Title
Mechanism Studies on Fungal Type I Iterative Polyketide Synthases and Nonribosomal Peptide Synthetases

Author
Qiao, Kangjian

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Mechanism Studies on Fungal Type I Iterative Polyketide Synthases and Nonribosomal Peptide Synthetases

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Chemical Engineering

by

Kangjian Qiao

2012

ABSTRACT OF THE DISSERTATION

Mechanism Studies on Fungal Type I Iterative Polyketide Synthases and Nonribosomal Peptide Synthetases

By

Kangjian Qiao

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2012

Professor Yi Tang, Chair

Filamentous fungi have long been recognized as prolific producers of natural products due in a considerable part to their strong track record in producing blockbuster drugs such as penicillin and lovastatin. The biosynthetic enzymes from fungi that assemble these molecules, such as polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are large highly complex multifunctional megasynthetases. Different from their well-studied bacterial counterparts, to date, the biosynthetic programming rules utilized by the fungal PKSs and NRPSs remain largely unknown. Despite the fact that a number of fungal PKSs and NPRSs genes have been discovered over the last three decades, few of them have been characterized and thus directly linked to the biosynthesis of specific fungal secondary metabolites. To fully understand the biosynthetic logic of PKSs and NRPSs and decipher the relationship between sequences and

the structures of fungal polyketides and nonribosomal peptides, the thesis focus on three independent fungal biosynthetic systems involving PKSs, NRPS and PKS-NRPS.

In the first project, the biochemistry of tandem fungal PKSs for the formation of nanomolar HSP 90 inhibitor radicicol are reconstituted *in vivo* and *in vitro* and extensively investigated with the help of tool compounds acyl thioesters. Secondly, a cryptic, NRPS-like enzyme (NRPS325) mined from *Aspergillus terreus* was reconstituted in vitro and was shown to synthesize thiopyrroles and thiopyrazines via unprecedented mechanisms. The remarkable substrate promiscuity of NRPS325 towards different amino acids, free thiols and β-ketoacyl substrates were explored to produce hundreds of new compounds. Lastly, the genome sequence of *Aspergillus clavatus* was analyzed and the 30 kb cytochalasin gene cluster was identified based on the presence of the PKS-NRPS and a putative Baeyer-Villiger monooxygenase. Deletion of the central PKS-NRPS gene, *ccsA*, abolished the production of cytochalasin E and K, demonstrating the direct association between the natural products and the gene cluster. Overexpression of the pathway specific regulator ccsR greatly elevated the titer of cytochalasins.

The dissertation of Kangjian Qiao is approved.

Neil Garg

James C. Liao

Harold Monbouquette

Yi Tang, Committee Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

1. Section 1.2. contains the material from Xie, X., **Qiao, K.**, Tang, Y., "Metabolic Engineering of Natural Product Biosynthesis." In: Tringali, C., ed. *BIOACTIVE COMPOUNDS FROM NATURAL SOURCES. Second Edition: Natural Products as Lead Compounds in Drug Discovery*. CRC Press. July, 2011.

2. Section 2.1. contains the material from Zhou, H., **Qiao, K**., Gao, Z., Vederas, J. C., Tang, Y., "Insights into Radicicol Biosynthesis via Heterologous Synthesis of Advanced Intermediates and Analogs." *J. Biol. Chem*. 2010, 285, 41412-4121.

3. Section 2.2. contains the material from **Qiao, K.**, Zhou, H., Xu, W., Zhang, W., Neil, G., Tang, Y., "Identification of a Fungal Nonribosomal Peptide Synthetase Module that can Synthesize Thiopyrazines" *Org. Lett.* 2011, 13, 1758-1761.

4. Section 2.3. is a version of **Qiao, K.**, Chooi, Y., Tang, Y., "Identification and Engineering of Cytochalasin E Gene Cluster from *Aspergillus clavatus* NRRL 1." *Metab. Eng.* 2011, 13, 723-732.

The work described in this dissertation is supported by National Institute of Health Grants (YT: 1R01GM085128 and 1R01GM092217), and the David and Lucile Fellowship in Science and Engineering to Y.T., in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair in Bioorganic and Medicinal Chemistry.

There are many individuals without whom the work in this thesis would not have been accomplished and to whom I greatly indebted.

First and foremost, I would like to sincerely thank my supervisor Professor Yi Tang for recruiting me from China and offering me the opportunity to conduct the natural product research of my own interest; and for his guidance, patience and encouragement throughout my five-year PhD training. I would like to express my thanks to Professor John C. Vederas and his graduate student Zhizeng Gao for the chemical synthesis in our collaborated projects on resorcylic acid lactones. I am very grateful to Professor James C. Liao, Associate Professor Neil Garg and Professor Harold Monbouquette for their advice and encouragement. Also, I am grateful to the University of California, Los Angeles, Dissertation Year Fellowship.

Similarly, I thank Dr. Jixun Zhan, Dr. Xinkai Xie and Dr. Wenjun Zhang for teaching me molecular biology and chemical analysis techniques. I am very grateful to Dr. Yit-Heng Chooi for the insightful discussion and showing me how to perform fungal genetics. Many thanks to my other colleagues, Dr. Suzanne Ma, Dr. Zhen Gu, Dr. Lauren Pickens, Yanran Li, Wei Xu, Jingjing Wang, Angelica Zabala, Anuradha Biswas, Peng Wang, Xue Gao, Ralph Cacho and Muxun Zhao, in the lab and department for their company, friendship and support. I would like to thank my friend Chih-Ning Pao, Haorui (Patrick) Zhang and Xiaolei Wang for make my life fruitful and enjoyable.

I would like to thank my parents and my twin sister for being proud of me. Their unconditional love and support have carried me along my academic journey over the past five years while I am 9950 miles away from home. Lastly, I would like to fondly thank my girlfriend Hui Zhou, for being the most supportive and helpful individual in my life, without whom the study would not have been possible.

VITA

2007	B. S., Chemical Engineering Tianjin University Tianjin, China
2008-2011	Teaching Assistant Department of Chemical and Biomolecular Engineering Unive rsity of California, Los Angeles
2007-2011	Research Assistant Department of Chemical and Biomolecular Engineering University of California, Los Angeles
2011-2012	Dissertation Year Fellowship University of California, Los Angeles Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

5. Zhan, J., **Qiao, K.**, Tang, Y.* "Investigation of Post-PKS Tailoring Modifications in Pradimicin Biosynthesis." *ChemBioChem*, 2009, 10, 1447-1452.

6. Zhou, H., **Qiao, K.**, Gao, Z.; Meehan, M. J., Li, J., Dorrestein, P. C., Vederas, J. C.*, Tang, Y.*, "Enzymatic Synthesis of Resorcylic Acid Lactones by Cooperation of Fungal Iterative Polyketide Synthases Involved in Hypothemycin Biosynthesis." *J. Am. Chem. Soc*. 2010, 132, 4530-4531.

7. Zhou, H.[†], **Qiao, K**.[†], Gao, Z., Vederas, J. C.*, Tang, Y.*, "Insights into Radicicol Biosynthesis via Heterologous Synthesis of Advanced Intermediates and Analogs." *J. Biol. Chem*. 2010, 285, 41412-4121.

8. **Qiao, K.**, Zhou, H., Xu, W., Zhang, W., Neil, G., Tang, Y.*, "Identification of a Fungal Nonribosomal Peptide Synthetase Module that can Synthesize Thiopyrazines" *Org. Lett.* 2011, 13, 1758-1761.

9. **Qiao, K.**, Chooi, Y.*, Tang, Y.*, "Identification and Engineering of Cytochalasin E Gene Cluster from *Aspergillus clavatus* NRRL 1." *Metab. Eng.* 2011, 13, 723-732.

10. Xie, X., **Qiao, K.**, Tang, Y., "Metabolic Engineering of Natural Product Biosynthesis." In: Tringali, C., ed. *BIOACTIVE COMPOUNDS FROM NATURAL SOURCES. Second Edition:*

Natural Products as Lead Compounds in Drug Discovery. CRC Press. July, 2011.

11. Zhou, H.[†], Gao, Z.[†], **Qiao, K.**, Wang, J., Vederas, J. C.*, Tang, Y.*, "A Fungal Ketoreductase Domain that Displays Substrate-dependent Stereospecificity." *Nat. Chem. Biol.*, 2012, 8, 331-333.

"A Thiopyrazine Synthetase Discovered from Genome Mining of *Aspergillus terreus*" *SIM annual meeting, San Francisco, CA. 2010, August*

"Discovery and Characterization of A Thiopyrazine Synthetase from Genome Mining of *Aspergillus terreus*" *241th ACS national meeting, Anaheim, CA. 2011, March*

"A Fungal Nonribosomal Peptide Synthetase Module that can Synthesize Thiopyrazines" *ASP annual meeting, San Diego, CA. 2011, July-August*

1. Introduction

1.1. Background of Natural Products

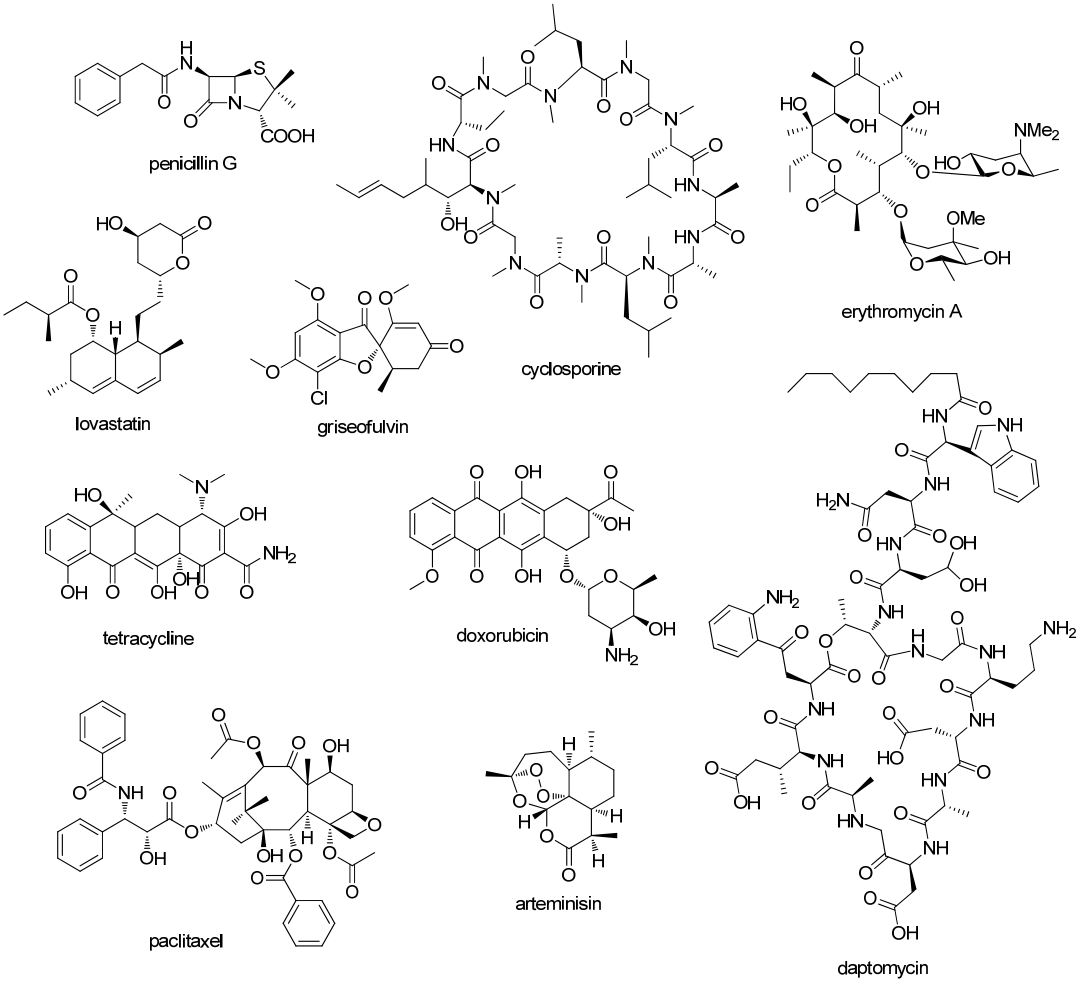


Figure 1. Examples of medicinally important natural products.

Natural products are a family of low molecular weight chemical compounds isolated from all kinds of living organisms, including bacteria, fungi, plants and animals. Natural products play significant roles in drug discovery and development [1,2]. In ancient times, mankind had known to search natural sources for poisons and other herb mixture for curing diseases. Over two hundred years ago, inspired by the isolation of pure morphine from *Papaver somniferum* [3], people started realizing that medicinal molecules could be purified from natural organisms. As a

result, numerous microbes and plants have been cultured, extracted and screened for bioactive compounds. This effort was greatly encouraged by the fortuitous discovery of the antibiotic penicillin by Alexander Fleming in 1928 [2]. By the year of 1990, about 70% of approved drugs were either natural products themselves or derived from natural products [4], including antibiotics (e.g. penicillin, tetracycline, erythromycin), anti-malarial agent (e.g. artemisinin), cholesterol-lowering drug (e.g. lovastatin, compactin), antifungal agents (e.g., grisefulvin), immunosuppressants (e.g. cyclosporine, rapamycin) and anti-cancer drugs (e.g., doxorubicin, taxol) (Figure 1). In modern drug market, natural products are still the major sources for developing innovative pharmaceuticals. From year 1991 and 2010, 5% of the total 1355 newly approved drugs were natural products, and 23% of them were natural product derivatives [1] (Figure 2).

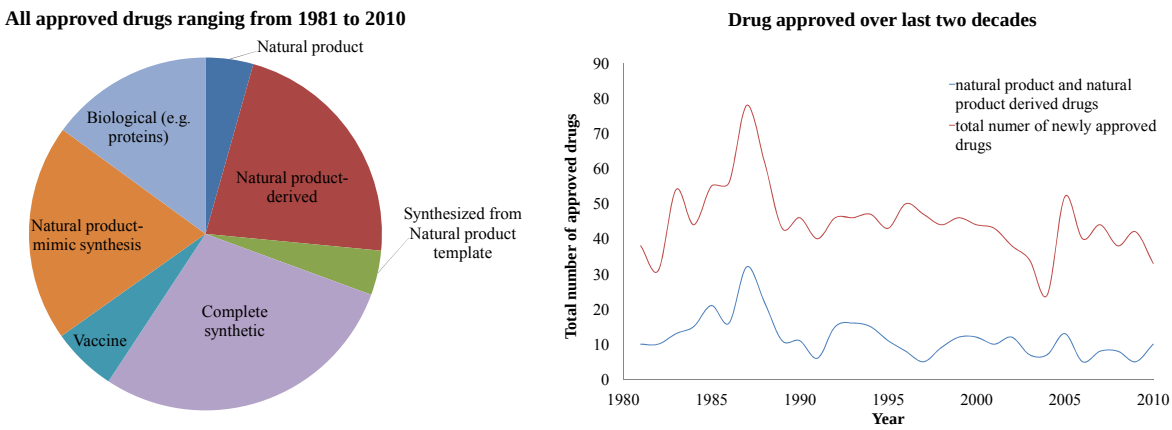


Figure 2. All 1355 new small-molecule approved drugs from 1981 to 2010 by (A) source and (B) year.

1.2. Biosynthesis of Natural Products

Natural products all belong to secondary metabolites that are not necessary for the growth of cells. The enzymatic machinery of living systems have evolved over several billion years, leading to these carefully controlled the biosynthetic pathways of metabolites with a pyramid of

basic skeletons and functional groups. Different from the self-sustaining primary metabolites, secondary metabolites are known to endow the selective advantages to the producing organisms over evolution. Therefore, secondary metabolites vary greatly among different species and exhibit a wide range of diversity and complexity in their structures. According to the scaffolding building blocks shared in the biosynthesis, natural products can be categorized into four major classes- alkaloids, terpernoids, polyketides, and nonribosomal peptides.

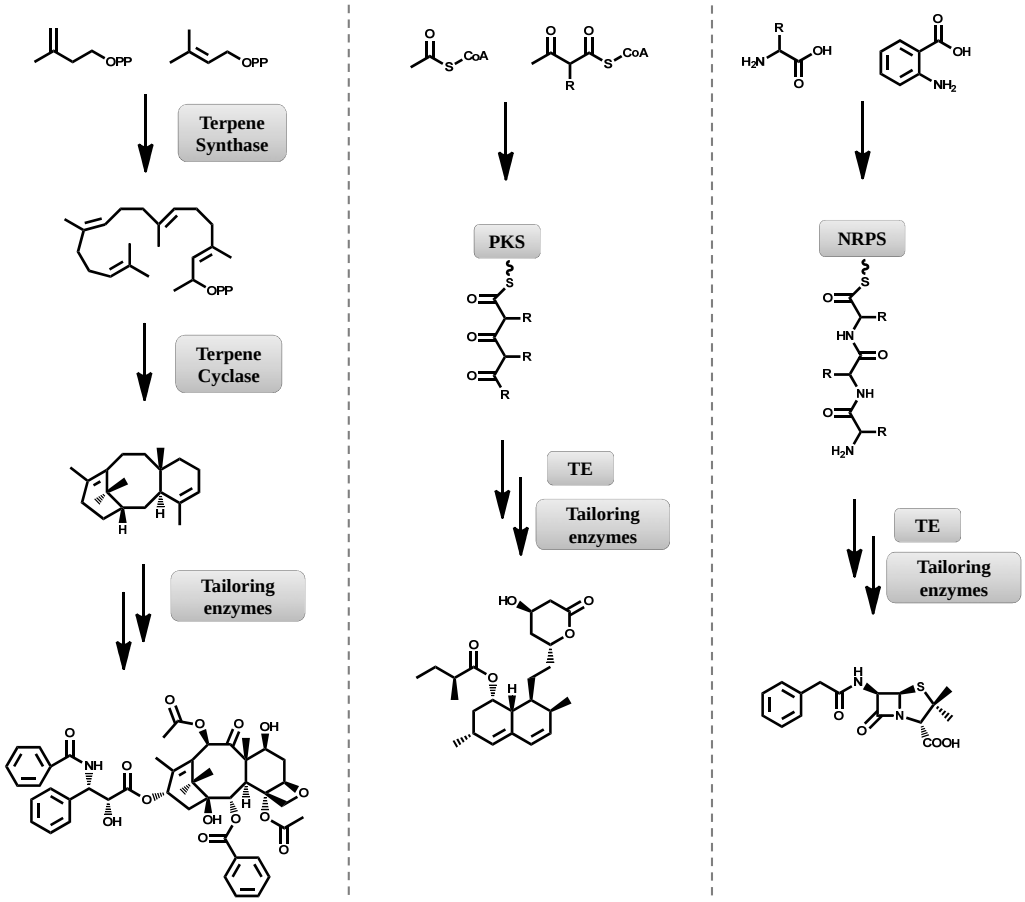


Figure 3. General biosynthetic pathways for terpenoids (exemplified by paclitaxel), polyketides (exemplified by lovastatin), and nonribosomal peptides (exemplified by penicillin).

Alkaloids are a class of plant natural products that contain one or multiple nitrogen atoms in the structure. They are generally derived from some nitrogen-containing precursors, e.g. amino acids. Isoprenoids (terpenes) are a large and diverse class of natural products produced by

nearly all species with approximately 50,000 known structures. All of them are derived from five carbon isoprene building blocks- isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). According to the numbers of isoprene unit, isoprenoids are classified as hemiterpenoids (C5), monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), etc.

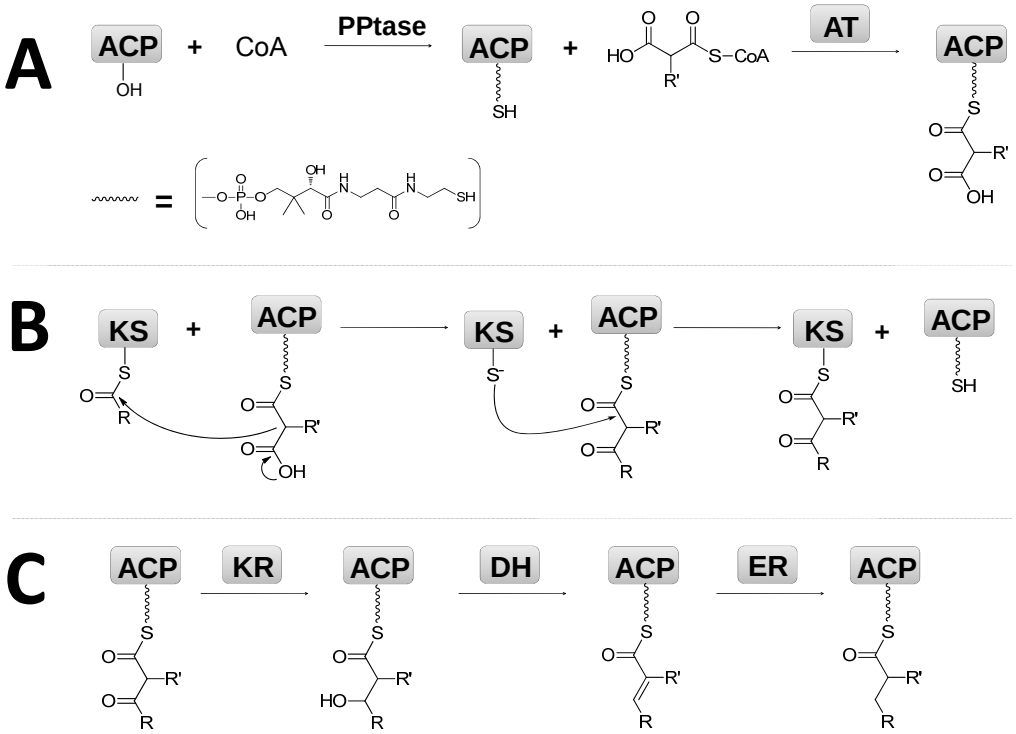


Figure 4. Basic mechanism involved in the fatty acid/ polyketide chain elongation. (A) Phosphopantetheinyl transfer modification of apo-ACP is catalyzed by a PPTase to form the holo-ACP. (B) The extension of the polyketide chain is achieved by successive decarboxylative Claisen condensations. (C) The growing polyketide chain can be selectively modified by KR, DH, and ER to a reduced plolyketide backbone.

Polyketides are synthesized by a group of enzymes called polyketide synthases (PKSs). Polyketide synthases operate in a similar fashion to fatty acid synthases (FASs) in building a carbon backbone from simple carbon building blocks, such as acetate. Both PKSs and FASs are composed of catalytic domains that work coordinately to condense units of activated acyl-CoAs. The polyketide synthases perform successive decarboxylative Claisen condensations between the growing polyketide chain attached to the keto-synthase (KS) domain and the extender unit

attached to acyl-carrier protein (ACP) domain to synthesize the polyketide backbone (Figure 4). The malonyl-CoA transacylase (MAT) is responsible for carrying out the loading of specific extender unit onto the terminal thiol of ACP domain. Following every elongation of the polyketide backbone, other accessory tailoring domains such as β-keto-reductase (KR), dehydrase (DH), and enoyl-reductase (ER) can selectively reduce the β-keto of the growing polyketide chain. A methyl group may be added onto nucleophilic α position of the growing chain in the occurrence of the methyltransferase (MT) domain. These tailoring domains can be used in different combinations to produce backbone β-carbons of different oxidation states. All polyketide ACP domains must first be posttranslationally modified at the active site serine within the conserved DSL motif by attaching phosphopantetheine to the hydroxyl. This modification of the ACP domain is typically catalyzed by a phosphopantetheinyl transferase (PPTase), either dedicated to the gene cluster, or shared with FAS or other PKS clusters in the host. The MAT domain selectively activates the extender unit and loads it to the ACP domain for chain extension.

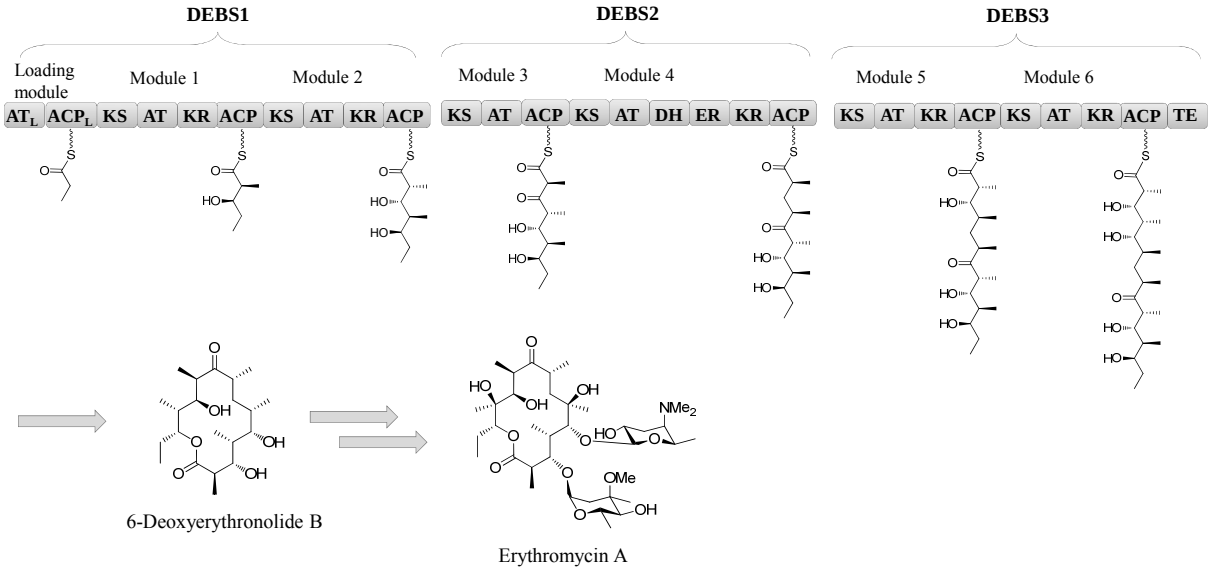


Figure 5. Erythromycin biosynthetic pathway.

Based on the domain organization and mode of action of biosynthetic enzymes, PKSs can be classified into three types [5,6]: type I PKSs represent very large multifunctional protein containing covalently linked function domains, while type II PKSs are a group of dissociated mono-functional proteins. This classification is also cleared represented by the phylogenetic analysis on the amino acid sequence of KS domains [7]. Moreover, types III PKSs contain single KS domains that catalyze condensations between acyl-CoAs in the absence of ACPs. Like their FASs counterparts, type I PKSs are commonly found in bacteria and fungi, type II PKSs exclusive exist in bacteria, whereas type III PKSs are widely distributed in plants, bacteria and fungi.

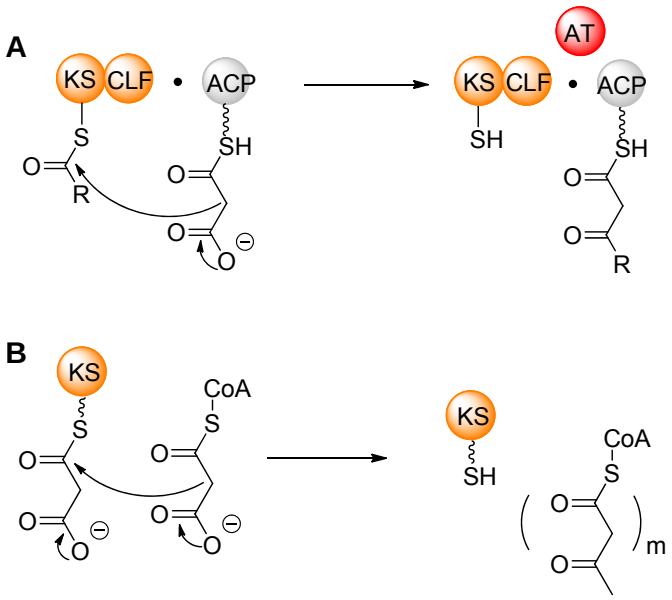


Figure 6. General scheme of the biosynthesis of polyketides by (A) type II PKSs and (B) type III PKSs.

Type I PKSs can also be divided into iterative and modular (non-iterative) systems dependent on whether KS domain can catalyzed more than one chain elongation cycle or not. Type I modular PKSs are common in all kinds of bacteria. As one of the well-studied type I modular PKSs, the biosynthesis of erythromycin by 6-deoxyerythronolide synthase (DEBS) is shown in Figure 5 as a representative [8]. In type I modular PKSs, each catalytic domain or ACP

domain is used only once as the growing polyketide chain is delivered from the N-terminal module to the C-terminal module. The number of the modules defines the number of elongation, in other words, the chain length of the nascent polyketide chain, while the presence of KR, DH and ER determine the degree of β-keto reduction [9,10]. The canonical correlation between the domain organizations and final polyketide structures can be summarized as the principle of colinearity. There is also another large group of type I modular PKSs, which lack the AT domains in each modules but share a standalone AT commonly appeared within the same gene cluster [11]. In trans-AT systems, the KS domains showed very strict substrate specificity towards start unit and the growing nascent polyketide intermediates, allowing researcher to derive an algorism to predict the precise product merely based on the primary sequence of KS domains [12]. However, some of the bacterial type I modular PKSs don't completely obey the colinearity rule but occasionally skip or repetitively use one module. Though it seems not to fit every modular PKS, the colinearity principle was not only widely employed to predict the polyketide structure based on the architecture of type I modular PKSs, but also provide us with paradigm to reprogram the biosynthetic logic for novel bacterial polyketides.

Iterative type I PKSs, such as the lovastatin synthase, was initially known to exclusively exist in fungi until the recent discovery of several bacterial iterative type I PKSs, including PKSs that synthesize members of the enediyne family [13], like orsellinate synthase for calicheamicin [14] and 2-hydroxyl-5-methyl-naphthoate synthase for neocarzinostatin [15]. Due to the fact that these multi-domain enzymes act in an iterative fashion, the programming rules are hidden and cannot be simply predicted from domain organizations. Different starter units and extenders can be recursively utilized by iterative type I PKSs to yield up to twenty-carbon polyketide backbone. Besides, the degree of the reduction varies in that the β-processing domains KR, DH and ER

along with the MT are optionally functions in each elongation cycle. To date, knowledge on this specific group of megasynthases largely remains unknown.

Nonribosomal peptides represent a large family of secondary metabolites from a variety of organisms. A vast majority of these natural products exhibit important biological activities, such as antibiotics penicillin [16] and vancomycin [17,18], siderophore enterobactin [19,20] and immunosuppressive agent cyclosporine A [21] (Figure 1).

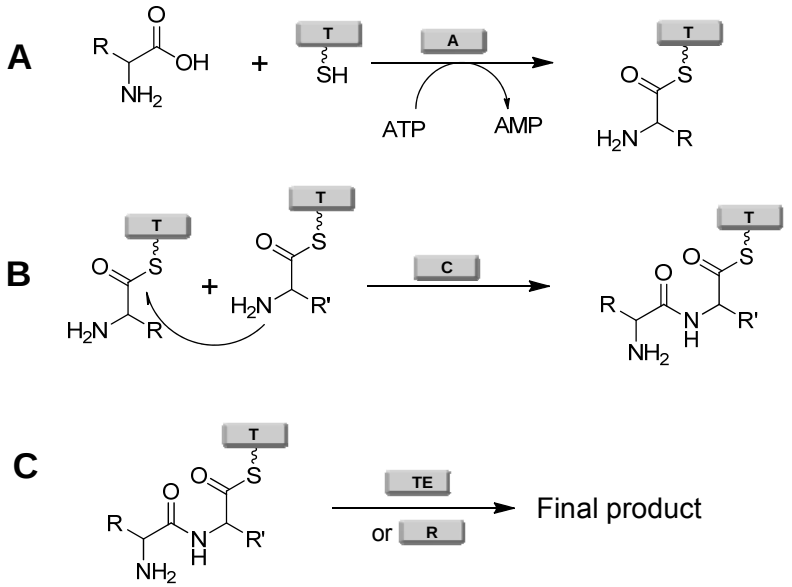


Figure 7. Biosynthesis of nonribosomal peptides. (A) Recognition, activation and loading of amino acid. (B) Condensation between the downstream and upstream aminoacyl-S-T. (C) Product realsing from thiolation domain.

Nonribosomal peptides are biosynthesized by a group of megasynthetases called Nonribosomal Peptide Synthetases (NRPSs) [22]. Typical NRPSs share the similar programming logic with PKSs but utilize a more diverse set of building blocks- amino acids [23]. Each NPRS is composed of a series of modules, whereby each module harbors unique domain organization and catalyzes one round of addition of a specific amino acid [24]. A classical NRPS module organized as C-A-T. Adenylation (A) domain activates one amino acid to aminoacyl-O-AMP by consumption of ATP and then tethers it to the phosphapantethiol arm of Thiolation (T) domain,

while Condensation (C) domain catalyzes the formation of amide bond between two amino acids (Figure 3.6). Some tailoring domains are also involved to amplify the diversity of NRPs. For example, Oxidase (Ox) domain in Epothilone biosynthesis is used to transform thiozole into thiozale. Epimerase (E) domain alters the stereochemistry of α positions of aminoacyl groups. The releasing domain, either a thioesterase (TE) or a reduction (R) domain, is responsible for releasing the nascent aminoacyl chain from the T domain of the last module. Three different releasing mechanisms are present (Figure 3.7): first, a linear peptide can be formed via direct hydrolysis by TE; second, macrocyclization catalyzed TE domains can afford peptidyl macrolactones or macrolactams; and lastly, an R domain can catalyze release of products on many PKS/NRPS hybrid assembly lines via either reduction or Dieckman cyclization [25] (Figure 3.6). Detailed reviews of NPRSs are available in numerous review papers [22,26].

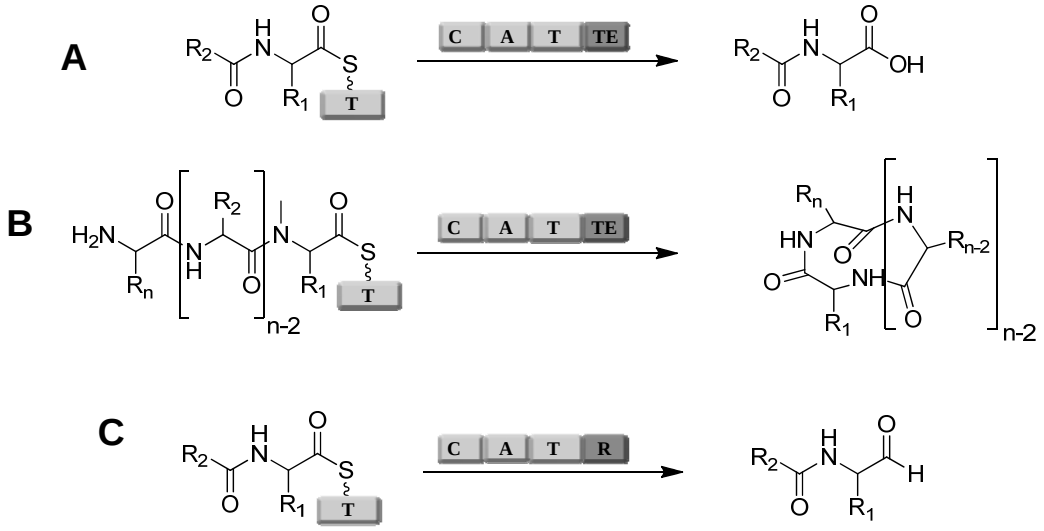


Figure 8. Different releasing mechanisms of NRPSs. (A) TE-catalyzed hydrolytic release (B) TE-mediated intra-molecular macrolactamization. (C) R domain directed reductive release.

1.3. Fungal Polyketide Synthase and Polyketide Synthase-Nonribosomal Peptide Synthetase Hybrid

The filamentous fungi are abundant resources of natural products, of which polyketides (PKs) and nonribosomal peptides (NRPs) constitute two large classes with diverse structural features [5]. Many PKs and NRPs together with derivatives thereof, play key roles in pharmaceuticals, agricultural, environmental applications, for instance, the famous antibiotics penicillin from *Penicillium* phyla [27,28] and Cholesterol-lowering medicine such as lovastatin [29] from *Aspergillus terreus*. Other fungal important secondary metabolites include mycotoxins aflatoxin and sterigmatocystin [30-32] from *Aspergillus* species, and fumonisin [33] from *Fusarium verticillioides*, and antitumor compound bikaverin [34] from *Gibberella zeae*. Furthermore, Radicicol [35,36] from *Pochonia chlamydosporia*, which belongs to a so-called resorcylic acid lactone family, is recently discovered to be a nanomolar inhibitor of the chaperone Hsp90. In contrast with the commonly existed PKS and NRPS pathways, fungi are also capable of producing numerous compounds that synthesized by the combination of the two. For example, 2-pyrrolidinones such as fusarin C [37,38] from *Fusarium venenatum*, equisetin [39,40] from *Fusarium heterosporum*, chaetoglobosin A [41,42] from *Penicillium expansum* and pseurotin A [43,44] from *Pseudeurotium ovalis* have been initially demonstrated by classical isotope feeding experiments to be derived from polyketide and amino acid components. Other compounds such as the fungal 2-pyridones including tenellin [45] from *Beauveria bassiana*, and aspyridone A [46] from *Aspergillus nidulans*.

Fungal polyketides are biosynthesized by fungal type I iterative polyketide synthases (PKSs) [47]. Interestingly, it is now clear that all fungal PKSs so far known belong to type I iterative PKSs (except for the type III PKSs and diketide synthase LovF). There are three stages involved

in iterative fungal polyketide synthesis: initiation, elongation, and termination, which are also employed by bacterial type I PKSs. However, in sharp contrast to modular type I PKSs, those catalytic domains on fungal PKSs are iteratively utilized during the product synthesis, which make it extremely hard to predict the fungal polyketide products based on the domain organization of fungal PKSs. Based on the presence of the β-processing domains- KR, DH and ER, the fungal type I iterative PKSs can be further divided into three subgroups [5]: Non-reducing PKSs, in which no reductive domains are present, mainly produce aromatic fungal polyketide products [48]. Partially-reducing (PR) PKSs, usually containing a single KR domain or KR-DH didomain, catalyze limited reductions during the formation of the polyketide and interestingly a Core domain is present with unknown function within PR-PKSs [49-51]. Highly-reducing PKSs enclose full set of β-processing domains- KR, DH and ER [52,53].

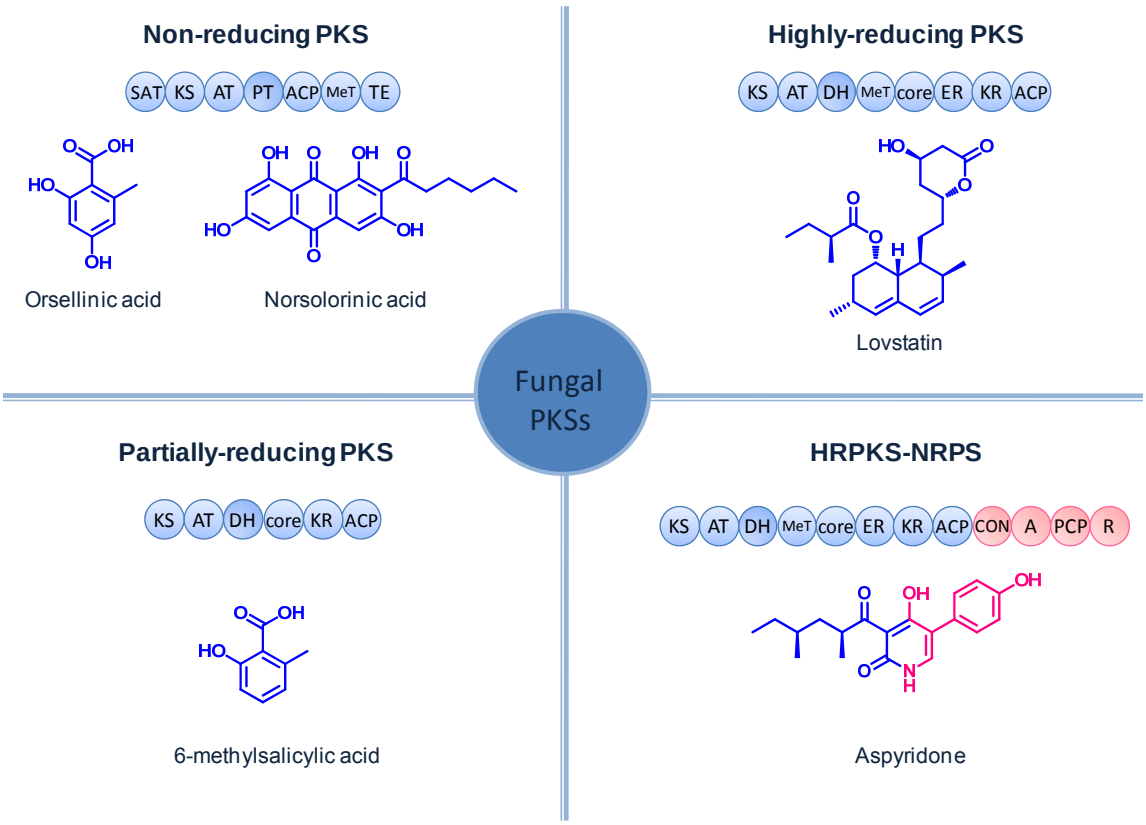


Figure 9. Examples of the four subgroups of fungal iterative type I PKSs. Norsolorinic acid and orsellinic acid are produced by a nonreducing PKS (NR-PKS); 6-methylsalicylic acid is produced by a partially reducing PKS; lovastatin is produced by two highly reducing PKSs (HR-PKS); and aspyridone is produced by a HR-PKS-NRPS hybrid.

1.3.1. Non-reducing Polyketide Synthase

The first discovery of fungal NR-PKSs dates back to 1968, when the isolation of orsellinic acid synthase (OSAS) performed in *Penicillium madriti* [54]. Although nothing on the sequence and the domain organization of the megasynthase were known by then, this really provides us with insights that a single megasynthase is able to assemble the tetraketide orsellinic acid in a programmed manner. Benefiting from the sequencing and protein identification techniques, the genes involved into the biosynthesis of fungal aromatic polyketides were identified and a general organization of all catalytic domains were summarized. At the N-terminus, a domain is shown to involve in the starter unit selection. The key members of the minimum PKS (KS-AT) were followed and the ACP domain was separated from KS-AT by the so-called product template domain (PT). The KS domains play key roles in controlling the chain length of the polyketide product, which nearly all AT domains were found to selectively activate malonyl-CoA as the extender for polyketide synthesis. After the ACP domain, a diverse range of domains may be present, including thioesterase, cyclase, methyltransferase and reductase. Occasionally, ACP is the last domain of the fungal NR-PKSs (Table 1). The domain organization of fungal NR-PKSs is in accord with the type of secondary metabolites they synthesize. All domains described above normally are responsible of dictating the programming rules -starter unit selection, chain-length control, cyclization pattern determination and product release mechanism.

Table 1. Characterized fungal NR-PKSs examples.

Origin	*Gene*	*Protein*	*metabolite*	*Domain architecture*	*ref*
Aspergillus parasiticus	*pksA*	NSAS	Aflatoxin B_1	SAT-KS-AT-PT-ACP-TE/CLC	[55-57]
Aspergillus	*pksST*	NSAS	Sterigmatocystin	SAT-KS-AT-PT-ACP-TE/CLC	[55]

nidulans					
Dothistroma septosporum	*pksA*	NSAS	Dothistromin	SAT-KS-AT-PT-ACP-TE/CLC	[58]
Aspergillus nidulans	*wA*	WAS	YWA1	SAT-KS-AT-PT-ACP-TE/CLC	[59]
Aspergilus fugmitatus	*alb1*	alb1p	YWA1	SAT-KS-AT-PT-ACP-TE/CLC	[60]
Colletotrichum lagenarium	*pks1*	THNS	Tetrahydroxyl naphthalene	SAT-KS-AT-PT-ACP-TE/CLC	[61]
Wangiella dermatitidis	*WdPKS1*	WdPks1	acetyl-tetrahydroxyl naphthalene	SAT-KS-AT-PT-ACP-ACP-TE/CLC	[62,63]
Monascus purpureus	*pksCT*	CitS	Citrinin	SAT-KS-AT-PT-ACP-R	[64]
Gibberella zeae	*pks4*	PKS4	bikaverin	SAT-KS-AT-PT-ACP-MT-R	[65-67]
Aspergillus nidulans	*orsA*	OSAS	orsellinic acid	SAT-KS-AT-PT-ACP	[68,69]
Acremonium strictum	*Aspks1*	MOS	3-methylorcinaldehyde	SAT-KS-AT-PT-ACP-ACP-TE/CLC	[70]
Aspergillus nidulans	*afoE*	AfoE	asperfuranone	SAT-KS-AT-PT-ACP-MT-R	[71]
Aspergillus terreus RED1	*pks*	PKS	emodin anthrone	SAT-KS-AT-PT-ACP-ACP-TE/CLC	[72]
Aspergillus terreus NIH2624	*ACAS*	ACAS	atrochrysone carboxylic acid	SAT-KS-AT-PT-ACP-ACP-TE/CLC	[73]
Aspergillus nidulans	*aptA*	AptA	asperthecin	SAT-KS-AT-PT-ACP	[74]
Penicillium aethiopicum	*vrtA*	VrtA	viridicatumtoxin	SAT-KS-AT-PT-ACP	[75]
Penicillum aethiopicum	*gsfA*	GsfA	griseofulvin	SAT-KS-AT-PT-ACP	[75]

In the mid of last century, isotopic labeled acetate or malonate was supplemented into the cultures of a variety of fungi, allowing the understanding of the origins of polyketides. As a result of those isotopic labeled substrates feeding studies, researcher managed the incorporation

of ^{14}C labeled acetate into the final structure of 6-methylsalicylic acid [76,77]. The backbone of the unreduced polyketide was carefully analyzed through the chemical degradation of the parent polyketides. Later, facilities by the NMR techniques, fungal polyketides can be traced using non-radioactive elements such as 2H, 13C, 15N, 18O, etc. Learning by these feeding data collected over 60 years, most of the fungal PKSs choose acetyl-CoA as starter units. Fungal NR-PKSs can be distinguished from other fungal PKSs by the fact that they can accept more advanced starter unit than acetyl-CoA. For example, the biosynthesis of the fungal carcinogen aflatoxin in many *Aspergillus* species involved an NR-PKS (known as PksA), which can accept a C6 fatty acid starter unit from a dedicated fungal FAS systems [78]. Analysis of the aflotoxin gene cluster in *Aspergillus parasiticus* allow Townsend and his coworkers identified two genes encoded the α and β components (HexA and HexB) of a typical fungal FASs [79,80]. Both of the PksA and HexA/HexB were isolated and characterized by in vitro assay containing acetyl-CoA, malonyl-CoA and NADPH [81]. This biochemical observation led to the conclusion that HexA/HexB FAS system can synthesize the hexanoate starter unit in the presence of acetyl-CoA, malonyl-CoA and NADPH. Cox and his coworkers manage to show that the AT domain of PksA was solely in charge of loading malonyl-CoA onto the ACP to facilitate the chain elongation, ruling out the possibility of the involvement of AT domain in the starter unit chain transfer [82]. Additional domain within PksA is apparently required to fulfill the starter unit transfer from the FAS to the PKS. The N-terminal SAT domain was therefore found and defined in an effort to elucidate this specific reactions. Townsend and his coworkers first realized that the N-terminus of the PksA contains a canonical acyl transferase and could be a candidate for the chain transfer. In vitro characterization of the SAT domain along with the mutagenesis study on the proposed cysteine residue confirmed the presence of the SAT domain in the N-terminus of PksA [83].

Furthermore, the SAT domain is the common feature in a lot of NR-PKSs, even though they still choose acetyl-CoA over other more For instance, bikavaerin synthase PKS4 from *Gibberella fujikuroi* was initially discovered to be able to utilize malonyl-CoA, rather than acetyl-CoA, for produce SMa96 in vitro [84]. However, lack complex substrates as the starter unit.

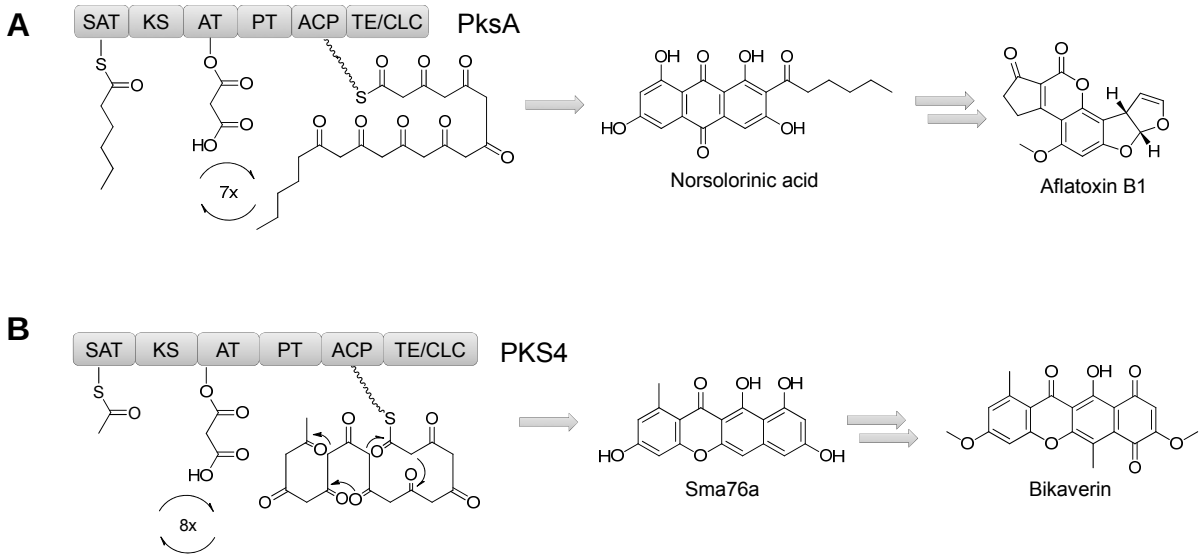


Figure 10. Biosynthetic pathways for (A) aflatoxin B1by PksA and (B) bikaverin by PKS4.

of the active site GXCXG in the SAT domain remind researcher to re-sequence and annotate the PKS4. By the method, an intact SAT domain containing the active site cysteine was recovered and shown to prefer acetyl-CoA over malonyl-CoA in the in vitro radio-labeling experiments. Same starter unit selection was also realized in lovastatin synthase LovB, which can use malonyl-CoA as starter unit in vitro while the sequencing analysis favors the acetyl-CoA [85]. Besides the SAT domains described above, a number of SAT domains were recently found in the NR-PKSs of HR-PKS/NR-PKS tandem systems such as those responsible for the biosynthesis of resorcylic acid lactones [86-88]. Extensive study of the SAT domain of fungal NR-PKSs not only can help us build the sequence-structure relationship but also provide us extra tool to diversity aromatic polyketides through engineering the starter unit alternations.

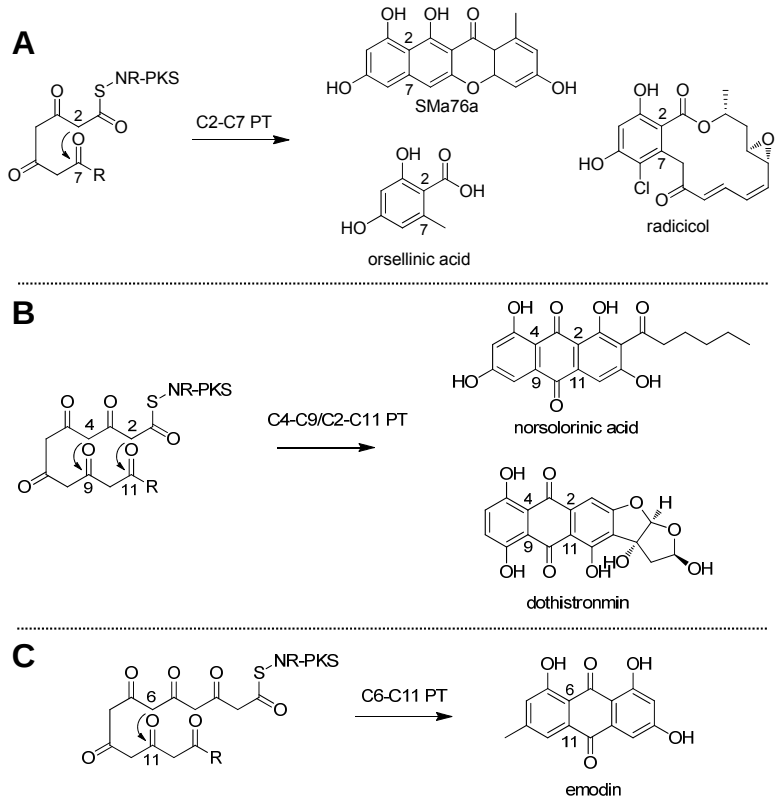


Figure 11. Schematic illustration of different cyclization patterns catalyzed by three types of PT domain.

Common to all PKSs, the chain extension of the fungal NR-PKSs rely on the so-called minimum PKS, including KS, AT and ACP. Due to the large size of fungal PKSs, they are usually difficult to be attained as purified proteins to sufficient quantity for investigating the enzymology. In spite of the fact that some megasynthases can be obtained in sufficient amount with the assistance of suitable heterologous workhorses, the function of single catalytic domain frequently is normally hidden. These problems started to be overcome by deserting the function type I PKSs megasynthases into catalytically functional single domains [89,90]. As SAT domain, the AT domains were generally accepted to load malonate instead of other acyl and hexanoate from CoA onto terminal thiol of the phosopantetheinyl arm of the ACP domain. The KS domain catalyzes the decarboxylation of malonyl group and the Claisen-type formation of carbon-carbon bond between the growing polyketinde nascent chain and the extender. The KS domain was

thought to largely, if not entirely, determine the chain length of the polyketide. In order to verify this notion, the KS, AT and ACP domains of PKS4 from *G. fujikuroi* were purified independently from *E. coli* and incubated in vitro in the present of the substrate malonyl-CoA, the primary products retained the correct chain length (nonakeitde) but with different cyclization patterns to that of bikaverin [90,91]. Similar results were obtained by deconstructing PksA [89].

The cyclization of the nascent polyketide chain produced by KS-AT-ACP was mediated controlled either by PT domain, or by TE domain or CLC domain, depending on the presence of them [92]. With the addition of PksA-PT domain into the minimal PKSs assay, obvious increases of the product with correctly cyclized first two rings were detected using HPLC-MS [89]. The PksA-PT domain therefore was thought to be responsible for catalyzing its stepwise cyclizations to generate mature polyketides. Based on the series of in vitro studies using PKS4-PT, the PT domain is in charge of catalyzing the first-ring cyclization by joining C2 and C7 via aldo condensation [84]. The cyclization of bikaverin also involves a terminal TE/CLC domain, which mainly mediate the cyclization of the rest of rings except for the first ring and finally releasing of the intermediate from the megasynthases. These biochemical findings indicate that PT domains are responsible for catalyze the first ring cyclization in producing non-reduced fungal polyketides. The cyclization regioselectivities of the PT and TE domains give rise to the F-mode cyclization observed for nearly all the fungal aromatic polyketides Based on the different PT domains' regioselectivity, the PT domain mediated cyclizations can easily be classified into three categories: C2-C7, C6-C11 and C4-C9 [92]. Both PT domains from PKS4 and the first discovered osellinic acid synthase were known to direct C2-C7 type of first-ring cyclization [93]. Other examples include the PT domains from the fungal NR-PKSs involving in the biosynthesis of resorcylic acid lactone family of fungal polyketides-zearalenone, hypothemycin and radicicol

[86]. Moreover, a subgroup of fungal NR-PKSs contains a reductive (R) domain and an additional methytransferase (MT) domain and is thus classified as Clade III NR-PKS via phylogenetic analysis [94]. For example, citrinin, produced in *Aspergillus*, *Penicillium* genera, is the first characterized member of the group of compounds [95]. The PT domains in this subset most likely catalyze the C2-C7 cyclization to afford the first aromatic ring. The R domain is proposed to catalyze the reductive release of the polyketide chain and the resulting aldehyde function group serves as the basis for the spontaneous cyclization of the second ring.

When the polyketide chain reaches longer than a pentaketide, the fungal PT domains can regioselectively cyclize the first ring with C4-C9 cyclization pattern. Among all polyketide products retaining this pattern, norsolorinic acid, which is the precursor of aflatoxin B was the most well-studied [80]. The total cyclization framework of norsolorinic acid were accomplished by PT mediated C4-C9 first and C2-C11 aldol cylizations, and the C14-C1 Claisen-like condensation performed by TE/CLC domain, followed by the product release. The protein structure of PksA PT domain was recently solved by Townsend and his coworkers [57]. The PksA PT domain adopts a 'double hot dog' fold resembling the structure of bacterial cyclase TcmN [96]. The substrate binding pocket can be divided into three regions: the phosphopantetheine binding channel, the ACP domain binding region and the central cyclization chamber. The inner end of the pocket is proposed to be a section that holds the hexyl moiety from the hexanoate starter unit. The linear arrangement of these binding regions is proposed to hold the linear poly-β-keto chain and facilitate the F-mode C4-C9 and C2-C11 cyclization. Therefore, PT domain functions as a template role for the cyclization. The proposed catalytic dyad, Asp1543 and His1345 were confirmed by mutagenesis studies, revealing the catalytic role

of PTs [57]. Other polyketides encompass the C4-C9 pattern include Xanthone structures such as demethylsterigmatocystin [97], and dothistronmin from *Dothistroma septosporum* [98].

A large family of fungal non-reducing polyketides is anthraqunones, among which emodin is the model compounds. It has been isolated from multiple fungal genera, including *Aspergillus, Penicillium,* and *Cladosporium.* The atrochrysone carboxylic acid synthase (ACAS) involving in the biosynthesis of emodin was realized from *A. terreus* NIH2624 and is proposed to fold the first ring into C6-C11, followed by C4-C13 and C2-C15 cyclizations to establish the second and third rings [99]. Although the function of the PT domain within ACAS has not been verified, it is likely that the PT catalyzed the first two ring closures. It is also noted that the C-terminus of ACAS lacks the TE or CLC domains. However, an *in trans* dissociated metal-dependent β-lacttamase (ACTE) gene was identified right downstream of the NR-PKS gene [100]. ACAS and ACTE collectively function to finally cyclize the last ring [100].

Although only three specific ring closure types were summarized above, nature still can surprise us with a few exceptions. The antifungal compound griseofulvin from *Penicillium griseofulvum* attains the C1-C6 and C8-C13 folding pattern according to the ^{13}C acetate labeling experiments [101]. Recently, the gene cluster of griseofulvin was identified from *Penicillium aethiopicum* and the bioinformatics analysis on the core NR-PKS reveals that the NR-PKS bears the typical NR-PKS organization but lacks the releasing domain in the N-terminus [102]. The PT domain was believed to catalyze Claisen-like condensation that connected C1 and C6 in the absence of the proper catalytic domain that would carry out the reaction.

Besides the regioselective cyclization, the PT domains were also known to control the chain length of nonreducing polyketide. This is probably accomplished by the cavity volume and cyclization template function that are required to accommodate the linear intermediates.

Sequence alignment among many PT domains unveils the relationship between the polyketide structure and PT domain sequences. A phylogenetic studies performed by Tang and his coworkers revealed that PT domains fall into five subgroups based on the cyclization regioselectivity and number of the rings [103]. This allows us to predict the unknown PT functions purely according to their sequence data and provide sufficient insights into diversifying nonreducing polyketide cyclization patterns.

1.3.2. Partial Reducing Polyketide Synthase

Table 2. Characterized PR-PKSs from fungi.

Origin	Gene	Protein	Secondary metabolite	Domain architecture	ref
Penicilllium patulum	*MSAS*	MSAS	6-MSA	KS-AT-DH-Core-KR-ACP	[104]
Aspergillus terreus	*atX*	MSAS	6-MSA	KS-AT-DH-Core-KR-ACP	[105,106]
Glarea Lozoyensis	*pks2*	MSAS	6-MSA	KS-AT-DH-Core-KR-ACP	[107]
Aspergillus westerdikiae	*aomsaa*	MSAS	6-MSA, asperlactione	KS-AT-DH-Core-KR-ACP	[108]

The enzymology of fungal PR-PKSs is less known comparing to fungal NR-PKSs. PR-PKSs are defined according to the fact that they only contain part of the β-keto processing domains-KR, DH and ER [109]. The domain organization of PR-PKs differs from NR-PKSs in terms of lack of releasing domain after ACP domain and presence of Core domain instead of PT domain. Much fewer examples of PR-PKSs are known so far. The 6-methylsalicylic acid synthase (MSAS) was firstly discovered from *Penicillium patulum* [110]. The MSAS is one the smallest type I PKSs (only 191 KDa). Surprisingly, the fungal MSAS genes are closely related to the discovered iterative type I PKSs from bacterial origins that are responsible of the

biosynthesis of orsellinic acid. The heterologous expression of MSAS has been accomplished using *Streptomyces coelicolor* and *Saccharomyces cerevisiae* [111-113]. The MSAS selectively choose acetyl-CoA instead of malonyl-CoA as the starter unit. When acetyl-CoA is not available, MSAS can accept other acyl-CoA as the starter units. The chain length of MSAS's product is partially determined by KR domain's function. In the absence of NADPH, MSAS can lead to the formation of triketide lactone instead of the full length tetraketide product [51,114]. In order to examine the function of the unique Core domain, a number of mutant single polypeptide subunit of MSAS were constructed and supplementation experiments were performed. A short region of the core domain was identified, of which is essential for the functions of the mutant MSASs [113].

1.3.3. Fungal Highly Reducing Polyketide Synthase

Table 3. Characterized fungal HR-PKSs examples.

Origin	Gene	Protein	Metabolite	Domain architecture	Ref
Aspergillus terreus	*lovB*	LNKS	Lovastatin	KS-AT-DH-MT-ER*-KR-ACP	[85,115]
A. terreus	*lovF*	LDKS	Lovastatin	KS-AT-DH-MT-ER-KR-ACP	[115,116]
Penicillium citrinum	*mlcA*	CNKS	compactin	KS-AT-DH-MT-ER*-KR-ACP	[117]
Penicillium citrinum	*mlcB*	CDKS	compactin	KS-AT-DH-MT-ER-KR-ACP	[117]
Phoma sp.	*Phpks1*	SQTKS	Squalestatin tetraketide	KS-AT-DH-MT-ER-KR-ACP	[118]
Cochliobolus heterostrophus	*pks1*	TTS1	T-toxin	KS-AT-DH-MT-ER-KR-ACP	[119]
Cochliobolus heterostrophus	*pks2*	TTS2	T-toxin	KS-AT-DH-MT-ER-KR-ACP	[120]
Gibberella zeae	*fum1*	FUMS	Fumonisin B1	KS-AT-DH-MT-ER-KR-ACP	[121]
Alternaria solani	*pksN*	PKSN	alternapyrone	KS-AT-DH-MT-ER-KR-ACP	[122]
Alternaria	*pksF*	PKSF	aslanipyrone,	KS-AT-DH-MT-ER-KR-ACP	[123]

solani			aslaniol		
Alternaria solani	*Pks-AS2*	PSS	solanapyrone	KS-AT-DH-MT-ER-KR-ACP	[124]
Aspergillus nidulans	*afoG*	AfoG	asperfuranone	KS-AT-DH-MT-ER-KR-ACP	[71]

*: inactive domain.

Fungal HR-PKSs are large, multifunctional enzymes that are least understood among PKSs. Comparing to the type I bacterial modular PKSs, fungal HR-PKSs is less studied due to the lack of functional expression hosts as well as the mysterious product release mechanisms. However, they can make very important fungal secondary metabolites including lovastatin [85], T-toxin [125], fumonisin B1 [126] and squalestatin [127]. As the well-known cholesterol-lowering drug from *Aspergillus terreus*, lovastatin is mostly used as the precursor of simvastatin (Zocor, Merck) with more than 4 billion annual sales before the patent expiration in 2006. Biosynthesis of lovastatin was performed by two HR-PKSs- LovB and LovF with the assistance of a disassociate ER-LovC [85]. LovB is a 335 kDa megasynthase consist of the typical fungal HR-PKS architecture: KS-AT-DH-ER*-MT-KR-ACP, while LovF is a diketide synthase that only contains KR as its reductive domain. The ER* is structural ER domain without the NADPH binding site, indicating it cannot fully reduce the double bond but play a role in the protein-protein recognition. The direct product of LovB and LovC is nonketide dihydromonacoli L, of which the structure feature was gained from up to 35 independent reactions, including a biological Diels-Alder reaction. The study of LovB remained in genetic level for two decade until recently Tang and his coworkers manage to purify sufficient amount of active LovB from the engineered S. cerevisiae strain BJ5464-NpgA. A copy of phosphopantetheinyl transferase *npgA* gene from *A. nidulans* was integrated into the chromosome of the engineered host. The following in vitro reconstitution demonstrate that LovB can catalyze the formation of dihydromonacolin L but the yield of it remain to be a low level due to the lack of correct off-load

partner. Substitution of LovC with MlcG from compactin biosynthesis restores the production of dihydromonacolin L, proving the gate-keeping role of the reductive partner enzyme [85].

1.3.4. Fungal Polyketide Synthases and Polyketide Synthase- Nonribosomal Peptide Synthetase Hybrid

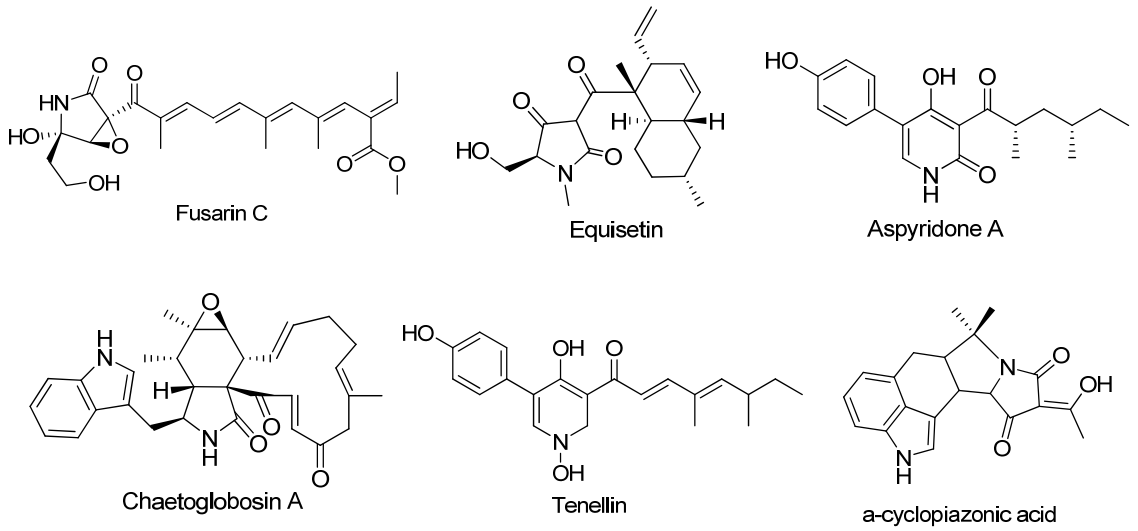


Figure 12. Examples of mixed polyketide-amino acids secondary metabolites produced by fungal PKS-NRPSs.

Among the diverse fungal reduced polyketide structures, many of them containing an amino acid derived heterocyclic five-member ring either belong to tetramic acids or related to pyrrolidine-2-ones. The biosynthetic origins of those metabolites are known to be acetate and amino acids and the biosynthetic enzymes were recently realized to be PKS-NRPSs hybrids. Benefiting from the recent fungal genome sequencing projects, a number of PKS-NRPS hybrids were identified and characterized over the last decade. For example, the biosynthetic PKS-NRPS enzymes for fusarin C from *Fusarium moniliforme* [38], equisetin from *Fusarium heterosporum* [128], aspyridone A from *A. nidulans* [129,130], pseurotin A from *A. famigatus* [131], cyclopiazonic acid from *A. oryzea* [132,133], and tenellin from *Beauveria bassiana* [134]

were extensively investigated either in natively producing host via genetic inactivation or in heterologous workhorses. The large hybrid enzymes (more than 450 kDa) appeared to be composed of a C-terminus HR-PKS fused to an N-terminal NRPS module. It is typical that PKS-NRPSs terminated with the reductive (R) domain and two proposed mechanisms were known to release the linear nascent aminoacyl chain from the PCP domain of the NRPS module [25]. In the first proposal, the R domain can release the intermediate in the reductive manner using NADPH as the cofactors, resulting in the formation of aldehydes or alcohols depending on how many electrons were transferred. On the other hand, Different from their relatives in the short chain dehydrogenase/reductases (SDRs), the Dieckmann-like R domain does not catalyze the redox reaction. Instead, R domain can catalyze a Dieckmann like condensation to directly give the tetramic acids [25].

Table 4. Examples of characterized fungal PKS-NRPS hybrids.

Origin	Gene	*Protein*	*metabolite*	*Domain architecture*	*ref*
Fusarium moniliforme	*ORF3*	FUSS	Fusarin C	KS-AT-DH-MT-ER*-KR-ACP-C-A-T-R	[38]
Fusarium heterosporum	*eqiS*	EqiS	equisetin	KS-AT-DH-MT-ER*-KR-ACP-C-A-T-R	[135,136]
Aspergillus fumigatus	*Afu8g00540*	Afu8g00540	Pseurotin A	KS-AT-DH-MT*-ER*-KR-ACP-C-A-T-R	[44]
Aspergillus oryzae	*CpaS*	CpaS	Cyclopiazonic acid	KS-AT-DH*-MT*-ER*-KR*-ACP-C-A-T-R	[137]
Penicillium expansum	*CheA*	CheA	Cheatoglobosin A	KS-AT-DH-MT-ER-KR-ACP-C-A-T-R	[138]
Beauveria bassiana	*ORF4*	TenS	Tenellin	KS-AT-DH-MT-ER*-KR-ACP-C-A-T-R	[45,139,140]
Aspergillus nidulans	*apdA*	ApdA	aspyridone	KS-AT-DH-MT-ER*-KR-ACP-C-A-T-R	[129,141]
Aspergillus terreus	*ATEG00325*	ATEG325	isoflavipucine	KS-AT-DH*-MT*-ER*-KR*-ACP-C-A-T-R	[142,143]

*: inactive domain

1.4. Strategies for Investigating and Engineering Fungal Natural Product Pathways

Given the fact that combinatorial synthetic libraries of compounds cover less chemical space comparing to natural products, natural products are still the major sources for new drug discovery [144]. As one of the less explored source, fungal natural products are becoming more and more popular among scientists. As the selective advantages that are conferred to the producing fungi, natural product pathway have been evolved to be active only under appropriate environmental conditions, such as lack of specific nutrients and challenging by other co-growing strains [145]. As a result, a number of second metabolite pathways were cryptic under the standard lab culture conditions [146]. Following the nature's rule, those silent pathways can be activated under certain conditions or in the presence of specific inducers [147]. The metabolites synthesized by these pathways represent a new source of natural products that have not been explored by the traditional drug discovery and development process. Furthermore, natural products are normally synthesized in a well-controlled manner to give just enough amounts to generate and maintain the selective advantage for the producing strains [148]. Hence, the titers of the wild type strains are much lower than what is needed for the industrial-scale production [148]. Besides, the chemical total synthesis of natural products was often hampered by the structural complexity of biologically active molecules, such as occurrence of multiple chiral centers and fused heterocyclic rings. Significant strain improvements have to be made in order to awake the silent natural products pathways and harvest the sufficient amount of target natural products. Conventionally, this aim is achieved through screening the randomly generated libraries of mutant strains. The higher producing strain was selected for the next cycle of mutation and screening. For instance, this 'black box' approach allow people to obtain a strain of *P. chrysogenum*, which is capable of producing penicillin 100 fold higher than its wild-type parent

strain [149]. Although great accomplishment have been made using this approach, in particular, when large portion of the knowledge regarding to genetics and natural product pathways remained unknown, yet the process for selection is very labor intensive and always result in unstable phenotypes.

With the advances of our understanding of natural product biosynthetic pathways and the development of genetics toolbox, the strain improvement was performed in a rational fashion. Metabolic engineering therefore emerged as a powerful solution toward promoting the natural product yield by redirect metabolic flux to the synthesis of a desired secondary metabolite via genetic strategies, including increasing precursors or substrate supply, overexpression of the bottleneck enzymes, upregulation of specific metabolic pathway, reducing the shunt or other similar products by inactivating the unwanted gene or competing pathways, and reconstructing the whole pathway in engineered heterologous hosts [148].

1.4.1. Targeting and Cloning of Fungal Secondary Metabolite Biosynthetic Genes

Targeting a specific PKS or NRPS gene from its producing strain is required as the first step for the biosynthetic studies. Before the occurrence of genome sequencing, genomic DNA (cosmid) and cDNA (phage) libraries are created and screened using either Sourthern hybridization or PCR with degenerated primers in order to locate specific secondary metabolite biosynthetic genes. This approach is widely used in targeting genes in bacterial hosts, however in fungi, due to the lack of properly designed primers initially; it has not been very successful and become an effective method until recently. For example, MSAS homologue was located in an effort to target lovastatin synthases in *A. terreus* using KS domain of the *P. patulum* 6-MSAS gene as the sourthern hybridization probe [110]. Alternatively, the lovastatin synthases were

identified by the supplementation of cosmid library into mutant *A. terreus* strain that lost the production of lovastatin [150]. T-toxin PKS gene locus was identified using tagged mutation cassette that was generated using restriction enzymes [119]. In mid-1990s, with newly designed degenerated primers based on the KS domain sequence of the previously identified PKS genes, Keller and his coworkers manage to amplify the KS region of the fumonisin polyketide synthase gene *fum5* from cDNA of *G. fujikuroi* [151]. The KS domains of WA-type and MSAS-type PKSs can be amplified respectively using the two pairs of degenerated primers- LC1 and LC2c, LC3 and LC5c [152]. In addition to the KS region, a number of degenerated primers can be also designed on the basis of other domains, such as KR domain and MT domain, allowing the rapid targeting specific fungal PKS genes. For example, primers designed based on MT domain was demonstrated to be a valuable tool for targeting fungal HR-PKSs by the isolation of the PKS genes responsible for squalestatin biosynthesis [153,154].

Besides the methods described above, there are several other viable approaches in determining the region of biosynthetic genes. For instance, by screening the cDNA library built under the repression of nitrogen in the culture media, the anticancer compound bikaverin gene locus was detected in *G. fujikuroi* [155]. Moreover, the PKS gene in charge of ochratoxin A biosynthesis was isolated by a suppression subtractive PCR-base hybridization method [156].

With the advances of next generation technology, a number of fungal genomes were extensively sequenced over the last decade, allowing that the identification of biosynthetic genes do not rely on labor intensive library construction and screening. On the basis of biochemistry insights gained in elucidating PKSs or NRPSs pathways, a few bioinformatics methods emerged as the efficient alternative tools to target unknown biosynthetic genes. Given the fact that the genes conducting similar reactions share the primary sequence to some extent, different

alignment algorithms were developed to analyze the fungal genomes. Notably, an increasing number of secondary metabolites can be linked to their biosynthetic gene cluster through retro-biosynthetic analysis and match the structural features to the corresponding enzyme functions. In order to confirm the gene-compound connections as a consequence of the bioinformatics analysis, the targeted biosynthetic genes need to be either functional verified in their native hosts via genetics tools or reconstituted in heterologous hosts.

1.4.2. Functional Characterization of Fungal Natural Product Gene in the Native Hosts

The functional analysis of the biosynthetic genes in native hosts relies considerably on the transformation of heterologous DNA into fungal strains to make perturbations. Starting from the year of 1979 when genetic manipulation of *Neurospora crassa* was accomplished by Case and his coworkers [157], many fungal transformation protocols have been developed and used to manipulate the genome of the model fungal strain- *A. nidulans* [158,159]. A typical fungal transformation method consist of two processes- preparation of the protoplast and introducethe heterologous DNA into the protoplast. The former process is mainly achieved by digesting the fungal cell wall using lytic enzymes to give protoplasts with permeable cell membranes, while the heterologous DNAs were introduced via polyethylene glycol (PEG)- mediated method in the presence of calcium chloride. Later, electroporation of germinating conidia have been reported in multiple fungal hosts, including *A. nidulans* [160], *A. oryzae* [161] and *A. fumigatus* [162]. Because not all the fungal strains can be transformed by the above-described methods, an alternative transformation approach was developed on the basis of *Agrobacterium tumefaciens*, which is capable of injecting its own DNA into plants and fungi in nature [163]. The selection of transformants was accomplished by the use of selective markers. Some selection markers were

derived from one type critical nutrient biosynthetic genes, in the presence of which could restore the growth of auxotrophic fungal strain under the nutrient limited cultural conditions. Most of the markers are antibiotic resistence against hygromycin, phelomycin and sulfonyurea.

After the transformation of heterologous DNAs, those DNAs can be integrated onto the genome of the fungal host via homologous recombination. However, the integration frequency varies among different fungal strains dependent on the length of the homologous regions [164]. Furthermore, the recombination specificity is not guaranteed in most cases; as a result, the heterologous DNA can be incorporated either onto the target locus or ectopically in other regions of the genome. The mis-integration usually led to the disruption of other functional genes of the fungal host and greatly decreased the stability of the transformants. In this context, a typical designed inactivation cassettes including a marker gene in the middle flanking with large regions of homologous regions in both sides. Sometimes, a bi-part knock-out cassette was used in order to increase the efficiency of targeting [165]. With the assistance of these methods, the function of a PKS gene (*pks12*) from *F. graminearum* was functionally confirmed to synthesize the red pigment rubiofusarin [166]. Complementations of the mutant strains are required to further confirm the obtained phenotypes. Many of the methods have been applied to functionally characterize PKS biosynthetic pahways in the context of a variety of fungal hosts, including aflotoxin from *A. nidulans* , lovastatin from *A. terreus* [115], DHN-melanin in *G. lozoyensis* [167].

In addition to the gene disruption, the inactivation of target gene can also be achieved via RNA-silencing strategy. This method utilized the phenomenon of posttranscriptional RNA interfering, in which double-strand RNAs trigger their degradations. The prove-of-concept RNA-silencing experiments were successfully conducted in *Magnaporthe oryzae* by Nakayashiki and

his coworkers [168]. In the year of 2007, Hertweck and his coworkers used the same strategy to inactivate the PKS-NRPS gene that is responsible for the biosynthesis of chaetoglobosin A in *P. expansum* [138].

1.4.3. Heterologous Expression of Fungal Polyketide Synthases and Nonribosomal Peptide Synthetases

Although molecular genetic studies along with the current fungal genome sequencing projects in the last few decades have revealed a pyramid of PKS and NRPS genes, biochemical studies on PKSs and NRPSs have been impeded for years due to the low abundance and instability in their native hosts [164,169]. Functional expression of these genes in heterologous workhorses started draws many researchers' attention and has been developed as a viable tool for elucidating the biochemistry and producing medicinally important fungal secondary metabolites. Many fungi host are considered as good hosts as they can thrive on cheap nutrients in a fast growing pace. Indeed, *A. oryzae* host have been developed as an expression system for active expression of the HR-PKS- squalestatin S1 synthase. The transcription of the gene was greatly improved under the starch inducible a-amylase promoter promoter. Later, the NR-PKS PKSN was reported to be functional characterized to synthsize alternapyrone by the same expression systems [52]. In addition to the examples of HR- and NR-PKSs, Cox and his coworkers successfully reconstitute the four-gene tenellin pathway including the PKS-NRPS TenS, in *A. oryzae* [45].

The main difficulty of express fungal enzymes in bacterial hosts or yeast lies in the fact that mRNAs have undergo a splicing step to gain the ability of produce function protein. This make fungal hosts an attractive option, as fungal hosts normally shared the same sets of the

splicing enzymes. However, proper cloning strategies can overcome the difficulty and remove the intron regions at the DNA level, ensuring the function expression in *E. coli* and yeast. The first step of cloning a fungal biosynthetic gene is identifying the intron regions. A few algorithms, such as HMM based gene structure prediction, have been developed to greatly facilitate the annotation step. However, due to the limited resources of characterized PKS and NRPS genes in the database, those softwares are usually not well-trained and researchers have to analyze the gene sequence by themselves in most cases. After the annotation of the gene, the exons can be joined together via Splicing by overlap extension PCR or utilize recombination mechanism to connect them in yeast cells or in vitro. Based on the appropriately cloned genes, heterologous expression of fungal megasynthases can be fulfilled in *E. coli* and yeast [170].

E. coli has been previously demonstrated to express the large megasynthases from bacterial type I modular PKSs, such as erythromycin [171,172] and epothilone [173,174] PKSs to fungal type I PKSs, such as gzPKS13 and gfPKS4. Yeast are another attractive hosts for the heterologously expressed proteins[175], in that, unlike the prokaryotic systems, their eukaryotic subcellular organization enables them to perform many of the post-translational folding, processing and modification events which are essential for the bioactivities of the heterologous proteins. Furthermore, relatively rapid growth and ease of genetic manipulation enhance their ability to be a heterologous platform. As one of the already well-characterized engineered yeast, *Saccharomyces cerevisiae* [176], was widely used in lab researches. First reported in 1981[177], *S. cerevisiae* was demonstrated as a well-functioned host to successfully express a number of fungal PKS megasynthases. Examples include MSAS expression, and PKSN expressions. *S. cerevisiae* was demonstated to successfully reconstitute the activity of a highly unreduced

polyketide synthase HpmHRPKS and a nonreduced polyketide synthase HpmNRPKS from *Hypomyces subiculosus* [178].

1.4.4. Genome Mining Strategies for Novel Fungal Natural Products

Filamentous fungi are prolific producers of a large variety of natural products and have a strong record in producing medicinally important drugs, including the well-known β-lactam antibiotics penicillin and cephalosporin, cyclic peptide immunosuppressant cyclosporine and cholesterol-lowering agent lovastatin. However, these microorganisms are widely considered to be underachievers in natural product biosynthesis. The recent fungal genomes sequencing projects unveiled that filamentous fungal genomes harbor massive number of natural product biosynthetic genes that remain silent under standard laboratory conditions [179,180]. Table 5 shows a list of fungal strains that have been recently sequenced. Careful annotation of each of the genome revealed more than PKS genes and several PKS-NRPS genes, with less than 30% of them have been characterized. The appearances of these cryptic biosynthetic genes suggest that many useful natural products are undiscovered from conventional natural product isolations. Exploring these untapped natural product pathways is therefore an important objective towards discovery of new bioactive molecules and novel enzymatic machineries [181]. Over the last decade, various analytical and genomic approaches (genome mining strategies) have been developed and employed to intentionally activate these cryptic pathways in fungi [182].

Table 5. A sub-list of Sequenced Ascomycetes till 2010.

Organisms	***Genome size (Mb)***	***Sequence Source***	***PKS (characterized)***	***PKS-NRPS hybrid (characterized)***
Aspergillus clavatus NRRL 1	*31.0*	TIGR	13 (1)	4 (1)
Aspergillus nidulans FGSC A4	*30.0*	Broad-MIT	26 (15)	1 (1)
Aspergillus niger	*33.9*	DOE-JGI	15 (4)	5 (0)

CBS 513.88				
Aspergillus oryzae RIB40	*37.0*	CADRE, NITE, AIST	30 (3)	3 (1)
Aspergillus terreus NIH 2624	*29.3*	Broad-MIT	28 (5)	3 (1)
Aspergillus flavus NRRL 3357	*36.3*	TIGR	13 (1)	4 (1)
Aspergillus fumigatus Af293	*28.8*	TIGR	26 (15)	1 (1)
Botrytis cinerea B05.10	*43.0*	Broad-MIT	17	5
Chaetomium globosum CBS 148.57	*34.9*	Broad-MIT	20	5
Coccidioides imitis H538.4	*28.8*	Broad-MIT	10	1
Cochliobolus heterostrophus C5	*34.9*	DOE-JGI	23	2
Fusarium graminearum PH-1	*36.3*	Broad-MIT	13	2

Generally, the genome mining strategies can be classified into three categories: 1) changing or creating different growth environment, 2) by-passing the existing native regulatory systems, 3) eliminating competing metabolic pathways, 4) complete reconstitution in heterologous hosts. Changing the environmental pressure can be simply achieved by systematically altering the easily accessible cultivation parameter, including media composition, aeration, culture vessel, addition of enzyme inhibitors and so on. For example, a systematic variation of the culture conditions led to the discovery of the aspoquinolones A-D [183]. Addition of inducers was also effective in activating silent fungal pathways. Moreover Fungi naturally grow in microbial communities and secondary metabolites are produced as the chemical signals or defending factors. Specific fungal secondary metabolites are very likely to be synthesized as a result of exposure to other types of fungi or bacteria. Induced expression of silent gene clusters in *A. nidulans*, including orsellinic acid, F-9775A and F-9775B, were observed through the interaction of a set of actinomycetes that share the similar habitat [93]. Based on the advanced molecular biology, Hoffmeister and Keller were able to identify

terequinone gene cluster in *A. nidulans* through altering the expression of the global regulator LaeA [184]. Furthermore, the modification of the chromatin in *A. nidulans* by demethylation of histone led to the activation of a cryptic biosynthetic pathway for emodin and led to the unexpected production of monodictyphenone and a number of derivatives of emodin [185]. Discovery of fungal pathway-specific regulators, such as $Zn(II)_2Cys_6$ allow researchers to by-pass the global regulatory pathways and specifically activating one or two fungal secondary metabolite pathway via rational design. This was exemplified by the production of aspyridones via the overexpression of ApdR in *A. nidulans* [186]. Change the weak naturally occurred promoter with engineered consecutive promoter can greatly elevate the expression level of target genes. For instances, exchange the native promoter of AfoR with GpdA allow Wang and his coworkers discover asperfuranone from *A. nidulans* [187]. Although many genome mining approaches have been designed and prove to be effective in activating one or two pathways, yet most of them still remain in the proof-of-concept stage and more bio-activity assay need to be employed in order to functionally screen the newly mined natural products.

2. Results and Disscussion

2.1. Investigation of Radicicol Biosynthesis via Heterologous Synthesis of Intermediates and Analogs[1]

2.1.1. Introduction

Fungal polyketides represent an important family of natural products that display a wide range of biological activities [188,189]. One structurally distinct group of fungal polyketides is the resorcylic acid lactones (RALs) (Fig. 13A), of which several members have clinically relevant bioactivities [190,191]. A well-known RAL is radicicol (**1**) (Fig. 13B), which is a nanomolar (IC_{50}=20 nM) inhibitor of the heat shock protein 90 (Hsp90) [192,193]. Hsp90 chaperones the maturation of a wide range of oncogenic proteins [194], and is therefore an attractive target for anticancer drug development. Radicicol inhibits the ATPase activity of Hsp90 via competitivel binding to the ADP/ATP binding pocket, leading to the inactivation of Hsp90 chaperoning ability [195]. Despite the highly potent activity, radicicol has not been developed as a drug due to its poor activity in vivo. Radicicol can be readily inactivated through attack at the strained C7'-C8' epoxide, as well as facile Michael addition at C6' facilitated by the conjugated dienone [196]. To overcome these limitations, chemically derived radicicol analogues that do not contain these labile moieties have been pursued [197-200]. Recently, C2' oxime derivatives of *R*-monocillin II (**6**) and pochonin D (**11**) screened from a chemically synthesized library showed greatly enhanced in vivo activity [201]. A phosphate prodrug strategy was developed based on a lead compound to further increase the oral bioavailability [202]. These promising results suggest that the radicicol-based RAL scaffold remains an attractive starting point for development of Hsp90 inhibitors, and prompted us to examine the biosynthesis of the compound in detail.

[1] Compounds are numbered independently in this section.

Like other RALs studied to date, such as hypothemycin (**2**) [203] and zearalenone (**4**) [204,205], the carbon scaffold of radicicol is synthesized by the collaborative functions of two type I iterative polyketide synthases (IPKSs). IPKSs are megasynthases in which linearly juxtaposed catalytic domains function in an iterative and a highly programmed fashion [206]. Genetic knockout experiments in the two different radicicol producing fungi, *Pochonia chlamydosporia* and *Chaetomium chiversii* [203,207], confirmed that a highly-reducing PKS (HRPKS) and a nonreducing PKS (NRPKS) are involved in the biosynthesis of radicicol. The domain structures and putative functions of the two *P. chlamydosporia* PKSs are shown in Fig. 13B. The HRPKS Rdc5 contains the following domains: ketosynthase (KS) that performs the decarboxylative condensation; malonyl-CoA:ACP transacylase (MAT) that selects the building block malonyl-CoA; and acyl carrier protein (ACP) that serves as the tether of the growing polyketide via its phosphopantetheinyl arm. It also has the complete ensemble of β-keto reductive domains, which include ketoreductase (KR); dehydratase (DH); and enoyl reductase (ER). Via iterative condensation and selective combinations of β-keto reduction, Rdc5 is proposed to synthesize the reduced portion of the radicicol scaffold. In place of the reductive domains, the NRPKS Rdc1 contains a *N*-terminus starter-unit:ACP transacylase (SAT) [208] that transfers the completed reduced polyketide from Rdc5 to Rdc1; a product template (PT) domain [89] putatively involved in the cyclization of the completed nonaketide to yield the resorcylate core; and a *C*-terminus thioesterase (TE) domain that performs the macrolactonization to release the RAL product (Fig. 13B).

Although the assignment of Rdc5 and Rdc1 to synthesize the two chemically distinct portions of radicicol parallels the "bi-module" strategy employed by the hypothemycin and zearalenone biosynthetic pathways, the proposed *rdc* pathway shown in Fig. 13B contains

several unique features not present in other RAL pathways. Four of these intriguing mechanisms, which are highlighted in Fig. 13B in shaded boxes, in turn lead to unique structural features of radicicol:

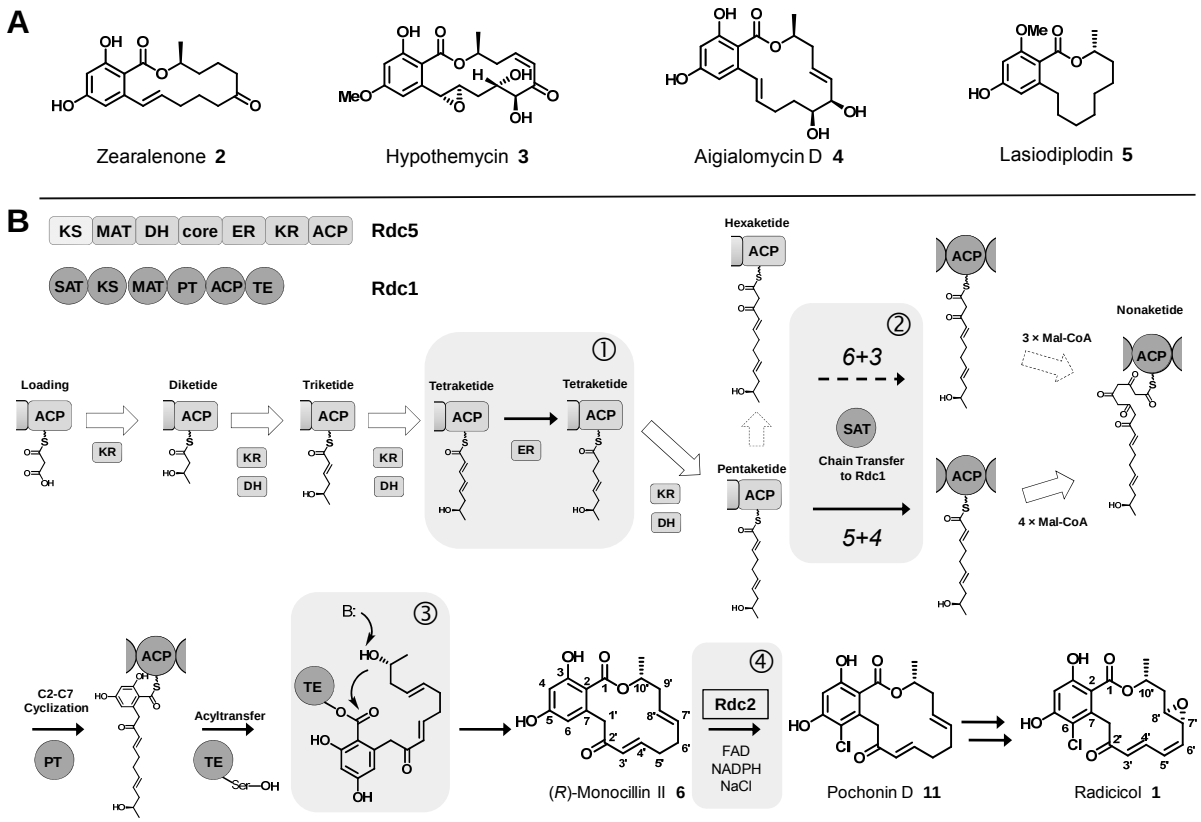


Figure 13. Resorcylic acid lactones. (**A**) The structures of RAL natural prodcuts; (**B**) The proposed biosynthetic pathway of *R*-monocillin II **6** by Rdc5 (HRPKS) and Rdc1 (NRPKS) as well as pochonin D (**6**). **6** is an early intermediate in the biosynthetic pathway of the natural product radicicol. The four featured biosynthetic steps are highlighted in shaded box.

1) <u>The C2'-C6' dienone.</u> The three-dimensional structure of radicicol is largely determined by the *trans-cis* dienone moiety and the adjacent C7'-C8' epoxide [209]. Since the epoxide is probably produced by oxidation of a double bond, it is possible that the reduced portion of the radicicol frame may be derived from a trienone. This then would suggest that the ER domain in Rdc5 may not be active during the iterative process and may instead serve as a C5'-C6' isomerase. Alternatively, the ER may function once at the tetraketide stage (Fig. 13B, Box 1) to

afford the C5'-C6'single bond and lead to the synthesis of a possible intermediate **6**, which is then desaturated during post-PKS modifications to afford the trienone.

2) The C2' ketone. This is a unique structural feature of radicicol compared to other RALs, which has been explored as a reactive handle for oximation reactions. Two possible pathways leading to the installation of the C2' ketone has been proposed (Fig. 13B, Box 2). In the first case, the ketone is synthesized by Rdc5 in the form of a β-keto group in a hexaketide intermediate, which is elongated by Rdc1 by three additional ketides (6+3 combination). This distribution of ketides between the two PKSs is identical to that of some other RALs [205,210,211]. Alternatively, Rdc5 could synthesize a pentaketide, which is transferred by the SAT domain to Rdc1 followed by chain extension of four additional ketides (5+4 combination) to complete the nonaketide. The C2' ketone would be synthesized by Rdc1 in this case.

3) The C10' *R*-OH. The *R*-stereochemistry of the ester-forming, terminal hydroxyl group in radicicol differs from the *S* configuration present in hypothemycin or zearalenone. While the differences in stereochemistry between these structurally related RALs is most likely the result of the stereospecific β-reduction by the KR domain, it remains unknown if the macrolactonizing TE domain of Rdc1 has evolved to be stereoselective towards the terminal *R*-hydroxyl group (Fig 13B, Box 3).

4) C6'-Chlorination. One of the key tailoring steps in radicicol biosynthesis is chlorination of the resorcylate core by a putative chlorinase Rdc2 (Fig. 13B, Box 4). Knockout of the Rdc2 homolog in *C. chiversii* resulted in the accumulation of the non-halogenated version of radicicol, monocillin I [207]. Rdc2 shares sequence homology to a variety of halogenases found in bacteria [212], fungi [75,207] and protozoa [213]. Structure-activity relationship studies have revealed the chlorine contributes to the potent activity of radicicol [214]. The presence of halide may

attenuate the electron density on the aromatic ring and stabilize the compound. The chlorination step has been proposed to take place immediately following the carbon scaffold synthesis; however the function of Rdc2 has not been verified. Understanding the substrate specificity of Rdc2 may therefore lead to the synthesis of other chlorinated RALs.

Comprehensive understanding of these highlighted steps of the *rdc* pathway is therefore an important goal towards the synthesis of radicicol analogs. In this work, we express Rdc5 and Rdc1 using an engineered *Saccharomyces cerevisiae* strain and completely reconstitute the activities of the two PKSs to demonstrate that **6** is the initial IPKS product. Using a combination of heterologous pathway reconstitution, precursor-directed biosynthesis and domain dissection, we provide insights into the unique functions of enzymes in the *rdc* pathway.

2.1.2. Methods and Materials

2.1.2. 1. Strains and General Techniques for DNA Manipulation.

P. chlamydosporia ATCC 16683 (previously *Verticillium chlamydosporium* var. *catenulatum*) was obtained from the American Type Culture Collection. *P. chlamydosporia* genomic DNA was prepared using the ZYMO (Orange, CA) ZR fungal/bacterial DNA kit according to supplied protocols. Total RNA was isolated from *P. chlamydosporia* using ZYMO ZR Fungal/Bacterial RNA MiniPrep™ kit. RT-PCR ImProm-II™ Reverse Transcription System kit was used for RT-PCR. The gene-specific primers are listed in Table 6. *E. coli* XL1-Blue (Stratagene) and *E. coli* TOPO10 (Invitrogen) were used for cloning. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs). PCR was performed using Platinum Pfx DNA polymerase (Invitrogen). Sequences of PCR products were confirmed by DNA sequencing (Laragen, CA). *E. coli* BL21(DE3) (Novagen) was used for protein

expression. *Saccharomyces cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*) was used as expression host [215,216].

Table 6. Sequences of primers used in the construction of plasmids in section 2.1.

No.	Primers	Sequences[a]
P1	Rdc1-f	5'-AA*CAT**ATG***AAGTCGCATACTTCGGC-3'
P2	Rdc1-r	5'-TT*GTTTAAAC*TCAGTGGTGGTGGTGGTGGTGTCGTCCAAAGTGCTCAAACGCC-3'
P3	Rdc5-f	5'-AA*GCTAGC***ATG**CCCTCCGCAACTGCCCA-3'
P4	Rdc5-E1-r	5'-TT*AGGCCT*TTGGTGTATTGCCAAGGAGATAAACTGCATGAGCTTATCGTCCGAAGAAG-3'
P5	Rdc5-E2-f	5'-TGGCAATACACCAA*AGGCCT*AATCTCGAGA-3'
P6	Rdc5-r	5'-TT*GTTTAAAC*TCATGTGGTGGTGGTGGTGGTGAGAACCTCCCACCAACT-3'
P7	Rdc1-flag-f	5'- AT*ACTAGT***ATG**AAGTCGCATACTTCGGC-3'
P8	Rdc1-flag-r	5'- TT*ATTTAAAT*TCATCGTCCAAAGTGCTCAA-3'
P9	Rdc5-flag-f	5'- AT*ACTAGT***ATG**CCCTCCGCAACTGCCCA-3'
P10	Rdc5-flag-r	5'- TT*ATTTAAAT*TCAAGAACCTCCCACCAACT-3'
P11	Rdc1ΔTE-r	5'- TT*CACGTG*TTAGTGGTGGTGGTGGTGGTGAGACCTGGGTTGCTGAGGTG-3'
P12	Rdc1TE0-f	5'-GGAGGGTTTGCAGGAGGCGCCATGATTG*CATATG*AAATTGCGCGCCAGCTATCCG-3'
P13	Rdc1TE0-r	5'- CGGATAGCTGGCGCGCAATTT*CATATG*CAATCATGGCGCCTCCTGCAAACCCTCC-3'
P14	Rdc2-f	5'-AC*CAT**ATG***TCGGTACCCAAGTCTTG-3'
P15	Rdc2-r	5'- TGGTTTAAACAACTTTGTTGAGGCCAAGTTTGC-3'

[a] The introduced restriction sites are shown in italics. The start codons are in bold.

2.1.2.2. Spectroscopic Analyses.

Nuclear magnetic resonance (NMR) spectra were obtained on Varian Inova 500 MHz, 600 MHz, 700 MHz (cryoprobe) instruments and a Bruker 500 MHz spectrometer. ^{1}H NMR chemical shifts are reported in parts per million (ppm) using the residual proton resonance of solvents as reference: $CDCl_3$ δ 7.26, CD_3OD δ 3.30 and CD_3COCD_3 δ 2.09. ^{13}C NMR chemical shifts are reported relative to $CDCl_3$ δ 77.0, CD_3OD δ 49.0 and CD_3COCD_3 δ 29.9. LC-MS was conducted with a Shimadzu 2010 EV liquid chromatography mass spectrometer by using both positive and negative electrospray ionization, and a Phenomenex Luna 5 μm 2.0 x 100 mm C18 reverse phase column. Samples were separated on a linear gradient of 5 to 95% CH_3CN (*v/v*) in H_2O supplemented with 0.05% (*v/v*) formic acid at a flow rate of 0.1 mL/min. Chiral HPLC analysis was performed using a Lux 3 μm Cellulose-1

column (4.60 × 250 mm, Phenomenex) with an isocratic condition of 25% isopropanol in n-hexane (*v/v*) and a flow rate of 0.75 mL/min.

2.1.2.3. Construction of Plasmids.

Genomic DNA from *P. chlamydosporia* was used as the template for PCR amplifications. The two exons of *rdc5* gene were amplified and ligated into the 2μ-based yeast-*E. coli* shuttle vector YEpADH2p with *URA3* marker to give pKJ61. The gene encoding Rdc1 was amplified and ligated into YEpADH2p with *TRP1* marker to give pKJ91. An engineered FLAG-tag expression vector was constructed by introducing DNA sequence 5'-CATATGGCTAGCGATTATAAGGATGATGATGATAAGACTAGTC-3' into YEpADH2p with *URA3* marker between *Nde*I and *Spe*I sites. Then, both *rdc5* and *rdc1* genes were amplified using primers pairs P7/P8 and P9/P10. The genes were digested by *Spe*I and *Swa*I, and ligated into the FLAG-tag expression vector to yield pZH223 and pZH232. Rdc1ΔTE gene was amplified by primer pair P1 and P11. The PCR product was ligated into YEpADH2p with *TRP1* marker to give pZH228. The Ser1889 to Ala mutation was introduced into pKJ91 by primer pair P12 and P13 to give pHZ252.

Table 7. Plasmid constructs and the resulting protein products in section 2.7.

Plasmid	Vector Source	Genes	Marker	Protein products	Reference
pZH78	YEpADH2p	*hpm8*	*URA3, Amp*	C-terminal hexahistidine tagged Hpm8	This work
pZH74	YEpADH2p	*hpm3*	*TRP1, Amp*	C-terminal hexahistidine tagged Hpm3	[87]
pKJ61	YEpADH2p	*rdc5*	*URA3, Amp*	Rdc5	This work
pKJ91	YEpADH2p	*rdc1*	*TRP1, Amp*	Rdc1	This work
pZH200	pET24a	*rdc2*	*Kan*	C-terminal hexahistidine tagged Rdc2	This work
pZH208	YEpADH2p	*rdc2*	*LEU2, Amp*	Rdc2	This work
pZH223	YEpADH2p	*rdc5*	*URA3, Amp*	N-terminal FLAG tagged Rdc5	This work
pZH232	YEpADH2p	*rdc1*	*URA3, Amp*	N-terminal FLAG tagged Rdc5	This work
pZH228	YEpADH2p	*rdc1ΔTE*	*TRP1, Amp*	C-terminal hexahistidine tagged Rdc1ΔTE	This work

pZH252	YEpADH2p	*rdc1_S1889A*	*TRP1, Amp*	C-terminal hexahistidine tagged Rdc1_S1889A	This work

Total RNA was isolated from *P. chlamydosporia* grown for 3 days on potato dextrose agar plate. P14 was used as a primer for RT-PCR to amplify *rdc2* cDNA and P14 paired with P15 were used or regular PCR later. The intron-free *rdc2* was ligated into pET24a vector to give *E. coli* expression plasmid pZH200 or into YEpADH2p with *LEU2* marker to yield pZH208. All the plasmids in this work are listed in Table 7.

2.1.2.4. Protein Purification and in vitro Assays.

C-terminal hexahistidine tagged Rdc1ΔTE, Rdc1_S1889A were expressed and purified from *S. cerevisiae* strain BJ5464-NpgA harboring plasmid pZH228 and pZH252, respectively. *C*-terminal hexahistidine tagged Rdc2 was expressed and purified from *E. coli* BL21(DE3) strain harboring pZH200 (see supplemental data for details). *N*-terminal FLAG tagged Rdc1 and Rdc5 were expressed and purified from *S. cerevisiae* strain BJ5464-NpgA harboring plasmid pZH232 and pZH223, respectively. For 1 L of yeast culture, the cells were grown at 25°C in YPD media with 1% dextrose for 72 hours. The cells were harvested by centrifugation (3500 rpm, 15 minutes, 4°C), resuspended in 20 mL lysis buffer (50 mM NaH_2PO_4, 0.15 M NaCl, 10 mM imidazole, pH=8.0) and lysed by sonication on ice. Cellular debris was removed by centrifugation (15000 g, 1 hour, 4°C). The supernatant was loaded onto a gravity flow column containing 3 mL anti-FLAG M2 affinity resin and washed with 10 column volumes TBS Buffer (150 mM Tris-HCl, 50 mM NaCl, pH=7.4). Purified proteins were eluted with five column volumes of TBS Buffer with 100 mg/mL FLAG peptide. Purified proteins were concentrated and buffer exchanged into Buffer A (50 mM Tris-HCl, 50 mM NaCl, pH=8.0) +10% glycerol,

concentrated, aliquoted and flash frozen. Protein concentrations were determined using the Bradford dye-binding assay (Biorad).

For in vitro assays, 10 μM of Rdc1 and 10 μM of Rdc5 were incubated with 2 mM of NADPH and malonyl-CoA. In the assays for Rdc1, the chemically synthesized starter units were added to 2 mM final concentration. For halogenation assays with Rdc2, 1 mM RAL substrate was incubated with 50 μM Rdc2, 10 μM SsuE, 100 μM FAD, 2 mM NADPH and 50 mM NaCl. SsuE was used as a FAD reductase, which is cloned and expressed as reported [217]. All the reactions were quenched and extracted twice with 99% ethyl acetate (EtOAc)/1% acetic acid (AcOH). The resultant organic extracts were evaporated to dryness, re-dissolved in methanol, and then analyzed by LC-MS.

2.1.2.5. Isolation of (R)-monocillin II (6), Isocoumarin (8) and (S)-monocillin II (10).

To purify **6** for structural analysis, *S. cerevisiae* harboring pKJ61 and pKJ91 plasmids was cultured in YPD liquid media (4 L) for 4 days. The pH of culture supernatant was adjusted to 5.0 and then extracted three times with equal volume of EtOAc. The resultant organic extracts were combined and evaporated to dryness, redissolved in methanol, and purified by reverse-phase HPLC (XTerra Prep MS C18 5 μm, 19 mm × 50 mm) on a linear gradient of 5 to 95% CH_3CN (*v/v*) over 15 min and 95% CH_3CN (*v/v*) further for 15 min in H_2O supplemented with 0.1% (*v/v*) trifluoroacetic acid at a flow rate of 2.5 mL/min. The eluent was extracted with EtOAc, and dried *in vacuo* to give pure solid **6** (approximate yield of 15 mg/L). To purify **8,** *S. cerevisiae*/pKJ61+pHZ228 was cultured in YPD (20 L) media for 4 days followed by the same purification steps as above. To produce **10**, *S. cerevisiae* strain harboring pKJ91 were cultured in YPD media (400 mL) for 2 days. The cell culture was supplemented with 100 mg/L (*R,2E,6E*)-9-

hydroxydeca-2,6-dienoyl-N-acetyl cysteamine (**9**) and continue growing for another day. Compound **10** was purified as described above for **6** with an approximate yield of 1.3 mg/L.

2.1.2.6. Isolation of Pochonin D (11) and 6-chloro, 7', 8'-dehydrozearalenol (Cl-DHZ) (14).

S. cerevisiae/pKJ61+pKJ91+pHZ208 was cultured in YPD media (4 L) for 4 days and **11** was purified to a final yield of 14.3 mg/L. *S. cerevisiae* strain harboring pKJ61+pHZ74+pHZ208 plasmids were cultured in YPD media (4 L) for 7 days and **14** was purified to a final yield of 9 mg/L.

2.1.3. Results

2.1.3.1. Heterologous Reconstitution of the Enzymatic Activities of Rdc5 and Rdc1.

Although the *rdc* gene cluster has been genetically verified in two different fungal hosts, the biochemical characterization of the key IPKSs has not been performed and the earliest RAL intermediate has not been conclusively identified. As a starting point to study the unique features of the *rdc* pathway, we aimed to reconstitute the functions of Rdc5 and Rdc1 in vitro and in the heterologous host *S. cerevisiae*. To do so, yeast 2μ expression plasmids encoding Rdc1 and Rdc5 under the *ADH2* promoter were separately transformed into *S. cerevisiae* BJ5464-NpgA. The use of this vacuolar protease-deficient yeast strain with genome-integrated NpgA [218] enables heterologous expression of intact and phosphopantetheinylated PKSs [215]. Both megasynthases were solubly expressed and were each purified to single-band purity by using anti-FLAG affinity chromatography (Fig. 14).

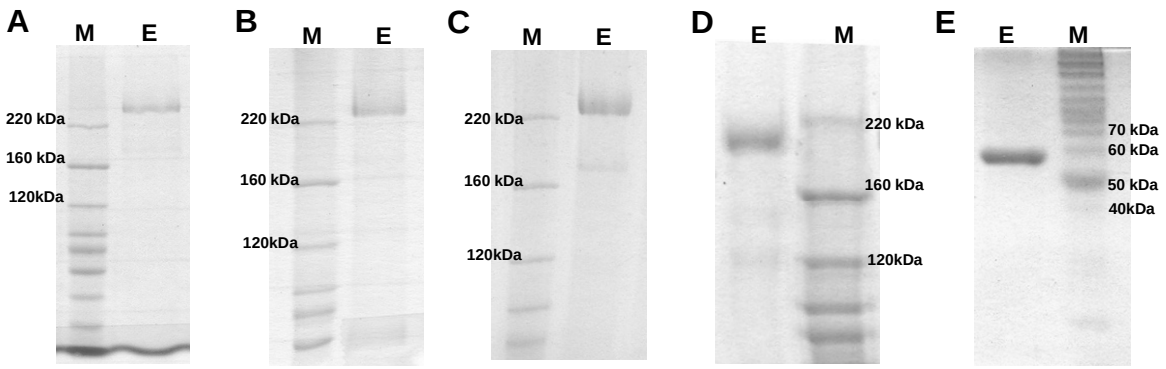


Figure 14. SDS-PAGE gel of proteins purified from *S. cerevisiae* and *E. coli* BL21(DE3). (A) 6% SDS-PAGE gel of N-terminal FLAG tagged Rdc5 (260kDa) purified from *S. cerevisiae* after FLAG affinity chromatography. (B) 6% SDS-PAGE gel of N-terminal FLAG tagged Rdc1 (229kDa) from *S. cerevisiae* after FLAG affinity chromatography. (C) 6% SDS-PAGE gel of C-terminal hexahistidine tagged Rdc1_S1889A (229kDa) purified from *S. cerevisiae* after Ni-NTA chromatography. (D) 6% SDS-PAGE gel of C-terminal hexahistidine tagged Rdc1ΔTE (195kDa) purified from *S. cerevisiae* after Ni-NTA chromatography. (E) 12% SDS-PAGE gel of C-terminal hexahistidine tagged Rdc2 (58kDa) purified from *E. coli* BL21(DE3) after Ni-NTA chromatography. In all the gels, lane M: Invitrogen Benchmark Protein Ladder, Lane E: Final elution of protein after chromatography.

When the purified Rdc5 and Rdc1 were incubated with 2 mM malonyl-CoA and NADPH, a major compound emerged with a mass of 316, which is identical to that of **6**. The UV absorbance of the compound is characteristic of resorcylate chromophore with λ_{max} at 217, 260 and 301 nm. In order to determine the structure of this RAL product, two plasmids harboring *rdc1* and *rdc5* with different auxotrophic selection markers were co-transformed into *S. cerevisiae* BJ5464-NpgA. After three days of culturing in YPD media, both the culture broth and cell pellet were extracted with organic solvent and analyzed with LC-MS. Compared to BJ5464-NpgA expressing either Rdc1 or Rdc5 individually, the co-expression strain synthesized the identical *m/z* 316 compound as a predominant product at a titer of 15 mg/L (Figure 15).

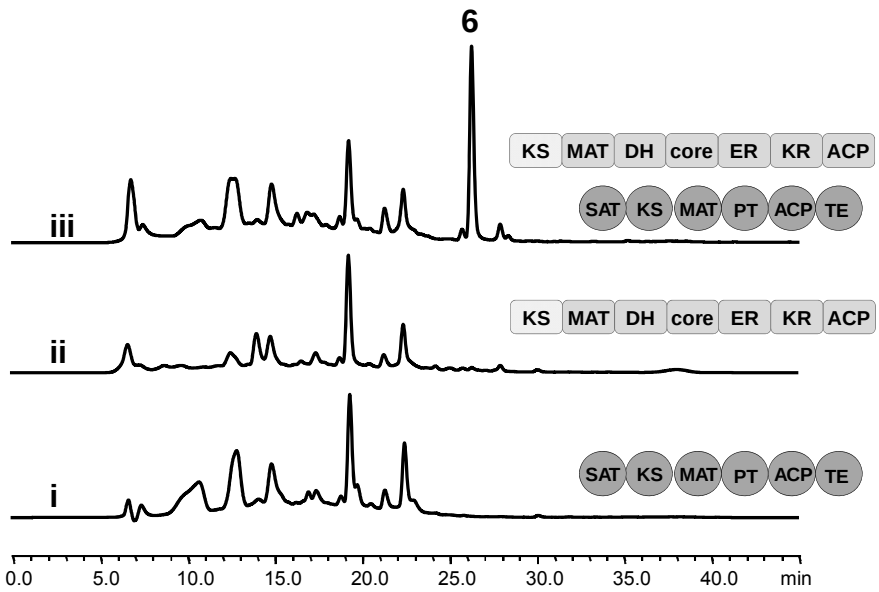
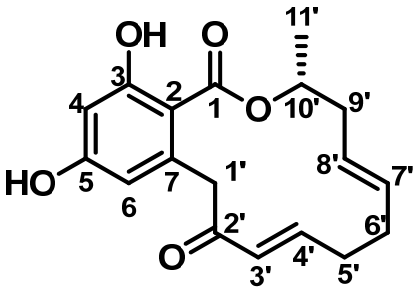


Figure 15. In vivo reconstitution of biosynthesis of *R*-monocillin II (**6**). LC-MS profiles of organic extract from (i) *S. cerevisiae* strain BJ5464-NpgA/pKJ91 expressing Rdc1; (ii) *S. cerevisiae* strain BJ5464-NpgA/pKJ61 expressing Rdc5; and (iii) *S. cerevisiae* strain BJ5464-NpgA/pKJ91+pKJ61 co-expressing Rdc5 and Rdc1. All traces are monitored at 300 nm.

The new compound was then purified and analyzed by extensive NMR characterization. The ^{13}C spectrum showed 18 carbon signals, indicative of a nonaketide backbone. One carbon signal at δ 196.5 ppm suggested the existence of an aliphatic ketone. Combining all spectroscopic information, we assigned the structure of this RAL as that of **6**, which has been isolated from different radicicol producing strains (Table 8) [207,219,220]. To confirm the structure of **6** and especially to verify the *R* stereochemistry of the lactone at C1, an authentic standard of **6** was chemically synthesized (Table 8). The UV pattern, retention time on chiral column, and all NMR data of **6** are all identical to that of the standard.

The enzymatic synthesis of **6** by purified Rdc1 and Rdc5 conclusively indicates **6** is the first RAL compound in the *rdc* pathway, and is the product directly offloaded from Rdc1 as shown in Fig. 13B. Interestingly, a fully reduced ketide unit at C5'-C6' is found in **6** instead of the possible trienone structure, which indicates that the ER domain in Rdc5 is indeed functional

Table 8 NMR Data for **6** [a]



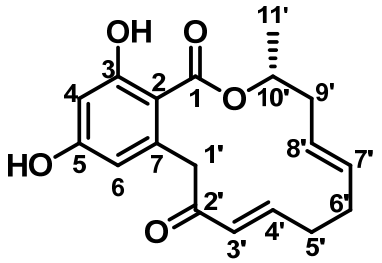
No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	171.1	-	
2	106.4	-	
3	166.2	-	
4	102.9	6.33 (d, 1H, 2.5)	C2, C3, C5, C6
5	163.2	-	
6	113.3	6.31 (d, 1H, 2.5)	C2, C5, C1’
7	141.3	-	
1’	49.0	4.06 (d, 1H, 16.9) 3.89 (d, 1H, 16.9)	C2, C6, C7, C2’
2’	196.5	-	
3’	131.2	5.84 (d, 1H, 15.6)	C2’, C5’
4’	147.4	6.61-6.67 (m, 1H)	C2’, C5’
5’	31.8	2.22-2.28 (m, 2H)	C4’, C6’
6’	31.7	2.13-2.27 (m, 2H)	C4’, C5’, C8’
7’	128.2	5.30-5.36 (m, 1H)	C6’, C5’
8’	132.7	5.23-5.29 (m, 1H)	
9’	37.4	2.19-2.21 (m, 1H) 2.62-2.68 (m, 1H)	C7’, C8’, C10’, C11’
10’	73.1	2.62-2.68 (m, 1H)	C1, C8’, C9’, C11’
11’	18.4	1.30 (d, 3H, 6.6)	C10’, C9’

[a] Spectra were obtained at 500 MHz for proton and 125 MHz for carbon and were recorded in CD_3COCD_3.

and can completely reduce the β-keto of the tetraketide to a methylene following the actions of KR and DH. However, the ER domain is not active during the other reductive tailoring iterations, thereby giving rise to the enone at C2’-C4’ and the *trans* double bond at C7’ and C8’ that eventually becomes epoxidized. The additional C5’-C6’*cis* double bond present in radicicol must therefore be introduced by post-PKS processing. This is consistent with the genetic studies with

C. chiversii, which suggested that C7'-C8' epoxidation and C5' hydroxylation may be catalyzed a single P450 Rdc4 (RadP), although this has not been biochemically confirmed [207].

Table 9 NMR Data for Chemical Synthesized **6** [a]



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))
1	171.0	-
2	106.4	-
3	166.1	-
4	113.3	6.32 (d, 1H, 2.43)
5	163.1	-
6	102.8	6.30 (d, 1H, 2.42)
7	141.3	-
1'	48.9	4.05 (d, 1H, 16.8) 3.88 (d, 1H, 16.8)
2'	196.5	-
3'	131.3	5.83 (d, 1H, 15.6)
4'	147.3	6.63 (dt, 1H, 15.4, 7.54)
5'	31.8	2.12-2.29 (m, 4H)
6'	31.7	2.12-2.29 (m, 4H)
7'	132.7	5.23-5.35 (m, 3H)
8'	128.2	5.23-5.35 (m, 3H)
9'	37.5	2.64 (ddd, 1H, 14.5, 8.24, 4.15) 2.33 (m, 1H)
10'	73.0	5.23-5.35 (m, 3H)
11'	18.4	1.29 (d, 3H, 6.59)

[a] Spectra were obtained at 700 MHz for proton and 175 MHz for carbon and were recorded in CD_3COCD_3.

2.1.3.2. Polyketide Chain Length Control of Rdc5 and Rdc1.

We next probed whether the starter unit of the NRPKS Rdc1 is a β-keto hexaketide or a pentaketide as shown in Fig. 13B. This would reveal the "distribution of labor" in the synthesis of **6** by the two *rdc* PKSs. To verify the feasibility of the "5+4" pathway, we tested whether chemically synthesized *N*-acetyl-cysteamine thioester (SNAC) of (*2E, 6E,* 9*R*)-9-hydroxydeca-2,6-dienoic acid (**7**) (supplemental data) can mimic the proposed pentaketide product of Rdc5 as shown in Fig. 13B. We have previously shown that KS domains of certain NRPKSs can accept specific acyl-SNAC substrates as the starter unit to prime the functions of the remaining steps [44,210,211]. We therefore reasoned that if the pentaketide is indeed the starter unit for Rdc1, the RAL product **6** should be observed in the reaction mixture. However, if Rdc1 is only programmed to perform three additional iterations as in the "6+3" model, either no product will be observed since the KS does not accept the shorter starter unit, or a 12-membered RAL compound would be synthesized.

The in vitro assay was performed by incubating purified Rdc1 with starter unit **7** and 2 mM malonyl-CoA. LC-MS analysis confirmed that indeed a product with the same *m/z* and UV absorbance spectrum as **6** was synthesized. To further demonstrate the precursor-directed biosynthesis of **6**, pentaketide thioester **7** was supplemented to the Rdc1-expression yeast culture at a concentration of 100 mg/L. After one additional day of culturing, the production of **6** was seen in the organic extracts (Fig. 16A, trace iii). When purified from the yeast culture, the identity of **6** was further confirmed by comparing to the authentic standard. Using selected ion monitoring in MS, we searched for additional metabolites in both the in vitro and in vivo extracts to determine if other RAL compounds or biosynthetic intermediates could be found. However, **6** is the only product that can be inferred to derive from **7**.

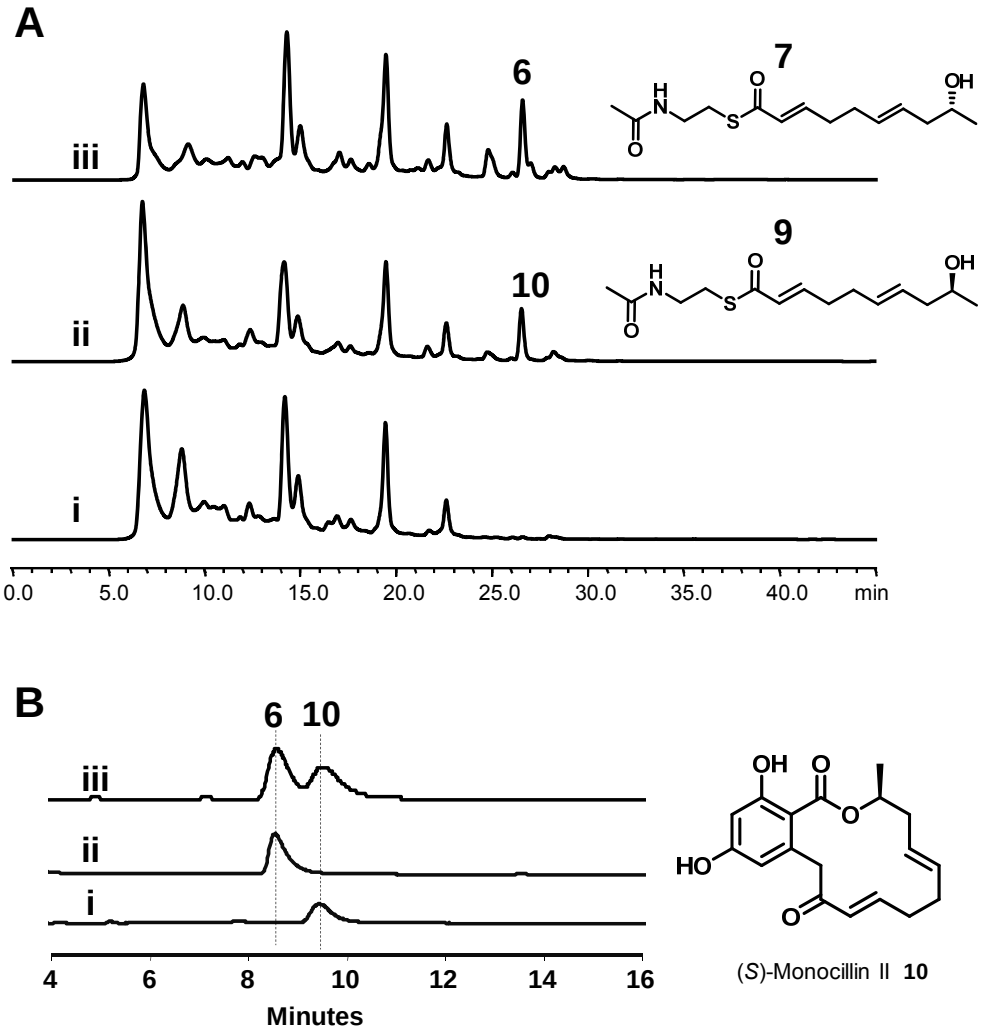


Figure 16. Precursor directed feeding using SNAC substrates. (A) LC-MS profiles of organic extracts from *S. cerevisiae* strain BJ5464-NpgA/pKJ91 expressing Rdc1 and i) without substrate feeding; ii) supplementation with starter unit **9**; iii) supplementation with starter unit **7**. (B) Chiral HPLC separation of the enantiomers 1 and 5: i) purified 5; ii) purified 1; iii) co-injection of a mixture containing purified 5 and 1. All traces are monitored at 300 nm.

The biosynthesis of **6** by Rdc1 alone in the presence of **7** demonstrated that the synthetic starter unit can replace the natural role of Rdc5 in the tandem IPKS system. Rdc1 is able to precisely elongate the acyl chain of **7** by four additional ketide units to afford the nonaketide backbone, with the first unreduced ketide being the C2' ketone in **6**. This result therefore

authenticates the “5+4” pathway shown in Fig. 13B as the most likely mechanism of **6** biosynthesis.

2.1.3.3. Functionality of Rdc1 TE Domain and polyketide chain release.

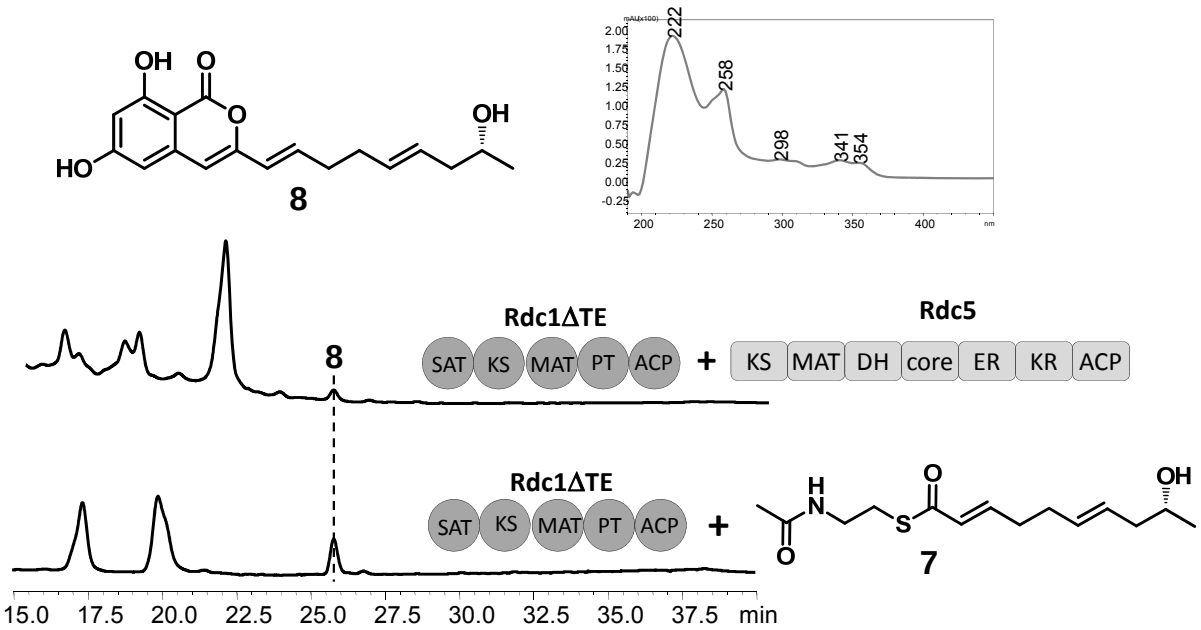
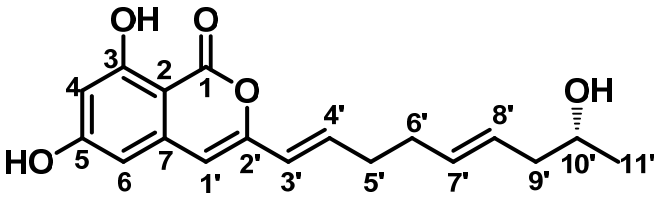


Figure 17. Reconstitution of Rdc1_S1889A in vivo and in vitro. (A) HPLC analysis (300nm) of polyketides synthesized by (i) expressing Rdc1_S1889A and Rdc5 in vivo and (ii) incubating Rdc1_S1889A with compound **7** in vitro. (B) UV spectrum and compound structure of **8**.

As demonstrated in studies on PKS13 and Hpm3 [210,211,221], the TE domains are responsible for the final macrolactonization using the secondary hydroxyl group on the highly reduced polyketide portion. This acts as a nucleophile to attack the TE-bound carboxyl ester to facilitate chain release. To verify the function of Rdc1 TE domain, the active site Ser1889 of the catalytic triad identified from sequence alignment with PKS13TE was mutated to Ala. This Rdc1_S1889A mutant was solubly expressed and purified from BJ5464-NpgA (Fig. 14). When Rdc1_S1889A was incubated with thioester **7** and malonyl-CoA, we could no longer detect the synthesis of **6**. Instead, a new compound **8** with the same mass of **6** (*m/z* of 316) was observed

(Fig. 15). Compound **8** was also present in the culture extract of BJ5464-NpgA co-expressing Rdc1_S1889A and Rdc5. However the titer of **8** is >100 fold less than that of **6** synthesized in Fig. 16. Since the starter unit **7** or Rdc5 is required for the synthesis of **8**, we inferred that the pentaketide acyl chain is part of **8**. The UV spectrum of **8** does not display the characteristic peaks observed from resorcylate-containing compounds, but is more complex and similar to isocoumarin compounds SMA76b and SMA76c produced by the NRPKS PKS4 from *Gibberella fujikuroi* [44]. To solve the structure of **8**, we scaled up the yeast culture and purified **8** for NMR characterization (Table 10). The NMR data is consistent with that of a pentaketide-primed isocoumarin **8** shown in Fig. 17.

Table 10. NMR Table for **8** [a].



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	166.2	-	
2	99.8	-	
3	164.4	-	
4	102.5	6.46 (d, 1H, 2.11)	C5, C2
5	166.4	-	
6	104.1	6.38 (d, 1H, 2.10)	C5, C2, C4
7	129.1	-	
1’	105.5	6.45 (s 1H)	C2’, C3’
2’	152.9		
3’	123.1	6.19 (dt, 1H, 15.6, 1.42)	C2’, C5’
4’	136.5	6.63 (dt, 1H, 15.6, 7.09)	C2’, C5’
5’	33.3	2.30-2.34 (m, 2H)	C3’, C4’, C6’
6’	32.6	2.18-2.23 (m, 2H)	C7’, C5’
7’	128.9	5.46-5.58 (m, 1H)	C6’, C9’
8’	131.9	5.46-5.58 (m, 1H)	C6’, C9’
9’	43.4	2.12-2.17 (m, 2H)	C10’, C8’

10’	67.5	3.71 (1H, 6.1 Hz)	
11’	23.2	1.08 (d, 3H, 6.2)	C9’, C10’

[a] ^{1}H spectrum was obtained at 600 MHz for proton and ^{13}C spectrum was obtained at 150 MHz for carbon. Both spectra were recorded in CD_3COCD_3.

To exclude any possible involvement of Rdc1 TE in the formation of **8**, a truncated Rdc1 without the TE domain (Rdc1ΔTE) was constructed. Rdc1ΔTE was similarly expressed and purified from yeast (Fig. 14). Both the in vitro assay of Rdc1ΔTE supplemented with starter unit **7** and in vivo product of Rdc1ΔTE co-expressed with Rdc5 showed that **8** was synthesized (data not shown). Our results confirm the role of the Rdc1 TE domain in catalyzing the macrocyclization reaction during the biosynthesis of **6**. When the TE activity is compromised, **8** is released as a shunt product of Rdc1 through enolization of the C2’ ketone and nucleophilic attack on the thioester carbonyl to form the benzopyrone (Fig. 18). Interestingly, we did not find any pentaketide-primed resorcylic acids in the reaction mixture, suggesting releasing by hydrolysis of the thioester linkage is limited for Rdc1. The mechanism shown in Fig. 18 may also account for the biosynthesis of structurally similar isocoumarin co-metabolites from other RAL producing fungi strains, such as paraphaeosphaerins and chaetochiversins [222], which can accumulate as a result of derailment of NRPKS TE functions.

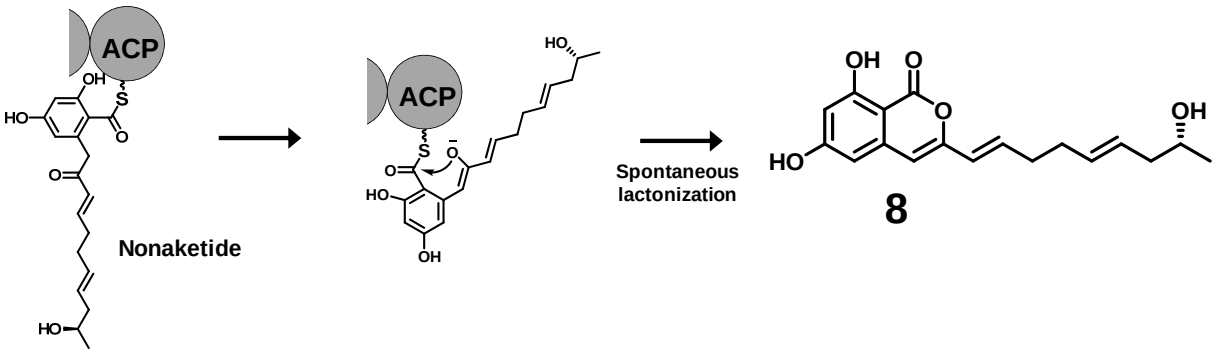
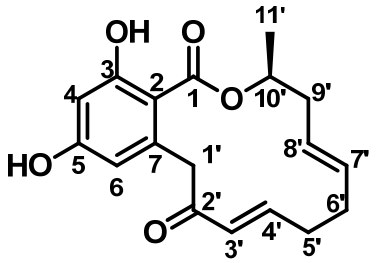


Figure 18. Proposed biosynthetic pathway of the isocoumarin **8** upon inactivation or deletion of the Rdc1 TE domain.

2.1.3.4. Probing the Stereoselectivity of the TE Domain.

Having confirmed the involvement of the Rdc1 TE domain in macrolactone formation, we then examined whether the stereochemistry of the C10' hydroxyl nucleophile can affect the recognition and cyclization of the resorcylate thioester by the TE domain (Fig. 13B, Box 3). To probe the stereochemical requirement of TE domain, we chemically synthesized (*2E, 6E, 9S*)-9-hydroxydeca-2,6-dienoyl-*S*-NAC (**9**)**,** which is the enantiomer of the usual starter unit **7**. After supplementing **9** into the Rdc1-expressing BJ5464-NpgA culture, the organic extract contained a compound that is indistinguishable from **6** based on retention time, UV absorbance and mass spectrum (Fig. 16A, trace ii). This compound was purified to homogeneity and all NMR spectra perfectly matched to that of **6** (Table 11). However, the compound can be clearly separated from **6** when co-injected onto a chiral HPLC column as shown in Fig. 16B, suggesting that it is **10**, which contains an *S*-stereocenter at C10' (Fig. 16B). Indeed, the circular dichroism spectra of **6** and **10** were mirror images of each other (Fig. 17), which is expected as **6** and **10** are enantiomers.

Table 11 NMR Data for **10** [a]



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))
1	170.9	-
2	106.4	-
3	165.8	-
4	113.2	6.33 (d, 1H, 2.58)
5	163.1	-
6	102.7	6.31 (d, 1H, 2.32)
7	141.3	-

1’	48.9	4.06 (d, 1H, 16.9) 3.88 (d, 1H, 16.8)
2’	196.4	-
3’	131.2	5.83 (d, 1H, 15.6)
4’	147.3	6.63 (dt, 1H, 15.1, 7.82)
5’	31.8	2.11-2.29 (m, 4H)
6’	31.7	2.11-2.29 (m, 4H)
7’	132.6	5.23-5.35 (m, 3H)
8’	128.2	5.23-5.35 (m, 3H)
9’	37.4	2.64 (ddd, 1H, 14.4, 8.12, 3.99) 2.33 (m, 1H)
10’	72.9	5.23-5.35 (m, 3H)
11’	18.4	1.29 (d, 3H, 6.71)

[a] Spectra were obtained at 700 MHz for proton and 175 MHz for carbon and were recorded in CD_3COCD_3.

The titer of **10** from the in vivo precursor feeding study is nearly the same as that of **6** when supplemented with **7**. Furthermore, in vitro experiments also gave similar levels of the enantiomers. Therefore, the Rdc1 TE domain appears insensitive towards the stereochemical configuration of the terminal nucleophile in the macrocyclization reaction. It is possible that the linear portion of the unlactonized molecule is flexible enough to allow the hydroxyl in either configuration to be deprotonated by the histidine that serves as the general base, which can then lead to formation of the *S* or *R* configured RAL compounds with equal efficiency. This observation thus expands the already broad substrate specificities of the RAL TE domains characterized to date and further hints at a structurally very loose cyclization chamber for this unique class of TE domains. Lastly, the biosynthesis of **10** also reveals the KS and PT domains of the Rdc1 are not specific towards the chirality of the hydroxyl group.

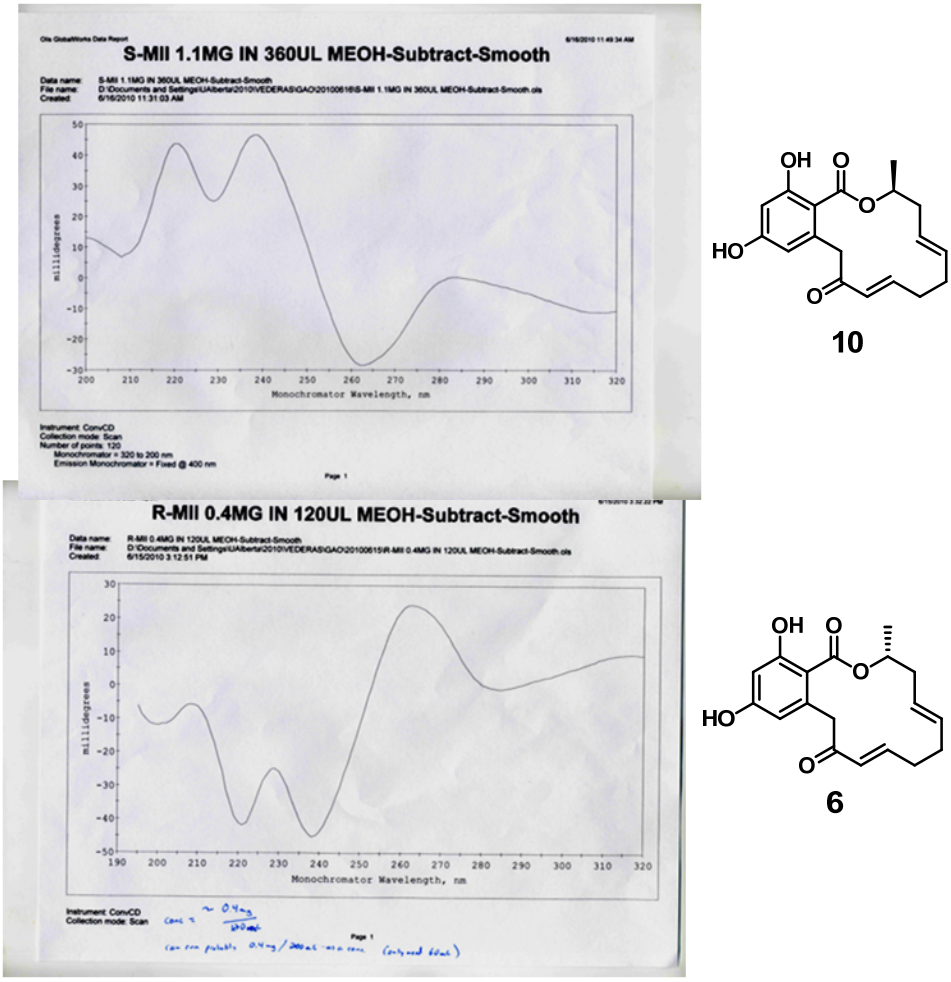


Figure 19. Circular dichroism (CD) spectra of compound **10** (up) and compound **6** (bottom).

2.1.3.5. In vivo Reconstitution of pochonin D Synthesis.

Establishment of **6** as the earliest RAL compound in the radicicol biosynthetic pathway represents the starting point to investigate the downstream post-PKS tailoring enzymes. Pochonin D (**11**), which is the chlorinated version of **6**, appears to be biosynthetically accessible through the action of the FAD-dependent halogenase Rdc2 (Fig. 13B, Box 4). Compound **11** displays good affinity towards Hsp90 and its C2' oxime derivatives are considerably more stable than radicicol. Therefore, reconstituting the biosynthesis of **11** in *S. cerevisiae* may provide a convenient route of affording this potential drug lead.

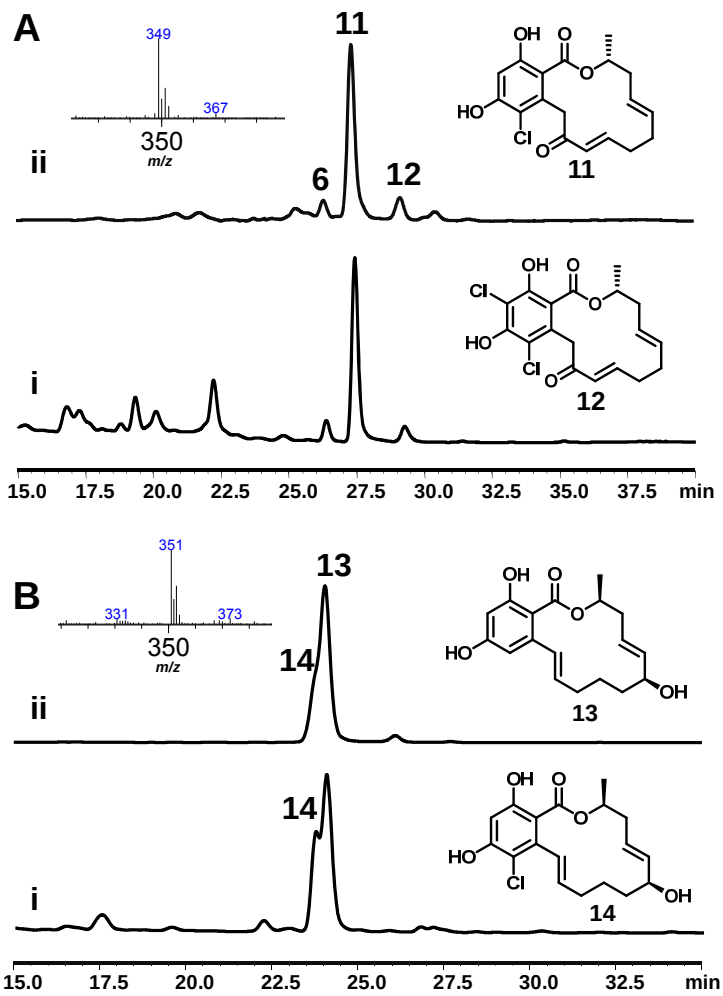
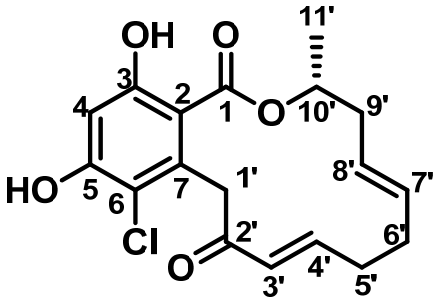


Figure 20. Biosynthesis of chlorinated RAL compounds using the halogenase Rdc2. (**A**) LC-MS profiles for production of **11** from: i) BJ5464-NpgA expressing Rdc5, Rdc1 and Rdc2; ii) in vitro assay of 50 µM Rdc2 incubated with 1 mM **6**, 100 µM FAD, 2 mM NADPH, 50 mM NaCl and 10 µM SsuE. Inset shows observed MS spectrum of **11**; (**B**) LC-MS profiles for production of **14**: i) BJ5464-NpgA expressing Hpm8, Hpm3 and Rdc2, ii) in vitro assay of 50 µM Rdc2 incubated with 1 mM **7**, 100 µM FAD, 2 mM NADPH, 50 mM NaCl and 10 µM SsuE. Inset shows observed MS spectrum of **14**. All traces are monitored at 300 nm.

The cDNA of Rdc2 was obtained by RT-PCR and inserted into a yeast expression plasmid that contains a *LEU2* selection marker, as well as into an *E. coli* pET28a expression vector. When the BJ5464-NpgA host is transformed with three expression plasmids separately encoding Rdc1, Rdc2 and Rdc5, we observed the emergence of a major RAL product from the third day culture extract (Fig. 20A, trace i). While a small amount of **6** remained; the titer was less than 5%

of the overall RAL pool at the end of four days of continued culturing. The mass spectrum of the new compound matched precisely to that of the estimated isotope mass pattern of **11**. To verify the identity of the compound and the regioselectivity of Rdc2, we scaled up the yeast culture and purified the compound at a final titer of 14.3 mg/L. One and two dimensional NMR was performed to verify that the new compound is indeed **11** (Table 12). The ^{1}H NMR displayed a single aromatic proton signal at δ 6.62 ppm instead of the typical spin system observed for the resorcylate core of **6**, indicative of loss of one proton. Further comparison of the 2D NMR data to **6** confirmed that no proton is linked to C6 in the HSQC spectrum; while correlation between C1' proton and C6 remained in HMBC spectrum. Therefore, co-expression of the three *rdc* genes in yeast led to robust synthesis of **11** and thereby confirmed the role of Rdc2 as a C6-specific chlorinase. This high in vivo conversion of **6** to **11** also demonstrated the robust activity of the heterologously expressed Rdc2.

Table 12. NMR Table for **11**[a].



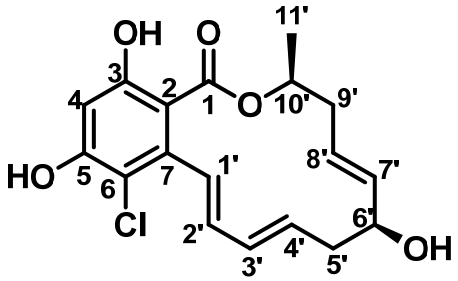
No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	170.7	-	
2	108.3	-	
3	163.6	-	
4	103.8	6.62 (s, 1H)	C2, C3, C5, C6
5	158.9	-	
6	116.2	-	
7	138.0	-	
1'	46.3	4.14 (d, 1H, 17.5) 4.36 (d, 1H, 17.5)	C2, C6, C7, C2'

2’	195.3		
3’	131.0	5.85 (d, 1H, 15.5)	C2’
4’	147.4	6.69-6.73 (m, 1H)	C2’, C5’, C6’
5’	31.9	2.17-2.34 (m, 2H)	C3’, C4’, C6’, C7’
6’	31.9	2.17-2.34 (m, 2H)	C5’
7’	127.7	5.34-5.36 (m, 1H)	C5’, C6’, C8’, C9’
8’	133.0	5.25-5.31 (m, 1H)	C6’
9’	37.3	2.26-2.29 (m, 1H) 2.59-2.64 (m, 1H)	C7’, C8’, C10’, C11’
10’	73.4	5.40-5.45 (m, 1H)	C1, C9’
11’	18.2	1.32 (d, 3H, 6.6)	C9’, C10’

[a] Spectra were obtained at 500 MHz for proton and 125 MHz for carbon and were recorded in CD_3COCD_3.

We also investigated the in vitro conversion of **6** to **11** using purified Rdc2 from *E. coli* and the *E. coli* SsuE as a flavin reductase accessory enzyme. We were able to observe near complete conversion of **6** to **11** in the presence of NADPH and NaCl in four hours (Fig. 20A, trace ii). By LCMS analysis, we also detected a compound **12** that appeared to be the dichlorinated **6** from both in vitro assay and in vivo extract at a low yield. The structure of **12**, however, has not been confirmed by NMR analysis.

Table 13. NMR Table for **14** [a].



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	170.46	-	
2	110.49	-	
3	157.75	-	
4	103.19	6.51 (s, 1H)	C2, C3, C5, C6
5	160.23	-	
6	112.62	-	
7	139.88	-	

1’	126.94	6.52 (d, 2H, 16.1)	C7, C2’, C3’
2’	137.28	5.84 (dt, 1H, 16.1, 6.75)	C3’, C4’, C7
3’	31.65	2.22-2.31 (m, 2H)	
4’	23.65	1.41-1.46 (m, 1H)	C2’, C3’, C5’, C6’
		1.68-1.74 (m, 1H)	C2’, C3’, C5’, C6’
5’	35.09	1.55-1.61 (m, 1H)	C3’, C4’, C6’, C7’
		1.68-1.70 (m, 1H)	C3’, C4’, C6’, C7’
6’	72.37	4.18 (m, 1H)	C8’
7’	137.66	5.54-5.59 (m, 1H)	C6’, C9’, C8’
8’	126.29	5.61-5.67 (m, 1H)	C6’, C7’, C9’, C10’
9’	38.17	2.32-2.37 (m, 1H)	C11’, C10’, C8’, C7’
		2.50-2.54 (m, 1H)	C11’, C8’, C7’
10’	73.36	5.27 (m, 1H)	C8’
11’	20.19	1.35 (d, 3H, 6.3)	C9’, C10’

[a] Spectra were obtained at 500 MHz for proton and 125 MHz for carbon and were recorded in CD_3COCD_3.

2.1.3.6. Engineered Biosynthesis of Chlorinated RAL Analogs.

Since RALs share the similar structural scaffold and the core resorcylate, we tested whether Rdc2 may display substrate tolerance towards other RAL compounds. Dehydrozearalenol (**13**), the earliest RAL product from the hypothemycin pathway [211], serves as an ideal candidate for the assay. As shown in Fig. 20B (trace ii), the in vitro mixture of **13**, Rdc2, SsuE and NADPH led to the emergence of a new peak **14** of which the isotopic mass pattern is indicative of a chlorinated compound. To characterize **14**, we introduced the yeast Rdc2 expression plasmid into the DHZ-producing strain, which co-expresses the HRPKS Hpm8 and the NRPKS Hpm3. Following organic extraction of the culture, we also observed the presence of **14** that accumulated to approximately the same level as **13** after seven days of culturing. Compound **14** was purified from the yeast culture to a final titer of 9 mg/L and was confirmed by NMR to be 6-chloro, 7’,8’-dehydrozearalenol (Cl-DHZ) **8** (Table 13).

2.1.4. Discussion

In this work, we fully reconstituted the enzymatic activities of Rdc5 and Rdc1 from radicicol biosynthetic pathway with the production of monocillin II **6** as an early intermediate. Following the same biosynthetic model as that of other RAL compounds, the assembly of **6** from malonyl-CoA requires one HRPKS (Rdc5) and one NRPKS (Rdc1). The tandem activities of the two IPKSs represent the fungal version of the well-studied modular PKS assembly lines found in bacteria. However, the fungal strategy is significantly more compact and efficient, utilizing only two "modules" and a combined 11 active sites to complete the >30 steps of catalytic events. The workload is divided based on the capabilities of each IPKS: the upstream module Rdc5 is solely responsible for assembling the reducing portion of **6**, while the downstream module Rdc1 is responsible for the aromatic portion and cyclization of **6**. The iterative nature of the fungal IPKSs may initially appear to make them less amendable to combinatorial biosynthesis approaches that have been successfully implemented on modular type I PKSs. However, the relaxed substrate specificity of the NRPKS enzymes and their ability to accept smaller thioesters as starter units should permit formation of many interesting compounds. This highly efficient manner of synthesizing and tailoring the polyketide backbone is evolutionarily more advantageous, especially for the metabolically more complex eukaryotic organisms. However, the IPKSs, especially the HRPKS, require considerably more complex programming rules as discussed below. The use of a "bi-modular" strategy to synthesize polyketides containing structurally distinct subunits is also found in other fungal biosynthetic pathways. For example in the aflatoxin biosynthetic pathways, a FAS-like module synthesizes the hexanoyl starter unit, which is then transferred to the NRPKS PksA to synthesize the anthraquinone portion of the precursor norsolorinic acid [223]. Similarly, in the recently uncovered biosynthetic pathway of

asperfuranone [224], an HRPKS and a NRPKS synthesize the linear tetraketide and the β-resorcylic aldehyde subunits, respectively.

Using the pentaketide SNAC **7** and precursor directed biosynthesis, we showed that the "5+4" distribution of effort between Rdc5 and Rdc1 is a highly plausible mechanism of assembling **6**. This programming rule is apparently different from the "6+3" split employed by the zearalenone (PKS4 and PKS13) and hypothemycin (Hpm8 and Hpm3) pathways [205,210,211]. The "5+4" mechanism maintains the aforementioned chemical modularity of the two IPKSs. Each β-keto position synthesized by Rdc5 is reduced, while those synthesized by Rdc1 remain completely unreduced. With this split, the KR of Rdc5 reduces the β-carbon in a nondiscriminatory fashion after each chain extension step, and therefore requires no additional programming rule to distinguish a β-keto hexaketide. Indeed, this feature of KR can be found in all RAL HRPKSs. We previously suggested that PKS4 KR is inactive at the tetraketide step in the synthesis of zearalenone scaffold, thereby, giving rise to the C6' ketone. However, further studies with the highly parallel hypothemycin pathway [211] suggests that the KR is indeed active at this step and affords the secondary alcohol at C7' that is eventually oxidized to the ketone by an alcohol oxidase encoded in both pathways. To maintain this orthogonal chemical modularity between the HRPKS and NRPKS in the *rdc* pathway, the substrate specificity of the Rdc1 SAT domain must therefore be different from that of Hpm3 and PKS13. The SAT domain is the key enzyme that facilitates acyl transfer of the completed, reduced precursor from the ACP domain of the HRPKS to the ACP domain of the NRPKS, thereby ensuring the correctly substrate is passed between the two PKSs [208,211]. The Rdc1 SAT domain therefore must display specificity for the pentaketide acyl chain instead of the hexaketide chain. Following

chain transfer, the KS domain of Rdc1 maintains strict chain length control to synthesize the complete nonaketide.

Although the KR domain appears to function during each cycle, the DH and the ER domains are programmed to be different, thereby lead to the different reduced chains among zearalenone, hypothemycin and radicicol. All three DH domains are programmed to be inactive towards the β-hydroxybutyryl diketide intermediate, which is different from LNKS and LDKS in the lovastatin pathway [85,225]. While the Rdc5 DH does not discriminate against the remaining β-hydroxyl intermediates, the PKS4 and Hpm8 DH domains do not dehydrate the β-hydroxyl position of the tetraketide, which is a required process to install the C6' hydroxyl in zearalenol and **13**. Similar analysis of the substrate specificities of the ER domains also reveals subtle, yet important differences between the three RAL HRPKSs. For example, our work here demonstrates that Rdc5 ER can only reduce the double bond of the tetraketide during biosynthesis of **6**. In contrast, the Hpm8 ER only functions on a α, β-unsaturated pentaketide whereas the PKS4 ER acts on both the diketide and pentaketide. These differences in substrate specificity of a single DH/ER domain towards different intermediates of different chain lengths, as well as the differences between homologous DH/ER domains towards substrates of the same chain lengths are faithfully maintained to produce the different RAL compounds. Understanding the structural basis of these catalytic differences will be a key to decode HRPKS programming rules and allow precise prediction of product structures from protein sequences.

Although the three RAL KR domains can reduce β-keto functionality in all intermediates, the programming rules dictating stereochemical outcomes are different. In contrast to the *S*-stereochemistry found in **13** and zearalenol, the first β-ketoreduction of acetoacetyl diketide by Rdc5 takes place to give an *R*-configuration. Compared to the KR domains of bacterial type I

modular PKSs, of which the biochemical and structural basis of stereochemical control has been well established [226-229], the stereospecificity of KR in fungal PKSs remains enigmatic. For the type I modular KRs, it has been proposed that β-keto polyketides gain access to the catalytic site via different paths, which are oriented and controlled by characteristic residues in the active site. From structural and mutational analysis, B-type KR contains the LDD tripeptide fingerprint and reduces with *R* stereochemistry [229,230]. Sequence alignment between Rdc5 KR with B-type KR revealed a similarly placed patch of LRD in Rdc5 KR instead of the LDD sequence. However, this LRD motif is also conserved in Hpm8 and PKS4, both of which reduce the same acetoacetyl diketide with *S* stereochemistry. Therefore, the classification of type I modular KR stereochemistry based on sequence analysis does not apply to fungal IPKS KRs. Another degree of complexity associated with HRPKS KR domains may be their abilities to reduce substrates of different chain lengths with differing stereochemistry. This is not apparent in our current study with Rdc5 since the other four β-hydroxyl products of the KR are all dehydrated by the DH domain, thereby masking the stereochemistry of these keto reductions. However, we previously observed that during the synthesis of **13** by Hpm8 and Hpm3, the two nondehydrated hydroxyl groups at C6' and C10' were apparently the results of β-keto reduction by Hpm8 KR with opposite stereochemistry. Therefore, it is possible that the masked reduction steps in Rdc5 may also proceed via different stereochemistries, which may result from orientation of acyl chains of different sizes divergently in the cavity. We are currently in the process of establishing the correlation between chain length and stereospecificity of the various RAL KR domains using synthetic β-keto substrate mimics. It may also be possible that the differences in β-hydroxyl stereochemistry may influence the selectivity of the subsequent DH-catalyzed dehydrations as discussed previously [231,232].

Regardless of the stereochemistry of the terminal hydroxyl nucleophile, the Rdc TE domain can complete the macrocyclization to yield either *R*- or *S*-monocillin II. Substrate tolerance was also observed in TEs from type I modular PKS, such as that of DEBS. Engineered bi-modules fused with cognate TE domain can produce different 6-membered triketide lactones with inversion in hydroxyl group stereochemistry [233,234]. However, the ability to form a same-sized macrolactone using nucleophiles of opposite stereochemistry has not been demonstrated for a bacterial TE domain. Our results here, combined with our previous studies with RAL TEs from Hpm3 and PKS13 [210,211,221], further illustrates the remarkable substrate tolerance of this class of fungal macrolactonizing TE domains.

In addition to reconstituting the combined >30 individual reactions catalyzed by Rdc5 and Rdc1, we also confirmed the function of Rdc2 as a flavin-dependent halogenase in the biosynthesis of **11**. Although the catalytic mechanisms of halogenases in this family have been well-studied for homologs in bacteria, this represents the first heterologous reconstitution of a fungal aromatic chlorinase. It has been shown that certain aromatic chlorinases generate hypochlorous acid through the flavin cofactor and molecular oxygen [212]. The nucleophilic hypochlorous acid is then covalently linked to a lysine residue in the active pocket to form a reactive chloramine, which can attack the electron rich aromatic substrate [235,236]. Sequence alignment of Rdc2 with selected bacterial halogenases revealed the presence of a likely active site containing the lysine residue K74. Interestingly, Rdc2 appears to be a very efficient chlorinase from in vitro analysis, which is in contrast to most bacterial chlorinases reconstituted to date. Rdc2 also functioned efficiently in the heterologous host as evidenced by the nearly complete conversion of **6** to **11**. Furthermore, Rdc2 was readily combined with heterologous

RAL IPKSs to generate a different chlorinated RAL **14**, demonstrating its potential utility in combinatorial biosynthesis.

This work on reconstitution of the *rdc* enzymes further showcases the usefulness of the BJ5464-NpgA heterologous host. In addition to being a host capable of expressing the fungal megasynthases at preparative quantities, it can be considered as a versatile system for engineered biosynthesis of fungal polyketides. We first demonstrated that precursor directed biosynthesis can be performed with the host through the synthesis of *R*-monocillin II **6** and its enantiomer **10**. The large acyl-SNAC substrates **7** and **9** were able to penetrate the yeast cell membrane and were available intracellularly to prime Rdc1. This can be highly useful in the probing of fungal PKS programming rules, as well as generation of polyketide analogs. The synthesis of **11** and **14** demonstrates that the yeast host is efficient in coexpression of multiple fungal biosynthetic enzymes through the multi-plasmid approach. This feature can be particularly useful in the heterologous reconstitution of a multistep fungal biosynthetic pathway and the study of fungal-specific tailoring enzymes.

2.2. A Fungal Nonribosomal Peptide Synthetase Module that can Synthesize Thiopyrazines and Thiopyrroles[2]

2.2.1. Introduction

Although filamentous fungi have a strong track record in producing blockbuster drugs such as penicillin and lovastatin [237], these microorganisms are widely considered to be underachievers in natural product biosynthesis. This is evidenced by the presence of a large number of silent secondary metabolism pathways in recently sequenced fungal genomes [179,238]. As a result, the vast biosynthetic potential of species in the *Aspergillus, Penicillium, Gibberella,* etc genera is far from being realized. Mining these uncharacterized pathways is therefore an important objective towards discovery of new bioactive molecules and novel enzymatic machineries. While different strategies have been employed to activate cryptic pathways in the native hosts [239-241], a general strategy that can directly harness the hidden enzymatic power is the heterologous expression and reconstitution of fungal enzymes in genetically well-established microorganisms, such as *Escherichia coli* [242,243] and *Saccharomyces cerevisiae* [85,244]. These hosts offer a clean background to study the biosynthetic products of the transplanted pathways and enzymes, as well as a means to produce abundant amounts of purified proteins for in vitro analysis [245].

Among the natural products biosynthesized by fungi, polyketides (such as lovastatin) and nonribosomal peptides (NRPs, such as penicillin) are of particular interest because of their diverse biological activities. The biosynthetic enzymes that assemble these molecules, such as polyketide synthases (PKSs) and NRP synthetases (NRPSs) are large megasynthases and the genes encoding them can be readily identified from the genome. The PKSs iteratively condense acetate-derived building blocks using domains such as ketosynthase (KS), malonyl-CoA:ACP

[2] Compounds are numbered independently in this section.

acyltransferase (MAT) and acyl carrier protein (ACP); and modify the nascent polyketide chain using tailoring domains such as ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and methyltransferase (MT) [109]. On the other hand, NRPSs catalyze the formation of amino acid-derived products using domains such as condensation (C), adenylation (A) and thiolation (T) [246]. Each set of C-A-T catalyzes one round of i) adenylation of an amino acid by the A domain and transfer to the phosphopantetheinyl (pPant) arm of T domain; and ii) condensation between the nucleophilic amino group and an electrophilic carbonyl by the C domain. Using this biosynthetic logic, fungal NRPS modules can synthesize tetramic acids when fused to a PKS [45,239,247,248]; diketopiperizines when paired in tandem [242,249]; and other oligopeptides [250].

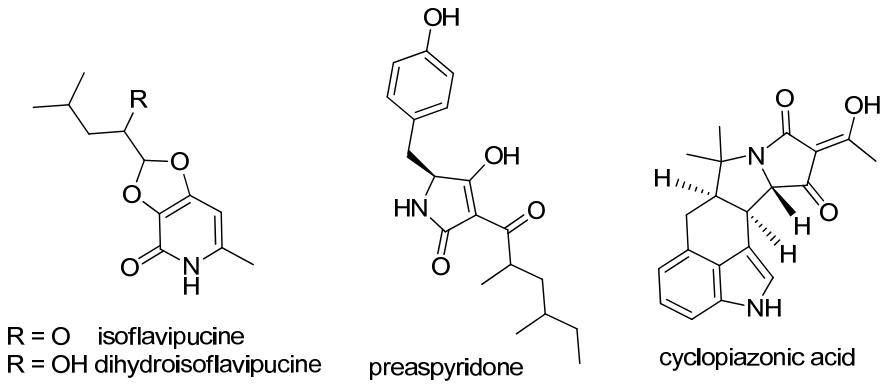


Figure 21. Selected tetramic acid natural products.

We previously reconstituted the activities of ApdA, an *Aspergillus nidulans* PKS-NRPS megasynthetase that synthesizes the tetramic acid preaspyridone (Figure 21) [251]. We demonstrated that the ApdA PKS and NRPS modules of the megasynthetase can be dissected and can interact functionally in trans. When ApdA PKS module and cyclopiazonic acid synthetase (CpaS) NRPS module from *Aspergillus flavus* [137] were combined, a tryptophan-containing preaspyridone analog was obtained. Inspired by this result of combinatorial biosynthesis, we set out to funcationally identify other heterologous NRPS modules from

different sequenced fungal species. In this study, we show that serendipidously, the NRPS module (NRPS325) of the only PKS-NRPS megasynthetase (ATEG00325) in *Aspergillus terreus* can synthesize thiol-substituted pyrazines and pyrroles. The thiopyrazine synthetase activities were independent of any upstream PKS activities. The natural role of ATEG00325 was genetically identified to be involved in the biosynthesis of isoflavipucine and dihydroisoflavipucine (Figure 21) [252], highlighting the unexpected biosynthetic potential of this NRPS module unlocked during our genome mining studies.

2.2.2. Materials and Methods

2.2.2.1. Bioinformatics Analysis

Putative PKS-NRPSs were identified by performing a feature search of Aspergillus Comparative Database (Broad Institute, Fungal Genome Initiative) using polyketide synthase as the key word query and a BLASTP search using equisetin adenylation domain sequence as the query sequence. Both results enabled us to identify a single PKS-NRPS ATEG00325 in the genome of *Aspergillus terreus*. The adjacent genes of ATEG00325 were used as queries to perform BLASTP search in NCBI database and were assigned based on its closest homolog. Predicted conserved domains of ATEG00325 were assigned by Pfam analysis (http://pfam.sanger.ac.uk/) [253]. The adenylation domain 10 amino acid code of ATEG00325 was identified by submitting NRPS325 sequence to the automated web-based resources NRPS predictor (http://www-ab.informatik.uni-tuebingen.de/software/)[254]. All primary sequence alignments including ATEG00325 PKS and NRPS domains were performed using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2).

2.2.2.2. Strains and General Techniques for DNA Manipulation

E. coli XL1-Blue and *E. coli* TOP10 (Invitrogen) were used for cloning following standard recombinant DNA techniques. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs). PCR was performed using Platinum Pfx DNA polymerase (Invitrogen). The constructs of pCR-Blunt vector (Invitrogen) containing desired PCR products were confirmed by DNA sequencing. *E. coli* BL21(DE3) (Novagen) and BAP1 were used for protein expression.

2.2.2.3. Cloning of C-terminus Hexahistidine NRPS325 Expression Plasmid

Genomic DNA from *A. terreus* was used as the template for PCR amplification. The gene encoding NRPS325 (residues 2447-3890) was amplified with primer pair NRPS325-NheI-f (5'AAGCTAGCATGACTCCACACACCGAAATCA3') and NRPS325-NotI-r (5'AAGCGGC CGCCTGCCTCAATCCCTTA3'), flanking the gene product with restriction sites *Nhe*I and *Not*I. The PCR product was first placed into a pCR-Blunt vector and sequenced to create pKJ74. Then, the gene product was prepared by digesting pKJ74 with *Nhe*I and *Not*I, and ligated into pET23a vector to generate NRPS325 expression plasmid pKJ75.

2.2.2.4. Expression and Purification of NRPS325 in *E. coli*.

The expression plasmid pKJ75 was transformed into *E. coli* BL21(DE3) strain and *E. coli* BAP1 strain for *apo*- and *holo*- protein expression, respectively. For 1 L of LB liquid culture, the cells were grown at 37 °C in LB medium with 100 μg/mL ampicillin to an OD_{600} of 0.4-0.6. At which time the cells were incubated on ice for 10 minutes, and then induced with 0.1 mM isopropyl thio-β-D-galactoside (IPTG) for 16 hours at 16°C. The cells were harvested by

centrifugation (3500 rpm, 15 minutes, 4°C), re-suspended in 30 mL lysis buffer (50 mM Tris-HCl, 2 mM EDTA, 2 mM DTT, 500 mM NaCl, 5 mM imidazole, pH=7.9) and lysed using sonication on ice. Cellular debris was removed by centrifugation (30000 g, 30 min, 4 °C). Ni-NTA agarose resin was added to the supernatant and the solution was stirred at 4°C for at least 2 hours. The protein resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH=7.9). Purified NRPS325 were concentrated and buffered exchanged into buffer E (50 mM Tris-HCl, 100 mM NaCl, pH=7.9) + 10% glycerol. The final enzyme was concentrated, aliquoted and flash frozen. Protein yield was determined to be 25 mg/L, with the Bradford assay using BSA as a standard.

2.2.2.5. NRPS325 PPi Releasing Assays

Kinetic analysis of NRPS325 was performed using the Enzymatic Determination of Pyrophosphate Kit (Sigma-Aldrich), essentially following the manufacturer's protocol but with 100 uL assay volumes in a quartz 96-well microplate. In this assay, pyrophosphate (PPi) was hydrolyzed to inorganic phosphate (Pi) in the presence of D-fructose-6-phosphate (F-6-P) catalyzed by fructose-6-phosphate kinase (PPi-PFK). The byproduct D-fructose-1,6-phosphate (F-1,6-P) was further decomposed to D-glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) catalyzed by aldolase. Under the control of Triosephosphate isomerase, GAP and DHAP can be transformed to each other. DHAP can be further reduced to glycerol-3-phosphate by glycerophosphate dehydrogenase (GDH) with the consumption of NADH (maximum absorption at 340nm). Therefore, the detection of PPi was

indirectly connected to the consumption of NADH, which can be monitored spectrophotometrically at 340nm[255].

Reactions (100 µL) contained 75 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM ATP, 0.25 mM NADH, 4 mM F-6-P, 0.2 units/mL PPi-PFK, 2.5 units/mL aldolase, 1.8 units/mL GDH and 20 units/mL TPI and 1 µM NRPS325 was used with 2 mM L-amino acid. Absorbance at 340 nm was measured over a 10-min interval in Biotek Microplate Spectrophotometer PowerWave XS. The spectrophotometer recorded data points every 20 s. Each assay was performed in duplicate. A linear reaction velocity was obtained by using a minimum of 20 co-linear data points and an extinction coefficient of 6400 $M^{-1}cm^{-1}$.

2.2.2.6. NRPS325 in vitro Turnover Assays

10 µM of NRPS325 was incubated with 2 mM acetoacetyl-*S*-NAC, 10 mM L-leucine, 10 mM $MgCl_2$, 20 mM ATP and 2 mM NADPH in phosphate buffer (100 mM NaH_2PO_4, pH=7.4) at room temperature to produce compound 1a. Reaction (50 µL) contained 10 µM of NRPS325, 2 mM SNAC, 10 mM L-leucine, 10 mM $MgCl_2$, 20 mM ATP and 2mM NADPH in phosphate buffer was incubated at room temperature to produce 2a. All the in vitro assays were quenched after overnight reaction and extracted twice with an equal volume of ethyl acetate (EA). The organic phase was separated, evaporated to dryness, and re-dissolved in 20 µL of methanol. The organic residue was analyzed by LC-MS. LC-MS was conducted with a Shimadzu 2010 EV Liquid Chromatography Mass Spectrometer by using both positive and negative electrospray ionization and a Phenomenex Luna 5µ 2.0 x 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5 to 95% CH_3CN (vol/vol) over 30 min and further 95% CH_3CN (vol/vol) for 20 min in H_2O supplemented with 0.05% (vol/vol) formic acid at a flow

rate of 0.1mL/min at room temperature. The LC retention time of 1a was 24.4 min, while the retention time of 2a was 29.3 min.

2.2.2.7. NRPS325 Transesterification Assays

Reaction (50 μL) containing 10 μM NRPS325 was incubated with 5 mM SNAC, 2 mM acetoacetyl-CoA in phosphate buffer at room temperature. In addition, the control reaction (50 μL) containing 2 mM acetoacetyl-CoA and 5 mM SNAC was incubated in phosphate buffer at room temperature. All the in vitro assays were quenched hourly and extracted twice with an equal volume of 99% ethyl acetate (EA)/1% trifluoroacetic acid (TFA). The organic phase was separated, evaporated to dryness, and re-dissolved in 20 μL of methanol. The organic residue was analyzed by LC-MS. LC-MS was conducted with a Shimadzu 2010 EV Liquid Chromatography Mass Spectrometer by using both positive and negative electrospray ionization and a Phenomenex Luna 5μ 2.0 x 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5 to 95% CH_3CN (v/v) over 30 min and 95% CH_3CN (v/v) further for 20 min in H_2O supplemented with 0.05% (v/v) formic acid at a flow rate of 0.1mL/min at room temperature.

2.2.2.8. In vitro Large Scale Synthesis of Compound 1a and NMR.

During the scaled-up reaction, the following reagents were added to phosphate buffer (10 mL): 2 mM acetoacetyl-*S*-NAC, 10 mM $MgCl_2$, 20 mM ATP, 10 mM L-leucine and 10 μM NRPS325. The reaction mixture was stirred gently at room temperature and the reaction progress was followed by HPLC. After the product level had maximized, the reaction mixture was extracted three times with EA. The resultant organic extracts were combined and evaporated to

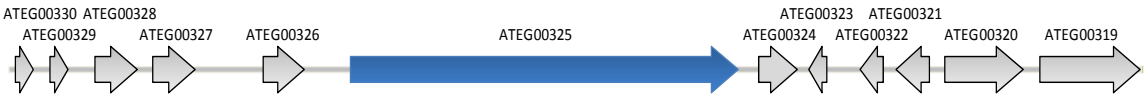
dryness, re-dissolved in methanol and purified by reverse-phase HPLC (XTerra Prep MS C18 5µm, 19 mm X 50 mm) on a linear gradient of 5 to 95% CH_3CN (v/v) over 15 min and 95% CH_3CN (v/v) further for 15 min in H_2O supplemented with 0.1% (v/v) trifluoroacetic acid at a flow rate of 2.5 mL/min. High resolution mass analysis of 1a was performed on VG ZAB2SE (1996) high resolution mass spectrometer, with Opus V3.1 and DEC 3000 Alpha Station. 1D and 2D NMR of 1a were performed on Bruker DRX-500 spectrometer using $CDCl_3$ as the solvent.

2.2.2.9. In vitro Large Scale Synthesis of Compound 1a and NMR.

During the scaled-up reaction, the following reagents were added to phosphate buffer (10 mL): 5 mM SNAC, 10 mM $MgCl_2$, 20 mM ATP, 10 mM L-leucine and 10 µM NRPS325. The reaction mixture was stirred gently at room temperature and the reaction progress was followed by HPLC. After the product level had maximized, the reaction mixture was extracted three times with EA. The resultant organic extracts were combined and evaporated to dryness, re-dissolved in methanol and purified by reverse-phase HPLC (XTerra Prep MS C18 5µm, 19 mm X 50 mm) on a linear gradient of 5 to 95% CH_3CN (v/v) over 15 min and 95% CH_3CN (v/v) further for 15 min in H_2O supplemented with 0.1% (v/v) trifluoroacetic acid at a flow rate of 2.5 mL/min. High resolution mass analysis of 2a was performed on VG ZAB2SE (1996) high resolution mass spectrometer, with Opus V3.1 and DEC 3000 Alpha Station. Both 1D and 2D NMR of 2a were performed on Bruker DRX-500 spectrometer using CD_3COCD_3 as the solvent.

2.2.3. Results and Discussion

Table 14. Bioinformatics study on ATEG00325 and its adjacent genes.



Gene No.	Gene Size (bp)	Closest Homolog	Identity/ Similarity (%)	Conserved Domain*	E value
ATEG00330	483	*A. terreus*, ATEG00329	44/63	No conserved domain	
ATEG00329	489	*A. fumigatus*, Afu8G00430	45/68	No conserved domain	
ATEG00328	1101	*A. terreus*, ATEG00326	27/40	cd00067, GAL4, GAL4-like Zn2Cys6 binuclear cluster DNA-binding domain	$3x10^{-6}$
ATEG00327	1320	*A. fumigatus*, Afu8G00530	65/79	pfam06500, Alpha/beta hydrolase of unknown function	$3x10^{-5}$
ATEG00326	1275	*T. stipitatus*, TSTA_060580	42/57	cd00067, GAL4, GAL4-like Zn2Cys6 binuclear cluster DNA-binding domain	$2x10^{-4}$
ATEG00325	11673	*A. fumigatus*, Afu8g00540	39/57	KS-AT-KR-DH-ER-ACP-C-A-T-R	
ATEG00324	1143	*A. flavus,* AFLA068350	65/73	cd00195, UBCc, Ubiquitin-conjugating enzyme E2, catalytic (UBCc) domain	$6x10^{-21}$
ATEG00323	480	*A. niger,* An08g03510	38/52	No conserved domain	
ATEG00322	351	*N. fishcheri,* NFIA014430	84/92	pfam10270, Tmemb_32, Transmembrane protein precursor 32	$8x10^{-14}$
ATEG00321	681	*A. oryzae,* AO090038000287	82/93	cd00292, EF1B, Elongation factor 1 beta (EF1B) guanine nucleotide exchange domain	$4x10^{-20}$
ATEG00320	2421	*A. oryzae,* AO090038000288	89/93	TIGR03346, chaperone_ClpB, ATP-dependent chaperone ClpB	$<1x10^{-180}$
ATEG00319	2379	*A. niger,* An08g03470	66/79	TIGR01070, mutS1, DNA mismatch repair protein MutS	$5x10^{-64}$

* Based on NCBI Conserved Domain Search.

During bioinformatics analysis of the sequenced *A. terreus* genome, we were intrigued by the possible product(s) of the only encoded PKS-NRPS hybrid (ATEG00325) (Figure 21A). This gene has no assigned function and no tetramic acid compounds are known to be biosynthesized by *A. terreus*. Sequence analysis of the encoded protein showed that while the KS, MAT and ACP domains align well with other fungal PKSs, the tailoring domains such as KR, ER and MT do not contain well-conserved cofactor binding sites and are thus likely inactive (Figure 22). The

NRPS module (NRPS325) is capped with a *C*-terminus NADPH-binding reductase (R) domain that is homologous to those found in equisetin [247] and tenellin synthetases [45]. Since the PKS module may only be able to synthesize an acetoacetyl-*S*-ACP precursor, ATEG00325 appears to have comparable activities as the α-cyclopiazonic acid synthetase CpaS that synthesizes the tetramic acid cyclo-acetoacetyl-L-tryptophan [137,248]. Alignment of the NRPS325 A domain with other A domains of known amino acid specificity shows a few matches among the consensus sequences that are important for substrate specificity. To analyze the activities of NRPS325, we expressed and purified the 158 kDa *holo*-NRPS module consisting of C-A-T-R from the engineered *E. coli* BAP1 (Figure 22) [256].

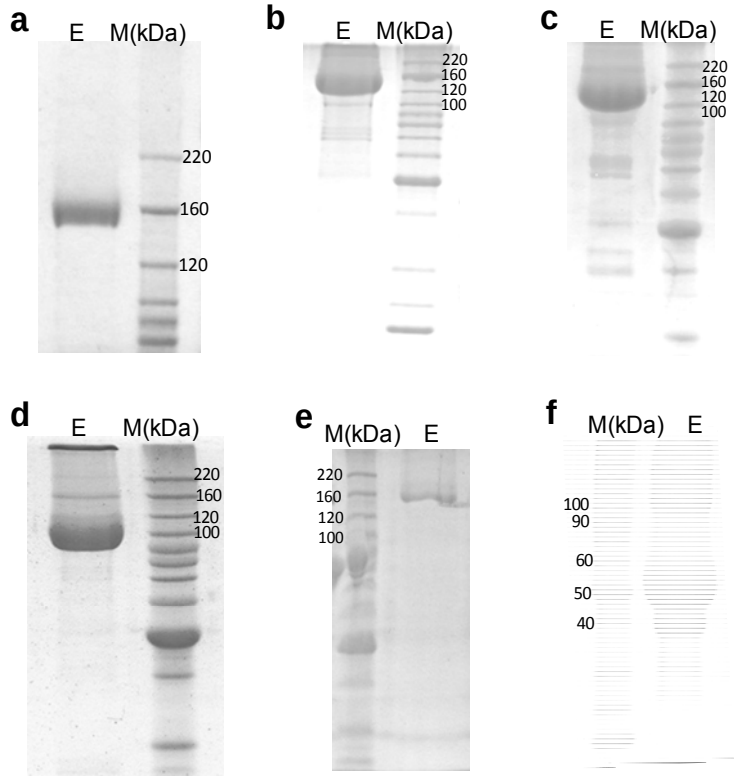


Figure 22. SDS-PAGE gels of proteins purified from *E. coli* BAP1 and BL21(DE3). (a) 6% SDS-PAGE gel of C-terminal hexahistidine tagged NRPS325 (158 kDa) from *E. coli* BAP1 after Ni-NTA chromatography. (b) 12% SDS-PAGE gel of C-terminal hexahistidine tagged NRPS325-H193A (158 kDa) from *E. coli* BAP1 after Ni-NTA chromatography. (c) 12% SDS-PAGE gel of C-terminal hexahistidine tagged NRPS325-H194A (158 kDa) from *E. coli* BAP1 after Ni-NTA chromatography. (d) 12% SDS-PAGE gel of C-terminal hexahistidine tagged ATR (107 kDa) from *E. coli* BAP1 after Ni-NTA

chromatography. (e) 12% SDS-PAGE gel of C-terminal hexahistidine tagged NRPS325-G1132A-G1135A (158 kDa) from *E. coli* BAP1 after Ni-NTA chromatography. (f) 12% SDS-PAGE gel of C-terminal hexahistidine tagged TR (51 kDa) from *E. coli* BAP1 after Ni-NTA chromatography (there is a 40 kDa truncated R domain existed). In all the gels, lane M: Invitrogen Benchmark Protein Ladder, Lane E: Final elution of protein after chromatography.

The relative specificity of the A domain was assessed using a pyrophosphate release assay in the presence of different amino acids (Figure 24). NRPS325 displayed substrate promiscuity towards nearly all the natural aliphatic and aromatic amino acids, including L-Ile, L-Met, L-Leu, L-Val, L-Phe, L-Tyr and L-Trp. Towards L-Leu, which is the predicted amino acid substrate by NRPSpredictor [257], the A domain displayed k_{cat} of 17.2 ± 2.4 min^{-1} and K_m of 310.0 ± 15.5 μM (Figure 25).

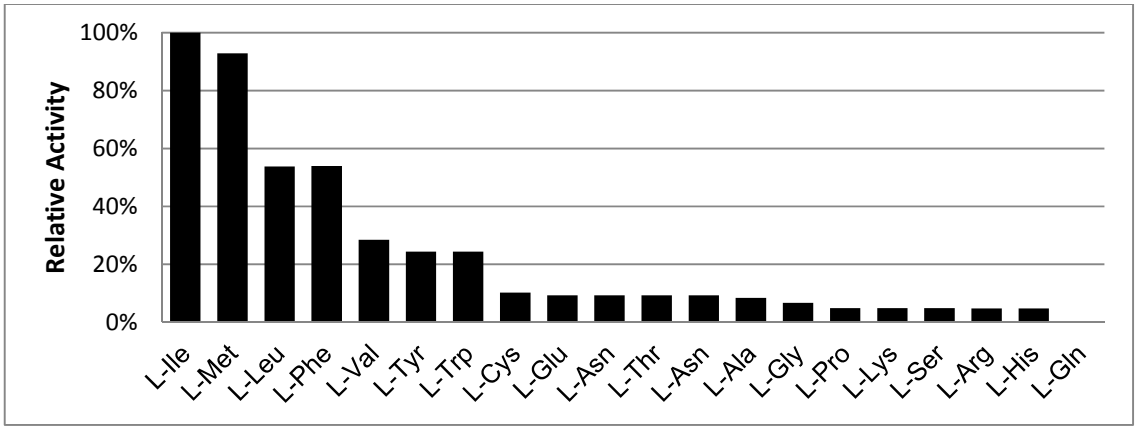


Figure 24. Relative activities of NRPS325 towards different amino acid substrates as determined by the PPi releasing assay. The y axis indicates the relative activity for various amino acid compared to the activity for L-isoleucine (100% activity).

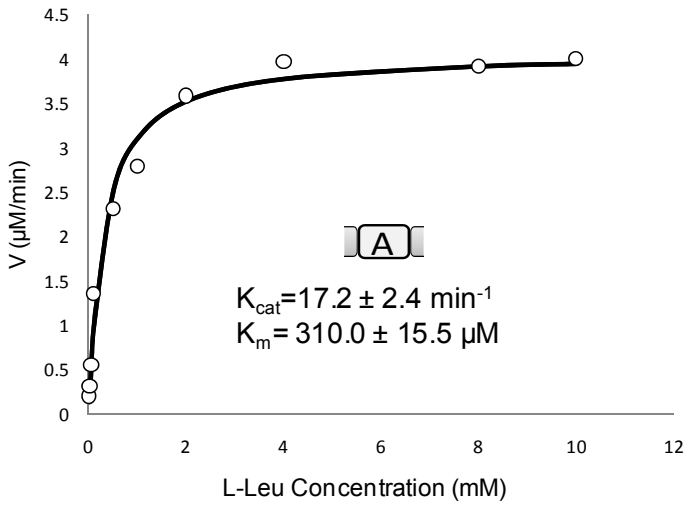


Figure 25. Kinetic analysis of NRPS325 catalyzed formation of L-leucyl-AMP. Reaction (100 μL) contains 75 mM Tris-HCl (pH=7.5), 10 mM $MgCl_2$, 5 mM ATP and was incubated with 500 nM NRPS325 in the presence of various concentrations (from 20 μM to 10 mM) of L-leucine at 25 °C.

We then attempted to elucidate the function of NRPS325 in vitro in the presence of l-Leu, NADPH, ATP and acetoacetyl-*S*-*N*-acetylcysteamine (acetoacetyl-*S*-NAC, **3**) (Figure 26). The thioester **3** was chosen to represent the likely polyketide precursor synthesized by the upstream PKS module. The reaction extract contained essentially a single product **1a** (Figure 26, i) with UV absorption maxima (λ_{max}) at 296 nm, which indicates the presence of a conjugated chromophore. In the negative controls, excluding either L-Leu (Figure 26B, ii) or ATP from the above reaction mixture, or using the apo-form of NRPS325 expressed from BL21 (DE3), resulted in no product formation. Finally, formation of **1a** was absolutely dependent on NADPH (Fig. 26B, iii), which points to a likely role of the R domain in reductive product release.

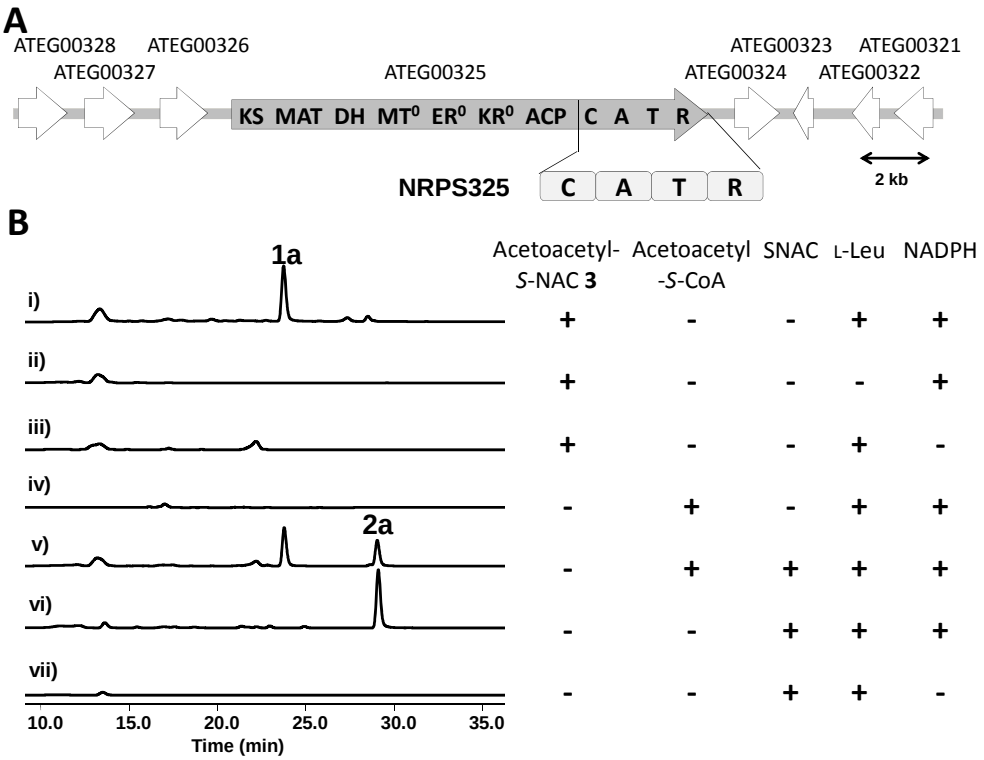
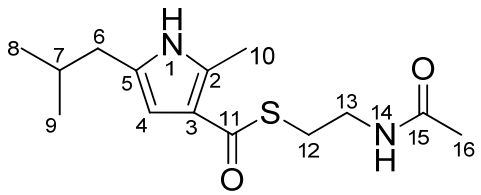


Figure 26. Identification and in vitro reconstitution of NRPS325. (**A**) Organization of ATEG00325 gene cluster from *A. terreus*. Highlighting in grey, the 12.4kb hybrid PKS-NRPS gene ATEG00325 is the only PKS-NRPS encoded in the *A. terreus* genome. (**B**) HPLC analysis (300 nm) of ethyl acetate extracts from the in vitro assays i-vii. Each assay contains 10 μM NRPS325 in the presence of different reagents shown on the right of the traces. The final concentrations of the different components, when added, were: acetoacetyl-*S*-NAC **3**: 5 mM; SNAC: 5 mM; acetoacetyl-CoA: 5 mM; L-Leu: 10 mM; NADPH: 2 mM; and ATP: 20 mM. All reactions were performed at room temperature for 12 hr in phosphate buffer (pH=7.4).

The molecular formula of **1a** was determined to be $C_{14}H_{22}N_2O_2S$ by High-Resolution Mass Spectrometry The surprising presence of the sulfur atom is indicative that one molecule of the thioester carrier SNAC is incorporated, which was indeed confirmed from the ^{1}H and ^{13}C NMR spectra of **1a** (Table 15). The NMR also displays characteristic up-field signals that correspond to the isobutyl side chain of L-Leu. The structure of **1a** was fully elucidated as 5-isobutyl-2-methyl-*1H*-pyrrolyl-3-carboxyl-*S*-NAC. Presence of the substituted pyrrole ring is fully supported by ^{1}H-^{15}N HMBC signals. While a number of pyrrole biosynthetic pathways have

been found in secondary metabolism and mostly involve the four-electron oxidation of L-proline by a dedicated dehydrogenase [258], the synthesis of **1a** from one molecule each of L-Leu and **3** by NRPS325 as proposed in Figure 27 is unprecedented. Following activation and transfer of the leucyl moiety to the T domain, the C domain catalyzes the nucleophilic attack on the β-ketone carbonyl of **3** by the leucyl α-amino group to form a tetrahedral intermediate which can dehydrate to yield the imine **4**. Attack on the β-carbonyl is in sharp contrast to the canonical reaction catalyzed by all NRPS C domains, in which attack of the amine on the thioester carbonyl precedes formation of the peptide bond. The selectivity towards the β position by the nucleophile is analogous to the β-branching reaction catalyzed by 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (HCS)-cassettes of an increasing number of PKSs [259-261]. The NRPS325 R domain is then proposed to catalyze the two-electron reduction of the thioester **4** to yield the aldehyde intermediate **5**, which can readily cyclize and dehydrate to arrive at **1a**.

Table 15. NMR Data for **1a**.



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	-	6.26 (d, 1H, 2.8)	-
2	119.4	-	-
3	132.6	-	-
4	106.6	6.26 (d, 1H, 2.8)	C3, C5
5	130.3	-	-
6	36.6	2.38 (d, 2H, 7.1)	C4, C5, C7, C8, C9, N1
7	28.8	1.80-1.85 (m, 1H)	C8, C9
8, 9	22.2	0.93 (d, 6H, 6.6)	C6, C7, C8, C9
10	13.6	2.50 (s, 3H)	C2, C3, N1
11	186.7	-	-

12	27.5	3.13 (t, 2H, 6.0)	C11, C13
13	40.5	3.50 (dt, 2H, 9.3, 4.6)	C12, C15
14	-	6.51 (br, 1H)	-
15	170.3	-	-
16	23.1	1.97 (s, 3H)	C15

Spectra were obtained at 500 MHz for protons and 125 MHz for carbons and were recorded in $CDCl_3$

Synthesis of **1a** shows that NRPS325 may function in tandem with the upstream PKS module to elaborate a simple diketide precursor into the substituted pyrrolyl-3-carboxyl-*S*-NAC. Interestingly, when acetoacetyl-CoA was introduced in place of **3**, we were unable to detect the synthesis of the corresponding pyrrolyl-3-carboxyl-CoA (Figure 26B, iv). However, addition of free SNAC readily restored the synthesis of **1a,** along with the appearance of a conjugated product **2a** (λ_{max} at 319 nm) (Figure 26B, v;). We therefore hypothesized that a transesterification reaction between acetoacetyl-CoA and free SNAC had occurred to first generate **3** (Figure 27). Indeed, when NRPS325 was incubated with acetoacetyl-CoA and SNAC, we were able to detect the accumulation of **3** in the reaction mixture at a rate significantly higher than that of the spontaneous thioester exchange. Furthermore, while enhancement in thiol-exchange rates can be observed in the presence of a number of free thiols, such as dithiothreitol, mercaptoethanol and 4-mercapto-1-butanol, etc; only background exchange rates were detected with alkylthiols and cysteine. The observed substrate selectivity is therefore evident of an enzyme-facilitated transesterification, most probably catalyzed by the C domain. The activities displayed by NRPS325 towards different free thiols also hint that different thioesters may be incorporated into the final products, as is reported subsequently (Figure 28).

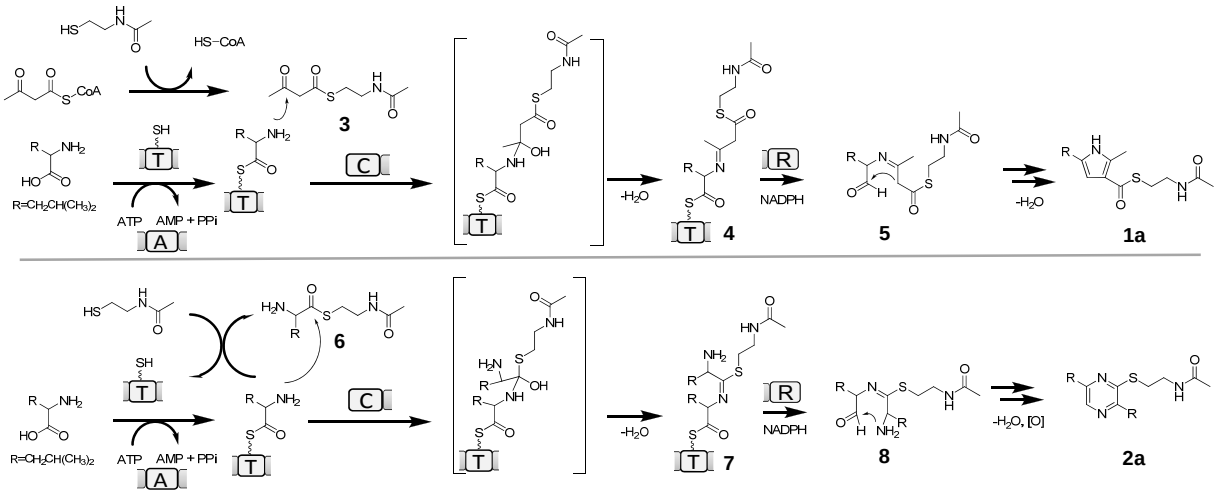
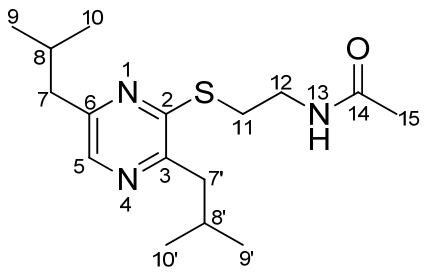


Figure 27. Proposed biosynthetic pathways of compounds **1a** and **2a**.

The molecular formula of **2a** was determined to be $C_{16}H_{27}N_3OS$ and is also suggestive of a SNAC-containing compound. We found that synthesis of **2a** is not dependent on the polyketide precursor, as removal of acetoacetyl-CoA resulted in the exclusive synthesis of **2a** in high yield (Figure 26B, vi). The synthesis of **2a** also requires NADPH (Figure 26B, vii), hinting parallel product release and cyclization mechanisms between **1a** and **2a**. From extensive NMR analysis (Table 16), the structure of **2a** was solved to be the 2-(*S*-NAC)-3,6-diisobutylpyrazine (Figure 27). NRPS325 can therefore synthesize a pyrazine instead of a diketopiperazine from two molecules of l-Leu and SNAC, resulting in the SNAC being attached to C-2 of the pyrazine via an aryl sulfide linkage.

Table 16. NMR Data for **2a**.



No.	^{13}C δ (ppm)	^{1}H δ (ppm) (m, area, J_{HH} (Hz))	HMBC
1	-	-	-
2	153.9	-	-
3	153.6	-	-
4	-	-	-
5	139.0	8.25 (s, 1H)	C3, C6, C7, N1, N4
6	151.7	-	-
7 7’	44.2 33.1	2.56 (d, 2H, 7.1), 2.57 (d, 2H, 7.1)	C5, C6, C8, C9, C10, N1 C3, C2, C8’, C9’, C10’, N4
8 8’	29.1 27.7	2.09 (m, 1H), 2.19 (m, 1H)	C7, C9, C10 C7’, C9’, C10’
9 10	22.5	0.87 (d, 6H, 6.7)	C7, C8, C9, C10
9’ 10’	22.7	0.89 (d, 6H, 6.6)	C7’, C8’, C9’, C10’
11	29.1	3.29 (t, 2H, 6.7)	C2, C12
12	39.4	3.45 (t, 2H, 6.5)	C11, C14, N13
13	-	7.35 (br, 1H)	-
14	169.69	-	-
15	22.8	1.87 (s, 3H)	C14, N13

Spectra were obtained at 500 MHz for protons and 125 MHz for carbons and were recorded in CD_3COCD_3.

The proposed mechanism of pyrazine synthesis is shown in Figure 27. Since two molecules of L-Leu must be activated sequentially by a single A domain, we propose NRPS325 must transfer the first leucyl moiety from the T domain to the free thiol of SNAC to form leucyl-*S*-NAC **6**. In the mixture containing L-Leu, ATP, SNAC and NRPS325, a compound with *m/z* $[M+H]^+$ 233 can indeed be identified using selected ion monitoring, and its RT matched precisely to a chemically synthesized **6** (Figure 27). This transesterification reaction frees up the T domain to be loaded with the second leucyl group, and is consistent with the thiol-dependent formation of **2a**. **6** as an intermediate in the reaction can be further supported through the synthesis of **2a** by NRPS325 in the presence of only **6** and NADPH (Figure 29). In this case, the T domain can be loaded with leucyl through transthioesterification between the pPant arm and **6**,

hence no ATP or free l-Leu is required for activation. Attack of the α-amino group of the second leucyl moiety on the carbonyl of **6** then yields a tetrahedral intermediate, which is dehydrated to afford the ethanimidothioate **7**. This step is reminiscent of the cyclodehydration reactions catalyzed by the cyclization (Cy) domain in some bacterial NRPSs to afford oxazole and thiazole rings [262,263]. Subsequently, reductive release by the R domain and cyclization of aldehyde **8**, followed by dehydration and air oxidation result in the formation of **2a**.

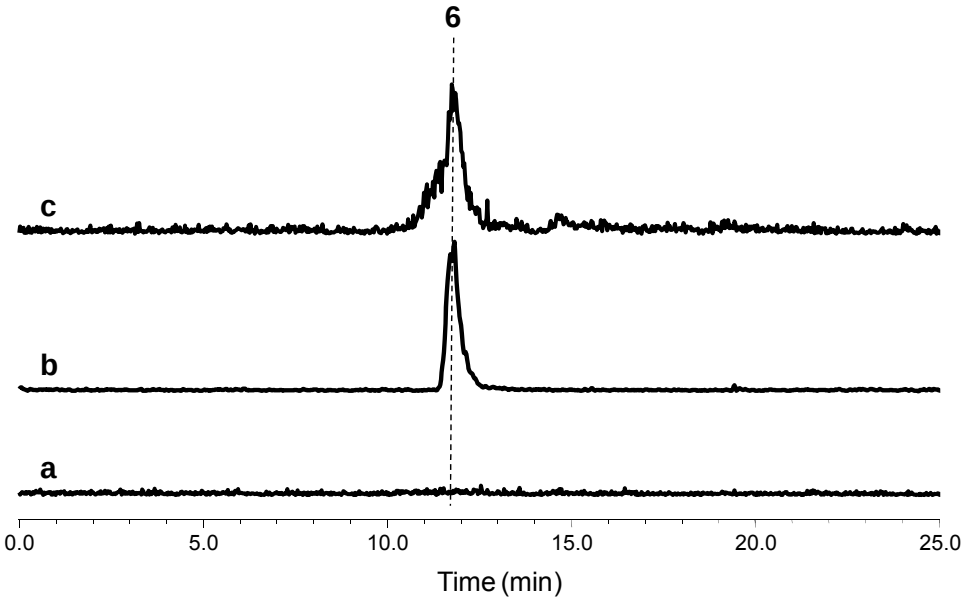


Figure 28. LC-MS traces for the appearance of leucyl-*S*-NAC **6** when NRPS325 is incubated with L-leucine in the presence of ATP and NADPH. (a) Mass filter of m/z $[M+H]^+$ 233 in the assay without NRPS325. (b) Mass filter of m/z $[M+H]^+$ 233 of standard **6**. (c) Mass filter of m/z $[M+H]^+$ 233 in the NRPS325 assay. Authentic standard of leucyl-*S*-NAC **6** was synthesized according to the standard procedure[264]. Reactions (100 µL) were incubated at 25 °C for 12 h with 10 mM L-leucine, 10 mM ATP and 2 mM NADPH in the presence or absence of 10 µM NRPS325. The reaction mixtures were quenched and extracted in pure EA and analyzed by LC-MS.

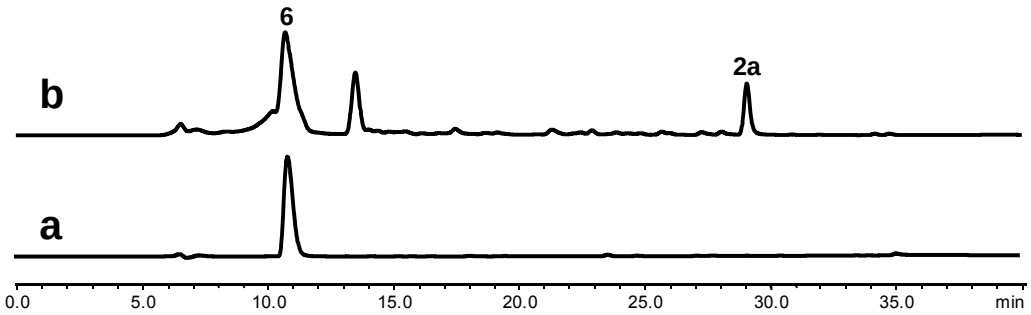


Figure 29. LC-MS traces for the turnover of leucyl-*S*-NAC **6** by NRPS325. (a) Compound **2a** was not produced in the control assay without NRPS325. (b) Production of compound **2a** in the assay containing NRPS325 and NADPH. Reactions (100 μL) were carried out at 25 °C with 10 mM leucyl-*S*-NAC **6**, 2 mM NADPH in the presence or absence of 10 μM NRPS325. The reactions were incubated for 12 h and extracted with pure EA and analyzed by LC-MS.

To gain further insights into the unusual mechanism of thiopyrazine synthesis as shown in Figure 27, the roles of the individual domains were probed. First, *apo*-NRPS325 lost the ability to produce **2a**, confirming the dependence on the pPant arm of the loaded T domain. The C domain of NRPSs catalyzes the canonical C-N bond formation, and therefore should play a role in the formation of the tetrahedral intermediate in the proposed pathway [246]. The two histidine residues located within the signature motif of C domain HHxxxDG were mutated to Ala separately [265,266]. Both H193A and H194A mutants of NRPS325 were impaired in the synthesis of thiopyrazine compounds, as the apparent rate of **2a** formation was <10% of the wild type NRPS325 (Figure 28). A comparable ~13-fold reduction in the rate of **2a** synthesis was also observed when we truncated the C domain and used the ATR tridomain (Figure 22) for the synthesis of **2a**. Therefore, the proposed nucleophilic attack of the free amine leucyl-*S*-T on the carbonyl of **6** can take place spontaneously, but its rate can be significantly enhanced in the presence of the C domain. The C domain may achieve rate enhancement through favored binding of the two substrates. Similar observations in reduction in C-N bond formation rates upon C domain inactivation were also observed in the vibriobactin biosynthetic enzymes VibF (C2) and the free-standing VibH C domain [267,268].

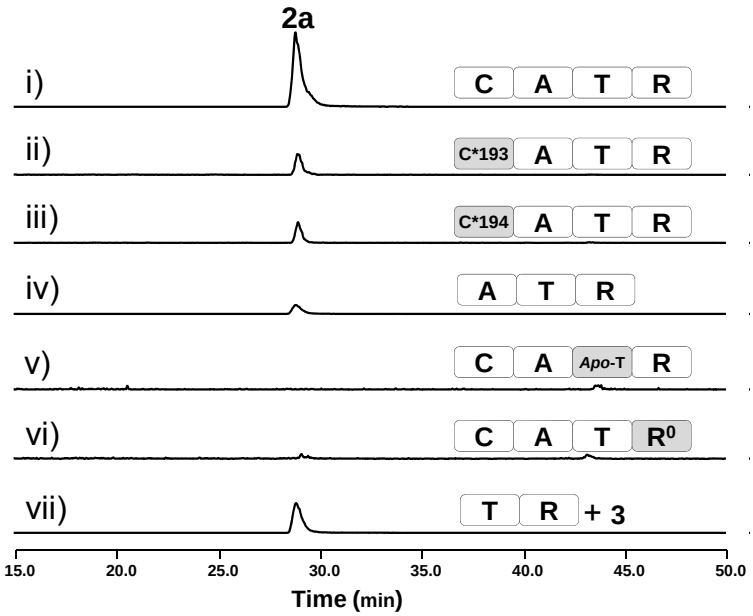


Figure 30. Synthesis of **2a** by intact NRPS325 or dissociated NRPS domains. Traces shown are the selected ion monitoring of desired ion (m/z = 310) in the positive ionization mode. Intensity of the peaks in *vii* was amplified on purpose by 100-fold for clear presentation. Products recovered from in vitro assay with i) NRPS325, ii) NRPS mutant 193, iii) NRPS325 mutant 194, iv) NRPS325-G1132A-G1135A, v) *apo*-NRPS325 in the presence of L-Leu, NAC, ATP and NADPH. The final concentrations of the different components, when added, were: 10 µM enzyme; NAC: 5 mM; L-Leu: 10 mM; NADPH: 2 mM; and ATP: 20 mM. All reactions were performed at room temperature for 12 hr in phosphate buffer (pH=7.4).

Clearly, the proposed reductive release of aldehyde **8** by the R domain is a critical requirement for thiopyrazine formation. The R domain contains the intact catalytic triad Ser-Tyr-Lys and the well-conserved NADPH binding site GxxGxxG found in short chain dehydrogenase/reductase.[5b, 10, 11] Mutation of the GxxGxxG motif to GxxAxxA completely abolished the production of **2a** (Figure 28). The requirement of NADPH by NRPS325 was also monitored spectrometrically at 340 nm (Figure 30). Consumption of NADPH was only observed in the presence of all the required building blocks, including amino acid (L-Leu), ATP and the free thiol (NAC). Only background change in absorbance at 340 nm was observed in the absence of the free thiol, thereby excluding the possibility of direct reduction of aminoacyl-*S*-T by the R domain. The reductive release of **7** observed here is consistent with the proposed role

of the R domain in the synthesis of isoflavipucine [252]. A number of R domains in fungal PKS-NRPSs were previously identified to lack reductive function and instead catalyze Dieckmann condensation to form tetramic acids. As expected, no trace of thiopyrazines was detected when the NRPS modules from ApdA and CpaS were assayed as standalone enzymes. Lastly, formation of **2a** by the truncated TR didomain in the presence of **6** and NADPH (Figure 28) suggests that the R domain may also be involved in the dehydration of the tetrahedral intermediate to form the ethanimidothioate, instead of the thermodynamically favored peptide bond. However, this putative new function of the R domain remains unverified.

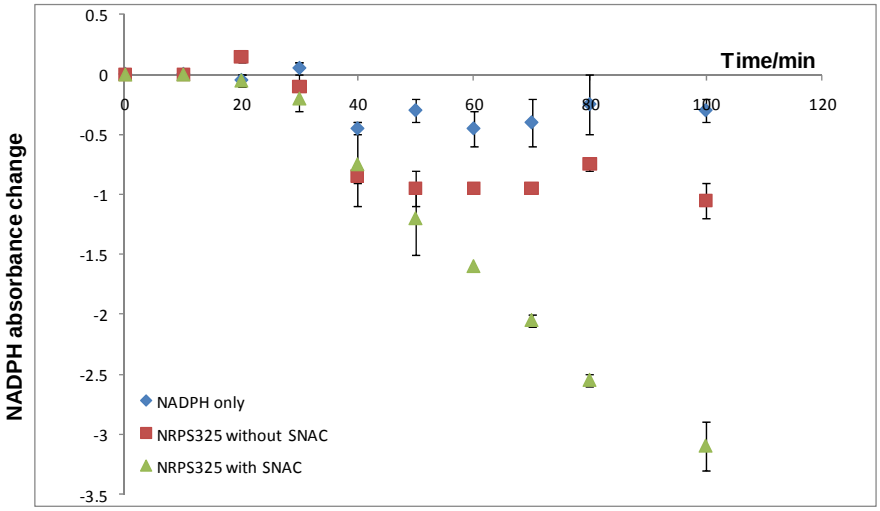


Figure 30. NADPH detection assays**.** 10 µM of NRPS325 was incubated with 10 mM L-Leucine, 10 mM $MgCl_2$, 20 mM ATP and 2 mM NADPH in the presence (green) and absence (red) of 5 mM NAC in phosphate buffer. 2 mM NADPH in phosphate buffer was also monitored and served as a control (blue). Reactions initiated with addition of enzyme and record every 10 min.

Given the broad substrate specificities of the A domain in activating different amino acids, the C domain in performing transesterification using different free thiols, and the remarkable ability of the ensemble to catalyze the synthesis of two different heterocycles with nearly equal efficiency, we tested the biocatalytic prowess of NRPS325 in the synthesis of a library of pyrrolyl-3-carboxyl thioesters and trisubstituted pyrazines. Using different combinations of amino acids, free thiols and β-ketoacyl substrates, we showed that 126 different compounds can

be synthesized by this single NRPS in good yields (Table 17). A subset of this is shown in Figure 31, in which eight different amino acids, four different thiols, and acetoacetyl-CoA were combinatorially mixed to produce 64 compounds in vitro. Notably, the unnatural amino acids trifluoroleucine (Tfl) and azidohomoalanine (Aha) were each incorporated into both the pyrrole and pyrazine scaffolds efficiently.

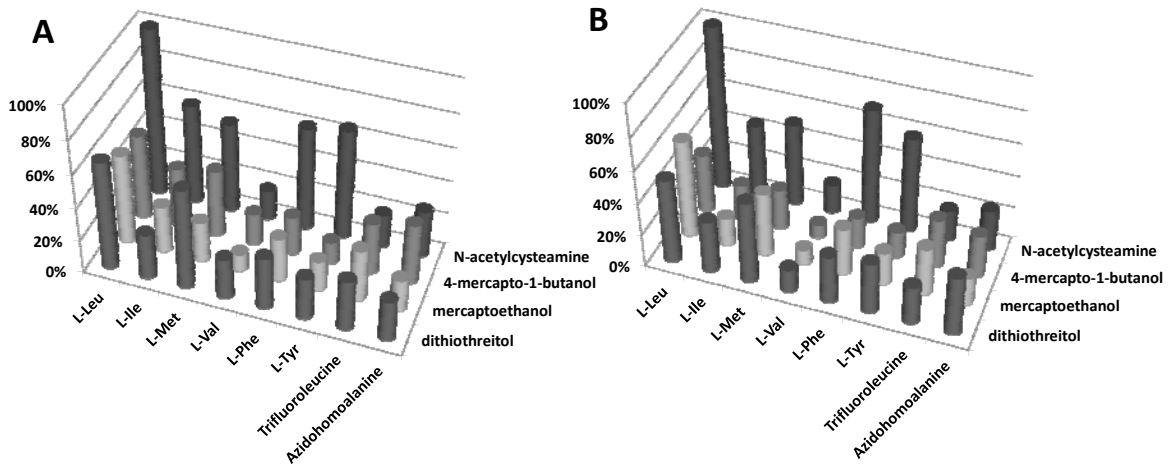


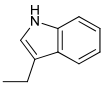
Figure 31. Comparison of the yields of pyrrolyl-3-carboxyl thioesters and pyrazines synthesized by NRPS325. (**A**) A subset of pyrrole products that are synthesized by NRPS325 using 8 different amino acids and 4 different free thiols. The percentages indicate the relative yields normalized to that of **1a**. (**B**) A subset of pyrazine products that are synthesized by NRPS325. The percentages indicate the relative yields normalized to that of **2a**.

In summary, our genome mining efforts have resulted in the discovery of a highly flexible enzyme that can produce compounds that are previously not isolated as natural products. Although the natural role of NRPS325 in *A. terreus* is unknown without reconstituting the entire ATEG00325 megasynthase, our work here highlights the power of heterologous expression as a method to awake and harvest the biocatalytic power of silent biosynthetic enzymes.

Table 17. Matrices of compounds biosynthesized by NRPS325.

	N-acetylcysteamine R_2=$CH_2NHCOCH_3$	dithiothreitol (DTT) R_2= $CH(OH)CH(OH)CH_2SH$	mercaptoethanol R_2=CH_2OH	4-mercapto-1-butanol R_2= $CH_2CH_2CH_2OH$
L-Leu R_1=$CH_2CH(CH_3)_2$	**1a**, 100% **2a**, 100%	**9a**, 65% **10a**, 52%	**11a**, 54% **12a**, 60%	**13a**, 51% **14a**, 36%
L-Ile R_1=$CH(CH_3)CH_2CH_3$	**1b**, 60% **2b**, 45%	**9b**,27% **10b**, 32%	**11b**, 28% **12b**, 17%	**13b**, 36% **14b**, 22%
L-Met R_1=$CH_2CH_2SCH_3$	**1c**, 54% **2c**, 51%	**9c**, 60% **10c**, 50%	**11c**, 24% **12c**, 39%	**13c**, 40% **14c**, 25%
L-Val R_1=$CH(CH_3)_2$	**1d**, 18% **2d**, 18%	**9d**, 24% **10d**, 14%	**11d**, 10% **12d**, 9%	**13d**, 19% **14d**, 8%
L-Phe R_1=CH_2Ph	**1e**, 62% **2e**, 71%	**9e**, 31% **10e**, 29%	**11e**, 26% **12e**, 29%	**13e**, 23% **14e**, 19%
L-Tyr R_1=CH_2PhOH	**1f,** 66% **2f**, 59%	**9f**, 25% **10f**, 31%	**11f**, 18% **12f**, 20%	**13f**, 13% **14f**, 16%
Trifluoroleucine R_1=$CH_2CH(CH_3)CF_3$	**1g**, 20% **2g**, 19%	**9g**, 30% **10g**, 23%	**11g**, 31% **12g**, 29%	**13g**, 31% **14g**, 30%
Azidohomoalanine R_1=$CH_2CH_2N_3$	**1h**, 28% **2h**, 25%	**9h**, 24% **10h**, 36%	**11h**, 19% **12h**, 17%	**13h**, 36% **14h**, 26%

[a]: The percentages in parenthesis indicate the relative yields of pyrroles and pyrazines normalized to the yields of **1a** and **2a**, respectively.

[b]: C_9H_8N= 

2.3. Identification and Engineering of the Cytochalasin Gene Cluster from Aspergillus clavatus NRRL 1[3]

2.3.1. Introduction

Cytochalasins are a group of polyketide-amino acid hybrid compounds belong to the cytochalasan family of fungal secondary metabolites, which have significant commercial and research values due to their diverse arrays of biological activities and complex molecular structures (Figure 32) [269]. The cytochalasans are characterized structurally by their tricyclic core, which consists of a macrocyclic ring fused to an isoindolone moiety derived from a higly-reduced polyketide backbone and an amino acid (phenylalanine for cytochalasins) [269,270]. To date, over 80 different cytochalasans have been isolated from a number of the fungal genera, including *Aspergillus*, *Phomopsis*, *Penicillium*, *Zygosporium*, *Chaetomium*, *Rosellinia*,

[3] Compounds are numbered independently in this section.

Metarrhizium, etc [271]. Many cytochalasins, such as the earliest isolated cytochalasin A and B [272], are capable of inhibiting the polymerization of actin and are thus widely used as tools in studying the division and motility of mammalian cells [273]. Besides the well-known actin binding characteristics, cytochalasins are also recognized for their other biological activities. For example, cytochalasin A and B were reported to repress glucose transport in human erythrocytes membrane [274]; cytochalasin D was shown to be a reversible inhibitor of protein synthesis in HeLa cells [275] and derivatives of cytochalasin H can regulate plant growth [276]. Cytochalasin E, the molecule of interest in this study, was shown to display strong anti-angiogenic activities (Figure 32) [277]. The remarkable structural complexity of cytochalasin E and the potent biological activities make this molecule an interesting target for biosynthetic study.

Previous isotope labeling studies have revealed the mixed malonate and amino acid origin of cytochalasans [278-280], as well as the source of oxygen atoms in these molecules [281,282]. By contrast, the genetic and molecular basis for cytochalasan biosynthesis have only been revealed recently in the context of the (indol-3-yl)methyl-bearing cytochalasan, chaetoglobosin A, from *Penicillium expansum* (Figure 32) [283]. Using RNA-silencing, the biosynthesis of chaetoglobosin A was shown to involve the *cheA* gene that encodes for a hybrid iterative type I polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS), and a biosynthetic pathway that could be generalized to other cytochalasans was proposed based on the *che* gene cluster.

To date, the *che* gene cluster responsible for chaetoglobosin A biosynthesis remained the only example of a gene cluster that encodes for cytochalasan production [283]. CheA is among the several fungal PKS-NRPS pathways that have been identified and characterized during the last few years, as exemplified by equisetin [128], aspyridone A [129,130], pseurotin A [131], cyclopiazonic acid [132,133], tenellin [134] and etc. The PKS module of the PKS-NRPS

responsible for the synthesis of a polyketide chain typically consists of several catalytic domains including ketosynthase (KS), malonyl-CoA:ACP transacylase (MAT), dehydratase (DH), methyltransferase (MT), enoylreductase (ER), ketoreductase (KR), and acyl-carrier protein (ACP), arranged in an assembly-line fashion from the N- to C-terminus. A downstream NRPS module, with the canonical set of condensation (C), adenylation (A) and thiolation (T) domains, amidates the carboxyl end of the polyketide with a specific amino acid. Typically, a reductase-like (R) domain is typically found at the C-terminus and can release the PKS-NRPS products via either a Dieckmann cyliczation reaction [133,284] or as an aldehyde in a NADPH dependent fashion [143].

Both chaetoglobosin A and cytochalasin E contain a substituted perhydroisoindolone scaffold fused with a macrocyclic ring that is the hallmark of cytochalasans, which is proposed to be derived from an intramolecular Diels-Alder reaction of the PKS-NRPS product following its release and formation of the pyrrolinone dienophile [283]. Unlike chaetoglobosin A however, cytochalasin E is derived from a shorter polyketide chain (octaketide instead of nonaketide), a different amino acid building block (phenylalanine instead of tryptophan), and contains a unique vinyl carbonate moeity, which all warrants further biosynthetic investigation at the molecular genetics levels.

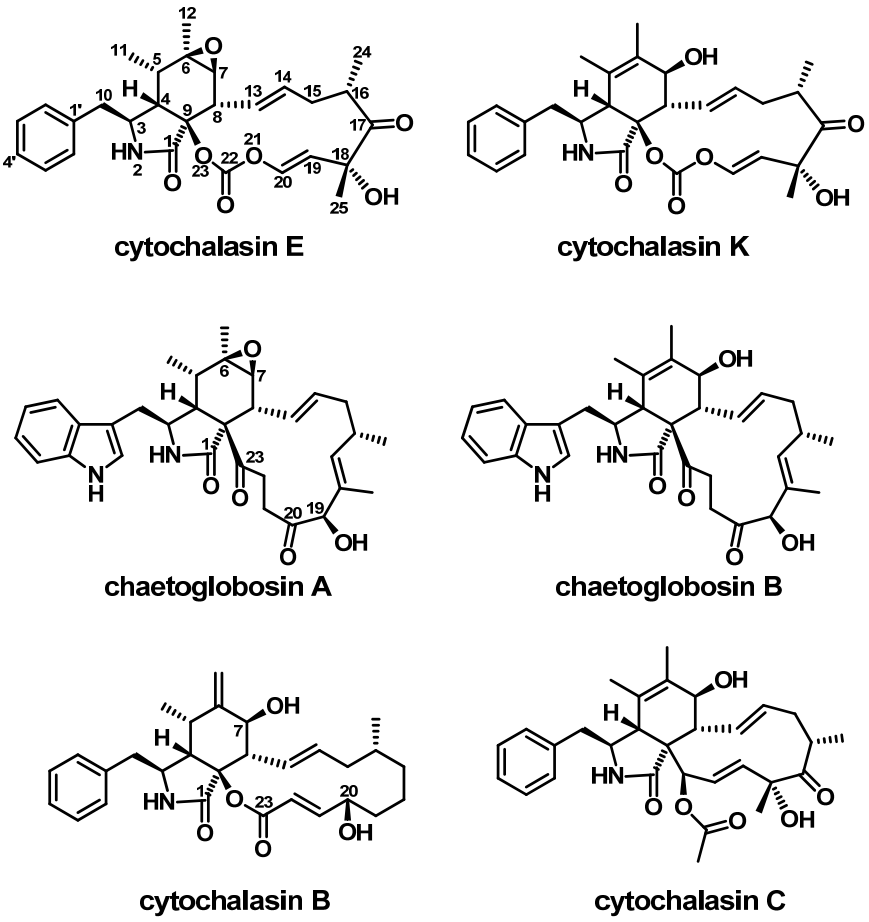


Figure 32. Chemical structures of selected cytochalasans.

In this work, we report the discovery of the *ccs* gene cluster involved in the biosynthesis of cytochalasin E and K from *A. clavatus* NRRL 1 by genome mining. Involvement of the PKS-NRPS (CcsA) was confirmed by gene disruption. Bioinformatic analysis of the genes encoded in the *ccs* gene cluster revealed insights into the biosynthesis of the unique features present in cytochalasin E and K. With the genetic blueprint in hand, we were able to significantly increase the titer of the cytochalasin products through overexpression of the pathway-specific regulator CcsR.

2.3.2. Materials and Methods

2.3.2.1. Strains and Culture Condition

The *A. clavatus* NRRL1 obtained from the Agriculture Research Service (NRRL) Culture Collection was used as the parental strain in this study. The wild type and mutant strains were maintained on Potato Dextrose Agar (PDA). For sporulation, wild type *A. clavatus* was grown on malt extract peptone agar (MEPA) (30 g/L malt extract, 3 g/L papaic digest of soybean meal and 15 g/L agar) for 3 days at 25 ºC. *Escherichia coli* strain XL1-Blue (Stratagene) and *E. coli* TOPO10 (Invitrogen) were used for cloning.

2.3.2.2. Analyses of Genome Sequence of *A. clavatus* NRRL1

The genome sequence of *A. clavatus* NRRL1 was obtained from NCBI database [285]. Gene predictions were performed via FGENESH program (www.softberry.com) and manually checked based on homologous gene/protein sequences in the GenBank database. Protein domain functions were deduced using Conserved Domain Search (NCBI).

2.3.2.3. DNA Manipulation and Construction of Plasmids

Table 18. Sequences of primers used in the construction of plasmids in section 2.3.

Primers	*Sequences*[a]
ccsA-1	5'- TTGACTGTGTTCTAACAGACTTGAGAGGG -3'
ccsA-2	5'- AATCCAGTCACTCGAGCAGCAGCG -3'
ccsA-3	5'- GTCCTGCCCGTCACCGAGATTTAGCACTCATCACCCCGACATACACGGCC -3'
ccsA-4	5'- GGCCGTGTATGTCGGGGTGATGAGTGCTAAATCTCGGTGACGGGCAGGAC -3'
ccsA-5	5'-TGGCAATACACCAAAGGCCTAATCTCGAGA-3'
ccsA-6	5'- AAGTGCTCCTTCAATATCATCTTCTGTCGCTCGGCCGCGTCCGAAACTGC -3'
ccsA-7	5'- CAATCAAGCCGCGGGCTATGCCAAGCTTA -3'
ccsA-8	5'- GGATTGGATGGCTGCATCCAGAGTGGC -3'
ACLA_078640-f	5'- AA*GATATC***ATG**GATCTTTACCGTCGAAGTGCCTGTGAT -3'
ACLA_078640-r	5'- TT*CTCGAG*TCATTTATGAACTGTCAACTCTGCTCTGAT -3'
ACLA_078740-f	5'- AA*GATATC***ATG**TATCATGCATCCATTGCCCT -3'
ACLA_078740-r	5'- TT*CTCGAG*TTATAATTGTATCCCTTTCTTACACC -3'

[a] The introduced restriction sites are shown in italics. The start codons are in bold.

High molecular weight genomic DNA of *A. clavatus* NRRL1 was prepared according to the protocol described previously [286]. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs). PCR was performed using Platinum Pfx DNA polymerase (Invitrogen). Sequences of PCR products were confirmed by DNA sequencing (Laragen, CA). The plasmids pBARKS1 and pBARGPE1 [287] were obtained from the Fungal Genetics Stock Center (FGSC). The gene-specific primers in this work are listed in Table 18. The selection marker *bar* gene with *trpC* promoter was amplified from the plasmid pBARKS1. Construction of the knock-out cassette for *ccsA* gene was achieved using the fusion PCR method described previously [288], and cloned into pCRblunt (Invitrogen) vector (Table 19). PCR was used to produce up to 10 μg DNA for fungal transformation. For regulator overexpression, the *ccsR* gene was amplified using PCR and digested with *EcoR*I/*Xho*I. The vector was prepared by digesting pBARGPE1 with the same set of restriction enzymes. The digested *ccsR* gene product was ligated into vector pBARGPE1 to give pKJ150 (Table 19). The other overexpression plasmid for the regulator ACLA_078740 (pKJ151) was cloned using the same strategy as pKJ150.

Table 19. Plasmid constructs used in this study.

Plasmid	***Vector Source***	***Genes***	***Marker***	***Reference***
pKJ145	pCRblunt vector	*ccsA knock-out cassette*	*Kan*	This work
pKJ150	pBARGPE1	*ACLA_078640*	*Amp, Bar*	This work
pKJ151	pBARGPE1	*ACLA_078740*	*Amp, Bar*	This work

2.3.2.4. DNA Transformation of *A. clavatus*

Preparation and transformation of *A. clavatus* protoplasts were performed as described previously [289]. Transformants were selected on glucose minimal medium agar supplemented

with glufosinate (8 mg/mL) and sorbitol (1.2 M) as osmotic stabilizer. Miniprep genomic DNA from *A. clavatus* transformants was used as templates for PCR screening of gene deletants and was prepared as described for *Aspergillus nidulans* [286].

Table 20. *Aspergillus clavatus* strains used in this study.

Strain	*Genotype*	*Reference*
***A. clavatus* NRRL1**	Parental cytochalasin E/K producer	[290]
A. clavatus ΔccsA-3	Δ*ccs*A	This work
A. clavatus T2	*Ccs*R overexpressed	This work
A. clavatus ACLA_078740 KI	ACLA_078740 overexpressed	This work

2.3.2.5. SouthernBlot Hybridization

High molecular weight genomic DNA (10 µg) was digested with restriction enzyme *Pvu*II, separated onto 0.8% agarose gel and blotted onto the positively charged nylon membranes (Roche Applied Science). For Southern blot analysis, the DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche Applied Science) was used. Hybridizations were carried out following the manufacturer's protocol, except using 0.4% NaOH as transferring buffer instead of 20 x SSC buffer.

2.3.2.6. Chemical Analysis and Characterization of Compounds from *A. clavatus*

For small scale analysis, wild type *A. clavatus* and transformants were grown in stationary liquid surface culture as a surface mat on 100 x 15 mm Petri dishes with 5 mL malt extract peptone (MEP) medium for 4 days at 25ºC. The cultures were extracted with equal volumes of ethyl acetate (EtOAc) and evaporated to dryness. The dried extract was dissolved in methanol and analyzed by liquid chromatography mass spectrometry (LC-MS). LC-MS was conducted with a Shimadzu 2010 EV liquid chromatography mass spectrometer by using both positive and

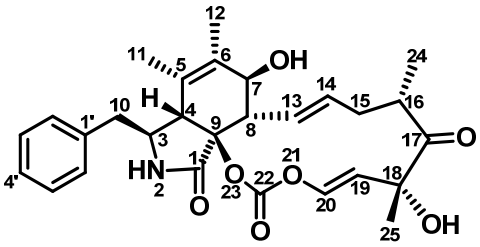
negative electrospray ionization, and a Phenomenex Luna 5 µm 2.0 x 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5 to 95% CH_3CN (*v/v*) in H_2O supplemented with 0.05% (*v/v*) formic acid at a flow rate of 0.1 mL/min. The identity of cytochalasin E was confirmed by comparing the UV spectra, retention time and *m/z* value to the authentic standard (Sigma-Aldrich). To purify cytochalasin K for structural analysis, wild type *A. clavatus* was grown in stationary liquid surface culture condition in two liters MEP liquid medium divided into 20 large 150 x 15 mm Petri dishes for 7 days at 25 ºC. The resulting mycelial mats along with the culture medium were pooled together and extracted three times with equal volume of EtOAc. The organic extracts were combined and evaporated to dryness, redissolved in methanol, and purified by reverse-phase HPLC (XTerra Prep MS C18 5 µm, 19 mm × 50 mm) on a linear gradient of 50% to 95% CH_3CN (*v/v*) over 20 min and 95% CH_3CN (*v/v*) further for 15 min in H_2O at a flow rate of 2.5 mL/min. The eluent was extracted with EtOAc, and dried *in vacuo* to give cytochalasin K as a pure solid (approximate yield of 18 mg/L). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker 500 MHz spectrometer using pyridine-*d5* as solvent. For comparison of the cytochalasin E production between wild type and *ccsR*-overexpressing strains, the same stationary liquid surface culture condition as above was employed; while the submerged shaking flask liquid cultures were performed in two liter flasks containing 500 mL MEP medium (250 rpm, 25 ºC).

2.3.3. Results and Discussion

2.3.3.1 Identification of a Cytochalasin Biosynthetic Gene Cluster from Genome-wide Analysis of PKS-NRPS Genes from *A. clavatus* NRRL 1

Both cytochalasin E and K have been reportedly isolated from the *A. clavatus* species [291,292]. The production of cytochalasin E in the specific *A. clavatus* NRRL 1 (CBS 513.65) strain of which the genome has been sequenced [285], has also been detected in a chemotaxanomic study [293]. To confirm the production of cytochalasins in *A. clavatus* NRRL 1, the strain was grown in stationary liquid surface culture in MEP medium for 4 days. LC-MS analysis of the ethyl acetate extract of the culture revealed two primary cytochalasin compounds with identical molecular weight (*m/z* 518 $[M + Na]^+$) and UV absorption (λ_{max}=235 nm) (Figure 35). The compound that eluted at retention time (RT) of 27 min matched with the authentic standard of cytochalasin E. To confirm the identity of the metabolite with RT = 25 min, the compound was purified from a large scale stationary liquid surface culture of *A. clavatus* NRRL1 with an approximate yield of 17 mg/L. The ^{1}H NMR spectrum is consistent with that previously reported for cytochalasin K [292,294] (Table 21). Cytochalasin K is an isomer of cytochalasin E in which the 6,7-epoxide is isomerized to a C7 hydroxyl and a C5-C6 double bond, therefore both compounds are expected to be originated from the same gene cluster.

Table 21. NMR Data for cytochalasin K[a].

	 Cytochalasin K	
No.	^{1}H δ (ppm) (m, J_{HH} (Hz))	
	cytochalasin K from *Spicaria elegans*	cytochalasin K from *A. clavatus*
1	-	-

2	9.84 (br, s)	9.83 (br, s)
3	3.84 (br, t, 6.9)	3.84 (br, tr, 6.9)
4	4.31 (br, s)	4.31 (br, s)
5	-	-
6	-	-
7	3.48 (dd, 10.0, 9.6)	3.48 (dd, 10.0, 9.8)
8	-	-
9	-	-
10a	3.04 (m)	3.05 (m)
10b	2.99 (m)	2.98 (m)
11	1.48 (br, s)	1.48 (br, s)
12	1.85 (br, s)	1.85 (s)
13	6.91 (dd, 15.1, 10.0)	6.91 (dd, 15.0, 10.0)
14	5.51 (ddd, 15.1, 11.0, 3.2)	5.51 (ddd, 15.0, 11.0, 3.2)
15a	2.83 (br, dd, 13.7, 11.0)	2.83 (dd, 13.7, 11.0)
15b	2.05 (br, d, 13.7)	2.06 (br, d, 13.7)
16	3.02 (m)	3.01 (m)
17	-	-
18	-	-
19	6.08 (dd, 11.9, 2.3)	6.07 (dd, 11.9, 2.3)
20	7.21 (d, 11.9)	7.22 (d, 11.9)
21	-	-
22	-	-
23	-	-
24	1.00 (d, 6.9)	1.00 (d, 6.9)
25	1.57 (s)	1.58 (s)
1’	-	-
2’, 6’	7.25 (m, 2H)	7.22-7.26 (m)
3’, 5’	7.26 (m, 2H)	7.22-7.26 (m)
4’	7.20 (m)	7.18 (m)

[a] Spectra were obtained at 500 MHz for proton was recorded in pyridine-d5.

Given that the polyketide-amino acid backbone of cytochalasin E is likely biosynthesized by a PKS-NRPS, we searched the sequenced genome of *A. clavatus* NRRL 1 for genes encoding PKS-NRPS with the BLASTP program. Using the chaetoglobosin PKS-NRPS CheA as a query sequence [283], the *A. clavatus* NRRL 1 genome was found to encode four putative PKS-NRPS genes (ACLA_004770, ACLA_077660, ACLA_023380 and ACLA_078660), none of which have been characterized previously (Figure 33 and Table 22). Among these four hits, ACLA_004770 is orthologous to pseurotin A synthetase from *Aspergillus fumigatus* (92%

similarity, 86% identity) [131], while ACLA_023380 is most closely related to equisetin synthetase from *Fusarium heterosporum* (64% similarity, 50% identity) [128]. Due to the significant structural differences of pseurotin A and equisetin to cytochalasin E and K, these two PKS-NRPS loci were considered unlikely to be the desired candidates. Since fungal secondary biosynthetic genes are often clustered together [295], adjacent genes of the remaining two PKS-NRPS candidates were scrutinized for additional clues (Figure 33).

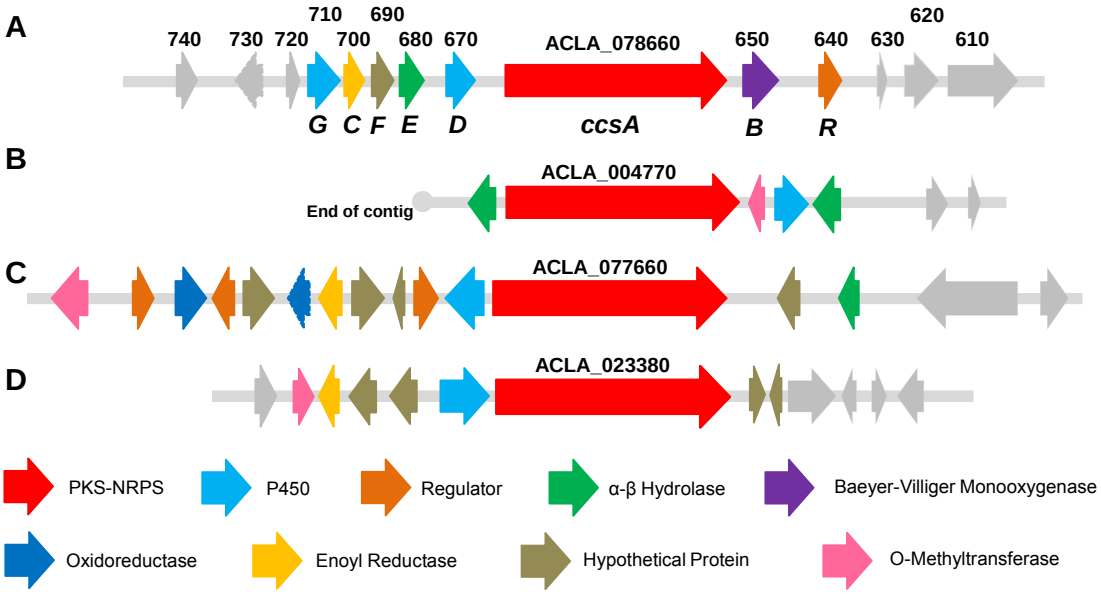


Figure 33. Organizations of the *ccs* gene cluster and other PKS-NRPS gene loci in the genome of *Aspergillus clavatus* NRRL 1. (**A**) The ccs gene cluster. (**B**) PKS-NRPS gene (ACLA_004770) locus. (**C**) PKS-NRPS gene (ACLA_077660) locus. (**D**) PKS-NRPS gene (ACLA_023380) locus.

The vinyl carbonate moiety present in cytochalasins E and K is a unique structural feature not found in other cytochalasans and is rare among known natural products. In a previous feeding study, the successful incorporation of the 13-membered carbocyclic deoxaphomin to yield the 14-membered macrolactone-containing cytochalasin B implied the involvement of an enzymatic Baeyer-Villiger-type oxygen insertion between the C9 bridging carbon and C23 carbonyl [296]. Accordingly, the 13-membered macrocyclic carbonate in cytochalasin E was proposed to be originated from a corresponding 11-membered carbocyclic cytochalasan,

whereby the unusual insertion of two oxygen atoms may occur via two consecutive Baeyer-Villiger oxidations [296].

Table 22. Four PKS-NRPSs from *A. clavatus* NRRL 1.

No.	***Accession No.***	***Closest Homolog***	***Identity/ Similarity***
1	ACLA_004770	Afu8g00540 (Pseurotin A synthetase) *A. fumigatus*	86% / 92%
2	ACLA_077660	ANI_1_540074 (putative PKS-NRPS) *A. niger*	76% / 83%
3	ACLA_023380	AY700570 (Equisetin synthetase) *F. heterosporum*	50% / 64%
4	ACLA_078660	EFZ03354.1 (putative PKS-NRPS) *M. anisopliae ARSEF 23*	62% / 76%

In nature, such Baeyer-Villiger oxidations of ketones are known to be mediated by flavin-containing proteins collectively known as Baeyer-Villiger monooxygenases (BVMOs), which can be classified into type I, type II and type "O" with the type I BVMOs being the most commonly found [297]. Preliminary analysis of the locus containing the PKS-NRPS gene (ACLA_078660) revealed a gene (ACLA_078650) directly downstream that encodes for a putative flavoprotein (Figure 33A), which exhibits homology to the well-characterized type I BVMOs cyclohexanone monooxygenase (CHMO) from *Acinetobacter* sp. and cyclopentanone monoxygenase (CPMO) from *Comamonas* sp. (23 % identity and 24 % identity respectively) [298,299]. Like all type I BVMOs, the flavoprotein encoded by ACLA_078650 contains the conserved FxGxxxHxxxWP fingerprint motif, which serves as a linker that connects the FAD-binding domain to the NADP-binding domain and is important for the catalysis [300,301]. The presence of BVMO afforded a strong indication that this gene cluster may encode the targeted cytochalasin E pathway. Furthermore, two cytochrome P450 oxygenases are encoded by genes in close proximity to ACLA_078660, which are consistent with the required enzymatic installation of the C6-C7 epoxide, the C17 keto and the C18 hydroxyl of cytochalasin E (Figure 33A). In contrast, the remaining PKS-NRPS (ACLA_077660) locus lacks a BVMO candidate in

the vicinity, but encodes for an *O*-methyltransferase that is not required in the biosynthesis of cytochalasins E and K (Figure 33C). Therefore, our genome-wide survey of PKS-NRPS concluded that ACLA_078660 (renamed to *ccsA*) and the enzymes encoded in the corresponding gene cluster (designated as *ccs* cluster) are most likely to be involved in the biosynthesis of cytochalasins E and K.

Table 23. Genes within and flanking the *ccs* gene cluster.

Assigned gene name	Gene locus (*ACLA_*)	Gene size (bp)	Closest characterized homolog	Identity/ Similarity (%)	Conserved domains	Deduced function
-	078610	3980	*Aspergillus fumigates* XP_746424.2*	77/86	HATPase_c[cd00075], Histidine kinase-like ATPases;	Sensor histidine kinase
-	078620	1970	*Neosartorya fischeri* XP_001258671.1*	70/83	Glyco_hydro_15 [pfam00723], Glycosyl hydrolases family 15.	Glucan 1,4-alpha-glucosidase
-	078630	585	n/a	n/a	n/a	n/a
ccsR	078640	1593	*Aspergillus fumigatus* XP_747734.1 *	27/43	GAL4-like Zn(II)2Cys6 binuclear cluster DNA-binding domain.	Transcription regulator
ccsB	078650	1967	*Pseudomonas sp. HI-70* BAE93346	42/57	PLN02172, flavin-containing monooxygenase FMO GS-OX.	Baeyer-Villiger monooxygenase
ccsA	078660	12374	*Fusarium heterosporum* AAV66106.2	42/61	KS-AT-DH-MT-ER^0-KR-ACP-C-A-T-R	PKS-NRPS hybrid
ccsD	078670	1598	*Aspergillus fumigatus* AAW03300.1 (Afu8g00560)	32/52	P450 super family [cl12078], Cytochrome P450.	P450 epoxidase
ccsE	078680	1381	*Metarhizium anisopliae* EFY94436.1*	43/61	Abhydrolase_1[pfam00561], alpha/beta hydrolase fold.	Esterase
ccsF	078690	1190	*Metarhizium anisopliae* EFZ03366.1*	73/83	n/a	Unknown
ccsC	078700	1154	*Aspergillus terreus* 3B6Z_A (LovC)	42/59	Enoyl_reductase_like [cd08249], enoyl reductase of the MDR family	Enoyl reductase
ccsG	078710	1797	*Gibberella moniliformis* CAQ16961.1 (GA-P450-4)	39/55	P450 [pfam00067], Cytochrome P450.	P450 monooxygenase

-	078720	730	*Aspergillus oryzae* XP_001827639.2*	72/88	DsbA_FrnE [cd03024], DsbA family, FrnE subfamily.	Thioredoxin
-	078730	1509	*Neosartorya fischeri* XP_001266605.1*	88/94	Mannitol_dh [pfam01232], Mannitol dehydrogenase Rossmann domain;	Mannitol dehydrogenase
-	078740	1132	*Metarhizium anisopliae* EFY97644.1*	34/49	n/a	Transcription regulator

* No close characterized homologs. Uncharacterized closest homologs based on BLASTP search are shown.

2.3.3.2. Targeted Gene Disruption of the *ccsA* Gene in *A. clavatus* NRRL 1

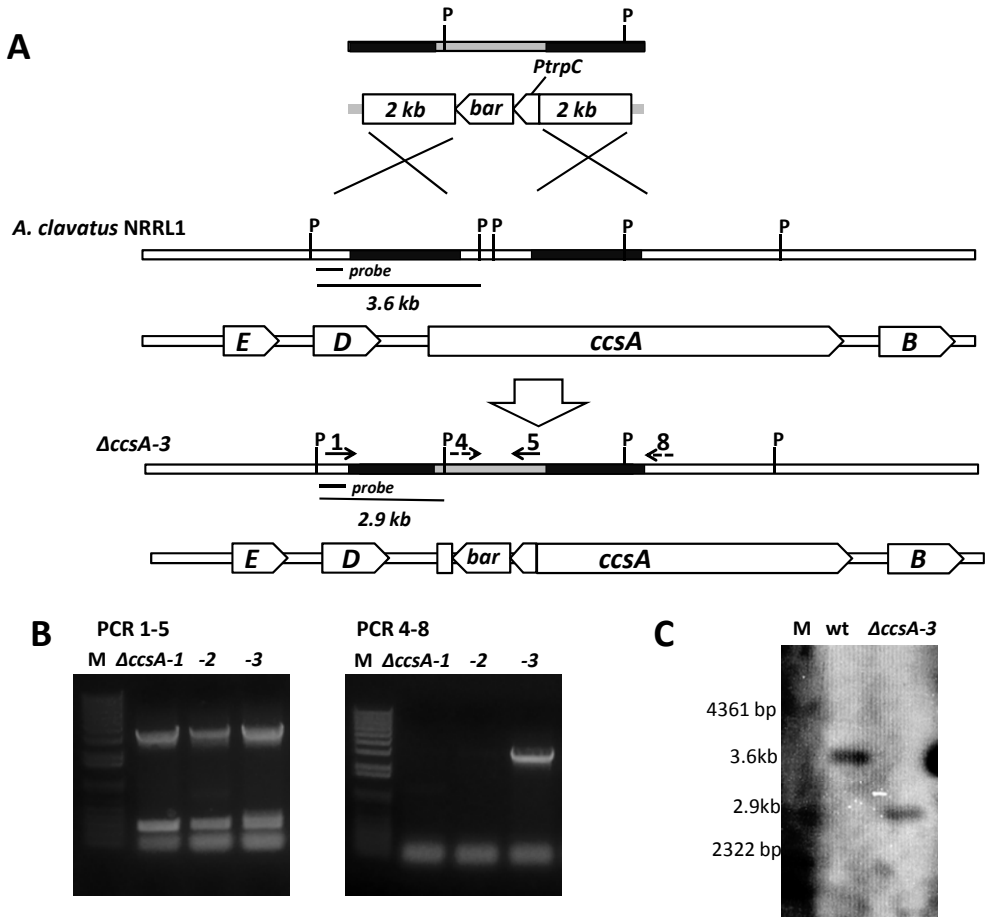


Figure 34. Functional deletion of *ccsA*. (A) Homologous recombination scheme for deletion of *ccs*A gene in *A. clavatus.* P stands for *Pvu*II sites and 1, 4, 5, 8 represent primers ccsA-1, ccsA-4, ccsA-5, ccsA-8 respectively. (**B**) Diagnostic PCR analysis of mutant *ΔccsA-1*, *ΔccsA-2* and *ΔccsA-3* using primer pairs described above. (**C**) Southern blotting analysis of transformant *ΔccsA-3* and wild type *A. clavatus*. Probe positions are shown in (**A**).

To verify the proposed association between the putative *ccs* gene cluster and cytochalasin E biosynthesis, a protoplast-based transformation system for *A. clavatus* and a gene inactivation strategy were developed with the fungal selection marker *bar,* which confers host resistance towards the selection marker glufosinate (Figure 34A) [289]. Successful incorporation of the *ccsA* deletion cassette by double-crossover recombination expects the replacement of an internal 1 kb region of the KS domain of *ccsA* gene with the *bar* resistant marker (Figure 34). Out of 51 glufosinate-resistant transformants, three positive transformants were found to have completely lost the production of cytochalasins E and K by LC-MS screening. The remaining 48 cytochalasin-producing transformants were most likely resulted from the ectopic integration of *bar* gene cassette in the *A. clavatus* genome. Further examination by diagnostic PCR showed that only one (*Δccs*A-*3*) out of the three non-producing mutants yielded PCR products with the band sizes expected for a correct double-crossover recombination (Figure 34B). Disruption of cytochalasins production in *Δccs*A-*1* and *Δccs*A-*2* may have been a result of single crossover insertion of the deletion cassette as indicated by the diagnostic PCR results (Figure 34B). The disruption of *ccs*A by double homologous recombination in *Δccs*A-*3* was further confirmed by southern-blot hybridization of the genomic DNA digested with *Pvu*II with a DIG-labeled probe. As a *Pvu*II site is present in the *bar* resistant cassette, a 2.9 kb band, instead of the 3.6 kb in wild type, is expected to be detected from the integration of the *ccsA* deletion cassette into the correct site (Figure 34C). The loss of cytochalasin E and K production upon disruption of *ccsA* confirmed its essential role in cytochalasin biosynthesis in *A. clavatus* NRRL 1.

2.3.3.3. Overexpression of Pathway-specific Regulator Encoded Gene ACLA_078740 and ACLA_078640

Two genes (ACLA_078740 and ACLA_078640) encoding for putative fungal transcriptional factors with $Zn(II)_2Cys_6$ motif were found in the *ccs* gene cluster in the vicinity of the PKS-NRPS gene *ccsA*. The $Zn(II)_2Cys_6$ binuclear cluster proteins have so far been identified solely in fungi and were demonstrated to play crucial roles in transcriptional regulation [302]. $Zn(II)_2Cys_6$-type fungal transcriptional regulator genes located inside secondary metabolite gene clusters of filamentous fungi has been shown to regulate the secondary metabolic genes in a pathway-specific manner [303,304] and have been used as tools for genome mining [71,129].

To functionally examine the two putative fungal transcription factors adjacent to the PKS-NRPS gene *ccs*A, we cloned both of them into pBARGPE1 vector under the control of *A. nidulans gpdA* (glyceraldehyde phosphate dehydrogenase gene) promoter. The resultant plasmids, pKJ150 and pKJ151 (containing ACLA_78640 and ACLA78740 respectively; Table 19) were randomly integrated into the genome of *A. clavatus* NRRL 1. The transformants selected for glufosinate resistance were analyzed by PCR using primer pair gpdA-f/ACLA_078640-r and gpdA-f/ACLA_078740-r, respectively. All transformants with the correct integration of intact overexpression cassettes were grown in stationary liquid surface culture with MEP medium. LC-MS analyses of the culture extracts showed that overexpression of ACLA_078740 did not affect the growth of *A. clavatus* nor influence the production of cytochalasins, suggesting that ACLA_078740 is not involved in the regulation of cytochalasin biosynthesis. On the other hand, significantly elevated production of both cytochalasin E and K were detected in all four transformants (T2, T6, T10 and T12) carrying the intact ACLA_078640-overexpression cassette,

with no significant effect observed on the cell growth. This established that ACLA_078640 (*ccsR*) is the *ccs* pathway-specific transcriptional regulator and further confirmed that cytochalasin E and K are the products of the *ccs* gene cluster.

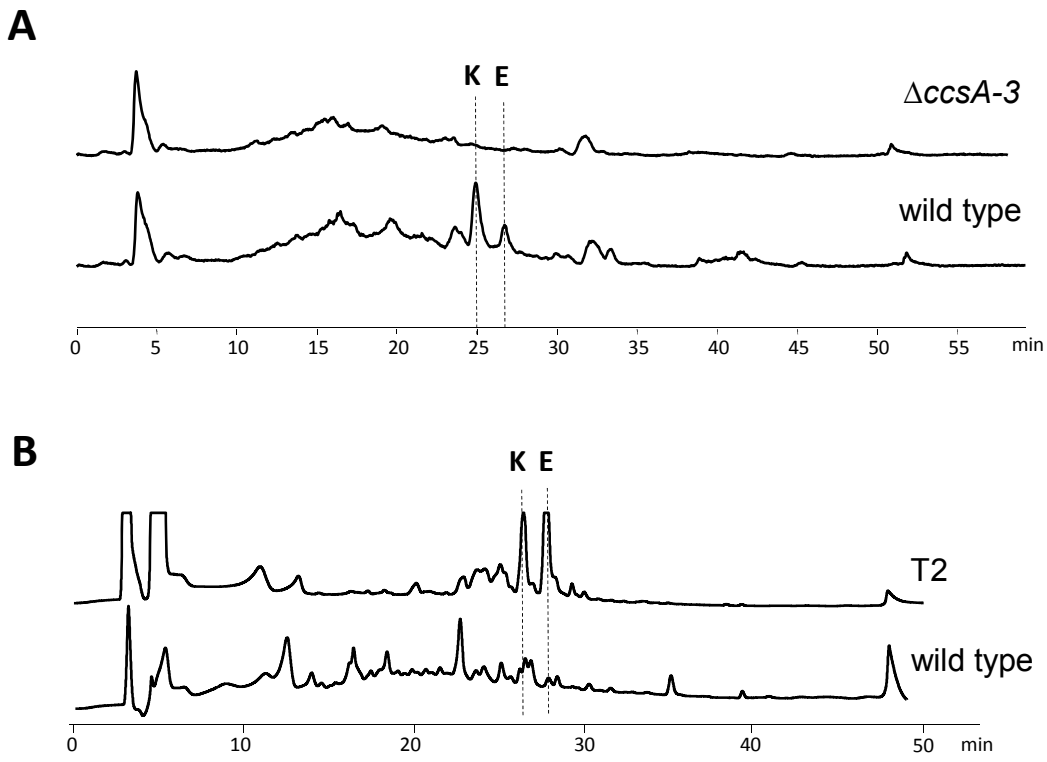


Figure 35. Metabolic extracts of *A. clavatus* strains. (A) LC-MS analysis of metabolites produced by wild type and Δ*ccsA* mutant *A. clavatus* strains. Total ion current chromatogram (*m/z* range 100-800) is shown. (**B**) HPLC analysis (210 nm) of metabolites produced by wild type *A. clavatus* and *ccsR*-overexpressed strain (T2). E and K stand for cytochalasin E and K, respectively.

The titers of cytochalasin E in the four transformants were elevated at least by 3 fold, with T2 as the highest producer of cytochalasin E. The titer of cytochalasin E in a 8-day stationary liquid surface culture of T2 was estimated to be 250 mg/L based on ion current integration of *m/z* 518 $[M + Na]^+$ in LCMS analysis using cytochalasin E standard curve, and a total of 175 mg/L of cytochalasin E was successfully purified. Therefore, the production of cytochalasin E in *ccsR*-overexpressing strain T2 had increased by approximately seven fold compared to the titer of wild

type *A. clavatus* NRRL 1 strain (25 mg/L was isolated under the same culture condition and starting inoculation amount). The variation in the production levels of cytochalasin E are likely due to the different copy number of *ccsR* overexpression cassette present in the individual transformants. We observed during our culture condition testing experiments that cytochalasin E and K were only produced under stationary liquid surface culture condition by wild type *A. clavatus* NRRL 1, but not in submerged shaking flask condition. This is consistent with previous cytochalasin fermentation optimization experiments that utilized solid substrates without submerging the cells in liquid medium, in which a series of grains were used to culture the fungus in agitated flasks for 2 weeks with barley affording the highest yield (35 mg/kg barley) [305]. We observed that overexpression of *ccsR* in *A. clavatus* T2 strain elevated the titer of cytochalasin E in a 6-day submerged shaking flask culture from undetectable levels in wild type to ~35 mg/L (estimated by MS ion current integration). Although the titer of cytochalasin E for T2 in submerged shaking flask culture is significantly lower than in stationary surface liquid culture, the successful production of cytochalasins in submerged culture condition presents significant advantages for fermentation scale-up in industrial stirred bioreactors.

Regulatory genes often clustered together with the biosynthetic genes in microorganisms. However, not all gene clusters contain a pathway-specific regulator, whereas some gene clusters may contain more than one pathway-specific regulator. The pathway-specific regulator can either play a positive (activator) or a negative (repressor) role in regulating the expression of biosynthetic genes within the cluster. Overexpression of pathway-specific activators is considered a useful metabolic engineering strategy to improve the titer of secondary metabolites. In bacterial systems, this strategy was extensively exploited: overexpression of the *Streptomyces* antibiotic regulatory protein (SARP) family has been commonly used to improve the production

of pharmaceutically important secondary metabolites in actinomycetes [306]. In fungi, the $Zn(II)_2Cys_6$ finger domain-containing proteins are the most common transcriptional activators; it has been demonstrated that overexpression of the corresponding activators AflR and CtnR led to upregulated transcription of the pathway genes and increased production of aflatoxin and citrinin respectively [304,307]. More recently, a novel regulator CefR, which contains the similar "fungal_trans" conserved domain in CtnR but lack the $Zn(II)_2Cys_6$ finger domain, was identified in the cephalosporin biosynthetic gene cluster [308]. Overexpression of CefR elevated the cephalosporin C production in the industrial *Acremonium chrysogenum* strain. Therefore, this approach should prove useful for improving production of pharmaceutically important fungal natural products, provided that such a pathway-specific activator exists in the corresponding biosynthetic gene cluster. In the absence of pathway-specific activator, overexpression of bottleneck enzymes in the pathway or increasing precursor and cofactor supply by engineering of primary metabolic pathways are two alternative strategies that may be considered for increasing fungal secondary metabolite production [309,310].

2.3.3.4. Proposed Biosynthetic Pathway for Cytochalasin E/K Biosynthesis

In the *ccs* gene cluster, a total of eight putative genes were predicted to be involved in biosynthesis of cytochalasin E and K (Table 23). At the two boundaries of the *ccs* gene cluster is a gene encoding for P450 monooxygenase (*ccsG*) and a gene encoding for the pathway-specific regulator (*ccsR*) (Figure 34). Immediately upstream of *ccsG* is a thioredoxin-encoding gene (ACLA_078720) and a mannitol dehydrogenease gene (ACLA_078730), which are both highly conserved in the genomes of other *Aspergillus* spp. (Table 23). Downstream of *ccsR*, are ACLA_078630, ACLA_078620 and ACLA_078610, which respectively encode for a

hypothetical protein with no significant similarity to any protein sequence in the GenBank database (maybe a misannotation), a glucan 1,4-α-glucosidase (>67% identity to the homologs in *A. fumigatus* and *N. fischeri*), and a sensor histidine kinase (60-77% identity to homologs in *Aspergillus* spp.). As genes further upstream of *ccsG* and downstream of *ccsR* are mostly close orthologs shared among *Aspergillus* spp. and do not appear to involve in secondary metabolic biosynthesis, we reason that these genes are unlikely to participate in cytochalasin biosynthesis and may involve in primary metabolism or housekeeping roles.

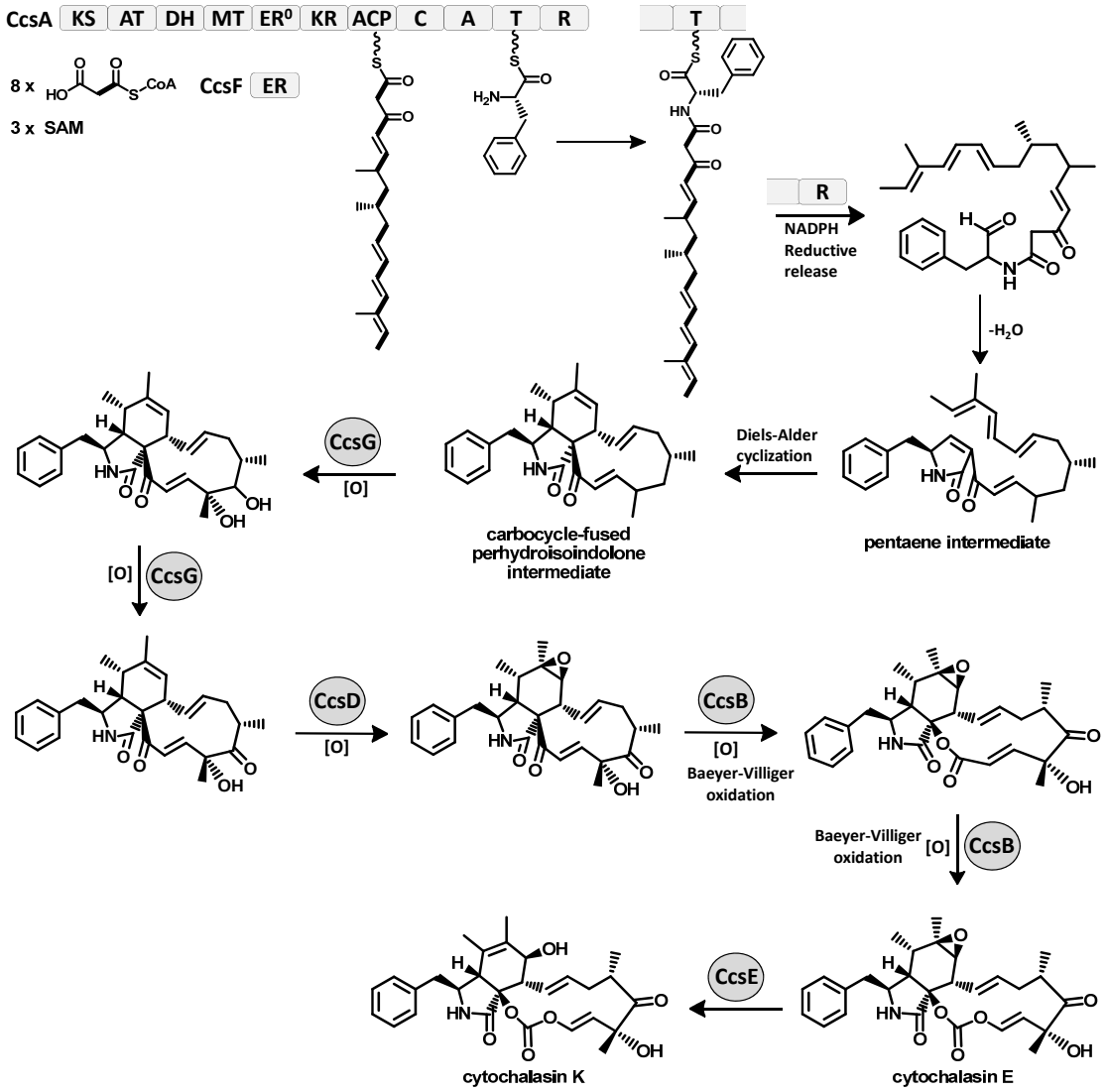


Figure 36. Proposed biosynthetic pathway for cytochalasin E and K.

Based on the previous isotope labeling studies and deduced gene functions of *ccs* gene cluster, the biosynthetic pathway for cytochalasin E and K can be proposed (Figure 36) [269,278]. The domain organization of the PKS-NRPS encoded by *ccs*A is similar to that of other reported PKS-NRPSs genes, including CheA in *P. expansum* (35% identity to CcsA), TenS in *Beauveria bassiana* (33 % identity to CcsA) and ApdA in *A. nidulans* (36 % identity to CcsA) [129,134,283]. Due to the lack of the key NADPH binding motif (LXHXG(A)XGGVG), the ER domain of CcsA is proposed to be inactive. Correspondingly, a dissociated ER (encoded by gene *ccsC*) downstream of *ccs*A is proposed to function *in trans* with the PKS module of CcsA to synthesize a reduced *β*-keto octaketide backbone, analogous to the roles played by TenC and ApdC during the biosynthesis of tenillin and aspyridone respectively [130,134]. Subsequently, one molecule of phenylalanine is selectively activated by the A domain of the NRPS module and transferred to the phosphopantetheinyl (pPant) arm of T domain. C domain then catalyzes the condensation between the nucleophilic amino group of phenylalanine and the electrophilic carbonyl of the upstream nascent octaketide chain to yield the T domain-tethered amide intermediate (Figure 36).

The R domain at the end of the NRPS module is proposed to catalyze the reductive release of the nascent aminoacyl-thioester to give an aminoaldehyde intermediate, which can readily undergo an intramolecular Knoevenagel condensation to yield the 3,5-disubstituted 3-pyrrolin-2-one (Figure 36). The terminal R domains of some PKS-NRPSs were demonstrated to offload the polyketide-amino acid intermediates via a Dieckmann condensation to afford a tetramic acid moiety [133,284]. Under this scenario, the released tetramic acid intermediate from CcsA would have to undergo further keto reduction and dehydration to form the proposed 3-pyrrolin-2-one dienophile. The lack of the required reductase encoded in *ccs* gene cluster is inconsistent with

this possibility, while the reductive release catalyzed by CcsA R domain is supported by the protein sequence analysis. The CcsA R domain harbors an intact conserved NADPH binding motif GXSXXG and the catalytic triad Ser-Tyr-Lys shared among the short chain dehydrogenase/reductase SDR superfamily proteins, whereas the characterized Dieckmann R* domains contain a leucine or phenylalanine in place of the tyrosine in the catalytic triad [133].

The intramolecular [4+2] Diels-Alder *endo* cycloaddition of the pentaene intermediate was proposed to occur between the terminal diene of the polyketide chain and the 3-pyrrolin-2-one dienophile to afford the 11-membered carbocycle-fused perhydroisoindolone scaffold (Figure 36). This proposed enzymatic reaction step, which may involve a so-called "Diels-Alderase", has been shown to be feasible as demonstrated by previous biomimetic synthesis of cytochalasins [282,311], although the identity and enzymatic mechanism of such a "Diels-Alderase" remains enigmatic. In biosynthesis of lovastatin, the Diels-Alder reaction was proposed to occur at the hexaketide stage where the polyketide intermediate is still tethered to the ACP of the PKS LovB [312]. In contrast, the Diels-Alder cycloaddition in cytochalasan biosynthesis is predicted to occur after the release of the hybrid polyketide-amino acid molecule from the PKS-NRPS. Therefore, a discrete post PKS-NRPS tailoring enzyme, such as in solonapyrone and spinosyn A biosyntheses [124,313], may be required for the post PKS-NRPS Diels-Alder reaction in cytochalasin biosynthesis.

Upon formation of the 11-membered carbocycle-fused perhydroisoindolone intermediate, a number of oxidative steps are required to afford the final cytochalasin E and K, including two hydroxylations at C17 and C18, one alcohol oxidation at C17, one epoxidation at C6 and C7 and two Baeyer-Villiger oxidations. Based on the previous ^{13}C-labeled acetate feeding experiments of cytochalasin B pathway in *Phoma exigua* [281], the carbonyl (C-22) on the vinyl carbonate

and the lactam carbonyl (C-1) of cytochalasin E and K are likely to be retained from the acetate units. Previous cytochrome P450 inhibitory experiment in chaetoglobosin-producing fungus *Chaetomium subaffine* showed that the oxygen atoms on the C20 carbonyl, C19 hydroxyl and the 6,7-epoxide are all added by cytochrome P450 enzymes [282]. Similarly, we propose that the oxidative modification at C17, C18 and the C6-C7 epoxidation are likely to be catalyzed by the two cytochrome P450 oxygenases (CcsD and CcsG) found in the *ccs* gene cluster. CcsD exhibits similarity to GliF from *A. fumigatus* (42 % protein identity) that possibly catalyzes the formation of an arene oxide in gliotoxin biosynthesis [314]; hence, CcsD may be responsible for the epoxidation of the C6-C7 double bond. The closest characterized homolog of CcsG is GA-P450-4 (39% identity) from *Gibberella moniliformis,* which is an *ent*-kaurene oxidase that catalyzes the multiple oxidations of *ent*-kaurene to *ent*-kaurenoic acid [315]. This hints that CcsG may be responsible for the successive oxidative modifications at C17 and C18.

The double Baeyer-Villiger oxidations are among the final steps leading to cytochalasin E and K (Figure 36). Multiple sequence alignment of the primary protein sequences of CcsB and previously characterized BVMOs indicated that CcsB contains the two intact conserved Rossmann fold motifs GxGxxG and GxGxxA, as well as the so-called BVMO signature motif FxGxxxHxxxW [297,316]. Compared to the well-known BVMOs that function on smaller cyclic compounds, CcsB exhibits a significantly higher similarity toward the more recently characterized *Pseudomonas* sp. HI-70 cyclopentadecanone monooxygenase (CPDMO) CpdB (41% identity to CcsB) and *Rhodococcus ruber* SC1 cyclododecanone monooxygenase (CDMO) CddA (38% identity to CcsB), which are capable of lactonizing the C_{15} and C_{12} cyclic ketones respectively [316]. This coincides with the proposed role of CcsB being involved in the oxidations of the 11-membered carbocyclic intermediate (Figure 36). Given that there are no

additional gene that encodes for BVMO-like enzyme in the *ccs* gene cluster beside *ccsB*, CcsB may be responsible for the consecutive Baeyer-Villiger insertion of two oxygen atoms into the 11-membered carbocyclic intermediate to generate the unique vinyl carbonate moiety in cytochalasin E and K. A two-step oxidation is proposed: the first oxidation step is consistent with the mechanism in the biosynthesis of cytochalasin B [296], while the second Baeyer oxidation is likely occurred on an acrylate moiety (Figure 36). Examples of such oxygen insertion between a ketone and an *α-β* double bond are rare but was reportedly observed in a microbial steroid degradation study where the A-ring of cholest-4-en-3-one was cleaved presumably via an enzymatic Baeyer-Villiger oxidation [317]. If proven, CcsB may represent the first example of a BVMO that can catalyze such a double Baeyer-Villiger oxidation.

The epoxide-containing cytochalasin E is likely to be the compound that precedes cytochalasin K in the pathway. CcsE, which belongs to the large family of α/β hydrolase, may catalyze hydrolysis of epoxide bond to afford cytochalasin K. The similarity of CcsE to the Afu8g00540 (46% identity) in the pseurotin A pathway, in which the α/β hydrolase was proposed to hydrolyze the epoxide bond on the pyrrolidinone ring, supports the proposed function of CcsE. *ccsF* encodes for a protein that has no significant similarity to characterized proteins and no conserved domains are detected, thus its function has not been assigned. Interestingly, most CcsF homologs seems to be in the vicinity of PKS-NRPS genes in several sequenced fungal genomes, including *Metarhizium anisopliae* ARSEF 23, *Magnaporthe oryzae* 70-15, *Phaeosphaeria nodorum* SN15 and *Chaetomium globosum* CBS 148.51. Therefore, CcsF and its homologs may play a role in post-PKS-NRPS biosynthetic step (possibly Diels-Alder cyclization), resistance or transport of cytochalasins and related PKS-NRPS products.

2.3.3.5. *ccs*-like Gene Clusters in Other Fungal Genomes.

Closer examination of the structures of cytochalasin E and chaetoglobosin A suggests that the reduction, dehydration and methylation steps occur during the initial four polyketide iterations of both CcsA and CheA are identical. Therefore, it is surprising that CcsA shares the lowest head-to-tail protein similarity with CheA (39% identity between the PKS modules) among all the four PKS-NRPSs encoded in the *A. clavatus* NRRL 1 genome. In fact, CheA shares a slightly higher similarity (43% identity) to TenS and DmbS that produce pretenellin A and predesmethylbassianin A, both of which are structurally distinctive from cytochalasans. Furthermore, the tailoring enzymes in the *che* and *ccs* cluster do not appear to share close homology either (Table 2). Therefore, it is reasonable to speculate that the pathway for (indol-3-yl)-methyl-containing cytochalasans may have diverged significantly from the benzyl-containing cytochalasins, or alternatively the two pathways may have arisen convergently.

Table 24. Comparative analysis of *ccs, che* and *ccs*-like PKS-NRPS gene clusters.

***ccs* gene cluster from *Aspergillus clavatus* NRRL 1**		***che* gene cluster from *Penicillium expansum***		***ccs*-like gene cluster from *Metarhizium anisopliae* ARSEF 23**		***CHGG1239* gene cluster from *Chaetomium globosum***	
Assigned gene name/ Gene locus (ACLA_)	**Deduced function**	**Assigned gene name**	**Identity to Ccs homologs* (coverage)**	**Gene locus (MAA_)**	**Identity to Ccs homologs* (coverage)**	**Gene locus (CHGG_)**	**Identity to Ccs homologs ***
ccsA 078660	PKS-NRPS hybrid	cheA	40%	00428	59% (71%)	01239	47%
ccsB 078650	Baeyer-Villiger monooxygenase	n/a	n/a	n/a	n/a	n/a	n/a
ccsR 078640	Transcription activator	n/a	n/a	n/a	n/a	n/a	n/a
ccsD 078670	P450 expoxidase	cheD cheG	21% (40%) 22% (45%)	00429	67%	01243	47%
ccsE	Esterase	n/a	n/a	00441	71%	01246	55%

078680							
ccsF 078690	Unknown	n/a	n/a	00440	73%	01241	53%
ccsC 078700	Enoyl reductase	CheB	39%	00439	68%	01240	49%
ccsG 078710	P450 monooxygenase	cheD cheG	21% (36%) 25% (34%)	00438	51%	01243	41%

* Percentage values represent protein identity to Ccs homologs. Percentage coverage is indicated in parentheses if the protein alignment covers only part of the sequence.

The discovery and identification of *ccs* gene cluster opened up the possibility to use the PKS-NRPS gene *ccsA* and its associated tailoring genes for genome mining of gene clusters of both known and novel cytochalasins. Cytochalasin C and D have been isolated from the entomopathogenic fungus *M. anisopliae* [318,319]. Like cytochalasin E and K, both cytochalasin C and D are derived from octaketide chain and contain a benzyl group originated from phenylalanine but lack the vinyl carbonate moeity (Figure 33). Cytochalasin D in particular, has been used extensively to study cellular processes and is known to impair maintenance of long term potentiation (LTP) of actin filaments [320]. Since the genome of *M. anisopliae* ARSEF 23 has been sequenced recently [321], we examined the genome sequence for candidate cytochalasin gene cluster using CcsA as a query sequence in BLASTP search. A PKS-NRPS gene MAA_00428 was found to share the highest similarity to CcsA among all the PKS-NRPS genes in the GenBank database at the time of query. The corresponding homologs of the post PKS-NRPS tailoring genes (*ccsCDEFG*) in the *ccs* gene cluster can also be found in vicinity of MAA_00428 PKS-NRPS gene (Table 24). As expected, the BVMO *ccsB* homolog is missing in the gene cluster, while a putative acetyltransferase is present. However, the MAA_00428 PKS-NRPS appear to be truncated after the C domain of the NRPS module. Therefore, it is likely that this proposed cytochalasin-producing gene cluster from *M. anisopliae* ARSEF 23

strain may not have the ability to produce cytochalasin C and D due to truncation of the A-T-C domain in a relatively recent mutation event. Although this speculation remains to be proven, it is possible that such a homologous gene cluster with the complete CcsA homolog maybe present in cytochalasin-producing *M. anisopliae* strains. Besides *M. anisopliae*, *ccs* homologs (*ccsACDEFG*) are also found clustered in other fungal genomes, including the chaetoglobosin-producing *Chaetomium globosum* (CHGG_01239, 49% identity) (Table 24) and several plant pathogenic fungi, such as *Magnaporthe grisea* (syn2, 55% identity) and *Phaeosphaeria nodurum* (SNOG_00308, 50% identity), which may responsible for production of cytochalasin-related PKS-NRPS products.

3. Conclusion

In conclusion, a series of core megasynthetases from fungal secondary metabolite biosynthetic pathways were biochemically characterized and the hidden programming rules of fungal PKSs and PKS-NRPS hybrids were investigated. First of all, we were able to reconstitute a tandem PKS system (radicicol pathway) in vitro and in *S. cerevisae* and the mechanism regarding to chain initiation, elongation and product releasing were probed with the assistance of the chemically synthesized tool compounds. Besides, the *S. cerevisae* heterologous host BJ5464-NpgA can be developed to a versatile system for engineered biosynthesis of fungal polyketides, as demonstrated in the combinatorial biosynthesis of halogenated resorcylic acid lactones. Secondly, our work on characterization of NRPS325 uncovers the cryptic capabilities of a fungal NRPS module in the synthesis of thiopyrazines and thiopyrroles, which are vastly different from the recently confirmed natural role of the parent megasynthetase. Such unexpected findings further underscore the untapped biocatalytic potential of megasynthetases from natural product

biosynthetic pathways. In the last project, we have identified the gene cluster for biosynthesis of cytochalasin E and K from the genome of *A. clavatus* NRRL 1. Disruption of *ccs*A provided the direct evidence that the PKS-NRPS gene *ccsA* is essential for the biosynthesis of cytochalasin E and K. Overexpression of the cytochalasin pathway-specific regulator *ccs*R led to a significantly increased cytochalasin production. The detail mechanistic steps of Diels-Alder reaction and the double Baeyer-Villiger oxidations of ketone to carbonate are currently under investigation. The identification of *ccs* gene cluster not only allows for continuous investigation of the molecular basis of the structural diversity generated by fungal PKS-NRPSs, but also opens the door for genome mining of cytochalasin gene cluster and application of molecular engineering to generate novel cytochalasan derivatives.

4. References

1. Newman, D. J., and Cragg, G. M. (2012) *Journal of Natural Products* **75**, 311-335
2. Li, J. W., and Vederas, J. C. (2009) *Science* **325**, 161-165
3. Hamilton, G. R., and Baskett, T. F. (2000) *Can J Anaesth* **47**, 367-374
4. Butler, M. S. (2005) *Nat Prod Rep* **22**, 162-195
5. Cox, R. J. (2007) *Org Biomol Chem* **5**, 2010-2026
6. Hertweck, C. (2009) *Angew Chem Int Ed Engl* **48**, 4688-4716
7. Jenke-Kodama, H., Sandmann, A., Muller, R., and Dittmann, E. (2005) *Mol Biol Evol* **22**, 2027-2039
8. Cane, D. E. (2010) *J Biol Chem* **285**, 27517-27523
9. Cane, D. E., and Walsh, C. T. (1999) *Chem Biol* **6**, R319-325
10. Fischbach, M. A., and Walsh, C. T. (2006) *Chem Rev* **106**, 3468-3496
11. Piel, J. (2002) *Proc Natl Acad Sci U S A* **99**, 14002-14007
12. Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., and Piel, J. (2008) *Nat Biotechnol* **26**, 225-233
13. Shen, B. (2003) *Curr Opin Chem Biol* **7**, 285-295
14. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K. X., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M., and Thorson, J. S. (2002) *Science* **297**, 1173-1176
15. Liu, W., Nonaka, K., Nie, L. P., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) *Chemistry & biology* **12**, 293-302

16. Smith, D. J., Burnham, M. K., Edwards, J., Earl, A. J., and Turner, G. (1990) *Biotechnology (N Y)* **8**, 39-41
17. Hubbard, B. K., and Walsh, C. T. (2003) *Angew. Chem. Int. Edit.* **42**, 730-765
18. Oberthur, M., Leimkuhler, C., Kruger, R. G., Lu, W., Walsh, C. T., and Kahne, D. (2005) *J. Am. Chem. Soc.* **127**, 10747-10752
19. Gehring, A. M., Bradley, K. A., and Walsh, C. T. (1997) *Biochemistry* **36**, 8495-8503
20. Gehring, A. M., Mori, I., and Walsh, C. T. (1998) *Biochemistry* **37**, 2648-2659
21. Weber, G., Schorgendorfer, K., Schneiderscherzer, E., and Leitner, E. (1994) *Curr. Genet.* **26**, 120-125
22. Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) *Nat. Prod. Rep.* **20**, 275-287
23. von Dohren, H., Dieckmann, R., and Pavela-Vrancic, M. (1999) *Chem. Biol.* **6**, R273-R279
24. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) *Chem. Rev.* **97**, 2651-2673
25. Sims, J. W., and Schmidt, E. W. (2008) *J Am Chem Soc* **130**, 11149-11155
26. Fischbach, M. A., and Walsh, C. T. (2006) *Chem. Rev.* **106**, 3468-3496
27. Drews, J. (2000) *Science* **287**, 1960-1964
28. Feldmann, E. G. (2002) *J Am Pharm Assoc (Wash)* **42**, 828-830
29. Tobert, J. A. (2003) *Nat Rev Drug Discov* **2**, 517-526
30. Sweeney, M. J., and Dobson, A. D. (1999) *FEMS Microbiol Lett* **175**, 149-163
31. Minto, R. E., and Townsend, C. A. (1997) *Chem Rev* **97**, 2537-2556
32. Zhan, J., Burns, A. M., Liu, M. X., Faeth, S. H., and Gunatilaka, A. A. (2007) *J Nat Prod* **70**, 227-232
33. Zaleta-Rivera, K., Xu, C., Yu, F., Butchko, R. A., Proctor, R. H., Hidalgo-Lara, M. E., Raza, A., Dussault, P. H., and Du, L. (2006) *Biochemistry* **45**, 2561-2569
34. Linnemannstons, P., Schulte, J., del Mar Prado, M., Proctor, R. H., Avalos, J., and Tudzynski, B. (2002) *Fungal Genet Biol* **37**, 134-148
35. Prodromou, C., Nuttall, J., Millson, S., Roe, S., Pearl, L., Workman, P., Tiow-Suan, S., Tan, D., and Piper, P. (2009) *ACS Chem Biol*
36. Wang, S., Xu, Y., Maine, E. A., Wijeratne, E. M., Espinosa-Artiles, P., Gunatilaka, A. A., and Molnar, I. (2008) *Chem Biol* **15**, 1328-1338
37. Rees, D. O., Bushby, N., Cox, R. J., Harding, J. R., Simpson, T. J., and Willis, C. L. (2007) *Chembiochem* **8**, 46-50
38. Song, Z., Cox, R. J., Lazarus, C. M., and Simpson, T. T. (2004) *Chembiochem* **5**, 1196-1203
39. Burke, L. T., Dixon, D. J., Ley, S. V., and Rodriguez, F. (2005) *Org Biomol Chem* **3**, 274-280
40. Sims, J. W., Fillmore, J. P., Warner, D. D., and Schmidt, E. W. (2005) *Chem Commun (Camb)*, 186-188
41. Ohtsubo, K., Saito, M., Sekita, S., Yoshihira, K., and Natori, S. (1978) *Jpn J Exp Med* **48**, 105-110
42. Shinohara, C., Chikanishi, T., Nakashima, S., Hashimoto, A., Hamanaka, A., Endo, A., and Hasumi, K. (2000) *J Antibiot (Tokyo)* **53**, 262-268
43. Ishikawa, M., Ninomiya, T., Akabane, H., Kushida, N., Tsujiuchi, G., Ohyama, M., Gomi, S., Shito, K., and Murata, T. (2009) *Bioorg Med Chem Lett* **19**, 1457-1460
44. Maiya, S., Grundmann, A., Li, X., Li, S. M., and Turner, G. (2007) *Chembiochem* **8**, 1736-1743

45. Halo, L. M., Marshall, J. W., Yakasai, A. A., Song, Z., Butts, C. P., Crump, M. P., Heneghan, M., Bailey, A. M., Simpson, T. J., Lazarus, C. M., and Cox, R. J. (2008) *Chembiochem* **9**, 585-594
46. Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A. A., and Hertweck, C. (2007) *Nature Chemical Biology* **3**, 213-217
47. Hoffmeister, D., and Keller, N. P. (2007) *Nat Prod Rep* **24**, 393-416
48. Schmitt, I., Martin, M. P., Kautz, S., and Lumbsch, H. T. (2005) *Phytochemistry* **66**, 1241-1253
49. Lu, P., Zhang, A., Dennis, L. M., Dahl-Roshak, A. M., Xia, Y. Q., Arison, B., An, Z., and Tkacz, J. S. (2005) *Mol Genet Genomics* **273**, 207-216
50. Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., and Sankawa, U. (1996) *Molecular & General Genetics* **253**, 1-10
51. Spencer, J. B., and Jordan, P. M. (1992) *Biochem J* **288 (Pt 3)**, 839-846
52. Fujii, I., Yoshida, N., Shimomaki, S., Oikawa, H., and Ebizuka, Y. (2005) *Chem Biol* **12**, 1301-1309
53. Graziani, S., Vasnier, C., and Daboussi, M. J. (2004) *Appl Environ Microbiol* **70**, 2984-2988
54. Gaucher, G. M., and Shepherd, M. G. (1968) *Biochem Bioph Res Co* **32**, 664-&
55. Minto, R. E., and Townsend, C. A. (1997) *Chem. Rev.* **97**, 2537-2556
56. Crawford, J. M., Thomas, P. M., Scheerer, J. R., Vagstad, A. L., Kelleher, N. L., and Townsend, C. A. (2008) *Science* **320**, 243-246
57. Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpeh, O., Tsai, S. C., and Townsend, C. A. (2009) *Nature* **461**, 1139-1143
58. Bradshaw, R. E., and Zhang, S. (2006) *Mycopathologia* **162**, 201-213
59. Watanabe, A., Fujii, I., Sankawa, U., Mayorga, M. E., Timberlake, W. E., and Ebizuka, Y. (1999) *Tetrahedron Letters* **40**, 91-94
60. Watanabe, A., Fujii, I., Tsai, H., Chang, Y. C., Kwon-Chung, K. J., and Ebizuka, Y. (2000) *Fems Microbiology Letters* **192**, 39-44
61. Fujii, I., Mori, Y., Watanabe, A., Kubo, Y., Tsuji, G., and Ebizuka, Y. (2000) *Biochemistry* **39**, 8853-8858
62. Feng, B., Wang, X., Hauser, M., Kaufmann, S., Jentsch, S., Haase, G., Becker, J. M., and Szaniszlo, P. J. (2001) *Infect. Immun.* **69**, 1781-1794
63. Wheeler, M. H., Abramczyk, D., Puckhaber, L. S., Naruse, M., Ebizuka, Y., Fujii, I., and Szaniszlo, P. J. (2008) *Eukaryot. Cell* **7**, 1699-1711
64. Shimizu, T., Kinoshita, H., Ishihara, S., Sakai, K., Nagai, S., and Nihira, T. (2005) *Appl. Environ. Microbiol.* **71**, 3453-3457
65. Linnemannstöns, P., Schulte, J., Prado, M. D., Proctor, R. H., Avalos, J., and Tudzynski, B. (2002) *Fungal Genet. Biol.* **37**, 134-148
66. Wiemann, P., Willmann, A., Straeten, M., Kleigrewe, K., Beyer, M., Humpf, H. U., and Tudzynski, B. (2009) *Mol. Microbiol.* **72**, 931-946
67. Ma, S. M., Zhan, J., Watanabe, K., Xie, X., Zhang, W., Wang, C. C., and Tang, Y. (2007) *J. Am. Chem. Soc.* **129**, 10642-10643
68. Bok, J. W., Chiang, Y. M., Szewczyk, E., Reyes-Dominguez, Y., Davidson, A. D., Sanchez, J. F., Lo, H. C., Watanabe, K., Strauss, J., Oakley, B. R., Wang, C. C., and Keller, N. P. (2009) *Nat. Chem. Biol.* **5**, 462-464

69. Schroeckh, V., Scherlach, K., Nutzmann, H. W., Shelest, E., Schmidt-Heck, W., Schuemann, J., Martin, K., Hertweck, C., and Brakhage, A. A. (2009) *Proceedings of the National Academy of Sciences of the United States of AmericaP Natl Acad Sci USA* **106**, 14558-14563
70. Bailey, A. M., Cox, R. J., Harley, K., Lazarus, C. M., Simpson, T. J., and Skellam, E. (2007) *Chem. Commun. (Camb)*, 4053-4055
71. Chiang, Y. M., Szewczyk, E., Davidson, A. D., Keller, N., Oakley, B. R., and Wang, C. C. (2009) *J. Am. Chem. Soc.* **131**, 2965-2970
72. Couch, R. D., and Gaucher, G. M. (2004) *J. Biotechnol.* **108**, 171-178
73. Awakawa, T., Yokota, K., Funa, N., Doi, F., Mori, N., Watanabe, H., and Horinouchi, S. (2009) *Chem. Biol.* **16**, 613-623
74. Szewczyk, E., Chiang, Y. M., Oakley, C. E., Davidson, A. D., Wang, C. C., and Oakley, B. R. (2008) *Appl. Environ. Microbiol.* **74**, 7607-7612
75. Chooi, Y. H., Cacho, R., and Tang, Y. (2010) *Chem. Biol.* **17**, 483-494
76. Birch, A. J., Massywestropp, R. A., and Moye, C. J. (1955) *Aust J Chem* **8**, 539-544
77. Birch, A. J. (1968) *Ann Rev Plant Physio* **19**, 321-&
78. Chang, P. K., Cary, J. W., Yu, J., Bhatnagar, D., and Cleveland, T. E. (1995) *Mol Gen Genet* **248**, 270-277
79. Watanabe, C. M., Wilson, D., Linz, J. E., and Townsend, C. A. (1996) *Chem Biol* **3**, 463-469
80. Watanabe, C. M., and Townsend, C. A. (2002) *Chem Biol* **9**, 981-988
81. Townsend, C. A., Christensen, S. B., and Trautwein, K. (1984) *Journal of the American Chemical Society* **106**, 3868-3869
82. Ma, Y., Smith, L. H., Cox, R. J., Beltran-Alvarez, P., Arthur, C. J., and Simpson, T. J. (2006) *Chembiochem* **7**, 1951-1958
83. Crawford, J. M., Dancy, B. C. R., Hill, E. A., Udwary, D. W., and Townsend, C. A. (2006) *P Natl Acad Sci USA* **103**, 16728-16733
84. Ma, S. M., Zhan, J., Watanabe, K., Xie, X., Zhang, W., Wang, C. C., and Tang, Y. (2007) *J Am Chem Soc* **129**, 10642-10643
85. Ma, S. M., Li, J. W., Choi, J. W., Zhou, H., Lee, K. K., Moorthie, V. A., Xie, X., Kealey, J. T., Da Silva, N. A., Vederas, J. C., and Tang, Y. (2009) *Science* **326**, 589-592
86. Winssinger, N., and Barluenga, S. (2007) *Chem Commun (Camb)*, 22-36
87. Zhou, H., Qiao, K., Gao, Z., Meehan, M. J., Li, J. W., Zhao, X., Dorrestein, P. C., Vederas, J. C., and Tang, Y. (2010) *J Am Chem Soc* **132**, 4530-4531
88. Zhou, H., Zhan, J., Watanabe, K., Xie, X., and Tang, Y. (2008) *Proc Natl Acad Sci U S A* **105**, 6249-6254
89. Crawford, J. M., Thomas, P. M., Scheerer, J. R., Vagstad, A. L., Kelleher, N. L., and Townsend, C. A. (2008) *Science* **320**, 243-246
90. Zhang, W., Li, Y., and Tang, Y. (2008) *Proc Natl Acad Sci U S A* **105**, 20683-20688
91. Ma, S. M., and Tang, Y. (2007) *FEBS J* **274**, 2854-2864
92. Zhou, H., Li, Y., and Tang, Y. (2010) *Nat Prod Rep* **27**, 839-868
93. Schroeckh, V., Scherlach, K., Nutzmann, H. W., Shelest, E., Schmidt-Heck, W., Schuemann, J., Martin, K., Hertweck, C., and Brakhage, A. A. (2009) *Proc Natl Acad Sci U S A* **106**, 14558-14563
94. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C., and Turgeon, B. G. (2003) *Proc Natl Acad Sci U S A* **100**, 15670-15675

95. Shimizu, T., Kinoshita, H., Ishihara, S., Sakai, K., Nagai, S., and Nihira, T. (2005) *Appl Environ Microbiol* **71**, 3453-3457
96. Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S. C. (2008) *Proc Natl Acad Sci U S A* **105**, 5349-5354
97. Ehrlich, K. C., Montalbano, B., Boue, S. M., and Bhatnagar, D. (2005) *Appl Environ Microbiol* **71**, 8963-8965
98. Bradshaw, R. E., Jin, H., Morgan, B. S., Schwelm, A., Teddy, O. R., Young, C. A., and Zhang, S. (2006) *Mycopathologia* **161**, 283-294
99. Couch, R. D., and Gaucher, G. M. (2004) *J Biotechnol* **108**, 171-178
100. Awakawa, T., Yokota, K., Funa, N., Doi, F., Mori, N., Watanabe, H., and Horinouchi, S. (2009) *Chem Biol* **16**, 613-623
101. Tanabe, M., and Detre, G. (1966) *Journal of the American Chemical Society* **88**, 4515-4517
102. Chooi, Y. H., Cacho, R., and Tang, Y. (2010) *Chem Biol* **17**, 483-494
103. Li, Y., Image, II, Xu, W., Image, I., and Tang, Y. (2010) *J Biol Chem* **285**, 22764-22773
104. Beck, J., Ripka, S., Siegner, A., Schiltz, E., and Schweizer, E. (1990) *Eur. J. Biochem.* **192**, 487-498
105. Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., and Sankawa, U. (1996) *Mol. Gen. Genet.* **253**, 1-10
106. Fujii, I. (2010) *J. Antibiot.* **63**, 207-218
107. Lu, P., Zhang, A., Dennis, L. M., Dahl-Roshak, A. M., Xia, Y. Q., Arison, B., An, Z., and Tkacz, J. S. (2005) *Mol. Genet. Genomics* **273**, 207-216
108. Bacha, N., Dao, H. P., Atoui, A., Mathieu, F., O'Callaghan, J., Puel, O., Liboz, T., Dobson, A. D., and Lebrihi, A. (2009) *Fungal Genet Biol* **46**, 742-749
109. Cox, R. J. (2007) *Organic & Biomolecular Chemistry* **5**, 2010-2026
110. Beck, J., Ripka, S., Siegner, A., Schiltz, E., and Schweizer, E. (1990) *Eur J Biochem* **192**, 487-498
111. Bedford, D. J., Schweizer, E., Hopwood, D. A., and Khosla, C. (1995) *J Bacteriol* **177**, 4544-4548
112. Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C., and Barr, P. J. (1998) *Proc Natl Acad Sci U S A* **95**, 505-509
113. Moriguchi, T., Ebizuka, Y., and Fujii, I. (2006) *Chembiochem* **7**, 1869-1874
114. Light, R. J. (1967) *J Biol Chem* **242**, 1880-1886
115. Kennedy, J., Auclair, K., Kendrew, S. G., Park, C., Vederas, J. C., and Hutchinson, C. R. (1999) *Science* **284**, 1368-1372
116. Xie, X., Meehan, M. J., Xu, W., Dorrestein, P. C., and Tang, Y. (2009) *J. Am. Chem. Soc.* **131**, 8388-8389
117. Abe, Y., Suzuki, T., Ono, C., Iwamoto, K., Hosobuchi, M., and Yoshikawa, H. (2002) *Mol. Genet. Genomics* **267**, 636-646
118. Cox, R. J., Glod, F., Hurley, D., Lazarus, C. M., Nicholson, T. P., Rudd, B. A., Simpson, T. J., Wilkinson, B., and Zhang, Y. (2004) *Chem. Commun. (Camb)*, 2260-2261
119. Yang, G., Rose, M. S., Turgeon, B. G., and Yoder, O. C. (1996) *Plant Cell* **8**, 2139-2150
120. Baker, S. E., Kroken, S., Inderbitzin, P., Asvarak, T., Li, B. Y., Shi, L., Yoder, O. C., and Turgeon, B. G. (2006) *Mol. Plant Microbe In.*
19, 139-149

121. Proctor, R. H., Desjardins, A. E., Plattner, R. D., and Hohn, T. M. (1999) *Fungal Genet. Biol.* **27**, 100-112
122. Fujii, I., Yoshida, N., Shimomaki, S., Oikawa, H., and Ebizuka, Y. (2005) *Chem. Biol.* **12**, 1301-1309
123. Kasahara, K., Fujii, I., Oikawa, H., and Ebizuka, Y. (2006) *Chembiochem* **7**, 920-924
124. Kasahara, K., Miyamoto, T., Fujimoto, T., Oguri, H., Tokiwano, T., Oikawa, H., Ebizuka, Y., and Fujii, I. (2010) *Chembiochem* **11**, 1245-1252
125. Baker, S. E., Kroken, S., Inderbitzin, P., Asvarak, T., Li, B. Y., Shi, L., Yoder, O. C., and Turgeon, B. G. (2006) *Mol Plant Microbe Interact* **19**, 139-149
126. Proctor, R. H., Desjardins, A. E., Plattner, R. D., and Hohn, T. M. (1999) *Fungal Genet Biol* **27**, 100-112
127. Bate, C., Salmona, M., Diomede, L., and Williams, A. (2004) *J Biol Chem* **279**, 14983-14990
128. Sims, J. W., Fillmore, J. P., Warner, D. D., and Schmidt, E. W. (2005) *Chem. Commun.*, 186-188
129. Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A. A., and Hertweck, C. (2007) *Nat. Chem. Biol.* **3**, 213-217
130. Xu, W., Cai, X., Jung, M. E., and Tang, Y. (2010) *J. Am. Chem. Soc.* **132**, 13604-13607
131. Maiya, S., Grundmann, A., Li, X., Li, S. M., and Turner, G. (2007) *Chembiochem* **8**, 1736-1743
132. Tokuoka, M., Seshime, Y., Fujii, I., Kitamoto, K., Takahashi, T., and Koyama, Y. (2008) *Fungal. Genet. Biol.* **45**, 1608-1615
133. Liu, X., and Walsh, C. T. (2009) *Biochemistry* **48**, 8746-8757
134. Halo, L. M., Marshall, J. W., Yakasai, A. A., Song, Z., Butts, C. P., Crump, M. P., Heneghan, M., Bailey, A. M., Simpson, T. J., Lazarus, C. M., and Cox, R. J. (2008) *Chembiochem* **9**, 585-594
135. Sims, J. W., Fillmore, J. P., Warner, D. D., and Schmidt, E. W. (2005) *Chem. Commun. (Camb)*, 186-188
136. Sims, J. W., and Schmidt, E. W. (2008) *Journal of the American Chemical Society* **130**, 11149-11155
137. Liu, X., and Walsh, C. T. (2009) *Biochemistry* **48**, 8746-8757
138. Schumann, J., and Hertweck, C. (2007) *J Am Chem Soc* **129**, 9564-9565
139. Eley, K. L., Halo, L. M., Song, Z. S., Powles, H., Cox, R. J., Bailey, A. M., Lazarus, C. M., and Simpson, T. J. (2007) *Chembiochem* **8**, 289-297
140. Heneghan, M. N., Yakasai, A. A., Halo, L. M., Song, Z. S., Bailey, A. M., Simpson, T. J., Cox, R. J., and Lazarus, C. M. (2010) *Chembiochem* **11**, 1508-1512
141. Xu, W., Cai, X. L., Jung, M. E., and Tang, Y. (2010) *Journal of the American Chemical Society* **132**, 13604-13607
142. Gressler, M., Zaehle, C., Scherlach, K., Hertweck, C., and Brock, M. (2011) *Chem. Biol.* **18**, 198-209
143. Qiao, K. J., Zhou, H., Xu, W., Zhang, W. J., Garg, N., and Tang, Y. (2011) *Org. Lett.* **13**, 1758-1761
144. Dobson, C. M. (2004) *Nature* **432**, 824-828
145. Challis, G. L., and Hopwood, D. A. (2003) *P Natl Acad Sci USA* **100**, 14555-14561
146. Sanchez, J. F., Somoza, A. D., Keller, N. P., and Wang, C. C. C. (2012) *Natural Product Reports* **29**, 351-371

147. Brakhage, A. A., and Schroeckh, V. (2011) *Fungal Genetics and Biology* **48**, 15-22
148. Pickens, L. B., Tang, Y., and Chooi, Y. H. (2011) *Annu Rev Chem Biomol* **2**, 211-236
149. Rokem, J. S., Lantz, A. E., and Nielsen, J. (2007) *Natural Product Reports* **24**, 1262-1287
150. Hendrickson, L., Davis, C. R., Roach, C., Nguyen, D. K., Aldrich, T., McAda, P. C., and Reeves, C. D. (1999) *Chem Biol* **6**, 429-439
151. Proctor, R. H., Desjardins, A. E., Plattner, R. D., and Hohn, T. M. (1999) *Fungal Genetics and Biology* **27**, 100-112
152. Bingle, L. E. H., Simpson, T. J., and Lazarus, C. M. (1999) *Fungal Genetics and Biology* **26**, 209-223
153. Cox, R. J., Glod, F., Hurley, D., Lazarus, C. M., Nicholson, T. P., Rudd, B. A. M., Simpson, T. J., Wilkinson, B., and Zhang, Y. (2004) *Chem Commun*, 2260-2261
154. Nicholson, T. P., Rudd, B. A. M., Dawson, M., Lazarus, C. M., Simpson, T. J., and Cox, R. J. (2001) *Chemistry & biology* **8**, 157-178
155. Linnemannstons, P., Schulte, J., Prado, M. D., Proctor, R. H., Avalos, J., and Tudzynski, B. (2002) *Fungal Genetics and Biology* **37**, 134-148
156. O'Callaghan, J., Caddick, M. X., and Dobson, A. D. W. (2003) *Microbiol-Sgm* **149**, 3485-3491
157. Case, M. E., Schweizer, M., Kushner, S. R., and Giles, N. H. (1979) *P Natl Acad Sci USA* **76**, 5259-5263
158. Ballance, D. J., Buxton, F. P., and Turner, G. (1983) *Biochem Bioph Res Co* **112**, 284-289
159. Yelton, M. M., Hamer, J. E., and Timberlake, W. E. (1984) *P Natl Acad Sci-Biol* **81**, 1470-1474
160. Sanchez, O., Navarro, R. E., and Aguirre, J. (1998) *Molecular and General Genetics* **258**, 89-94
161. Chakraborty, B. N., Patterson, N. A., and Kapoor, M. (1991) *Can J Microbiol* **37**, 858-863
162. Weidner, G., d'Enfert, C., Koch, A., Mol, P. C., and Brakhage, A. A. (1998) *Curr Genet* **33**, 378-385
163. de Groot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. (1998) *Nature Biotechnology* **16**, 839-842
164. Schumann, J., and Hertweck, C. (2006) *J Biotechnol* **124**, 690-703
165. Davis, C., Carberry, S., Schrettl, M., Singh, I., Stephens, J. C., Barry, S. M., Kavanagh, K., Challis, G. L., Brougham, D., and Doyle, S. (2011) *Chem Biol* **18**, 542-552
166. Maier, F. J., Malz, S., Losch, A. P., Lacour, T., and Schafer, W. (2005) *FEMS Yeast Res* **5**, 653-662
167. Bell, A. A., and Wheeler, M. H. (1986) *Annu Rev Phytopathol* **24**, 411-451
168. Nakayashiki, H., Hanada, S., Quoc, N. B., Kadotani, N., Tosa, Y., and Mayama, S. (2005) *Fungal Genetics and Biology* **42**, 275-283
169. Fujii, I. (2009) *Nat Prod Rep* **26**, 155-169
170. Lubertozzi, D., and Keasling, J. D. (2009) *Biotechnol Adv* **27**, 53-75
171. Pieper, R., Ebert-Khosla, S., Cane, D., and Khosla, C. (1996) *Biochemistry* **35**, 2054-2060
172. Luo, G., Pieper, R., Rosa, A., Khosla, C., and Cane, D. E. (1996) *Bioorg Med Chem* **4**, 995-999

173. O'Connor, S. E., Walsh, C. T., and Liu, F. (2003) *Angew Chem Int Ed Engl* **42**, 3917-3921
174. Chen, H., O'Connor, S., Cane, D. E., and Walsh, C. T. (2001) *Chem Biol* **8**, 899-912
175. Buckholz, R. G., and Gleeson, M. A. G. (1991) *Bio-Technology* **9**, 1067-1072
176. Keller, N. P., Turner, G., and Bennett, J. W. (2005) *Nat Rev Microbiol* **3**, 937-947
177. Hitzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G., and Hall, B. D. (1981) *Nature* **293**, 717-722
178. Zhou, H., Qiao, K. J., Gao, Z. Z., Meehan, M. J., Li, J. W. H., Zhao, X. L., Dorrestein, P. C., Vederas, J. C., and Tang, Y. (2010) *Journal of the American Chemical Society* **132**, 4530-+
179. Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G. H., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G. H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J. H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E. U., Archer, D. B., Penalva, M. A., Oakley, B. R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W. C., Denning, D. W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M. S., Osmani, S. A., and Birren, B. W. (2005) *Nature* **438**, 1105-1115
180. von Dohren, H. (2009) *Fungal Genet Biol* **46 Suppl 1**, S45-52
181. Hertweck, C. (2009) *Nat Chem Biol* **5**, 450-452
182. Brakhage, A. A., and Schroeckh, V. (2011) *Fungal Genet Biol* **48**, 15-22
183. Scherlach, K., and Hertweck, C. (2006) *Org Biomol Chem* **4**, 3517-3520
184. Bouhired, S., Weber, M., Kempf-Sontag, A., Keller, N. P., and Hoffmeister, D. (2007) *Fungal Genet Biol* **44**, 1134-1145
185. Bok, J. W., Chiang, Y. M., Szewczyk, E., Reyes-Dominguez, Y., Davidson, A. D., Sanchez, J. F., Lo, H. C., Watanabe, K., Strauss, J., Oakley, B. R., Wang, C. C., and Keller, N. P. (2009) *Nat Chem Biol* **5**, 462-464
186. Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A. A., and Hertweck, C. (2007) *Nat Chem Biol* **3**, 213-217
187. Chiang, Y. M., Szewczyk, E., Davidson, A. D., Keller, N., Oakley, B. R., and Wang, C. C. (2009) *J Am Chem Soc* **131**, 2965-2970
188. Hoffmeister, D., and Keller, N. P. (2007) *Nat. Prod. Rep.* **24**, 393-416
189. Bräse, S., Encinas, A., Keck, J., and Nising, C. F. (2009) *Chem. Rev.* **109**, 3903-3990
190. Winssinger, N., and Barluenga, S. (2007) *Chem. Commun. (Camb)*, 22-36
191. Kashima, T., Takahashi, K., Matsuura, H., and Nabeta, K. (2009) *Biosci. Biotechnol. Biochem.* **73**, 1118-1122
192. Schulte, T. W., Akinaga, S., Soga, S., Sullivan, W., Stensgard, B., Toft, D., and Neckers, L. M. (1998) *Cell Stress Chaperones* **3**, 100-108
193. Soga, S., Shiotsu, Y., Akinaga, S., and Sharma, S. V. (2003) *Curr. Cancer Drug Targets* **3**, 359-369
194. Workman, P. (2003) *Curr. Cancer Drug Targets* **3**, 297-300
195. Roe, S. M., Prodromou, C., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1999) *J. Med. Chem.* **42**, 260-266
196. Agatsuma, T., Ogawa, H., Akasaka, K., Asai, A., Yamashita, Y., Mizukami, T., Akinaga, S., and Saitoh, Y. (2002) *Bioorg. Med. Chem.* **10**, 3445-3454

197. Yang, Z. Q., Geng, X., Solit, D., Pratilas, C. A., Rosen, N., and Danishefsky, S. J. (2004) *J. Am. Chem. Soc.* **126**, 7881-7889
198. Ikuina, Y., Amishiro, N., Miyata, M., Narumi, H., Ogawa, H., Akiyama, T., Shiotsu, Y., Akinaga, S., and Murakata, C. (2003) *J. Med. Chem.* **46**, 2534-2541
199. Dakas, P. Y., Barluenga, S., Totzke, F., Zirrgiebel, U., and Winssinger, N. (2007) *Angew. Chem. Int. Ed. Engl.* **46**, 6899-6902
200. Day, J. E., Sharp, S. Y., Rowlands, M. G., Aherne, W., Workman, P., and Moody, C. J. (2010) *Chemistry* **16**, 2758-2763
201. Barluenga, S., Wang, C., Fontaine, J. G., Aouadi, K., Beebe, K., Tsutsumi, S., Neckers, L., and Winssinger, N. (2008) *Angew. Chem. Int. Ed. Engl.* **47**, 4432-4435
202. Wang, C. H., Barluenga, S., Koripelly, G. K., Fontaine, J. G., Chen, R. H., Yu, J. C., Shen, X. D., Chabala, J. C., Heck, J. V., Rubenstein, A., and Winssinger, N. (2009) *Bioorg. Med. Chem. Lett.* **19**, 3836-3840
203. Reeves, C. D., Hu, Z., Reid, R., and Kealey, J. T. (2008) *Appl. Environ. Microbiol.* **74**, 5121-5129
204. Kim, Y. T., Lee, Y. R., Jin, J., Han, K. H., Kim, H., Kim, J. C., Lee, T., Yun, S. H., and Lee, Y. W. (2005) *Mol. Microbiol.* **58**, 1102-1113
205. Gaffoor, I., and Trail, F. (2006) *Appl. Environ. Microbiol.* **72**, 1793-1799
206. Cox, R. J. (2007) *Org. Biomol. Chem.* **5**, 2010-2026
207. Wang, S., Xu, Y., Maine, E. A., Wijeratne, E. M., Espinosa-Artiles, P., Gunatilaka, A. A., and Molnar, I. (2008) *Chem. Biol.* **15**, 1328-1338
208. Crawford, J. M., Dancy, B. C., Hill, E. A., Udwary, D. W., and Townsend, C. A. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16728-16733
209. Moulin, E., Zoete, V., Barluenga, S., Karplus, M., and Winssinger, N. (2005) *J. Am. Chem. Soc.* **127**, 6999-7004
210. Zhou, H., Zhan, J., Watanabe, K., Xie, X., and Tang, Y. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6249-6254
211. Zhou, H., Qiao, K., Gao, Z., Meehan, M. J., Li, J. W., Zhao, X., Dorrestein, P. C., Vederas, J. C., and Tang, Y. (2010) *J. Am. Chem. Soc.* **132**, 4530-4531
212. Neumann, C. S., Fujimori, D. G., and Walsh, C. T. (2008) *Chem. Biol.* **15**, 99-109
213. Neumann, C. S., Walsh, C. T., and Kay, R. R. (2010) *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5798-5803
214. Barluenga, S., Fontaine, J. G., Wang, C., Aouadi, K., Chen, R., Beebe, K., Neckers, L., and Winssinger, N. (2009) *ChemBioChem* **10**, 2753-2759
215. Lee, K. K., Da Silva, N. A., and Kealey, J. T. (2009) *Anal. Biochem.* **394**, 75-80
216. Jones, E. W. (1991) *Methods Enzymol.* **194**, 428-453
217. Eichhorn, E., van der Ploeg, J. R., and Leisinger, T. (1999) *J. Biol. Chem.* **274**, 26639-26646
218. Mootz, H. D., Schorgendorfer, K., and Marahiel, M. A. (2002) *FEMS Microbiol. Lett.* **213**, 51-57
219. Hellwig, V., Mayer-Bartschmid, A., Muller, H., Greif, G., Kleymann, G., Zitzmann, W., Tichy, H. V., and Stadler, M. (2003) *J. Nat. Prod.* **66**, 829-837
220. Wicklow, D. T., Jordan, A. M., and Gloer, J. B. (2009) *Mycol. Res.* **113**, 1433-1442
221. Wang, M., Zhou, H., Wirz, M., Tang, Y., and Boddy, C. N. (2009) *Biochemistry* **48**, 6288-6290

222. Wijeratne, E. M. K., Paranagama, P. A., and Gunatilaka, A. A. L. (2006) *Tetrahedron* **62**, 8439-8446
223. Watanabe, C. M., and Townsend, C. A. (2002) *Chem. Biol.* **9**, 981-988
224. Chiang, Y. M., Szewczyk, E., Davidson, A. D., Keller, N., Oakley, B. R., and Wang, C. C. (2009) *J. Am. Chem. Soc.* **131**, 2965-2970
225. Xie, X., Meehan, M. J., Xu, W., Dorrestein, P. C., and Tang, Y. (2009) *J. Am. Chem. Soc.* **131**, 8388-8389
226. Caffrey, P. (2003) *ChemBioChem* **4**, 654-657
227. Castonguay, R., He, W., Chen, A. Y., Khosla, C., and Cane, D. E. (2007) *J. Am. Chem. Soc.* **129**, 13758-13769
228. Castonguay, R., Valenzano, C. R., Chen, A. Y., Keatinge-Clay, A., Khosla, C., and Cane, D. E. (2008) *J. Am. Chem. Soc.* **130**, 11598-11599
229. Keatinge-Clay, A. T. (2007) *Chem. Biol.* **14**, 898-908
230. O'Hare, H. M., Baerga-Ortiz, A., Popovic, B., Spencer, J. B., and Leadlay, P. F. (2006) *Chem. Biol.* **13**, 287-296
231. Keatinge-Clay, A. (2008) *J. Mol. Biol.* **384**, 941-953
232. Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H., and Smith, J. L. (2010) *Structure* **18**, 94-105
233. Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005) *Nat. Biotechnol.* **23**, 1171-1176
234. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007) *Annu. Rev. Biochem* **76**, 195-221
235. Dong, C., Flecks, S., Unversucht, S., Haupt, C., van Pee, K. H., and Naismith, J. H. (2005) *Science* **309**, 2216-2219
236. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L., and Walsh, C. T. (2007) *Biochemistry* **46**, 1284-1292
237. Tobert, J. A. (2003) *Nat. Rev. Drug Discovery* **2**, 517-526
238. von Dohren, H. (2009) *Fungal Genet. Biol.* **46**, S45-S52
239. Bergmann, S., Schumann, J., Scherlach, K., Lange, C., Brakhage, A. A., and Hertweck, C. (2007) *Nat. Chem. Biol.* **3**, 213-217
240. Bok, J. W., Chiang, Y. M., Szewczyk, E., Reyes-Domingez, Y., Davidson, A. D., Sanchez, J. F., Lo, H. C., Watanabe, K., Strauss, J., Oakley, B. R., Wang, C. C., and Keller, N. P. (2009) *Nat. Chem. Biol.* **5**, 462-464
241. Schroeckh, V., Scherlach, K., Nutzmann, H. W., Shelest, E., Schmidt-Heck, W., Schuemann, J., Martin, K., Hertweck, C., and Brakhage, A. A. (2009) *Proc. Natl. Acad. Sci. USA* **106**, 14558-14563
242. Balibar, C. J., and Walsh, C. T. (2006) *Biochemistry* **45**, 15029-15038
243. Zhou, H., Zhan, J., Watanabe, K., Xie, X., and Tang, Y. (2008) *Proc. Natl. Acad. Sci. USA* **105**, 6249-6254
244. Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C., and Barr, P. J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 505-509
245. Sattely, E. S., Fischbach, M. A., and Walsh, C. T. (2008) *Nat. Prod. Rep.* **25**, 757-793
246. Finking, R., and Marahiel, M. A. (2004) *Annu. Rev. Microbiol.* **58**, 453-488
247. Sims, J. W., and Schmidt, E. W. (2008) *J. Am. Chem. Soc.* **130**, 11149-11155
248. Seshime, Y., Juvvadi, P. R., Tokuoka, M., Koyama, Y., Kitamoto, K., Ebizuka, Y., and Fujii, I. (2009) *Bioorg. Med. Chem. Lett.* **19**, 3288-3292

249. Maiya, S., Grundmann, A., Li, S. M., and Turner, G. (2006) *ChemBioChem* **7**, 1062-1069
250. von Dohren, H. (2004) *Adv. Biochem. Eng. Biotechnol.* **88**, 217-264
251. Xu, W., Cai, X., Jung, M. E., and Tang, Y. (2010) *J Am Chem Soc* **132**, 13604-13607
252. Gressler, M., Zaehle, C., Scherlach, K., Hertweck, C., and Brock, M. (2011) *Chem Biol*
253. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010) *Nucleic Acids Res.* **38**, D211-222
254. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D. H. (2005) *Nucleic Acids Res.* **33**, 5799-5808
255. O'Brien, W. E. (1976) *Anal. Biochem.* **76**, 423-430
256. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001) *Science* **291**, 1790-1792
257. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D. H. (2005) *Nucleic Acids Res.* **33**, 5799-5808
258. Walsh, C. T., Garneau-Tsodikova, S., and Howard-Jones, A. R. (2006) *Nat. Prod. Rep.* **23**, 517-531
259. El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J., and Thomas, C. M. (2003) *Chem. Biol.* **10**, 419-430
260. Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. (2004) *J. Nat. Prod.* **67**, 1356-1367
261. Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T., and Dorrestein, P. C. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 8977-8982
262. Duerfahrt, T., Eppelmann, K., Muller, R., and Marahiel, M. A. (2004) *Chem. Biol.* **11**, 261-271
263. Kelly, W. L., Hillson, N. J., and Walsh, C. T. (2005) *Biochemistry* **44**, 13385-13393
264. Ehmann, D. E., Trauger, J. W., Stachelhaus, T., and Walsh, C. T. (2000) *Chem. Biol.* **7**, 765-772
265. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L. O. (2007) *Structure* **15**, 781-792
266. Bergendahl, V., Linne, U., and Marahiel, M. A. (2002) *Eur J Biochem* **269**, 620-629
267. Marshall, C. G., Hillson, N. J., and Walsh, C. T. (2002) *Biochemistry* **41**, 244-250
268. Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002) *Nat Struct Biol* **9**, 522-526
269. Scherlach, K., Boettger, D., Remme, N., and Hertweck, C. (2010) *Nat. Prod. Rep.* **27**, 869-886
270. Binder, M., and Tamm, C. (1973) *Angew. Chem. Int. Ed* **12**, 370-380
271. Cole, R. J., Schweikert, M. A., and Jarvis, B. B. (2003) *Handbook of secondary fungal metabolites*, Academic, Amsterdam; Boston
272. Aldridge, D. C., Armstrong, J. J., Speake, R. N., and Turner, W. B. (1967) *J. Chem. Soc. C*, 1667-1676
273. Cooper, J. A. (1987) *J. Cell. Biol.* **105**, 1473-1478
274. Rampal, A. L., Pinkofsky, H. B., and Jung, C. Y. (1980) *Biochemistry* **19**, 679-683
275. Ornelles, D. A., Fey, E. G., and Penman, S. (1986) *Mol. Cell. Biol.* **6**, 1650-1662
276. Cox, R. H., Cutler, H. G., Hurd, R. E., and Cole, R. J. (1983) *J. Agr. Food. Chem.* **31**, 405-408

277. Udagawa, T., Yuan, J., Panigrahy, D., Chang, Y. H., Shah, J., and D'Amato, R. J. (2000) *J. Pharmacol. Exp. Ther.* **294**, 421-427
278. Probst, A., and Tamm, C. (1981) *Helv. Chim. Acta.* **64**, 2065-2077
279. Binder, M., Kiechel, J. R., and Tamm, C. (1970) *Helv. Chim. Acta.* **53**, 1797-1812
280. Vederas, J. C., and Tamm, C. (1976) *Helv. Chim. Acta.* **59**, 558-566
281. Vederas, J. C., Nakashima, T. T., and Diakur, J. (1980) *J. Med. Plants. Res.* **39**, 201-202
282. Oikawa, H., Murakami, Y., and Ichihara, A. (1992) *J. Chem. Soc. Perk Trans. 1*, 2955-2959
283. Schumann, J., and Hertweck, C. (2007) *J. Am. Chem. Soc.* **129**, 9564-9565
284. Sims, J. W., and Schmidt, E. W. (2008) *J. Am. Chem. Soc.* **130**, 11149-11155
285. Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., Crabtree, J., Silva, J. C., Badger, J. H., Albarraq, A., Angiuoli, S., Bussey, H., Bowyer, P., Cotty, P. J., Dyer, P. S., Egan, A., Galens, K., Fraser-Liggett, C. M., Haas, B. J., Inman, J. M., Kent, R., Lemieux, S., Malavazi, I., Orvis, J., Roemer, T., Ronning, C. M., Sundaram, J. P., Sutton, G., Turner, G., Venter, J. C., White, O. R., Whitty, B. R., Youngman, P., Wolfe, K. H., Goldman, G. H., Wortman, J. R., Jiang, B., Denning, D. W., and Nierman, W. C. (2008) *PLoS. Genet.* **4**, e1000046
286. Chooi, Y. H., Stalker, D. M., Davis, M. A., Fujii, I., Elix, J. A., Louwhoff, S. H., and Lawrie, A. C. (2008) *Mycol. Res.* **112**, 147-161
287. Pall, M., and Brunelli, J. (1993) *Fungal. Genet. Newsl.* **40**, 59
288. Szewczyk, E., Nayak, T., Oakley, C. E., Edgerton, H., Xiong, Y., Taheri-Talesh, N., Osmani, S. A., and Oakley, B. R. (2006) *Nat. Protoc.* **1**, 3111-3120
289. Chooi, Y. H., Cacho, R., and Tang, Y. (2010) *Chem. Biol.* **17**, 483-494
290. Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., Crabtree, J., Silva, J. C., Badger, J. H., Albarraq, A., Angiuoli, S., Bussey, H., Bowyer, P., Cotty, P. J., Dyer, P. S., Egan, A., Galens, K., Fraser-Liggett, C. M., Haas, B. J., Inman, J. M., Kent, R., Lemieux, S., Malavazi, I., Orvis, J., Roemer, T., Ronning, C. M., Sundaram, J. P., Sutton, G., Turner, G., Venter, J. C., White, O. R., Whitty, B. R., Youngman, P., Wolfe, K. H., Goldman, G. H., Wortman, J. R., Jiang, B., Denning, D. W., and Nierman, W. C. (2008) *PLoS Genet* **4**, e1000046
291. Buchi, G., Kitaura, Y., Yuan, S. S., Wright, H. E., Clardy, J., Demain, A. L., Ginsukon, T., Hunt, N., and Wogan, G. N. (1973) *J. Am. Chem. Soc.* **95**, 5423-5425
292. Steyn, P. S., Vanheerden, F. R., and Rabie, C. J. (1982) *J. Chem. Soc. Perk Trans. 1*, 541-544
293. Varga, J., Due, M., Frisvad, J. C., and Samson, R. A. (2007) *Stud. Mycol.* **59**, 89-106
294. Liu, R., Gu, Q., Zhu, W., Cui, C., Fan, G., Fang, Y., Zhu, T., and Liu, H. (2006) *J. Nat. Prod.* **69**, 871-875
295. Keller, N. P., Turner, G., and Bennett, J. W. (2005) *Nat. Rev. Microbiol.* **3**, 937-947
296. Robert, J. L., and Tamm, C. (1975) *Helv. Chim. Acta.* **58**, 2501-2504
297. Leisch, H., Morley, K., and Lau, P. C. K. (2011) *Chem. Rev.* **111**, 4165-4222
298. Mirza, I. A., Yachnin, B. J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P. C., and Berghuis, A. M. (2009) *J. Am. Chem. Soc.* **131**, 8848-8854
299. Iwaki, H., Hasegawa, Y., Wang, S., Kayser, M. M., and Lau, P. C. (2002) *Appl. Environ. Microbiol.* **68**, 5671-5684
300. Malito, E., Alfieri, A., Fraaije, M. W., and Mattevi, A. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 13157-13162

301. Fraaije, M. W., Kamerbeek, N. M., van Berkel, W. J., and Janssen, D. B. (2002) *FEBS Lett.* **518**, 43-47
302. MacPherson, S., Larochelle, M., and Turcotte, B. (2006) *Microbiol. Mol. Biol. Rev.* **70**, 583-604
303. Ehrlich, K. C., Montalbano, B. G., and Cary, J. W. (1999) *Gene* **230**, 249-257
304. Shimizu, T., Kinoshita, H., and Nihira, T. (2007) *Appl. Environ. Microbiol.* **73**, 5097-5103
305. Demain, A. L., Hunt, N. A., Malik, V., Kobbe, B., Hawkins, H., Matsuo, K., and Wogan, G. N. (1976) *Appl. Environ. Microbiol.* **31**, 138-140
306. Olano, C., Lombo, F., Mendez, C., and Salas, J. A. (2008) *Metab. Eng.* **10**, 281-292
307. Flaherty, J. E., and Payne, G. A. (1997) *Appl. Environ. Microbiol.* **63**, 3995-4000
308. Teijeira, F., Ullan, R. V., Fernandez-Aguado, M., and Martin, J. F. (2011) *Metab. Eng* **13**, 532-543
309. Wattanachaisaereekul, S., Lantz, A. E., Nielsen, M. L., and Nielsen, J. (2008) *Metab. Eng.* **10**, 246-254
310. Thykaer, J., and Nielsen, J. (2003) *Metab. Eng.* **5**, 56-69
311. Vedejs, E., Campbell, J. B., Gadwood, R. C., Rodgers, J. D., Spear, K. L., and Watanabe, Y. (1982) *J. Org. Chem.* **47**, 1534-1546
312. Auclair, K., Sutherland, A., Kennedy, J., Witter, D. J., Van den Heever, J. P., Hutchinson, C. R., and Vederas, J. C. (2000) *J. Am. Chem. Soc.* **122**, 11519-11520
313. Kim, H. J., Ruszczycky, M. W., Choi, S.-h., Liu, Y.-n., and Liu, H.-w. (2011) *Nature* **473**, 109-112
314. Gardiner, D. M., and Howlett, B. J. (2005) *FEMS Microbial. Lett.* **248**, 241-248
315. Bomke, C., Rojas, M. C., Hedden, P., and Tudzynski, B. (2008) *Appl. Environ. Microbiol.* **74**, 7790-7801
316. Iwaki, H., Wang, S., Grosse, S., Bergeron, H., Nagahashi, A., Lertvorachon, J., Yang, J., Konishi, Y., Hasegawa, Y., and Lau, P. C. (2006) *Appl. Environ. Microbiol.* **72**, 2707-2720
317. Turfitt, G. E. (1948) *Biochem. J.* **42**, 376-383
318. Aldridge, D. C., Armstron, J. J., Speake, R. N., and Turner, W. B. (1967) *Chem. Commun. (London)*, 26-27
319. Aldridge, D. C., and Turner, W. B. (1969) *J. Chem. Soc. C*, 923-928
320. Krucker, T., Siggins, G. R., and Halpain, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6856-6861
321. Gao, Q., Jin, K., Ying, S. H., Zhang, Y., Xiao, G., Shang, Y., Duan, Z., Hu, X., Xie, X. Q., Zhou, G., Peng, G., Luo, Z., Huang, W., Wang, B., Fang, W., Wang, S., Zhong, Y., Ma, L. J., St Leger, R. J., Zhao, G. P., Pei, Y., Feng, M. G., Xia, Y., and Wang, C. (2011) *PLoS Genet.* **7**, e1001264