Aspergillus oryzae* rekonstituiert und die biosynthetischen Schlüsseltransformationen wurden durch Kombination von *in vivo* und *in vitro* Studien erforscht.

Vorherige Arbeiten zeigten, dass Terrein **94** durch eine Abfolge von oxidativer Decarboxylierung/ Ringkontraktion des Vorläufers 6-Hydroxymellein **113** gebildet wird. Die Produktion von **113** in *Aspergillus oryzae* NSAR1 wurde duch Co-Expression zweier Gene erreicht, die eine nicht-reduzierende Polyketidsynthase (PKS) TerA und ein PKS-ähnliches Multidomänen-Protein TerB kodieren. Gezielte Punktmutationen in beiden Proteinen zeigten, dass TerB ein neuartiges *trans*-agierendes Enzym ist. Die katalytisch aktive Ketoreduktase Domäne in TerB reduziert das wachsende Polyketidintermediate welches an der PKS TerA gebunden ist; die katalytisch inaktive Dehydratase Domäne scheint Protein-Protein Kontakt mit TerA zu vermitteln.

Mittels heterologer Expression in *A. oryzae* und *in vitro* Assays wurde gezeigt, dass die Flavin-abhängige Monooxygenase (FMO) TerC eine Lactonöffnung in **113** *via* oxidativer Decarboxylierung katalysiert. Das instabile aromatische Produkt wird ferner durch die zweite FMO TerD hydroxyliert, jedoch war eine weiterführende Rekonstitution des Biosynthesewegs nicht erfolgreich. Heterologe Expression eines homologen Biosynthese-Genclusters (BGC) lieferte ein verwandtes aromatisches Intermediat. Obwohl keine Ringkontraktion beobachtet wurde, wurde die zentrale Lactonöffnung während der Biosynthese von **94** und verwandten Naturstoffen zum ersten Mal experimentell aufgezeigt.

Co-Expression dreier Gene des Sorbicillinoid BGC aus *Trichoderma reesei* rekonstituierte den Biosyntheseweg bis zu dem hochreaktiven Intermediate Sorbicillinol **148a**, welches effektiv in dem heterologen Wirt *A. oryzae* reduziert wurde. Dimere Sorbicillinoide wurden ausschließlich in Anwesenheit eines vierten Genes, kodierend für die FMO TrSorD produziert. Fütterungsstudien bekräftigten dass TrSorD eine präzedenzlose FMO ist, die Epoxidation und *intermolekulare* Dimerisierung zwischen **148a** mit sich selbst oder anderen geeigneten Molekülen *via* Diels-Alder oder Michael-Addition katalysiert. Die nicht-homologe FMO PcSorD aus *Penicillium chrysogenum* übernimmt dieselbe Rolle wie TrSorD, und oxidiert ferner **148a** zu Oxosorbicillinol **173**. Diese Ergebnisse beantworten die offenstehende Frage, ob ein *in vivo* Katalysator an den Dimerisierungsreaktionen beteiligt ist. Mutationsstudien deuteten an, dass kovalente Flavinylierung für die enzymatische Aktivität von TrSorD und PcSorD essentiell ist.

Publications Included in this Cumulative Thesis

Partial results obtained in the course of this PhD thesis were published in advance in the following articles approved by the corresponding mentor.

1.   **L. Kahlert**, E. F. Bassiony, R. J. Cox, E. J. Skellam

       Diels-Alder Reactions During the Biosynthesis of Sorbicillinoids

       Angewandte Chemie International Edition, **2020**, *59*, 5816-5822

       (DOI: 10.1002/anie.201915486)

2.   **L. Kahlert**, R. J. Cox, E. Skellam

       The same but different: multiple functions of the fungal flavin dependent monooxygenase SorD from *Penicillium chrysogenum*

       Chemical Communications, **2020**, *56*, 10934-10937

       (DOI: 10.1039/d0cc03203d)

3.   **L. Kahlert**, M. Villanueva, R. J. Cox, E. J. Skellam

       Biosynthesis of 6-Hydroxymellein Requires a Collaborating Polyketide Synthase-like Enzyme

       Angewandte Chemie International Edition, **2021**, *60*, 11423-11429

       (DOI: 10.1002/anie.202100969)

4.   **L. Kahlert**, D. Bernardi, M. Hauser, L. P. Ióca, R. G. S. Berlinck, E. J. Skellam, R. J. Cox

       Early Oxidative Transformations During the Biosynthesis of Terrein and Related Natural Products

       Chemistry – A European Journal, **2021**, in press

       (DOI: 10.1002/chem202101447)

Acknowledgement

First of all I would like to thank Prof. Russell J. Cox for giving me the opportunity to continue with my PhD after I finished my master's thesis in his group. It is safe to say that the terrein project haunted me throughout the last four years. Yet, I always enjoyed working on such challenging biosynthetic questions and learned a lot throughout my time in your lab. I highly appreciate that you were always accessible, had an open ear whenever I needed your advice and that you never pressured me. I am grateful for all the discussions we had and your valuable input when I needed specialist help. I also acknowledge that you gave me a free hand to a large extent in pursuing my own scientific ideas. You were a fantastic mentor.

Of course none of this work would have been accomplished without the tremendous support by Dr. Elizabeth J. Skellam (Liz). You stood by my side since the very first day I started my master's thesis and you have always been there when I had a question or needed someone to cheer me up. I enjoyed all the discussions we had and I am very happy for you to pursue your own independent academic carrer now.

I must also give a huge compliment to all the technicians that keep our institute running, especially Katja (the good soul), Doreen, Birgit and Serge. I thank Monika, Dagmar and Dr. Jörg Fohrer for their excellent NMR service as well as Dr. Gerald Dräger and the HRMS team for doing their best to find the masses I was looking for.

It was a pleasure to work along with the powerhouses from lab 201: Carsten (nice shirt), Henrike (oh boi) and Chongqing ("Jürgen", I am sure you extracted even more tranformants than me). I also thank Raissa for giving me such a nice introduction into the group and all the former and current group members for helping me in my early months and later years, especially Eric, Maurice, Olli, Haili and Yunlong. Another honourable mention goes out to our coffee maschine that kept me alive through many days.

Last but not least I thank my family for all their support and inviting me for dinner when I once more was too lazy to cook by myself.

Abbreviations and Units

6-HM	6-hydroxymellein	IPTG	isopropyl-β-D-1-thiogalactopyranoside
6-MSA	6-methylsalicylic acid	k	kilo-
Å	angstrom	kan	kanamycin
aa	amino acid	KO	knockout
ACP	acyl carrier protein	KR	ketoreductase
APS	ammonium persulfate	KS	ketosynthase
AT	acyl transferase	L	litre
ATP	adenosine triphosphate	LCMS	liquid chromatography - mass spectrometry
BGC	biosynthetic gene cluster	m	milli-
BLAST	basic local alignment search tool	M	mega-
bp	base pairs	M	molar
cam	chloramphenicol	MeOH	methanol
carb	carbenicillin	min	minutes
CD	conserved domain	mRNA	messenger RNA
cDNA	complementary DNA	*m/z*	mass to charge ration
cf.	confer	μ	micro-
CFE	cell free extract	n	nano-
C-MeT	*C*-methyltransferase	NAD(P)H	nicotinamid adenine dinucleotide (phosphate)
CoA	coenzyme A		
COSY	correlation spectroscopy	NHEJ	non-homologous end joining
CRISPR	clusterred regulary interspaced short palindromic repeats	NMR	nuclear magnetic resonance
		nr	non-reducing
Da	Dalton	NRP(S)	non-ribosomal peptide (synthetase)
DA	Diels-Alder	PAGE	polyacrylamide gel electrophoresis
DH	dehydratase	PAM	protospacer adjacent motif
DMAPP	dimethylallyl-pyrophosphate	PCR	polymerase chain reaction
DMF	dimethyl formamide	PEG	polyethylene glycol
DNA	deoxyribonucleic acid	pH	*potentia hydrogenii*
EAS	electrophilic aromatic substitution	PKS	polyketide synthase
e.g.	*exempli gratia*; for example	ppm	parts per million
ER	enoyl reductase	PPTase	phospopenthetheinyl transferase
ES	electro-spray	pr	partially reducing
EtOAc	ethyl acetate	PT	product template
FAD	flavin adenine dinucleotide	R	reductive release domain
(m)FAS	(mammalian) fatty acid synthase	RNA	ribonucleic acid
FMN	flavin mononucleotide	rpm	revolutions per minute
FMO	flavin-dependent monooxygenase	RT	room temperature
FPP	farnesyl-pyrophosphate	RT-PCR	reverse transcriptase-PCR
gDNA	genomic DNA	s	seconds
gRNA	guide RNA	SAM	*S*-adenosyl methionine
GOI	gene of interest	SAT	starter unit-ACP transacylase
h	hours	SDR	short-chain dehydrogenase/ reductase
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid	SDS	sodium dodecyl sulphate
		sp.	species
HMBC	heteronuclear multiple bond correlation	TE	thiolesterase
HPLC	high-performance liquid chromatography	TEMED	N,N,N′,N′-Tetramethylethane-1,2-diamine
hr	highly reducing	TH	thiolester hydrolase
HRMS	high-resolution mass spectrometry	THF	tetrahydrofuran
HSQC	heteronuclear single quantum correlation	TLC	thin layer chromatography
hyg	hygromycin	Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
Hz	hertz		
i.e.	*id est*; that is	UV	ultraviolet
IMAC	immobilized metal-ion affinity chromatography	*v/v*	volume per volume
		wt	wild type
IPP	isopentenyl-pyrophosphate	*w/v*	weight per volume

Table of Contents

1 Introduction

1.1 Fungal Secondary Metabolites

Filamentous fungi are well known for the production of secondary metabolites, also referred to as natural products. Unlike primary metabolites, these structurally heterogenous small molecules are not directly required for the growth their producers, but rather grant a competitive advantage in their specific micro-ecosystem.[1,2] Secondary metabolites do not only repel or kill environmental competitors (*e.g.* bacteria or protozoa), but also act as signalling molecules that modulate interaction between the producers and their immediate surroundings.[3]

Towards humankind natural products are curse and blessing at the same time due to their "Janus faced"[4] biological activities that may either promise pharmaceutical applications or threaten animal and human health due to their toxic properties. One of the most infamous mycotoxins is aflatoxin B1 **1**, a food contaminant produced by *Aspergillus* species that is responsible for a significant loss of crops each year and that possesses very high carcinogenic potential upon ingestion.[5,6] In contrast, molecules like the famous penicillin G **2**, whose serendipitous discovery almost one century ago[7] initiated the age of antibiotics, are indispensable for the treatment of bacterial infections. In fact, the rational use of penicillin G **2** during World War II reduced the mortality from infected wounds tremendously and greatly promoted the post war economy.[8] Between 1981 and 2010 around two thirds of all small-molecule approved drugs originated from natural products, further illustrating their enormous pharmaceutical potential.[9]

Fungal secondary metabolites can be classified either based on their function (*e. g.* antibacterial, anti-inflammatory, phytotoxic) or much more usefully based on their biosynthetic origin, since all secondary metabolites are based on precursor molecules that derive from the primary metabolism. Therefore, secondary metabolites are divided into four main classes: polyketides, non-ribosomal peptides (NRP), terpenes and alkaloids.[10] Polyketides, the most abundant class of fungal natural products, are biosynthezised by polyketide synthases (PKS) and are derived from acetate **3** or other small carboxylic acid building blocks.[11] Both the previously mentioned aflatoxin B1 **1** and the hypercholesterolemic agent lovastatin **4** are among the most well-known representatives of fungal polyketides.[12]

The structural backbones of non-ribosomal peptides like penicillin G **2** or the mycotoxin gliotoxin **5** are assembled from amino acid precursors **6** by non-ribosomal peptide synthetases (NRPS) and often modified by accessory enzymes to yield the final product. [13,14]

The two isomers dimethylallyl-pyrophosphate **7** (DMAPP) and isopentenyl-pyrophosphate **8** (IPP) are the building blocks for the assembly of linear polyprenyl-diphosphates of different chain length that are subsequently cyclised by terpene cyclases resulting in a manifold of structurally distinct terpenes such as the bicyclic aristolochene **9** (C_{15}-backbone) or the tricyclic brassicicene C **10** whose C_{20}-backbone is further modified by oxidative enzymes.[10,15–17]

In contrast to the three previous classes, alkaloids do not share a common biosynthetic background but are characterized by a nitrogen-containing heterocyclic ring that is usually derived from an amino acid building block.[18] They might be as simple in structure as the psychoactive alkaloid psilocybin **11** or as complex as the cytotoxic communesin A **12**.[19,20]
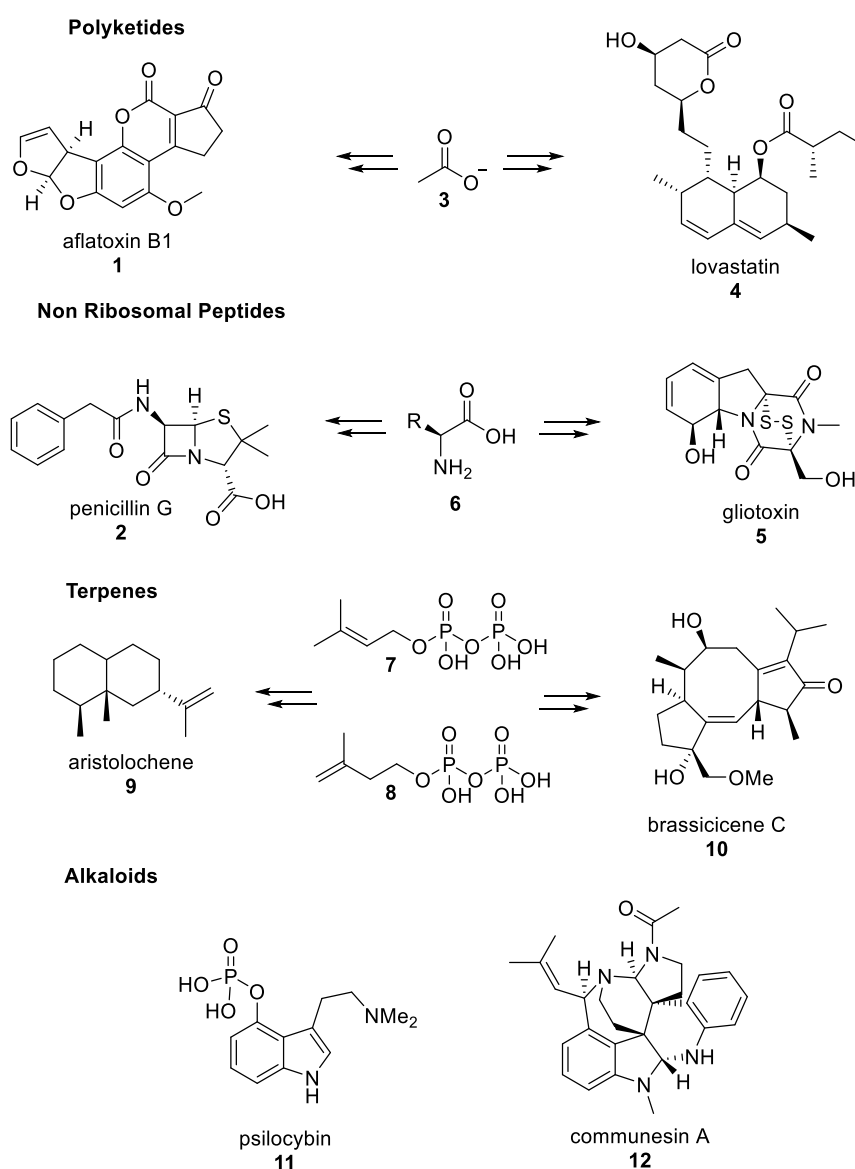

Figure 1.1.1: The four main classes of secondary metabolites including exemplary compounds and common biosynthetic building blocks.

1.2 Polyketide Biosynthesis in Filamentous Fungi

1.2.1 Biosynthetic Gene Clusters in the Era of Genomics

Following the discovery of penicillin G **2**, the isolation of new natural products was primarily driven by their potential utility as antibiotics. Although most of the antibiotic classes in use for medical treatment until today were discovered during the 'golden era' between the 1930's and 1950's, extensive research towards the identification of novel natural products was and is still being conducted.[21,22] Much effort has gone into natural product discovery and as much labour has been spent on biosynthetic investigations to understand how these small molecules are assembled. Early experiments mainly relied on feeding studies with ^{14}C-labelled precursors whose incorporation/ modification was tracked by chemical degradation (radiotracing).[23,24] Based on this approach Birch *et al.* first demonstrated in 1955 that polyketides are derived from head-to-tail linkage of acetate units.[25] With the introduction and refinement of Nuclear Magnetic Resonance Spectroscopy (NMR) techniques radioactive labels have been replaced with stable isotopic labels (^{13}C, ^{18}O, ^{15}N, ^{2}H). However, it was ultimately the emergence of recombinant deoxyribonucleic acid (DNA) methodology that enabled the biosynthetic characterization of natural products on a genetic level.[10]

Despite earlier beliefs, and in contrast to the majority of plant secondary metabolite genes,[26] genes that are involved in the production of a distinct secondary metabolite are usually clustered on a single genetic locus in filamentous fungi.[1,27–30] Although the origin for this phenomenon has not yet been finally clarified, hypotheses include horizontal gene transfer and gene rearrangements as means for explanation.[31,32] Gene clustering may provide the advantage of transcriptional co-regulation and is presumably maintained by evolutionary pressure that disfavours the dispersion of biosynthetically cohesive genes throughout the genome.[33,34]

The presence of biosynthetic gene clusters (BGC) also offers researchers a very beneficial advantage as it allows predictions about potential pathway end products and intermediates based on the contained genes. The cost and time scale of DNA sequencing has dropped significantly during the last years, rendering sequencing affordable for individual researchers while next generation technologies facilitate high-quality throughput of whole genomes.[35] Bioinformatic software programs like antiSMASH (antibiotics & Secondary Metabolite Analysis Shell)[36,37] analyse the query DNA sequence towards the presence of potential BGC and identify associated genes based on conserved motifs in the corresponding protein sequences. Further refinements of gene boundary predictions or intron patterns based on user-defined constraints may be performed by additional programs such as AUGUSTUS[38] or FGENESH.[39] These algorithms represent a crucial tool for both the

discovery of new natural products and their allocation to a certain BGC. Collaborative efforts between researchers worldwide promote the gathering and exchange of BGC information and have led to projects like the MIBiG database (Minimum Information about a Biosynthetic Gene cluster) that pools reference BGC and links them to known natural products.[40]

During the 21st century genome mining has become a key technology for the discovery and analysis of BGC and novel bioinformatic tools are steadily developed to exploit natural products diversity and support biosynthetic research.[41]

1.2.2 Chemistry and Architecture of FAS and PKS

Despite their enormous structural variety all polyketides are assembled by one distinct class of multifunctional enzymes, the polyketide synthases.[29]

The biosyntheses of polyketides and fatty acids are very closely related as realized by Birch in 1953, two years before providing the experimental evidence that polyketides are built up by formal condensation of acetate molecules.[25,42] Yet, the first proposal to explain the formation of fatty acids and 'a very large number of interesting compounds' (as the term polyketides was not established at that time) by the formation of a key polyketone intermediate **13** was already published by Collie in 1907 based on his experiment to explain the formation of orcinol **14** from dimethylpyrone **15** by treatment with a barium hydroxide in 1893.[43,44] This synthetically generated polyketone **13** resembles the poly-β-keto thiolester intermediate **16**, now known to be involved during polyketide/ fatty acid biosynthesis (Scheme 1.2.1).[29]
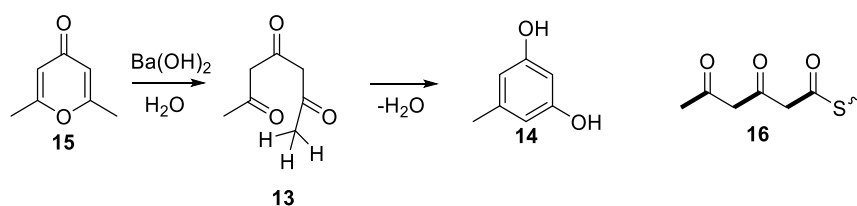
Scheme 1.2.1: Synthesis of orcinol **14** from dimethylpyrone **15** *via* polyketone-intermediate **13** resembling the poly-β-keto-thiolester intermediate **16** during polyketide/fatty acid biosynthesis. Bold lines in **16** indicate acetate units.

Fatty Acid Synthases

Fatty acids are assembled by repetitive decarboxylative Claisen reactions between an acetate **3** starter and malonate **17** extender unit. All steps are catalyzed by a multidomain fatty acid synthase (FAS).[11]

In order to be catalytically active the acyl carrier protein (ACP) domain of the FAS has to be converted from its *apo*-form into its *holo*-form. This reaction is catalysed by phosphopantetheinyl transferases (PPTases) that transfer the phosphopantetheinyl-moiety of Coenzyme A (CoA) **18** onto a conserved serine residue within the ACP (Scheme 1.2.2 A).[45,46] The terminal thiol group is therefore not part of the native FAS, but derived from posttranslational modification.

Both **3** and **17** are activated into their corresponding CoA-thiolesters **3b** and **17b** prior to chain extension (cf. Scheme 1.2.2). While **3b** is the central C_2-unit derived from primary metabolism, **17b** is made from **3b** through formal carboxylation catalysed by the acetyl-CoA carboxylase.[47]
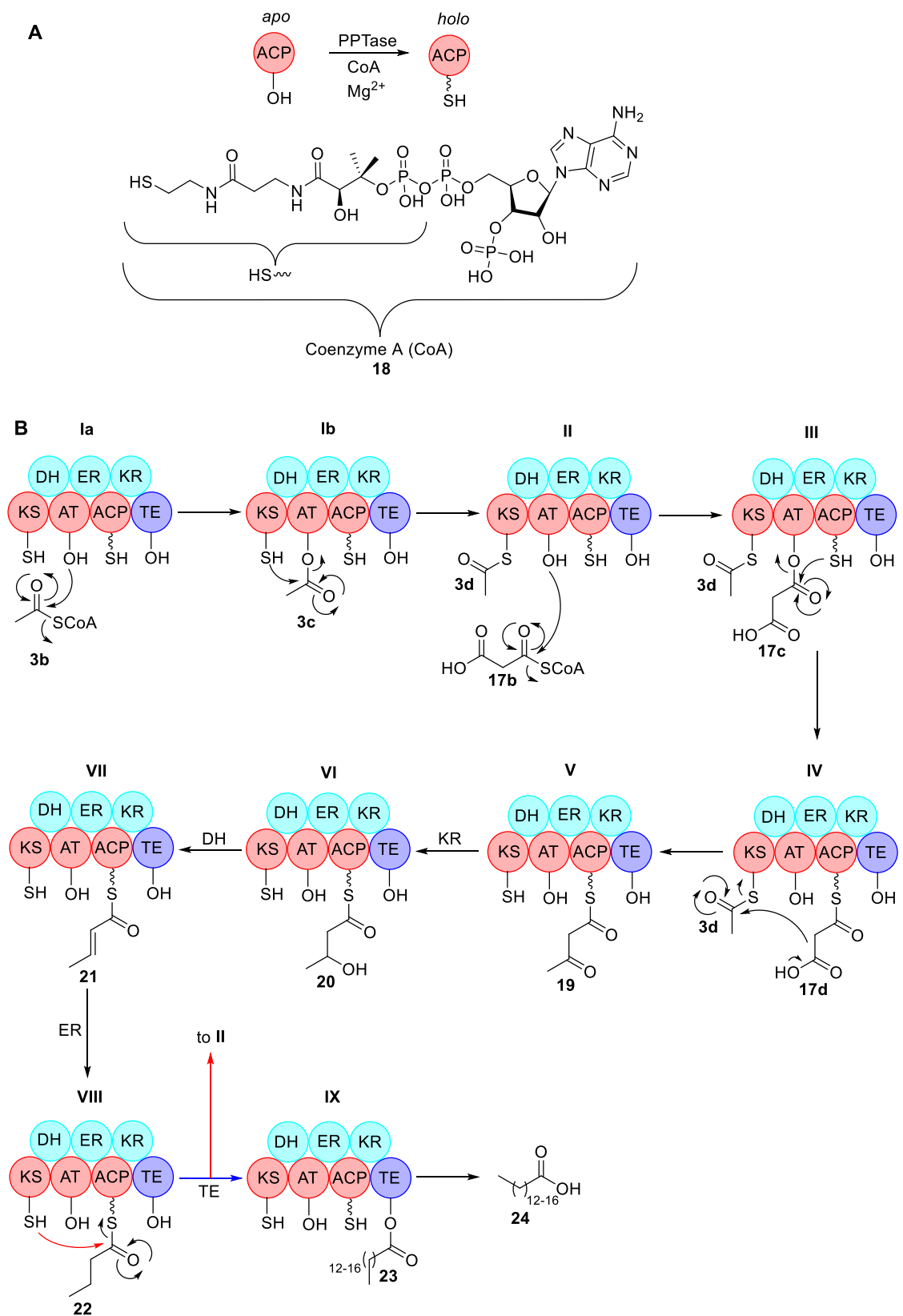
Fatty acid biosynthesis is initiated by shuttling of the starter acyl-unit **3b/3c** onto a conserved cysteine residue within the ketosynthase (KS) domain forming the first thiolester-species **3d**, a process that is mediated by the acyl transferase (AT) domain (Scheme 1.2.2 step Ia and Ib). The AT subsequently loads the first malonyl-CoA extender unit **17b** and shuttles it onto the *holo*-ACP (step II and III).[11] The following key biosynthetic step establishes a new carbon-carbon bond between the acyl-thiolester **3d** tethered to KS and the malonyl-thiolester **17d** bound to ACP by decarboxylative Claisen-condensation (step IV).[48] The linear β-keto-thiolester intermediate **19**, elongated by one C2-unit, remains covalently bound to the ACP domain (step V) and is set up for further β-processing by the ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains.

The flexibility of the phosphopantetheinyl linker allows dynamic movement of the ACP which is essential to facilitate interactions with both the β-processing domains (KR, DH, ER) and the domains involved the chain elongation (KS, AT) and -termination (thiolesterase, TE).[46,48]

Initial NADPH-dependent reduction of the β-carbonyl **19** by the KR-domain yields the secondary alcohol **20** (step VI) that is subsequently eliminated by the DH-domain to give the (*E*)-alkene moiety in **21** (step VII). As will be outlined later ketoreduction during polyketide biosynthesis is a stereoselective process that is controlled by the respective KR-domain. Ultimately the alkene is reduced by the NADPH-dependent ER domain to yield the saturated thiolester **22** (step VIII). Alkanoyl-intermediate **22** is then transferred back to the KS domain and the cycle continues in an iterative fashion until eventually the carbon chain reaches a distinct length (usually 14, 16 or 18 carbon atoms) and is released as the fatty acid **24** by the TE-domain.[11]

Based on structure and domain organization FAS are subdivided into two types: Type I FAS constitute large multifunctional megasynthases that combine all catalytic domains required for fatty acid biosynthesis; Type II FAS consist of complexes formed by individual enzymes.

Scheme 1.2.2: Overview of fatty acid biosynthesis catalyzed by FAS. **A**, Conversion of *apo*-ACP into *holo*-ACP; **B**, Mechanism of fatty acid biosynthesis.

While the type I fungal FAS exhibits a 4.6 MDa heterododecameric structure consisting of two polypeptide chains ($\alpha_6\beta_6$),[49] the type I mammalian FAS (mFAS) shows a catalytically active homodimeric structure composed of two 270 kDa subunits that each comprise all seven domains. The crystal structure of mFAS reveals an intricate architecture with a distinct spatial partitioning between condensing- and β-processing domains that further show structural remnants of a *C*-methyltransferase (ψC-MeT) domain at the periphery (Figure 1.2.1).[50] Due to its inherent flexibility the ACP and adjacent TE domain are not resolved in the crystal structure.

Type II FAS found in bacteria and plants are composed of dissociated, individual proteins that are each encoded by a discrete gene.[51] While structures of individual components of Type II FAS are known,[52,53] there is no structure of a type II FAS complex.
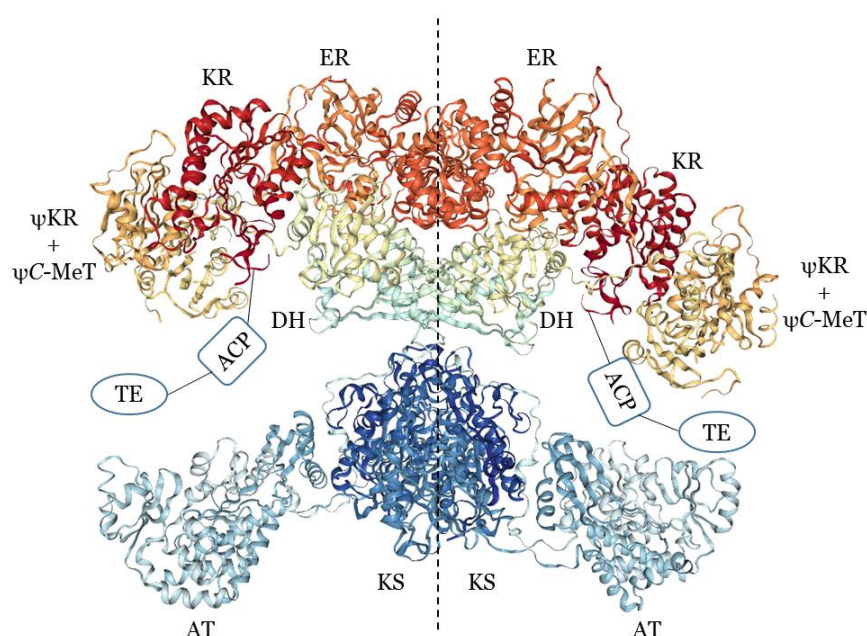

Figure 1.2.1: Crystal structure of the homodimeric mFAS resolved at 3.2 Å (PDB: 2VZ8).[50] ACP and TE domain not resolved in the crystal structure are indicated by blue boxes. Image created with NGL viewer.[54]

Polyketide Synthases

In addition to following the same basic chemistry, PKS and FAS also show related structural features.[55] Accordingly the majority of PKS are subdivided into type I PKS (multidomain synthases) and type II PKS (individual stand-alone proteins).[29,55] The much less frequent type III PKS, that are often found in plants, use a single domain to catalyze chain elongation in an ACP-independent process acting on free CoA-thiolester substrates.[56]

The majority of fungal PKS belong to the iterative type I subfamily in which (like FAS) the same set of domains is reused during each cycle to assemble the polyketide chain *via*

consecutive decarboxylative Claisen-condensations.[57] The difference to fatty acid biosynthesis is that fungal PKS are programmed to exhibit a distinct extent of β-processing during each iteration, resulting in additional keto-, hydroxyl- and enoyl-functionalities along the polyketide backbone.

PKS are furthermore able to accept alternative starter and extender units other than acetyl-CoA **3b** and malonyl-CoA **17b** and control the degree of chain methylation premised a functional *C*-MeT domain is included in the PKS.[29,58,59] This domain catalyzes incorporation of α-methyl branches into the polyketide chain using the cofactor *S*-adenosyl methionin (SAM; section 1.2.3).[29,60]

Fungal iterative type I PKS are furthermore classified into non-reducing (nr), partially reducing (pr) and highly reducing (hr) reflecting the presence/absence of particular β-processing domains.[29]

nrPKS

Non-reducing PKS possess the core domains essential for chain-extension (KS, AT, ACP) but do not possess any β-processing domains (KR, DH, ER). However, nrPKS usually feature additional starter unit–ACP transacylase (SAT), product template (PT) and releasing domains.[57] The SAT domain loads the starter unit acetyl-CoA **3b** or any other acyl-derived starter onto the ACP, whereas the AT-domain only transfers malonyl-CoA **17b** for chain elongation. Alternative starter units are often supplied by a partnering hrPKS as for example during the biosynthesis monocillin II **25** (Scheme 1.2.3 A): the hrPKS CcRADS1 produces the pentaketide intermediate **26** that is subsequently transferred onto the SAT domain of the nrPKS CcRADS2 and elongated to yield the nonaketide intermediate **27**.[61]

The role of the PT domain is to mediate the regioselective cyclization of the highly reactive poly-β-keto-intermediate tethered to the PKS resulting in formation of an aromatic ring system.[62,63]

The releasing domain ultimately catalyzes release of the PKS-bound intermediate, often by hydrolysis or (trans)-esterification.[64] Alternatively, reductive (R) domains use NADPH to release the products as aldehydes like the product of thr nrPKS TropA, 3-methylorcinaldehyde **28** (Scheme 1.2.3 B).[65]

prPKS

Very few fungal partially reducing PKS have been attributed with a final polyketide product. One important is that of MSA-synthase (MSAS), 6-methylsalicylic acid **30** (6-MSA).[29,66] The architecture shows lack of the ER domain and the previously assigned DH domain has been

shown to actually function as a thiolester hydrolase (TH) domain that is involved in product release of the ACP-tethered intermediate 31 (Scheme 1.2.3 C).[67,68]
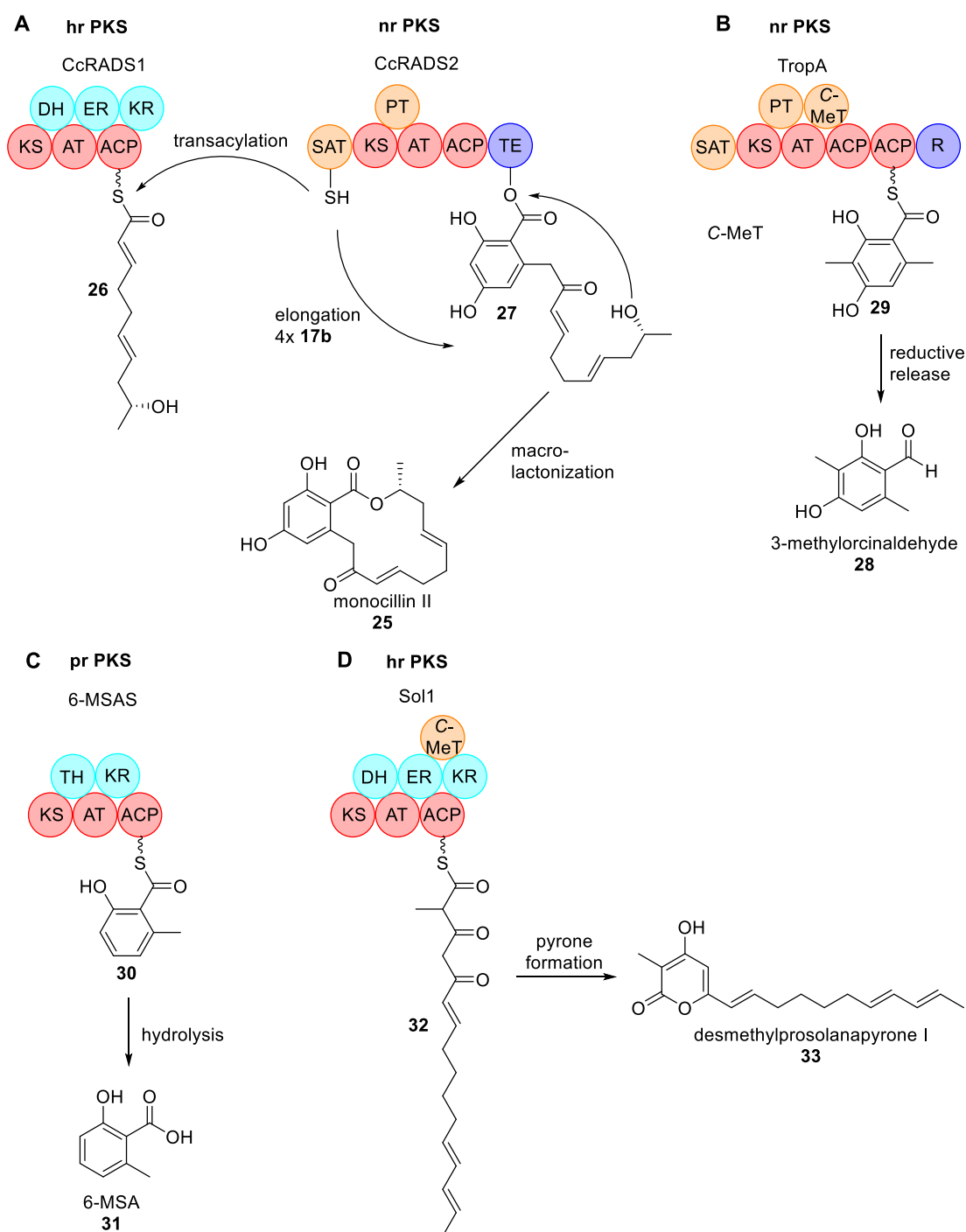
Scheme 1.2.3: Examples of domain architecture and polyketides produced by fungal iterative type I PKS: **A**, collaborating hrPKS-nrPKS; **B**, nrPKS; **C**, prPKS; **D**, hrKS. Each example **A-D** displays a different release mechanism.

hrPKS

The class of highly reducing PKS possesses the complete repertoire of β-processing domains (KR, DH, ER). However, unlike FAS, not every of those β-processing domains is active during

each iteration, resulting in a distinct degree of saturation along the polyketide-backbone.[57,69] Additional *C*-Met domains are frequently found among hrPKS.

The hrPKS Sol1 from *Alternaria solani* shows the typical domain architecture that furthermore lacks any releasing domain. The polyketide intermediate is released from the enzyme *via* intramolecular lactonization of the ACP-tethered intermediate **32** to yield the octaketide product desmethylprosolanapyrone I **33** (Scheme 1.2.3 D).[70,71]

Recently, the structure of the nrPKS LovB that is involved in the biosynthesis of lovastatin **4** was resolved at 2.91 Å using cryo-electron microscopy.[72] As anticipated for fungal iterative type I PKS, LovB forms a homodimeric architecture that closely resembles the structure of mFAS (cf. Fig. 1.2.1).

In contrast to modular PKS,[73–75] the molecular factors that determine stereoselectivity of a fungal iterative hrPKS ketoreductase domain are much less understood.[76,77] Since fungal iterative type I hrPKS only possess a single KR one would expect that all the β-ketoreductions proceed with the same relative stereochemistry. Yet, Tang and Vederas reported a fungal KR-domain of the hrPKS Hpm3 that displays a substrate dependent stereoselectivity.[77] The authors concluded that this trait is modulated by a $\beta_5\alpha_5\alpha_6$ motif in that the catalytic site is located. However, the general applicability of this finding remains to be investigated; in particular whether the $\beta_5\alpha_5\alpha_6$ motif serves as a determinant for stereoselectivity of KR-domains in all fungal iterative type I hrPKS.

Trans-acting Proteins

In fungal polyketide biosynthesis several examples for *trans*-acting proteins have been reported. *Trans*-acting proteins are external catalysts that are not part of the core PKS, but physically interact with the PKS during assembly. Such proteins include *trans*-ER enzymes that substitute broken ER catalytic domains in the core PKS,[78,79] *trans*-release enzymes that are necessary for chain length determination and product release[80] as well as *trans*-*C*-MeT enzymes that introduce α-methyl branches to the PKS-tethered intermediate.[81,82] Due to a lack of crystal structures the precise interaction points and involved amino acid residues are often yet to be determined.

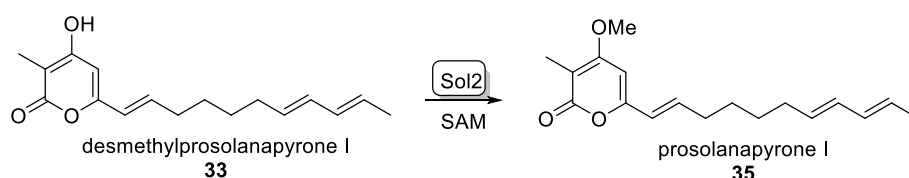### 1.2.3 Polyketide Biosynthesis: Post-Assembly Modifications

After initial release from the PKS the polyketide product is frequently modified by ancillary enzymes, a process referred to as *tailoring*.[83] In fact, the immense structural variety of polyketides and their scope of biological activities is largely derived from these post-assembly tailoring steps rather then from the PKS themselves.[84]
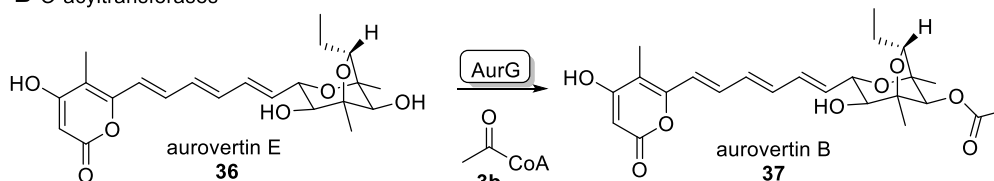
Group-transferases are enzymes that decorate the target molecules with additional carbon units, often *via* nucleophilic addition. *O*-methylation is frequently catalyzed by SAM **34**-dependent *O*-methyltransferases, as for example during the conversion of **33** into prosolanapyrone I **35** (Scheme 1.2.4 A).[70,71] *O*-acyltransferases attach longer carbon chains to the polyketide backbone that might be tetraketide derived[85] or as short as acetate that is used to convert aurovertin E **36** into aurovertin B **37** (Scheme 1.2.4 B).[86]

Meroterpenoids are structurally often highly complex natural products that share subunits of both polyketide and terpene origin.[87] The terpene unit is connected with the polyketide moiety by prenyltransferases that, for instance, transfer farnesyl-pyrophosphate **38** (FPP) onto the polyketide 3,5-dimethylorsellinic acid **39** (DMOA) resulting in the initial meroterpenoid intermediate **40** (Scheme 1.2.4 C).[88]

A *O*-methyltransferases

desmethylprosolanapyrone I
33

prosolanapyrone I
35

B *O*-acyltransferases

aurovertin E
36

aurovertin B
37

C prenyltransferases

DMOA
39

FPP
38

farnesyl-DMOA
40

S-adenosyl-methionine (SAM)
34

**Scheme 1.2.4**: Overview of tailoring group-transferases. **A**, *O*-methyltransferases; **B**, *O*-acyltransferases; **C**, prenyltransferases.

Oxidoreductases are by far the most abundant group of tailoring enzymes. They are further divided into oxygenases, oxidases, peroxidases, reductases and dehydrogenases based on their catalytic activity. Their general function is the incorporation of oxygen into the substrate (*e.g.* hydroxylation or epoxidation) as well as the interconversion of oxygen-containing groups by redox-reactions (*e.g.* oxidation of a secondary hydroxyl-group into a ketone).[89] Since these reactions frequently introduce new stereocenters and affect the physico-chemical properties of the substrate, oxidoreductases potentially have a tremendous impact on the biological activity by altering the molecule-ligand binding affinity of a distinct molecule.[83]

Short chain dehydrogenases/ reductases (SDR) constitute a large family of NAD(P)H **41/42** dependent oxidoreductases that are often involved in polyketide tailoring.[90] The trichoxide BGC encodes five SDR proteins, that all catalyze regiospecific keto- or aldehyde reductions as depicted for the conversion of 5-deoxyaurocitrin **45** into virensol B **46** (Scheme 1.2.5).[91]
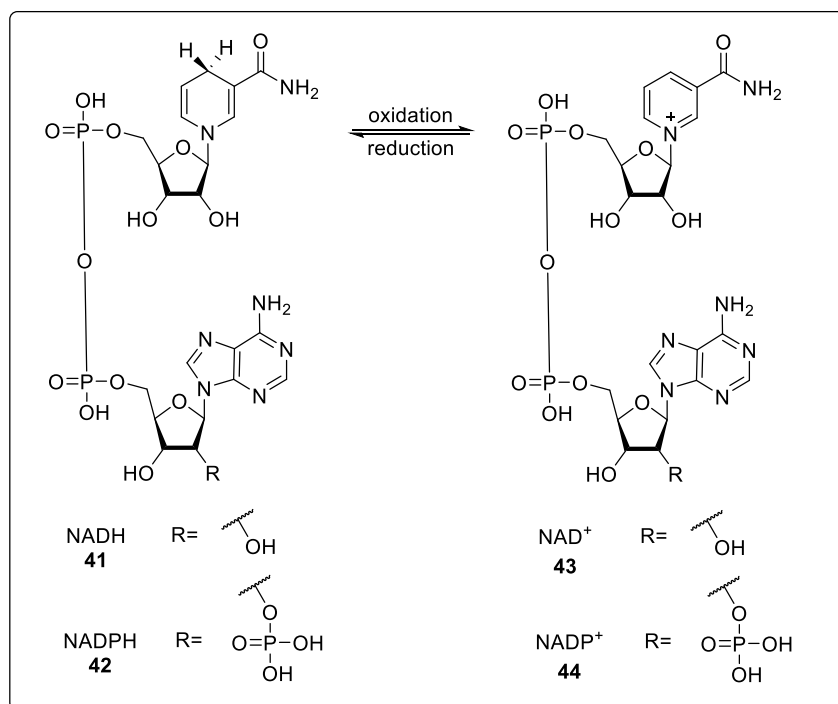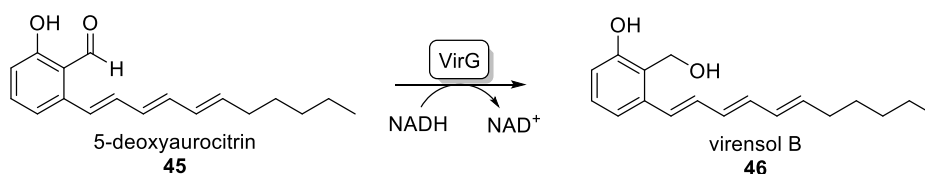
Scheme 1.2.5: Exemplary reduction catalysed by an NADH-dependent SDR tailoring enzyme during polyketide biosynthesis. The structures of involved cofactors are shown below.

Oxygenases are able to reductively activate molecular dioxygen and catalyse its subsequent regio- and stereoselective insertion into suitable substrates.[83] Many of these reactions are

unprecedented using classical synthetic methods and hence a lot of research has been conducted to understand the underlying mechanism of these valuable biocatalysts.[92,93]

Oxygenases are divided into monooxygenases that catalyse the insertion of one oxygen atom into the substrate while the other is reduced to H_2O, and dioxygenases where the second oxygen atom is not converted into H_2O.[83,93,94] Although the oxidation of hydrocarbons by molecular oxygen is an exothermic process, dioxygen containing two unpaired electrons is spin-forbidden to react with spin paired substrate molecules in the absence of a suitable catalyst. Oxygenases therefore use transition metal ions (*e.g.* iron or copper) or organic cofactors for means of dioxygen activation. Furthermore, uncontrolled reduction of molecular oxygen would often generate highly reactive radicals that react non-specifically with cellular components.[95] Hence, in addition to reductive activation of dioxygen, the respective biocatalysts also control the fate of activated oxygen intermediates by highly selective target oxidation to prevent toxic side reactions.[94] The most prominent types of oxygenases will be discussed below.[93,96]
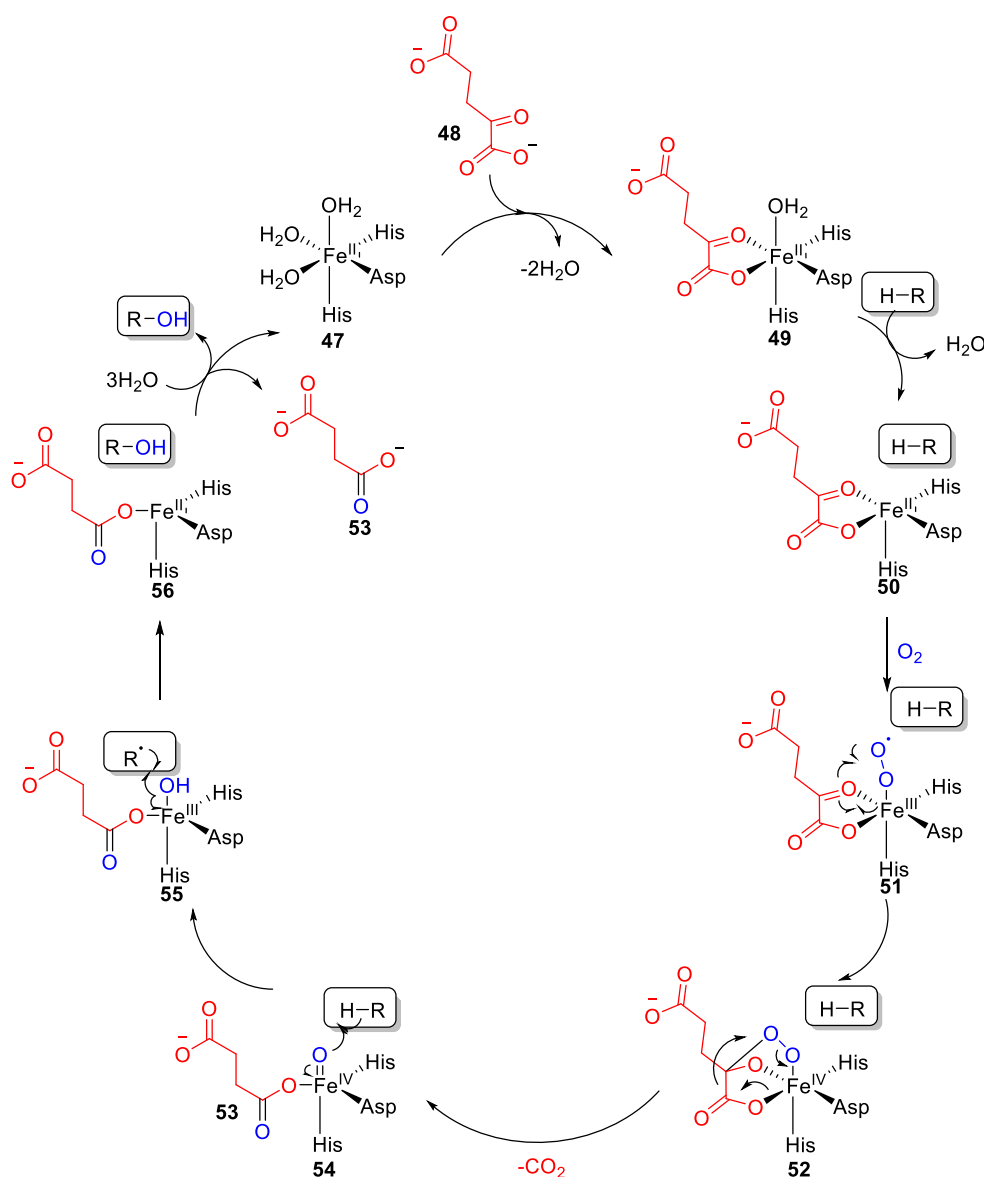
Non-heme-Fe(II)-α-ketoglutarate dependent dioxygenases

Non-heme-Fe(II)-α-ketoglutarate dependent dioxygenases feature a central Fe(II) that is coordinated by two histidine and one aspartate residue (facial triad) for activation of dioxygen. These enzymes catalyse a wide array of oxidative transformations on unactivated substrate molecules including hydroxylation, epoxidation and desaturation.[96,97] Halogenation might also be catalysed, characterized by an unusual exchange of aspartate with the respective halide in the facial triad.[98]

Substrate hydroxylation is the best studied mechanism and is initiated by the Fe(II) species **47** that is coordinated by the facial triad and three additional water molecules (Scheme 1.2.6). Binding of the organic co-substrate α-ketoglutarate **48** replaces two of those water molecules and the third is subsequently displaced by the substrate hydrocarbon entering the active site to successively yield the Fe(II) species **49** and **50**, respectively. This step furthermore opens a binding site for molecular oxygen that creates the Fe(III)-peroxyl radical **51**.

The distal oxygen atom of the Fe(III)-peroxyl radical **51** then forms a bicyclic peroxo-hemiketal intermediate **52** with **48**, that *via* oxidative decarboxylation decomposes into succinate **53** bound to the central Fe(IV)-oxo species **54**, also denoted as the ferryl intermediate.[99] The high valence ferryl intermediate can abstract a hydrogen atom from the substrate with concomitant formation of the Fe(III)-OH species **55** and a substrate radical. Radical rebound from the hydroxyl-radical onto the substrate radical yields the now hydroxylated substrate molecule and returns the central iron ion into its Fe(II) oxidation

state **56**. In the last step both product (R-OH) and succinate **53** are released and the coordination sites are re-occupied with water molecules.[96]

Scheme 1.2.6: Catalytic mechanism of non-heme-Fe(II)-α-ketoglutarate dependent dioxygenases depicting hydroxylation of an unactivated hydrocarbon substrate.

Cytochrome P450 monooxygenases

Another type of oxygenases that rely on a central metal ion to activate dioxygen are cytochrome P450 monooxygenases (P450). In eukaryotes these enzymes are commonly membrane bound. The Fe(III) within the essential heme prosthetic group **57** (the iron-protoporphyrin IX complex) is coordinated to a conserved cysteine of the protein backbone.[100,101] Among others, the scope of reactions catalysed by P450s include highly regio- and stereoselective hydroxylation of unactivated hydrocarbons and aromatic

substrates, olefin epoxidation, aldehyde- and heteroatom oxidation as well as *N*-or *O*-dealkylation.[101]

The catalytic cycle is triggered as the substrate molecule enters the active site of the P450, replacing an axial water molecule in **58** (Scheme 1.2.7). Single electron transfer from a reductase partner (ultimately the electrons are derived from NAD[P]H **41/42**) reduces Fe(III) **59** to Fe(II) **60** and subsequent reaction with dioxygen forms the Fe(III)-peroxyl radical **61**. A second single electron transfer yields the peroxide anion **62** that is protonated to give the electrophilic iron-hydroperoxide **63**.[101]

Protonation of the distal oxygen atom in **63** induces heterolytic cleavage resulting in release of water and formation of the high valence Fe(V)-oxo species **64** that is a strong oxidant and might catalyse hydroxylation (**65**) or epoxidation (**66**) of the substrate molecule following a radical rebound. Finally, the oxygenated product is released and replaced by a water molecule to set the Fe(III) species **58** for another catalytic cycle.[96,100,101]
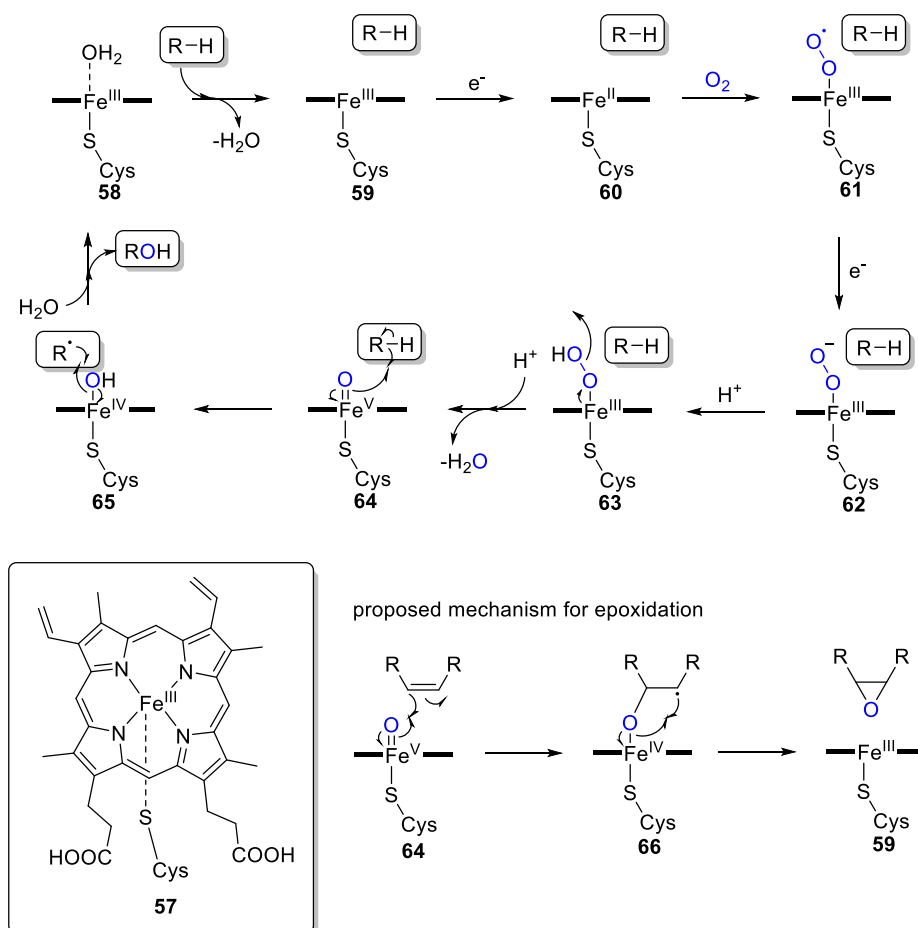

Scheme 1.2.7: Catalytic mechanism of cytochrome P450 monoxygenases depicting activation of dioxygen and incorporation into the substrate by either hydroxylation or epoxidation. Horizontal bold lines indicate the organic protoporphyrin IX moiety group as shown in **57**.

15

Flavin-dependent monooxygenases

Flavin-dependent monooxygenases (FMO) employ a purely organic cofactor to activate molecular oxygen. The cofactor is either flavin adenine dinucleotide **67** (FAD) or flavin mononucleotide **68** (FMN) in its corresponding reduced hydroquinone form FADH⁻ **69** or FMNH⁻ **70**, respectively (Scheme 1.2.8).[94] Throughout the scientific literature there is some inconsistency regarding the protonation state of the reduced flavin cofactor that is either displayed in the anionic form with the negative charge delocalized between N-1 and C-2 (*e.g.* FADH⁻ **69**) or in the neutral form with protonation at N-1.[94,102–105]

Generally, the physico-chemical properties (*e.g.* pK_a) of the flavin cofactor depend on the chemical- and host protein environment.[106] The anionic form of the reduced flavin has been experimentally deduced for several FMOs.[107–113] For instance, Moonen *et al.* reconstituted the *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* with [¹⁵N]-enriched FAD **67** *in vitro*. ¹⁵N-NMR analysis of the flavin cofactor after reduction revealed a negative charge at N-1 (the most acidic position in **69**) at pH 6-8 indicated by a chemical shift of 180 ppm, distinguishable from the neutral N-1 wich resonates at 128 ppm.[107] Therefore, the anionic form of the reduced flavin will the displayed throughout this section.
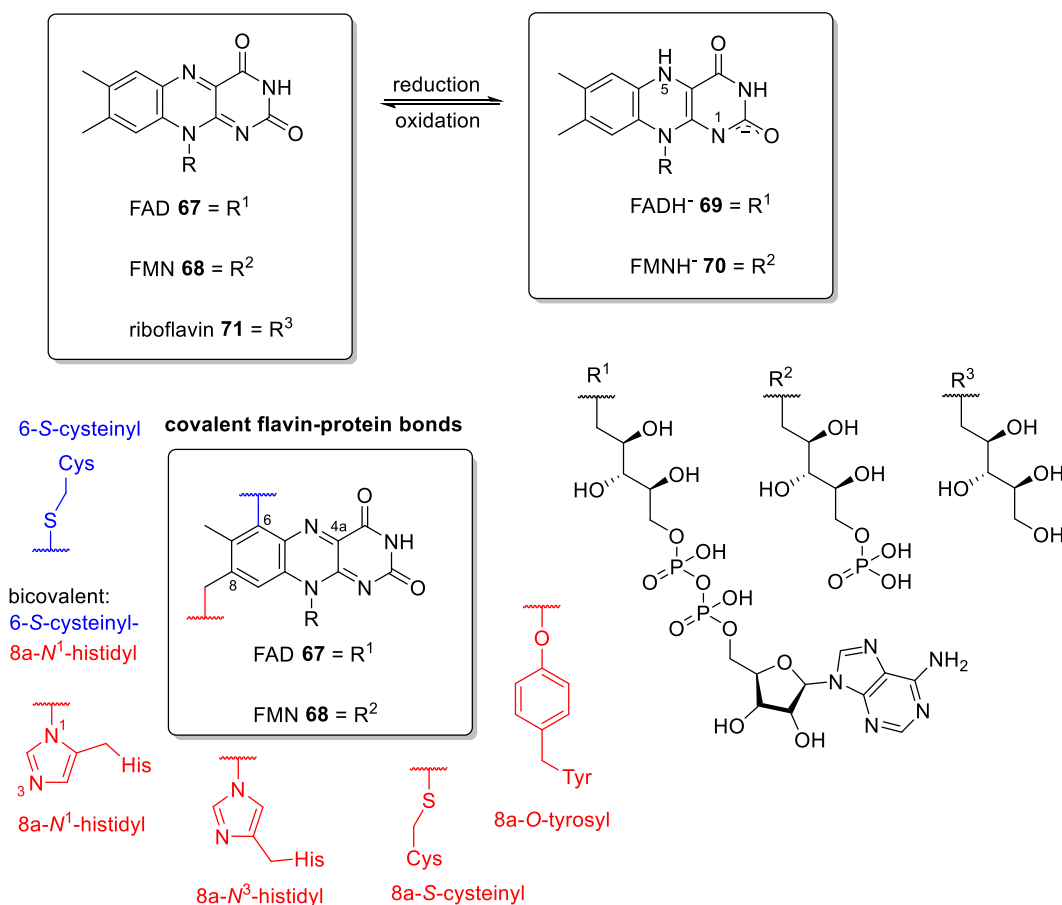
Scheme 1.2.8: Structure of flavin cofactors and amino acid residues involved in mono-or bicovalent bonding.

Due to the absorption maximum (375 and 450 nm) of the catalytically active isoalloxazine moiety of **67** and **68**, FMOs display a yellow colouration.[103] In fact, the term flavoprotein is derived from the latin word for yellow, "*flavus*".[114] The flavin cofactor is made from riboflavin **71** (vitamin B2) and is synthesized *de novo* in bacteria, fungi and plants. However, humans and animals rely on exogenous uptake of riboflavin **71** that is first phosphorylated to yield **68** and upon transfer of adenosine-monophosphate (AMP) further converted into **67**.[115,116]

While the flavin cofactor is found to be bound to most FMOs in a non-covalent manner, an increasing number of flavoproteins have been identified that form covalent flavin-protein linkage(s).[117–121] Among those proteins, monocovalent linkage to the flavin isoalloxazine moiety *via* either nitrogen-atom of the histidine side chain is most common. Bicovalent tethering (6-*S*-cysteinyl-8α-N^1-histidyl-flavin) has been reported for some flavoproteins, first observed in the gluco-oligosaccharide oxidase from *Acremonium strictum* (Scheme 1.2.8).[122,123]

Covalent flavinylation is usually a post-translational auto-catalytic protein modification[124–127] and accomplishes more functions than solely preventing dissociation of the cofactor from the enzyme. These functions include: 1, the electron-withdrawing amino acid substituent(s) increase the flavin redox potential substantially[128,129] and facilitate more efficient oxidative catalysis;[123] 2, covalent binding is often, but not always,[129] necessary for structural integrity of the enzyme as mutation of the binding amino acid residues into alanine might lead to protein instability and/or misfolding;[130–132] and 3, as *holo*-enzymes have a longer life-time than *apo*-enzymes[133] covalent flavinylation might counteract protein aging. In addition, it has been suggested the 6-*S*-cysteinyl-flavin linkage in a distinct monocovalent flavoprotein prevents a side reaction that induces C-6 hydroxylation of the flavin cofactor, rendering the enzyme inactive. Therefore covalent flavinylation might represent a self-protection mechanism against enzyme autoxidation in some circumstances.[134]

In order to activate molecular oxygen the flavin cofactor must be in its reduced hydroquinone form, a process that is usually mediated by electron transfer from NAD(P)H **41/42**,[135] but in rare examples cofactor reduction occurs concomitantly with substrate oxidation.[136]

The reduced flavin **69** (and **70**, here exemplarily described for **69**) and dioxygen are believed to initially undergo a single electron-transfer (SET) to generate a radical pair between the neutral flavin semiquinone **72** (one electron reduced) and the superoxide radical (Scheme 1.2.9). This step is required to overcome the spin-inversion barrier between singlet (flavin **69**) and triplet (O_2) state.[104] However, despite some spectroscopic

evidence[137] this paired intermediate **72** has never been observed due to its intrinsic instability to decompose into the C-4a-(hydro)peroxyflavin intermediates **73/74** that represent the catalytically active flavin-oxygen species to perform distinct monooxygenation reactions on various substrates, including oxygenation of heteroatoms (Scheme 1.2.9).[94,104,135]

After insertion of the distal oxygen atom into the target molecule the C-4a-hydroxyflavin **75** is recycled into the oxidized flavin **67** by elimination of H_2O. Anew reduction by NAD(P)H **41/42** sets the flavin cofactor for another catalytic cycle.
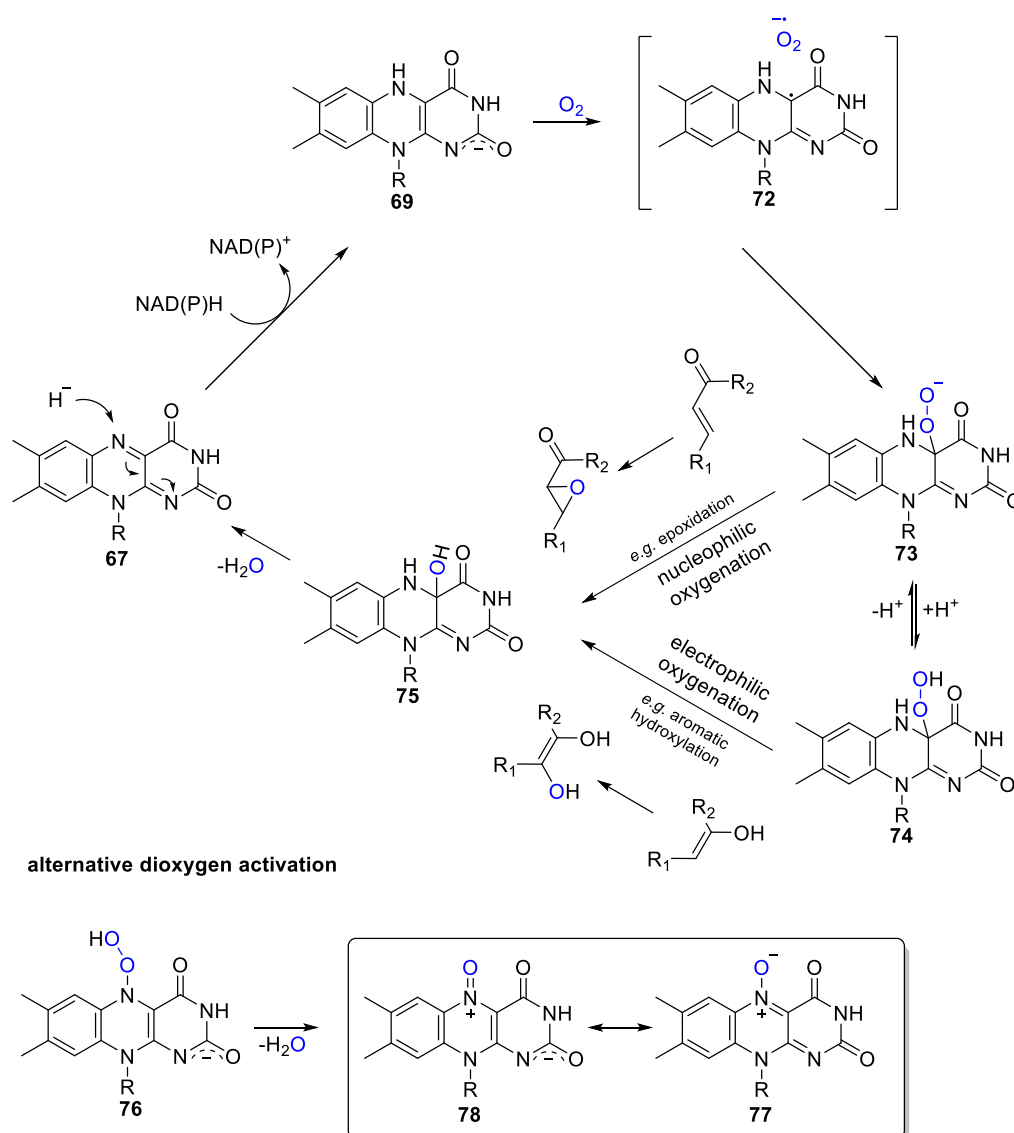
Scheme 1.2.9: Catalytic cycle of FMOs including possible substrate oxygenation reactions and alternative dioxygen activation *via* the N5-hydroperoxide **76**.

In 2013, an up to then unprecedented FMO was reported[138]. Instead of the C-4a-peroxyflavin intermediate **73**, this flavoenzyme forms an N-5-hydroperoxide intermediate **76**. Upon elimination of water **76** would form the N-5-oxide **77**. Experimental data that

support formation of the N-5-oxide **77** are based on spectroscopic data and high resolution mass spectrometry (HRMS).[139]

After this initial discovery two additional flavoproteins have been attributed the N-5-oxide **77** as the catalytically active species.[140,141] Since the catalytic potential of **77** remains elusive and the actual number of flavoproteins employing the same flavin species is unexplored, one can only speculate why specific enzymes form the N-5-oxide **77** rather than the C-4a-(hydro)peroxyflavin **73/74**, that furthermore requires a different orientation of the substrate towards the flavin cofactor.[103]

Even though FMOs may share the same fold topology and a similar active site architecture, their mode of operation may be distinctly different and therefore no reliable catalytic prediction based on protein sequence can currently be made.[94]

Typically, the flavin cofactor is bound deep within the core of the respective flavoprotein. Access is provided by hydrophilic water-filled tunnels that allow selective filtering of substrate molecules.[94] However, as such an environment does not favour the diffusion of rather hydrophobic dioxygen into the active site, thorough investigations revealed that separate tunnels facilitate a "protein-guided" oxygen diffusion within the enzyme.[142,143] Remarkably, the rate of oxygen uptake often is regulated by one or more gatekeeper-residues and it has been shown that mutations at these specific positions can greatly increase or decrease enzymatic activity.[144] Moreover, the majority of FMOs contain a Rossmann fold for binding of the nicotinamide cofactors. This protein domain consists of alternating α-helices and β-sheets and commonly exhibits a GxGxxG motif for binding of NAD(P)H.[145–148]

Halogenation by FMOs

In addition to insertion of oxygen, FMOs may also catalyse halogenation of substrates.[149] Mechanistically, the best studied example is the conversion of L-tryptophan **79** into 7-L-chlorotryptophan **80** by the two bacterial enzymes PrnA and RebH, respectively.[150–153]

Upon reductive activation of dioxygen the highly reactive C-4a-hydroperoxyflavin **74** cannot directly interact with the substrate **79** whose binding site within the enzyme is separated from the flavin cofactor by a 10 Å long tunnel. Instead, a chloride ion located adjacent to the cofactor binding site is believed to perform a nucleophilic attack on the peroxide of **74** to release a free Cl-OH species (hypochlorous acid) that diffuses through the tunnel towards **79**.

A conserved lysine at the end of the tunnel interacts with and activates the free Cl-OH species either *via* hydrogen bonding or formation of a covalent bond (chloramine).[150,151,154] Either of these chlorine species could then chlorinate the substrate by electrophilic aromatic

substitution (EAS, Scheme 1.2.10 A). Therefore, halogenation, as well as oxygenation described previously, is only performed on electron rich substrates (*e.g.* alkenes, aromatics) that act as a nucleophile. Flavin-dependent halogenases are also involved in the tailoring process of fungal polyketides, improving biological activity and increasing structural variety of the respective target molecules.[155–157]
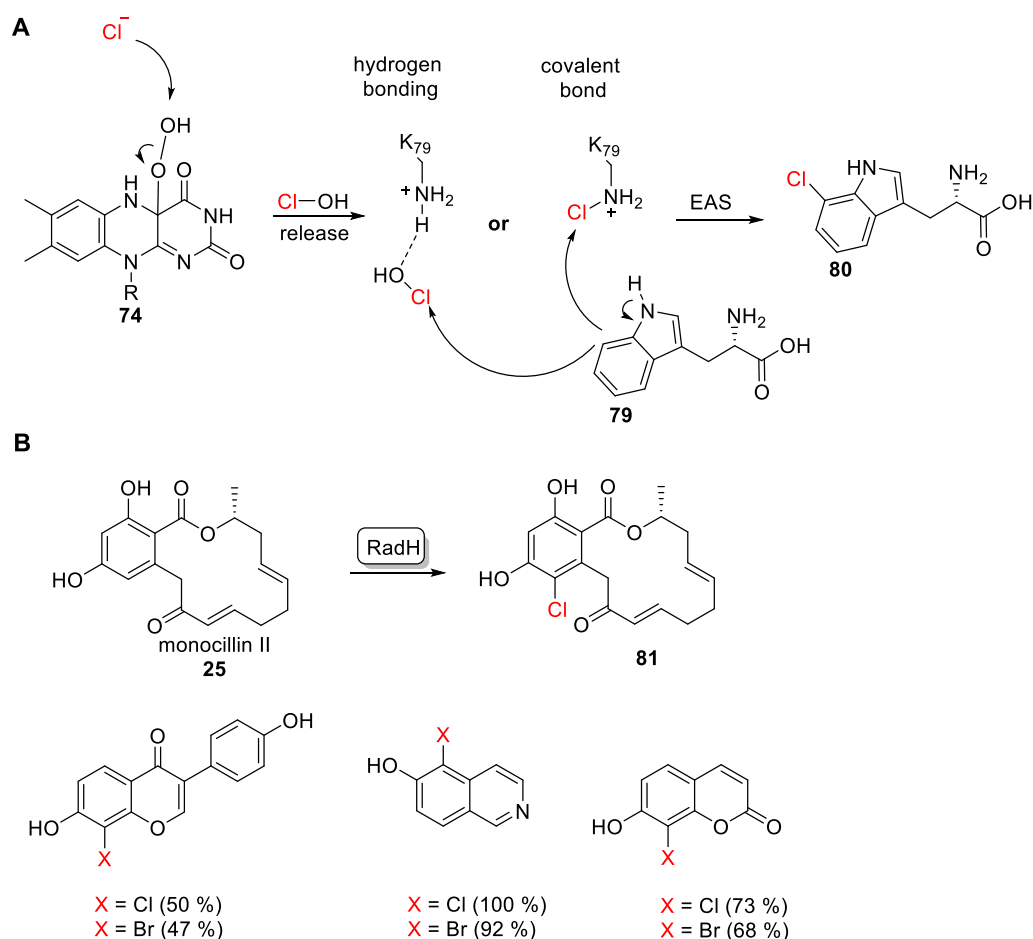
Scheme 1.2.10: Substrate halogenation catalysed by FMOs. **A**, Proposed mechanism for the chlorination of L-tryptophan **79** by PrnA/RebH involving a free Cl-OH species; **B**, Chlorination of the native substrate monocollin II **25** by the fungal flavin-dependent halogenase RadH and alternative substrate molecules with conversion rates according to Micklefield *et al.*[158]

Recently, a highly promiscuous flavin-dependent halogenase RadH was characterized that naturally converts monocillin II **25** into its chloro-congener **81** (Scheme 1.2.10 B).[158] By testing a variety of different potential alternative substrates the biocatalytic potential of RadH was unveiled, incorporating both chlorine and, to a slightly lower extent, bromine into phenolic substrates.

Further structural analysis and site-directed mutagenesis experiments suggested a mechanistic background similar to that of PrnA/RebH: An essential N-terminal lysine residue is required for activation of the chlorine species and another crucial aspartate residue probably acts as a general base to activate the phenol moiety.

1.3 Methods to Investigate Fungal Natural Product Biosynthesis

The serendipitous discovery of penicillin G **2** by Sir Alexander Flemming[7] as well as the isolation of the aminoglycoside antibiotic streptomycin by Waksman and co-workers[159] gave rise to a period from the early 1950s to the late 1960s known as the "Golden Age" of natural product drug discovery.[160]Although high-throughput screening of (semi)-synthetic compound libraries and the contemporary decline in new natural product derived lead structures led to a partial de-emphasis on natural product drug discovery,[9,161] recent advantages in both microbial genomics and bioinformatics (cf. section 1.2.1) as well as synthetic biology and analytical methods may pave the way towards a "New Age" of natural product drug discovery.[160] It is therefore crucial to understand how a distinct natural product is biosynthesized by a certain microorganism. Once the biosynthetic machinery is unveiled, purposeful engineering or exchange of participating enzymes eventually allows the modification of functional groups or the entire scaffold backbone towards the creation of novel natural product analogues that are unprecedented in nature.[162]

Before the current "Age of Genomics" biosynthetic studies often relied on isotope labelling studies, in which the incorporation of small radioactively labelled precursors into larger target molecules was traced and elucidated.[163] Pioneering work in this field was made by Birch and co-workers in the 1950s.[164] Nevertheless, isotope labelling studies (nowadays with stable NMR-active isotopes) still represent a highly useful method to investigate the incorporation of subunits into target molecules,[165,166] for structure elucidation or revision[165,167] as well as to elucidate biosynthetic mechanisms.[168,169]

Over the past decades a multitude of molecular biological techniques have been established, allowing the deciphering of fungal biosynthetic pathways in detail. In general it can be distinguished between *in vivo* (investigations using the whole living organism) and *in vitro* (investigations on isolated proteins) approaches, some of the most common and relevant for this thesis to be discussed below.

1.3.1 Targeted Gene Disruption in Filamentous Fungi

Gene disruption or gene knockout (KO) is a facile technique to elucidate the function of a target gene within a dedicated BGC. It is also useful to link a BGC to a final pathway product as KO of the biosynthetic core gene(s) will abolish or alter production of the respective secondary metabolite. However, two requirements must be fulfilled: first, the target BGC/gene must be expressed in the native fungal producer under laboratory cultivation conditions. Second, the native fungal producer must be susceptible to genetic manipulation including compatibility with the chosen selection marker.

Most commonly, the target gene is disrupted by homologous recombination with an exogeneous DNA fragment carrying a suitable selection marker.[170] The exogeneous DNA fragment is therefore introduced into the fungal cell using various strain specific transformation protocols including polyethylene glycol (PEG)/$CaCl_2$ mediated DNA uptake,[171,172] electroporation[173] or *Agrobacterium* mediated host transformation.[174]

Generally, a preceding step is necessary to generate protoplasts by enzymatic digestion (usually β-1,3-gluconases and chitinases) of the fungal cell wall. This process is often very tedious as protoplasts are highly osmotically sensitive and must be stored in an appropriate buffer.[175]

Even though the target gene to be disrupted and the exogeneous DNA fragment to be inserted usually require at least 1 kb of homologous sequence overlap[176], ectopic DNA integration driven by non-homologous end joining (NHEJ) is the prevalent mechanism and site-specific recombination occurs at frequencies less than 1% (Scheme 1.3.1 A).[170] To overcome this hurdle that results in many false-positive KO-transformants, a split marker (bipartite) system was developed[177] that relies on two individual exogenous DNA fragments that only when recombined by homologous recombination *in vivo* confer an active selection marker to the fungal host organism (Scheme 1.3.1 B).[178,179]

The split within the selection marker gene is designed such that ectopic integration of either the left- or right fragment alone will not produce a functional gene product. Often the hygromycin B resistance gene *hph* is used as a selection maker. The respective phosphotransferase confers resistance to the aminoglycoside antibiotic hygromycin B that inhibits protein biosynthesis in both pro- and eukaryotes.[180,181]

Using the bipartite gene KO strategy Cox and co-workers verified a BGC linked with tropolone biosynthesis in *Talaromyces stipitatus* as disruption of the nrPKS gene *tropA* abolished production of any tropolone related metabolites[65]. Further gene knockout experiments revealed that two additional enzymes TropB and TropC are required to convert the PKS product 3-methylorcinaldehyde **28** into the first tropolone stipitaldehyde **82** *via* intermediate formation of **83** by TropB (Scheme 1.3.1 C).
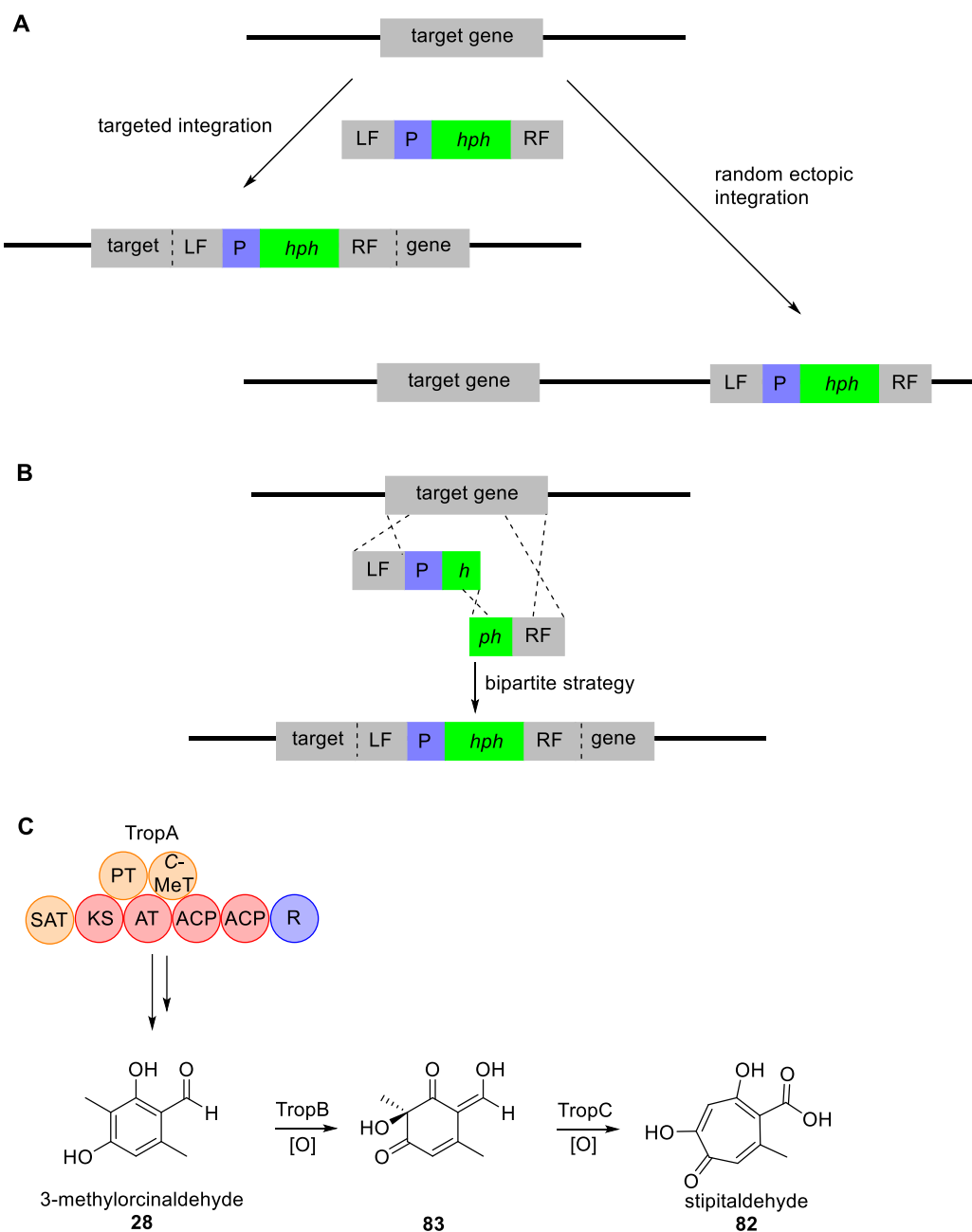
Scheme 1.3.1: Gene disruption by insertion of an exogeneous DNA cassette carrying the *hph* selection marker under control of a suitable promotor (P). **A**, Despite homologous regions (LF and RF) with the target gene, integration into the fungal often proceeds randomly; **B**, A bipartite knockout strategy significantly reduces the amount of false positive transformants; **C**, Biosynthetic steps towards the production of stipitaldehyde **81** in *Talaromyces stipitatus*. The function of each depicted enzyme was elucidated by disruption of the corresponding gene using the bipartite KO strategy.[65]

Another highly efficient method for targeted gene disruption in fungi utilizes the by now well established CRISPR/Cas9 system (Clustered Regularly Interspaced Short Palindromic Repeats).[182,183] Briefly, the gene to be disrupted is targeted by a single stranded guide RNA (gRNA) that possesses a 20bp segment of sequence homology, whereas this part of the gRNA can be edited to match the desired target sequence. Located upstream adjacent to the complementary region, the protospacer adjacent motif 5'-NGG-3' (PAM) is recognized by

the Cas9 endonuclease that induces a double strand break within the targeted double stranded genomic DNA sequence. The double strand break might either be repaired by error prone NHEJ, often resulting in frame shifts and therefore inactivation of the respective gene, or by homologous recombination in presence of a suitable (exogeneous) DNA fragment.

Generally the efficiency of CRISPR/CAS9 mediated gene editing varies greatly between different fungal species (10-100%) depending on several factors including accessibility of the target gene and Cas9 protein expression levels.[183,184] However, in established fungal host organisms that display nearly perfect efficiency this approach has the advantage that no additional selection marker and tedious selection/screening steps are required.[185]

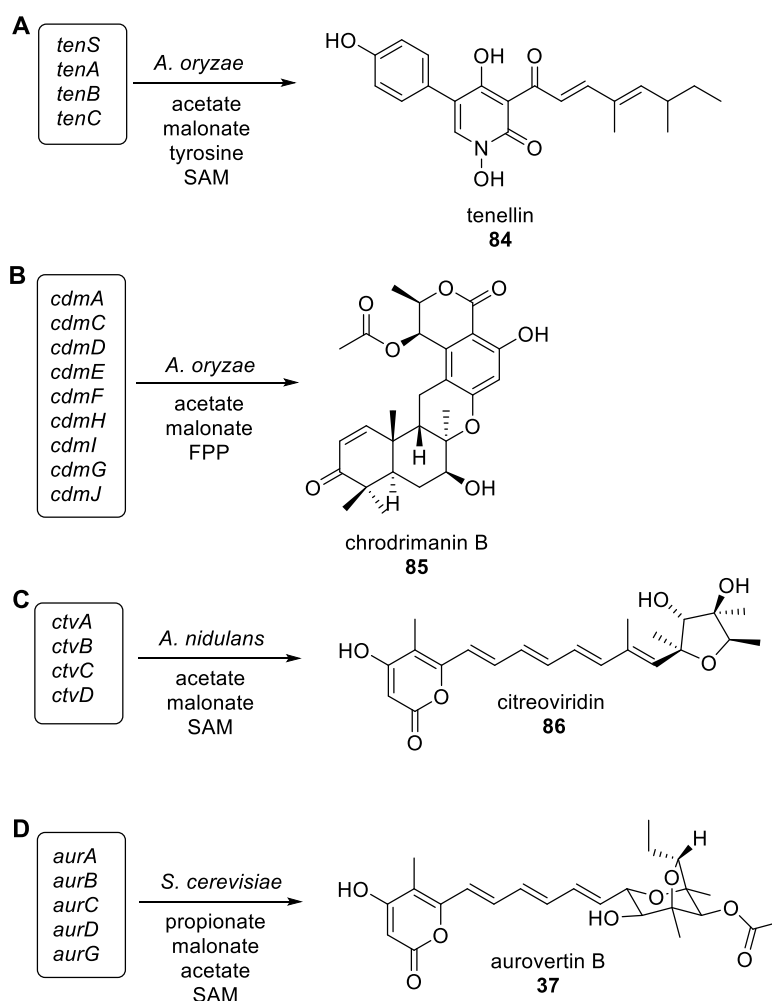## 1.3.2 Heterologous Expression in a Fungal Host System

Heterologous expression is a technique that allows the characterization of single genes or even complete BGCs in a fungal host that (ideally) does not contain any related genes and therefore does not naturally produce any compounds that are linked to the target gene/BGC. This method is especially valuable when the native producer is not easily cultured under laboratory conditions, the respective BGC is not expressed under laboratory conditions or represents a cryptic BGC, or when the native producer is not readily susceptible to genetic engineering.

In addition, the co-expression of genes from different biosynthetic pathways harbours a tremendous potential towards the construction of novel hybrid compounds that are unprecedented in nature and possibly feature unexplored biological activities.[186]

The fungal host strain must be easily transformable, easily selectable (often by auxotrophy), grow quickly, and show little production of native secondary metabolites to allow convenient detection of the heterologously produced compounds of interest. Because of these requirements, only a handful of model organisms have been established as fungal heterologous hosts. These are commonly *Aspergillus* species as well as the yeast *Saccharomyces cerevisiae*.[187,188]

First successful reconstruction of a complete fungal BGC was reported in 2010 by Lazarus and co-workers[189] who expressed four genes from *Beauveria bassiana* in the auxotropic strain *A. oryzae* M2-3 to yield tenellin **84** (Scheme 1.3.2 A).

By using the alternative quadruply-auxotrophic strain *A. oryzae* NSAR1 (that was also used for all heterologous expression studies performed in this thesis), even more genes can be co-expressed simultaneously as reported by Abe *et al.*[190] who reconstituted the biosynthetic pathway towards chrodrimanin B **85** involving nine genes from the native producer *Penicillium verrucolosum* (Scheme 1.3.2 B).

Scheme 1.3.2: Examples for reconstitution of fungal biosynthetic pathways by heterologous expression in different fungal host organisms. Respective genes are in boxes. **A**, Production of tenellin **84** in *A. oryzae*; **B**, Production of chrodrimanin B **85** in *A. oryzae*; **C**, Production of citreoviridin **86** in *A. nidulans*; **D**, Production of aurovertin B **37** in *S. cerevisiae*.

Other examples include heterologous expression of the citreoviridin **86** BGC in *A. nidulans*[191] or production of the related aurovertin B **37** in *S. cerevisiae* (Scheme 1.3.2 C/D).[86] It should be noted that in many examples the titre of heterologously produced compounds is much higher in comparison to native production due to the usage of strong promotors (see below). For instance, tenellin **84** was produced in the heterologous host *A. oryzae* in approx. fivefold enhanced titres compared to the native producer *B. bassiani* (243 mg·L⁻¹ *vs* 47 mg·L⁻¹).[189]

Aspergillus oryzae has been used for production of edible fermentation products (soy sauce, sake) for over 1000 years and has obtained the GRAS status (Generally Regarded As Safe) that allows its usage for the production of food and pharmaceutical products.[192] The genetically engineered quadruply auxotrophic strain *A. oryzae* NSAR1 that is deficient in

nitrate reduction ($niaD^-$) and sulfate assimilation (sC^-), as well as arginine ($argB^-$) and adenine ($adeA^-$) anabolism allows the co-expression of up to 16 genes and is therefore often the host organism of choice for the reconstitution of larger BGCs.[193,194] If required, two additional genetic markers conferring resistance against phleomycin (*ble*) and glufosinate ammonium (*bar*) can also be used to increase the number of genes that can be heterologously expressed at the same time to 24.[195]

For each of those genetic markers a fungal expression plasmid of the pTYGS-series (*e.g.* pTYGSarg) is available that harbours three constitutive (P_{adh}, P_{gpdA}, P_{eno}) and one starch-inducible (P_{amyB}) promoter with the respective terminator regions for the insertion of up to four target genes (Scheme 1.3.3). Furthermore, each expression plasmid contains origins of replication (*ori*) to propagate vector replication in *E. coli* (*colE1*) and *S. cerevisiae* (*2µ*) as well as the corresponding marker genes that confer resistance against ampicillin/ carbenicillin (amp^R) for selection and propagation of *E. coli* transformants and *ura3* that complements uracil deficiency in uracil-auxotrophic yeast strains. The *ccdB* cassette that is flanked by a chloramphenicol resistance gene (cam^R) and attachment sites (attR1 and attR2) for optional Gateway® recombination is located downstream of P_{amyB}.[196]

Insertion of the target genes into any pTYGS expression vector is ideally accomplished by exploiting *in vivo* homologous recombination in yeast. An *Asc*I restriction site located immediately downstream of each constitutive promotor (P_{adh}, P_{gpdA}, P_{eno}) is utilized to linearize the vector backbone yielding 3 individual fragments. Upon a PEG/ LiOAc mediated transformation of linear DNA fragments into yeast, free homologous ends are recombined *in vivo* to yield a circular product.[197–199] During this step additional DNA fragments are co-transformed that represent the genes of interest sharing each 30bp sequence overlap with the respective promotor/terminator region at both 5' and 3' ends.[200] The homologous overhangs are conveniently introduced during the preceding polymerase chain reaction (PCR) using overhang primers.

Positive yeast transformants are selected by growth on uracil-deficient media and the assembled multigene expression vector is transformed into *E. coli ccdB* survival cells to enable higher replication rates. The presence of all desired genes is ultimately monitored by colony PCR and/or DNA sequencing. The *ccdB* cassette may be exchanged for another gene of interest (GOI) by *in vitro* Gateway® recombination with a suitable donor plasmid.

Here, the shuttle vector pE-YA that harbours a multiple cloning site flanked by two attachment sites (AttL1 and AttL2) as well as *ura3*, *2µ* and the kanamycin resistance gene (kan^R) and *pUC ori* for selection and replication of *E. coli* transformants serves as the donor plasmid.[195] The GOI is assembled into the pE-YA vector by yeast homologous recombination and transferred into the target vector (*e.g.* pTYGSarg) *in vitro* using an

enzyme mix (integrase, integration hast factor, excisionase) that swaps the GOI with the *ccdB* cassette of the target vector.[196,201]
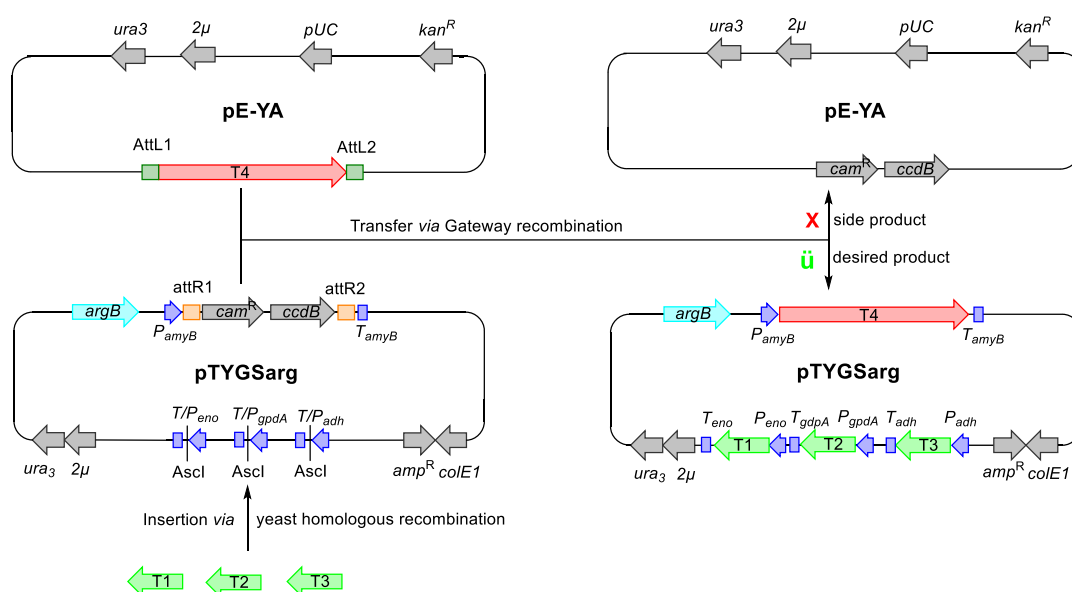
Scheme 1.3.3: Construction of fungal multigene expression vectors. Target genes (T1-T4) are assembled into both pTYGSarg (T1-T3) and pE-YA (T4) exploiting homologous recombination in yeast. Subsequent Gateway® recombination swaps the *ccdB* cassette of the destination vector with T4 of the donor plasmid. Double selection yields only bacterial colonies that carry the desired multigene expression vector (T1-T4) shown at the bottom right hand side.

Subsequently, positive *E. coli* transformants are subjected to double selection on media supplemented with carbenicillin using Top10 host cells to ensure only colonies carrying the desired multigene expression vector survive (Scheme 1.3.3). Large genes (encoding *e.g.* PKS) are often amplified as smaller individual fragments that are recombined in yeast to minimize PCR-derived errors and optimize PCR product yields.[166,186]

In contrast to bacteria, fungal genes frequently contain introns that are spliced during mRNA maturation.[202] Although it is possible that exogenous genes are spliced correctly in the heterologous host,[203] it is often expedient to already clone their intron-free congeners from cDNA of the native producer to avoid intron-related complications that might cause non-functional protein products.[204]

Once the desired fungal expression vector is available, it is transformed into the host as described in the previous section where it integrates ectopically into the fungal genome.[176] It is therefore often necessary to screen several transformants as the location of genomic integration can have an influence on gene expression.[205]

1.3.3 *In vitro* Studies with Isolated Proteins

In vivo approaches offer several benefits such as convenient upscaling of the fermentation volume for the purpose of product isolation or the verification of designated BGC by gene

disruption of the biosynthetic core gene. Still, some disadvantages, including product modification/ shunt reactions by native host enzymes[70,204,206], general chemical instability of intermediates or toxicity towards the host organism are potential drawbacks associated with *in vivo* characterisation of biosynthetic pathways.

In order to circumvent these obstacles, or for means of more profound enzymatic studies (*e.g.* reaction mechanisms, testing alternative substrates, kinetic studies), individual enzymes can be purified from any suitable producer and investigated *in vitro*, ideally in an environment free from any external interference.[207] The most established host organism for recombinant protein production is *E. coli*, with a myriad of different expression strains, vectors and protocols available that are designed to maximize production and purification of soluble target protein.[208–210]

In general, the GOI is cloned into an expression vector that is suitable for the selected host strain. Synthetic DNA is often used because it additionally allows codon optimization for *E. coli* or any other chosen host organism. In the very popular and wide-spread pET-vector series used for *E. coli* BL21 (DE3) derived host strains the GOI is cloned behind the T7-promotor that is recognized by the strong T7-RNA polymerase (T7 RNAP). The T7 RNAP itself is encoded in the host strain under control of the *lac*UV5 promoter that is in turn is repressed by LacI. Addition of the non-hydrolyzable lactose analogue isopropyl β-D-1-thiogalactopyranoside induces expression of the gene encoding for T7 RNAP and subsequent transcription of the GOI.[210,211]

Additional fusion-tags are frequently used to enhance solubility[212] or to enable affinity purification of the recombinant protein.[213] For the later, Ni^{2+}-immobilized-metal affinity chromatography (Ni^{2+}-IMAC) is a commonly used technique that uses immobilized Ni^{2+} ions (Lewis acid) to trap the target gene that is equipped with a terminal hexa-histidin sequence tag (Lewis base).[214] After multiple washing steps the target protein is eluted by applying a high concentration of imidazole. With soluble protein in hand one must ultimately find suitable *in vitro* reaction conditions (buffer, pH, cofactors *etc.*) in order to obtain and optimize enzymatic activity.

In vitro characterisation is not limited to single enzymes. In 2013 Tang *et al.* reconstituted the entire biosynthesis of the fungal polyketide griseofulvin **87** using nine isolated enzymes *in vitro*.[215] Besides step-wise reassembly of the pathway the total *in vitro* biosynthesis of **87** was achieved by simultaneous co-incubation of all respective enzymes in 100 mM sodium phosphate buffer pH 7.4 with the addition of NaCl, the building block malonyl-CoA as well as the cofactors SAM and NADPH (Scheme 1.3.3). The initial polyketide **88** is assembled by the nrPKS GsfA that was purified from *S. cerevisiae*. Two consecutive *O-*

methylations by GsfB and GsfC (purified from *E. coli*) yield the methylated congeners **89** and **90**, respectively.
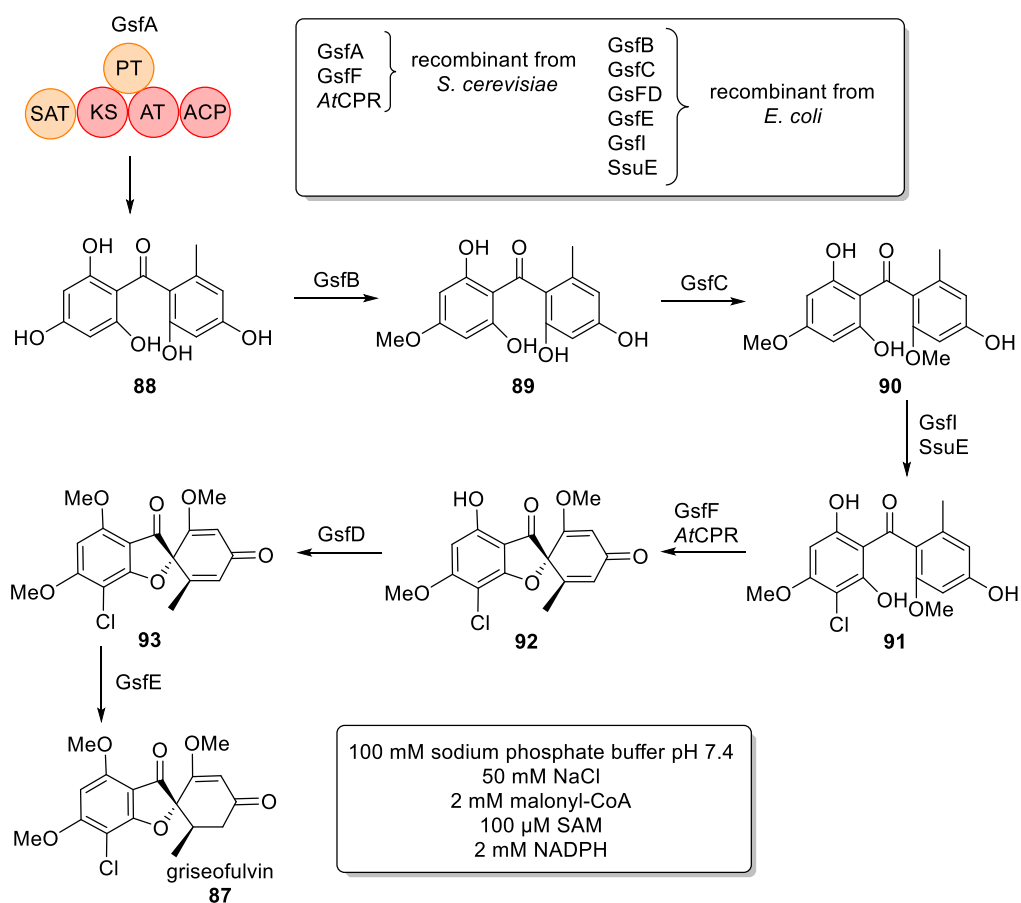
Scheme 1.3.3: Total biosynthesis of the fungal polyketide griseofulvin **87** *in vitro*. Respective recombinant enzymes produced by *S. cerevisiae* or *E. coli* are depicted in the top box. *In vitro* reaction buffer components are shown in the bottom box. New functionalities introduced by each enzyme are displayed in red.

Next, the flavin-dependent halogenase GsfI catalyzes regioselective chlorination. This step requires the ancillary NADPH-dependent flavin reductase SsuE to regenerate the flavin-cofactor of GrfI. GsfF induces an oxidative rearrangement of **91** to introduce the spirocycle moiety into **92**. As eukaryotic cytochrome P450 monooxygenases are usually membrane-bound GrsF was not purified. Instead, the microsomal fraction of *S. cerevisiae* expressing recombinant *grfF* was used that furthermore contained the ancillary cytochrome P450 reductase *At*CPR.

A third *O*-methylation (GsfD) and a final reduction of **93** by GsfE complete the total biosynthesis of griseofulvin **87** *in vitro*. Although this example showcases the potency of *in vitro* studies that often constitute very "clean" experiments, it must be kept in mind that the acquisition of soluble proteins is often the bottleneck. Furthermore, the successful co-incubation of several enzymes under the same reaction conditions is never a trivial process to achieve.

1.4 Terrein and Related Polyketides

1.4.1 Introduction

The mycotoxin (+)-terrein **94** was first isolated from the ascomycete *Aspergillus terreus* by Raistrick and Smith in 1935[216] as a "new mould metabolic product" with the chemical composition $C_8H_{10}O_3$. Despite earlier speculations,[217] the constitution and configuration of the *trans*-diol in **94** was revealed 20 years later by Barton and Miller *via* conversion into a derivative of (+)-tartaric acid.[218] The first total syntheses of racemic and optically pure **94** were reported in 1974 and 1990, respectively.[219–221]

Meanwhile **94** has been isolated from other *Aspergilli* including *A. lentulus*,[222] *A. novofumigatus*,[222] *A. flavipes* MM2[223] and *Neosartorya fischeri*,[224,225] from a marine-derived strain of *Emericella variecolor*[226] as well as from *Penicillium* sp. 20135[227] and the endolichenic *Eupenicillium javanicum*.[228]

The promising biological activities of terrein **94** have been subject to many studies: In addition to anti-inflammatory[229,230] and antimicrobial[231] effects, **94** exhibits anti-cancer properties by suppression of tumour angiogenesis or induction of apoptosis[232,233]. Moreover, **94** has been reported to inhibit melanogenesis[227] and biofilm formation in *Pseudomonas aeruginosa*[234] and was suggested as a new molecular fragment for the development of synthetic proteasome inhibitors.[235] The phytotoxicity of **94** is illustrated by inhibition of root growth and the cause of surface lesions on plants.[236,237]

Notably, while terrein was isolated from natural sources only as the 2*S*, 3*R*-isomer there is one report each on the isolation of either *cis*-diol diastereomer. In 2006 2*R*, 3*R*-isoterrein **95** was isolated along with **94** from *Neosartorya fischeri*[225] while in 2019 2*S*, 3*S*-isoterrein **96**, again together with **94**, was isolated from *Eupenicillium javanicum (*Figure 1.4.1).[228] Both **95** and **96** were in agreement with spectroscopic- and NMR data reported for synthetic standards.[238] The *trans*- and *cis*-diastereomers are best distinguished by the coupling constants between H-7 and H-8 ($^3J_{2,3}$) at the diol-positon (cf. Figure 1.4.1 A). The carbon numbering throughout this section and as displayed in Figure 1.4.1 is based on biosynthetic considerations to be introduced in section 1.4.2 and 1.4.3.

A few structurally (and as will be outlined later also biosynthetically) related chlorinated congeners of **94** have been described that share the characteristic cyclopent-2-en-1-one/ol moiety. In contrast to terrein **94** that has only been isolated as its 2*S*, 3*R*-enantiomer many of the related compounds have been isolated as different stereoisomers, depending on the producing organism. All compounds and stereoisomers mentioned in the following section are illustrated in Figure 1.4.1 B.
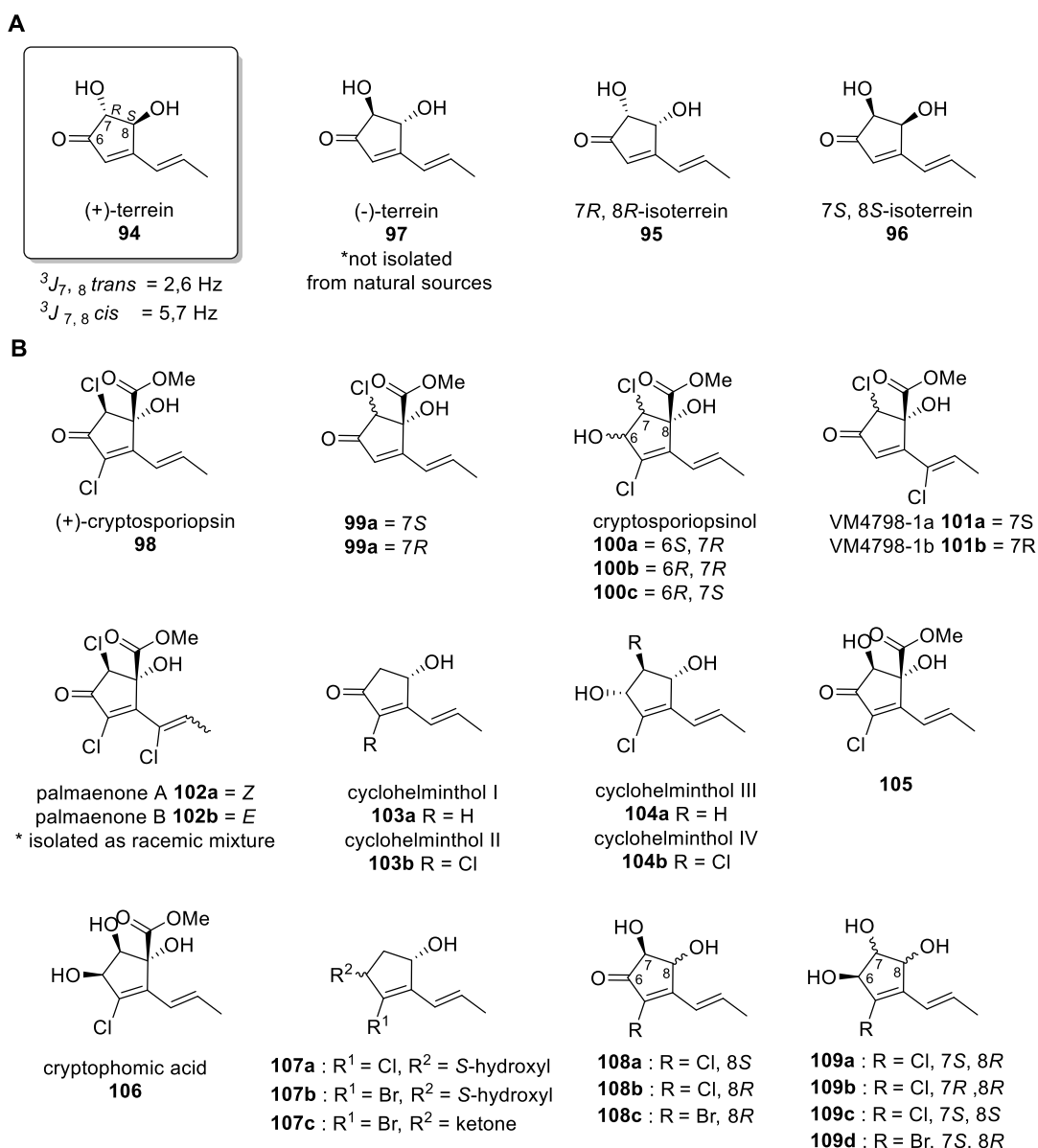
A

B

Figure 1.4.1: Structural overview of terrein **94** and related metabolites. **A**, **94** and its stereoisomers **95 – 97**; **B**, related small molecules **98 – 109** that share the characteristic cyclopent-2-en-1-on/ol moiety. Carbons are numbered based on their biosynthetic origin from the carbon skeleton of precursor 6-hydroxymellein **113**. Since terreins (**A**) and related polyketides (**B**) are products of different ring rearrangements they are numbered differently.

(+)-Cryptosporiopsin **98** was first isolated in 1968 independently from the two fungi *Sporormia affinis*[239] and *Cryptosporipsis* sp.[240,241] and was described as a broad-range antibiotic due to its antifungal and partially antibacterial properties. Along with **98** the two mono-chlorinated diastereomers **99a** and **99b** were also isolated from *S. affinis*.[239] **98** was further purified from an endophytic *Pezicula* sp.[242], *Periconia macrospinosa*[243] and *Phialophora* sp.[244] Interestingly, the later fungus also produced the (-)-enantiomer of cryptosporiopsin **98**. The reduced form of **98**, the alcohol-containing cryptosporiopsinol **100a** was concomitantly isolated from *P. macrospinosa*[243]. In contrast, two additional diastereomers **100b** (C-5 epimer) and **100c** were claimed to be produced by *Cryptosporipsis*

sp.[245] In 2004 the two diastereomers VM 4798-1a **101a** and VM 4798-1b **101b** were obtained as a 3:1 inseparable mixture from *Dasyscyphus* sp. A47-98.[246] The two diastereomers of **101**, that exhibit cytotoxicity against different cancer cell lines, distinguish from the previous compounds **98**-**100** as they are the first examples to show additional chlorination of the side chain.

The first natural trichlorinated congeners palmaenone A **102a** and palmaeonone B **102b** were isolated as a racemic mixture, as indicated by X-ray diffraction analysis, from the discomycete *Lachnum palmae* in 2011. They showed moderate antibacterial activity.[247] **102b** is the only example reported so far that features an isomerization of the C-8/ C-9 alkene.

Another four related congeners, cyclohelminthols I-IV **103ab** and **104ab** were isolated from the culture broth of *Helminthosporium velutinum* in 2016. These are the first examples that do not feature a formal methyl formate functionality at C-2, similar to terrein **94**.[248] In fact, cyclohelminthol I **103a** structurally corresponds to a des-hydroxy analogue of **95** or **97**, respectively (cf. Figure 1.4.1 A). Recently **105**, the C-4 hydroxy analogue of **98**, was observed in the fermentation media of *Pezicula sporulosa,*[249] yet was also found to be produced by *Cryptosporiopsis* sp.[250]
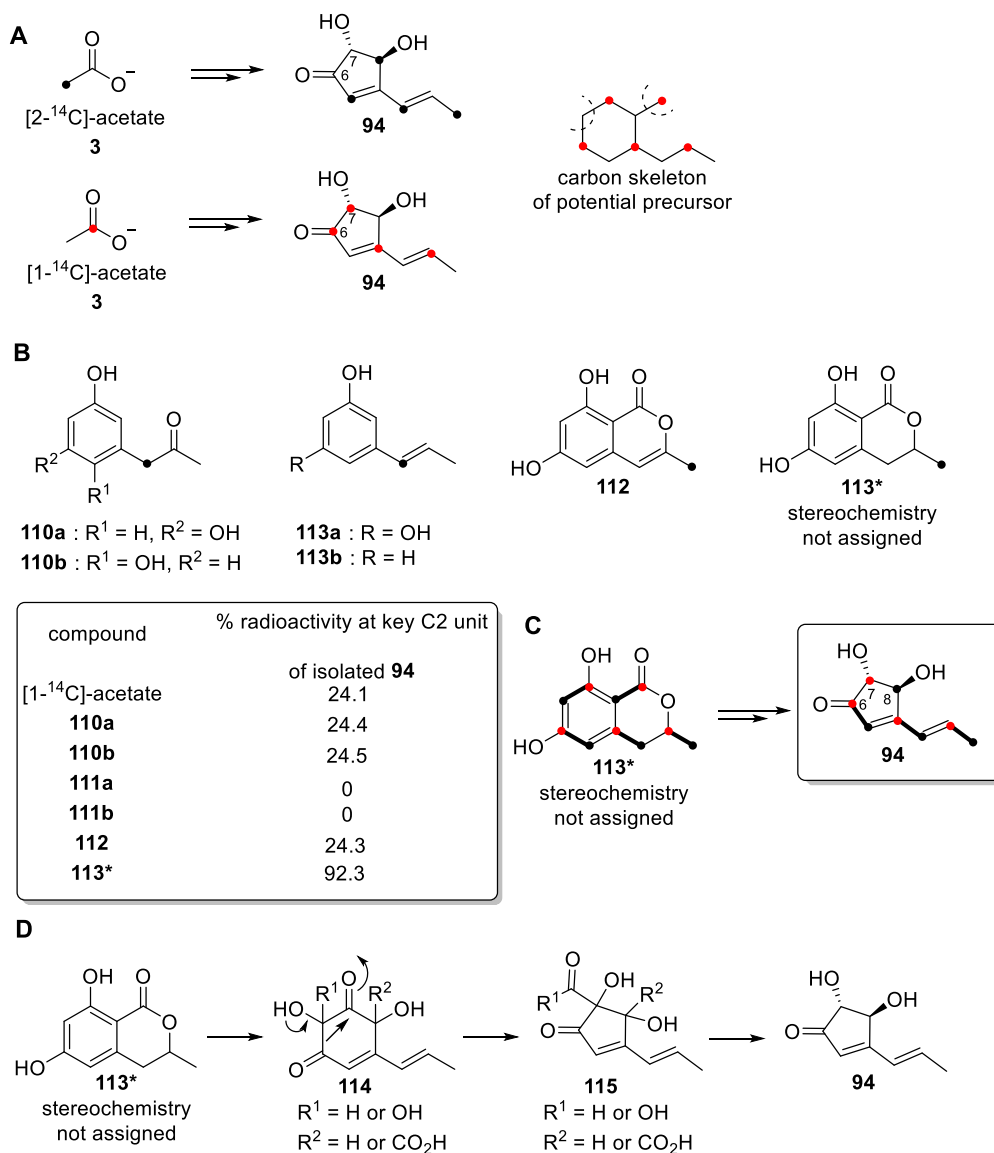
While the majority of the cyclopentenones discussed above possess an *O*-methyl ester, cryptophomic acid **106** isolated from a marine *Phoma* sp. is a rare example in which the free carboxylic acid moiety is present.[251] Finally, in 2020 a total of 9 new congeners **107 – 109** were isolated from the endophytic fungus *Saccharicola bicolor* following the OSMAC (one strain – many compounds) strategy (**109b** was already reported previously).[251,252] These small molecules, termed bicolorins, feature the same carbon skeleton as **94**, but with an additional halogenation at C-6. Notably, there is a high flexibility within the stereochemistry of the C-2- and C-4 hydroxyl substituents although the 2*R*, 4*S*-configuration predominates (cf. Figure 1.4.1 bottom row; carbon numbering based on **98** and other chlorinated congeners), but no biosynthetic evidence available so far.

Consideration of the structures of compounds **94** to **109** shows that the stereochemistry of equivalent substituents is not always uniform, but rather depends on the producing organism.

1.4.2 Isotope-Labelling Studies and Biosynthetic Proposals

The polyketide nature of terrein **94** was first suggested by Birch *et al.* in 1965 based on its chemical formula $C_8H_{10}O_3$ expected for a tetraketide, yet the authors were intrigued by the unusual five-membered ring system.[253] Complementary feeding studies with either [1-^{14}C]- or [2-^{14}C]-labelled acetate **3** were in agreement with a polyketide origin. However, the

experiment revealed two linked [1-14C] signals at the C-3 and C-5 position that represent a break in the commonly found alternating labelling pattern (Scheme 1.4.1 A). Birch *et al.* therefore proposed that the five membered ring is formed by contraction of a six membered precursor. Although not mentioned in the original publication the expected carbon skeleton according to an alternating pattern is further depicted in Scheme 1.4.1 A.

compound	% radioactivity at key C2 unit of isolated **94**
[1-14C]-acetate	24.1
110a	24.4
110b	24.5
111a	0
111b	0
112	24.3
113*	92.3

Scheme 1.4.1: Labelling studies to investigate the biosynthetic origin of terrein **94**. **A**, Labelling pattern of **94** derived from feeding of either [1-14C]- or [2-14C]-labelled acetate **3** and carbon skeleton of a proposed precursor molecule according to Birch *et al.*;[253] **B**, Incorporation of radioactivity into **94** by feeding singly [1-13C]- labelled potential precursors **110 – 113*** providing strong evidence that only **113*** serves as an intact precursor of **94**;[254] **C**, Summarized labelling pattern of **113*** and **94** obtained from feeding of 14C- and 13C-labelled acetate; **D**, Biosynthetic proposal for the formation of **94** based on precursor **113*** by Hill *et al.*[254]

Based on this carbon skeleton and the aforementioned labelling studies Hill *et al.* synthesized a small selection of potential precursors **110 – 113*** that were each labelled with [14C] at one distinct position as indicated in Scheme 1.4.1 B.[254] After these potential precursors were individually fed to *A. terreus* the biosynthesized **94** was isolated and the

terminal C-9/ C-10 (**112** and **113***) or C-7/ C-10 (**110** and **111**) unit was degraded into acetate by Kuhn-Roth oxidation followed by product isolation. Compounds **110** and **112** were found to carry 25% of the total radioactivity of isolated **94** within the respective C_2 key unit (cf. Scheme 1.4.1 B). This result was consistent with degradation of the administered compound by *A. terreus* into acetate that was subsequently re-incorporated into **94** resulting in a radioactivity distributed equally along the compound. This is further supported by the control experiment in which [1-^{14}C]-acetate was fed that yielded the same percentage of relative radioactivity.

In contrast, when the dihydroisocoumarin **113*** (6-hydroxymellein; 6-HM) was supplemented to the fungus, the acetate unit that was derived from degradation of the terminal C-9/ C-10 fragment carried almost 100% of the radioactivity (2.27 % total incorporation of radioactivity into **94**).

This sophisticated experiment demonstrated that 6-HM **113*** is a direct precursor to terrein **94** as it remained intact. Ancillary labelling studies using [1-^{13}C]-, [2-^{13}C]- or [1, 2-^{13}C$_2$]-labelled acetate **3** confirmed previous results by Birch *et al.*[254,255] In fact, **113*** is evidently produced by *A. terreus* and follow-up feeding studies with [1, 2-^{13}C$_2$]-labelled **113*** in 1982 confirmed its direct conversion into **94** (Scheme 1.4.1 C).[245] Based on their observations Hill *et al.*[254] envisaged a biosynthetic proposal for the formation of **94** from **113*** by initial di-hydroxylation to yield an intermediate similar to **114** in that the lactone ring was oxidatively cleaved (Scheme 1.4.1 D). Subsequent carbon-carbon bond shift would yield the cyclopent-2-en-1-one skeleton of **115** that would convert into **94** by deformylation/ decarboxylation.

Following an initial proposal[243] that cryptosporiopsinol **100a** was derived from a dihydroisocoumarin-like precursor (*i.e.* polyketide origin), Holker and Young performed extensive studies in 1975 by feeding [1-^{13}C]-, [2-^{13}C]- or [1, 2-^{13}C$_2$]-labelled acetate **3** to the producing fungus *P. macrospinosa*.[256]

In addition to **100a** the dihydroisocoumarin **116*** was isolated in all three feeding experiments and based on the labelling pattern a biosynthetic proposal was made in that **100a** and **116*** would diverge from a shared intermediate **118** (Scheme 1.4.2 A). In the original publication both proposed intermediates **120** and **121** feature a secondary hydroxyl-group at C-5 whose introduction was not discussed by the authors.[256] Since a ketone at C-5, formed by tautomerisation, would furthermore help to stabilize the expected transition state during the rearrangement of **121** it is shown in Scheme 1.4.2 instead.

Interestingly, **100a** exhibits a labelling pattern different from **94** as the C-4 position was retained within the cyclopentenol-moiety. Accordingly, the pivotal ring contraction must

follow a different mechanism in both compounds by extruding a different carbon-position from the six membered precursor (Scheme 1.4.2 B)
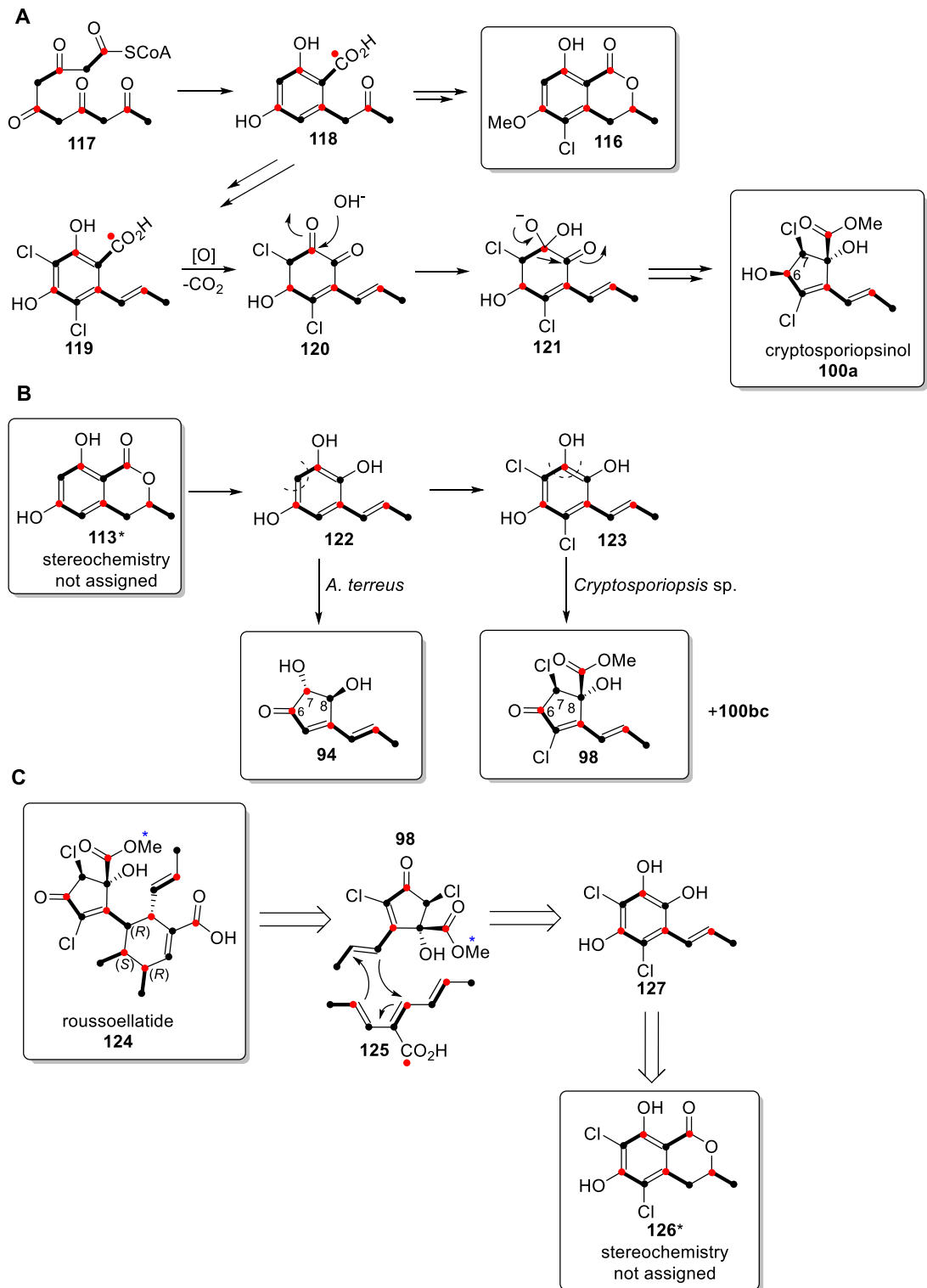
Scheme 1.4.2: Labelling studies to investigate the biosynthetic origin of cryptosporiopsin **98** and related congeners. All compounds in boxes were isolated during the respective studies. **A** Labelling pattern of **100a** and **116** derived from feeding of either [1-^{13}C]-, [2-^{13}C]- or [1, 2-^{13}C$_2$]-labelled acetate including a biosynthetic proposal modified from Holker and Young;[256] **B**, Labelling pattern and biosynthetic proposal for the formation of **94** and **98** from the shared precursor molecule **113***;[245] **C**, Labelling pattern of roussoellatide **124** and biosynthetic proposal based on precursor **98**.[257] Methyl group indicated by blue asterisk is derived from SAM.

In 2015 Berlinck *et al.* isolated an unprecedented derivative from the sponge derived fungus *Roussoella* DLM33 that was named roussoellatide **124**.[257] Labelling studies utilizing [1-^{13}C]-, [2-^{13}C]- and [1, 2-^{13}C$_2$]-labelled acetate suggested that the dichlorinated cyclopent-2-en-1-one moiety of **124** was derived from cryptosporiopsin **98** that was possibly subjected to an intermolecular Diels-Alder reaction with a second pentaketide-derived molecule **125** (Scheme 1.4.2 C). The dichlorinated isocoumarin **126*** that was isolated from the same strain might therefore constitute a precursor to **98**, suggesting that, opposed to previous assumptions, chlorination would take place prior to oxidative decarboxylation. In addition, it was experimentally proven that the *O*-methyl ester was derived from SAM ([*methyl*-^{13}C]-methionine labelling), suggesting the involvement of an *O*-methyltransferase during the biosynthesis of **98** and related congeners.

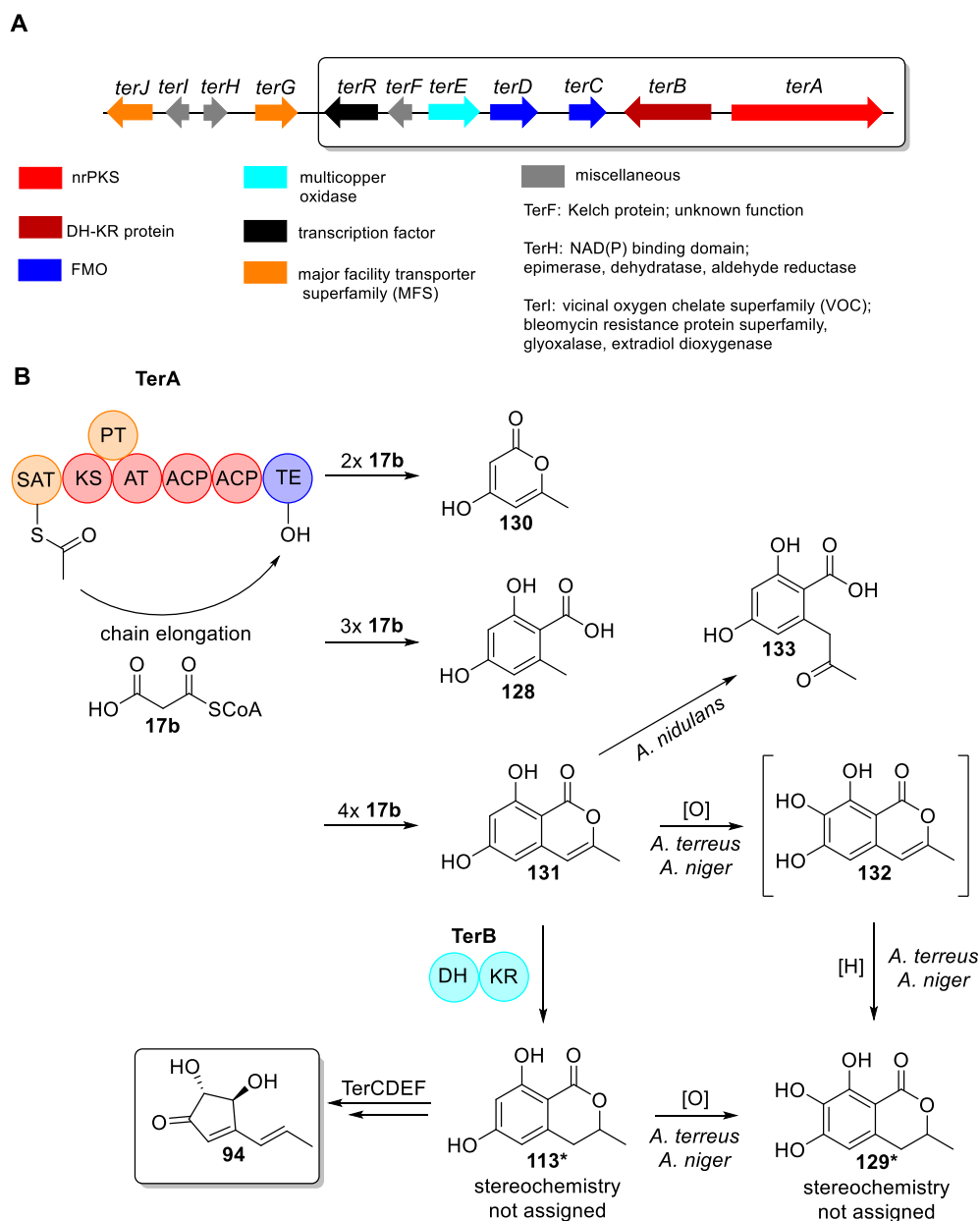### 1.4.3 Biosynthetic Gene Cluster and Proposed Biosynthesis

The terrein BGC was serendipitously discovered by Brock *et al.*[237] during studies that were originally meant to investigate the molecular basis for conidia pigmentation in *A. terreus*. It encodes for a nrPKS (TerA), an unusual DH-KR di-domain protein (TerB), two FMOs (TerC and TerD), a multicopper oxidase (TerE), a transcriptional regulator (TerR) and a Kelch protein of unknown function (Scheme 1.4.3). Further genes encode for transporter proteins (TerG and TerJ), an NAD(P)H-binding protein (TerH) and a potential resistance protein (TerI).

Gene knockout of the nrPKS encoding gene ATEG_00145 (Δ*terA*) completely abolished production of terrein **94**. Semiquantitative RT-PCR of 24 adjacent genes revealed that the expression of 10 additional genes (*terB-terJ*, Scheme 1.4.3 A) was co-regulated with expression of *terA*.

Individual knockout-transformants Δ*terB*-Δ*terJ* concluded that, in total, six genes *terA-terF*, as well as the transcription factor TerR, are essential for production of **94** in *A. terreus*. In fact, later studies demonstrated that TerR is a very potent transcriptional activator that binds to its promotor region with high affinity and is very useful as used in a high-performance fungal heterologous expression system.[258]

Knockout of the DH-KR di-domain protein TerB lead to accumulation of the tetraketide orsellinic acid **128** and the pentaketide **129*** in the fermentation medium. The same two polyketides were produced by heterologous expression of TerA in *A. niger* along with the triketide lactone **130** (Scheme 1.4.3 B). Feeding experiments with [1-^{13}C]-labelled acetate and various [^{13}C]-labelled alternative starter units revealed that TerA is exclusively primed with acetyl-CoA. Swapping and deletion of the TE-domain revealed that although this

domain is necessary for any product release it does not determine the chain length. Hence the formation of different chain length products **128** - **130** was attributed to the KS domain.

Scheme 1.4.3: Terrein biosynthesis in *A. terreus*. **A**, The terrein BGC in *A. terreus*. Genes proposed to be essential for production of **94** based on knockout studies are framed. The function of corresponding proteins is based on the NCBI conserved domain database (CDD);[259] For further analysis on adjacent genes see appendix Table S1; **B**, Characterization of the nrPKS TerA and proposed biosynthesis of terrein **94**.[237]

Mutants Δ*terC*-Δ*terF* all showed production of 6-HM **113*** that would therefore be derived from TerAB, but no additional putative pathway intermediates could be isolated due to their chemical instability. Interestingly, Δ*terD*-Δ*terF* strains exhibited a red media colouration that might be derived from accumulation of biosynthetic intermediates and/ or shunts.

Fractionated feeding experiments showed that only **113*** restored production of **94** in the Δ*terA* strain, but not **129***. Further experiments showed that **113*** was converted into **129***

by both *A. terreus* and *A. niger*. The authors therefore concluded that **129*** was a shunt product that would result from modification of the actual pentaketide product **131** (cf. Scheme 1.4.3 B).

Additional expression of *terB* in *A. niger* was not successful, but it was proposed that TerB would open the lactone ring of **131**, perform a reduction and yield **113*** by re-lactonization.[237] Through heterologous expression of *terA* in a different host organism, *A. nidulans*, Oakley *et al*. furthermore observed production of the pentaketide in both its lactone **131** and corresponding acid form **133**.[260]

Although knockout experiments showed that the transformation of 6-HM **113*** into terrein **94** requires the action of two FMOs (TerC, TerD), one multicopper oxidase/laccase (TerE) and one Kelch protein with no obvious function (TerF) the detailed steps remain obscure to date.

Using bioinformatic analysis Oikawa *et al*. recently identified putative BGCs for the family of palmaenones **102** (*plo*) and cyclohelminthols **103/104** (*chm*) in the genomes of *L. palmae* strain NBRC106495 and *H. velutinum* strain yone96, respectively.[261] Both BGC contain genes that share significant homology with each of *terA*, *terB, terC* and *terR* as well as two halogenase genes (*ploKN* and *chmKN*), but no genes homologous to *terDEF* or any of the allegedly non-essential genes *terG-terJ* are present in these two BGC. Heterologous expression of *chmAB* in *A. oryzae* resulted in high production of 6-HM that was shown to be the *R*-enantiomer **113**, thereby confirming the BGCs.

In contrast to previously mentioned *Aspergillus* sp., **113** was converted into the hydroxylated **134** by native oxidative enzymes in *A. oryzae* (Scheme 1.4.4). Combining heterologous expression and *in vitro* studies with purified enzymes the regioselectivity of the flavin-dependent halogenases PloKN and ChmKN towards the substrate **113** was revealed (Scheme 1.4.4). PloK thereby is the only halogenase that is further able to chlorinate outside of the aromatic ring at C-8.

After chlorination the oxidative decarboxylation of **126/138** was proposed to be catalysed by the FMO PloC or ChmC to yield intermediate **127/139**, respectively, that also features the alkene functionality in the side chain that is present in the final pathway products. Accordingly, TerC would fulfil the same function during the biosynthesis of **94**. However, experimental data regarding the characterization of any homologous FMO are not yet available.

In summary, all the studies carried out indicate that the biosynthetic pathways of terrein **94** and congeners **98 – 106** diverge after the three first biosynthetic steps: In each case the

polyketide backbone is assembled by an nrPKS and a highly unusual multidomain protein, followed by oxidative decarboxylation catalysed by a distinct FMO.
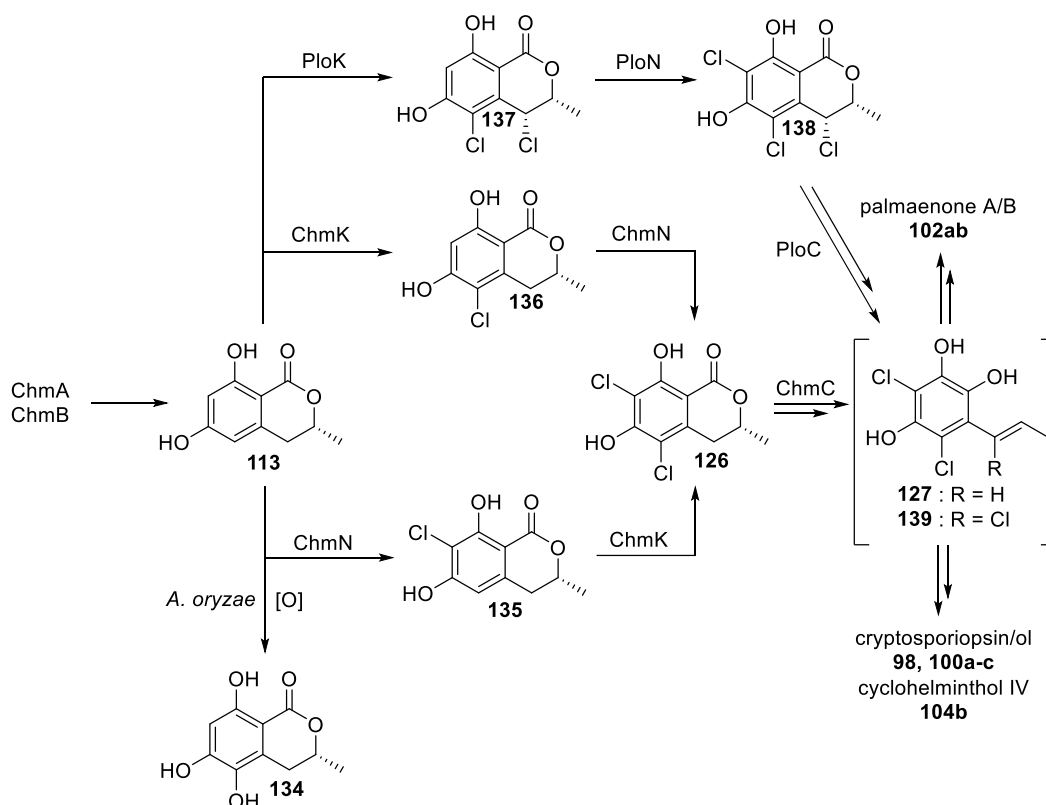

Scheme 1.4.4: Characterization of early biosynthetic steps towards chlorinated cyclopentenones.[237]

It appears very reasonable that particularly the additional (oxidative) genes in the terrein BGC (*terDEF*) cause a different mode of ring contraction. It remains to be seen whether the bicolorins **107** – **109** originate from a "hybrid" BGC as their structure shows features that are found in both **94** and **98** – **106** and no labelling studies are available so far.

1.4.4 Biosynthesis of 6-Hydroxymellein

Although both an nrPKS (TerA) and a DH-KR di-domain (TerB) are required for the formation of the universal precursor molecule 6-HM **113**, the exact biosynthetic background remains to be elucidated. Two possibilities are as follows: TerA releases the pentaketide lactone **131** that might interconvert with the free acid **133**, or **133** is directly released from the PKS (Scheme 1.4.5 A). TerB would subsequently catalyse conversion into alcohol **140** that yields **113** by lactonization. Alternatively, TerB could perform ketoreduction on the TerA-tethered polyketide intermediate **141** in *trans* during chain elongation. Finally, the pentaketide alcohol **142** would be released from the PKS by intramolecular lactonization to give **113**. It further stands to question what the DH-domain in TerB is needed for.
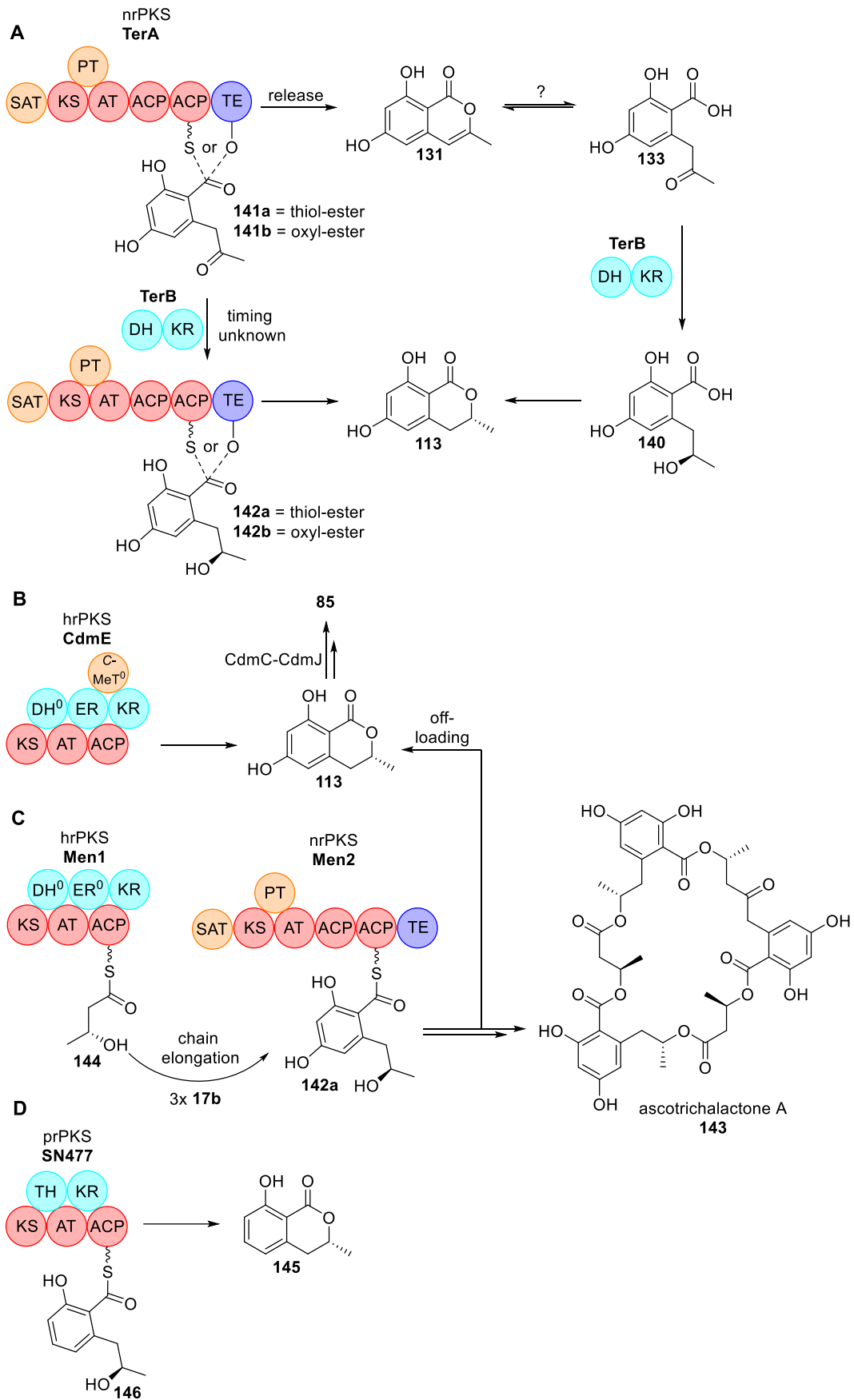
Scheme 1.4.5: Biosynthesis of **113** by fungal PKS systems. **A**, The nrPKS TerA and the DH-KR di-domain protein TerB produce **113**, the exact mechanisms is not yet elucidated; **B**, the hrPKS CdmE directly releases **113** that is converted into **85** by tailoring enzymes CdmC-CdmJ; **C**, **113** is produced as a derailment product by the hrPKS Men1 and the nrPKS Men2; **D**, Mellein **145** itself is produced by a prPKS (SN477).

Interestingly, two other systems in fungi have been reported for the biosynthesis of **113** so far: The hrPKS CdmE directly releases **113** during the biosynthesis of the meroterpenoid chrodrimanin B **85** (Scheme 1.4.5 B, Scheme 1.3.2 B).[190] During the biosynthesis of the macrocyclic polylactone ascotrichalactone A **143** the diketide **144** provided by the hrPKS Men1 is elongated by the nrPKS Men2. Successive esterification and cyclolactonization reactions carried out by these two PKS are sufficient for the production of **143**. 6-HM **113** would therefore constitute a premature off-loading or derailment product, respectively.[262]

In addition, the dihydroisocoumarin mellein **145** itself, lacking the "6-hydroxyl" group of **113**, is produced by a fungal iterative prPKS (Scheme 1.4.5 D).[263] The fact that fungal melleins are produced by (at least) four different PKS systems highlights the general importance of these intriguing small molecules for theirs producers with regards to biochemical and ecological processes.[264]

1.5 Sorbicillinoids

1.5.1 Classification of Sorbicillinoids

Sorbicillinoids constitute an important family of hexaketides produced by both terrestrial and marine ascomycetes including the genera *Trichoderma*, *Penicillium*, *Verticillium* and *Aspergillus*.[265,266] The eponymous molecule sorbicillin **147a** was itself isolated from *Penicillium notatum* by Cram and Tishler in 1948.[267,268] A review[266] from 2016 lists about 90 sorbicillinoids and novel congeners are still frequently isolated,[269–271] indicating the number of sorbicillinoids identified up to date is over one hundred.

During the past decades a lot of research regarding the structure, biosynthesis and biological activities of these fungal pigments was conducted.[265,266] Sorbicillinoids display a distinct hexaketide backbone with the characteristic C-1'/ C-6' sorbyl sidechain as presented by sorbicillin **147a**. Based on their structural architecture sorbicillinoids are classified into monomeric, dimeric, trimeric and hybrid sorbicillinoids (Figure 1.5.1).

As with many of the over 30 described monomeric members, the hexaketide origin of sorbicillin **147a** as well as of its 2',3'-dihydro congener **147b**[272] is easily recognized. In fact, most of the sorbicillinoids have been reported both with a 2', 3'-alkane and a 2', 3'- *E*-alkene side chain. For clarity all structures within this chapter will be shown with the 2', 3'-*E*-alkene side chain (sorbyl).

Both **147a** and **147b** show inhibitory activity on HeLa and HepG2 cells in the low μM range.[273] The highly reactive sorbicillinol **148a**, an intermediate during the formation of dimeric sorbicillinoids, is formed through oxidative dearomatization of **147a**.[274]

Further oxidative steps furnish transformation into epoxysorbicillinol **149a**[275]. Often the phenyl-substituent is subjected to modifications that include epoxidation, lack[276] and/ or oxidation of an α-methyl branch as well as *O*-methylation as exemplified by **150**, produced by a fungicolous *Phaeoacremonium* sp..[277]
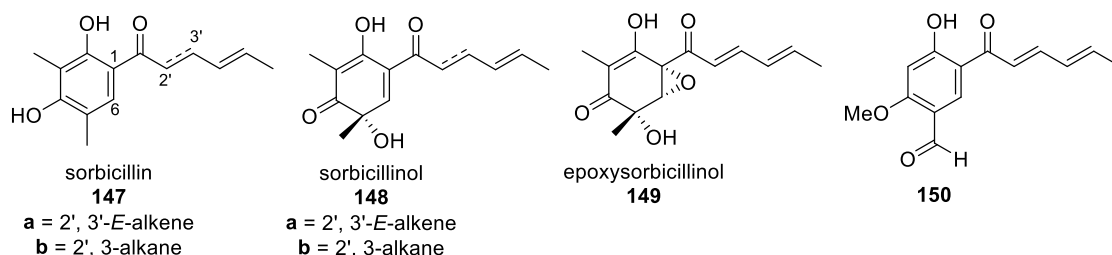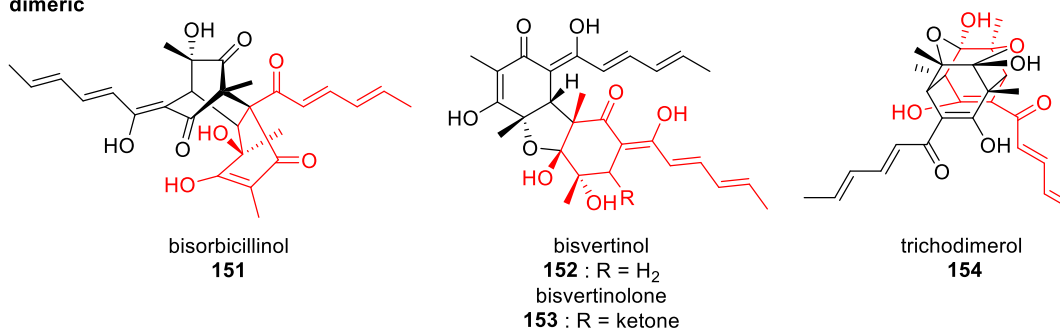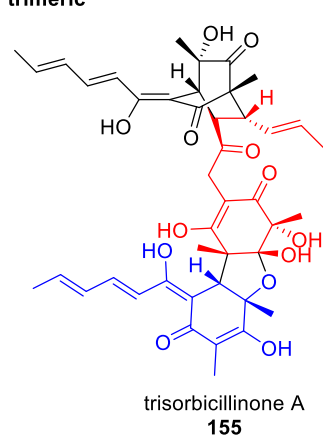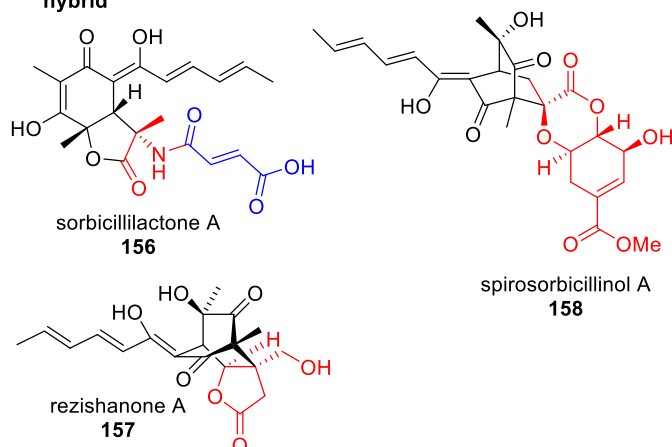
Figure 1.5.1: Structural overview of monomeric, dimeric, trimeric and hybrid sorbicillinoids. Subunits are indicated by different colours.

Dimeric sorbicillinoids consist of two subunits derived from **148** and are divers in structure. The Diels-Alder derived bisorbicillinol **151**, first isolated by Abe *et al.*, was shown to possess radical-scavenging activity almost matching that of α-tocopherol (vitamin E).[278,279] Bisvertinol **152**, with an architecture distinctly different from **151**, was first isolated by Dreiding and co-workers[280] from *V. intertextum* in 1986. Its oxidized congener bisvertinolone **153**, isolated from *T. longibrachiatum*,[281] was claimed to be the first natural

β-1,6-glucan biosynthesis inhibitor.[282] Trichodimerol **154**, with its a cage-like structure, is a potent inhibitor of the prostaglandin H synthase-2 which results in a down-regulation of prostaglandin biosynthesis and subsequently reduces the presence of TNF-α, a cytokine involved in many diseases.[283,284]

The first trimeric sorbicillinoid trisorbicillone A **155**, combining structural features of both **151** and **153**, was isolated from a marine *Phialocephala* sp. and characterized in 2007.[285] Three further trimeric derivatives which all show cytotoxic activities were described a few years later isolated from the same strain, along with a total of 14 other sorbicillinoids.[286]

Hybrid sorbicillinoids are easily distinguishable by incorporation of a non-sorbicillin derived subunit. Among the roughly 25[266] characterized hybrid molecules, sorbicillilactone A **156** is one of the rare examples containing a nitrogen atom. By feeding experiments Bringmann *et al.*[287] demonstrated that the nitrogen containing subunit is derived from L-alanine, whereas the additional C_4-unit might be derived from a fumaric acid derivative. Notably, **156** exhibited anti-HIV properties and was considered to be a good lead candidate for the inhibition of the Ebola virus matrix protein VP40, based on computer assisted drug screening.[288]

The family of rezishanones, *e.g.* **157**, constitute the first natural sorbicillinoids reported that are formed by Diels-Alder reaction with a dienophile not related to sorbicillinol **148**. **157** showed weak antimicrobial activity against *S. aureus* and *B. subtilis* while being non-toxic to fungi.[289] Finally, spirosorbicillinol A **158**, isolated from both soil and marine *Trichoderma* sp., has been proposed to be derived from dimerization with scytolide, that in turn is assumed to be related to the primary metabolite shikimic acid.[290–292]. Similar to bisorbicillinol **151**, **158** exhibited prominent radical scavenging activity.

Taken together, sorbicillinoids cover a broad spectrum of pharmaceutically important properties including anticancer, antimicrobial, antiviral and antioxidant activities. Although highly diverse in structure, the core of every sorbicillinoid is based on a distinct hexaketide backbone.

1.5.2 Isotope-Labelling Studies and Biosynthetic Proposals

Dreiding and co-workers characterized several new monomeric and dimeric sorbicillinoids produced by *Verticillium intertextum* between 1981 and 1986.[272,280,293] Their biosynthetic proposal (Scheme 1.5.1), premised on the assumption that the methylated hexaketide sorbicillin **147a**, which arose from cyclization of the PKS-bound intermediate **159** at the carboxyl-terminus, was the key biosynthetic precursor for all the sorbicillinoids. The next proposed intermediate, epoxide **160**, would be able to undergo two different epoxide opening reactions. Formal tautomerisation and epoxide opening would yield intermediate

148a (sorbicillinol) that forms the dimeric bisorbicillinol **151** *via* Diels-Alder reaction with itself. Alternatively, reductive epoxide opening would yield intermediate **161**. Nucleophilic attack of **161** onto the epoxide moiety in **160** followed by a consecutive nucleophilic attack of the newly installed alcohol-functionality would explain the formation of bisvertinol **152**.
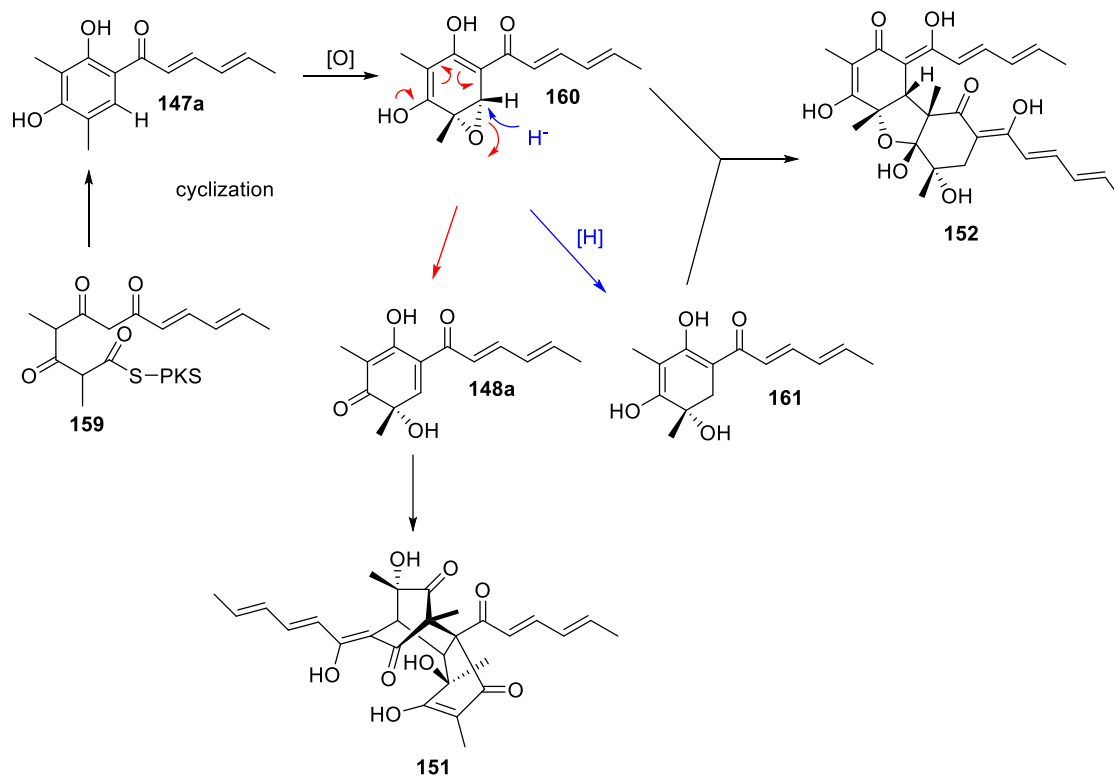
Scheme 1.5.1: Biosynthetic proposal for the formation of dimeric sorbicillinoids **151** and **152** based on initial oxidation of sorbicillin **147a**.

As even more (dimeric) sorbicillinoids were discovered, the monomeric sorbicillinol **148**, which was postulated by Trifonov and co-workers,[280] was deemed the central compound from which the branching towards the majority of (dimeric) sorbicillinoids would occur. In fact, the presaged sorbicillinol **148a** was identified *in vivo* for the first time by Hirota *et al.* in 2000 from *Trichoderma* USF-2690.[294] By HPLC the authors noticed that the area of the corresponding peak in the chromatograms decreased over the time, while the peaks corresponding to later dimeric compounds increased in area. Although **148a** was stable in aqueous solution, it readily dimerized to bisorbicillinol **151** during extraction of the culture broth using ethyl acetate, or into trichodimerol **154** by lyophilisation, highlighting its inherent reactivity.

This observation was in line with previous results by Corey and Barnes-Seeman who obtained **154** after *in vacuo* concentration of synthetic **148a** followed by TLC in hexane-ethyl acetate (3:1).[295] Due to the instability of **148a** upon chemical workup and

concentration, the structure was confirmed by *O*-acetylation and comparison with a synthetic standard and complementarily by hydrolysis of the synthetic standard.[294,296]

Following their confirmation of the existence of sorbicillinol **148a**, the Abe group performed a series of elegant feeding studies with [^{13}C]-labelled acetate that proved the anticipated polyketide origin of sorbicillin **147a**, **148a** and bisorbicillinol **151** among other dimeric sorbicillinoids.[297–300] However, despite thorough attempts Abe *et al.*[299] were not able to assuredly deduce the precise order of the oxidative steps required for the formation of **148a** (Scheme 1.5.2). Based on their key observation that only one isotopomer of **148a** was obtained from feeding [1,2-^{13}C$_2$]-labelled acetate to *Trichoderma* sp. USF-2690, the authors realized that the release of the PKS-bound hexaketide intermediate **159** could not follow a Claisen-like cyclization as the resulting product **162** was symmetrical and would therefore cause a scrambled labelling pattern in **148a**. The release of the PKS-tethered intermediate would accordingly involve a reduction at the C-1 position that leads to formation of an asymmetrical released product.

The following second key question was the timing of oxidation at the C-2 position. Oxidation of **159** that is still attached to the PKS would yield intermediate **163** that directly yields sorbicillinol **148a** upon reductive release and dehydration of intermediate **164**. Alternatively, reductive release of **159** would give sorbicillin **147a** first and proceed to **148a** *via* oxidation of **147a**. Attempting to answer this question the fungal fermentation medium was supplemented with [1-^{13}C]-labelled **148a** and in fact the isolated **147a** showed significant [^{13}C]-enhancement at the six expected positions (C-1, C-3, C-5, C-1', C-3', C-5'). Although this experiment demonstrated that **148a** serves as a precursor molecule to **147a** in *Trichoderma*, it does not rule out that the opposite conversion *i.e.* oxidation of **147a** into **148a** is also possible. A few years later Bringmann *et al.* proved by feeding of (*methyl*-^{13}C)-methionine that the two methyl groups of **148a** at C-2 and C-4 are derived from SAM.[287]

Although structurally different at the first glance, the vertinolides share distinct similarities with monomeric sorbicillinoids including the sorbyl side-chain and the presence of a tertiary oxygen-atom with *S*-configuration.[265,280] Vertinolide **165** itself was first isolated in 1981 along with **147ab** and **148b**.[293] In 2007 the Abe group obtained [^{13}C]-labelled 5-hydroxyvertinolide **166** from a culture of *Trichoderma* sp. USF-2690 that was supplemented with either [1-^{13}C]-, [2-^{13}C]- or [1, 2-^{13}C$_2$]-labelled acetate, respectively.[300] The group speculated that intermediate **164** would be both the immediate precursor for **148a** (dehydration) and **166** (Scheme 1.5.2).

Upon tautomerisation **164** could putatively undergo an intramolecular rearrangement by nucleophilic attack of the C-2 hydroxyl group onto the C-5 carbonyl group, followed by C-5/C-6 ring cleavage of the strained bicyclic system. However, this proposal does not link to the

formation of vertinolide **165** that does not show the respective hydroxyl-group that would be introduced during assembly of the polyketide backbone in **164**.

only one ⇌ isotopomer of **148a** observed
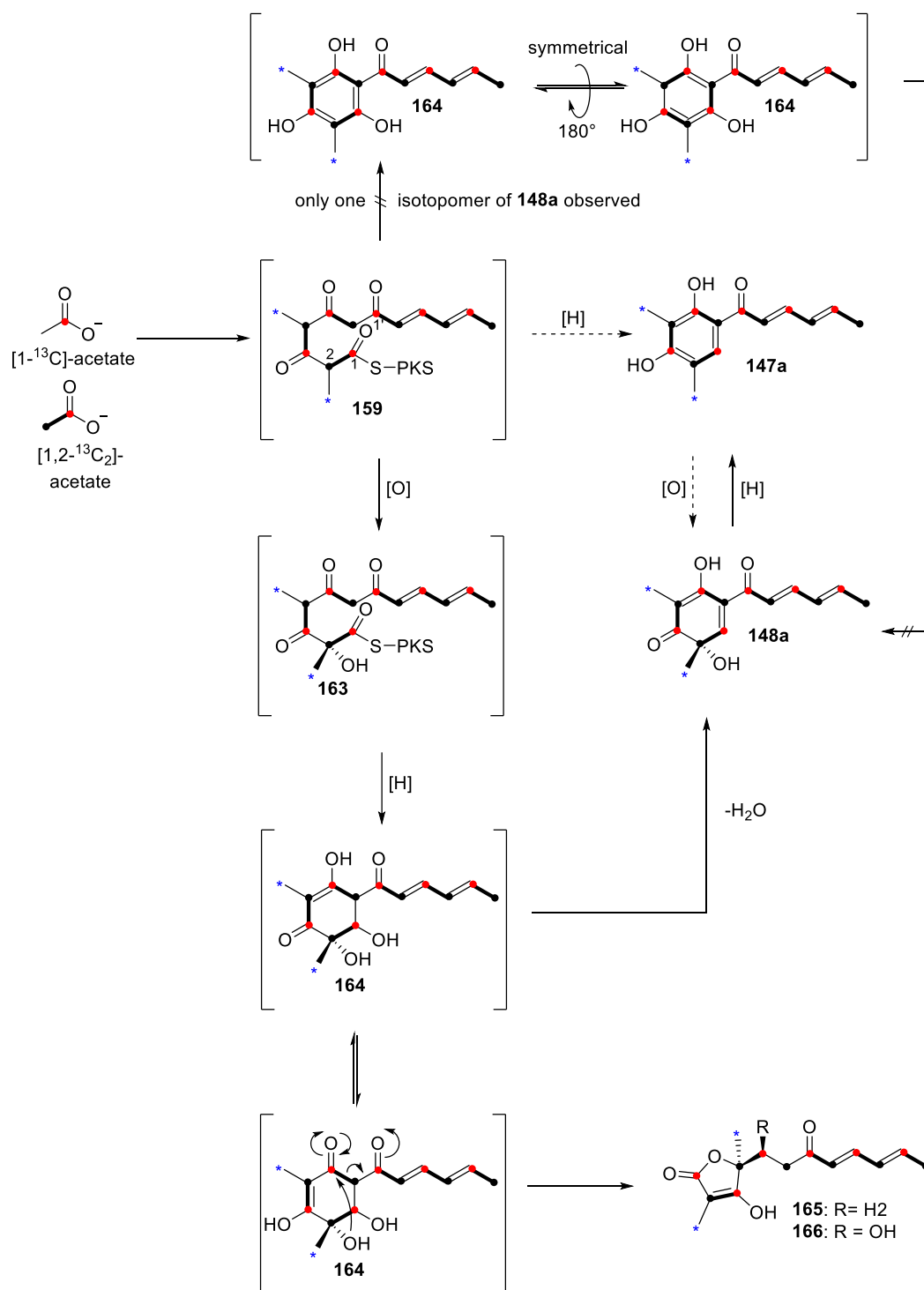
Scheme 1.5.2: Proposed biosynthesis of sorbicillinol **148a** and 5-hydroxyvertinolide **166** based on labelling studies by *Abe et al.*[299,300] Methyl groups indicated with blue asterisk are derived from SAM.[287]

Already upon its first isolation in 1981 bisorbicillinol **151** was proposed to derive from an intermolecular Diels-Alder reaction between two molecules of sorbicillinol **148**, that would serve as both the diene and dienophile.[293] Both the Abe group and Nicolaou *et al.* took two

different approaches to prove the origin of **151**. The approach by Abe and coworkers[297,299] was based on initial feeding experiments using [1-13C]- and [2-13C]-labelled acetate, whereas in a follow up experiment the fermentation medium was supplemented with [1-13C]- and [2-13C]-labelled sorbicillinol **148a**. The labelling pattern of **151** isolated from the respective fungal cultures was identical suggesting that it was indeed derived from two molecules of **148a** (Scheme 1.5.3). The lacking observable 13C-enrichement at the C-9 and C-11 position of **151** was explained by rapid keto-enol-tautomerisation that also caused a non-visibility of C-9 and C-11 in the 13C-spectra of unlabelled **151**.
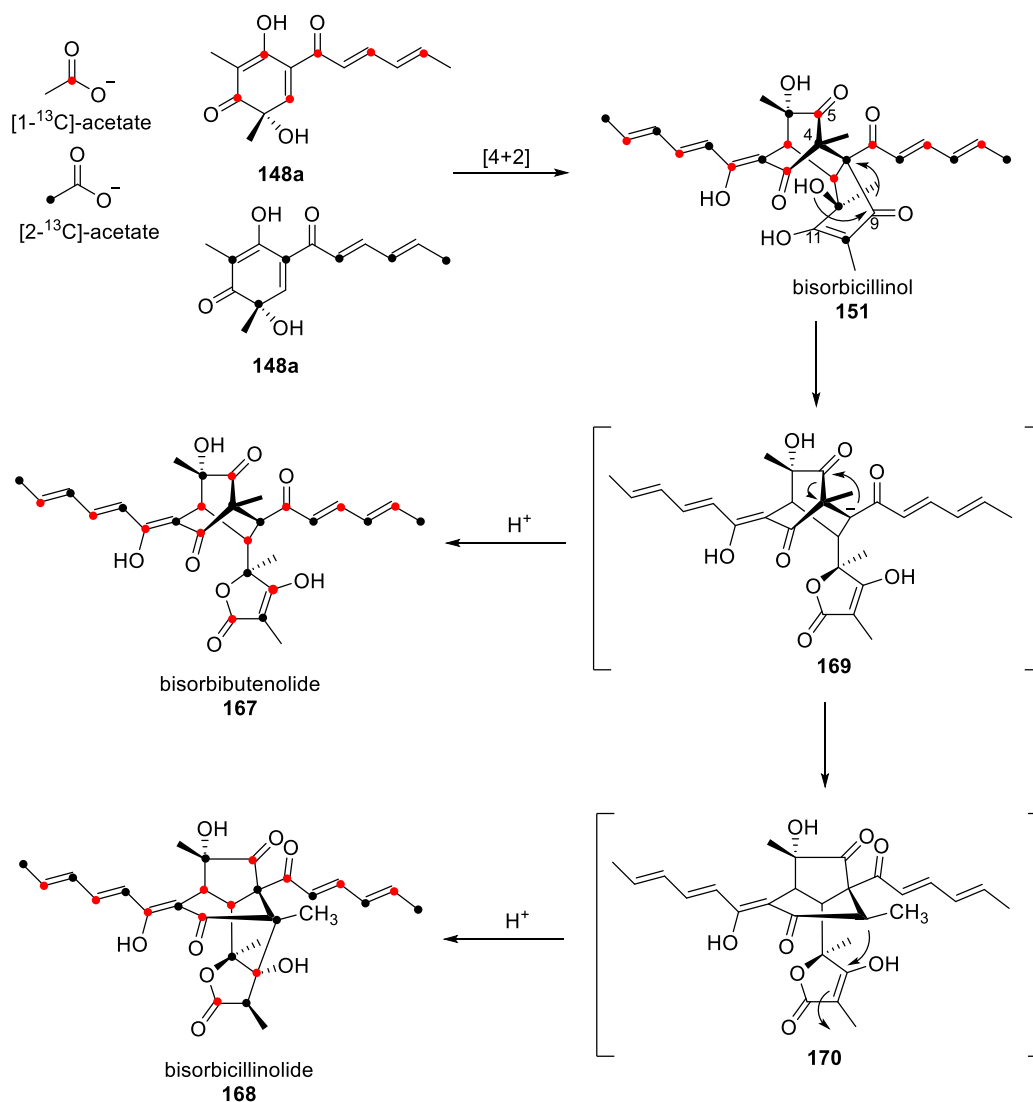
Scheme 1.5.3: Biosynthesis of **151** and rearrangement into **167** and **168** based on labelling studies by Abe *et al.*[297,299]

Interestingly, **151** was further converted into bisorbibutenolide **167** and bisorbicillinolide **168** in *Trichoderma* sp. USF-2690. Mechanistically, an initial nucleophilic attack of the C-12 hydroxyl group onto the C-9 ketone would give intermediate **169** that could form **167** by protonation of the C-8 carbanion. Alternatively, the C-8 carbanion could attack the C-5

ketone inducing a break between C-5 and C-4. The resulting C-4 carbanion in intermediate **170** would then attack the enolic C-9 position to yield **168** upon re-protonation (Scheme 1.5.3).

In contrast, Nicolaou *et al.*[279] obtained **151** in a biomimetic approach where racemic **148a** was first generated *in situ* by base hydrolysis (KOH in THF/ H$_2$O [9:1]; 0.05 M) of the corresponding acetyl-ester. Subsequent acidic quenching yielded racemic **151** with "remarkable regio- and diastereocontrol (*endo* selectivity)". This observation is supported by studies of Deng *et al.* who generated **151** by trifluoroacetic acid treatment (TFA in CH$_2$Cl$_2$) of optically pure *S-p*-methoxybenzyl (PMB)-protected sorbicillinol **171** (Scheme 1.5.4).[301]

Interestingly, while direct conversion of **151** into **167** was induced by treatment with the strong, non-nucleophilic base KHMDS (potassium bis[trimethylsilyl]amide),[279] simple incubation of **151** in MeOH for 48 hours led to formation of both **167** and **168** in 10% and 64% yield, respectively (Scheme 1.5.4).[301] These results both highlight the inherent reactivity of sorbicillinol **148a** towards intermolecular [4+2] cyclization as well as the apparent susceptibility of **151** towards intramolecular rearrangement upon storage in MeOH.
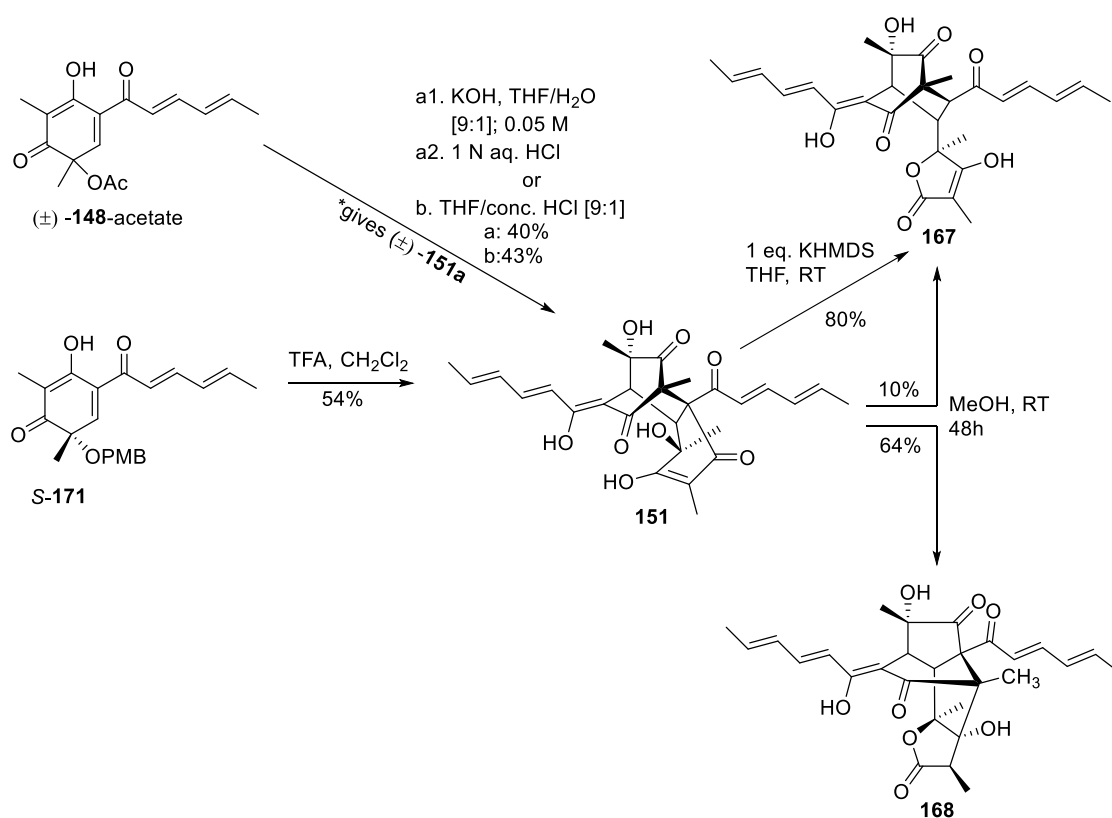
Scheme 1.5.4: Biomimetic synthesis of dimeric sorbicillinoids **151**, **167** and **168** based on *in situ* generation of **148a**.

The striking regioselectivity during the intermolecular Diels-Alder reaction of **148a** might be rationalized by invoking the "syn oxygen phenomenon" that has been experimentally investigated by Paquette *et al.*[302,303] The group demonstrated that the *endo*-regioisomer is preferably formed in which the transition state exhibits a syn-orientation of the heteroatom (here oxygen) towards the π-system of the diene.

With regards to these studies a transition-state reminiscent of **172,** in that the C-5 hydroxyl substituent is *syn* towards the newly forming bond, would be expected to precede the formation of **151** (Scheme 1.5.5).[265]
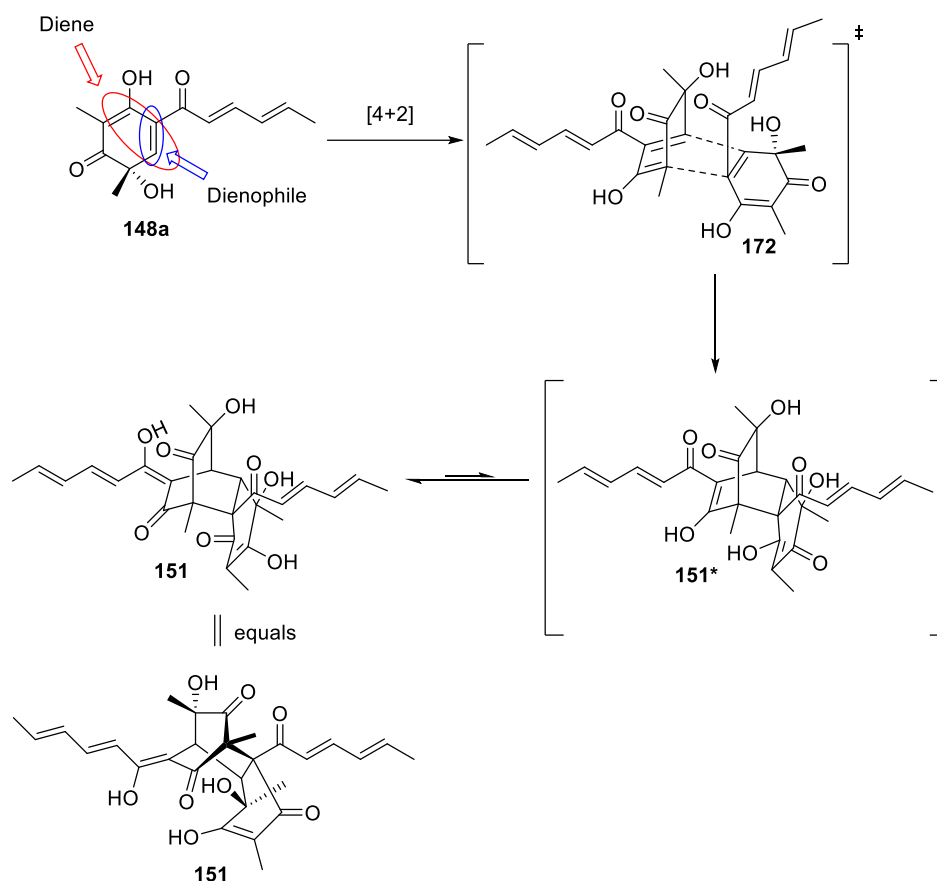
Scheme 1.5.5: Proposed transition state during the intermolecular Diels-Alder reaction of **148a** towards **151**.

During their screening of *Trichoderma* sp. USF-2690 towards the discovery of novel compounds with desirable biological activities the Abe group isolated the dimeric bisvertinolone **153** concomitantly with sorbicillinol **148a** and oxosorbicillinol **173**, along with other related metabolites. The authors realized the structural relationship between these three compounds and proposed a biosynthetic route from **148a** and **173** to **153** *via* a Michael-addition like mechanism (Scheme 1.5.6 A).[298] Initially, **148a** would act as a Michael-acceptor leading to the formation of intermediate **174**. Intramolecular ketalization and tautomerization would furnish the structure of **153**.

Feeding of [1-¹³C]-labelled acetate resulted in ¹³C-enrichment at the twelve expected positions that were in agreement with the proposed mechanism. Further evidence was obtained by feeding [1-¹³C]-labelled oxosorbicillinol **173** back into the fungal culture, yielding bisvertinolone **153** that showed ¹³C-enrichment only at the subunit that was expected to be derived from **173** (Scheme 1.5.6 B). One could speculate that the very similar dimeric bisvertinol **152** (lack of oxo-functionality) might follow the same biosynthetic route, utilizing an analogue of **173** that exhibits complete reduction at the C-6 position.
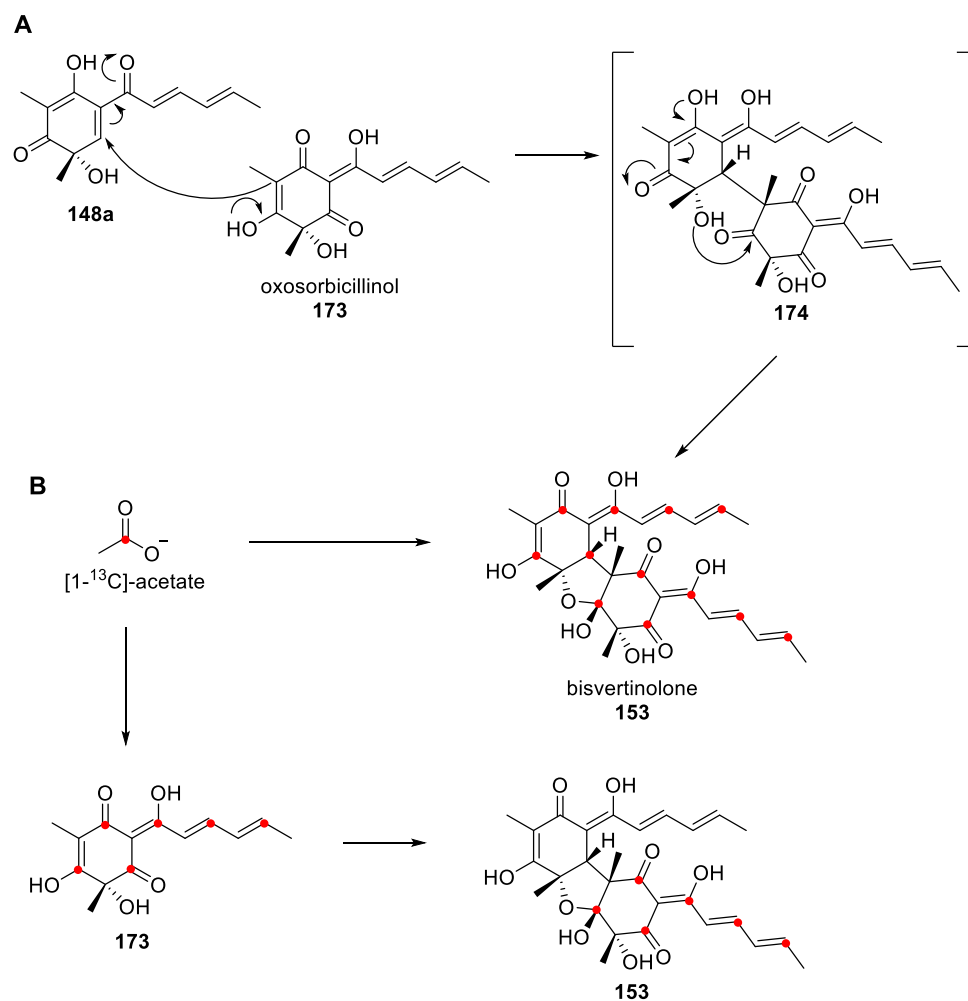
Scheme 1.5.6: Biosynthetic origin of bisvertinolone **153**. **A**, Proposed mechanism based on Michael-type addition between **148a** and **173** followed by intramolecular ketalization; **B**, Labelling studies demonstrating the incorporation of **173** into **153**.

In conclusion, pioneering labelling studies by the Abe group supported by biomimetic studies clearly demonstrated that the highly reactive sorbicillinol **148a** constitutes a key intermediate towards the formation of dimeric sorbicillinoids and might also serve as a precursor molecule to monomeric congeners.

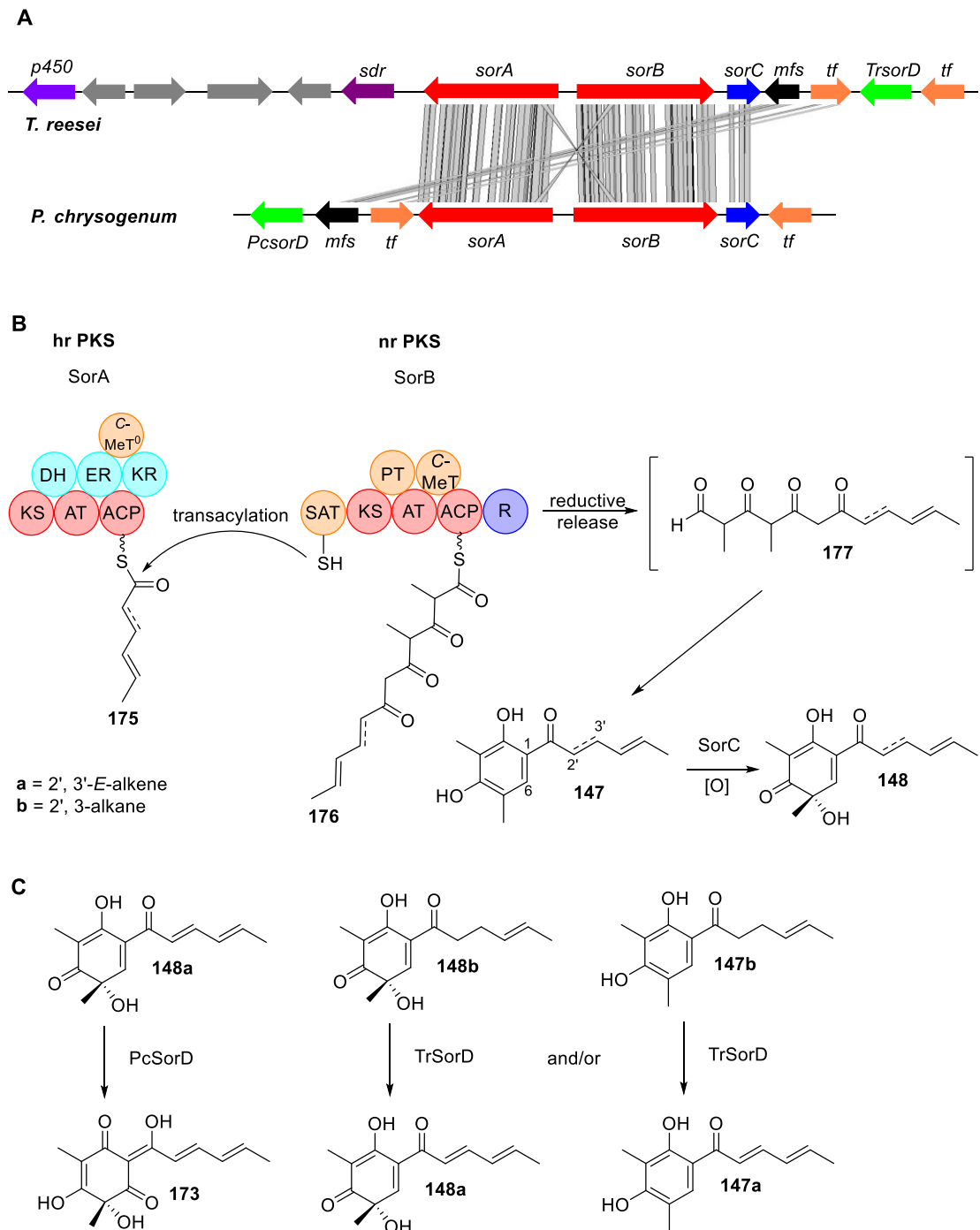### 1.5.3 Biosynthetic Gene Cluster and the Origin of Dimerization

Despite the previous elaborate studies the genetic basis of sorbicillinoid biosynthesis remained obscure for a long time. A breakthrough was achieved in 2014 when Cox and Simpson *et al.* identified the sorbicillinoid BGC in *P. chrysogenum* E01-10/3 encoding both a hrPKS (*sorA*) and a nrPKS (*sorB*).[274] In addition to the core domains of a typical nrPKS, SorB features a *C*-MeT domain as well as a reductive release (R) domain that are both in agreement with the chemistry that is required for in-line modification of any presumed PKS-tethered precursor towards sorbicillin **147**. The cluster furthermore encodes two transcription factors (TF), a transporter of the major facilitator superfamily (MFS), a FMO (SorC) and a second flavin-dependent enzyme classified as an oxidoreductase based on *in silico* analysis (henceforth PcSorD; Scheme 1.5.7 A).

Since all efforts to link the BGC to the production of sorbicillinoids by gene knockout of either PKS were unsuccessful the authors turned their attention to the FMO SorC. This enzyme shares high similarity to TropB, a FMO that catalyzes oxidative dearomatization during tropolone biosynthesis (section 1.3.1) and SorC therefore appeared a suitable candidate for the conversion of sorbicillin **147** to sorbicillinol **148**. In fact, when recombinant SorC was incubated with either **147a** or **147b**, rapid conversion into **148a** or **148b** was observed.

Combining their experimental results with bioinformatic analysis of the BGC the authors proposed a simplified biosynthetic route in which SorA produces the triketide **175** that is transferred onto the SAT-domain of SorB (Scheme 1.5.7 B). After chain extension the intermediate **176** would be released as a free aldehyde **177** by the R-domain of SorB and is readily subjected to Knoevenagel-condensation to yield the first isolatable product sorbicillin **147**, that is further oxidized into **148** by SorC. Depending on whether the ER-domain of SorA operates during the last iteration either a fully or partially unsaturated C-1/C-6 side chain would be installed. This proposal fits with all previously reported labelling studies (section 1.5.2) and circumvents the unlikely oxidation of the PKS-bound intermediate (**159** to **163**) to avoid formation of the symmetrical intermediate **162** (cf. Scheme 1.5.2).

Individual gene knockouts of both PKS in *P. chrysogenum* by the Driessen group[304,305] a few years later completely abolished production of any sorbicillinoid-related compound (*ΔsorA* or *ΔsorB*), in agreement with the biosynthetic proposal. Controversially, while the *ΔsorC* strain was reported to show diminished production of sorbicillinol **148a** and oxosorbicillinol **173**, the titre of dihydrosorbicillinol **148b** was 1.3 times higher compared to the wild type strain.[305] Knockout of the second flavin-dependent oxidase (*ΔPcsorD*)

almost completely abolished production of **173**, hence the role of PcSorD was assumed to oxidize **148a** into **173** (Scheme 1.5.7 C).[305]

Scheme 1.5.7: Biosynthesis of sorbicillin **147** and sorbicillinol **148**. **A**, Sorbicillinoid BGC in *T. reesei* and *P. chrysogenum*. Homology analysis was performed using Artemis.[306] Genes depicted in grey show either no conserved domains or are not likely to be involved in the biosynthetic pathway. A detailed gene analysis for the BGC of *T. reesei* is given in the appendix Table S2; **B**, Formation of **147** by the hr-nr-PKS pair SorA and SorB with ensuing oxidation into **148** by the FMO SorC; **C**, Proposed functions for the flavin-dependent oxidoreductases PcSorD and TrSorD.

In 2017 a homologous BGC was identified in *Trichoderma reesei* QM6a by Mach-Aigner *et al*. and confirmed to be responsible for the production of sorbicillinoids by gene knockout of

the homologous *sorA* and *sorC*.[307] Besides further homologous genes encoding for TF and MFS, the BGC in *T. reesei* also encodes for a second flavin-dependent enzyme (TrSorD; *in silico* oxidoreductase) that is not homologous to the flavin-dependent oxidase PcSorD found in the BGC of *P. chrysogenum* (Scheme 1.5.7 A; based on Artemis comparison).[306] TrSorD and PcSorD show 16.9 % pairwise identity by Clustal Omega[308] (18.7 % by global- and 19.4 % by local protein alignment in using the software Geneious). As a broad rule proteins sharing overall sequence identity >30 % are considered homologous.[309–311]

Gene Knockout of *TrsorD* led to accumulation of two compounds with a molecular mass matching that of dihydrosorbicillinol **148b**, along with a reduced titre of sorbicillinol **148a**. The authors therefore suggested that TrSorD would be able to oxidize the 2',3'-alkane **148b** into the 2',3'-*E*-alkene **148a** and/or **147b** into **147a**, respectively (Scheme 1.5.7 C).[307] Further genes that are potentially involved during the biosynthesis are encoding for a SDR and a P450-dependent protein, respectively (Scheme 1.5.7 A).

Both groups only lightly linked their gene knockout results to the formation of dimeric sorbicillinoids, although both were in agreement that sorbicillinol **148a** constitutes the key biosynthetic intermediate. In addition to a lacking production of oxosorbicillinol **173** Driessen *et al.* furthermore could not detect the dimeric bisvertinolone **153** and corresponding 2', 3'-reduced congeners in the *ΔPcsorD* strain (*P. chrysogenum*), in line with previous studies by the Abe group that showed incorporation of **173** into **153**.[298,305] For the *ΔTrsorD* strain (*T. reesei*) Mach-Aigner *et al.* additionally reported a diminished production of bisorbicillinol **151** and bisvertinolone **153**, that was attributed to the low production of the precursor sorbicillinol **148a** in the same strain.[307] The involvement of an enzyme catalyst during the formation of dimeric sorbicillinoids could be neither confirmed nor ruled out.

Very recently, the Gulder group reported the total synthesis of several (un)natural dimeric sorbicillinoids based on SorC-dependent stereoselective oxidative dearomatization of sorbicillin **147** and unnatural substrate analogues *in vitro*.[312–314] Remarkably, dimerization of enzymatically formed sorbicillinol **148a** (and corresponding analogues) was induced and controlled by the organic co-solvent added to the reaction as well as by the subsequent chemical workup protocol. By using acetone as the co-solvent (to solubilize the substrate **147a**) *in vitro* formation of **148a** yielded the Diels-Alder derived bisorbicillinol **151** after extraction with dichloromethane as the only dimeric product (Scheme 1.5.8 A).[312] In contrast, when the co-solvent was changed to the polar dimethylformamide (DMF), the major product was the cage-like trichodimerol **154** with concomitant formation of **151** as the minor product.[312]
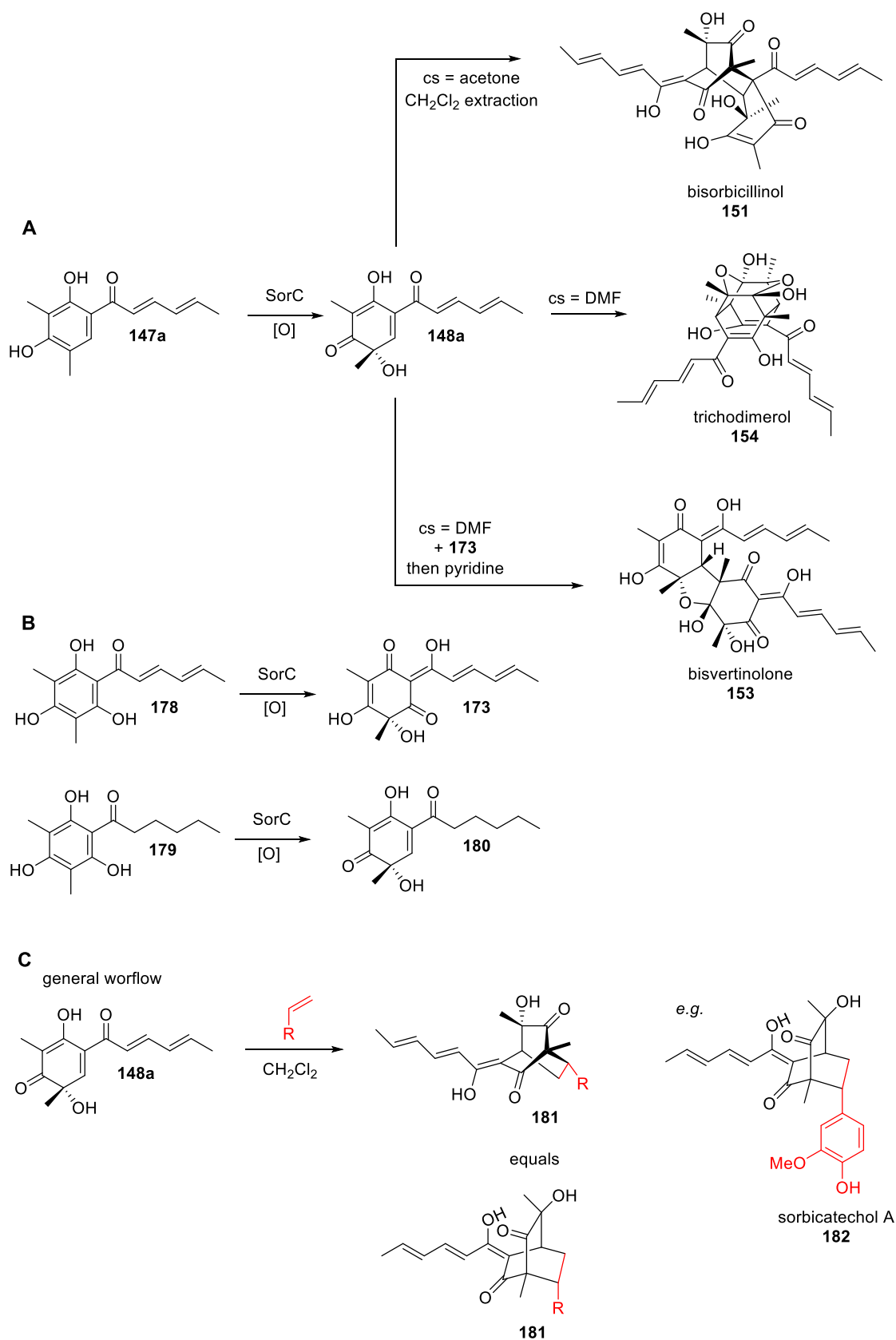
Scheme 1.5.8: Chemoenzymatic synthesis of dimeric sorbicillinoids described by Gulder *et al.*[312–314] **A**, Enzymatic formation of sorbicillinol **148a** *in vitro* results in formation of different dimeric sorbicillinoids depending on the chemical workup conditions; cs = co-solvent; **B**, Alternative substrates for SorC; **C**, Synthesis of hybrid sorbicillinoids by co-incubation of **148a** with a general dienophile and treatment with dichloromethane.

The Gulder[312,313] and Narayan groups[315] further demonstrated that SorC accepts substrate analogues that may differ in the substitution pattern along the aromatic ring (**178**) or exhibit saturation of the side chain (**179**, Scheme 1.5.8 B). By subjecting the saturated analogue **180** to the conditions described above, the corresponding Diels-Alder and Michael-type dimers were obtained.[312] Notably, **178** was directly oxidized into oxosorbicillinol **173** by SorC. Co-incubation of **173** with **148a** in DMF and pyridine yielded the Michael-like dimer bisvertinolone **153**.[313]

The potential of this "one-pot chemoenzymatic approach" towards the construction of novel sorbicillinoids was demonstrated by addition of various dienophiles to SorC-derived **148a**, followed by workup with dichloromethane. This simple workflow succeeded in the formation of 15 hybrid sorbicillinoids derived from [4+2] cycloaddition including sorbicatechol A **182** (Scheme 1.5.8 C).[314] As with bisorbicillinol **151** these compounds were shown to be the *endo* diastereomers.

The spontaneous formation of various dimeric/ hybrid sorbicillinoids without any additional catalyst again questions the origin of dimeric sorbicillinoids isolated from natural sources. Synthetic and biomimetic studies imply that these dimers represent artefacts that are formed during the chemical extraction procedure due to the inherent reactivity of sorbicillinol **148**. Still, nature might also use an undiscovered biocatalyst to direct the selective formation of specific dimers *in vivo*.

In fact, a recent survey suggests that more than 400 natural products potentially involve a [4+2] cycloaddition reaction during their biosynthesis.[316] However, the *in silico* prediction and identification of dedicated Diels-Alderases proves difficult due to a lack of sequence homology. Although enzyme-catalysed Diels-Alder reactions are expected to proceed with high regio- and diastereoselectivity[317] this is not a conclusive evidence as shown for the spontaneous formation of dimeric sorbicillinoids like bisorbicillinol **151**. Nevertheless, during the last decades several natural Diels-Alderases have been described *e.g.* for the biosyntheses of lovastatin,[318] pyrichalasin H[319], spinosyn A[320] and the fungal meroterpenoid xenovulene A.[186]
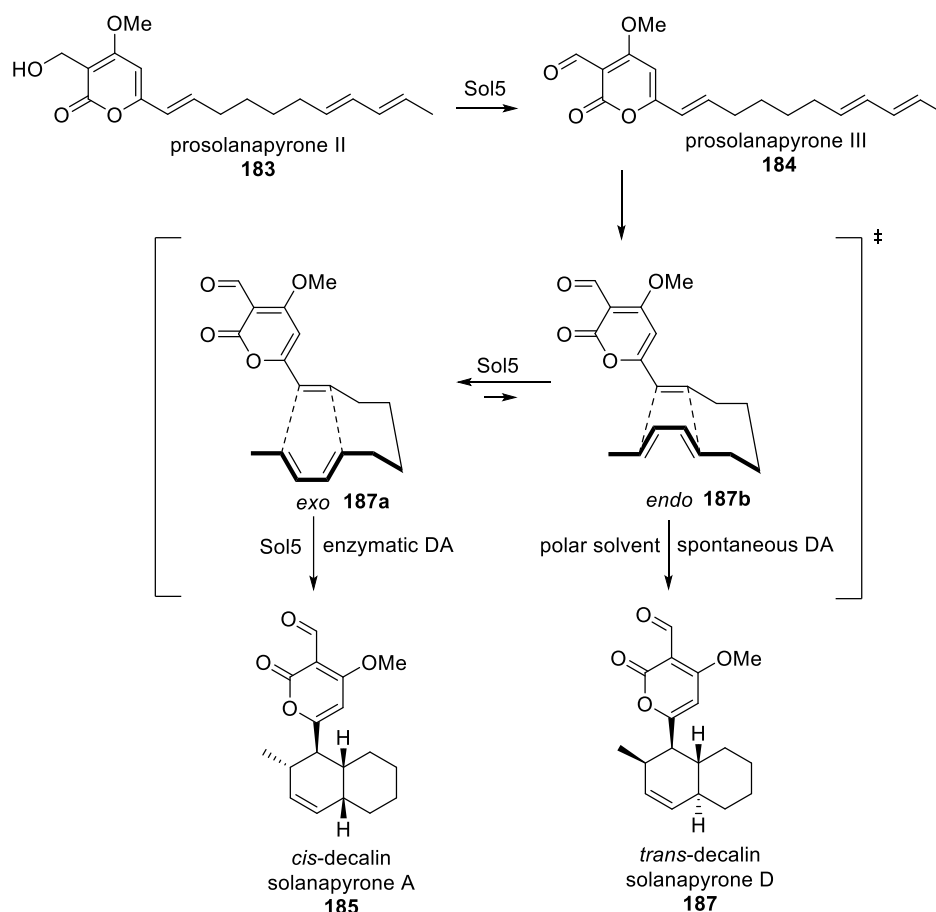
Scheme 1.5.9: Enzyme-assisted Diels-Alder reaction during the formation of solanapyrone A **185**.

A notable example is the flavin-dependent oxidase Sol5 that does not only oxidize the hydroxyl-substituent in prosolanapyrone II **183** to the aldehyde prosolanapyrone III **184**, but also governs the subsequent intramolecular [4+2] cycloaddition to give the *exo* diastereomer solanapyrone A **185** (Scheme 1.5.9).[70,71] In the absence of Sol5 **184** undergoes spontaneous Diels-Alder reaction to primarily yield the *endo* diastereomer solanapyrone D **186** (23:1 in H$_2$O).[321] Sol5 is proposed to deliver a stereochemical template that traps the reactive intermediate in an *exo* transition state **187a** as opposed to the kinetically favoured *endo* transition state **187b**.[70] Based on the *in situ* classification of Sol5 as a flavin-dependent oxidase one would not perceive its additional Diels-Alderase activity without experimental data.

1.6 Research Aims

The overall focus of this thesis is dedicated to decipher the biosyntheses of terrein and dimeric sorbicillinoids. Despite previous research the catalytic key steps in both biosynthetic pathways remain obscure.

Although the terrein BGC in *A. terreus* had already been identified, knockout studies in the producing strain have not been able to shed light on the oxidative transformation of the precursor molecule 6-hydroxymellein **113** into terrein **94**. Such oxidative ring contractions (and ring expansions) are mechanistically very interesting and further examples from the terrein pathway will help to illuminate this under-investigated area. By utilizing the complementary approach, heterologous expression in the fungal host *A. oryzae* NSAR1, new insights into the catalytic role of the oxidative enzymes TerC, TerD, TerE as well as the uncharacterized Kelch protein TerF might be obtained. Due the potential reactivity of any novel intermediates further *in vitro* studies with isolated enzymes appear expedient. Recombinant enzymes will therefore be obtained from overexpression in suitable *E. coli* strains.

In addition, the formation of 6-hydroxymellein **113** itself needs further consideration. While previous experiments clearly showed that an nrPKS (TerA) and an unusual DH-KR di-domain protein (TerB) are required for production of **113**, mechanistic details are lacking. Deliberate point mutations of catalytic key residues within both enzymes followed by *in vivo* expression will be designed to tackle the question whether TerB acts on the released pentaketide or constitutes a new member of *trans*-acting enzymes that operates *during* chain elongation of the PKS-tethered intermediate. While some fungal PKS are known to interact with *trans*-acting proteins, a multidomain *trans*-acting enzyme such as TerB, whose architecture is reminiscent of a truncated hrPKS, has not yet been characterized.

Through intermolecular dimerization with itself or other suitable small molecules sorbicillinol **148a** is the undoubted precursor molecule of various dimeric and hybrid sorbicillinoids. Despite elaborate research the involvement of an *in vivo* catalyst during these prominent dimerization reactions is still at issue. The inherent reactivity of **148a** towards dimerization was impressively demonstrated by the Gulder group and is in line with early biomimetic studies. With further regards to the intramolecular rearrangement of bisorbicillinol **151a** that proceeds upon storage in methanol one is tempted to believe that the direction of spontaneous dimerization is dependent on the chemical workup conditions.

Initially, heterologous expression of *sorABC* in *A. oryzae* should reconstitute the biosynthetic pathway up to sorbicillinol **148a** and its tendency towards spontaneous dimerization with respect to the extraction methods will be investigated. Since with *T. reesei* QM6a our group has a prolific producer of various sorbicillinoids in-hand, the BGC of this strain will used for amplification of target genes. Upon established production of **148a** attention will be directed to survey an additional enzyme that is required for dimerization cascades *in vivo*, if any.

Furthermore, conclusive evidence for the activity of the two flavin-dependent enzymes TrSorD (*T. reesei*) and PcSorD (*P. chrysogenum*) are lacking, meaning that the two best-studied sorbicillinoid BGC are still not fully understood. Both TrSorD and PcSorD are therefore potential candidates as possible Diels-Alderases. Although both proteins are generally classified as flavin-dependent oxidases based on phylogenetic analysis, either proposed function *i.e.* oxidation of **148a** into **173** (insertion of oxygen is expected for a monooxygenase) or oxidation of **148b** into **148a** (oxidation of alkane into alkene) is rather unexpected for a classical flavin-dependent oxidoreductase and needs further inspection.

1.7 References

[1]   A. A. Brakhage, *Nat. Rev. Microbiol.* **2013**, *11*, 21–32.

[2]   P. Vaishnav, A. L. Demain, *Biotechnol. Adv.* **2011**, *29*, 223–229.

[3]   G. Yim, H. H. Wang, J. Davies, *Philos. Trans. R. Soc. B Biol. Sci.* **2007**, *362*, 1195–1200.

[4]   D. Hoffmeister, N. P. Keller, *Nat. Prod. Rep.* **2007**, *24*, 393–416.

[5]   L. M. Forrester, G. E. Neal, D. J. Judah, M. J. Glancey, C. R. Wolf, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8306–8310.

[6]   P. J. Cotty, R. Jaime-Garcia, *Int. J. Food Microbiol.* **2007**, *119*, 109–115.

[7]   A. Fleming, *On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzae*, Wiley-Blackwell, **1929**.

[8]   N. Kardos, A. L. Demain, *Appl. Microbiol. Biotechnol.* **2011**, *92*, 677–687.

[9]   D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–335.

[10]   N. P. Keller, G. Turner, J. W. Bennett, *Nat. Rev. Microbiol.* **2005**, *3*, 937–947.

[11]   J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416.

[12]   J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas, C. R. Hutchinson, *Science.* **1999**, *284*, 1368–1372.

[13]   R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.

[14]   D. H. Scharf, T. Heinekamp, N. Remme, P. Hortschansky, A. A. Brakhage, C. Hertweck, *Appl. Microbiol. Biotechnol.* **2012**, *93*, 467–472.

[15]   M. B. Quin, C. M. Flynn, C. Schmidt-Dannert, *Nat. Prod. Rep.* **2014**, *31*, 1449–1473.

[16]   J. M. Caruthers, I. Kang, M. J. Rynkiewicz, D. E. Cane, D. W. Christianson, *J. Biol. Chem.* **2000**, *275*, 25533–25539.

[17]   A. Minami, N. Tajima, Y. Higuchi, T. Toyomasu, T. Sassa, N. Kato, T. Dairi, *Bioorganic Med. Chem. Lett.* **2009**, *19*, 870–874.

[18]   J. Ziegler, P. J. Facchini, *Annu. Rev. Plant Biol.* **2008**, *59*, 735–769.

[19]   J. Fricke, F. Blei, D. Hoffmeister, *Angew. Chemie - Int. Ed.* **2017**, *56*, 12352–12355.

[20]   A. Numata, C. Takahashi, Y. Ito, T. Takada, K. Kawai, Y. Usami, E. Matsumura, M. Imachi, T. Ito, T. Hasegawa, *Tetrahedron Lett.* **1993**, *34*, 2355–2358.

[21]   M. A. Fischbach, C. T. Walsh, *Science* **2009**, *325*, 1089–1093.

[22]   C. T. Walsh, M. A. Fischbach, *J. Am. Chem. Soc.* **2010**, *132*, 2469–2493.

[23]   A. J. Birch, *Annu. Rev. Plant Physiol.* **1968**, *19*, 321–332.

[24]   J. M. Crawford, C. A. Townsend, *Nat. Rev. Microbiol.* **2010**, *8*, 879–889.

[25]   A. J. Birch, R. A. Massy-Westropp, C. J. Moye, *Aust. J. Chem.* **1955**, *8*, 539–544.

[26]   Z. Liu, J. Cheema, M. Vigouroux, L. Hill, J. Reed, P. Paajanen, L. Yant, A. Osbourn, *Nat. Commun.* **2020**, *11*, 1–11.

[27]   N. P. Keller, T. M. Hohn, *Fungal Genet. Biol.* **1997**, *21*, 17–29.

[28] J. W. Bok, D. Noordermeer, S. P. Kale, N. P. Keller, *Mol. Microbiol.* **2006**, *61*, 1636–1645.

[29] R. J. Cox, *Org. Biomol. Chem.* **2007**, *5*, 2010–2026.

[30] P. Schläpfer, P. Zhang, C. Wang, T. Kim, M. Banf, L. Chae, K. Dreher, A. K. Chavali, R. Nilo-Poyanco, T. Bernard, et al., *Plant Physiol.* **2017**, *173*, 2041–2059.

[31] J. C. Slot, in *Adv. Genet.*, Academic Press Inc., **2017**, pp. 141–178.

[32] N. Khaldi, J. Collemare, M. H. Lebrun, K. H. Wolfe, *Genome Biol.* **2008**, *9*, 1–10.

[33] J. W. Bok, N. P. Keller, *Eukaryot. Cell* **2004**, *3*, 527–535.

[34] J. D. Walton, *Fungal Genet. Biol.* **2000**, *30*, 167–171.

[35] J. Shendure, E. L. Aiden, *Nat. Biotechnol.* **2012**, *30*, 1084–1094.

[36] M. H. Medema, K. Blin, P. Cimermancic, V. De Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucleic Acids Res.* **2011**, *39*, W339–W346.

[37] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81–W87.

[38] M. Stanke, B. Morgenstern, *Nucleic Acids Res.* **2005**, *33*, W465–W467.

[39] V. Solovyev, P. Kosarev, I. Seledsov, D. Vorobyev, *Genome Biol.* **2006**, *7 Suppl 1*, DOI 10.1186/gb-2006-7-s1-s10.

[40] S. A. Kautsar, K. Blin, S. Shaw, J. C. Navarro-Muñoz, B. R. Terlouw, J. J. J. Van Der Hooft, J. A. Van Santen, V. Tracanna, H. G. Suarez Duran, V. Pascal Andreu, et al., *Nucleic Acids Res.* **2020**, *48*, D454–D458.

[41] J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, et al., *Nat. Chem. Biol.* **2020**, *16*, 60–68.

[42] A. J. Birch, F. W. Donovan, *Aust. J. Chem.* **1953**, *6*, 360–368.

[43] J. N. Collie, *J. Chem. Soc., Trans.* **1907**, *91*, 1806–1813.

[44] N. Collie, W. S. Myers, *J. Chem. Soc., Trans.* **1893**, *63*, 122–128.

[45] R. J. Cox, J. Crosby, O. Daltrop, F. Glod, M. E. Jarzabek, T. P. Nicholson, M. Reed, T. J. Simpson, L. H. Smith, F. Soulas, et al., *J. Chem. Soc. Perkin 1* **2002**, *14*, 1644–1649.

[46] J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart, *Nat. Prod. Rep.* **2014**, *31*, 61–108.

[47] M. Hunkeler, E. Stuttfeld, A. Hagmann, S. Imseng, T. Maier, *Nat. Commun.* **2016**, *7*, 1–11.

[48] J. Beld, D. J. Lee, M. D. Burkart, *Mol. Biosyst.* **2015**, *11*, 38–59.

[49] S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolásek, N. Ban, *Science* **2007**, *316*, 254–261.

[50] T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315–1322.

[51] S. W. White, J. Zheng, Y. M. Zhang, C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791–831.

[52] C. E. Christensen, B. B. Kragelund, P. von Wettstein-Knowles, A. Henriksen, *Protein Sci.* **2006**, *16*, 261–272.

[53] R. Biswas, A. Dutta, D. Dutta, D. Hazra, D. R. Banerjee, A. Basak, A. K. Das, *Biochem. Biophys. Res. Commun.* **2015**, *458*, 369–374.

[54] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlic, P. W. Rose, *Bioinformatics* **2018**, *34*, 3755–3758.

[55] S.-C. C. Tsai, *Annu. Rev. Biochem.* **2018**, *87*, 503–531.

[56] M. B. Austin, J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79–110.

[57] Y. H. Chooi, Y. Tang, *J. Org. Chem.* **2012**, *77*, 9933–9953.

[58] C. Hertweck, *Angew. Chemie - Int. Ed.* **2009**, *48*, 4688–4716.

[59] R. Nofiani, K. de Mattos-Shipley, K. E. Lebe, L. C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson, R. J. Cox, *Nat. Commun.* **2018**, *9*, 3940.

[60] P. A. Storm, D. A. Herbst, T. Maier, C. A. Townsend, *Cell Chem. Biol.* **2017**, *24*, 316–325.

[61] Y. Xu, T. Zhou, Z. Zhou, S. Su, S. A. Roberts, W. R. Montfort, J. Zeng, M. Chen, W. Zhang, M. Lin, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5398–5403.

[62] Y. Li, W. Xu, Y. Tang, *J. Biol. Chem.* **2010**, *285*, 22764–22773.

[63] L. Liu, Z. Zhang, C. L. Shao, J. L. Wang, H. Bai, C. Y. Wang, *Sci. Rep.* **2015**, *5*, 10463.

[64] L. Liu, Z. Zhang, C. L. Shao, C. Y. Wang, *Front. Microbiol.* **2017**, *8*, 1685.

[65]   J. Davison, A. Al Fahad, M. Cai, Z. Song, S. Y. Yehia, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7642–7647.

[66]   R. J. Cox, E. Skellam, K. Williams, in *Physiol. Genet.*, Springer International Publishing, **2018**, pp. 385–412.

[67]   T. Moriguchi, Y. Ebizuka, I. Fujii, *ChemBioChem* **2006**, *7*, 1869–1874.

[68]   T. Moriguchi, Y. Kezuka, T. Nonaka, Y. Ebizuka, I. Fujii, *J. Biol. Chem.* **2010**, *285*, 15637–15643.

[69]   D. M. Roberts, C. Bartel, A. Scott, D. Ivison, T. J. Simpson, R. J. Cox, *Chem. Sci.* **2017**, *8*, 1116–1126.

[70]   K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, I. Fujii, *ChemBioChem* **2010**, *11*, 1245–1252.

[71]   R. Fujii, T. Ugai, H. Ichinose, M. Hatakeyama, T. Kosaki, K. Gomi, I. Fujii, A. Minami, H. Oikawa, *Biosci. Biotechnol. Biochem.* **2016**, *80*, 426–431.

[72]   J. Wang, J. Liang, L. Chen, W. Zhang, L. Kong, C. Peng, C. Su, Y. Tang, Z. Deng, Z. Wang, *Nat. Commun.* **2021**, *12*, 867.

[73]   P. Caffrey, *ChemBioChem* **2003**, *4*, 654–657.

[74]   A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 277–285.

[75]   A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[76]   I. Abe, *Nat. Chem. Biol.* **2012**, *8*, 322–323.

[77]   H. Zhou, Z. Gao, K. Qiao, J. Wang, J. C. Vederas, Y. Tang, *Nat. Chem. Biol.* **2012**, *8*, 331–333.

[78]   A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *J. Am. Chem. Soc.* **2011**, *133*, 10990–10998.

[79]   B. D. Ames, C. Nguyen, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. H. Li, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11144–11149.

[80]   A. O. Zabala, Y. H. Chooi, M. S. Choi, H. C. Lin, Y. Tang, *ACS Chem. Biol.* **2014**, *9*, 1576–1586.

[81]   M. Chen, Q. Liu, S. S. Gao, A. E. Young, S. E. Jacobsen, Y. Tang, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 5499–5504.

[82]   S. Kishimoto, Y. Tsunematsu, T. Matsushita, K. Hara, H. Hashimoto, Y. Tang, K. Watanabe, *Biochemistry* **2019**, *58*, 3933–3937.

[83]   U. Rix, C. Fischer, L. L. Remsing, J. Rohr, *Nat. Prod. Rep.* **2002**, *19*, 542–580.

[84]   C. Olano, C. Méndez, J. A. Salas, *Nat. Prod. Rep.* **2010**, *27*, 571–616.

[85]   B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Chem. Commun.* **2016**, *52*, 6777–6780.

[86]   X. M. Mao, Z. J. Zhan, M. N. Grayson, M. C. Tang, W. Xu, Y. Q. Li, W. B. Yin, H. C. Lin, Y. H. Chooi, K. N. Houk, et al., *J. Am. Chem. Soc.* **2015**, *137*, 11904–11907.

[87]   R. Geris, T. J. Simpson, *Nat. Prod. Rep.* **2009**, *26*, 1063–1094.

[88]   T. Itoh, K. Tokunaga, E. K. Radhakrishnan, I. Fujii, I. Abe, Y. Ebizuka, T. Kushiro, *ChemBioChem* **2012**, *13*, 1132–1135.

[89]   C. Olano, C. Méndez, J. A. Salas, *Nat. Prod. Rep.* **2010**, *27*, 571–616.

[90]   K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, *Cell. Mol. Life Sci.* **2008**, *65*, 3895–3906.

[91]   L. Liu, M. C. Tang, Y. Tang, *J. Am. Chem. Soc.* **2020**, *141*, 19538–19541.

[92]   J. B. Van Beilen, W. A. Duetz, A. Schmid, B. Witholt, *Trends Biotechnol.* **2003**, *21*, 170–177.

[93]   T. D. H. Bugg, *Tetrahedron* **2003**, *59*, 7075–7101.

[94]   E. Romero, J. R. Gómez Castellanos, G. Gadda, M. W. Fraaije, A. Mattevi, *Chem. Rev.* **2018**, *118*, 1742–1769.

[95]   B. G. Malmström, *Annu. Rev. Biochem.* **1982**, *51*, 21–59.

[96]   M. C. Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang, *Chem. Rev.* **2017**, *117*, 5226–5333.

[97]   S. S. Gao, N. Naowarojna, R. Cheng, X. Liu, P. Liu, *Nat. Prod. Rep.* **2018**, *35*, 792–837.

[98]   L. C. Blasiak, F. H. Vaillancourt, C. T. Walsh, C. L. Drennan, *Nature* **2006**, *440*, 368–371.

[99]   S. Martinez, R. P. Hausinger, *J. Biol. Chem.* **2015**, *290*, 20702–20711.

[100] P. R. Ortiz De Montellano, *Chem. Rev.* **2010**, *110*, 932–948.

[101] B. Meunier, S. P. de Visser, S. Shaik, *Chem. Rev.* **2004**, *104*, 3947–3980.

[102] C. T. Walsh, Y. -C J. Chen, *Angew. Chemie Int. Ed. English* **1988**, *27*, 333–343.

[103] M. Toplak, A. Matthews, R. Teufel, *Arch. Biochem. Biophys.* **2021**, *698*, 108732.

[104] P. Chaiyen, M. W. Fraaije, A. Mattevi, *Trends Biochem. Sci.* **2012**, *37*, 373–380.

[105] A. Mattevi, *Trends Biochem. Sci.* **2006**, *31*, 276–283.

[106] S. G. Mayhew, *Eur. J. Biochem.* **1999**, *265*, 698–702.

[107] J. Vervoort, W. J. H. van Berkel, F. Müller, C. T. W. Moonen, *Eur. J. Biochem.* **1991**, *200*, 731–738.

[108] A. C. Campbell, K. M. Stiers, J. S. Martin Del Campo, R. Mehra-Chaudhary, P. Sobrado, J. J. Tanner, *J. Biol. Chem.* **2020**, *295*, 13239–13250.

[109] S. Visitsatthawong, P. Chenprakhon, P. Chaiyen, P. Surawatanawong, *J. Am. Chem. Soc.* **2015**, *137*, 9363–9374.

[110] A. Alfieri, F. Fersini, N. Ruangchan, M. Prongjit, P. Chaiyen, A. Mattevi, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1177–1182.

[111] J. P. Roth, J. P. Klinman, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 62–67.

[112] C. Tongsook, J. Sucharitakuls, K. Thotsaporns, P. Chaiyen, *J. Biol. Chem.* **2011**, *286*, 44491–44502.

[113] C. Sanner, P. Macheroux, H. Rüterjans, F. Müller, A. Bacher, *Eur. J. Biochem.* **1991**, *196*, 663–672.

[114] P. Macheroux, *Methods Mol. Biol.* **1999**, *131*, 1–7.

[115] A. Bacher, S. Eberhardt, M. Fischer, K. Kis, G. Richter, *Annu. Rev. Nutr.* **2000**, *20*, 153–167.

[116] M. Fischer, A. Bacher, *Nat. Prod. Rep.* **2005**, *22*, 324–350.

[117] P. Halada, C. Leitner, P. Sedmera, D. Haltrich, J. Volc, *Anal. Biochem.* **2003**, *314*, 235–242.

[118] T. Rand, T. Halkier, O. C. Hansen, *Biochemistry* **2003**, *42*, 7188–7194.

[119] L. Motteran, M. S. Pilone, G. Molla, S. Ghisla, L. Pollegioni, *J. Biol. Chem.* **2001**, *276*, 18024–18030.

[120] I. Efimov, C. N. Cronin, W. S. McIntire, *Biochemistry* **2001**, *40*, 2155–2166.

[121] T. A. Ewing, M. W. Fraaije, A. Mattevi, W. J. H. van Berkel, *Arch. Biochem. Biophys.* **2017**, *632*, 104–117.

[122] C. H. Huang, W. L. Lai, M. H. Lee, C. J. Chen, A. Vasella, Y. C. Tsai, S. H. Liaw, *J. Biol. Chem.* **2005**, *280*, 38831–38838.

[123] D. P. H. M. Heuts, N. S. Scrutton, W. S. McIntire, M. W. Fraaije, *FEBS J.* **2009**, *276*, 3405–3427.

[124] N. S. Scrutton, L. C. Packman, F. S. Mathews, R. J. Rohlfs, R. Hille, *J. Biol. Chem.* **1994**, *269*, 13942–13950.

[125] I. Efimov, C. N. Cronin, D. J. Bergmann, V. Kuusk, W. S. McIntire, *Biochemistry* **2004**, *43*, 6138–6148.

[126] A. Hassan-Abdallah, R. C. Bruckner, G. Zhao, M. S. Jorns, *Biochemistry* **2005**, *44*, 6452–6462.

[127] J. Jin, H. Mazon, R. H. T. Van Den Heuvel, A. J. Heck, D. B. Janssen, M. W. Fraaije, *FEBS J.* **2008**, *275*, 5191–5200.

[128] D. E. Edmondson, S. Ghisla, *Flavin and Flavoproteins* **1999**, *76*, 71–76.

[129] M. W. Fraaije, R. H. H. Van Den Heuvel, W. J. H. Van Berkel, A. Mattevi, *J. Biol. Chem.* **1999**, *274*, 35514–35520.

[130] D. P. H. M. Heuts, E. W. Van Hellemond, D. B. Janssen, M. W. Fraaije, *J. Biol. Chem.* **2007**, *282*, 20283–20291.

[131] R. K. Nandigama, D. E. Edmondson, *J. Biol. Chem.* **2000**, *275*, 20527–20532.

[132] M. Eschenbrenner, L. J. Chlumsky, P. Khanna, F. Strasser, M. S. Jorns, *Biochemistry* **2001**, *40*, 5352–5367.

[133] M. H. Hefti, J. Vervoort, W. J. H. Van Berkel, *Eur. J. Biochem.* **2003**, *270*, 4227–4242.

[134] X. Lu, D. Nikolic, D. J. Mitchell, R. B. Van Breemen, J. A. Mersfelder, R. Hille, R. B. Silverman, *Bioorganic Med. Chem. Lett.* **2003**, *13*, 4129–4132.

[135] W. J. H. van Berkel, N. M. Kamerbeek, M. W. Fraaije, *J. Biotechnol.* **2006**, *124*, 670–689.

[136] W. B. Sutton, *J. Biol. Chem.* **1957**, *226*, 395–405.

[137] A. Pennati, G. Gadda, *Biochemistry* **2011**, *50*, 1–3.

[138] R. Teufel, A. Miyanaga, Q. Michaudel, F. Stull, G. Louie, J. P. Noel, P. S. Baran, B. Palfey, B. S. Moore, *Nature* **2013**, *503*, 552–556.

[139] R. Teufel, F. Stull, M. J. Meehan, Q. Michaudel, P. C. Dorrestein, B. Palfey, B. S. Moore, *J. Am. Chem. Soc.* **2015**, *137*, 8078–8085.

[140] S. Adak, T. P. Begley, *J. Am. Chem. Soc.* **2016**, *138*, 6424–6426.

[141]   S. Adak, T. P. Begley, *Biochemistry* **2017**, *56*, 3708–3709.

[142]   T. Hiromoto, S. Fujiwara, K. Hosokawa, H. Yamaguchi, *J. Mol. Biol.* **2006**, *364*, 878–896.

[143]   R. Baron, C. Riley, P. Chenprakhon, K. Thotsaporn, R. T. Winter, A. Alfieri, F. Forneris, W. J. H. Van Berkel, P. Chaiyen, M. W. Fraaije, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10603–10608.

[144]   D. Zafred, B. Steiner, A. R. Teufelberger, A. Hromic, P. A. Karplus, C. J. Schofield, S. Wallner, P. Macheroux, *FEBS J.* **2015**, *282*, 3060–3074.

[145]   O. Dym, D. Eisenberg, *Protein Sci.* **2001**, *10*, 1712–1728.

[146]   C. A. Bottoms, P. E. Smith, J. J. Tanner, *Protein Sci.* **2009**, *11*, 2125–2137.

[147]   P. Laurino, Á. Tóth-Petróczy, R. Meana-Pañeda, W. Lin, D. G. Truhlar, D. S. Tawfik, *PLoS Biol.* **2016**, *14*, e1002396.

[148]   W. P. Dijkman, G. De Gonzalo, A. Mattevi, M. W. Fraaije, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 5177–5188.

[149]   J. Büchler, A. Papadopoulou, R. Buller, *Catalysts* **2019**, *9*, 1030.

[150]   C. Dong, S. Flecks, S. Unversucht, C. Haupt, K. H. Van Pée, J. H. Naismith, *Science* **2005**, *309*, 2216–2219.

[151]   E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan, C. T. Walsh, *Biochemistry* **2007**, *46*, 1284–1292.

[152]   X. Zhu, W. De Laurentis, K. Leang, J. Herrmann, K. Ihlefeld, K. H. van Pée, J. H. Naismith, *J. Mol. Biol.* **2009**, *391*, 74–85.

[153]   E. Bitto, Y. Huang, C. A. Bingman, S. Singh, J. S. Thorson, G. N. Phillips, *Proteins Struct. Funct. Genet.* **2008**, *70*, 289–293.

[154]   S. Flecks, E. P. Patallo, X. Zhu, A. J. Ernyei, G. Seifert, A. Schneider, C. Dong, J. H. Naismith, K. H. Van Pée, *Angew. Chemie - Int. Ed.* **2008**, *47*, 9533–9536.

[155]   J. Zeng, J. Zhan, *ChemBioChem* **2010**, *11*, 2119–2123.

[156]   P. Chankhamjon, D. Boettger-Schmidt, K. Scherlach, B. Urbansky, G. Lackner, D. Kalb, H. M. Dahse, D. Hoffmeister, C. Hertweck, *Angew. Chemie - Int. Ed.* **2014**, *53*, 13409–13413.

[157]   M. Sato, J. M. Winter, S. Kishimoto, H. Noguchi, Y. Tang, K. Watanabe, *Org. Lett.* **2016**, *18*, 1446–1449.

[158]   B. R. K. Menon, E. Brandenburger, H. H. Sharif, U. Klemstein, S. A. Shepherd, M. F. Greaney, J. Micklefield, *Angew. Chemie - Int. Ed.* **2017**, *56*, 11841–11845.

[159]   A. Schatz, E. Bugle, S. A. Waksman, *Proc. Soc. Exp. Biol. Med.* **1944**, *55*, 66–69.

[160]   B. Shen, *Cell* **2015**, *163*, 1297–1300.

[161]   J. W. H. Li, J. C. Vederas, *Science* **2009**, *325*, 161–165.

[162]   E. Kim, B. S. Moore, Y. J. Yoon, *Nat. Chem. Biol.* **2015**, *11*, 649–659.

[163]   R. Bentley, J. W. Bennett, *Annu. Rev. Microbiol.* **1999**, *53*, 411–446.

[164]   R. Bentley, *Crit. Rev. Biotechnol.* **1999**, *19*, 1–40.

[165]   T. J. Simpson, *Chem. Soc. Rev.* **1987**, *16*, 123–160.

[166]   D.-S. Tian, E. Kuhnert, J. Ouazzani, D. Wibberg, J. Kalinowski, R. J. Cox, *Chem. Sci.* **2020**, *11*, 12477–12484.

[167]   F. Trenti, R. J. Cox, *J. Nat. Prod.* **2017**, *80*, 1235–1240.

[168]   J. Rinkel, J. S. Dickschat, *Beilstein J. Org. Chem.* **2015**, *11*, 2493–2508.

[169]   P. Rabe, T. Schmitz, J. S. Dickschat, *Beilstein J. Org. Chem.* **2016**, *12*, 1839–1850.

[170]   U. Kück, B. Hoff, *Appl. Microbiol. Biotechnol.* **2010**, *86*, 51–62.

[171]   Z. Liu, T. L. Friesen, *Methods Mol. Biol.* **2012**, *835*, 365–375.

[172]   M. Penttilä, H. Nevalainen, M. Rättö, E. Salminen, J. Knowles, *Gene* **1987**, *61*, 155–164.

[173]   K. Ozeki, F. Kyoya, K. Hizume, A. Kanda, M. Hamachi, Y. Nunokawa, *Biosci. Biotechnol. Biochem.* **1994**, *58*, 2224–2227.

[174]   E. D. Mullins, X. Chen, P. Romaine, R. Raina, D. M. Geiser, S. Kang, *Phytopathology* **2001**, *91*, 173–180.

[175]   V. Olmedo-Monfil, C. Cortés-Penagos, A. Herrera-Estrella, *Methods Mol. Biol.* **2004**, *267*, 297–313.

[176]   R. J. Weld, K. M. Plummer, M. A. Carpenter, H. J. Ridgway, *Cell Res.* **2006**, *16*, 31–44.

[177]   C. Fairhead, B. Llorente, F. Denis, M. Soler, B. Dujon, *Yeast* **1996**, *12*, 1439–1457.

[178]   B. J. You, M. H. Lee, K. R. Chung, *Arch. Microbiol.* **2009**, *191*, 615–622.

[179]   C. Wang, V. Hantke, R. J. Cox, E. Skellam, *Org. Lett.* **2019**, *21*, 4163–4167.

[180]   L. Gritz, J. Davies, *Gene* **1983**, *25*, 179–188.

[181]   P. J. Punt, R. P. Oliver, M. A. Dingemanse, P. H. Pouwels, C. A. M. J. J. van den Hondel, *Gene* **1987**, *56*, 117–124.

[182]   M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816–821.

[183]   R. Song, Q. Zhai, L. Sun, E. Huang, Y. Zhang, Y. Zhu, Q. Guo, Y. Tian, B. Zhao, H. Lu, *Appl. Microbiol. Biotechnol.* **2019**, *103*, 6919–6932.

[184]   M. Schuster, R. Kahmann, *Fungal Genet. Biol.* **2019**, *130*, 43–53.

[185]   C. Pohl, J. A. K. W. Kiel, A. J. M. Driessen, R. A. L. Bovenberg, Y. Nygård, *ACS Synth. Biol.* **2016**, *5*, 754–764.

[186]   C. Schotte, L. Li, D. Wibberg, J. Kalinowski, R. J. Cox, *Angew. Chemie - Int. Ed.* **2020**, *59*, 23870–23878.

[187]   Y. He, B. Wang, W. Chen, R. J. Cox, J. He, F. Chen, *Biotechnol. Adv.* **2018**, *36*, 739–783.

[188]   F. Alberti, G. D. Foster, A. M. Bailey, *Appl. Microbiol. Biotechnol.* **2017**, *101*, 493–500.

[189]   M. N. Heneghan, A. A. Yakasai, L. M. Halo, Z. Song, A. M. Bailey, T. J. Simpson, R. J. Cox, C. M. Lazarus, *ChemBioChem* **2010**, *11*, 1508–1512.

[190]   T. Bai, Z. Quan, R. Zhai, T. Awakawa, Y. Matsuda, I. Abe, *Org. Lett.* **2018**, *20*, 7504–7508.

[191]   T. S. Lin, Y. M. Chiang, C. C. C. Wang, *Org. Lett.* **2016**, *18*, 1366–1369.

[192]   P. Barbesgaard, H. P. Heldt-Hansen, B. Diderichsen, *Appl. Microbiol. Biotechnol.* **1992**, *36*, 569–572.

[193]   F. J. Jin, J. I. Maruyama, P. R. Juvvadi, M. Arioka, K. Kitamoto, *FEMS Microbiol. Lett.* **2004**, *239*, 79–85.

[194]   C. M. Lazarus, K. Williams, A. M. Bailey, *Nat. Prod. Rep.* **2014**, *31*, 1339–1347.

[195]   K. A. K. Pahirulzaman, K. Williams, C. M. Lazarus, *Methods Enzymol.* **2012**, *517*, 241–260.

[196]   F. Katzen, *Expert Opin. Drug Discov.* **2007**, *2*, 571–589.

[197]   R. D. Gietz, R. A. Woods, *Methods Enzymol.* **2002**, *350*, 87–96.

[198]   T. L. Orr-Weaver, J. W. Szostak, R. J. Rothstein, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 6354–6358.

[199]   S. Kunes, P. J. Schatz, D. Botstein, *Gene* **1987**, *58*, 201–216.

[200]   S. bing Hua, M. Qiu, E. Chan, L. Zhu, Y. Luo, *Plasmid* **1997**, *38*, 91–96.

[201]   J. L. Hartley, G. F. Temple, M. A. Brasch, *Genome Res.* **2000**, *10*, 1788–1795.

[202]   D. M. Kupfer, S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu, D. W. Dyer, B. A. Roe, J. W. Murphy, *Eukaryot. Cell* **2004**, *3*, 1088–1100.

[203]   W. B. Yin, Y. H. Chooi, A. R. Smith, R. A. Cacho, Y. Hu, T. C. White, Y. Tang, *ACS Synth. Biol.* **2013**, *2*, 629–634.

[204]   Z. Song, W. Bakeer, J. W. Marshall, A. A. Yakasai, R. M. Khalid, J. Collemare, E. Skellam, D. Tharreau, M. H. Lebrun, C. M. Lazarus, et al., *Chem. Sci.* **2015**, *6*, 4837–4845.

[205]   J. M. Palmer, N. P. Keller, *Curr. Opin. Microbiol.* **2010**, *13*, 431–436.

[206]   Y. He, R. J. Cox, *Chem. Sci.* **2016**, *7*, 2119–2127.

[207]   B. Lowry, C. T. Walsh, C. Khosla, *Synlett* **2015**, *26*, 1008–1025.

[208]   F. Baneyx, *Curr. Opin. Biotechnol.* **1999**, *10*, 411–421.

[209]   H. P. Sørensen, K. K. Mortensen, *J. Biotechnol.* **2005**, *115*, 113–128.

[210]   G. L. Rosano, E. A. Ceccarelli, *Front. Microbiol.* **2014**, *5*, 172.

[211]   Z. Zhang, G. Kuipers, Ł. Niemiec, T. Baumgarten, D. J. Slotboom, J. W. de Gier, A. Hjelm, *Microb. Cell Fact.* **2015**, *14*, 142.

[212]   E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, J. M. McCoy, *Bio/Technology* **1993**, *11*, 187–193.

[213]   D. Esposito, D. K. Chatterjee, *Curr. Opin. Biotechnol.* **2006**, *17*, 353–358.

[214]   V. Gaberc-Porekar, V. Menart, *J. Biochem. Biophys. Methods* **2001**, *49*, 335–360.

[215]   R. A. Cacho, Y. H. Chooi, H. Zhou, Y. Tang, *ACS Chem. Biol.* **2013**, *8*, 2322–2330.

[216]   H. Raistrick, G. Smith, *Biochem. J.* **1935**, *29*, 606–11.

[217]   P. W. Clutterbuck, H. Raistrick, F. Reuter, *Biochem. J.* **1937**, *31*, 987–1002.

[218]   D. H. R. Barton, E. Miller, *J. Chem. Soc.* **1955**, 1028–1029.

[219]   J. Auerbach, S. M. Weinreb, *J. Chem. Soc. Chem. Commun.* **1974**, 298.

[220] H. -J Altenbach, W. Holzapfel, *Angew. Chemie Int. Ed. English* **1990**, *29*, 67–68.

[221] H. C. Kolb, H. Martin, R. Hoffmann, *Tetrahedron: Asymmetry* **1990**, *1*, 237–250.

[222] T. O. Larsen, J. Smedsgaard, K. F. Nielsen, M. A. E. Hansen, R. A. Samson, J. C. Frisvad, *Med. Mycol.* **2007**, *45*, 225–232.

[223] M. M. Nagia, M. El-Metwally, M. Shaaban, S. M. El-Zalabani, A. G. Hanna, *Org. Med. Chem. Lett.* **2012**, *2*, 9.

[224] P. Phattanawasin, K. Pojchanakom, U. Sotanaphun, N. Piyapolrungroj, S. Zungsontiporn, *Nat. Prod. Res.* **2007**, *21*, 1286–1291.

[225] D. Wakana, T. Hosoe, T. Itabashi, K. Nozawa, K. Okada, G. M. d. C. Takaki, T. Yaguchi, K. Fukushima, K. I. Kawai, *Mycotoxins* **2006**, *56*, 3–6.

[226] J. Malmstrøm, C. Christophersen, A. F. Barrero, J. Enrique Oltra, J. Justicia, A. Rosales, *J. Nat. Prod.* **2002**, *65*, 364–367.

[227] S. H. Park, D. S. Kim, W.-G. Kim, I.-J. Ryoo, D.-H. Lee, C.-H. Huh, S.-W. Youn, I.-D. Yoo, K.-C. Park, *Cell. Mol. Life Sci.* **2004**, *61*, 2878–2885.

[228] K. Xu, G. Li, R. Zhu, F. Xie, Y. Li, W. Yang, L. Xu, T. Shen, Z. Zhao, H. Lou, *Phytochemistry* **2020**, *170*, 112191.

[229] H. Mandai, K. Omori, D. Yamamoto, T. Tsumura, K. Murota, S. Yamamoto, K. Mitsudo, S. Ibaragi, A. Sasaki, H. Maeda, et al., *Bioorganic Med. Chem.* **2014**, *22*, 5338–5344.

[230] J. C. Lee, M. K. Yu, R. Lee, Y. H. Lee, J. G. Jeon, M. H. Lee, E. C. Jhee, I. D. Yoo, H. K. Yi, *J. Endod.* **2008**, *34*, 433–437.

[231] J. Goutam, G. Sharma, V. K. Tiwari, A. Mishra, R. N. Kharwar, V. Ramaraj, B. Koch, *Front. Microbiol.* **2017**, *8*, 1334.

[232] M. Arakawa, T. Someno, M. Kawada, D. Ikeda, *J. Antibiot. (Tokyo).* **2008**, *61*, 442–448.

[233] W. Y. Liao, C. N. Shen, L. H. Lin, Y. L. Yang, H. Y. Han, J. W. Chen, S. C. Kuo, S. H. Wu, C. C. Liaw, *J. Nat. Prod.* **2012**, *75*, 630–635.

[234] B. Kim, J. S. Park, H. Y. Choi, S. S. Yoon, W. G. Kim, *Sci. Rep.* **2018**, *8*, 8617.

[235] M. Demasi, A. L. -Felicio, A. O. Pacheco, H. G. Leite, C. Lima, L. H. Andrade, *J. Braz. Chem. Soc.* **2010**, *21*, 299–305.

[236] S. Kamata, H. Sakai, A. Hirota, *Agric. Biol. Chem.* **1983**, *47*, 2637–2638.

[237] C. Zaehle, M. Gressler, E. Shelest, E. Geib, C. Hertweck, M. Brock, *Chem. Biol.* **2014**, *21*, 719–731.

[238] M. Mikołajczyk, M. Mikina, M. W. Wieczorek, J. Błaszczyk, *Angew. Chemie Int. Ed. English* **1996**, *35*, 1560–1562.

[239] W. J. McGahren, J. H. van den Hende, L. A. Mitscher, *J. Am. Chem. Soc.* **1969**, *91*, 157–162.

[240] M. A. Stillwell, F. A. Wood, G. M. Strunz, *Can. J. Microbiol.* **1969**, *15*, 501–507.

[241] G. M. Strunz, A. S. Court, J. Komlossy, M. A. Stillwell, *Can. J. Chem.* **1969**, *47*, 3700–3701.

[242] B. Schulz, J. Sucker, H. J. Aust, K. Krohn, K. Ludewig, P. G. Jones, D. Döring, *Mycol. Res.* **1995**, *99*, 1007–1015.

[243] D. Giles, W. B. Turner, *J. Chem. Soc. C Org.* **1969**, *0*, 2187–2189.

[244] R. J. J. C. Lousberg, Y. Tirilly, M. Moreau, *Experientia* **1976**, *32*, 331–332.

[245] L. O. Zamir, C. C. Chin, *Bioorg. Chem.* **1982**, *11*, 338–349.

[246] V. Mierau, O. Sterner, T. Anke, *J. Antibiot. (Tokyo).* **2004**, *57*, 311–315.

[247] T. Matsumoto, T. Hosoya, H. Tomoda, M. Shiro, H. Shigemori, *Chem. Pharm. Bull.* **2011**, *59*, 1559–1561.

[248] Y. Honmura, S. Uesugi, H. Maeda, K. Tanaka, T. Nehira, K. ichi Kimura, M. Okazaki, M. Hashimoto, *Tetrahedron* **2016**, *72*, 1400–1405.

[249] D. R. McMullin, B. D. Green, N. C. Prince, J. B. Tanney, J. David Miller, *J. Nat. Prod.* **2017**, *80*, 1475–1483.

[250] G. M. Strunz, P. I. Kazinoti, M. A. Stillwell, *Can. J. Chem.* **1974**, *52*, 3623–3625.

[251] M. F. Elsebai, H. A. Ghabbour, N. Legrave, F. Fontaine-Vive, M. Mehiri, *Med. Chem. Res.* **2018**, *27*, 1885–1892.

[252] M. Zhao, D. Le Guo, G. H. Liu, X. Fu, Y. C. Gu, L. S. Ding, Y. Zhou, *J. Agric. Food Chem.* **2020**, *68*, 185–192.

[253] A. J. Birch, A. Cassera, A. R. Jones, *Chem. Commun.* **1965**, *0*, 167.

[254] R. A. Hill, R. H. Carter, J. Staunton, *J. Chem. Soc. Perkin Trans. 1* **1981**, *0*, 2570–2576.

[255] M. J. Garson, R. A. Hill, J. Staunton, *J. Chem. Soc. Chem. Commun.* **1977**, *3*, 624–626.

[256] J. S. E. Holker, K. Young, *J. Chem. Soc. Chem. Commun.* **1975**, *0*, 525–526.

[257] E. L. F. Ferreira, D. E. Williams, L. P. Ióca, R. P. Morais-Urano, M. F. C. Santos, B. O. Patrick, L. M. Elias, S. P. Lira, A. G. Ferreira, M. R. Z. Passarini, et al., *Org. Lett.* **2015**, *17*, 5152–5155.

[258]   M. Gressler, P. Hortschansky, E. Geib, M. Brock, *Front. Microbiol.* **2015**, *6*, 184.

[259]   A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, et al., *Nucleic Acids Res.* **2017**, *45*, D200–D203.

[260]   Y. M. Chiang, C. E. Oakley, M. Ahuja, R. Entwistle, A. Schultz, S. L. Chang, C. T. Sung, C. C. C. Wang, B. R. Oakley, *J. Am. Chem. Soc.* **2013**, *135*, 7720–7731.

[261]   T. Ugai, A. Minami, S. Tanaka, T. Ozaki, C. Liu, H. Shigemori, M. Hashimoto, H. Oikawa, *ChemBioChem* **2020**, *21*, 360–367.

[262]   W. Bunnak, P. Wonnapinij, A. Sriboonlert, C. M. Lazarus, P. Wattana-Amorn, *Org. Biomol. Chem.* **2019**, *17*, 374–379.

[263]   Y. H. Chooi, C. Krill, R. A. Barrow, S. Chen, R. Trengove, R. P. Oliver, P. S. Solomon, *Appl. Environ. Microbiol.* **2015**, *81*, 177–186.

[264]   P. Reveglia, M. Masi, A. Evidente, *Biomolecules* **2020**, *10*, 772.

[265]   A. M. Harned, K. A. Volp, *Nat. Prod. Rep.* **2011**, *28*, 1790–1810.

[266]   J. Meng, X. Wang, D. Xu, X. Fu, X. Zhang, D. Lai, L. Zhou, G. Zhang, *Molecules* **2016**, *21*, 715.

[267]   D. J. Cram, M. Tishler, *J. Am. Chem. Soc.* **1948**, *70*, 4238–4239.

[268]   D. J. Cram, *J. Am. Chem. Soc.* **1948**, *70*, 4240–4243.

[269]   J. Wang, K. Li, X. Luo, Z. Wu, T. Gu, S. Liao, X. Lin, B. Yang, Y. Liu, W. Fang, et al., *Org. Biomol. Chem.* **2019**, *17*, 8721–8725.

[270]   J. Meng, G. Gu, P. Dang, X. Zhang, W. Wang, J. Dai, Y. Liu, D. Lai, L. Zhou, *Front. Chem.* **2019**, *7*, 435.

[271]   X. Pang, X. Zhou, X. Lin, B. Yang, X. Tian, J. Wang, S. Xu, Y. Liu, *Bioorg. Chem.* **2020**, 104600.

[272]   L. S. Trifonov, J. H. Bieri, R. Prewo, A. S. Dreiding, L. Hoesch, D. M. Rast, *Tetrahedron* **1983**, *39*, 4243–4256.

[273]   Y. M. Ying, Z. J. Zhan, Z. S. Ding, W. G. Shan, *Chem. Nat. Compd.* **2011**, *47*, 541–544.

[274]   A. Al Fahad, A. Abood, K. M. Fisch, A. Osipow, J. Davison, M. Avramović, C. P. Butts, J. Piel, T. J. Simpson, R. J. Cox, *Chem. Sci.* **2014**, *5*, 523–527.

[275]   S. Sperry, G. J. Samuels, P. Crews, *J. Org. Chem.* **1998**, *63*, 10011–10014.

[276]   L. Du, T. Zhu, L. Li, S. Cai, B. Zhao, Q. Gu, *Chem. Pharm. Bull.* **2009**, *57*, 220–223.

[277]   R. F. Reátegui, D. T. Wicklow, J. B. Gloer, *J. Nat. Prod.* **2006**, *69*, 113–117.

[278]   N. Abe, T. Murata, A. Hirota, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 661–666.

[279]   K. C. Nicolaou, K. B. Simonsen, G. Vassilikogiannakis, P. S. Baran, V. P. Vidali, E. N. Pitsinos, E. A. Couladouros, *Angew. Chemie - Int. Ed.* **1999**, *38*, 3555–3559.

[280]   L. S. Trifonov, H. Hilpert, P. Floersheim, A. S. Dreiding, D. M. Rast, R. Skrivanova, L. Hoesch, *Tetrahedron* **1986**, *42*, 3157–3179.

[281]   R. Andrade, W. A. Ayer, P. P. Mebe, *Can. J. Chem.* **1992**, *70*, 2526–2535.

[282]   M. Kontani, Y. Sakagami, S. Marumo, *Tetrahedron Lett.* **1994**, *35*, 2577–2580.

[283]   C. E. Mazzucco, G. Warr, *J. Leukoc. Biol.* **1996**, *60*, 271–277.

[284]   K. C. Nicolaou, R. Jautelat, G. Vassilikogiannakis, P. S. Baran, K. B. Simonsen, *Chem. - A Eur. J.* **1999**, *5*, 3651–3665.

[285]   D. Li, F. Wang, X. Xiao, Y. Fang, T. Zhu, Q. Gu, W. Zhu, *Tetrahedron Lett.* **2007**, *48*, 5235–5238.

[286]   D. Li, S. Cai, T. Zhu, F. Wang, X. Xiao, Q. Gu, *Tetrahedron* **2010**, *66*, 5101–5106.

[287]   G. Bringmann, G. Lang, T. A. M. Gulder, H. Tsuruta, J. Mühlbacher, K. Maksimenka, S. Steffens, K. Schaumann, R. Stöhr, J. Wiese, et al., *Tetrahedron* **2005**, *61*, 7252–7265.

[288]   S. Skariyachan, A. B. Acharya, S. Subramaniyan, S. Babu, S. Kulkarni, R. Narayanappa, *J. Biomol. Struct. Dyn.* **2016**, *34*, 1865–1883.

[289]   R. P. Maskey, I. Grün-Wollny, H. Laatsch, *J. Nat. Prod.* **2005**, *68*, 865–870.

[290]   K. Washida, N. Abe, Y. Sugiyama, A. Hirota, *Biosci. Biotechnol. Biochem.* **2009**, *73*, 1355–1361.

[291]   R. Marra, R. Nicoletti, E. Pagano, M. DellaGreca, M. M. Salvatore, F. Borrelli, N. Lombardi, F. Vinale, S. L. Woo, A. Andolfi, *Nat. Prod. Res.* **2019**, *33*, 3389–3397.

[292]   W. A. Ayer, Y. Fukazawa, H. Orszanska, *Nat. Prod. Lett.* **1993**, *2*, 77–82.

[293]   L. S. Trifonov, A. S. Dreiding, L. Hoesch, D. M. Rast, *Helv. Chim. Acta* **1981**, *64*, 1843–1846.

[294]   N. Abe, O. Sugimoto, K. Tanji, A. Hirota, *J. Am. Chem. Soc.* **2000**, *122*, 12606–12607.

[295]   D. Barnes-Seeman, E. J. Corey, *Org. Lett.* **1999**, *1*, 1503–1504.

[296]   N. Abe, O. Sugimoto, T. Arakawa, K. Tanjii, A. Hirota, *Biosci. Biotechnol. Biochem.* **2001**, *65*, 2271–2279.

[297]   N. Abe, K. Yamamoto, T. Arakawa, A. Hirota, *Chem. Commun.* **2001**, *0*, 23–24.

[298]   N. Abe, T. Arakawa, A. Hirota, *Chem. Commun.* **2002**, *2*, 204–205.

[299]   N. Abe, T. Arakawa, K. Yamamoto, A. Hirota, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 2090–2099.

[300]   K. Sugaya, H. Koshino, Y. Hongo, K. Yasunaga, J. ichi Onose, K. Yoshikawa, N. Abe, *Tetrahedron Lett.* **2008**, *49*, 654–657.

[301]   R. Hong, Y. Chen, L. Deng, *Angew. Chemie - Int. Ed.* **2005**, *44*, 3478–3481.

[302]   K. Ohkata, Y. Tamura, B. B. Shetuni, R. Takagi, W. Miyanaga, S. Kojima, L. A. Paquette, *J. Am. Chem. Soc.* **2004**, *126*, 16783–16792.

[303]   G. Silvero, M. J. Lucero, E. Winterfeldt, K. N. Houk, *Tetrahedron* **1998**, *54*, 7293–7300.

[304]   O. Salo, F. Guzmán-Chávez, M. I. Ries, P. P. Lankhorst, R. A. L. Bovenberg, R. J. Vreeken, A. J. M. Driessen, *Appl. Environ. Microbiol.* **2016**, *82*, 3971–3978.

[305]   F. Guzmán-Chávez, O. Salo, Y. Nygård, P. P. Lankhorst, R. A. L. Bovenberg, A. J. M. Driessen, *Microb. Biotechnol.* **2017**, *10*, 958–968.

[306]   T. Carver, S. R. Harris, M. Berriman, J. Parkhill, J. A. McQuillan, *Bioinformatics* **2012**, *28*, 464–469.

[307]   C. Derntl, F. Guzmán-Chávez, T. M. Mello-de-Sousa, H. J. Busse, A. J. M. Driessen, R. L. Mach, A. R. Mach-Aigner, *Front. Microbiol.* **2017**, *8*, 2037.

[308]   F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al., *Mol. Syst. Biol.* **2011**, *7*, 539.

[309]   A. Šali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, *Proteins Struct. Funct. Bioinforma.* **1995**, *23*, 318–326.

[310]   P. Aloy, H. Ceulemans, A. Stark, R. B. Russell, *J. Mol. Biol.* **2003**, *332*, 989–998.

[311]   Z. Xiang, *Curr. Protein Pept. Sci.* **2006**, *7*, 217–227.

[312]   A. Sib, T. A. M. Gulder, *Angew. Chemie - Int. Ed.* **2017**, *56*, 12888–12891.

[313]   A. Sib, T. A. M. Gulder, *Angew. Chemie - Int. Ed.* **2018**, *57*, 14650–14653.

[314]   A. Sib, T. M. Milzarek, A. Herrmann, L. Oubraham, J. I. Müller, A. Pichlmair, R. Brack-Werner, T. A. M. Gulder, *ChemBioChem* **2020**, *21*, 492–495.

[315]   S. A. Baker Dockrey, A. L. Lukowski, M. R. Becker, A. R. H. Narayan, *Nat. Chem.* **2018**, *10*, 119–125.

[316]   H. Oikawa, *Cell Chem. Biol.* **2016**, *23*, 429–430.

[317]   H. Oikawa, T. Tokiwano, *Nat. Prod. Rep.* **2004**, *21*, 321–352.

[318]   K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van den Heever, C. R. Hutchinson, J. C. Vederas, *J. Am. Chem. Soc.* **2000**, *122*, 11519–11520.

[319]   V. Hantke, E. J. Skellam, R. J. Cox, *Chem. Commun.* **2020**, *56*, 2925–2928.

[320]   H. J. Kim, M. W. Ruszczycky, S. H. Choi, Y. N. Liu, H. W. Liu, *Nature* **2011**, *473*, 109–112.

[321]   H. Oikawa, T. Kobayashi, K. Katayama, Y. Suzuki, A. Ichihara, *J. Org. Chem.* **1998**, *63*, 8748–8756.

2 Publications

Publication 1

2.1 Diels-Alder Reactions During the Biosynthesis of Sorbicillinoids

Lukas Kahlert[1], Eman F. Bassiony[1,2], Russell J. Cox[1] and Elizabeth J. Skellam[1,3*]

[1] Institute for Organic Chemistry and BMWZ, Leibniz University of Hannover, Schneiderberg 38, 30167 Hannover, Germany

[2] Biochemistry Department, Faculty of Science, Zagazig University, Zagazig, Ash Sharquia Governorate 44519, Egypt

[3] Current Adress: Department of Chemistry & BioDiscovery Institute, University of North Texas, 1155 Union Circle 305220, Denton, Texas, 76203, USA

Type of authorship:	First author
Type of article:	Research article
Share of the work:	80%
Contribution to the publication:	Planned, performed and analyzed all heterologous expression-, feeding- and *in vitro* experiments, analysed all NMR data, wrote the first manuscript and majority of the ESI
Journal:	Angewandte Chemie International Edition
5-Year impact factor:	15.336 (2020), 14.205 (5-year)
Date of publication:	accepted 13.01.2020
Number of citations (Google Scholar on 30.03.2021):	11
DOI:	10.1002/anie.201915486
Supplementary information:	accessible free of charge at https://doi.org/10.1002/anie.201915486

Diels–Alder Reactions During the Biosynthesis of Sorbicillinoids

Lukas Kahlert, Eman F. Bassiony, Russell J. Cox, and Elizabeth J. Skellam*

Abstract: *The sorbicillinoids are a class of biologically active and structurally diverse fungal polyketides arising from sorbicillin. Through co-expression of sorA, sorB, sorC, and sorD from Trichoderma reesei QM6a, the biosynthetic pathway to epoxysorbicillinol and dimeric sorbicillinoids, which resemble Diels–Alder-like and Michael-addition-like products, was reconstituted in Aspergillus oryzae NSAR1. Expression and feeding experiments demonstrated the crucial requirement of the flavin-dependent monooxygenase SorD for the formation of dimeric sorbicillinoids, hybrid sorbicillinoids, and epoxysorbicillinol in vivo. In contrast to prior reports, SorD catalyses neither the oxidation of 2′,3′-dihydrosorbicillin to sorbicillin nor the oxidation of sorbicillinol to oxosorbicillinol. This is the first report that both the intermolecular Diels–Alder and Michael dimerization reactions, as well as the epoxidation of sorbicillinol are catalysed in vivo by SorD.*

Introduction

Sorbicillinoids are an important family of hexaketides produced by terrestrial[1,2] and marine[3–5] fungi.[6] Sorbicillin (**1a**) was first isolated from *Penicillium notatum* by Cram and Tishler in 1948.[7,8] **1a** is oxidatively dearomatised to form sorbicillinol (**2a**; Scheme 1A),[9] which reacts with itself and other compounds to form various sorbicillinoids (Scheme 1), more than 90 of which are known.[10–12] Trifonov first proposed **2a** as the intermediate for self-dimerization due to its dual diene and dienophile character.[2,13,14] This was confirmed by Abe et al. through elegant feeding experiments.[15] Dimeric sorbicillinoids include bisorbicillinol **3a**, which is formed by an intermolecular Diels–Alder (DA) reaction (Scheme 1B) and displays radical-scavenging activity almost matching that of α-tocopherol.[16] Other dimers include the Michael-addition-like products bisvertinol **4a** (Scheme 1C)[17] and tricho-dimerol **5** (Scheme 1D), a potent inhibitor of prostaglandin

[*] L. Kahlert, E. F. Bassiony, Prof. Dr. R. J. Cox, Dr. E. J. Skellam
Institute for Organic Chemistry and BMWZ
Leibniz University of Hannover
Schneiderberg 38, 30167 Hannover (Germany)
E-mail: elizabeth.skellam@oci.uni-hannover.de

E. F. Bassiony
Biochemistry Department, Faculty of Science, Zagazig University
Zagazig, Ash Sharqia Governorate 44519 (Egypt)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201915486.

biosynthesis.[18] Hybrid sorbicillinoids such as spirosorbicillinols A and B **6a** and **6b** (Scheme 1E),[19] are formed by DA reaction of **2a** with different dienophiles.

A breakthrough was achieved when the sorbicillinoid biosynthetic gene cluster (BGC) was discovered in *P. chrysogenum*. The BGC encodes a highly reducing iterative polyketide synthase (hr-PKS, SorA), a non-reducing iterative PKS (nr-PKS, SorB), an FAD-dependent monooxygenase (FMO, SorC), and a second FMO (SorD).[9] SorA was proposed to synthesize the triketide intermediate **7**, which remains tethered to the ACP domain due to lack of any release domain (Scheme 1A). Two intermediates, **7a** and **7b**, were proposed to be formed depending on whether the SorA enoyl reductase (ER) domain acts during the second cycle of polyketide chain elongation. The starter acyl transferase (SAT) domain of SorB loads **7** and elongates this starter unit three more times, introducing two methyl groups with its C-methyltransferase (C-MeT) domain. Reductive release from SorB yields aldehyde **8**, which undergoes Knoevenagel condensation to give the first isolatable intermediates sorbicillin (**1a**) and 2′,3′-dihydrosorbicillin (**1b**). SorC was shown to stereoselectively catalyse the oxidative dearomatisation of **1a** to **2a** and **1b** to **2b**.[9] In subsequent studies, knockout of *sorA* by the Dreissen group abolished production of **1** and all related sorbicillinoids.[20]

The only other confirmed homologous sorbicillinoid BGC exists in *T. reesei* QM6a[21] and shares the three core genes *sorA*, *sorB* and *sorC* required for formation of **2** (Scheme 1E).[9,20–22] The *T. reesei* BGC also encodes a second FMO named SorD. Derntl et al.[21] proposed that *T. reesei* SorD catalyses the oxidation of **1b** to **1a** (Scheme 1F). In contrast, Guzmán-Chávez et al.[22] suggested that *P. chrysogenum* SorD catalyses the oxidation of **2a** into oxosorbicillinol **9a** (Scheme 1F), since increased amounts of **2a** were produced by their Δ*sorD* strain. ARTEMIS analysis shows that the two *sorD* genes are not homologous (Scheme 1E), thus indicating that they may catalyse different reactions in *T. reesei* and *P. chrysogenum*, despite both fungi synthesizing a similar variety of sorbicillinoids.[20,21] However, the precise role of SorD has not yet been elucidated.
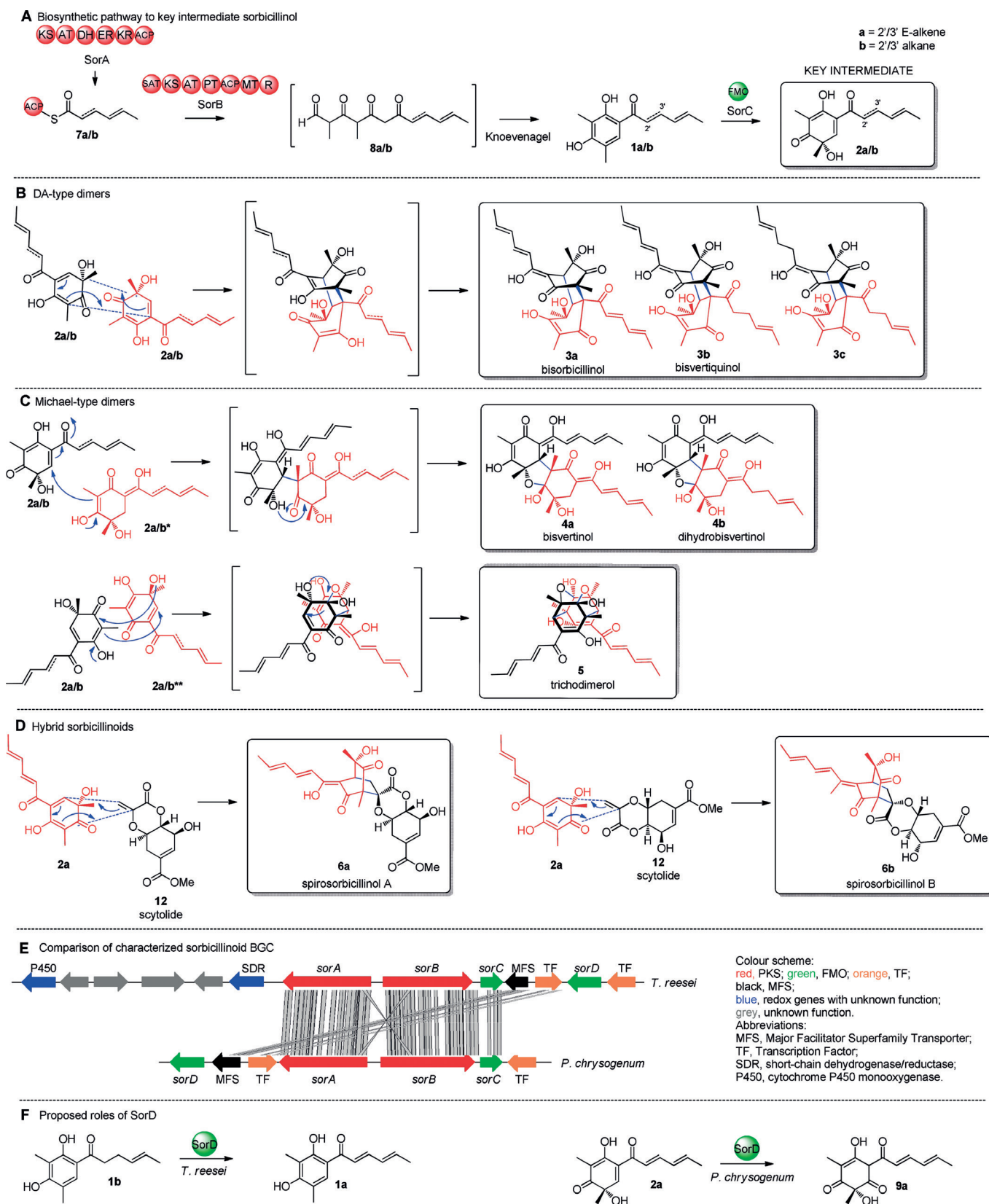
Dimeric sorbicillinoids are also formed during organic extraction and workup. Corey et al.[23] observed that **2a** is highly reactive and the dimer **5** was isolated after silica-gel chromatography with 3:1 hexane/EtOAc. Shortly afterwards, Abe et al.[10] reported spontaneous dimerization of **2a** to **3a** during liquid–liquid extraction with EtOAc and formation of **5** during lyophilization of **2a**. Recently, Gulder et al.[24,25] showed conclusively that in vitro production of **2a** by SorC, forms different dimeric sorbicillinoids depending on the organic co-solvent added to the reaction. For instance, addition of 20% *v/v* acetone yielded **3a** (27%) after

Scheme 1. A) Sorbicillinol biosynthetic pathway. B–D) Representative sorbicillinoid metabolites arising from the key intermediate sorbicillinol **2a**. E) Comparison of the *T. reesei* QMa6 and *P. chrysogenum* sorbicillinoid BGC. F) Proposed roles of SorD. New C–C bonds formed by Diels–Alder (DA)- or Michael-addition-like reactions are shown in blue. The **a/b** nomenclature denotes 2′/3′ E-alkene or 2′/3′ alkane functionality, respectively. **a/b*** and **a/b**** indicate tautomeric forms.

extraction with CH_2Cl_2, whereas addition of 20% *v/v* dimethylformamide (DMF) yielded **5** (27%) after extraction with CH_2Cl_2. However, the confirmed formation of dimeric sorbicillinoids from **2a** in vitro in the presence of organic solvents does not rule out the possibility of the existence of in vivo catalysts that can also form these compounds.

Since questions remain regarding the precise role of SorD and other enzymes, for example, a P450 and an SDR, encoded by genes adjacent to the *T. reesei* sorbicillinoid BGC (Scheme 1 E), we set out to investigate the BGC in more detail.

Results and Discussion

Sorbicillinoid production in *T. reesei* QM6a was confirmed by fermentation and analysis by liquid-chromatography mass spectrometry (LCMS, Figure 1 A). Numerous sorbicillinoid-related compounds were purified and identified based on UV, HRMS, and 1D and 2D NMR analysis (see the Supporting Information for details). These include the monomer **1a**, as well as epoxysorbicillinols **10a** and **10b**. The reactive intermediate **2a** could not be purified. The hybrid sorbicillinoid spirosorbicillinol B (**6b**) was also identified and fully eluci-

Figure 1. LCMS (DAD 210–600 nm) analysis of wt (A) and ΔsorA (B) *T. reesei* QM6a. * denotes uncharacterized sorbicillinoid-related compounds (Figure S4 and S5). Dimeric sorbicillinoids are highlighted in yellow; monomeric sorbicillinoids are highlighted in grey; non-sorbicillinoids are highlighted in blue.

dated. A number of other dimeric sorbicillinoids were identified based on their UV and MS profiles, including bisvertinolone (**11**),[14] **5**,[26] and **6a** (Figure S1–S3), as well as several dimers that could not be fully characterized (* in Figure 1 A, Figures S4, S5). Compounds **6a** and **6b** are proposed to arise through the reaction of **2a** with scytolide (**12**),[19] which has not previously been reported in *T. reesei*. Knockout of *sorA* completely abolished production of all sorbicillinoids, but led to the production of **12** (Figure 1 B).

The sorbicillinoid biosynthetic pathway was reconstituted in the heterologous host *Aspergillus oryzae* NSAR1 from *T. reesei* QM6a cDNA templates. Compared to the NSAR1 control (Figure 2 A), transformants expressing *sorA* (Figure 2 B) and *sorB* (Figure 2 C) individually did not produce any new compounds (Scheme 1 A). However, LCMS analysis of *A. oryzae* + *sorAB* transformants revealed four new peaks, of which two corresponded to **1a** and **1b** (Figure 2 D). The known trichopyrone (**13**)[27] and the related pyrone **14** were also produced. Pyrones **13** and **14** are likely premature off-loading shunt products of SorB, since similar products are known to be formed when methylation steps are not properly completed by nr-PKS.[28]

Next, *sorC* was added to *sorAB* in the expectation of forming **2a**. However, *A. oryzae* + *sorABC* produced **15a** as the major product (Figure 2 E). The UV absorption of **15a** is distinct from that of **2a**,[9] and the *m/z* $[M-H]^-$ of 249 indicated a reduction. Thorough NMR analysis of **15a** and its mono-methylated derivative (Me_3SiCHN_2)[29] confirmed the structure of the reduced sorbicillinol **15a**. The *A. oryzae* + *sorABC* transformants also produced the vertinolides **16a**[13] and **16b**. The absolute stereochemistries of **16a** and **16b** were assigned based on comparison with data reported by Takaia and Yamashita.[30]

SorC therefore catalyses the oxidative dearomatisation of **1** to **2** as expected, however **2** appears to be reduced to **15**. This is likely the result of an unknown enzyme in *A. oryzae* since this host is known to reduce other heterologously produced intermediates.[31–33] Vertinolides **16** are also derived from **15**. The furanone backbone of **16** presumably originates from a retro-aldol like ring opening or an intramolecular rearrangement of **15** (Scheme 2). A very similar skeletal rearrangement has been observed during the synthesis of paclitaxel derivatives.[34] No formation of any dimeric sorbicillinoids was observed in any of the +*sorABC* transformants.

Heterologous expression of +*sorABCD* resulted in many new peaks, which were identified as dimeric sorbicillinoids originating from either DA-like (**3a–c**) or Michael-addition-like (**4a,b**) dimerization of **2a** and/or **2b** (Figure 2 F). All compounds were fully elucidated using NMR, except for **4b**, which was characterized according to UV absorption and MS data (Figure S6). Compound **3c** must arise from a DA dimerisation of **2b**. In addition to the dimeric sorbicillinoids, oxidation to the epoxysorbicillinols **10a** and **10b** was observed for the first time (Figure 2 F). The production of **15a** is minimal, thus suggesting that in the absence of SorD, **2a** is quickly reduced by *A. oryzae*, however when SorD is present, **2a/b** can be efficiently converted into dimers **3a–c** and **4a,b** (Scheme 1 B,C) or oxidized to **10a/b**.
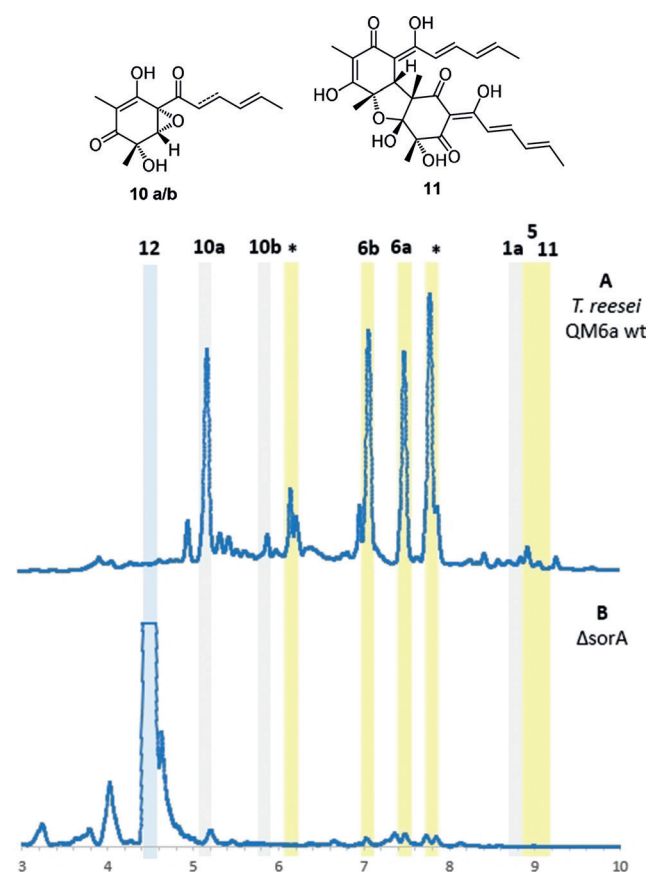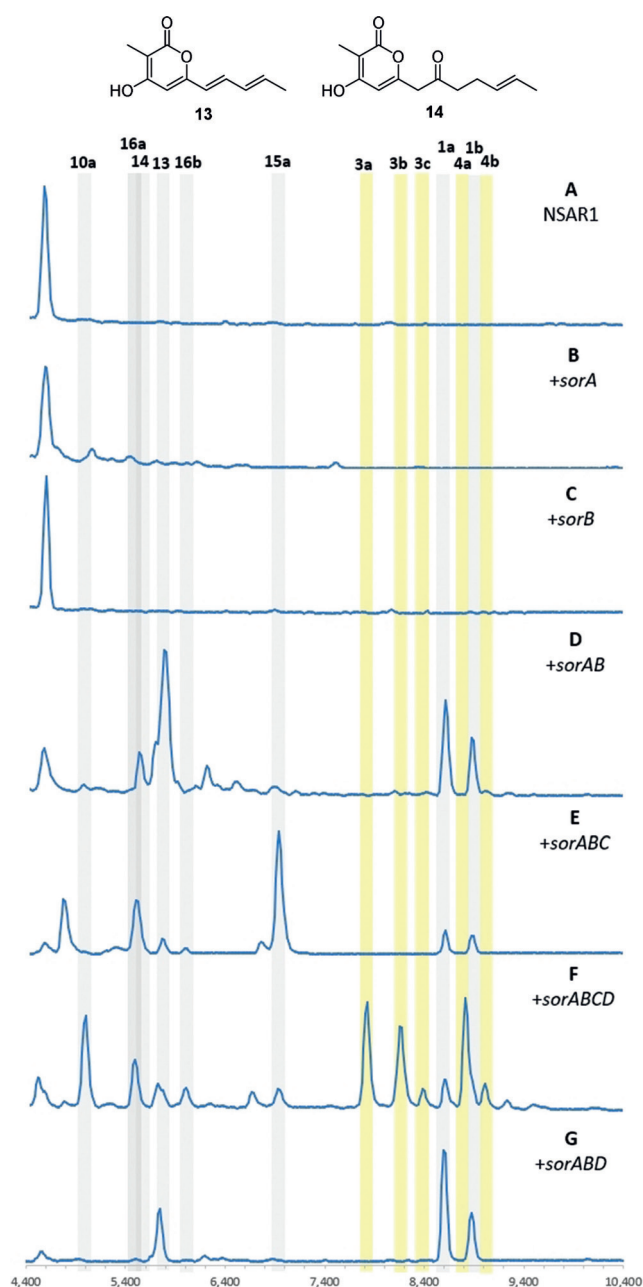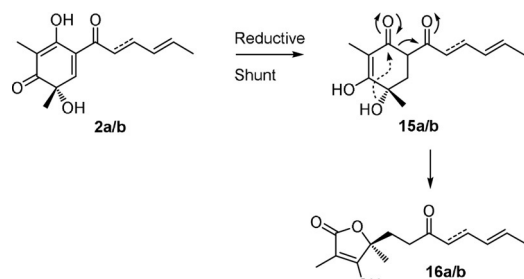
Figure 2. LCMS analysis [DAD 210–600 nm] of *A. oryzae* expressing different combinations of genes from the sorbicillinoid biosynthetic gene cluster. Dimeric sorbicillinoids are highlighted in yellow; monomeric sorbicillinoids are highlighted in grey.

Scheme 2. Proposed pathway for the reduction of **2** by *A. oryzae* leading to the formation of **15** and **16**.

Although **2a** is known to be inherently unstable, spontaneous formation of dimeric compounds during the extraction process is unlikely because the +*sorABC* and +*sorABCD* transformants had all been cultured and extracted under identical conditions, yet no dimeric compounds were observed in +*sorABC* transformants (Figure 2E). To exclude the possibility that these dimers formed during the extraction procedure, the crude culture supernatant of +*sorABCD* transformants was directly subjected to LCMS analysis prior to extraction, and it showed the same qualitative composition (Figure S10). Therefore, under the in vivo conditions used, dimerization of **2a/b** is independent of any added solvents and can be attributed to the presence of SorD.

LCMS chromatograms obtained from +*sorABD* transformants (Figure 2G) did not significantly differ from those for +*sorAB* transformants, thus indicating that **2a** and **2b** are the required substrates for SorD. The observation of both **1a** and **1b** in these transformants indicates that SorD does not oxidize **1b** to **1a** as proposed by Derntl et al.[21] Similarly, since the oxosorbicillinols **9a** and **9b** were not observed in any of the chromatograms, SorD does not seem to have a role in oxidizing **2** to **9** as proposed by Guzmán-Chávez et al.[22] This shows that SorD has two roles: dimerization of **2** to **3** and **4**, which does not require oxidation of **2**, and an independent role in oxidizing **2a/b** to epoxides **10a/b**.

To further investigate the role of SorD in dimerization, we attempted in vitro assays. Extensive attempts were made to obtain soluble SorD, but all efforts at expression in either *E. coli* or *S. cerevisiae* resulted in insoluble and inactive protein. Experiments with cell-free extract (CFE) or whole cells of these organisms expressing SorD also met with failure.

In contrast, we were able to obtain soluble his$_6$-SorC (52.8 kDa) in very high yields (180 mg L^{-1}), and in vitro assays were performed as described previously.[9] Substrates **1a/b** were purified from +*sorAB* transformants. As expected, substrates **1a** and **1b** were quickly converted into **2a** and **2b** (Figure S11). In contrast, when the assays were performed under the specific conditions as described by Gulder et al.[24] in the presence of 20% *v/v* acetone followed by extraction of the assay mixture with CH$_2$Cl$_2$ or CHCl$_3$, dimers **3a–c** were formed (Figure S11). These observations are in accordance with previous reports[10,23–25] that under specific conditions, dimerization of **2a** can be induced by organic solvents.

When an excess of NAD(P)H or prolonged incubation times were employed in the SorC assays, small amounts of **15a** or **10a/b** could be detected (Figure S12). Therefore, the reduction of **2** to **15** in the heterologous expression experiments may not even require an enzyme. Although the reduction of **2** prevents subsequent dimerization reactions, the reduction of **2** in vivo must be faster than any spontaneous dimerization reactions.

We next investigated the individual functions of SorC and SorD in vivo. Since **2** cannot be purified from our heterologous expression experiments due to its conversion to **15**, we used our in vitro assay with SorC to obtain **2**. Substrates **1a/b** were incubated with SorC leading to the formation of compounds **2a/b**. Dihydrosorbicillin (**2b**) was purified (Figure 3A) and supplemented to *A. oryzae* expressing either *sorC* or *sorD*. When **2b** was fed to *A. oryzae* expressing *sorC*
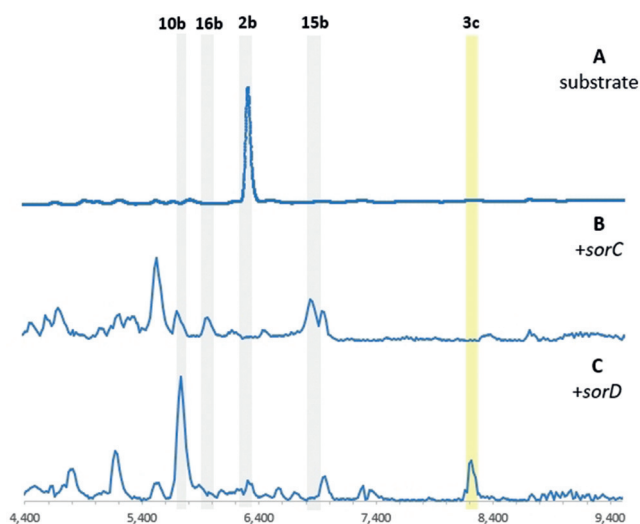
Figure 3. LCMS analysis [DAD 210–600 nm] of *A. oryzae* expressing either *sorC* (B) or *sorD* (C) fed with compound **2b**. Dimeric sorbicillinoids are highlighted in yellow; monomeric sorbicillinoids are highlighted in grey.

Figure 4. LCMS analysis (DAD 210–600 nm) of *A. oryzae* NSAR1 (A; control) and transformants expressing *sorABC* (B) or *sorABCD* (C) fed with scytolide (**12**). Dimeric sorbicillinoids are highlighted in yellow; monomeric sorbicillinoids are highlighted in grey.

and cultured and extracted using standard conditions, no dimeric compounds were detected. Instead, compounds **15b** and **16b** were detected as well as a small amount of **10b** (Figure 3B). **2b** is therefore directly converted into **15b** and **16b** (Scheme 2), thus proving that dimerization is not spontaneous and nor is it catalysed by SorC under the conditions investigated.

When **2b** was fed to *A. oryzae* expressing *sorD* under identical conditions, the dimeric sorbicillinoid **3c** was detected (Figure 3C). Since all other dimeric compounds form from at least one molecule of **2a**, no other dimeric molecules were detected. **10b** was also detected but **15b** was not (Figure 3C), thus demonstrating conclusively that in the presence of SorD, **2b** is efficiently dimerized to **3c** or epoxidized to **10b**. Clearly SorD, catalyzes distinct epoxidation or dimerization reactions, both of which are faster than the reduction of **2b** to **15b** either by *A. oryzae* or intracellular NAD(P)H.

The hybrid sorbicillinoids **6a/b** are considered to derive from dimerization of **2a** with **12** (Scheme 1D).[19] To further investigate the role of SorD in the dimerization of **2**, scytolide (**12**) was purified from *T. reesei* Δ*sorA* (Figure 1B) and fed to *A. oryzae* strains and then extracted using standard conditions. **12** remained unaltered in the NSAR1 control (Figure 4A). In strains expressing +*sorABC*, no conversion to **6** was observed. The major product remained the reduced shunt product **15a**, thus indicating that **2** does not spontaneously undergo a Diels–Alder reaction with **12**, nor does SorC catalyse the reaction (Figure 3B). However, when **12** was fed to +*sorABCD* transformants, **6a** and **6b** were observed (Figure 3C). These results demonstrate that **6a** and **6b** are exclusively formed in the presence of sorD.

Surprisingly, compounds **5** and **11**, which were observed in *T. reesei* QM6a, were not detected in any of our expression strains. Compound **5** is proposed to arise from two consecutive Michael-like additions (Scheme 1C). That **5** is not observed further supports these dimerization mechanisms not
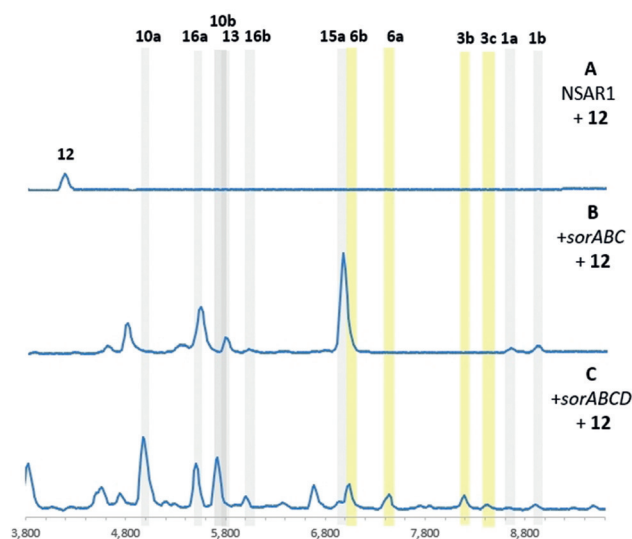
being spontaneous and requiring an enzyme catalyst. Compound **11** is a dimer of **2a** and **9a**, however the formation of **9a** appears to be independent of *SorABCD*. To be certain that the additional genes encoding an SDR and P450 are not required for the production of **5** and/or **11**, both genes were successfully disrupted in *T. reesei* using the bipartite knock-out strategy (Figure S18).[35] However, no changes in the metabolite profile were observed. Similarly, heterologous expression of the SDR and P450 genes with +*sorABCD* in *A. oryzae* did not yield any new pathway intermediates (Figure S19). This indicates that genes encoding the enzymes necessary for the formation of **5** and **11** are located elsewhere in the genome of *T. reesei*.

Conclusion

Our results demonstrate that the SorAB PKS system of *T. reesei* QM6a behaves as expected to produce **1** in the heterologous host *A. oryzae* (Scheme 3). The FMO SorC catalyses the oxidative dearomatisation of **1** to **2** in vitro as expected. Sorbicillinol (**2**) can also be reduced to **15** in vitro by NAD(P)H alone. In vivo, SorABC produces **2**, which is also reduced to **15**. Although SorC can catalyse consecutive oxidation of **2** to epoxysorbicillinol (**10**) in vitro, the second oxidation is very inefficient. This epoxidation is primarily catalysed in vivo by the FMO SorD. The requirement for two different enzymes may reflect the differences in mechanism of the two oxidation reactions. During the oxidative dearomatisation of **1** to **2**, the enzyme bound-hydro-peroxyflavin species acts as an electrophile, whereas during the epoxidation of **2** to **10**, the peroxyflavin species would be expected to act as a nucleophile (Scheme 4). Gulder et al. reported a similar facile reaction of **2a** with 'BuO₂H.[25]

Dimeric sorbicillinoids derived from **2** were only detected in vivo in the presence of SorD. SorD is also able to epoxidize
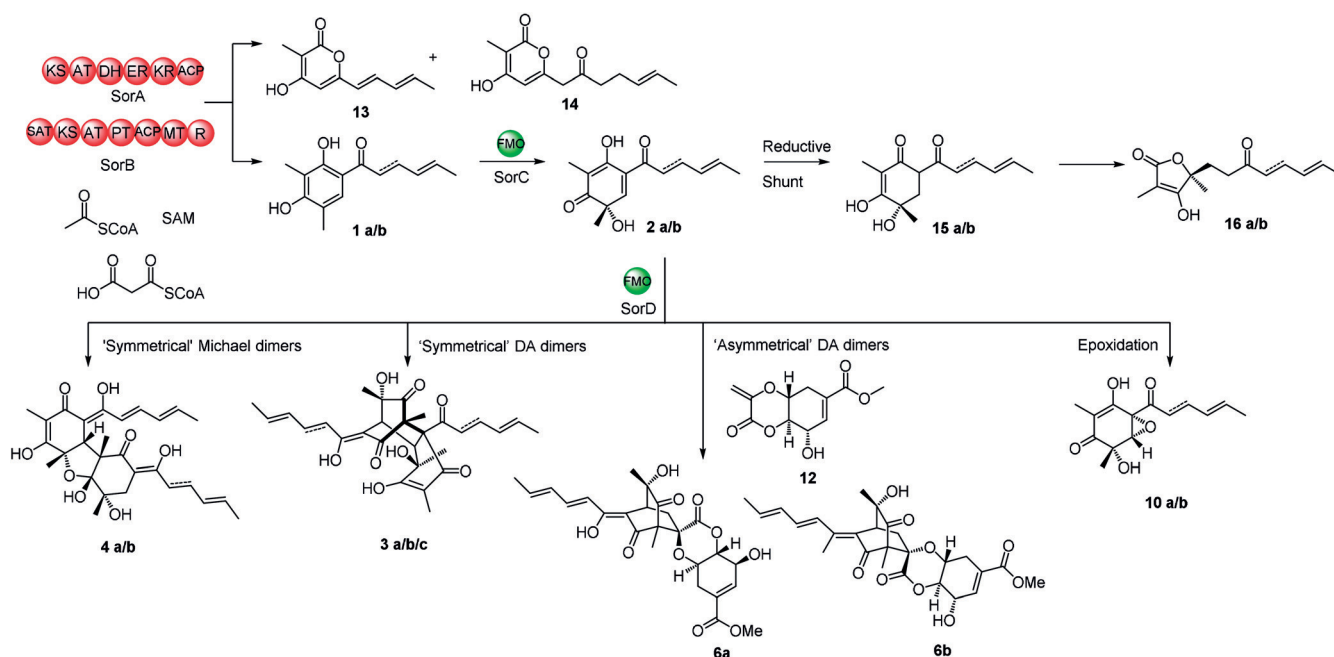
Scheme 3. Summary of the sorbicillinoid biosynthetic pathway in *T. reesei* QM6a. The dual-function FMO SorD catalyses both Diels–Alder- and Michael-addition-like dimerizations, as well as epoxidation of **2**.
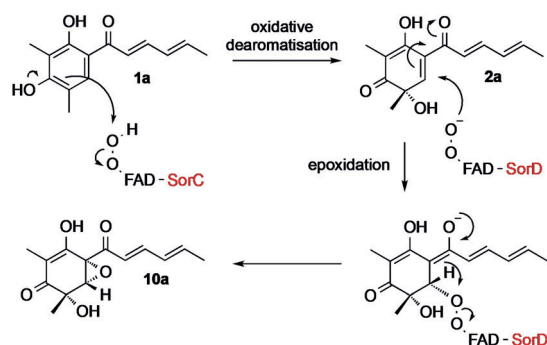
Scheme 4. Putative mechanism for the oxidative dearomatisation of **1a** to **2a** by SorC and its subsequent epoxidation to **10a** catalysed by SorD.

2 to **10**, which is independent of the dimerization mechanism. These dimeric compounds were previously described as resulting from organic-solvent-induced spontaneous reactions. However, in our hands, we developed extraction and purification conditions, which do not catalyse these dimerization reactions. Under these conditions but in the presence of SorD, dimeric products were observed, including "symmetrical" Michael-addition dimers such as **4**, "symmetrical" Diels–Alder reaction dimers such as **3**, and "asymmetrical" Diels–Alder reaction dimers such as **6a** and **6b** (Scheme 3). Therefore, we reveal a role of SorD in dimerizing the highly reactive intermediate **2** and demonstrate the scope of the reactions catalysed by this enzyme. Furthermore, we clearly demonstrate that the isomers spirosorbicillinol A (**6a**) and B (**6b**) arise from **2a** and the shikimic acid derivative **12**, thereby confirming their biosynthetic origin.

Formation of epoxides is a common reaction catalysed by FMOs,[36–41] but catalysis of DA reactions is much rarer. Fujii and co-workers reported a dual-acting flavin-dependent enzyme, Sol5, which catalyses both an alcohol oxidation and an *intramolecular* DA cyclisation during the biosynthesis of solanapyrone.[42] In contrast, SorD catalyses *intermolecular* DA and Michael reactions, which do not require initial oxidation. SorD is the first flavin-dependent enzyme reported to behave in this way.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] R. P. Maskey, I. Grün-Wollny, H. Laatsch, *J. Nat. Prod.* **2005**, *68*, 865–870.

[2] L. S. Trifonov, J. H. Bieri, R. Prewo, A. S. Dreiding, L. Hoesch, D. M. Rast, *Tetrahedron* **1983**, *39*, 4243–4256.

[3] S. Sperry, G. J. Samuels, P. Crews, *J. Org. Chem.* **1998**, *63*, 10011–10014.

[4] A. Abdel-Lateff, K. Fisch, A. D. Wright, *Z. Naturforsch. C* **2009**, *64*, 186–192.

[5] D. Li, S. Cai, T. Zhu, F. Wang, X. Xiao, Q. Gu, *Tetrahedron* **2010**, *66*, 5101–5106.

[6] A. M. Harned, K. A. Volp, *Nat. Prod. Rep.* **2011**, *28*, 1790–1810.

[7] D. J. Cram, M. Tishler, *J. Am. Chem. Soc.* **1948**, *70*, 4238–4239.

[8] D. J. Cram, *J. Am. Chem. Soc.* **1948**, *70*, 4240–4243.

[9] A. Al Fahad, A. Abood, K. M. Fisch, A. Osipow, J. Davison, M. Avramović, C. P. Butts, J. Piel, T. J. Simpson, R. J. Cox, *Chem. Sci.* **2014**, *5*, 523–527.

[10] N. Abe, O. Sugimoto, K. Tanji, A. Hirota, *J. Am. Chem. Soc.* **2000**, *122*, 12606–12607.

[11] N. Abe, K. Yamamoto, T. Arakawa, A. Hirota, *Chem. Commun.* **2001**, 23–24.

[12] J. Meng, X. Wang, D. Xu, X. Fu, X. Zhang, D. Lai, L. Zhou, G. Zhang, *Molecules* **2016**, *21*, 715.

[13] L. S. Trifonov, A. S. Dreiding, L. Hoesch, D. M. Rast, *Helv. Chim. Acta* **1981**, *64*, 1843–1846.

[14] L. S. Trifonov, H. Hilpert, P. Floersheim, A. S. Dreiding, D. M. Rast, R. Skrivanova, L. Hoesch, *Tetrahedron* **1986**, *42*, 3157–3179.

[15] N. Abe, T. Arakawa, K. Yamamoto, A. Hirota, *Biosci. Biotechnol. Biochem.* **2002**, *66*, 2090–2099.

[16] N. Abe, T. Murata, A. Hirota, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 661–666.

[17] K. C. Nicolaou, K. B. Simonsen, G. Vassilikogiannakis, P. S. Baran, V. P. Vidali, E. N. Pitsinos, E. A. Couladouros, *Angew. Chem. Int. Ed.* **1999**, *38*, 3555–3559; *Angew. Chem.* **1999**, *111*, 3762–3766.

[18] C. E. Mazzucco, G. Warr, *J. Leukocyte Biol.* **1996**, *60*, 271–277.

[19] K. Washida, N. Abe, Y. Sugiyama, A. Hirota, *Biosci. Biotechnol. Biochem.* **2009**, *73*, 1355–1361.

[20] O. Salo, F. Guzmán-Chávez, M. I. Ries, P. P. Lankhorst, R. A. L. Bovenberg, R. J. Vreeken, A. J. M. Driessen, *Appl. Environ. Microbiol.* **2016**, *82*, 3971–3978.

[21] C. Derntl, F. Guzmán-Chávez, T. M. Mello-de-Sousa, H. J. Busse, A. J. M. Driessen, R. L. Mach, A. R. Mach-Aigner, *Front. Microbiol.* **2017**, *8*, 2037.

[22] F. Guzmán-Chávez, O. Salo, Y. Nygård, P. P. Lankhorst, R. A. L. Bovenberg, A. J. M. Driessen, *Microb. Biotechnol.* **2017**, *10*, 958–968.

[23] D. Barnes-Seeman, E. J. Corey, *Org. Lett.* **1999**, *1*, 1503–1504.

[24] A. Sib, T. A. M. Gulder, *Angew. Chem. Int. Ed.* **2017**, *56*, 12888–12891; *Angew. Chem.* **2017**, *129*, 13068–13071.

[25] A. Sib, T. A. M. Gulder, *Angew. Chem. Int. Ed.* **2018**, *57*, 14650–14653; *Angew. Chem.* **2018**, *130*, 14861–14864.

[26] R. Andrade, W. A. Ayer, P. P. Mebe, *Can. J. Chem.* **1992**, *70*, 2526–2535.

[27] K. Washida, N. Abe, Y. Sugiyama, A. Hirota, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 1052–1057.

[28] P. A. Storm, D. A. Herbst, T. Maier, C. A. Townsend, *Cell Chem. Biol.* **2017**, *24*, 316–325.

[29] E. Kühnel, D. D. P. Laffan, G. C. Lloyd-Jones, T. Martínez Del Campo, I. R. Shepperson, J. L. Slaughter, *Angew. Chem. Int. Ed.* **2007**, *46*, 7075–7078; *Angew. Chem.* **2007**, *119*, 7205–7208.

[30] A. Takaiwa, K. Yamashita, *Agric. Biol. Chem.* **1983**, *47*, 429–430.

[31] Z. Song, W. Bakeer, J. W. Marshall, A. A. Yakasai, R. M. Khalid, J. Collemare, E. Skellam, D. Tharreau, M. H. Lebrun, C. M. Lazarus, et al., *Chem. Sci.* **2015**, *6*, 4837–4845.

[32] Y. He, R. J. Cox, *Chem. Sci.* **2016**, *7*, 2119–2127.

[33] T. Ugai, A. Minami, R. Fujii, M. Tanaka, H. Oguri, K. Gomi, H. Oikawa, *Chem. Commun.* **2015**, *51*, 1878–1881.

[34] R. Marder-Karsenti, J. Dubois, A. Chiaroni, C. Riche, D. Guénard, F. Guéritte, P. Potier, *Tetrahedron* **1998**, *54*, 15833–15844.

[35] M. L. Nielsen, L. Albertsen, G. Lettier, J. B. Nielsen, U. H. Mortensen, *Fungal Genet. Biol.* **2006**, *43*, 54–64.

[36] K. Otto, K. Hofstetter, M. Röthlisberger, B. Witholt, A. Schmid, *J. Bacteriol.* **2004**, *186*, 5292–5302.

[37] K. Büch, H. Stransky, A. Hager, *FEBS Lett.* **1995**, *376*, 45–48.

[38] R. Nofiani, K. de Mattos-Shipley, K. E. Lebe, L. C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson, R. J. Cox, *Nat. Commun.* **2018**, *9*, 3940.

[39] M. Tanaka, S. Tahara, *Phytochemistry* **1997**, *46*, 433–439.

[40] S. Li, J. M. Finefield, J. D. Sunderhaus, T. J. McAfoos, R. M. Williams, D. H. Sherman, *J. Am. Chem. Soc.* **2012**, *134*, 788–791.

[41] Y. Matsuda, T. Wakimoto, T. Mori, T. Awakawa, I. Abe, *J. Am. Chem. Soc.* **2014**, *136*, 15326–15336.

[42] K. Kasahara, T. Miyamoto, T. Fujimoto, H. Oguri, T. Tokiwano, H. Oikawa, Y. Ebizuka, I. Fujii, *ChemBioChem* **2010**, *11*, 1245–1252.

Publication 2

2.2   The same but different: multiple functions of the fungal flavin dependent monooxygenase SorD from *Penicillium chrysogenum*

Lukas Kahlert[1], Russell J. Cox[1] and Elizabeth Skellam*[1,2]

[1] Institute for Organic Chemistry and BMWZ, Leibniz University of Hannover, Schneiderberg 38, 30167 Hannover, Germany

[2] Current Adress: Department of Chemistry & BioDiscovery Institute, University of North Texas, 1155 Union Circle 305220, Denton, Texas, 76203, USA

Type of authorship:	First author
Type of article:	Research article
Share of the work:	90%
Contribution to the publication:	Planned, performed and analyzed all experiments, wrote the first manuscript and the ESI
Journal:	Chemical Communications
5-Year impact factor:	6.222 (2020), 6.008 (5-year)
Date of publication:	accepted 05.08.2020
Number of citations (Google Scholar on 30.03.2021):	1
DOI:	10.1039/d0cc03203d
Supplementary information:	accessible free of charge at https://doi.org/10.1039/D0CC03203D

ChemComm

Check for updates

The same but different: multiple functions of the fungal flavin dependent monooxygenase SorD from *Penicillium chrysogenum*†

Lukas Kahlert, Russell J. Cox [iD] and Elizabeth Skellam [iD]*

Sorbicillinoids are a large family of fungal secondary metabolites with a diverse range of structures and numerous bioactivites, some of which have pharmaceutical potential. The flavin-dependent monooxygenase SorD from *Penicillium chrysogenum* (PcSorD) utilizes sorbicillinol to catalyze a broad scope of reactions: formation of oxosorbicillinol and epoxysorbicillinol; intermolecular Diels–Alder and Michael-addition dimerization reactions; and dimerization of a sorbicillinol derivative with oxosorbicillinol. PcSorD shares only 18.3% sequence identity with SorD from *Trichoderma reesei* (TrSorD) and yet unexpectedly catalyzes many of the same reactions, however, the formation of oxosorbicillinol and bisvertinolone by PcSorD extends the range of reactions catalyzed by a single enzyme. Phylogenetic analysis indicates that PcSorD and TrSorD bind the flavin cofactor covalently but *via* different residues and point mutations confirm these residues are essential for activity.

Sorbicillinoids are a family of structurally diverse secondary metabolites produced exclusively by fungi (Fig. 1A and Fig. S1, ESI†).[1] Sorbicillinoids are named after the first characterized member sorbicillin **1** which is oxidatively dearomatized to give the key intermediate sorbicillinol **2**.[2] **2** is able to react with itself and other compounds to form complex dimers such as bisorbicillinol **3** and bisvertinol **4**, or be oxidized further to generate epoxysorbicillinol **5** and oxosorbicillinol **6**.[3] Many sorbicillinoids are biologically active including **3** with potent radical scavenging activities, **4** with anti-microbial properties, and bisvertinolone **7** with cytotoxic properties.[4–6]

Biosynthetic gene clusters (BGC) encoding the enzymes responsible for sorbicillinoid biosynthesis have been identified in *Trichoderma reesei* QM6a and *Penicillium chrysogenum*. These two BGC encode several highly homologous enzymes: a highly-reducing polyketide synthase (HR-PKS; SorA); a non-reducing polyketide synthase (NR-PKS; SorB); and a flavin-dependent

monooxygenase (FMO; SorC) (Fig. 1B). SorA and SorB synthesize the hexaketide precursor sorbicillin **1** and SorC catalyses the subsequent stereoselective oxidative dearomatisation of **1** to sorbicillinol **2**.[2,7] Recently, we showed that the second FMO of the sorbicillinoid BGC in *T. reesei* QM6a, TrSorD, is essential for the intermolecular Diels–Alder dimerization of **2** to form bisorbicillinol **3**, the Michael-like dimerization of **2** with **8** to form bisvertinol **4**, and furthermore independently oxidizes **2** to epoxysorbicillinol **5**.[8]

The *P. chrysogenum* BGC also encodes a second FMO, named PcSorD, but shares little sequence homology with TrSorD
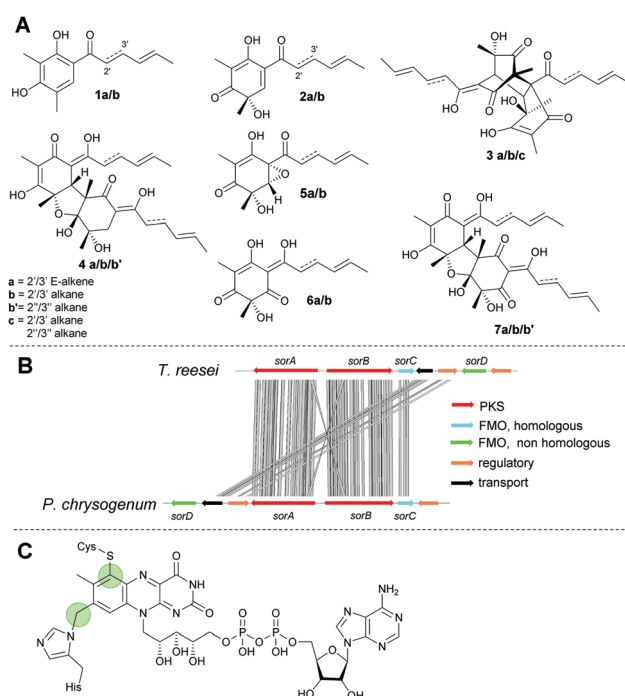

Fig. 1 (A) Sorbicillinoids produced by both *T. reesei* and *P. chrysogenum*; (B) comparison between the sorbicillinoid BGC in *T. reesei* and *P. chrysogenum*; (C) structure of FAD bi-covalently bound to conserved His and Cys residues in an exemplary FMO.

Institute for Organic Chemistry and BMWZ Leibniz University of Hannover Schneiderberg 38, 30167, Hannover, Germany.
E-mail: elizabeth.skellam@oci.uni-hannover.de

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cc03203d

This journal is © The Royal Society of Chemistry 2020

(<18.3% identity; Fig. S2, ESI†). FMOs are a class of versatile enzymes often involved in the oxygenation of secondary metabolites.[9] The scope of these biocatalysts covers hydroxylation,[10] epoxidation,[11] Baeyer-Villiger oxidation,[12] halogenation,[13] amine oxidation,[14] and also includes Diels–Alder reactions.[8] Most FAD/FMN enzymes bind the cofactor non-covalently, but examples have been reported where the cofactor is covalently bound.[15,16] Covalent linkage usually occurs between the 8α-methyl of the cofactor and a conserved Tyr,[17] His,[18] or Cys[19] of the enzyme (reviewed comprehensively elsewhere).[20] Alternatively, the flavin cofactor may be bi-covalently bound by forming a second linkage to a conserved Cys at its C-6-position in the isoalloxazine ring (Fig. 1C).[21]

Investigations into sorbicillinoid biosynthesis in *P. chrysogenum* by Guzmán-Chávez *et al.* led to the proposal that PcSorD is involved in converting **2a** into oxosorbicillinol **6a**.[22] However, our recent work showed that FMO TrSorD appears not to be involved in the biosynthesis of **6**.[8] Since both *T. reesei* and *P. chrysogenum* produce **1–7** and a similar spectrum of dimeric sorbicillinoids (Fig. S1, ESI†) yet have no additional homologous biosynthetic enzymes encoded within their BGC (Fig. 1B) we were interested in the function of the FMO PcSorD and how **6** and **7** are biosynthesized in both fungi.

To elucidate the function of PcSorD we first tried to obtain soluble enzyme for *in vitro* studies, however despite extensive attempts all efforts at expression in *E. coli* resulted in insoluble and inactive protein (Fig. S3–S5, ESI†). Alternative methods to study enzyme function include heterologous expression where the role of the enzyme can be deduced *in vivo*. Previously we reported the characterization of SorABC in *Aspergillus oryzae* through isolating the expected sorbicillins **1a/b**, and several shunts **8–10** (Fig. 2A; Scheme 1A).[8] When we co-expressed TrSorD we were able to isolate dimeric sorbicillinoids **3** and **4**, in addition to epoxysorbicillinols **5** (Fig. 2B). We therefore decided to use *A. oryzae* + *sorABC* as the heterologous host strain to investigate the role of PcSorD in a directly comparable environment to TrSorD. We obtained a synthetic gene encoding PcSorD (*PcsorD*) and co-expressed it with *sorABC* in *A. oryzae*. Transformants were grown, extracted and analysed using previously established methods.[8] LCMS analysis showed that the Diels–Alder and Michael-like dimeric sorbicillinoids **3** and **4**, as well as **5** were produced (Fig. 2C). These compounds were identified by comparing the physical data with the same compounds isolated in our previous study (Fig. S33–S46, ESI†).[8] In addition, two new major peaks were observed only when *PcsorD* was co-expressed (Fig. 2C). The respective compounds were purified by preparative LCMS and identified as oxosorbicillinol **6a** and bisvertinolone **7a** by thorough NMR analysis (Fig. S18–S32, ESI†).

Despite sharing little sequence homology PcSorD and TrSorD both catalyse intermolecular Diels–Alder dimerization of **2** to **3**, intermolecular Michael-addition of **2** and **8** to form **4**, as well as an independent oxidation of **2** to **5**. Remarkably, PcSorD is also involved in the formation of **6** and the Michael-like dimerization of **6** with **2** to form **7**, further expanding the versatility of this class of enzymes (Scheme 1A). We suspect that PcSorD is able to catalyze the Michael-like addition of **2** and **6**
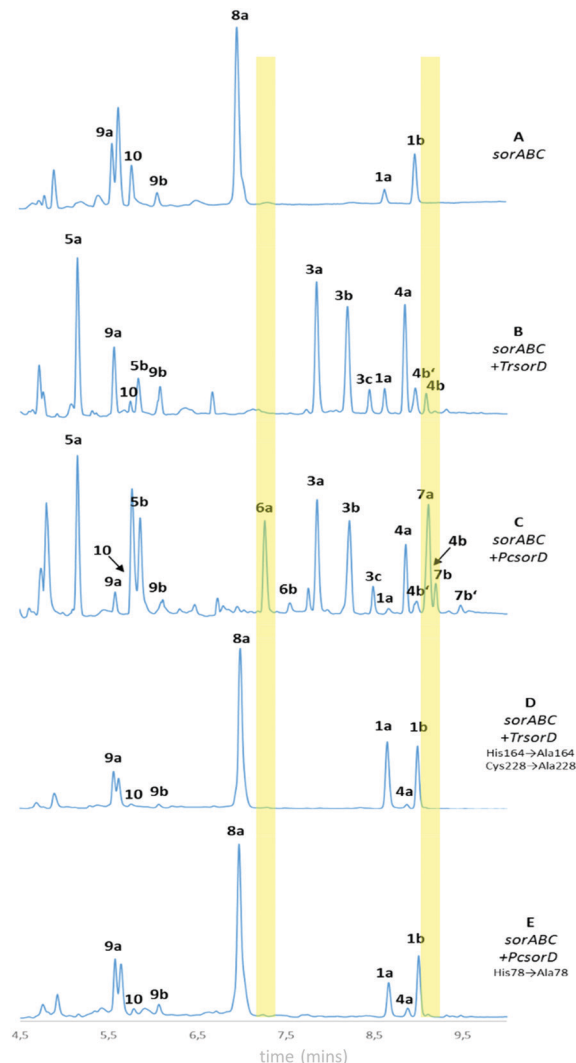

Fig. 2 LCMS chromatogram (DAD 210-600 nm) of transformants expressing solely *sorABC* (A) in addition with: (B), *TrsorD*; (C) *PcsorD*; (D) *TrsorD* (H164A/C228A); (E) PcSorD H78A. Compounds only produced in presence of PcSorD are indicated by yellow bars. **6b** and **7b/b'** are presumed to be the dihydro-forms based on physical data (Fig. S47–S51, ESI†) however could not be purified due to low titres.

in an analogous mechanism to the formation of **4** by both PcSorD and TrSorD (Scheme 1B).

PcSorD might be able to activate the epoxide-moiety of **5** and facilitate rearrangement into **6**. Rearrangements of epoxides into ketones following such a mechanism are well described reactions in organic synthesis.[23] Accordingly, we fed *A. oryzae* expressing *PcsorD* with **5a**, but no conversion to **6** was observed. To ensure that the enzyme itself was active we fed a mixture of **2ab** to the respective transformants, which resulted in formation of epoxides **5ab**, oxosorbicillinols **6ab** and dimeric sorbicillinoids **3** as anticipated for an active PcSorD (Fig. S8, ESI†). Experiments with cell-free extract (CFE) or whole cell extract (WCE) of these transformants expressing *PcsorD* also did not show conversion of **5** into **6** (Fig. S6 and S7, ESI†). This outcome might indicate that **5a** was unable to reach SorD *in vivo* or that **5** may not be the precursor to **6**.
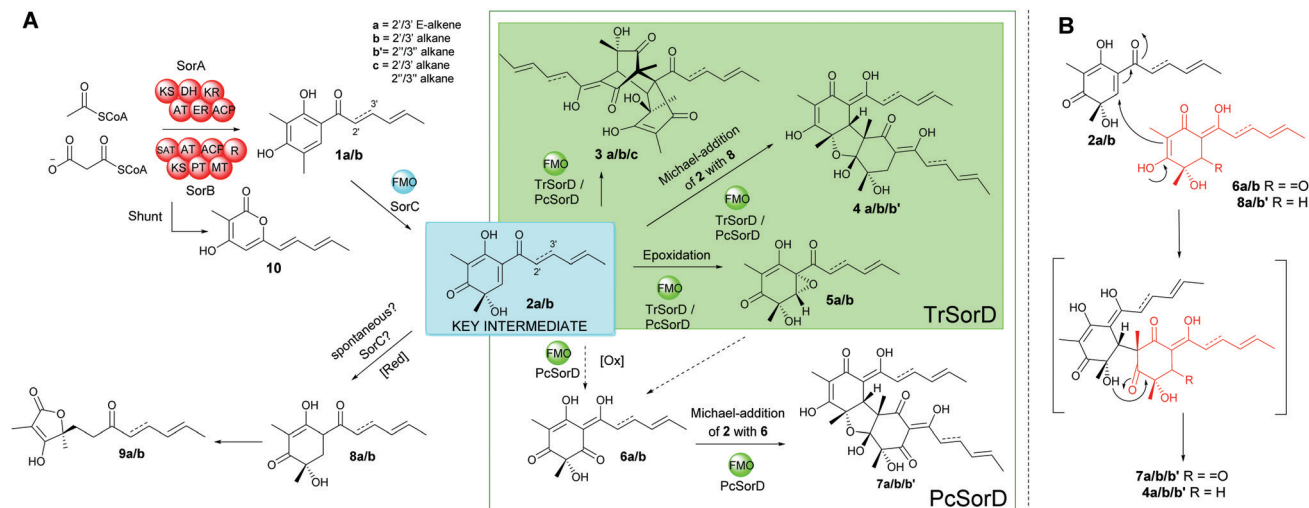
Scheme 1 (A) Overview of sorbicillinoid biosynthesis highlighting the range of reactions catalysed by TrSorD and the extended reactions catalysed by PcSorD; (B) mechanism of Michael-addition reaction between **2** and **6** or **8** to generate **7** and **4**, respectively. Stereochemistry of the Michael-dimers **4** and **7** are based on literature[29,30] and biosynthetic consideration.

In an attempt to understand the different catalytic capabilities of TrSorD and PcSorD we conducted a more detailed bioinformatic analysis using the web tool InterPro[24] which classified both PcSorD and TrSorD as vanillyl-alcohol oxidase (VAO)-type FMOs, that form a covalent bond to FAD *via* a conserved N-terminal histidine.[25] Using the protein fold recognition server Phyre2[26] the best template found for TrSorD was a VAO-type FMO MtVAO615 from the thermophilic fungus *Myceliophthora thermophila* (PDB 6F73).[27] This enzyme bi-covalently binds the flavin cofactor *via* H157 [His (N1)-8α-flavin] and C222 [Cys(S)-6-flavin]. The best template found for PcSorD was a different VAO-type FMO, XylO from *Myceliophthora thermophila* (PDB 5K8E) that exhibits the same bi-covalent linkage *via* H94 and C155.[28] Protein alignments including ancillary well studied FMO (Fig. S9 and S10, ESI†) revealed that TrSorD contains both conserved amino acid residues, H164 and C228. Strikingly, PcSorD from *P. chrysogenum* only shares the conserved N-terminal histidine residue H78, but lacks the conserved cysteine. A multiple sequence alignment including a total set of 30 FMO with available crystal structures was performed and used to build a phylogenetic tree (Fig. S11, ESI†).

As the enzyme mediated dimerization of sorbicillinol **2** is not expected to require a redox cofactor, in contrast to its oxidation to epoxysorbicillinol **5** and oxosorbicillinol **6**, we were curious whether mutating the amino acid residues covalently linking the flavin cofactor would result in the sole production of dimers **3** and **4**, while abolishing production of the oxidized molecules **5** and **6**. Accordingly, H164 and C228 (TrSorD) as well as H78 (PcSorD) were replaced with alanine to prevent covalent bonding to the flavin cofactor. Successful mutation was confirmed by sequencing (Fig. S14–S16, ESI†). Both *TrsorD* and *PcsorD* were then individually co-expressed with *sorABC* in *A. oryzae*. Cultures were grown, extracted and analysed according to standard protocols.[8] In all transformants formation of the oxidation products **5** and **6** was completely abolished

(Fig. 2D and E). Interestingly, although in significantly reduced intensity, the Michael-dimer **4a** was still detected in both TrSorD and PcSorD mutants (Fig. 2D, E and Fig. S17, ESI†). Apart from these changes, the metabolic profiles of the two mutant strains did not significantly differ from the control strain expressing *sorABC* (Fig. 2A).

These results clearly show that PcSorD and TrSorD are the enzymes required for dimerization and oxidation of **2** and that these specific conserved residues (H78 and H164/C228) are essential for enzymatic function. However, without a crystal structure we are unable to prove whether the introduced mutations completely prevent binding of the cofactor or whether we affected the overall structure of the proteins. Other studies into FMOs involving the mutation of the respective amino acid(s) responsible for covalent binding of the flavin cofactor have revealed a number of different outcomes: complete loss of catalytic activity,[31] completely insoluble protein,[32] or decreased enzyme activity due to the co-factor being bound tightly, but non-covalently within the enzyme.[33,34] In the case of TrSorD and PcSorD it appears that the respective mutations, while completely suppressing oxidation of **2**, still enable Michael-like dimerization of **2a** with **8a** to generate **4a** to a minor extent, but are not capable to facilitate the Diels–Alder like dimerization of **2** to **3**.

Finally, we attempted to identify the enzyme in *T. reesei* responsible for the formation of **6** and **7**, as although these compounds are also produced by *T. reesei*, their formation does not appear to be catalysed by TrSorD in our heterologous expression studies.[7,8] We therefore performed a protein BLAST search with PcSorD against the proteome of *T. reesei* QM6a. The top hit yielded an uncharacterized FMO (XP_006968264.1) with 35% sequence identity. Bioinformatic analysis indicates that XP_006968264.1 also putatively bi-covalently binds the flavin cofactor as sequence alignments revealed the conserved H82 and C123 residues (Fig. S9, ESI†). The corresponding gene

was cloned from *T. reesei* gDNA and co-expressed with *sorABCD* in *A. oryzae*, but neither **6**, **7**, nor any change in the metabolic profile was observed (data not shown). Therefore how **6** is formed in *T. reesei* remains unknown.

In conclusion, we elucidated the functions of the multi-functional FMO PcSorD by heterologous expression and confirmed the biosynthetic origin of oxosorbicillinol **6** and bisvertinolone **7** in *P. chrysogenum*. Surprisingly, despite being non-homologous to TrSorD, this enzyme catalyses inter-molecular Diels–Alder and Michael-like dimerization of the key intermediate sorbicillinol **2** as well as an independent oxidation of **2** to **5** and is therefore the second enzyme described to behave this way.[8] However, PcSorD is remarkable in that it clearly has additional roles in the formation of **6**, which in turn is a facile substrate for Michael-like dimerization with **2**, to form bisvertinolone **7**. Mutation studies revealed that the amino acid residues most likely responsible for covalent binding of the flavin cofactor, H78 (PcSorD) and H164/C228 (TrSorD), are crucial for enzyme oxidation activity.

Furthermore we demonstrate that enzymes that are difficult to study in conventional *in vitro* assays *e.g.* due to being insoluble can be functionally elucidated *in vivo* using heterologous gene expression and mutagenesis. This study also indicates that with extensive engineering FMOs may have the potential to be developed as specific dimerization catalysts for further diversifying small bioactive natural products.

Conflicts of interest

Notes and references

1 J. Meng, X. Wang, D. Xu, X. Fu, X. Zhang, D. Lai, L. Zhou and G. Zhang, *Molecules*, 2016, **21**, 715.
2 A. Al Fahad, A. Abood, K. M. Fisch, A. Osipow, J. Davison, M. Avramović, C. P. Butts, J. Piel, T. J. Simpson and R. J. Cox, *Chem. Sci.*, 2014, **5**, 523–527.
3 A. M. Harned and K. A. Volp, *Nat. Prod. Rep.*, 2011, **28**, 1790–1810.
4 N. Abe, T. Murata and A. Hirota, *Biosci., Biotechnol., Biochem.*, 1998, **62**, 661–666.
5 R. P. Maskey, I. Grün-Wollny and H. Laatsch, *J. Nat. Prod.*, 2005, **68**, 865–870.
6 L. Du, T. Zhu, Y. Li, S. Cai, B. Zhao and Q. Gu, *Chem. Pharm. Bull.*, 2009, **57**, 220–223.
7 C. Derntl, F. Guzmán-Chávez, T. M. Mello-de-Sousa, H. J. Busse, A. J. M. Driessen, R. L. Mach and A. R. Mach-Aigner, *Front. Microbiol.*, 2017, **8**, 2037.
8 L. Kahlert, E. F. Bassiony, R. J. Cox and E. J. Skellam, *Angew. Chem., Int. Ed.*, 2020, **59**, 5816–5822.
9 V. Massey, *J. Biol. Chem.*, 1994, **269**, 22459–22462.
10 R. Schor, C. Schotte, D. Wibberg, J. Kalinowski and R. J. Cox, *Nat. Commun.*, 2018, **9**, 1–9.
11 R. Nofiani, K. de Mattos-Shipley, K. E. Lebe, L. C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson and R. J. Cox, *Nat. Commun.*, 2018, **9**, 3940.
12 C. T. Walsh and Y.-C. J. Chen, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 333–343.
13 P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13843–13848.
14 S. Badieyan, R. D. Bach and P. Sobrado, *J. Org. Chem.*, 2015, **80**, 2139–2147.
15 K. M. Robinson and B. D. Lemire, *J. Biol. Chem.*, 1996, **271**, 4055–4060.
16 T. Rand, T. Halkier and O. C. Hansen, *Biochemistry*, 2003, **42**, 7188–7194.
17 W. McIntire, D. E. Edmondson, T. P. Singer and D. J. Hopper, *Biochemistry*, 1981, **20**, 3068–3075.
18 M. Brühmüller and K. Decker, *Eur. J. Biochem.*, 1973, **37**, 256–258.
19 E. B. Kearney, J. I. Salach, W. H. Walker, R. L. Seng, W. Kenney, E. Zeszotek and T. P. Singer, *Eur. J. Biochem.*, 1971, **24**, 321–327.
20 D. P. H. M. Heuts, N. S. Scrutton, W. S. McIntire and M. W. Fraaije, *FEBS J.*, 2009, **276**, 3405–3427.
21 C. H. Huang, W. L. Lai, M. H. Lee, C. J. Chen, A. Vasella, Y. C. Tsai and S. H. Liaw, *J. Biol. Chem.*, 2005, **280**, 38831–38838.
22 F. Guzmán-Chávez, O. Salo, Y. Nygård, P. P. Lankhorst, R. A. L. Bovenberg and A. J. M. Driessen, *Microbiol. Biotechnol.*, 2017, **10**, 958–968.
23 K. Maruoka, S. Nagahara, T. Ooi and H. Yamamoto, *Tetrahedron Lett.*, 1989, **30**, 5607–5610.
24 P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez and S. Hunter, *Bioinformatics*, 2014, **30**, 1236–1240.
25 E. De Jong, W. J. H. Van Berkel, R. P. Van der Zwan and J. A. M. De Bont, *Eur. J. Biochem.*, 1992, **208**, 651–657.
26 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protoc.*, 2015, **10**, 845–858.
27 A. R. Ferrari, H. J. Rozeboom, A. S. C. Vugts, M. J. Koetsier, R. Floor and M. W. Fraaije, *Molecules*, 2018, **23**, 111.
28 A. R. Ferrari, H. J. Rozeboom, J. M. Dobruchowska, S. S. Van Leeuwen, A. S. C. Vugts, M. J. Koetsier, J. Visser, M. W. Fraaije and G. Hart, *J. Biol. Chem.*, 2016, **291**, 23709–23718.
29 N. Abe, T. Arakawa and A. Hirota, *Chem. Commun.*, 2002, 204–205.
30 A. Sib and T. A. M. Gulder, *Angew. Chem., Int. Ed.*, 2018, **57**, 14650–14653.
31 J. Gottowik, A. M. Cesura, P. Malherbe, G. Lang and M. Da Prada, *FEBS Lett.*, 1993, **317**, 152–156.
32 R. Teufel, F. Stull, M. J. Meehan, Q. Michaudel, P. C. Dorrestein, B. Palfey and B. S. Moore, *J. Am. Chem. Soc.*, 2015, **137**, 8078–8085.
33 M. W. Fraaije, R. H. H. Van Den Heuvel, W. J. H. Van Berkel and A. Mattevi, *J. Biol. Chem.*, 1999, **274**, 35514–35520.
34 L. Motteran, M. S. Pilone, G. Molla, S. Ghisla and L. Pollegioni, *J. Biol. Chem.*, 2001, **276**, 18024–18030.

Publication 3

2.3 Biosynthesis of 6-Hydroxymellein Requires a Collaborating Polyketide Synthase-like Enzyme

Lukas Kahlert[1], Miranda Villanueva[1,2], Russell J. Cox*[1] and Elizabeth J. Skellam*[1,3]

[1] Institute for Organic Chemistry and BMWZ, Leibniz Universität Hannover, Schneiderberg 38, 30167 Hannover, Germany

[2] Current Address: The Molecular Biology Institute, UCLA, Los Angeles, California, 90095-1570, USA

[3] Current Address: Department of Chemistry & BioDiscovery Institute, University of North Texas, 1155 Union Circle 305220, Denton, Texas, 76203, USA

Type of authorship:	First author
Type of article:	Research article
Share of the work:	75%
Contribution to the publication:	Planned, performed and analyzed the majority of experiments, analysed all NMR data, wrote the first manuscript and the ESI
Journal:	Angewandte Chemie International Edition
5-Year impact factor:	15.336 (2020), 14.205 (5-year)
Date of publication:	accepted 04.03.2021
Number of citations (Google Scholar on 30.03.2021):	0
DOI:	10.1002/anie.202100969
Supplementary information:	accessible free of charge at https://doi.org/10.1002/anie.202100969

Biosynthesis

Biosynthesis of 6-Hydroxymellein Requires a Collaborating Polyketide Synthase-like Enzyme

*Lukas Kahlert, Miranda Villanueva, Russell J. Cox,* and Elizabeth J. Skellam**

Abstract: The polyketide synthase (PKS)-like protein TerB, consisting of inactive dehydratase, inactive C-methyltransferase, and functional ketoreductase domains collaborates with the iterative non reducing PKS TerA to produce 6-hydroxymellein, a key pathway intermediate during the biosynthesis of various fungal natural products. The catalytically inactive dehydratase domain of TerB appears to mediate productive interactions with TerA, demonstrating a new mode of trans-interaction between iterative PKS components.

Introduction

Fungal polyketide synthases (PKS) are iterative Type I systems in which catalytic domains form large programmed covalent complexes that can synthesise a very wide variety of structures.[1,2] Fungal highly reducing (hr) PKS resemble mammalian fatty acid synthases[3] (mFAS) in domain organisation, while fungal non-reducing (nr) PKS are simpler and lack ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) β-processing domains. In almost all known cases the catalytic domains are *cis*-acting: that is, they are part of the covalent complex. However, in some cases *trans*-acting enoyl reductase (ER)[4] and (very rarely) *trans*-*C*-methyl transferase (*C*-MeT)[5] and *trans*-KR domains[6] are known. Other *trans*-acting interactions are also possible. For example nr-PKS usually have a starter-unit acyl transferase (SAT) domain which can accept an acyl chain priming unit directly from the acyl carrier protein (ACP) of an hr-PKS or FAS.[7]

[*] L. Kahlert, M. Villanueva, Prof. R. J. Cox, Prof. E. J. Skellam
Institute for Organic Chemistry and BMWZ
Leibniz Universität Hannover
Schneiderberg 38, 30167 Hannover (Germany)
E-mail: russell.cox@oci.uni-hannover.de

M. Villanueva
Current address: The Molecular Biology Institute, UCLA
Los Angeles, CA 90095-1570 (USA)

Prof. E. J. Skellam
Current address: Department of Chemistry & BioDiscovery Institute
University of North Texas
1155 Union Circle 305220, Denton, TX 76203 (USA)
E-mail: Elizabeth.skellam@unt.edu

Recently, Brock and co-workers described the biosynthetic gene cluster (BGC) of the fungal polyketide terrein **1** in *Aspergillus terreus*, which is derived from the polyketide precursor 6-hydroxymellein **2**.[8] The *ter* BGC encodes TerA which is an apparently canonical nr-PKS consisting of SAT, ketosynthase (KS), acyl transferase (AT), product template (PT), twin ACP, and N-terminal thiolesterase (TE) domains (Scheme 1). Heterologous expression of TerA in *Aspergillus niger* results in the formation of three products: the triketide lactone **3**, the tetraketide orsellinic acid **4**, and the pentaketide 6,7-dihydroxymellein **5** (Scheme 1) that arises via shunt hydroxylation of proposed TerA product **6** to **7** followed by reduction.
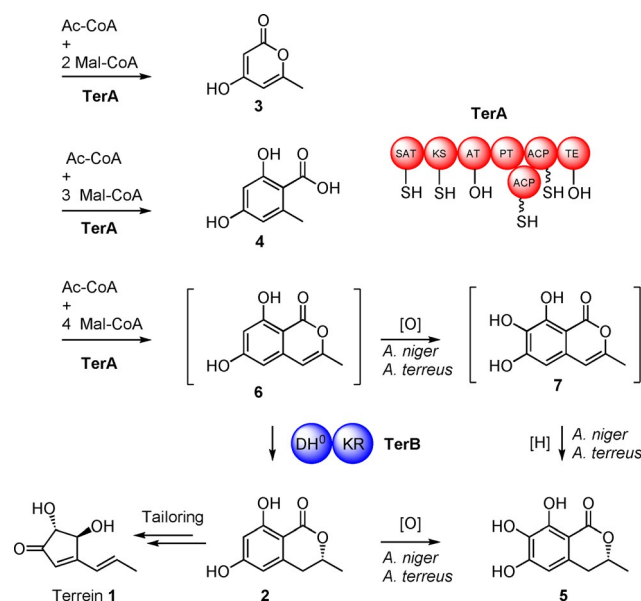
Scheme 1. Summary of previous observations. Compounds in brackets were not isolated. See main text for abbreviations.

A second protein, TerB, was shown to be essential for biosynthesis of **2** in *A. terreus* by gene knock-out. In the absence of TerB **3**, **4**, and **5** were isolated. However, co-expression of *terAB* in *A. niger* did not lead to the expected production of **2**. Due to the observation of **2** in additional knock-out studies of later tailoring enzymes in *A. terreus*, the role of TerB was proposed as opening lactone **6** followed by subsequent reduction and relactonisation. We have previously used heterologous expression of fungal BGC in *Aspergillus oryzae* to successfully investigate the biosynthesis of other metabolites without the complications of shunt pathways and we therefore decided to use it to investigate this intriguing system in vivo.[9,10]

Results

Expression of terA and terB

Genes encoding TerA and TerB were cloned from *A. terreus* cDNA and/or gDNA (see SI for all experimental details). The fungal expression vector pTYGSarg-*terA* was assembled by yeast homologous recombination and expressed in *A. oryzae* NSAR1.[11] Transformants expressing *terA* yield three compounds: the triketide pyrone **3** (71 mg L^{-1}), the tetraketide orsellinic acid **4** (153 mg L^{-1}), and the pentaketide acid **8** (91 mg L^{-1}, Figure 1 B, Figure 2), all confirmed by full structure elucidation by NMR and HRMS and comparison to literature values. Compound **8** is unshunted in *A. oryzae*, unlike in *A. niger* where it is evidently subjected to oxidative modifications by host enzymes. In contrast to previous reports by Oakley et al. the pentaketide lactone **6** was not observed (SI Figure S9).[12]

Close examination of the coding region of *terB* indicated that the NCBI sequence was likely truncated. Reanalysis identified an additional 193 amino acid residues at the C-terminus that contribute to a KR domain (SI Figures S1–S8). Conserved domain database (CDD)[13] analysis of the longer ORF identifies an N-terminal DH domain lacking key active site residues (DH^0) as well as the C-terminal KR. In addition, InterPro[14,15] and Phyre2[16] further identify an inactive *C*-methyltransferase (*C*-MeT^0) located centrally. Alignment with a number of characterised hr-PKS also indicates sequence homology with *C*-MeT domains. TerB thus appears to be a truncated hr-PKS that has lost KS, AT, ER, and ACP domains. Since reduction of the 9-ketone of **8** is required during biosynthesis of **2**, and TerA does not contain a KR domain, it seemed possible that TerB fulfils this role.

Coexpression of the longer *terB* with *terA* in *A. oryzae* results in production of *R*-6-hydroxymellein **2** (Figure 1 C, Figure 2, see SI for full characterisation) in very high titres (> 500 mg L^{-1}). Prolonged incubation times (> 4 days) result in hydroxylation of **2** by *A. oryzae* to give 5,6-dihydroxymellein as reported previously (SI Figure S11).[17] In addition to **2**, **3**, **4**, and **8**, a new minor compound was observed and 2D-NMR analysis revealed it to be the tetraketide lactone **9**, showing a secondary alcohol function in the aliphatic side chain (SI Figs S63–S67). The *R*-configuration was assigned by Mosher's esterification (SI Figure S68).[18,19]

Coexpression of *terAB* gives **2** at five-fold higher titre (530 mg L^{-1}) than production of pentaketide **8** by expression of *terA* alone. Based on this observation we envisaged two potential biosynthetic routes to **2** (Scheme 2). First, early release would provide free ketone **8** which could be reduced by the KR of TerB to alcohol **10**, followed by lactonisation. In a second possibility, ketoreduction could occur first while the polyketide, for example, **8A**, is still tethered to TerA. In this case final release could be spontaneous by intramolecular lactonisation, or release could be catalysed by the TE, and conceivably these possibilities could occur while the intermediate is bound to either ACP (**8A/10A**) or TE (**8T/10T**).
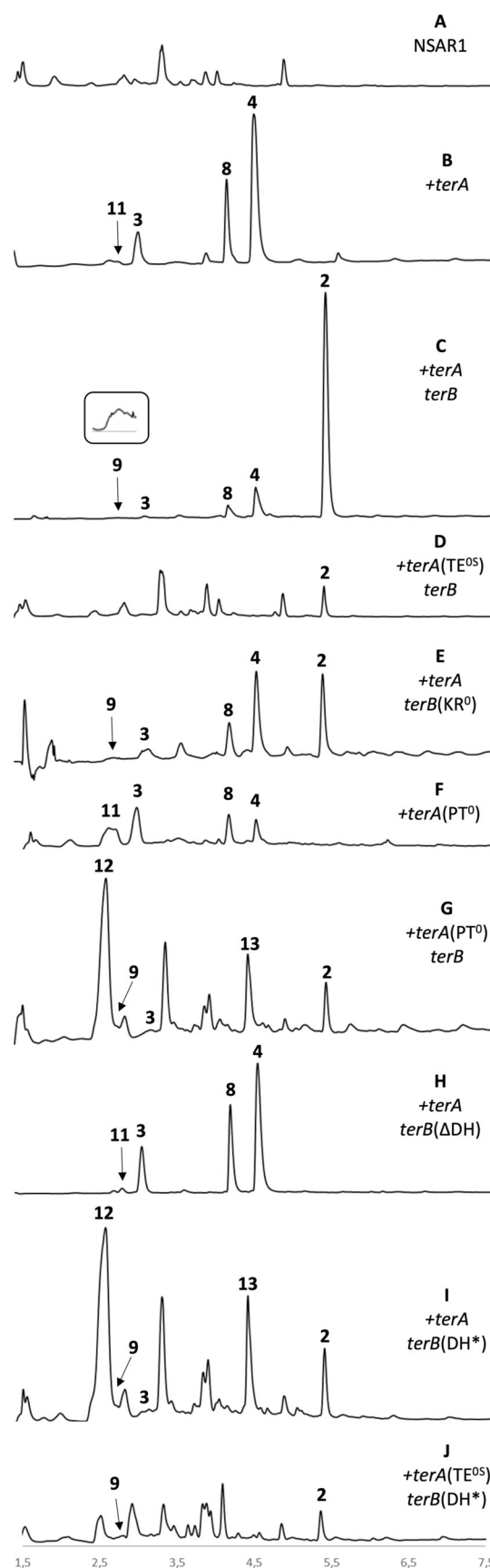

Figure 1. Selected DAD (210–600 nm) chromatograms of the indicated *A. oryzae* expression experiments. Unlabelled peaks are unrelated.
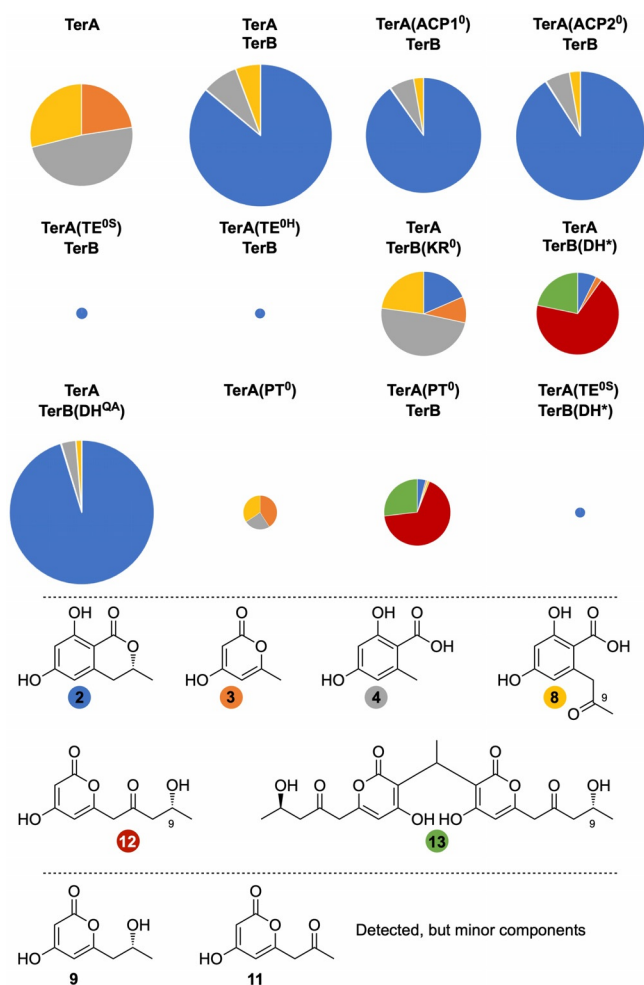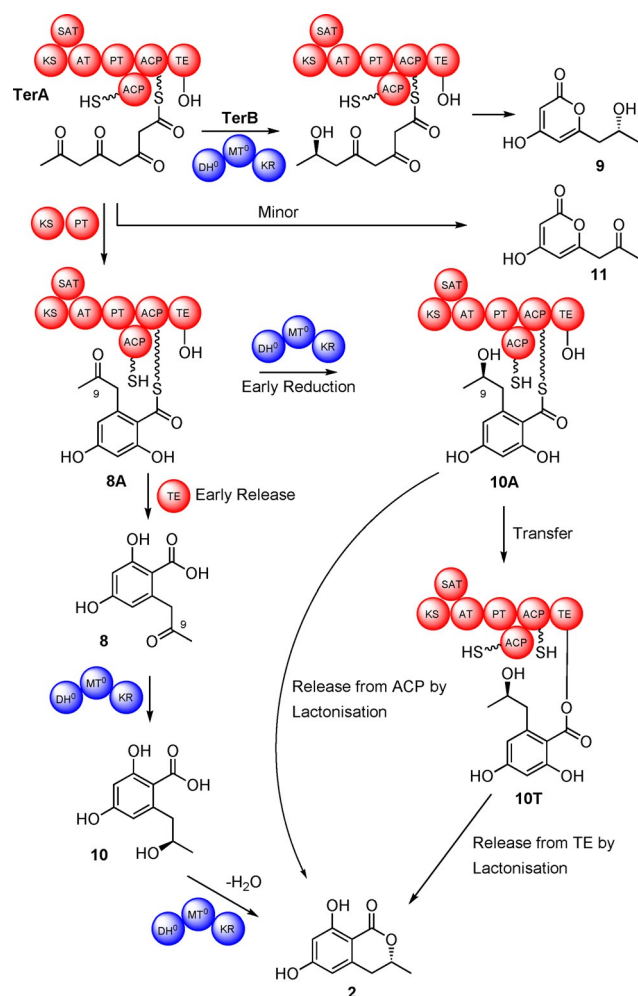
Figure 2. Titres of isolated compounds from the indicated heterologous expression experiments in *A. oryzae*. Area of pie charts is proportional to total titre in each case. Results of lower production experiments not shown. All titres are also summarised in Table S6.

Scheme 2. Possible reduction and release pathways during the formation of **2**.

In Vitro Study of Reduction and Lactonisation

In order to probe these release possibilities we generated *A. oryzae* transformants solely expressing *terB* and prepared a cell-free extract (CFE). The CFE was incubated with pentaketide **8** in the presence of NADPH for 3 h, followed by LCMS. Ketone **8** was converted to **2** only in the presence of TerB and NADPH, but not in the untransformed *A. oryzae* NSAR1 control (SI Figure S12). The intermediate secondary alcohol **10** was not detected suggesting that TerB also catalyses lactonisation. This is further supported by studies of the chemical properties of alcohol **10** generated by reduction of **8** with NaBH$_4$.

Alcohol **10** is stable in water and in the *A. oryzae* CFE (SI Figure S15), but upon acidification immediately converts to **2** (SI Figures S13,S14). These results show that the TerB KR can reduce free **8**, but do not rule out the possibility that TerB can also reduce enzyme-bound intermediates such as **8A** (Scheme 2).

ACP, TE, and KR Active Site Mutations

TerA contains tandem ACP domains and we were curious whether one ACP domain specifically interacts with TerB. In other PKS or polyunsaturated fatty acid synthase (PUFAS) systems individual ACP domains within a tandem repeat are often functionally equivalent.[20] The conserved serine in each TerA ACP was mutated. TerA(ACP1^0) carries an S1674A mutation; TerA(ACP2^0) has S1793A, while TerA(ACP12^0) has both mutations. The respective *terA* mutants were co-expressed with *terB* and production of **2**, **3**, **4**, and **8** was monitored and quantified (Figure 2). Production of **2** falls to approximately 65–80% in TerA(ACP1^0) and TerA(ACP2^0), however it is still 3.5–4.5 times higher than production of pentaketide **8** by native TerA alone (Figure 2). Production of all polyketides was abolished in TerA(ACP12^0). It appears that possession of two ACP domains increases the efficiency of TerA, but both ACP appear to be functionally identical in these experiments.

Next, we attempted to determine whether the TE domain in TerA is required for the formation of **2**, or whether TerB directs the release of the pentaketide intermediate. S1968 was identified as the active site nucleophile and H2117 as the

active site base of the TE domain (SI Figure S7) and we created TerA(TE^{0S}) and TerA(TE^{0H}), respectively, by mutation of these residues to alanine.[21] Expression of either *terA*(TE^{0S}) or *terA*(TE^{0H}) in *A. oryzae* almost completely abolishes production of TerA-derived polyketides: only orsellinic acid **4** was observed in extracted ion chromatograms (EIC) at barely detectable levels indicating that catalysis by the TE domain is required for release of *any* product, including pyrone **3**, confirming the previous observation of Brock and co-workers.[8]

Co-expression of *terA*(TE^0) with *terB* similarly significantly affects production of TerA-derived polyketides. These mutants produce mostly **2**, but overall titres drop 100-fold to below 5 mg L^{-1} (Figure 1 D, Figure 2). The tetraketide lactone **9** was also detected in minor amounts, however low overall titres prevented quantification. The striking observation is that once again co-expression of *terB* with *terA* is more productive than TerA alone: even with a non-functional TE domain, TerA(TE^0)/TerB leads to **2** while TerA(TE^0) alone cannot produce the corresponding pentaketide **8**. These results suggest that **8A** accumulates on the TerA ACP and is reduced by the TerB KR to **10A**. Release by lactonisation must be slow, blocking the productivity of TerA. This shows that the TerB KR is unlikely to be able to catalyse significant lactonisation and product release for ACP-bound intermediate **10A** and that the TE must catalyse the intramolecular release.

We next turned our attention to the KR domain of TerB. The conserved acid likely to activate the substrate carbonyl for reduction by protonation was identified as Y991 (SI Figure S5).[22] TerB(KR^0) was created by forming mutant Y991F. Co-expression of *terA* with *terB*(KR^0) results in reduction of the titre of **2** by 90% confirming that the KR domain in TerB is active and essential for biosynthesis of **2** (Figure 1 E, Figure 2).

PT and DH^o Mutations

To confirm the role of the TerA PT domain, its catalytic histidine, required for C−C bond formation, was identified by sequence alignment[23] as H1320 (SI Figure S8). The H1320A mutant was created and is referred to as TerA(PT^0). Expression of *terA*(PT^0) alone results in a significant decrease in production of **3**, **4**, and **8** as expected (Figure 1 F). Careful examination of the culture extract revealed formation of tetraketide ketone **11** (3 mg L^{-1}, Figure 1 F, Scheme 2) and also allowed it to be identified in the expression of *terA* alone (Figure 1 B) as a minor component.

The lack of any pentaketide pyrone products formed by the PT^0 mutant, together with the low titres of **3** and **11** in particular, suggests that TerA(PT^0) has stalled at the tetraketide stage and shows that the TE is unable to efficiently remove the linear tetraketide and pentaketide intermediates via catalysed pyrone formation or by intermolecular hydrolysis with H$_2$O.

When *terA*(PT^0) is co-expressed with *terB*, two new major compounds **12** and **13** appear in high titre (Figure 1 G, Figure 2). Purification and full structure elucidation (SI Figures S81–92) revealed that **12** is a pentaketide pyrone, while **13** is a dimer of **12** formally fused to acetaldehyde and both are reduced at C-9. Dimeric compounds related to **13** have been observed before,[24] but not previously in *A. oryzae*, indicating the presence of a highly unusual shunt pathway.

In order to probe the role of the TerB DH^0 domain, *terA* was co-expressed with *terB* lacking its initial 320 codons (*terBΔDH*). In this experiment no TerB-related compounds were produced (Figure 1 H, SI Figure S16) and the chemotype is identical to expression of *terA* alone (Figure 1 B), suggesting that the DH^0 domain is required for structural integrity of TerB.

Alignment of the TerB DH^0 with well-studied PKS DH domains such as that from the squalestatin tetraketide synthase SQTKS[25] reveals that the TerB DH^0 lacks the catalytic histidine within the highly conserved 46-H(X$_3$)G-(X$_4$)P-55 motif.[26] A glutamine residue replaces the histidine. In contrast, the second conserved 215-D(X$_3$)Q-219 motif is present. As no catalytic role for the DH^0 domain of TerB was immediately obvious, we attempted to re-activate the domain by introducing the active-site H46 and observe any effect on polyketide production. TerB(DH*) was created by introducing mutation Q46H. Co-expression of *terA* with *terB*(DH*) results in a 95% reduction in production of **2**, and **3**, **4**, and **8** are almost undetectable (Figure 1 I, Figure 2). No unsaturated congeners were observed, showing that TerB(DH*) remains deficient in dehydratase activity.

Remarkably, pyrone **12** and its related dimer **13** again constitute the major compounds produced. The combined titre of the pentaketides **12** (approx. 150 mg L^{-1}) and **13** (47 mg L^{-1}) is twice as high as the pentaketide product **8** naturally produced by TerA alone, and almost no tetraketide product is observed. Not only does the TerB(DH*) mutant change the product profile, it also increases the efficiency of TerA towards the formation of pentaketide products. The structures of **12** and the tetraketide **9** indicate that the TerA PT domain has failed to function, since neither of them arise via C-2/C-7 cyclisation evident in the TerA-derived aromatic products **4** and **8**.

The result of the DH* mutation is almost identical to the TerA(PT^0)/TerB experiment and shows that the effect of the DH* mutation is to inactivate the TerA PT. This is a significant finding and can only be explained by the collaborative action of TerB that enables the release of linear intermediates by pyrone formation when the PT domain of TerA is inactive. Additionally, since almost no **2** is produced it can be excluded that the DH domain in TerB functions as a PT domain that aids C-2/C-7 cyclisation. This result also suggests that reduction of C-7 at the tetraketide stage represses early C-2/C-7 cyclisation—all systems with defective or missing KR make considerably more tetraketide orsellinic acid **4**.

To determine whether the PT domain programming in TerA is disrupted by TerB(DH*) specifically due to the presence of H46, as opposed to the lack of the native Q46, we generated the Q46A mutation [TerB(DH^{QA})]. Co-expression of *terA* with *terB*(DH^{QA}) leads to slightly increased titres of **2** (600 mg L^{-1}, Figure 2), confirming that specifically the presence of H46 in TerB inactivates the PT domain of TerA.

To establish whether it is TerB that releases the linear tetraketide and pentaketide intermediates from the ACP or the TE domain of TerA, we co-expressed *terA*(TE^0S) with *terB*(DH*) that also effectively inactivates the TerA PT domain. Similarly to co-expression of *terA*(TE^0S) with the native *terB*, production of **2** is significantly affected and decreased by almost 95 % (< 3 mg L⁻¹) while tetraketide lactone **9** is still found in minor amounts (Figure 1 J, Figure 2). However, no pentaketide pyrone **12** (or its dimer **13**) is observed. This result demonstrates that TerB does not catalyse release of polyketide intermediates while they are attached to the ACP, and therefore shows that release occurs for intermediates bound to the TE.

Discussion and Conclusion

Taken together our results strongly support a collaborative model for TerA and TerB during the biosynthesis of **2** (Scheme 3). In all cases TerA is more productive in the presence of TerB suggesting a close interaction. The fact that mutation in TerB (e.g. DH*) affects chemistry catalysed by TerA (e.g. PT) strongly supports this conclusion.

Our results support a biosynthetic scheme in which triketide intermediates can only be extended, or released as the pyrone **3**, and that the TE is required for this step. Tetraketides are rarely released as a pyrone **11** for the WT system, but more often when the PT domain is disabled. Formation of **9** could come from release of a tetraketide 7-alcohol, or by reduction of **11**. In the presence of an active PT, tetraketides are cyclised to orsellinic acid **4** and then released by TE. This pathway is enhanced in systems lacking KR.

However, in the presence of KR, tetraketide intermediates prefer to be extended, suggesting that reduction occurs at the tetraketide stage which suppresses formation of **4**.

At the pentaketide stage the WT system prefers PT-catalysed C-2/C-7 Claisen cyclisation to form **10A**. Final transfer to the TE then leads to fast intramolecular cyclisation and release of **2**. Disruption of the Claisen cyclisation, either by mutation of PT, or by formation of the TerB DH* mutation, leads to effective off-loading of pentaketide 9-alcohol **12** as a pyrone. Importantly, pentaketide 9-keto pyrones are not observed, again suggesting that reduction occurs before release, either of the tetraketide or pentaketide intermediates.

This observation that mutation of PT and DH leads to almost identical results suggests that the TerA PT and TerB DH domains are likely to interact. This is an appealing hypothesis since, based on their structural (double hotdog folds) and catalytic features (His-Asp dyads), PT domains in nr-PKS are described as evolutionary ancestors of DH domains from hr-PKS.[27,28] Interestingly, it is known that while the overall fold of DH and PT domains is very similar, DH domains dimerise at their N-termini, while PT domains have a different dimerisation surface which leaves the N-termini free.[29] It is tempting to speculate that the N-terminus of the TerB DH may interact with the N-termini of the PT domains in TerA. Interaction via a PT–DH interaction is also supported by our observation that the TerB lacking the DH
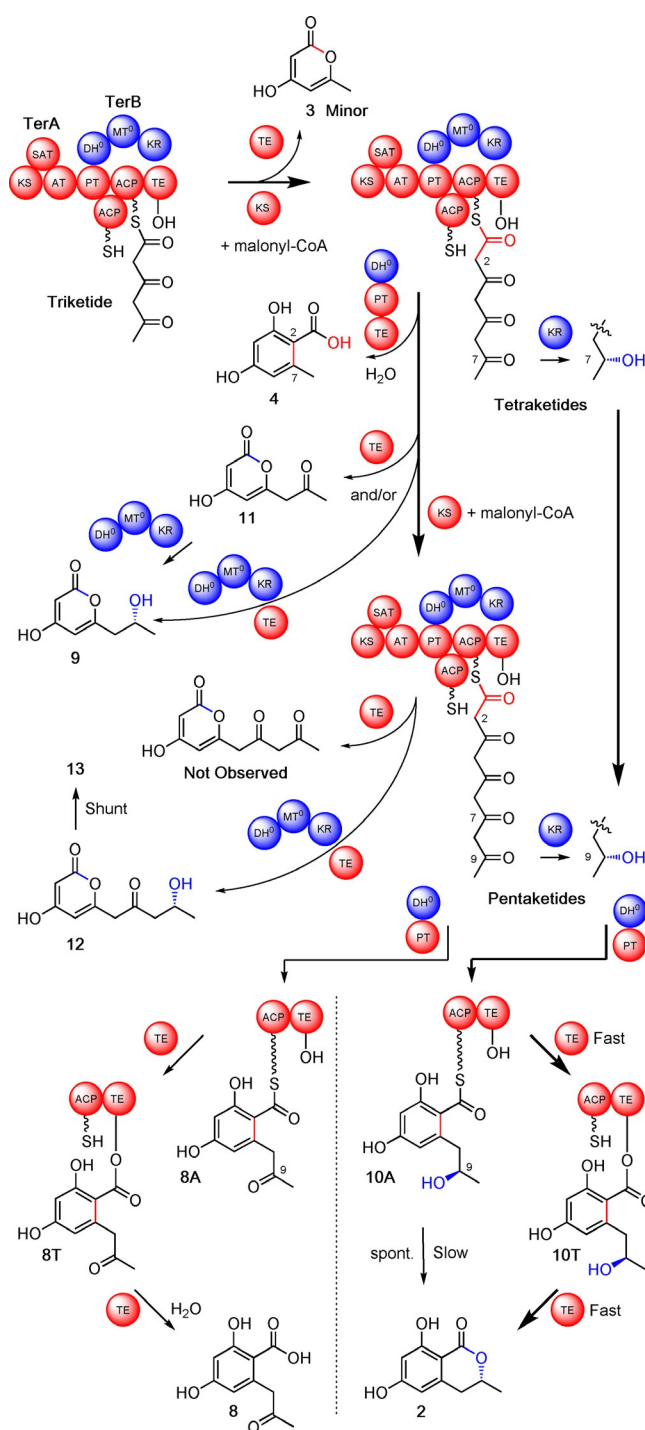
Scheme 3. Overall description of the interaction of TerA and TerB during the biosynthesis of **2**. Bold arrows show main pathway flux of the WT system.

domain is inactive in the presence of TerA. However, structural work, which is currently not possible due to non-availability of soluble proteins, will be required to confirm this hypothesis.

It is also noteworthy that TerB(DH*) is unable to catalyse dehydration as no unsaturated products were observed. This is perhaps unsurprising since PKS and FAS DH domains normally dehydrate β-hydroxy thiolesters[25] rather than β-

hydroxy ketones, consistent with TerAB reduction occuring at the tetraketide stage or later.

Mutation of the TE domain shows that it is required for all release modes, including intermolecular hydrolysis in the formation of **4** and **8**, and intramolecular reactions involved in the formation of pyrones (**3**, **9**, **11**, **12**) as well as release of **2** from **10T**. In the TE[0S] mutant, transfer of intermediates to the TE is not possible, and formation of **2** in this case must come from slow spontaneous intramolecular release from ACP-bound intermediate **10A**. While pyrone formation is often regarded as a default release mechanism when other catalysed routes are absent, some other nr-PKS TE domains, such as that of the cercosporin PKS described by Townsend and co-workers,[30] are also known to catalyse this type of release.

Other examples of collaborating and *trans*-acting enzymes have been reported in fungi, including: nr-PKS SAT domains,[31] *trans*-TE domains that are necessary for chain length determination and product release,[32] *trans*-ER domains that perform specific enoyl reduction prior to release from the PKS,[33, 34] and an unusual ψACP-*C*-MeT protein that binds to the KS domain of an nr-PKS enabling methylation of the polyketide intermediate.[35] Tan and co-workers recently reported a short-chain dehydrogenase/reductase (SDR) ChrB which interacts with an iterative hr-PKS ChrA during the biosynthesis of precursors of dalmanol,[6] but ChrB is unrelated to the KR of TerB ($< 15\%$ identity) which forms part of a truncated hr-PKS itself. Tang and co-workers also reported an interesting interaction between a bacterial Type II KR (act III) and the fungal nr-PKS4 from *G. fujikuroi*. Interestingly, this KR can also reduce at C-9, but it appears to reduce full-length octa- and nona-ketides which then prevents the function of PT,[36] unlike the case of the TerAB system where immature tetraketides are reduced and PT activity enhanced. Thus the proposed DH–PT interaction in TerAB appears unique. Although some of these *trans*-acting systems have been studied in detail, the precise interaction points have not yet been determined due to a lack of structural data. Future investigations will also be necessary to probe the proposed interactions in the TerAB system using a combination of in vitro domain assays with synthetic intermediates to determine the precise timing of reduction.[37]

Finally, these results also highlight the importance of melleins as fungal secondary metabolite intermediates. The pentaketide mellein skeleton is known to be produced by at least four independent systems in fungi. As shown here TerA and TerB form an unusual collaborative PKS system combining nr-PKS and hr-PKS components. CdmE is an hr-PKS from the chrodrimanin B pathway that directly synthesises **2**.[38] Men2 is a nr-PKS, involved in ascotrichalactone A biosynthesis, that synthesises **2** as a derailment product.[39] Finally, mellein itself is synthesised by a partially reducing PKS (pr-PKS)[1, 40] with a programmed *cis*-acting KR. All four systems contain enzymes with differently constructed and programmed KR domains. The TerAB-type architecture appears relatively common, being also known to be involved in the biosynthesis of other 6-HM-derived compounds such as roussoellatide,[41] cyclohelminthols, and palmaenones.[17] It remains to be seen whether nature also uses this architecture in other systems.

Conflict of interest

The authors declare no conflict of interest.

[1] R. J. Cox, *Org. Biomol. Chem.* **2007**, *5*, 2010–2026.

[2] Y.-H. Chooi, Y. Tang, *J. Org. Chem.* **2012**, *77*, 9933–9953.

[3] O. Piech, R. J. Cox, *RSC Adv.* **2020**, *10*, 18469–18476.

[4] J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas, C. R. Hutchinson, *Science* **1999**, *284*, 1368–1372.

[5] S. Kishimoto, Y. Tsunematsu, T. Matsushita, K. Hara, H. Hashimoto, Y. Tang, K. Watanabe, *Biochemistry* **2019**, *58*, 3933–3937.

[6] Z. Z. Zhou, H. J. Zhu, L. P. Lin, X. Zhang, H. M. Ge, R. H. Jiao, R. X. Tan, *Chem. Sci.* **2019**, *10*, 73–82.

[7] J. M. Crawford, B. C. Dancy, E. A. Hill, D. W. Udwary, C. A. Townsend, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16728–16733.

[8] C. Zaehle, M. Gressler, E. Shelest, E. Geib, C. Hertweck, M. Brock, *Chem. Biol.* **2014**, *21*, 719–731.

[9] R. Nofiani, K. de Mattos-Shipley, K. E. Lebe, L.-C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson, R. J. Cox, *Nat. Commun.* **2018**, *9*, 3940.

[10] L. Kahlert, E. F. Bassiony, R. J. Cox, E. J. Skellam, *Angew. Chem. Int. Ed.* **2020**, *59*, 5816–5822; *Angew. Chem.* **2020**, *132*, 5865–5871.

[11] K. A. K. Pahirulzaman, K. Williams, C. M. Lazarus, *Methods Enzymol.* **2012**, *517*, 241–260.

[12] Y.-M. Chiang, E. C. Oakley, M. Ahuja, R. Entwistle, A. Schultz, S.-L. Chang, C. T. Sung, C. C. Wang, B. R. Oakley, *J. Am. Chem. Soc.* **2013**, *135*, 7720–7731.

[13] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, et al., *Nucleic Acids Res.* **2017**, *45*, D200–D203.

[14] M. Blum, et al., *Nucleic Acids Res.* **2021**, *49*, D344—D354.

[15] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, S. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, *Bioinformatics* **2014**, *30*, 1236–1240.

[16] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, *10*, 845–858.

[17] T. Ugai, A. Minami, S. Tanaka, T. Ozaki, C. Liu, H. Shigemori, M. Hashimoto, H. Oikawa, *ChemBioChem* **2020**, *21*, 360–367.

[18] T. R. Hoye, C. S. Jeffrey, F. Shao, *Nat. Protoc.* **2007**, *2*, 2451–2458.

[19] J. M. Seco, E. Quiñoá, R. Riguera, *Chem. Rev.* **2004**, *104*, 17–117.

[20] A. Watanabe, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake, Y. Ebizuka, *Tetrahedron Lett.* **1999**, *40*, 91–94; T. Thongkawphueak, A. J. Winter, C. Williams, H. J. Maple, S.

Soontaranon, C. Kaewhan, D. J. Campopiano, M. P. Crump, P. Wattana-Amorn, *Biochemistry* **2021**, *60*, 219–230; T. A. M. Gulder, M. F. Freeman, J. Piel, *Top. Curr. Chem.* **2011**, *298*; L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis, D. H. Sherman, *Angew. Chem. Int. Ed.* **2011**, *50*, 2795–2798; *Angew. Chem.* **2011**, *123*, 2847–2850; H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen, B. Shen, *J. Am. Chem. Soc.* **2008**, *130*, 6336–6337.

[21] I. Fujii, A. Watanabe, U. Sankawa, Y. Ebizuka, *Chem. Biol.* **2001**, *8*, 189–197.

[22] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.

[23] Y. Xu, T. Zhou, Z. Zhou, S. Su, S. A. Roberts, W. R. Montfort, J. Zeng, M. Chen, W. Zhang, M. Lin, J. Zhan, I. Molnár, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5398–5403.

[24] H. V. Kemami Wangun, C. Hertweck, *Eur. J. Org. Chem.* **2007**, 3292–3295.

[25] E. Liddle, A. Scott, L.-C. Han, D. Ivison, T. J. Simpson, C. L. Willis, R. J. Cox, *Chem. Commun.* **2017**, *53*, 1727–1730.

[26] L. Tang, Y. J. Yoon, C. Y. Choi, C. R. Hutchinson, *Gene* **1998**, *216*, 255–265; A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941–953; D. Gay, Y. O. You, A. Keatinge-Clay, D. E. Cane, *Biochemistry* **2013**, *52*, 8916–8928.

[27] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S. C. Tsai, C. A. Townsend, *Nature* **2009**, *461*, 1139–1143.

[28] J. F. Barajas, K. Finzel, T. R. Valentic, G. Shakya, N. Gamarra, D. Martinez, J. L. Meier, A. L. Vagstad, A. G. Newman, C. A. Townsend, M. D. Burkart, S. C. Tsai, *Angew. Chem. Int. Ed.* **2016**, *55*, 13005–13009; *Angew. Chem.* **2016**, *128*, 13199–13203.

[29] D. A. Herbst, C. A. Townsend, T. Maier, *Nat. Prod. Rep.* **2018**, *35*, 1046–1069.

[30] A. G. Newman, A. L. Vagstad, K. Belecki, J. R. Scheerer, C. A. Townsend, *Chem. Commun.* **2012**, *48*, 11772–11774.

[31] J. M. Winter, D. Cascio, D. Dietrich, M. Sato, K. Watanabe, M. R. Sawaya, J. C. Vederas, Y. Tang, *J. Am. Chem. Soc.* **2015**, *137*, 9885–9893.

[32] A. O. Zabala, Y. H. Chooi, M. S. Choi, H. C. Lin, Y. Tang, *ACS Chem. Biol.* **2014**, *9*, 1576–1586.

[33] A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *J. Am. Chem. Soc.* **2011**, *133*, 10990–10998.

[34] B. D. Ames, C. Nguyen, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. H. Li, J. C. Vederas, Y. Tang, S. C. Tsai, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11144–11149.

[35] M. Chen, Q. Liu, S. S. Gao, A. E. Young, S. E. Jacobsen, Y. Tang, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5499–5504.

[36] S. M. Ma, J. Zhan, X. Xie, K. Watanabe, Y. Tang, W. Zhang, *J. Am. Chem. Soc.* **2008**, *130*, 38–39.

[37] J. S. Parascandolo, J. Havemann, H. K. Potter, F. Huang, E. Riva, J. Connolly, I. Wilkening, L. Song, P. F. Leadlay, M. Tosin, *Angew. Chem. Int. Ed.* **2016**, *55*, 3463–3467; *Angew. Chem.* **2016**, *128*, 3524–3528.

[38] T. Bai, Z. Quan, R. Zhai, T. Awakawa, Y. Matsuda, I. Abe, *Org. Lett.* **2018**, *20*, 7504–7508.

[39] W. Bunnak, P. Wonnapinij, A. Sriboonlert, C. M. Lazarus, P. Wattana-Amorn, *Org. Biomol. Chem.* **2019**, *17*, 374–379.

[40] Y. H. Chooi, C. Krill, R. A. Barrow, S. Chen, R. Trengove, R. P. Oliver, P. S. Solomon, *Appl. Environ. Microbiol.* **2015**, *81*, 177–186.

[41] E. L. Ferreira, D. E. Williams, L. P. Ióca, R. P. Morais-Urano, M. F. Santos, B. O. Patrick, L. M. Elias, S. P. Lira, A. G. Ferreira, M. R. Passarini, et al., *Org. Lett.* **2015**, *17*, 5152–5155.

Publication 4

2.4 Early Oxidative Transformations During the Biosynthesis of Terrein and Related Natural Products

Lukas Kahlert[1], Darlon Bernardi[1,2], Maurice Hauser[1], Laura P. Ióca[2], Roberto G. S. Berlinck[2], Elizabeth J. Skellam[1,3] and Russell J. Cox*[1]

[1] Institute for Organic Chemistry and BMWZ, Leibniz Universität Hannover, Schneiderberg 38, 30167 Hannover, Germany

[2] Instituto de Química de São Carlos, Universidade de São Paulo, CP 780, CEP 13560-970, São Carlos, SP, Brazil

[3] Current Address: Department of Chemistry & BioDiscovery Institute, University of North Texas, 1155 Union Circle 305220, Denton, Texas, 76203, USA

Type of authorship:	First author
Type of article:	Research article
Share of the work:	75%
Contribution to the publication:	Planned, performed and analyzed all heterologous expression-, feeding- and *in vitro* experiments, analysed NMR data, drafted the current manuscript and the ESI
Journal (targeted):	Chemistry-A European Journal
5-Year impact factor:	5.236 (2020), 4.843 (5-year)
Date of publication:	accepted 11.06.2021
Number of citations (Google Scholar on 11.06.2021):	0
DOI:	10.1002/chem202101447
Supplementary information:	accessible free of charge at https://doi.org/10.1002/chem202101447

Early Oxidative Transformations During the Biosynthesis of Terrein and Related Natural Products

Lukas Kahlert,[a] Darlon Bernardi,[a, b] Maurice Hauser,[a] Laura P. Ióca,[b]
Roberto G. S. Berlinck,[b] Elizabeth J. Skellam,[a, c] and Russell J. Cox*[a]

Abstract: The mycotoxin terrein is derived from the C_{10}-precursor 6-hydroxymellein (6-HM) via an oxidative ring contraction. Although the corresponding biosynthetic gene cluster (BGC) has been identified, details of the enzymatic oxidative transformations are lacking. Combining heterologous expression and in vitro studies we show that the flavin-dependent monooxygenase (FMO) TerC catalyzes the initial oxidative decarboxylation of 6-HM. The reactive intermediate is further hydroxylated by the second FMO TerD to yield a highly oxygenated aromatic species, but further reconstitution of the pathway was hampered. A related BGC was identified in the marine-derived *Roussoella sp.* DLM33 and confirmed by heterologous expression. These studies demonstrate that the biosynthetic pathways of terrein and related (polychlorinated) congeners diverge after oxidative decarboxylation of the lactone precursor that is catalyzed by a conserved FMO and further indicate that early dehydration of the side chain is an essential step.

Introduction

The mycotoxin terrein **1**, first isolated in 1935 from the ascomycete *Aspergillus terreus*,[1] and related fungi such as *Aspergillus lentulus*,[2] has been the focus of many studies due to its diverse biological activities. These include phytotoxic[3] and anti-inflammatory properties,[4] inhibition of tumour angiogenesis[5] and melanogenesis[6] as well as inhibition of biofilm formation in *Pseudomonas aeruginosa*.[7] Isotope labelling[8,9] and genetic experiments[3] revealed that the oxygenated cyclopent-2-en-1-one moiety, present in **1**, is derived via ring contraction of the precursor *R*-6-hydroxymellein **2** (6-HM, Scheme 1A). Although the terrein biosynthetic gene cluster (BGC) was discovered in 2014, and the biosynthesis of **2** by collaboration of the non-reducing PKS (nr-PKS) TerA and the PKS-like multidomain protein TerB has been investigated in detail,[10] the later steps remain obscure.

By gene-knockout experiments in *A. terreus* Brock et al. identified that the late-stage genes *terCDEF* (and *terR* encoding a transcriptional activator) are essential for production of terrein **1** in vivo. However, no pathway intermediates following 6-HM **2** could be isolated from individual knockout strains.[3] The group's alternative attempt to reconstitute the biosynthetic pathway by heterologous expression in *A. niger* also met with failure as native enzymes were thought to interfere with formation of **2**.

Structurally related (poly)chlorinated polyketides include the families of cyclohelminthols **3** (*Helminthosporium velutinum*),[11] palmaenones **4** (*Lachnum palmae*),[12] cryptosporiopsin **5** and cryptosporiopsinol **6** (*Cryptosporiopsis* sp., *Periconia macrospinosa* and *Phialophora asteris* f. sp. *Helianthi*)[13–15] as well as the putative Diels-Alder adduct of **5**, roussoellatide **7** (*Roussoella* sp. DLM33, Scheme 1).[16] These compounds are expected to utilize chlorinated congeners of **2**, such as the double chlorinated **8**, as biosynthetic precursors. All respective BGC share genes homologous to *terABC* and *terR*, but no homologs of *terDEF* of the terrein BGC are present in any of these systems (Scheme 1B).[17]

Remarkably, labelling studies showed that the cyclopent-2-en-1-one moieties of terrein **1** and cryptosporiopsin **5** display opposite isotopic incorporation at the α-hydroxyl carbon indicating a different course of the crucial ring contraction in each case (Scheme 1A).[16,18,19] It therefore appears that the biosynthetic pathways of **1** and its chlorinated congeners diverge at some point in the pathway, possibly after the action of TerC which is annotated as a flavin dependent monooxygenase (FMO). TerD and TerE are also likely oxidative enzymes (FMO and multi-copper oxidase, respectively), while TerF (a so-called Kelch protein) is of unknown function. The role of each

[a] L. Kahlert, D. Bernardi, M. Hauser, Dr. E. J. Skellam, Prof. R. J. Cox
     Institute for Organic Chemistry and BMWZ
     Leibniz Universität Hannover
     Schneiderberg 38, 30167, Hannover (Germany)
     E-mail: russell.cox@oci.uni-hannover.de

[b] D. Bernardi, Dr. L. P. Ióca, Prof. R. G. S. Berlinck
     Instituto de Química de São Carlos
     Universidade de São Paulo
     CP 780, CEP 13560-970, São Carlos, SP (Brazil)

[c] Dr. E. J. Skellam
     Department of Chemistry & BioDiscovery Institute
     University of North Texas
     1155 Union Circle 305220, Denton, Texas, 76203 (USA)

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing
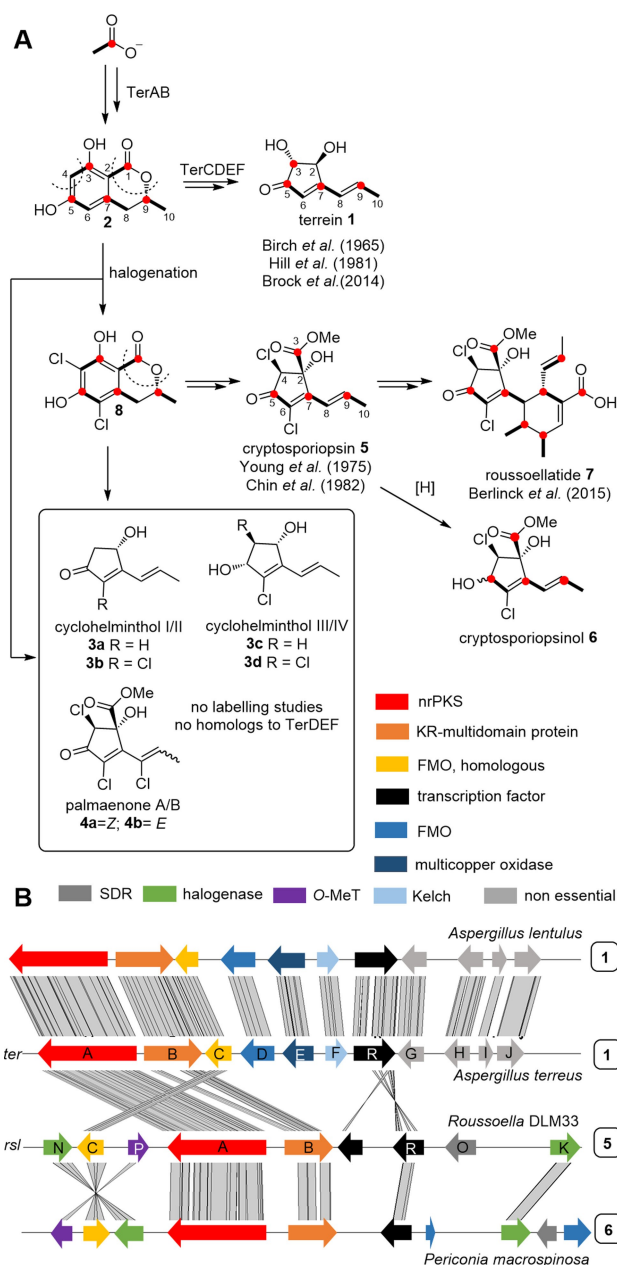
Scheme 1. **A**, Labelling patterns in terrein **1** and cryptosporiopsin **5**. Related compounds are boxed; **B**, Homology analysis[23] between the terrein **1** BGC in *A. lentulus* / *A. terreus* and the putative cryptosporiopsin **5** /-ol **6** BGC in *Roussoella* sp. DLM33 and *P. macrospinosa*, respectively. Confirmed or expected pathway end-products are boxed. For more detailed analysis of indicated genes within the terrein **1** BGC see Table S1.

pathway intermediates during the biosynthesis of these intriguing metabolites.[3]

Results

Characterization of the FMOs TerC and TerD

The BGC of terrein **1** and related natural products each harbor one homologous FMO encoding gene (*terC*). Bioinformatic analysis suggests that the sequence of TerC deposited at NCBI is truncated at the N-terminus, with the true *terC* sequence featuring an unusual intron starting after the first two 5′ codons (Figure S2). We therefore generated *A. oryzae* transformants co-expressing either the truncated (*terCᵗ*) or the extended sequence of *terC* with *terAB* following an established protocol.[24] TerA and TerB produce 6-HM **2**[25] that is the putative substrate for TerC.[3,10] Expression of the truncated *terCᵗ* with *terAB* led to no change over expression of *terAB* alone. However, transformants coexpressing the extended sequence of *terC* with *terAB* exhibited a distinct red colouration of the fermentation media that intensified with prolonged incubation (Figure 1), similar to that of △*terD*-△*terF* strains of *A. terreus* SBUG844.[3] In
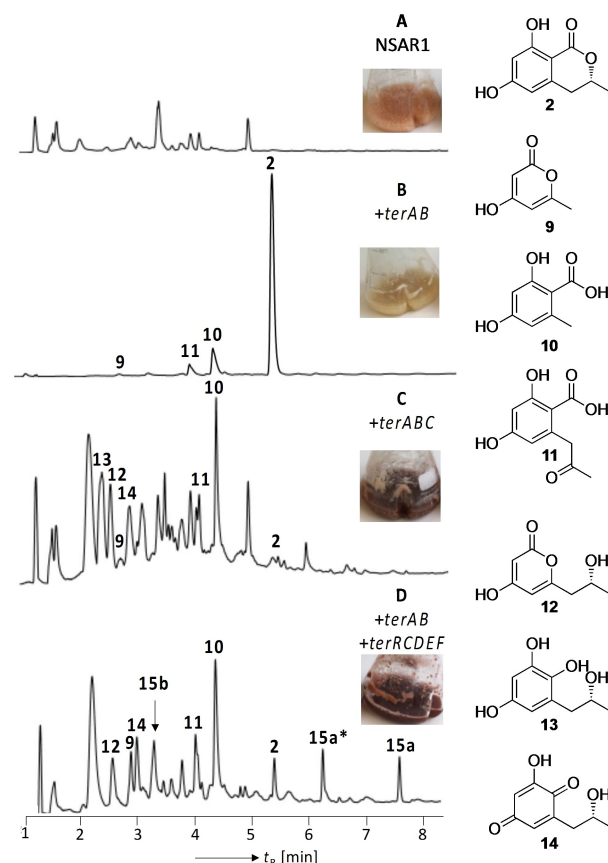
Figure 1. LCMS chromatograms (DAD 210–600 nm) of *A. oryzae* transformants expressing the indicated genes with example images of fungal liquid cultures. Fungal strains depicted in lane **A-C** were grown for 3 days; lane **D** was grown for 5 days. Note that the chromatogram in **C** is expanded 4-fold vs. **A** and unlabelled peaks in **C** are also present in the WT **A**.

enzyme beyond the formation of **2** is therefore speculative or unknown.

We,[20,21] and others,[22] have shown that reconstruction of cryptic fungal biosynthetic pathways in the host organism *Aspergillus oryzae* NSAR1 is an effective way to understand complex transformations. Here, we deploy those methods and in vitro studies of isolated enzymes aiming to overcome the hurdles that have previously hampered characterization of

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

addition to TerA-derived polyketide products **9–11** and the tetraketide pyrone **12**, LCMS analysis revealed that **2** was present in much lower quantities compared to transformants solely expressing *terAB* (Figures 1B and 1 C).

We initially identified two new peaks during early stages of cultivation of *terABC* transformants corresponding to the very polar compounds **13** ($C_9H_{12}O_4$; ($[M+H]^+$ calculated 185.0814, found 185.0802) and **14** ($C_9H_{10}O_4$). Based on their molecular weights, both **13** and **14** could be possible pathway intermediates. Compound **13** was purified to homogeneity (~ 12 mg from 1 L fungal culture) and its structure was revealed by full NMR analysis (Figures S34-S38). The *R*-configuration of **13** was assumed from the known configurations of the precursor 6-HM **2** and the off-pathway lactone **12**. Compound **14** proved even more unstable than **13** and could not be purified. Based on its mass, retention time and distinctive UV spectrum we propose **14** to be a quinone derived from **13**.

Interestingly, **13** is not red and no obvious new peaks in the respective LCMS chromatograms of *terABC*-transformants were observed concomitantly with **13**. We therefore increased the cultivation time to five days and separated the organic extract into 20 fractions by preparative HPLC. Although almost all fractions displayed a light red-brownish hue, three fractions in particular exhibited a deep red colouration. Two isomers **15 a** and **15 a*** were identified in this extract that show the same UV-absorption and molecular weight ($C_{18}H_{16}O_7$; ($[M+H]^+$ calculated 345.0974, found 345.0977). A third peak **15 b** ($C_{18}H_{18}O_8$; $[M–H]^-$ calculated 361.0923, found 361.0904) is more highly oxidised. However, despite many attempts the structure of these dimeric compounds could not be unambiguously assigned.

In order to corroborate that **2** is the precursor of **13** and **15 ab**, three individual *A. oryzae* transformants solely expressing *terC* were pulse fed with **2**. Each 50 ml fermentation was fed with 9 mg of **2** each day for 3 days. At the end of the experiment extracts were examined by LCMS. Very low levels of **13** and **15 a** were detected, but only in extracted ion chromatograms. Meanwhile, the peak for **2** completely disappeared in the supplemented *terC*-strains, but not in the supplemented control strain lacking *terC* (Figure S3). This observation emphasizes the in vivo instability of the new intermediate **13**. In addition, only the fermentation media of the *terC*-strain supplemented with **2** displayed a red colouration, but not the supplemented control strain (Figure S4).

In order to monitor the conversion of 6-HM **2** into **13** in detail, TerC was produced as soluble protein in *Escherichia coli* (Figure S5). Incubation of **2** with TerC in vitro, initially leads to production of a more polar compound **16** with 16 additional mass units ($C_{10}H_{10}O_5$; ($[M–H]^-$ calculated 209.0450, found 209.0457) that we propose to be the C-2 hydroxyl-congener of **2** (Figure 2B). Lactone **16**, in turn, readily converts to **13**. Upon prolonged incubation, the peak for **16** is completely absent and **14** and **15 b** are formed at the expense of **13** (Figure 2C). In contrast to previous in vivo experiments neither **15 a** nor **15 a*** are formed in vitro.

TerD is also an FMO, but is specific to the terrein BGC. We prepared TerD as recombinant protein in *E. coli* (Figures S6-S7). Incubation of TerD with 6-HM **2** yields a single new product **17**
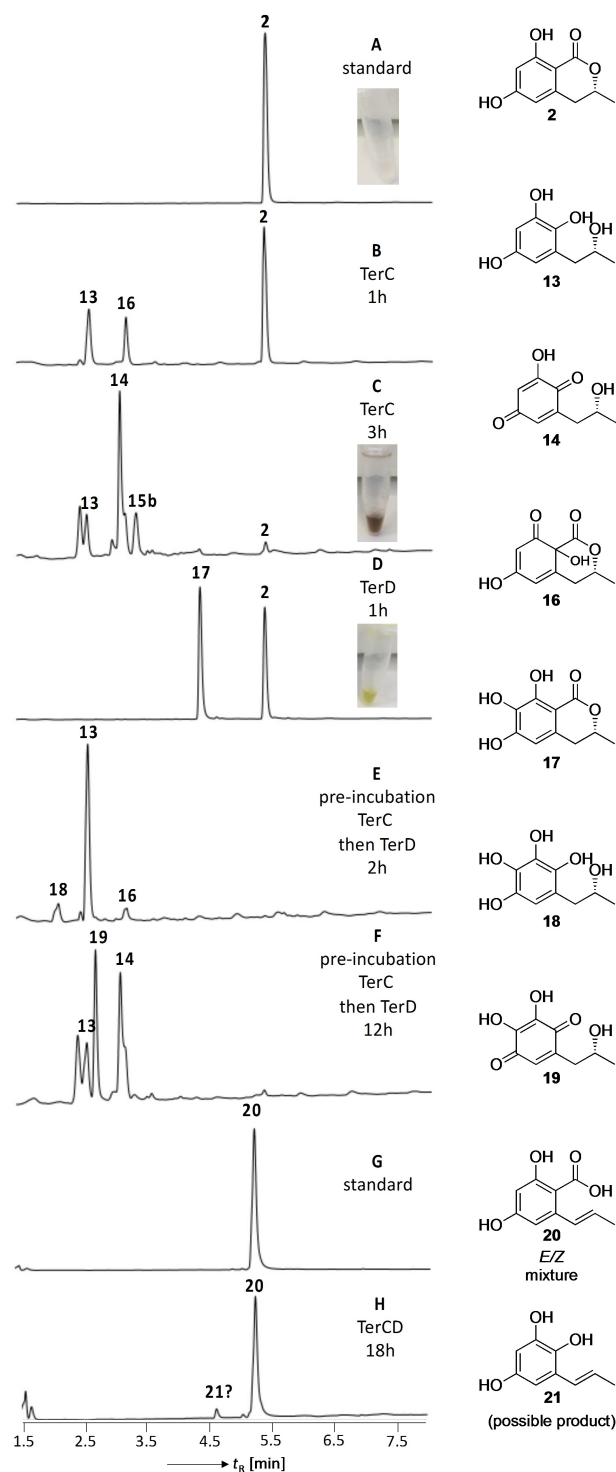
Figure 2. LCMS-chromatograms (DAD 210–600 nm) of in vitro assays with TerC/D using substrate 6-HM **2** or **20** under conditions as indicated. Colour changes of selected enzymatic conversions are given next to the chromatograms.

(Figure 2D) that was identified as the C-4-hydroxyl-congener of **2** by 2D NMR analysis (Figures S54-S58).

Co-incubation of TerC and TerD with **2**, as well as pre-incubation of 6-HM **2** with TerC followed by addition of TerD,

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

yields another very polar new minor product **18** (Figure 2E) in addition to the major product **13**. Compared to the TerC-derived tetra-ol **13**, compound **18** shows additional 16 mass units ($C_9H_{12}O_5$; ([M−H]$^-$ calculated 199.0606, found 199.0605). Based on UV, mass spectra, retention time and comparison to previously observed compounds, **18** appears to constitute the C-4 hydroxyl-congener of **13**. Compound **18** slowly oxidises to **19** (Figure 2F) similar to the oxidation of **13** to **14**. UV analysis of **19** is consistent with the quinone structure.[26]

Late steps during the biosynthesis of terrein 1

In order to reconstitute and decipher the late-stage biosynthesis of terrein **1** we individually co-expressed all of the proposed essential genes *terA-terF* in *A. oryzae*. In this experiment each structural gene was expressed from a promoter known to be functional in *A. oryzae*.[24] However, none of the transformants showed production of **1** and no other putative pathway intermediates other than **13/14** and the corresponding dimeric forms including **15a** could be identified (Figure S8). Unexpectedly, the TerD-derived compounds **17–19** were not observed in vivo, indicating an unknown incompatibility with our host organism.

Attempts to use alternative start sites for *terD-terF*, cloning individual genes from gDNA instead of cDNA, using different fermentation media and cultivation protocols, also met with failure to generate new pathway intermediates or the end product terrein **1**. Additional introduction of the four genes *terG-terJ* (previously shown to be non-essential for production of **1** in *A. terreus*),[3] also with known strong *A. oryzae* promoters, did not lead to any change in the metabolic output (Figure S8). Since **1** shows antifungal activity,[27] and its production could conceivably kill transformants successfully expressing the complete pathway, we tested its toxicity against *A. oryzae*. No inhibition of growth at a concentration up to 10 mg·mL^{-1} was observed on solid agar plates.

Earlier studies showed that the transcriptional regulator TerR induces high-level transcription of all essential terrein pathway genes *terA-terF* and is a useful tool for the construction of new fungal expression systems.[28] To exclude the possibility that cloning mistakes or uneven transcription levels between *terC-terF* caused by the use of different *Aspergillus* promotors (P_{amyB}, P_{eno}, P_{adh}, P_{gdpA}) hampered production of **1**, we constructed a *terR*-based expression vector. In contrast to the previously applied modular expression system, this construct (pTYGSade-*terRCDEF*) features *terR* cloned downstream of the starch-inducible P_{amyB}, followed by its native terminator and a contiguous *A. terreus* genomic fragment covering the promotor region of *terC* up to the terminator region of *terF* (Figure S9). This approach should ensure that *terCDEF* are expressed as closely as possible to the native situation.

In former *A. oryzae* transformants *terA-C* are equally functional when cloned from either gDNA or cDNA, showing that *A. terreus* introns are generally spliced correctly in *A. oryzae*. The vector pTYGSade-*terRCDEF* was co-transformed with pTYGSarg-*terAB* and positive transformants that integrated the desired

fragments were confirmed by PCR (Figure S10). However, the *terABRCDEF* transformants only produced the dimeric compounds related to **15** in elevated titres compared to previous transformants. Once again neither **1** nor any new pathway intermediates were observed (Figure 1D).

Unsuccessful reconstitution of the terrein pathway in vivo prompted us to revisit our in vitro strategy. However, attempts to obtain the multicopper oxidase TerE and the Kelch protein TerF as soluble proteins in *E. coli* met with failure (Figures S11-S12). We tried to simulate the general oxidative ability of TerE by addition of Cu^{2+} ions to the TerCD+**2** in vitro assay. In the absence of copper this reaction gives a faint red reaction mixture, but upon presence of copper (CuCl$_2$, 5 mM; Figure S13) it immediately adopts an intense red colouration. Yet, no new products were observed by LCMS.
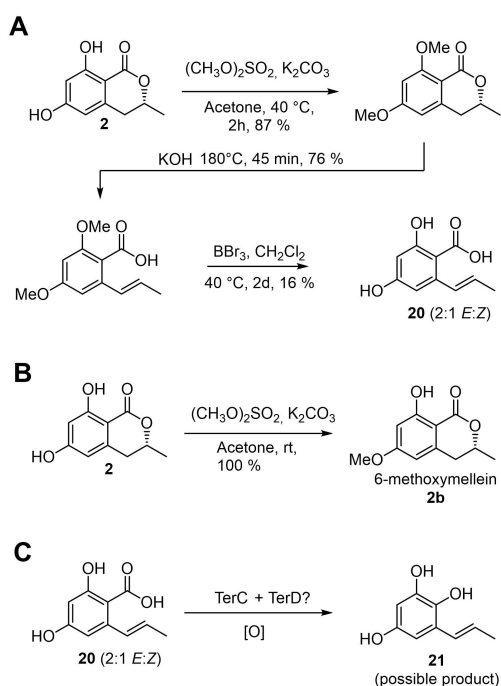
Unlike the final pathway product terrein **1**, the intermediates **13/14** and **18/19** possess a secondary hydroxyl-group at C-9 that has to be eliminated at some stage. The only candidate enzyme encoded in the terrein BGC that could conceivably catalyse this reaction is the SDR TerH. Yet, TerH has been shown non-essential for production of **1** in *A. terreus*.[3] TerH was obtained as soluble protein from *E. coli* and added to either TerC+**2** or TerCD+**2** in vitro. No new compounds were observed. Addition of the N-terminal catalytically inactive DH-domain of TerB also yielded no new intermediates (data not shown, Figures S14-S15).

We considered it possible that dehydration of any intermediate might have to occur before the oxidative steps, and that the dehydration might be catalysed by proteins encoded outside the known terrein BGC. We therefore prepared **20** in a three-step synthesis from **2** as an alternative substrate for in vitro studies (Scheme 2A). *O*-methylation of **2** was followed by base catalysed elimination, and then BBr$_3$-mediated demethylation. This afforded a ca. 2:1 mixture of inseparable *E:Z* isomers of **20**.

Neither TerC nor TerD alone show any conversion of **20**, but simultaneous incubation with both enzymes results in a new small peak in LCMS-chromatograms (Figures 2G and 2H) of assay components. HRMS confirmed a composition of $C_9H_{10}O_3$ ([M−H]$^-$ calculated 165.0552, found 165.0549) that would match the structure of tri-ol **21** that has been proposed as a potential biosynthetic intermediate.[17,19] However, very low production of **21** thwarted structural elucidation by NMR. In addition, the methoxy-derivative 6-methoxymellein **2b** (Scheme 2B) was probed as an alternative substrate, but neither TerC nor TerD exhibited any conversion (Figure S16).

Identification a homologous BGC in Roussoella sp. DLM33

Inability to progress understanding of the terrein system beyond the activity of TerC led us to consider the related biosynthetic systems which lack TerD and the later proteins. We hypothesized that such pathways may use more tractable rearrangement systems and that results from such studies might help to draw conclusions about the functions of TerE and

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

Scheme 2. Synthesis of alternative substrates. **A**, Synthesis of **20**; **B**, synthesis of 6-methoxymellein **2b**; **C**, possible product of enzymatic conversion of substrate **20**.

TerF that were previously shown to be essential for formation of **1**.[3]

FungiSMASH analysis[29] identified a candidate BGC in the genome of *Roussoella* sp. DLM33, a producer of the double chlorinated lactone **8** and the cryptosporiopsin **5** derived roussoellatide **7** (for details of genome sequencing see Supporting Information).[16] Refinement using the gene prediction tools AUGUSTUS[30] and FGENESH[31] associated nine genes with the proposed *rsl* BGC (~36 kb). The central genes encode RslA and RslB which are homologs of TerA and TerB and are expected to produce 6-HM **2**. Two additional genes encode an FMO (RslC, homologous to TerC) and a transcriptional activator (RslR, homologous to TerR) and are also conserved among related BGC (Scheme 1B). Further genes in the *rsl* BGC encode: two flavin-dependent halogenases (RslK and RslN); a short-chain dehydrogenase/ reductase (SDR, RslO); an *O*-methyltransferase (RslP); and another putative incomplete transcription factor. Seven of these genes are also conserved in the putative cryptosporiopsinol **6** BGC of *Periconia macrospinosa* (Scheme 1B).

The putative *rsl* BGC was confirmed by heterologous expression of *rslA* in *A. oryzae*. This yields the same three polyketides **9-11** as previously reported for expression of *terA*, in a combined titre of 125 mg L^{-1} (Figure 3B).[10] Addition of *rslB* results in formation of the key intermediate 6-HM **2** in high titres of approx. 500 mg L^{-1} (Figure 3B).
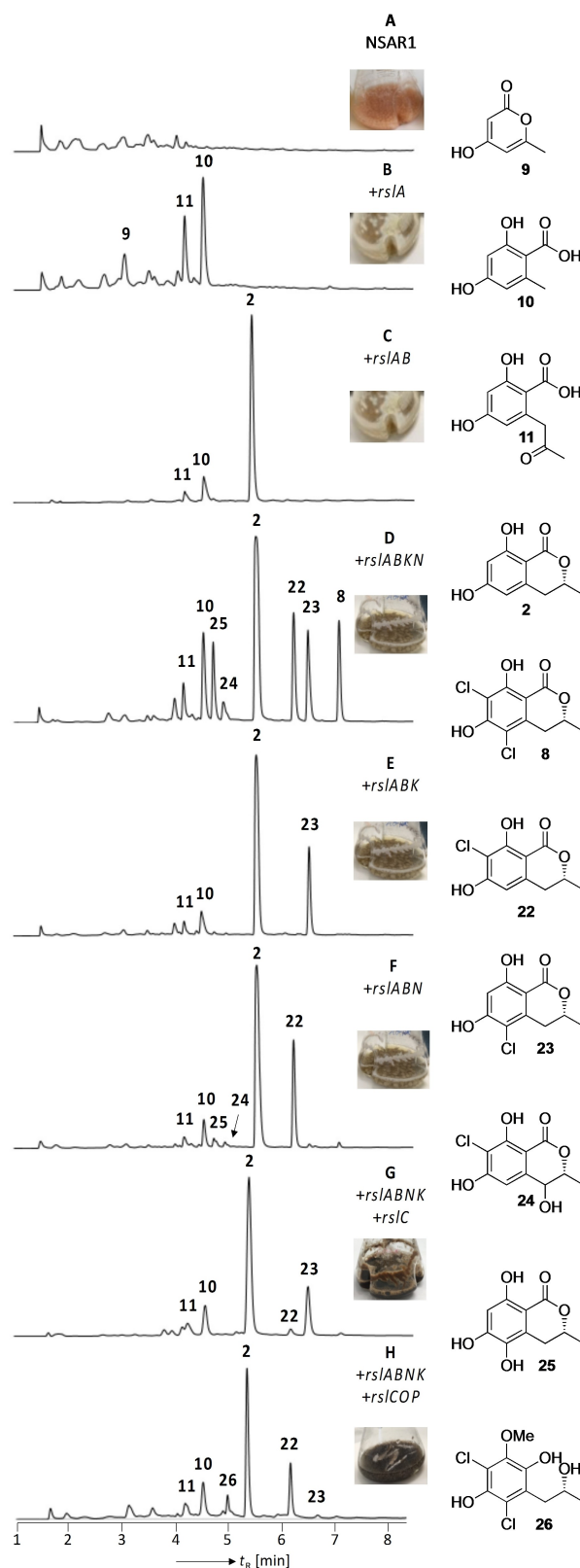

Figure 3. LCMS chromatograms (DAD 210–600 nm) of *A. oryzae* transformants expressing the indicated genes from the *rsl* BGC of *Roussoella* sp. DLM33. Example images of fungal liquid cultures are given next to the chromatograms. Fungal strains depicted in lane **A-G** were grown for 4 days. The fungal strain depicted in lane **H** was grown for 2 days to show a visible peak for **26**.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

Compatibility between TerA/RslA and TerB/RslB

TerB is a tridomain protein with an N-terminal catalytically inactive dehydratase (ψDH), a central catalytically inactive *C*-methyltransferase- (ψC-MeT) and a functional *C*-terminal keto-reductase (KR).[10] Interestingly, RslB, that is ca. 300 amino acid residues shorter than TerB, does not contain the central ψC-MeT domain (InterPro[32] analysis). Since TerA and TerB interact during the biosynthesis of 6-HM **2** in vivo we were interested whether any of these enzymes could be exchanged with its homologous protein RslA or RslB, respectively.[10] We therefore generated two transformant strains of *A. oryzae* expressing either *terA + rslB* or *rslA + terB*. Both transformant strains produce **2** very efficiently (ca. 400 mg L^{-1} and 450 mg L^{-1} vs. 500 mg L^{-1} for *terA + terB*; Figure S17).

Characterization of the flavin-dependent halogenases RslK and RslN

Addition of the two annotated halogenase genes (*rslK* and *rslN*) to the *A. oryzae* transformants leads to production of the di-chlorinated **8** in vivo, that was also isolated from cultures of *Roussoella* sp. DLM33.[16] The mono-chlorinated congeners **22** and **23** are also produced (Figure 3D; NMR in Supporting Information). Individual expression revealed regioselectivity for RslK at the C-6 and for RslN at the C-4 position, respectively (Figures 3E and 3F). These results are in agreement with recent in vitro characterization of the homologous halogenases ChmKN (cyclohelminthols) and PloKN (palmaeonones).[17]

Both **22** and **23**, but not **8**, are slowly converted into their C-8 hydroxyl-derivatives including **24** by endogenous enzymes in *A. oryzae*, supported by feeding studies (Figure S18, S81-S85). 6-HM **2** in turn is hydroxylated at C-6 to give **25** upon prolonged incubation which is also consistent with previous observations.[10,17] Supplementation of the fermentation medium with sodium bromide (50 μg·mL^{-1}) results in the successful formation of mono- and dibrominated derivatives of **2** as well as mixed chloro-bromo species based on distinct isotope ratios in the respective mass spectra (Figure S19). However, iodinated derivatives were not formed in the presence of NaI.

Characterization of the flavin-dependent monooxygenase RslC and the O-methyltransferase RslP

The TerC/RslC-type FMO are encoded by all related BGC (Scheme 1B). In fact, *rslC* appears to exhibit the same unusual intron pattern located directly at the 5-end of the corresponding mRNA analogous to *terC* (Figure S2).

Similar to previous in vivo expression of *terABC*, *rslABCKN*-transformants display a distinct reddish-brown colouration of the fermentation medium. However, no new compounds could be detected by LCMS analysis (Figure 3G). Shortened or prolonged cultivation also revealed no new products by LCMS.

All attempts to obtain RslC as soluble and active protein in *E. coli* met with failure. Alternatively, TerC was incubated with **8** in vitro and although no turnover was visible by LCMS analysis the samples displayed a reddish-brown colouration as observed in vivo (data not shown), indicative of oxidation.

Since potential RslC-derived products may also be chemically unstable or prone to degradation in *A. oryzae* (e.g. as observed for **13**), we proceeded to add the two ancillary genes *rslP* (*O*-methyltransferase) and *rslO* (SDR) to generate "full cluster" transformants. A new pale brownish di-chlorinated compound **26** was found in the fermentation media of respective transformants exclusively on the third day of cultivation (before the culture medium adopted a dark colouration, Figure 3H).

Compound **26** is two mass units heavier than the expected pathway end product cryptosporiopsin **5** ($C_{10}H_{12}O_4Cl_2$; ([M−H]$^-$ calculated 265.0034, found 265.0037). 2D NMR analysis revealed that **26** is a di-chlorinated derivative of the aromatic TerC product **13** with an additional *O*-methylation at the C-3 hydroxyl group (isolated 6 mg L^{-1}; Figures S87-S91). To exclude spontaneous methylation all analytical steps were repeated in the absence of MeOH whereby **26** was still produced. Interestingly, **26** has been previously isolated from *Periconia macrospinosa*, a producer of cryptosporiosinol **6**.[33] This suggests that *O*-methylation is introduced by RslP. No obvious function could be attributed to the SDR RslO.

In analogy to the TerC-derived **13**, the RslC-derived **26** features a secondary hydroxyl-group at C-9 that is not present in the proposed pathway end-product cryptosporiopsin **5**. Since no gene encoding an enzyme that would catalyze dehydration at this position is conserved among homologous BGC (Scheme 1B) this step remains obscure. Despite heterologous expression of seven genes (*rslABCKNOP*) in vivo, no compounds were identified that feature a re-arranged cyclopentenone/-ol skeleton.

Discussion

6-HM **2** is an intermediate during the biosynthesis of terrein **1** and related (poly-chlorinated) natural products such as **5**.[3,17] In the terrein pathway lactone **2** is formed by collaboration between an nrPKS and a multidomain protein with catalytically active KR-domain in vivo.[10] Bioinformatic analysis allowed the identification of a related BGC in the marine-derived *Roussoella* sp. DLM33 that shares the two corresponding genes *rslAB*, amongst others. Heterologous expression of either *rslA* alone or in combination with *rslB* revealed the same qualitative and quantitative product manifold reported for the homologous proteins of the terrein **1** BGC.[10] In contrast to TerB, RslB does not feature a central ψC-MeT domain. Despite this altered domain architecture combining homologous genes from two different species (*terA + rslB* and *rslA + terB*) yielded the product 6-HM **2** in titres nearly as high as obtained for the native system (Figure S17). These results show that collaboration is also facilitated between homologous proteins originating from different fungal species. Furthermore, since RslB physically lacks ψC-MeT it appears that this domain is neither involved in

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

protein–protein interactions, or fulfils a catalytic role, in line with previous findings.[10]

It is likely that collaboration during the biosynthesis of 6-HM **2** is a general feature of such homologous systems (e.g. also ChmAB from *H. velutinum*),[17] but further in-depth investigations are required to characterize the interacting elements in greater detail.

The *rsl* BGC encodes two flavin-dependent halogenases RslK and RslN that halogenate 6-HM **2** at the C-4 (RslN) and C-6 (RslK) positions (Scheme 3). The observed regioselectivity corresponds with previous reports on homologous enzymes.[17] The formation of brominated congeners in the presence of NaBr reveals that both halogenases show halide promiscuity. This is a potentially useful approach to generate novel brominated species.

Of primary interest was the elucidation of the function of the flavin-dependent monooxygenases that are common to all the related pathways (e.g. TerC and RslC, Scheme 1B) and that are likely to catalyse the final common step before divergence of the terrein and other pathways. Characterization of TerC shows that this enzyme initially hydroxylates 6-HM **2** adjacent to the carbonyl-position at C-2 to yield intermediate **16**. TerC thus matches the function expected for a classical flavin-dependent monooxygenase.[34,35] Intermediate **16** decarboxylates spontaneously to the tetra-ol **13** in vitro (Scheme 3A). An analogous mechanism was proposed by Townsend et al. for the FMO-dependent oxidative lactone opening during the biosynthesis of cercosporin and related perylenequinones.[36,37] Intermediate **13** is unstable both in vivo and in vitro and it is converted into the corresponding quinone **14** by autoxidation[38] that readily dimerises to **15b** that exhibits a

strong red colouration (Figure S13). Tetra-ol **13** appears to be a substrate for TerD which hydroxylates at C-4. But the product, **18**, does not proceed to ring-contracted species, giving only dimers by autoxidation (Scheme 3A). TerD appears inactive in *A. oryzae* expression experiments for unknown reasons. However, TerD is a potential tool for the in vitro stereoselective hydroxylation of isocoumarin/ phenol-derived compounds. Further investigations and screening of a substrate library will be necessary to evaluate its substrate promiscuity.

A significant difficulty in this study has been the instability of compounds beyond **13** and our inability to satisfactorily identify, isolate or characterise them. Dimers such as **15b** were identified by HRMS in vitro and in vivo, but do not appear to be on the pathway to **1**. Formation of **15a** exclusively in vivo suggests that a native enzyme in *A. oryzae* is involved in its formation, adding another layer of complexity to efforts to detect and identify pathway intermediates. Oxidative coupling of phenol-related compounds in fungi is often mediated by laccases or cytochrome P450-enzymes.[39,40] A recently described laccase catalyses the coupling of two subunits structurally very similar to **14**, yielding the bibenzoquinone oosporein that also exhibits a strong red colour.[41]

As anticipated, the homologous FMO RslC catalyses the same fundamental reaction, i.e. oxidative lactone opening of the di-chlorinated **8**, that we could only observe in vivo. However, the direct product **27** defied detection. Similar to previous transformants expressing *terABC*, *A. oryzae rslABCKN* transformants adopted a distinct dark colouration (cf. Figures 3G and 3H) indicative of further oxidation. Alternatively, production of **27** might have induced production of fungal
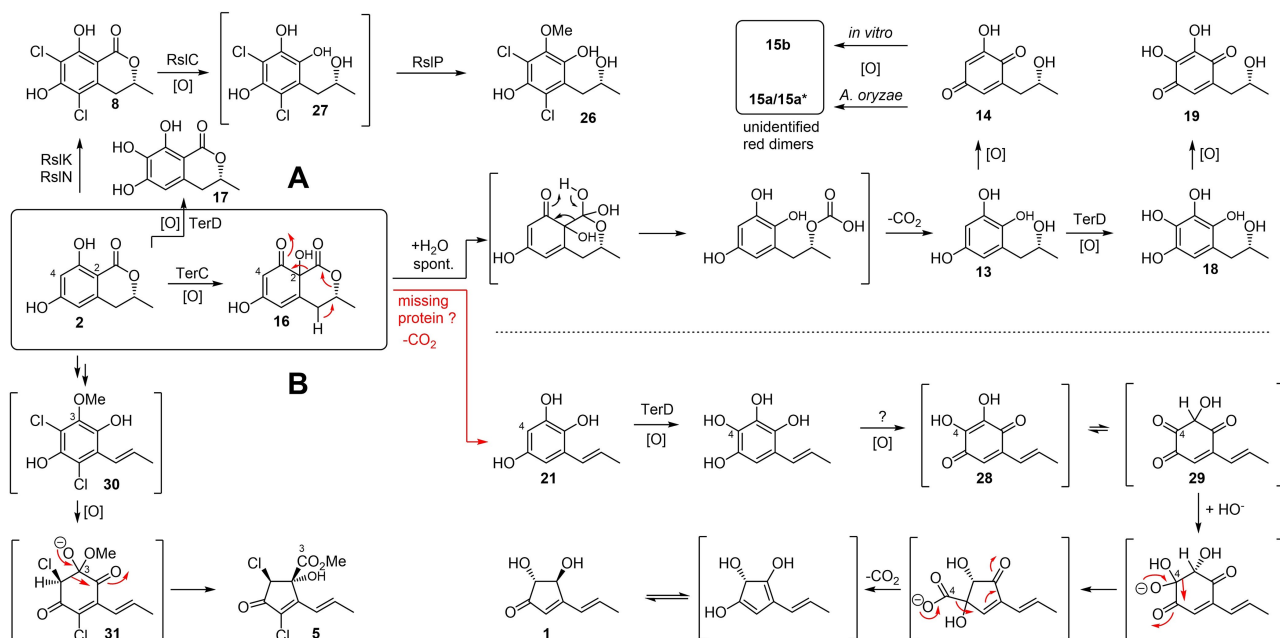
Scheme 3. Summary of observed and hypothetical reactions on the pathway to terrein **1** and cryptosporiopsinol **5**. **A**, Reactions and intermediates observed in vivo and in vitro. All compounds feature the C-9 hydroxyl which is not observed in **1** and **5**; **B**, hypothetical pathways to **1** and **5** involving early dehydration and late benzilic acid rearrangements. See Supporting Information for DFT calculations of the respective rearrangements in each case.

Chemistry–A European Journal

Full Paper
doi.org/10.1002/chem.202101447

Chemistry Europe
European Chemical
Societies Publishing

melanins that function as a defence mechanism against chemical environmental stressors.[42]

The *O*-methylated congener **26** was identified only in the presence of the *O*-methyltransferase RslP. Based on labelling studies *O*-methylation within **26** and cryptosporiopsin **5** shares the same position at C-3 (Scheme 1A). It remains elusive whether **26** constitutes a true pathway intermediate or an off-pathway shunt. Despite elevated titres of 6-HM **2** in the respective transformants (Figure 3G), the non-chlorinated **13** was not observed. This indicates that RslC requires chlorination at C-4 and C-6 for substrate recognition, and rules out the activity of RslC before chlorination.

Characterization of both FMOs TerC and RslC provides experimental evidence that the biosynthetic pathways towards terrein **1** and related (poly-chlorinated) congeners branch after oxidative lactone cleavage. The subsequent ring contractions follow a different pathway in each case as indicated by distinct labelling patterns (Scheme 1A). However the pathway beyond this decarboxylation remains unclear in both cases.

A key difficulty revealed by our results is the timing of the dehydration required to provide the *E*-propenyl side chain of terrein **1**, cryptosporiopsin **5** and the majority of other compounds in this class. The fact that this moiety is widely present suggests that the dehydration should occur early in the pathway. Since **13** is also produced by TerC in vitro, we can exclude that the secondary hydroxyl-group at C-9 is (re) introduced by native enzymes of our in vivo host *A. oryzae*.[43–46] Likewise **18**, that is derived from sequential catalysis of TerC and TerD in vitro, does not feature the *E*-propenyl side chain.

Furthermore, previous studies[3] imply that the TerD-derived lactone **17** is not an intermediate but a shunt metabolite: Feeding of **17** to a △*terB* strain of *A. terreus* did not restore production of terrein **1**. Similarly, the △*terD* strain of *A. terreus* did not show production of **1** in the presence of **17**.

However, neither TerAB nor TerC (or the respective homologs) appear capable of catalysing the required dehydration. An attractive possibility would be coupling the oxidative decarboxylation catalysed by TerC/RslC with the dehydration (Scheme 3B). The fact that we did not observe this in vitro or in vivo may indicate that TerC/RslC requires another protein to achieve such a transformation. However, since there are no more homologous genes in the two BGC a candidate is hard to identify. In addition, full expression of both clusters in *A. oryzae* could not progress either pathway further, suggesting that if such a protein exists, it is not encoded within the *ter* or *rsl* BGC. In other systems, such as those responsible for azaphilone biosynthesis in *Monascus* species, a specific *O*-acetylation and elimination sequence (MrPigM and MrPigO) provides the prop-2-enyl moiety.[47] However homologs of these proteins are not encoded within the terrein or related BGC.

For this reason, it appears unlikely that C-9-hydroxylated compounds such as **18** and **19** are true intermediates. Although we did not observe compounds such as **28** (the C-8/C-9 dehydrated homolog of **19**), this compound could be envisaged as a potential direct precursor of **1** via tautomerisation to **29** and then an α-ketol, or related benzilic acid rearrangement (Scheme 3B).[48] Such a mechanism is supported as feasible by

DFT calculations (see Supporting Information for details). Ring contractions in natural product biosynthesis have been previously proposed to proceed following such rearrangements. Examples include the biosynthetic pathways of xenovulene A,[49] preuisolactone A[50] (both C_7 to C_6) and fredericamycin A (C_6 to C_5).[51] In synthetic chemistry such reactions are often performed in the presence of metal-ion catalysts, pointing out a potential function of the predicted copper-dependent TerE.[52,53] Decarboxylation after rearrangement would give terrein **1** after tautomerisation.

A similar rearrangement mechanism could be envisaged during the biosynthesis of cryptosporiopsin **5** (Scheme 3B). Chlorination of C-4, however, in this case, prevents an identical mechanism. Here, 3-methoxy hydroquinone **30** could oxidise to benzoquinone **31**. Rearrangement of **31** by a benzilic acid reaction would also lead to ring-contraction, but via extrusion of C-3. This is consistent with the observed labelling pattern. Early methylation of O-3 would create a methyl ester after rearrangement and thereby prevent the decarboxylation which must occur during the rearrangement on the route to **1**. Again, DFT calculations (Supporting Information section 2.14) suggest this type of mechanism would be feasible.

Conclusion

The biosynthesis of terrein **1** and related compounds such as **5** and **6** has remained mysterious for more than 85 years. Despite the recent discovery of the BGC and molecular studies involving gene knockouts that confirmed the intermediacy of 6-hydroxy-mellein **2**, no more details on their biosynthesis could be found. Here we have advanced knowledge in this area by another small step. The FMO TerC (and its homologs) are involved in the oxidative decarboxylation which removes C-1 early in the biosynthesis, and this seems to be the last step in common between the terrein and (poly)chlorinated pentenone pathways. However we have not identified the origin of the required dehydration, but our studies suggest this must also occur early in the pathways. The highly oxygenated intermediates produced by TerC and TerD are unstable and difficult to characterise, but if **1** or **5** had been formed they would have been detected. The lack of these compounds suggests the likely importance of early dehydration. It is noteworthy that while our approach to sequentially reconstitute fungal secondary metabolite biosynthetic pathways by heterologous expression in *A. oryzae* NSAR1 is usually very expedient and reliable, it was not able to decipher the final transformation during the biosynthesis of **1** and **5**.[20,43,49,54,55] Inexplicably TerD seems inactive in heterologous expression experiments in *A. oryzae*, although it is active in vitro.

We propose that polyhydroxylated intermediates probably form quinones and then benzilic acid type rearrangements could connect the biosynthesis of terrein **1** and the related compounds such as **5**. However, we have so-far failed to discover the catalysts involved in the key dehydration and rearrangement steps and it is clear that significantly more work will have to be done to finally elucidate these steps.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] H. Raistrick, G. Smith, *Biochem. J.* **1935**, *29*, 606–11.
[2] T. O. Larsen, J. Smedsgaard, K. F. Nielsen, M. A. E. Hansen, R. A. Samson, J. C. Frisvad, *Med. Mycol.* **2007**, *45*, 225–232.
[3] C. Zaehle, M. Gressler, E. Shelest, E. Geib, C. Hertweck, M. Brock, *Chem. Biol.* **2014**, *21*, 719–731.
[4] J. C. Lee, M. K. Yu, R. Lee, Y. H. Lee, J. G. Jeon, M. H. Lee, E. C. Jhee, I. D. Yoo, H. K. Yi, *J. Endod.* **2008**, *34*, 433–437.
[5] M. Arakawa, T. Someno, M. Kawada, D. Ikeda, *J. Antibiot. (Tokyo).* **2008**, *61*, 442–448.
[6] S. H. Park, D. S. Kim, W. G. Kim, I. J. Ryoo, D. H. Lee, C. H. Huh, S. W. Youn, I. D. Yoo, K. C. Park, *Cell. Mol. Life Sci.* **2004**, *61*, 2878–2885.
[7] B. Kim, J. S. Park, H. Y. Choi, S. S. Yoon, W. G. Kim, *Sci. Rep.* **2018**, *8*, 8617.
[8] A. J. Birch, A. Cassera, A. R. Jones, *Chem. Commun.* **1965**, *9*, 167–168.
[9] R. A. Hill, R. H. Carter, J. Staunton, *J. Chem. Soc. Perkin Trans. 1* **1981**, 2570–2576.
[10] L. Kahlert, M. Villanueva, R. J. Cox, E. J. Skellam, *Angew. Chem. Int. Ed.* **2021**, *60*, 11423–11429.
[11] Y. Honmura, S. Uesugi, H. Maeda, K. Tanaka, T. Nehira, K. ichi Kimura, M. Okazaki, M. Hashimoto, *Tetrahedron* **2016**, *72*, 1400–1405.
[12] T. Matsumoto, T. Hosoya, H. Tomoda, M. Shiro, H. Shigemori, *Chem. Pharm. Bull.* **2011**, *59*, 1559–1561.
[13] G. M. Strunz, A. S. Court, J. Komlossy, M. A. Stillwell, *Can. J. Chem.* **1969**, *47*, 2087–2094.
[14] D. Giles, W. B. Turner, *J. Chem. Soc. C Org.* **1969**, *0*, 2187–2189.
[15] R. J. J. C. Lousberg, Y. Tirilly, M. Moreau, *Experientia* **1976**, *32*, 331–332.
[16] E. L. F. Ferreira, D. E. Williams, L. P. lóca, R. P. Morais-Urano, M. F. C. Santos, B. O. Patrick, L. M. Elias, S. P. Lira, A. G. Ferreira, M. R. Z. Passarini, L. D. Sette, R. J. Andersen, R. G. S. Berlinck, *Org. Lett.* **2015**, *17*, 5152–5155.
[17] T. Ugai, A. Minami, S. Tanaka, T. Ozaki, C. Liu, H. Shigemori, M. Hashimoto, H. Oikawa, *ChemBioChem* **2020**, *21*, 360–367.
[18] J. S. E. Holker, K. Young, *J. Chem. Soc. Chem. Commun.* **1975**, *0*, 525–526.
[19] L. O. Zamir, C. C. Chin, *Bioorg. Chem.* **1982**, *11*, 338–349.
[20] D.-S. Tian, E. Kuhnert, J. Ouazzani, D. Wibberg, J. Kalinowski, R. J. Cox, *Chem. Sci.* **2020**, *11*, 12477–12484.
[21] C. Schotte, L. Li, D. Wibberg, J. Kalinowski, R. J. Cox, *Angew. Chem. Int. Ed.* **2020**, *59*, 23870–23878; *Angew. Chemie.* **2020**, *132*, 24079–24087.
[22] L. Kahlert, C. Schotte, R. J. Cox, *Synthesis* **2021**, https://doi.org/10.1055/a-1401–2716.
[23] T. Carver, S. R. Harris, M. Berriman, J. Parkhill, J. A. McQuillan, *Bioinformatics* **2012**, *28*, 464–469.
[24] K. A. K. Pahirulzaman, K. Williams, C. M. Lazarus, in *Methods Enzymol.*, Academic Press Inc., **2012**, pp. 241–260.
[25] M. C. Tang, Y. Zou, K. Watanabe, C. T. Walsh, Y. Tang, *Chem. Rev.* **2017**, *117*, 5226–5333.
[26] T. Wilke, M. Schneider, K. Kleinermanns, *Open J. Phys. Chem.* **2013**, *03*, 97–102.
[27] J. Goutam, G. Sharma, V. K. Tiwari, A. Mishra, R. N. Kharwar, V. Ramaraj, B. Koch, *Front. Microbiol.* **2017**, *8*, 1334.
[28] M. Gressler, P. Hortschansky, E. Geib, M. Brock, *Front. Microbiol.* **2015**, *6*, 184.
[29] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81–W87.
[30] M. Stanke, B. Morgenstern, *Nucleic Acids Res.* **2005**, *33*, W465–W467.
[31] V. Solovyev, P. Kosarev, I. Seledsov, D. Vorobyev, *Genome Biol.* **2006**, *7*, S10.
[32] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, *Bioinformatics* **2014**, *30*, 1236–1240.
[33] J. Liu, M. Chen, R. Chen, K. Xie, D. Chen, S. Si, J. Dai, *J. Chinese Pharm. Sci.* **2020**, *29*, 244–251.
[34] E. Romero, J. R. Gómez Castellanos, G. Gadda, M. W. Fraaije, A. Mattevi, *Chem. Rev.* **2018**, *118*, 1742–1769.
[35] A. H. Westphal, D. Tischler, W. J. H. van Berkel, *Arch. Biochem. Biophys.* **2021**, *702*, 108820.
[36] A. G. Newman, C. A. Townsend, *J. Am. Chem. Soc.* **2016**, *138*, 4219–4228.
[37] J. Hu, F. Sarrami, H. Li, G. Zhang, K. A. Stubbs, E. Lacey, S. G. Stewart, A. Karton, A. M. Piggott, Y. H. Chooi, *Chem. Sci.* **2019**, *10*, 1457–1465.
[38] T. H. James, J. M. Snell, A. Weissberger, *J. Am. Chem. Soc.* **1938**, *60*, 2084–2093.
[39] C. Gil Girol, K. M. Fisch, T. Heinekamp, S. Günther, W. Hüttel, J. Piel, A. A. Brakhage, M. Müller, *Angew. Chem. Int. Ed.* **2012**, *51*, 9788–9791; *Angew. Chem.* **2012**, *124*, 9926–9929.
[40] L. S. Mazzaferro, W. Hüttel, A. Fries, M. Müller, *J. Am. Chem. Soc.* **2015**, *137*, 12289–12295.
[41] P. Feng, Y. Shang, K. Cen, C. Wang, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11365–11370.
[42] R. J. B. Cordero, A. Casadevall, *Fungal Biol. Rev.* **2017**, *31*, 99–112.
[43] Y. He, R. J. Cox, *Chem. Sci.* **2016**, *7*, 2119–2127.
[44] R. Fujii, T. Ugai, H. Ichinose, M. Hatakeyama, T. Kosaki, K. Gomi, I. Fujii, A. Minami, H. Oikawa, *Biosci. Biotechnol. Biochem.* **2016**, *80*, 426–431.
[45] J. Feng, F. Surup, M. Hauser, A. Miller, J. P. Wennrich, M. Stadler, R. J. Cox, E. Kuhnert, *Chem. Commun. (Camb.)* **2020**, *56*, 12419–12422.
[46] Z. Song, W. Bakeer, J. W. Marshall, A. A. Yakasai, R. M. Khalid, J. Collemare, E. Skellam, D. Tharreau, M. H. Lebrun, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Chem. Sci.* **2015**, *6*, 4837–4845.
[47] W. Chen, R. Chen, Q. Liu, Y. He, K. He, X. Ding, L. Kang, X. Guo, N. Xie, Y. Zhou, Y. Lu, R. J. Cox, I. Molnár, M. Li, Y. Shao, F. Chen, *Chem. Sci.* **2017**, *8*, 4917–4925.
[48] A. Burke, C. Marques, *Mini-Rev. Org. Chem.* **2007**, *4*, 310–316.
[49] R. Schor, C. Schotte, D. Wibberg, J. Kalinowski, R. J. Cox, *Nat. Commun.* **2018**, *9*, 1963.
[50] A. J. E. Novak, C. E. Grigglestone, D. Trauner, *J. Am. Chem. Soc.* **2019**, *141*, 15515–15518.
[51] Y. Chen, Y. Luo, J. Ju, E. Wendt-Pienkowski, S. R. Rajski, B. Shen, *J. Nat. Prod.* **2008**, *71*, 431–437.
[52] H. Brunner, H. B. Kagan, G. Kreutzer, *Tetrahedron: Asymmetry* **2003**, *14*, 2177–2187.
[53] A. J. Burke, A. Moutayakine, *Chem. Commun.* **2021**, *57*, 2585–2590.
[54] R. Nofiani, K. de Mattos-Shipley, K. E. Lebe, L. C. Han, Z. Iqbal, A. M. Bailey, C. L. Willis, T. J. Simpson, R. J. Cox, *Nat. Commun.* **2018**, *9*, 3940.
[55] L. Kahlert, E. F. Bassiony, R. J. Cox, E. J. Skellam, *Angew. Chem. Int. Ed.* **2020**, *59*, 5816–5822; *Angew. Chem.* **2020**, *132*, 5865–5871.

Appendix

Additional Bioinformatics

Table S1: Annotated functions for the gene products of *terA-terJ* and the 5 adjacent gene products up- and downstream of the terrein BGC in *A. terreus* as deposited at NCBI. No further genes of interest were identified using AUGUSTUS[38] and FGENESH[39].

Protein	Predicted function
ATEG_00130	DUF3807 superfamily; family of conserved fungal proteins of unknown function
ATEG_00131	HVSL superfamily; top hits: U6 snRNA phosphodiesterase
ATEG_00132	Pal1 cell morphology protein, conserved among *Aspergillus* sp.
ATEG_00133	transcription factor, conserved among *Aspergillus* sp.
ATEG_00134	no predicted protein; not found in other terrein producers
ATEG_00135, TerJ	major facility transporter superfamily (MFS)
ATEG_00136, TerI	vicinal oxygen chelate superfamily (VOC); bleomycin resistance protein superfamily, glyoxalase, extradiol dioxygenase
ATEG_00137, TerH	NAD(P) binding domain; epimerase, dehydratase, aldehyde reductase
ATEG_00138, TerG	major facility transporter superfamily (MFS)
ATEG_00139, TerR	Zn_2Cys_6 transcriptional regulator
ATEG_00140, TerF	protein with kelch motif
ATEG_00141, TerE	multicopper oxidase
ATEG_00142, TerD	flavin-dependent monooxygenase
ATEG_00143, TerC	flavin-dependent monooxygenase
ATEG_00144, TerB	DH-ψ*C*MeT-KR, multidomain protein with functional ketoreductase function
ATEG_00145, TerA	non-reducing polyketide synthase
ATEG_00146	rieske non-heme-iron oxygenase; conserved among many fungal species
ATEG_00147	sugar transporter
ATEG_00148	endoplasmic reticulum oxidoreductin
ATEG_00149	Peroxin 13, Src homology 3 domain of fungal peroxisomal membrane protein; protein import pathways into the peroxisomal matrix
ATEG_00150	Clr5 domain ;This domain is found at the N-terminus of the Clr5 protein which has been shown to be involved in silencing in fission yeast

Table S2: Analysis of proteins encoded within the sorbicillinoid BGC of *T. reesei* QM6a based on antiSMASH[36,37] analysis and gene refinement using FGENESH.[39] Functions were predicted using CD-analysis in NCBI. If applicable accession numbers at NCBI are given.

Protein/accession number NCBI	Predicted function
Sor "P450", XP_006961557	P450/ trans-cinnamate monooxygenase
XP_006961558	LysM superfamily/ peptidoglycan binding/ cell wall degradation
/	No conserved domains
XP_006961155	Chitinase/ glycosyl hydrolase
XP_006961559	protein kinase/ ankyrin repeats mediate protein-protein interactions
Sor "SDR", G0R6S7	SDR/ NAD dependent dehydrogenase
SorA, G0R6S8	hrPKS
SorB, G0R6S9	nrPKS
SorC, G0R6T0	FMO
XP_006961063	major facility transporter superfamily (MFS)
QEV86399	Zn_2Cys_6 transcriptional regulator
SorD, G0R6T3	FAD-binding domain/ oxidoreductase
XP_006961563	Zn_2Cys_6 transcriptional regulator
/	No conserved domains
/	DoxX superfamily/ SURF4/ membrane associated

Curriculum Vitae

January 2018 – June 2021	PhD student in organic chemistry, Gottfried Wilhelm Leibniz Universität Hannover
October 2015 – October 2017	Master of Science, Life Science, Gottfried Wilhelm Leibniz Universität Hannover
October 2012 – October 2015	Bachelor of Science, Life Science, Gottfried Wilhelm Leibniz Universität Hannover
June 2012	Abitur, Helene Lange Schule, Hannover