A Study of the Iterative Type I Polyketide Synthases in Enediyne and Mellein Biosynthesis

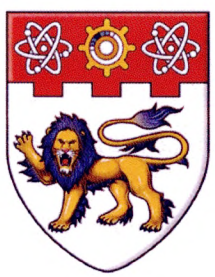
Sun Huihua

This dissertation is submitted for the degree of Doctor of Philosophy

Supervisor

Assoc. Prof. Liang Zhao-Xun

School of Biological Sciences



NANYANG TECHNOLOGICAL UNIVERSITY

Acknowledgements

Thanks must first go to my supervisor, Assoc. Prof. Liang Zhao-Xun, for giving me such a great chance to engage in my favorite research area in natural products. I am deeply grateful for his guidance and inspiration throughout my Ph.D. study.

Apart from my supervisor, I would like to acknowledge the great help and support given by my current and ex-lab mates, Lawrence, Mary, Siew Lee, Wu Long, Xu Linghui, Alolika, Jamila, Swathi, Kong Rong, Chong Wai, Yaning, Rao Feng, Ela and Jiqiang. Thank you for helping me out when I was at times confronted with problems in my experiment. I am extremely grateful to my friend Feiqing from the School of Physical and Mathematical Sciences. Without his generous and skillful help in TLC and NMR spectroscopy, the structure determination of a novel compound would have been impossible. I am very thankful to Assoc. Prof. Thirumaran Thanabalu for his kind advice and help on protein expression in yeast. I would like to thank Asst. Prof. Tang Kai, Dr. Li Bin and Dong Xueming for their help with LC-MS. I am indebted to Li Yan for the time he has spent in two-dimensional NMR experiments. Thanks also go to Dr Zhong Guofu's lab for the collaboration in chemical synthesis.

I am very grateful to the School of Biological Sciences, Nanyang Technological University for providing me with Ph.D. scholarship. My research is mainly funded by ARC grant from the Ministry of Education (Singapore).

Credit for this achievement must also go to my parents who loved, supported and inspired me through the years.

Table of Contents

Abstract

Iterative type I polyketide synthases (iPKSs) are large multifunctional enzymes that assemble polyketide products by using a single module composed of several catalytic domains. Although iPKSs utilize the same repertoire of catalytic domains as fatty acid synthases and modular PKSs, iPKSs are able to use a single set of catalytic domains to assemble the chemically and structurally diverse polyketide products in an iterative manner. How the iPKSs achieve the chemical and structural diversity by "programming" the catalytic domains remains one of the greatest mysteries in enzymology today. In this thesis, I describe the results from my studies on two iPKSs that aimed to understand the function and mechanism of the multifunctional iPKSs.

The first iPKS under investigation is SgcE, the PKS putatively responsible for the biosynthesis of the functional enediyne moiety of the antitumor compound C-1027. Naturally occurring enediynes are mostly known for their unique molecular structure and remarkable biological activities largely owing to the central bicyclic enediyne functionality. Following the recent sequencing of the gene clusters for the synthesis of several enediyne natural products, my research work aimed to understand the role of SgcE in the biogenesis of the enediyne functionality of C-1027. Together with the putative ancillary enzymes SgcE3 and SgcE10, SgcE were overexpressed in the heterologous expression systems that include *E. coli* and *Saccharomyces cerevisiae.* Several highly conjugated polyene products of SgcE generated by *in vitro* enzymatic reactions and *in vivo* protein co-expression were characterized to suggest that SgcE is

likely to be the enzyme responsible for assembling the acetate units of malonyl-CoA into the linear precursor of the enediyne structure. Comparative studies with the homologous CalE8 and DynE8 were also conducted to investigate a possible divergence of the pathways for 9- and 10-membered enediynes at the early stage of biosynthesis.

The second system is a functionally unknown iPKS (SACE5532) from *Saccharopolyspora erythraea*. SACE5532 shares sequence homology and domain composition with several known bacterial and fungal iPKSs for aromatic polyketide biosynthesis. SACE5532 was successfully overexpressed in *E. coli* as a soluble protein after optimization of the expression conditions. Enzymatic reactions and co-expression with the Sfp 4'-phosphopantetheinyl transferase (PPTase) were conducted to identify the enzymatic product. Structure determination by mass spectrometry and NMR spectroscopy suggested that SACE5532 produces mellein, a natural product that was previously found in fungi and the mandibular gland of insects. The results establish the polyketide origin of mellein and suggest that the compound is likely to be produced by the cyclization of a pentaketide intermediate. Given the predicted oxidation state of the pentaketide intermediate, the ketoreductase (KR) domain of SACE5522 is likely to selectively reduce two of the four keto groups *via* a programmed mechanism. Domain swapping between SACE5532 and another iPKS (NcsB) revealed that the ACP domain does not affect product formation, while TH-KR di-domain plays a role in controlling the keto-reduction pattern. Importantly, the substrate specificity exhibited by the stand-alone KR domains of SACE5532 and NcsB towards diketide analogs

strongly suggests that the KR domain alone is able to differentiate and selectively reduce the polyketide intermediates, which sheds light on the mechanism of the programmed keto-reduction for these partially reducing iterative PKSs.

Abbreviations

6-MSAS	6-methylsalcylic acid synthase
aa	Amino acid
Acetyl-CoA	Acetyl coenzyme A
ACP	Acyl carrier protein
APCI	Atmospheric pressure chemical ionization
AT	Acyl transferase domain
ATP	Adenosine-5'-triphosphate
β-ME	β-Mercaptoethanol
bp	Base pair
CAL	Calicheamicin
CoA	Coenzyme A
DEBS	6-deoxyerythronolide B synthase
DH	Dehydratase domain
DNA	5'-deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
DYN	Dynemycin
E. coli	*Escherichia coli*
Ek/LIC	Enterokinase(protease cleavage site)/ligation-independent cloning
ESI	Electrospray ionization

FA	Formic acid
FAS	Fatty acid synthase
FPLC	Fast protein liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
$(His)_6$-tag	Hexahistidine affinity tag
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iPKS	Iterative type I polyketide synthase
kb	Kilobase
kDa	Kilodalton
KR	Ketoreductase domain
KS	Ketosynthase domain
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectroscopy
Malonyl-CoA	Malonyl coenzyme A
MH^+	Protonated molecular ion
mPKS	Multi-modular type I polyketide synthase
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
Ni-NTA	Nickel-nitrilotriacetic acid resin

m/z	Mass-to-charge ratio
nm	Nanometer
NMR	Nuclear magnetic resonance spectroscopy
NOESY	Nuclear overhauser enhancement spectroscopy
NRPS	Nonribosomal peptide synthetase
OD	Optical density
OSAS	Orsellinic acid synthase
PCR	Polymerase chain reaction
PDA	Photodiode array
PKS	Polyketide synthase
PKSE	Enediyne PKS
PPTase	Phosphopantetheinyl transferase
rpm	Revolutions per minute
S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TE	Thioesterase domain
TH	Thioester hydrolase domain
TLC	Thin layer chromatography
TFA	Trifluoroacetic acid
Tris	Tris(hydroxymethyl)aminomethane
UV-Vis	Ultraviolet-Visible spectroscopy

CHAPTER 1

Background

1.1 Polyketide natural products

Polyketide-derived natural products are a large family of secondary metabolites found in bacteria, fungi and plants (**Figure 1.1**). This family of natural products exhibit astonishing structural diversities, ranging from simple aromatic molecules such as orsellinic acid to extremely complex structures such as macrocycles (erythromycin A), polyenes (amphotericin B), and bicyclic enediynes (calicheamicin, C-1027). Many polyketide natural products also contain multiple functional units with different polyketide origins. For instance, calicheamicin contains two moieties with polyketide origin: a 10-membered enediyne core and a monocyclic orsellinic acid moiety. With the modern nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS), more and more polyketides are being identified and characterized with many of them exhibiting potential therapeutic properties [1].

Some of the structurally diverse polyketide natural products boast unparalleled biological activities and have become a rich source of clinically valuable drugs. The polyketide-derived drugs include antibiotics (erythromycin A, monensin A, rifamycin S), immunosuppressants (rapamycin, FK506), antifungal (amphotericin B),

antiparasitic (avermectin) and antitumor (doxorubicin, C1027) drugs. The fungal polyketide-derived statins, such as lovastatin, are among the most successful cholesterol-lowering agents on the market. Currently, polyketide-derived drugs have generated huge revenues for drug developers with an estimated annual sale of 20 billion dollars.

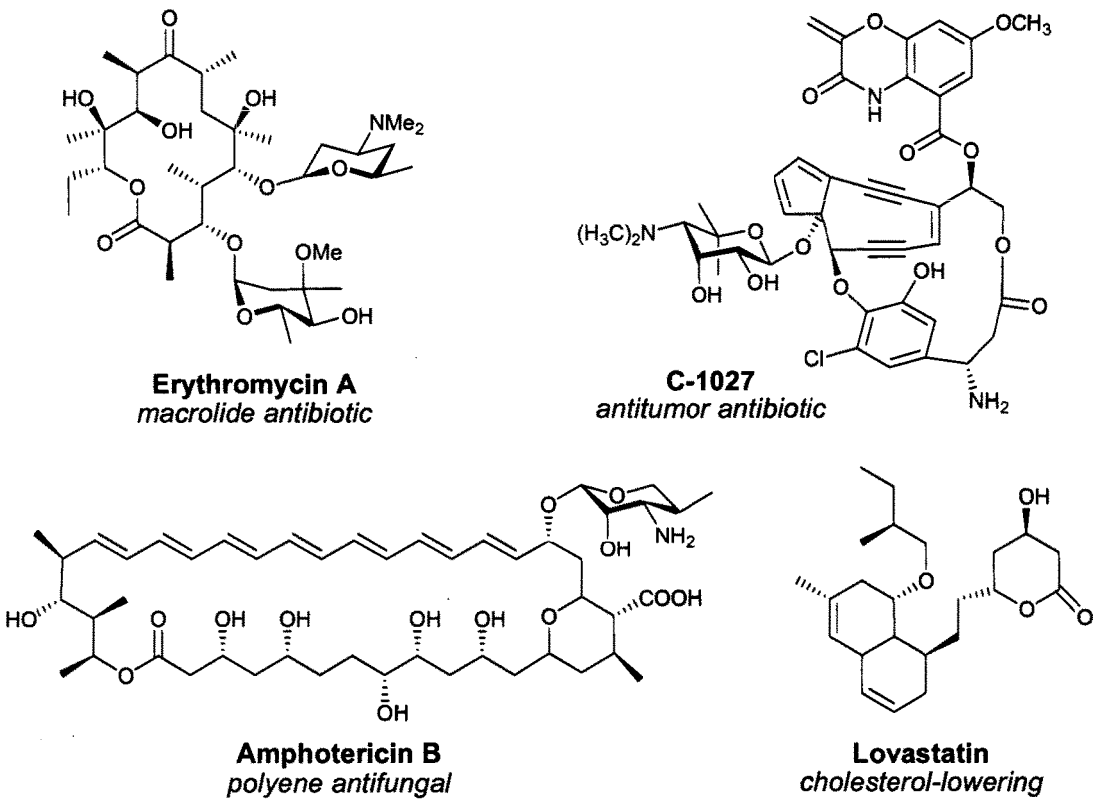


Figure 1.1. Examples of polyketide secondary metabolites. The name of the polyketide and its primary biological activity are indicated.

1.2 Fatty acid biosynthesis

Naturally occurring polyketides are enzymatically synthesized by a common mechanism of decarboxylative condensations of simple malonate derivatives by polyketide synthases (PKSs). PKSs share a very similar catalytic mechanism with bacterial or mammalian fatty acid synthases (FASs) [2-4]. As early as 1953, Birch proposed that polyketide biosynthesis is related to fatty acid biosynthesis [5]. Today, it is known that FAS and PKS both use acyl CoA substrates and share a similar overall strategy using similar catalytic domains for chain extension, reduction, dehydration and chain release. Insights into fatty acid biosynthesis from early studies have paved the way for the study of polyketide biosynthesis.

FASs are classified into two types according to the protein architectures. Type I FASs are large multifunctional and multidomain proteins, which contain ketosynthase (KS), acyl carrier protein (ACP), malonyl-acetyl transferase (MAT), ketoreductase (KR), dehydrotase (DH) and enoyl reductase (ER) and thioesterase (TE) domains. This type of FAS is usually found in animals and fungi. Type II FASs consist of discrete proteins with similar catalytic functions as the type I FAS domains. This type of FAS can be found in plants and bacteria.

The multidomain Type I FAS and PKS use the acyl carrier protein (ACP) domain to tether the growing acyl chain. The ACP is first modified post-translationally by a phosphopantetheinyl transferase (PPTase) through transferring the 4'-phosphopantetheine moiety derived from Coenzyme A to a conserved serine residue in ACP (**Figure 1.2**). The phosphopantetheine moiety has a terminal thiol group that

forms a thioester bond with extender unit or the growing acyl chain. The phosphopantetheine moiety functions as a flexible arm that shuttles the polyketide intermediates between various catalytic domains for chain elongation and processing.

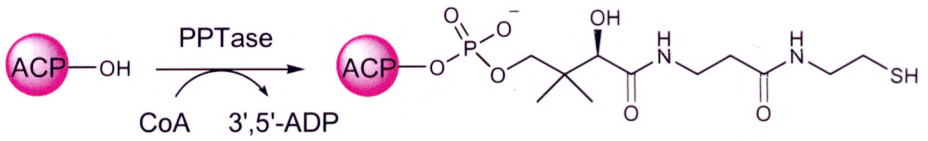


Figure 1.2. The modification of acyl carrier protein (ACP) by phosphopantetheinyl transferase (PPTase).

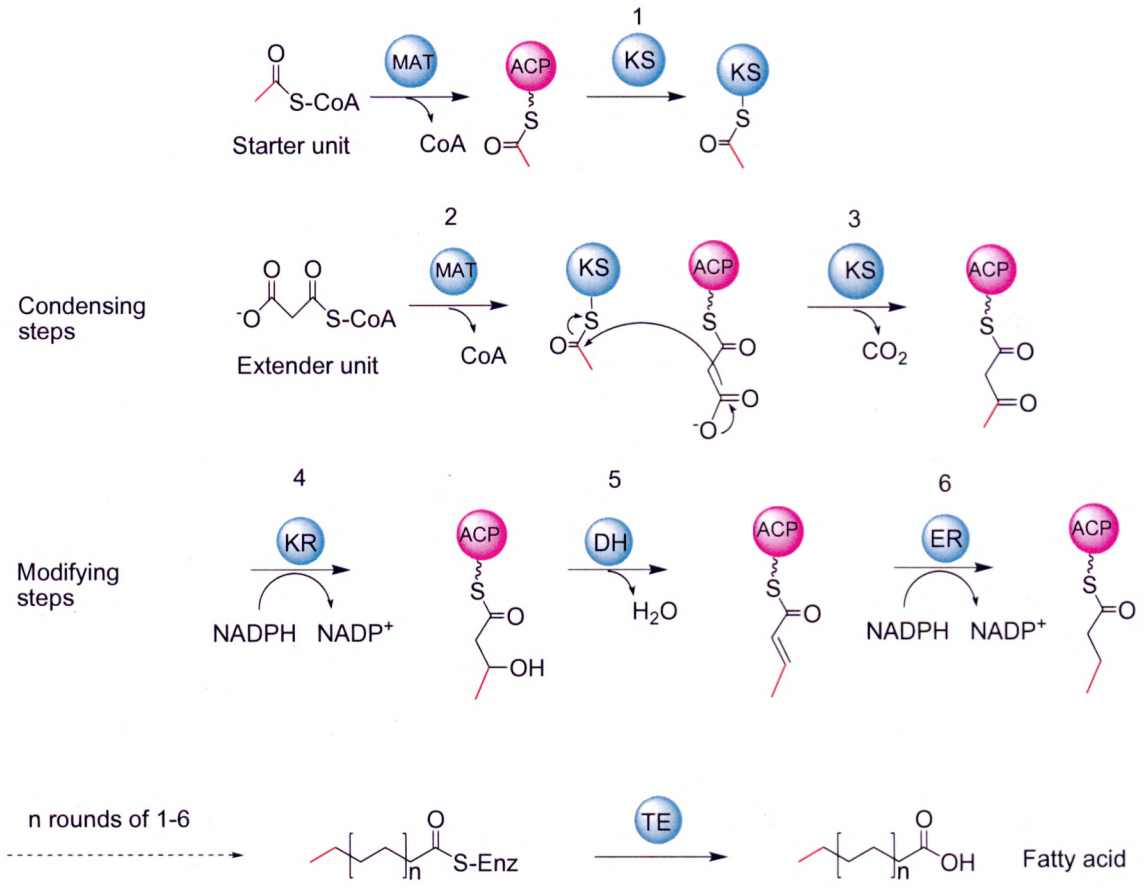


Figure 1.3. Typical reactions catalyzed by various functional domains of FASs. KS, ketosynthase; ACP, acyl carrier protein; MAT, malonyl-acetyl transferase; KR, ketoreductase; DH, dehydrotase; ER, enoyl reductase; TE, thioesterase.

In the case of fatty acid biosynthesis in mammals (**Figure 1.3**), the starter group acetyl-CoA is loaded onto the ACP domain as catalyzed by the malonyl-acetyl transferase (MAT) domain. The ACP, MAT, ketosynthase (KS), ketoreductase (KR), dehydrotase (DH) and enoyl reductase (ER) domains are used in an iterative fashion for chain elongation and processing. Each cycle of chain elongation includes the following steps: (1) The growing chain is handed over from the ACP domain to the KS domain as catalyzed by the KS domain. (2) The extender unit originated from malonyl-CoA is loaded onto the ACP domain as catalyzed by the AT domain. (3) The ACP-tethered malonyl group undergoes a Claisen condensation with the KS-bound polyketide intermediate to yield another intermediate with two additional carbons from one acetate unit. (4) The KR domain reduces the β-keto group to a β-hydroxy group. (5) The DH domain catalyzes a dehydration step to form the carbon-carbon double bond. (6) The double bond is further reduced to the final saturated product by the ER domain. This cycle is repeated to generate the final product with the length determined by the thioesterase (TE) domain. Finally, the linear fatty acid product is hydrolytically released from FAS by the TE domain.

High-resolution crystal structure of the type I mammalian FAS (mFAS) has been determined to reveal an X-shaped dimer that can be divided into two sections (**Figure 1.4**) [6]. The condensing section consists of KS and MAT domains, while the modifying section consists of DH, KR, ER, ψME (pseudo-methyltransferase) and ψKR (pseudo-KR) domains. ψME and ψKR domains with unknown function were suggested to play an important structural role in the modifying section. The ACP and

the adjacent TE domains were not observed in the crystal structure presumably due to conformational flexibility.

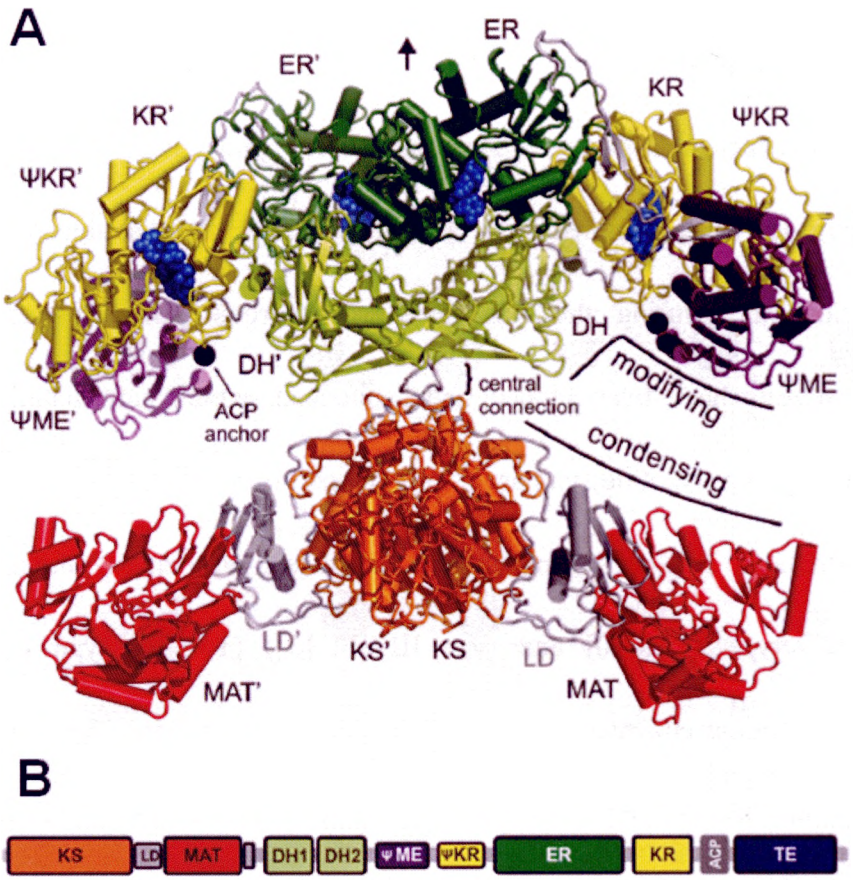


Figure 1.4. (A) The front view of the structure of mammalian FAS (mFAS), colored by domains as indicated. (B) Linear sequence organization of mFAS. (Figures are adopted from reference [6]).

As PKSs share sequence homology and domain organization with mFAS, the mFAS structure serves as an invaluable template for modeling PKS structures. The mFAS structure not only helps us to understand the catalysis performed by the individual domains, but also enables more accurate and reliable assignment of domain

boundaries for PKSs, which is critical for elucidating PKS chemistry.

1.3 Polyketide synthase (PKS) enzymology

The early understanding of polyketide biosynthesis came from sophisticated feeding experiments with selectively labeled precursors. It was established that polyketides are assembled from acetate precursor through decarboxylative Claisen reaction [1]. During the last decade, numerous gene clusters for polyketide biosynthesis have been cloned and sequenced, which significantly speeds up the discovery of polyketide synthases from diverse sources.

PKSs are categorized into three classes based on their enzyme architectures (type I [1,7], type II [8-10] and type III [11-13]) [14-17]. Type I PKSs are large multifunctional enzymes with a set of functional domains. These are further divided into multi-modular PKS (mPKS) and single-modular iterative type I PKS (iPKS). Each module consists of a set of catalytic enzyme domains, such as KS, AT, KR, DH and ACP. Type II PKs are multi-enzyme complexes that carry a single set of iteratively acting activities. Type III PKSs have a single KS-like domain, which carries out all the decarboxylation, condensation and cyclization reactions. This type of PKS does not depend on ACP and can use acyl-CoA thioester directly.

PKSs use the same array of chemical reactions as FASs to assemble the polyketide products (**Figure 1.5**). In brief, the starter unit is loaded onto the ACP domain as catalyzed by the acyl transferase (AT) domain. During chain elongation, the reactions

catalyzed by the ACP, AT, KS, KR, DH and ER domains proceed through just like in FAS. The final polyketide product can be finally released from the PKS by hydrolysis with or without cyclization. The TE domain or a discrete TE is usually needed to off-load the PKS-tethered polyketide from the ACP domain.

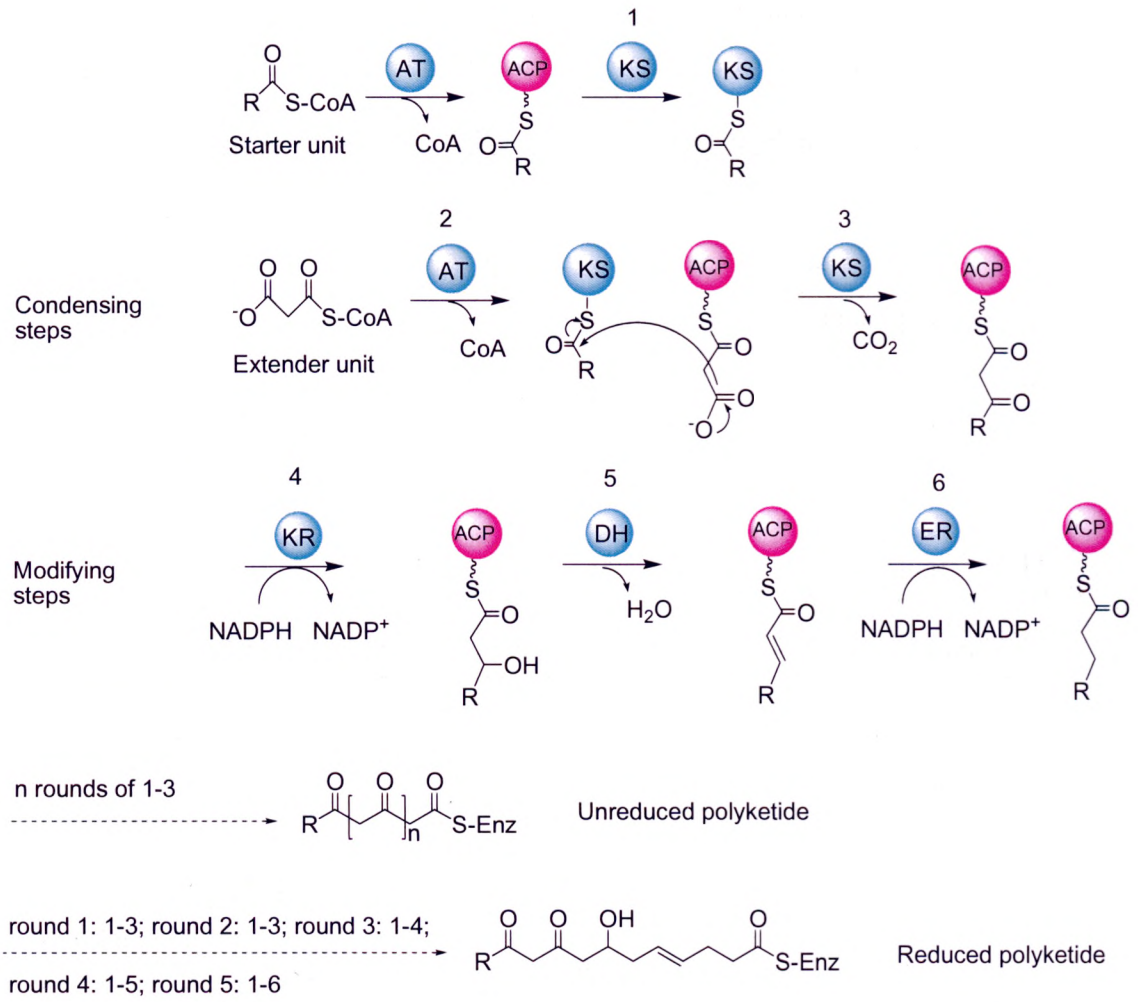


Figure 1.5. Typical reactions catalyzed by various functional domains of PKSs.

Although FAS and PKS utilize similar catalytic domains and mechanisms for assembling the products, PKSs can produce structurally diverse products in comparison to the structurally simple fatty acids produced by FASs. The greater

structural diversity of PKS products are attributed to several important features of PKSs: (1) PKSs are able to control the degree of reduction and dehydration during chain elongation. This leads to formation of ketone, hydroxyl, alkene or alkane functionality at different positions. (2) For many PKSs, the linear polyketide intermediates can be cyclized to form cyclic molecule of different ring sizes. (3) PKSs employ a wide variety of starter units and extender units (e.g. methylmalonyl-CoA and mthoxymalonyl-CoA) to create structural diversity. (4) In addition to the common KR, DH and ER processing domains shared between FAS and PKS, PKSs also use other catalytic domains such as methyltransferase domains for structural diversification.

1.4 Iterative type I PKS (iPKS)

Iterative type I PKSs (iPKSs) are PKSs that assemble the polyketide chain by using a single module composed of multiple catalytic domains iteratively. This is in sharp contrast to the multi-modular type I PKSs (e.g. DEBS for 6-deoxyerythronolide B biosynthesis) that employ multiple modules to assemble the final products. iPKSs are mainly found in fungi and bacteria with various protein sizes and domain compositions. The domain compositions of some representative iPKSs are shown in **Figure 1.6**. How these iPKSs, even sometimes with the same domain composition, generate different products by a programmed mechanism remains one of the most exciting topics in enzymology today.

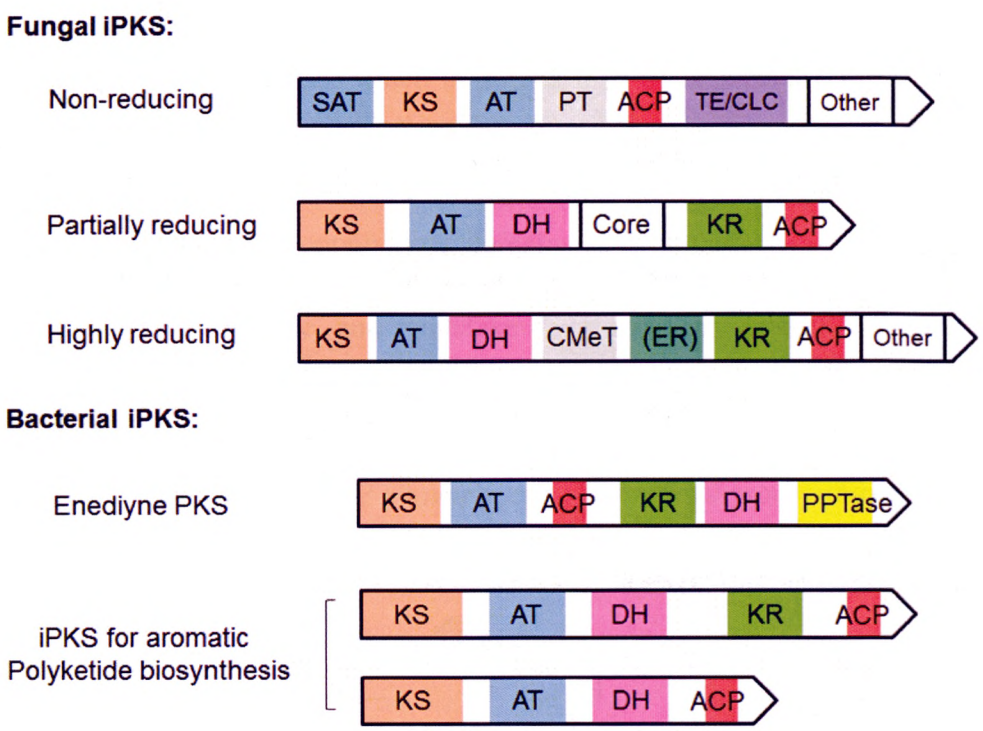


Figure 1.6. Domain organization of iterative type I PKSs from fungi and bacteria. SAT, starter unit acyl transferase; KS, ketosynthase; AT, acyl transferase; PT, product template; ACP, acyl carrier protein; TE/CLC, thioesterase/Claisen cyclase; DH, dehydrase; KR, ketoreductase; CMeT, C-methyltransferase; ER, enoyl reductase; PPTase, phosphopantetheinyl transferase.

1.4.1 Fungal iPKSs

Most identified iPKS are found in Fungi. The Fungal iPKSs are divided into three types based on the reduction pattern, including non-reducing (NR-PKS), partially reducing (PR-PKS) and highly reducing (HR-PKS) [18] (**Figure 1.7**).

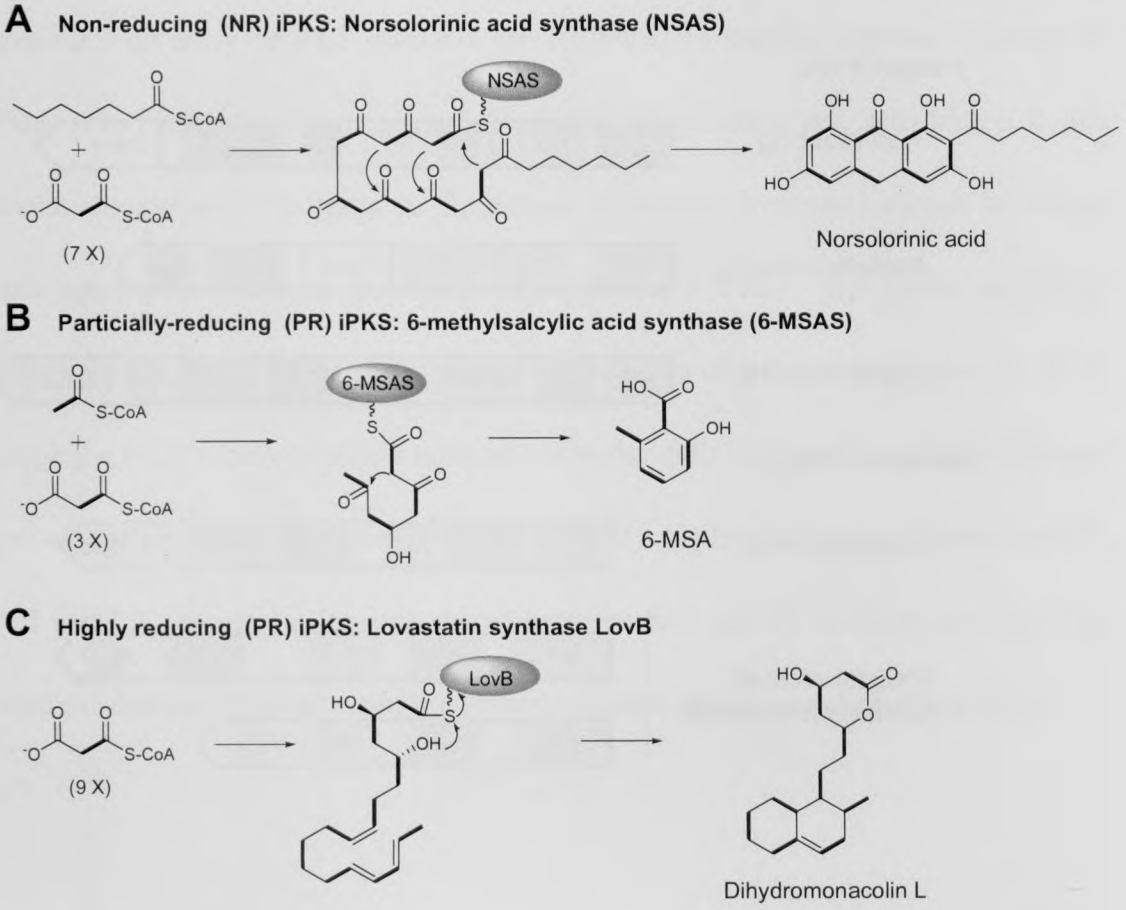


Figure 1.7. Three classes of fungal type I iterative polyketide synthases (iPKSs). (A) Non-reducing iPKS such as norsolorinic acid synthase (NSAS) for norsolorinic acid biosynthesis [19]. (B) Partially reducing iPKS such as 6-methylsalicyclic acid synthase (6-MSAS) for 6-methylsalicyclic acid biosynthesis [20]. (C) Highly reducing iPKS such as lovastatin synthase LovB for dihydromonacolin L biosynthesis [21].

NR-PKSs do not reduce the keto groups of the polyketide intermediates. The linear intermediates are cyclized into aromatic products such as orsellinic acid and norsolorinic acid at the end of the biosynthesis process. The N-terminal starter unit acyl transferase (SAT) domain is able to load a starter unit derived from a dedicated FAS or PKS. For example, the SAT domain of the norsolorinic acid synthase (NSAS) loads the hexanoate starter unit synthesized by a dedicated FAS (**Figure 1.7A**) [22]. The starter unit loading domain is followed by the KS and AT domains responsible for

chain extension and malonyl extender unit loading. Next to the AT domain is a unique product template (PT) domain, which is proposed to be responsible for chain length determination [23]. Although some NR-PKSs appear to have the ACP domain as the C-terminal domain, others contain extra domains such as thioesterase/Claisen-cyclase (TE/CLC), methyltransferase, and reductase domains. Thus, NR-PKSs appear to be arranged with an N-terminal loading component, a central chain extension component consisting of KS, AT, PT and ACP domains, and a C-terminal processing component.

PR-PKSs perform selective keto-reduction during the process of polyketide chain elongation. The domain organizations of PR-PKSs usually resembles that of mFAS, with an N-terminal KS domain and the AT, DH, a so-called 'core' domain, KR and ACP domains followed. The domain organizations seem to differ considerably from that of the NR-PKS, with the striking absence of SAT, PT domains and CLC/TE domain. Although a number of PR-PKS genes are known from genome sequencing projects, only three PR-PKSs have been characterized so far and all the three PKSs produces the same product - 6-methylsalicylic acid. The first PR-PKS to be characterized is the 6-methylsalicylic acid synthase (6-MSAS) found in *Penicillium patulum* [20]. Ebizuka and co-workers identified the *atX* gene from *Aspergillus terreus* as a 6-MSAS gene by heterologous expression [24]. Most recently, Tkacz's group have reported that the *pks2* gene from *Glarea lozoyensis* encodes a 6-MSAS [25].

In contrast to NR- and PR-PKSs, HR-PKSs produce highly reduced molecules such as lovastatin, T-toxin, fumonisin B1 and squalestatin. This class of fungal PKS generally program a complex set of keto-reduction, dehydration and enoyl reduction

[18,26]. A typical HR-PKS has an N-terminal KS domain, followed by AT, DH, C-methyltransferase (CMeT), KR and ACP domains. Some HR-PKSs such as lovastatin synthase LovB possess an ER domain. However, the need for a discrete ER domain, LovC, suggests that the ER domain within LovB is inactive [27]. HR-PKSs have no domain that shares similarity with the PT domain of NR-PKS or the "core" domain of PR-PKS; and no N-terminal SAT domain seems to be used by NR-PKSs. Although many putative HR-PKS genes are identified from the fungal genome sequencing projects, only a few proteins have been characterized with their biosynthetic products identified.

1.4.2 Bacterial iPKSs

Because bacterial type I PKSs are mostly found to be multi-modular PKSs, the single-modular iterative type I PKSs (iPKSs) were initially thought only present in fungi. However, recent genetic and biochemical studies have revealed the unexpected presence of iPKSs in bacteria as well [15,28,29].

A few bacterial iPKSs have been identified for aromatic polyketide biosynthesis. Because aromatic polyketide biosynthesis in bacteria is usually carried out by the iterative type II PKS, the discovery of the iterative type I PKSs with a single module for aromatic polyketide biosynthesis is surprising. AviM from *Streptomyces viridochromogenes* is the first bacterial iPKS discovered for orsellinic acid (OSA) synthesis [30]. The naphthoate synthase NcsB catalyzing

2-hydroxyl-5-methylnaphthoic acid involved in neocarzinostatin biosynthesis was found in *Streptomyces carzinostaticus* [31]. To date, seven bacterial iPKSs for aromatic polyketide biosynthesis have been reported [31-40]. Two of them, AviM and CalO5, catalyze the synthesis of OSA that will be later incorporated into avilamycin and calicheamicin. These two iPKS are NR-PKSs without a KR domain for keto-reduction; whereas the other five iPKSs display head-to-tail homology to fungal 6-MSASs and thus belong to PR-PKSs.

Another important family of bacterial iPKSs is for enediyne biosynthesis. Enediyne natural products are known for their unprecedented molecular structure, remarkable biological activities, and fascinating mode of action. The structurally unique enediyne moiety of enediyne natural products is composed of two acetylenic groups conjugated by a double bond within a 9- or 10-membered ring. Recently, the gene clusters for three 9-membered enediynes (C-1027 [41], neocarzinostatin [35] and maduropeptin [38]) and two 10-membered enediynes (calicheamicin [42] and dynemicin [43]) have been sequenced and annotated. The studies revealed a novel family of enediyne iPKSs (PKSE) that is putatively the major player responsible for the synthesis of the highly reactive enediyne core structure (or "warhead"). Six domains of PKSEs have been predicted based on sequence homology and active site mapping [44]. We will discuss this family of bacterial iPKSs in details in chapters 2 and 3.

1.4.3 Programmed polyketide synthesis in iPKSs

For bacterial modular type I PKS, the formation of product is pre-determined by the number of module and the domain composition of the modules. In contrast, the single module-containing iPKSs only have one set of catalytic domains that must be used iteratively. How the chain length is determined in some of the iPKSs is not known. Most intriguingly, the processing domains of some of the iterative PKSs, particularly the PR-PKSs, are able to selectively act on the polyketide intermediates at different stages to generate structural diversity in the final product. How the iPKSs are programmed to perform the selective reduction, dehydration or methylation is one of the unanswered questions in modern enzymology. Deciphering the programmed biosynthesis in iPKS will not only yield insight into these large protein machineries, but also open the door for generation of novel bioactive polyketides through rational enzyme engineering.

Townsend and co-workers have made great contributions to our understanding of fungal NR-PKSs. They characterized the N-terminal starter unit acyl transferase (SAT) domain and established that it plays a role in starter unit selection [22]. They also showed that the PT domain unique to NR-PKSs controls both chain length and cyclization pattern for the non-reduced polyketide biosynthesis by using an active-site cleft of defined geometry to fold the growing intermediate [23]. Taking advantage of the SAT domain's ability to control the starter unit selection among NR-PKSs, Wang and co-workers engineered a novel polyketide by replacing the SAT domain in AfoE (asperfuranone biosynthesis) with the SAT domain from StcA (sterigmatocystin

biosynthesis) [45].

Meanwhile, Du and co-workers reported the first successful domain swapping in fungal HR-PKS and showed that the KS domain of T-toxin PKS1 from *Cochliobolus heterostrophus* does not affect the programmed biosynthesis. [46,47]. Lastly, Cox and co-workers performed a series of rational domain swapping between closely related synthases to probe the programming mechanism in HR-PKSs. They found that the chain length and methylation pattern are controlled by the KR and CMeT-ψKR domains in tenellin synthase (TENS) and desmethylbassianin synthase (DMBS) [48]. The swapping experiments also suggested that the AT, KS and ACP domains play no role in programming (**Figure 1.8**).

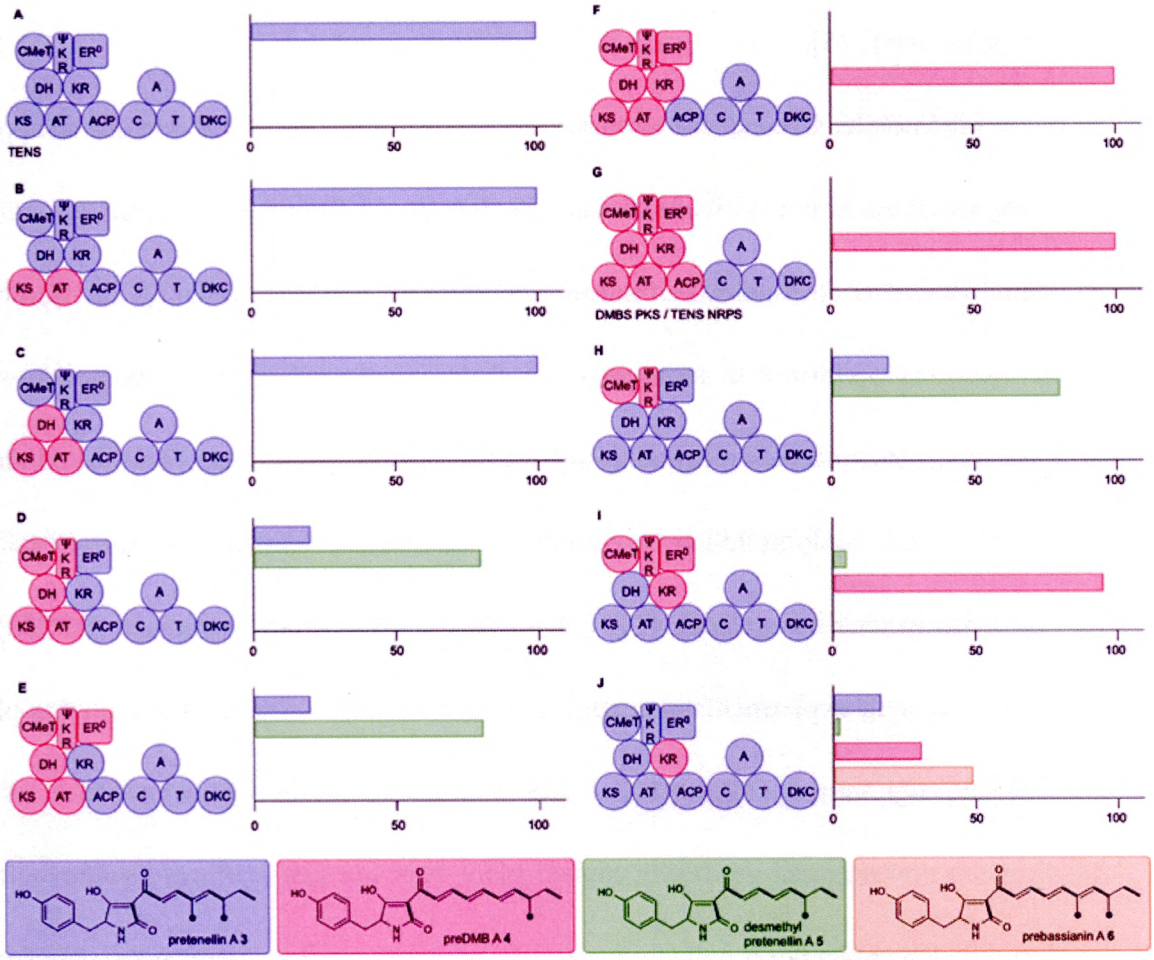


Figure 1.8. Domain swapping between tenellin synthase (TENS) for pretennellin A biosynthesis (purple colored) and desmethylbassianin synthase (DMBS) for predesmethylbassianin A biosynthesis (pink colored) performed by Cox and co-workers. Relative titers of small-molecule products are shown. For TENS and DMBS, the HR-PKS is fused to a single module of a nonribosomal peptide synthase (NRPS). PKS consists of ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP), dehydrase (DH), ketoreductase (KR), pseudo-KR (ψKR) and C-methyltransferase (CMeT) domains. NRPS consists of C, condensation (C), adenylation (A), thiolation (T) and Dieckmann cyclase (DKC) domains. (Figures are adopted from reference [48]).

1.5 Organization of the dissertation

This dissertation contains results from my studies on the bacterial type I iterative PKSs for enediyne and mellein biosynthesis. The results from the studies on the iPKSs CalE8, SgcE and DynE8 for enediyne biosynthesis are presented in chapter 2 and chapter 3; whereas the results from the studies on the iPKS SACE5532 for mellein biosynthesis are described in chapter 4.

The study described in **Chapter 2** aims to unravel the biosynthetic mechanism of the enediyne core of the 9-membered enediyne C-1027. The enediyne PKS SgcE, the thioesterase SgcE10 and the unknown protein SgcE3 were expressed in *E. coli*. My research work developed an alkaline hydrolysis method to release the PKS-tethered intermediate. *In vitro* enzymatic reactions of enediyne PKS were performed under various conditions to identify the products generated by SgcE in the absence or presence of ancillary proteins. The functions of SgcE3 and SgcE10 are also discussed.

Chapter 3 describes the comparative study of the PKSs from the biosynthetic pathways of 9-membered (C1027) and 10-membered enediynes (calicheamicin and dynemicin) to investigate a possible divergence of the pathways at the early stage of enediyne biosynthesis. The *in vitro* and *in vivo* products of SgcE, CalE8 and DynE8 were characterized to reveal the different properties of the homologous PKSs.

In **Chapter 4**, I introduce the iPKS SACE5532 from *Saccharopolyspora erythraea* that shares similar domain organization with several known bacterial iPKSs for aromatic polyketide biosynthesis. SACE5532 was successfully cloned and expressed in *E. coli*. Enzymatic assays of SACE5532 revealed that it generates a

surprising product (mellein) that is most likely formed by the cyclization of a pentaketide intermediate. SACE5532 was further studied by domain swapping with a homologous iPKS NcsB to probe the mechanism of programming. To understand the programmed keto-reduction, the stand-alone KR domains of SACE5532 and NcsB were cloned, expressed and characterized.

CHAPTER 2

Study of the iPKS and accessory enzymes in the biosynthesis of the enediyne antibiotic C-1027

2.1 Introduction

2.1.1 Enediyne natural products

Enediyne natural products are secondary metabolites produced by a variety of Gram-positive bacteria *actinomycetes* found in both soil and sea sediment. This class of natural products is of great interest and value because of their unprecedented molecular structure, remarkable biological activities, and fascinating mode of action [49-55].

Enediyne natural products exhibit potent antibiotic and antitumor activities[56] and have spawned considerable interest as anticancer agents in the pharmaceutical industry. It was hoped that the enediynes can be developed into powerful anticancer drugs if they can be delivered specifically to tumor cells. A variety of polymer-based delivery systems or enediyne antibody conjugates have shown great clinical promise in anticancer chemotherapy [51,53-55,57,58]. Despite the seemingly structural diversity, the natural enediynes share a remarkable mechanism of action with their potent cytotoxicity attributed to the bicyclic enediyne core. Upon the triggering by an

environmental stimulus, the enediyne core undergoes Bergman or Myers-Saito cyclization to form a benzenoid diradical. The highly reactive diradical would then strip hydrogen atoms from the sugar phosphate backbone of DNA and lead to chromosomal DNA cleavage [52,59-62].

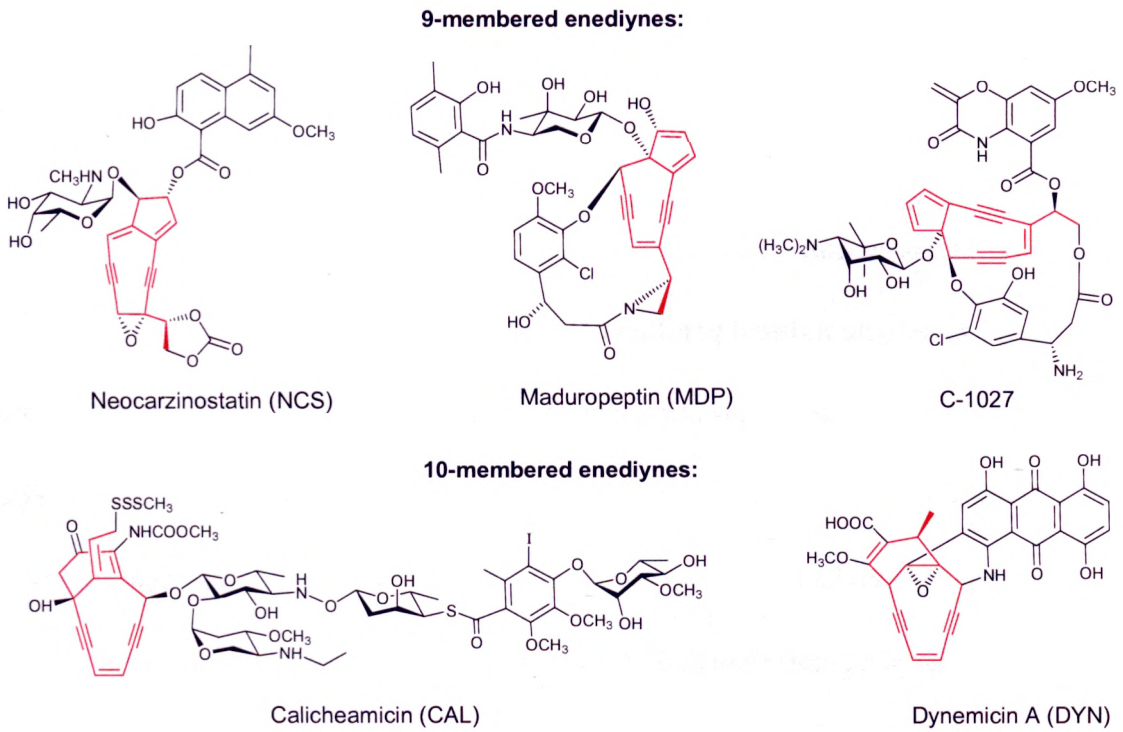


Figure 2.1. Representative 9- and 10-membered enediynes with the enediyne core highlighted in red.

The enediyne family of antibiotics displays fascinating structures that are highly unusual for natural products. All members share a characteristic enediyne core containing two acetylenic groups conjugated to a double bond within a nine-membered ring (e.g. C-1027) or ten-membered ring (e.g. calicheamicin) (**Figure 2.1**). To date, thirteen enediynes have been structurally elucidated [63-77], all of which can be

grouped into two categories according to the enediyne core structures. Members of the 9-membered enediyne category, such as C-1027, from a culture filtrate of *Streptomyces globisporus* C-1027 [67,78-81], are the chromoprotein enediynes that consist of the enediyne chromophore and an apo-protein for chromophore stabilization. C-1027 is a representative 9-membered enediyne natural product isolated from a culture filtrate of *Streptomyces globisporus* C-1027 [67,78-81]. Members of the 10-membered enediyne category, such as dynemicin (DYN) from *Micromonospora chersina* [66] and calicheamicin (CAL) from *Micromonospora echinospora ssp. calichensis* [65], contain the enediyne chromophore without any additional stabilizing proteins.

2.1.2 Biosynthesis of the enediyne core

Since the discovery of the first enediyne natural products, they have attracted many chemists and microbiologists to study the unparalleled biosynthetic mechanism of their unique molecular scaffolds. The study of the biosynthetic mechanism of the enediyne core is important not only because the enediyne core displays unprecedented structure but also because it is essential for the biological activity of enediynes. Biosynthetic studies on the enediyne family have been hindered by many factors, including their extreme instability and low production in fermentation that render them difficult to carry out in vivo feeding experiments [82,83]. Despite the difficulties, feeding experiments with ^{13}C-labeled precursors established that both the 9- and 10-membered enediyne cores were derived from eight head-to-tail acetate units [84-86],

though it was unclear until recently whether this the enediynes are derived from a polyketide or fatty acid precursor.

C-1027 is a representative 9-membered enediyne natural product isolated from a culture filtrate of *Streptomyces globisporus* C-1027 [67,78-81]. In addition to the 9-membered enediyne core, C-1027 contains four biosynthetic building blocks that are the enediyne core, deoxy aminosugar, β-amino acid and benzoxazolinate. Shen and co-workers cloned, sequenced and characterized the 85-kb C-1027 biosynthesis gene cluster and identified genes encoding the four biosynthetic building blocks, including *SgcE* to *sgcE11* that putatively encode the proteins for the biosynthesis of the enediyne core [41]. *SgcE* encodes a unique iterative type I PKS and the flanking genes *sgcE1* to *sgcE11* presumably encode accessory proteins for enediyne core biosynthesis.

Recent sequencing of the gene clusters for the production of other two 9-membered enediynes (neocarzinostatin [35] and maduropeptin [38]) and two 10-membered enediynes (calicheamicin [42] and dynemicin [43]) clearly revealed a common polyketide pathway for the biosynthesis of both 9- and 10-membered enediyne cores. Comparing the putative enediyne core biosynthetic genes of five enediyne biosynthetic clusters, a cassette of five genes are the most conserved (**Figure 2.2A**) [87]. The five genes are predicted to encode a novel enediyne PKS (PKSE), a thioesterase (TE) and three additional proteins of unknown function. Interestingly, a phylogenetic analysis of PKSs from the five gene clusters, including three from 9-membered enediynes and two from 10-membered enediynes, reveals a clear genotypic distinction between the two structural families (**Figure 2.2B**).

Based on the bioinformatic information, a possible mechanism for the C-1027 enediyne core biosynthesis includes the following steps: (a) the enediyne PKS is supposed to catalyze the assembly of a nascent linear polyunsaturated intermediate from the acyl CoA precursors in an iterative process, and (b) the nascent intermediate is subsequently desaturated to furnish the two yne groups and cyclized by accessory enzymes to yield either a 9- or 10-membered enediyne core. (c) the enediyne cores are further decorated by modifying enzymes. Such mechanism remains to be validated with most of the flanking genes encode proteins are of unknown function.

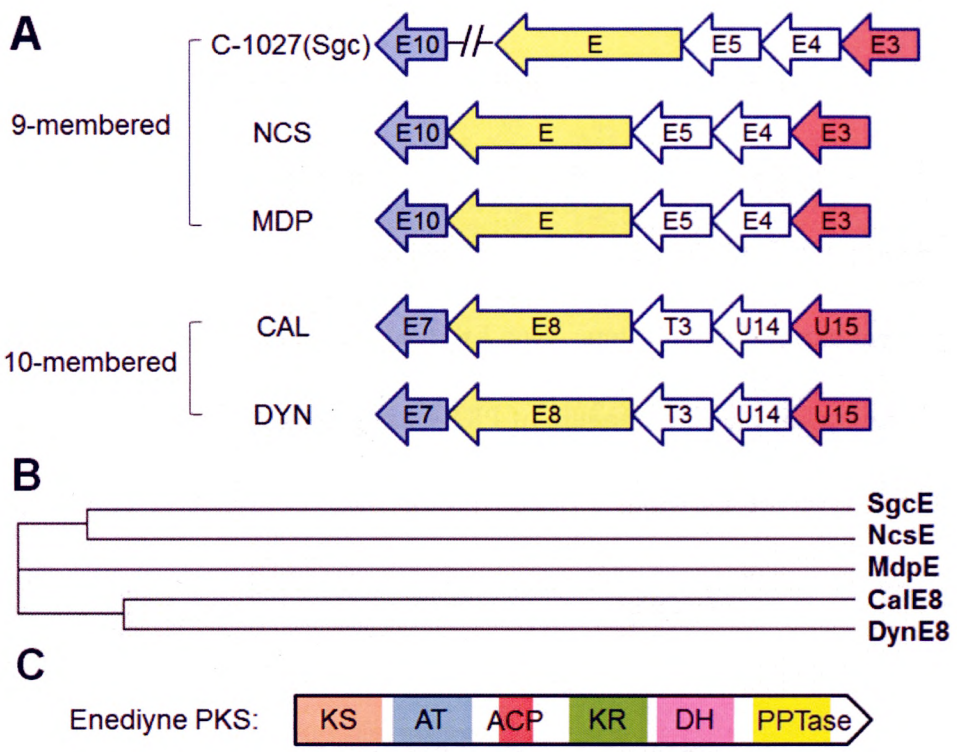


Figure 2.2. (A) Highly conserved gene cassettes for enediyne core biosynthesis. *pksE* (enediyne PKS) genes are in yellow, *TE* (thioesterase) genes are in blue, and *U15* genes are in red. (B) Phylogenetic analysis of PKSE among the five gene clusters. (C) Domain composition of enediyne PKS. From left to right: KS (ketosynthase), AT (acyltransferase), ACP (acyl carrier protein), KR (ketoreductase), DH (dehydratase) and PPTase (phosphopantetheinyl transferase).

Enediyne PKSs, such as SgcE, belong to iterative type I PKSs, which performed chain extension in an iterative fashion. Shen and co-workers reported that SgcE has six predicted domains based on sequence homology and active site mapping [44] (**Figure 2.2C**). The six domains include a ketosynthase (KS), an acyltransferase (AT), an acyl carrier protein (ACP), a ketoreductase (KR), a dehydratase (DH), and a phosphopantetheinyl transferase (PPTase) domain [44]. The ACP domain is first modified posttranslationally by covalent tethering of the 4'-phosphopantetheine (4'-PP) component of CoA onto a conserved Ser, which is presumably catalyzed by an unusual integral C-terminal PPTase domain.

The putative thioesterase SgcE10 shares similarity with the TEBC protein family as represented by 4-hydroxybenzoyl-CoA thioesterase (4HBT). 4HBT is a homotetrameric protein with four active sites and each subunit adopts a so-called hot-dog fold [88,89]. Two types of thioesterases are commonly found in natural product biosynthetic clusters. Type I thioesterases (TE-Is) are usually integrated at the C terminus of the final module of PKS or nonribosomal peptide synthase (NRPS) for the removal of the final product through macrocyclization or hydrolysis. Type II thioesterases (TE-IIs) are usually discrete proteins that remove aberrant units from carrier domains. SgcE10 and its homolog are discrete TEs but do not contain the two conserved motifs for TE-II. The recent biochemical and structural study of its homolog CalE7 in our lab revealed that CalE7 is a novel hotdog fold protein with thioesterase activity [90].

Meanwhile, SgcE3 and its homologs CalU15 and DynU15 belong to a family of

unknown protein (UNBL). The SgcE3 homologs exhibit no significant homology to other known proteins in the public databases. However, SgcE3 shares minimum sequence homology with a family of diiron cluster-containing desaturases or acetylenases found in plant and fungi [91,92]. The structure model of SgcE3 (SWISS-MODEL) based on Δ^9 stearoyl-acyl carrier protein desaturase (PDB: 1AFR) (identity 6.32%, similarity 12.75%) suggests a helical protein structure characteristic of a di-iron protein. Although the local features of the model are not very reliable, the predicted global structure of SgcE3 is likely to be meaningful. Multiple sequence alignment of SgcE3 and three di-iron cluster-containing desaturases revealed that the residues for coordinating the two iron ions are not strictly conserved (**Figure 2.3**).

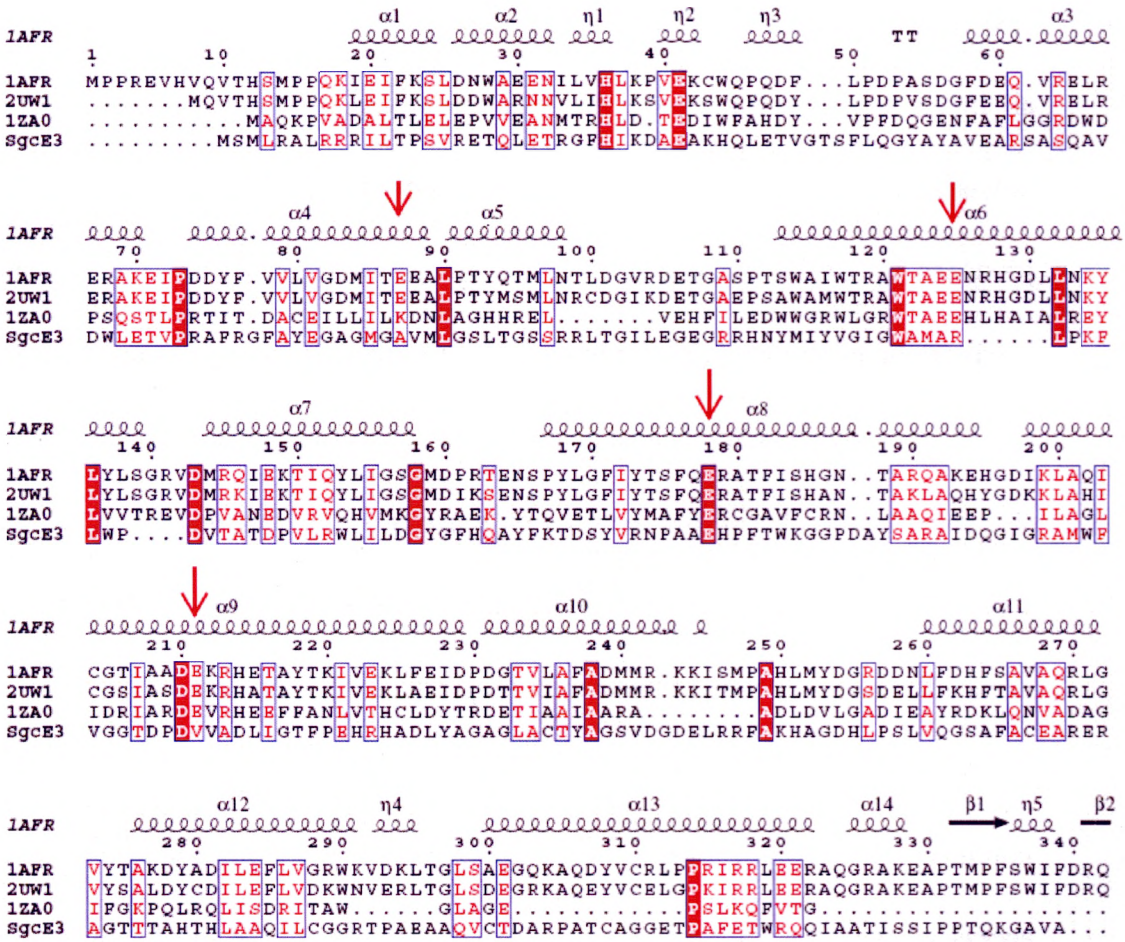


Figure 2.3. Multiple sequence alignment of SgcE3 and three di-iron cluster containing desaturases, Δ^9 stearoyl-acyl carrier protein desaturase (1AFR) [93], acyl-ACP desaturase DesA2 (1ZA0) [94] and ivy Δ^4-16:0-ACP desaturase (2UW1) [95]. The alignment was produced using CLUSTALW2. Strictly identical residues are highlighted by white characters in red box. Conserved residues are identified by red characters in blue frame. The secondary structure of 1AFR is showed according to the PDB file. The ligands to the iron cluster in 1AFR are indicated by arrows.

Guided by DNA sequencing results, considerable efforts have been made to elucidate the biosynthetic pathways of enediyne natural products in recent years. While the project is underway, Shen and co-workers reported a conjugated polyene 1,3,5,7,9,11,13-pentadecaheptaene extracted from the insoluble debris of the cells co-expressing SgcE and SgcE10 [44]. Later on, they isolated the same chemical from cell pellets co-expressing PKS-TE pairs from all five enediyne core biosynthetic clusters. Guo et al. confirmed that the six double bonds in the conjugated polyene are all in the *trans* configuration [96]. Our group reported that a carbonyl-conjugated polyene 3,5,7,9,11,13-pentadecen-2-one instead of the conjugated polyene can be produced by CalE8 and CalE7 from the biosynthetic pathway of the 10-membered enediyne calicheamicin [97]. Townsend's group observed both produced by CalE8 and CalE7 without quenching the reactions with trifluoroacetic acid (TFA) [98]. They suggested that neither of the reported products is a true intermediate in the enediyne core biosynthetic pathway and implied that one or more accessory enzymes may be needed for the production of the correct biosynthetic intermediate.

Most of these studies depend on a discrete thioesterase to release the PKS-tethered intermediate and little work is done in exploring the potential function of other accessory enzymes. In contrast to the *in vivo* studies carried out by Shen's lab, my

research aims to express and purify the PKS SgcE and its ancillary enzymes for *in vitro* enzymatic assays. *In vitro* reactions of PKS will be performed under various conditions to investigate whether SgcE can synthesize other products with and without other ancillary proteins. My research work develops an alkaline hydrolysis method to release the PKS-tethered intermediate and characterize the PKS product without reliance on the thioesterase. The protein SgcE3 will be studied both *in vitro* and *in vivo* to probe its function. In the next chapter, the biosynthetic pathways of 9-membered enediyne C-1027 will be compared with those of two 10-membered enediynes, calicheamicin and dynemicin to investigate a possible divergence of the pathways at the early stage of enediyne biosynthesis.

2.2 Materials and Methods

2.2.1 Strains, plasmids, and chemicals

Vectors pET-28b(+), pCDF-2 Ek/LIC, and LIC Duet™ Minimal Adaptor were obtained from Novagen. BL21 (DE3) competent cells were from Novagen. The genes that encode SgcE, SgcE10 and SgcE3 were provided by GenScript Corporation (NJ, USA) with sequences optimized for protein expression in *E. coli*. The reported protein sequences for SgcE, SgcE10 and SgcE3 have been deposited in Genbank under the accession numbers of AAL06699, AAL06692 and AAL06702. The genes were synthesized based on the reported protein sequences and were provided as pUC57-based plasmids. XhoI and NdeI restriction sites are added to the C-terminus and N-terminus respectively. BL21 (DE3) competent cells were from Novagen. Acetyl-CoA, malonyl-CoA, NADPH and other chemicals were purchased from Sigma-Aldrich and stored at -20 °C.

Yeast strain *Saccharomyces cerevisiae* 1000M/FY1679-08A (MATα; ura3-52; leu2Δ1; trp1Δ63; his3Δ200; GAL2) and vectors (**Table 2.1**) were kindly provided by Dr. Thiru's lab from the School of Biological Sciences in Nanyang Technological University.

Table 2.1. Plasmids used in protein expression in *S. cerevisiae*

Vector Name	Resistance	Selective marker	Remarks
YCplac111	Amp	Leu	Low copy
YEplac112	Amp	Trp	High copy
YEplac181	Amp	Leu	High copy

2.2.2 Cloning, expression and purification of enzymes in *E. coli*

2.2.2.1 Cloning and mutagenesis

The plasmid pUC57-SgcE was digested with NdeI-XhoI to yield SgcE. The 5.8-kb SgcE fragment was gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-SgcE. The ligation mixture was introduced to *E. coli* Top10 competent cells and cells were plated on LB medium supplemented with 50 μg/ml kanamycin. The colonies were screened by PCR to confirm the presence of pET28-SgcE. Then the plasmid was amplified in *E. coli* Top10 and purified for sequencing.

To generate the SgcE-C211A mutant, the PCR reaction was performed using pET28-SgcE as a template with the following primers: C211A (forward) 5'-CGTGGACGGCGCGGCCTCTTCAAGCCTG-3'/ (reverse) 5'-CAGGCTTGAAGAGGCCGCGCCGTCCACG-3'. Successful reaction mixtures consisted of 100 ng of template DNA, 300 nM each primer, 300 mM dNTPs, , 1x KAPAHiFi™ Fidelity Buffer, and 0.5 U of KAPAHiFi™ DNA Polymerase in a final volume of 25 μl. The PCR program was as follows: initial denaturing at 96°C for 5 minutes, followed by 18 cycles at 96°C for 50 seconds, 55°C for 50 seconds, and 68°C for 6 minutes, and completed by an additional 7 minutes at 68°C. Upon completion, 1 μl (10 U) of DpnI was added directly to the PCR mixture and digested at 37°C for 2 hours. An aliquot (5 μl) of the mixture was directly transformed into *E. coli* Top10 competent cells and plated on LB supplemented with 50 μg/ml kanamycin. The mutant constructs were confirmed by sequencing to give pET28-SgcE-C211A.

The genes for SgcE10 and SgcE3 were amplified by PCR using pUC-SgcE10 as a template for sgcE10 and pUC-SgcE3 for SgcE3. Reactions were performed using the following primers: SgcE10 (forward) 5'-GACGACGACAAGATTCATATGATGACCGCG-3'/ (reverse) 5'-GAGGAGAAGCCCGGTCTCGAGTTACGCCG-3'; SgcE3 (forward) 5'-GACGACGACAAGATTCATATGATGAGCATGCTGCGC-3'/ (reverse) 5'-GAGGAGAAGCCCGGTCTCGAGTTACGCCAC-3'. The gel-purified PCR product was treated with T4 DNA polymerase and annealed into pCDF-2 Ek/LIC vector as described by Novagen, affording pCDF-SgcE10 and pCDF-SgcE3. The LIC annealing reaction was directly transformed into *E. coli* BL21 (ED3) competent cells and cells were plated on LB medium supplemented with 50 μg/ml streptomycin. The colonies were screened by PCR to confirm the presence of pCDF-SgcE10 and pCDF-SgcE3. The plasmids were amplified in *E. coli* BL21 (ED3) and purified for sequencing.

2.2.2.2 Protein expression and purification

Expression and purification of the enediyne polyketide synthases SgcE, and SgcE-C211A mutant. pET28-SgcE or pET28-SgcE-C211A was introduced into *E. coli* BL21(DE3), and the resultant recombinant strains were plated on LB medium (per liter: 10 g NaCl, 10 g tryptone and 5 g yeast extract) supplemented with 50 μg/ml kanamycin. A single colony was used to inoculate 20 ml of LB medium supplemented with 50 μg/ml kanamycin and incubated at 37°C for 12 hours. A 5 ml aliquot was

transferred to 500 ml of LB medium supplemented with 50 μg/ml kanamycin and grown an additional 3h at 37°C at 210 rpm. The cultures were incubated at 16°C and induced with IPTG (final concentration of 0.2 mM) when OD_{600} reached ~0.8 (~3 hours). After incubation at 16°C for additional 18 hours at 160 rpm, cells were harvested and spun at 8,000 rpm. The cell pellet was resuspended in lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, 5 mM β-ME and 10% (v/v) glycerol] and lysed by sonication. After centrifugation at 20,000 rpm for 30 minutes at 4°C, the supernatant was filtered by 0.45 μm membrane and loaded onto HiTrap™ Ni^{2+}-NTA column (GE Healthcare). The column was then washed by lysis buffer and wash buffer containing 40 mM imidazole before eluted with elution buffer containing 500 mM imidazole. The eluted protein was further purified by gel filtration using a HiLoad™ 16/60 Superdex™ 200 column (GE Healthcare). Proteins were desalted into Tris buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT and 10% (v/v) glycerol]. Its purity was determined to be > 90 % by SDS-PAGE. The protein was concentrated, flash frozen in liquid nitrogen, and stored in -80°C freezer.

Expression and purification of the thioesterase SgcE10 and the unknown protein SgcE3. pCDF-SgcE10 or pCDF-SgcE3 was introduced into *E. coli* BL21(DE3), and the resultant recombinant strains were plated on LB medium (per liter: 10 g NaCl, 10 g tryptone and 5 g yeast extract) supplemented with 50 μg/ml streptomycin. The purification of these proteins was performed identically to SgcE except that cells were grown in LB media supplemented with 50 μg/ml streptomycin.

Co-expression of SgcE and SgcE10. pET28-SgcE and pCDF-SgcE10 were

co-transformed into *E. coli* BL21(DE3) for SgcE/E10 co-expression. The cells were plated on LB medium supplemented with 50 μg/ml kanamycin and 50 μg/ml streptomycin. The colonies were screened by PCR to confirm the existence of both pET28-SgcE and pCDF-SgcE10. A single colony was used to inoculate 20 ml of LB medium supplemented with 50 μg/ml kanamycin and 50 μg/ml streptomycin, and incubated at 37°C for 12 hours. A 5 ml aliquot was transferred to 500 ml of LB medium supplemented with 50 μg/ml kanamycin and 50 μg/ml streptomycin, and grown an additional 3h at 37°C at 210 rpm. The cultures were incubated at 16°C and induced with IPTG (final concentration of 0.4 mM) when OD_{600} reached ~0.8 (~3 hours). After incubation at 16°C for an additional 18 hours at 160 rpm, cells were harvested and lysed for protein purification using the procedures described above. As both SgcE and SgcE10 contain $(His)_6$-tag, the purification by Ni^{2+}-NTA affinity chromatography yielded a solution that contains both proteins. The SgcE and SgcE10 were subsequently separated and purified by size exclusion chromatography using a HiLoad™ 16/60 Superdex™ 200 column. The protein concentration and profile were examined by UV-Vis spectrophotometer before the proteins were flash frozen in liquid nitrogen and stored in -80°C freezer.

Co-expression of SgcE and SgcE3. pET28-SgcE and pCDF-SgcE3 were co-transformed into *E. coli* BL21(DE3) for SgcE/E3 co-expression. The protein purification was carried out following the similar procedures described for co-expression of SgcE/E10.

2.2.3 Cloning and expression of enzymes in yeast

2.2.3.1 Plasmid constructions for yeast expression

The plasmid YCplac111-C17 was digested with HindIII-BamHI to yield the fragment C17, which including GAL promoter. C17 was gel-purified and ligated into the identical sites of YEplac112 and YEplac181 to give YEplac112-C17 and YEplac181-C17. The gene sequence of SgcE10 was optimized for yeast expression and provided as pUC57-SgcE10yeast. XhoI, $(His)_6$-tag, stop codon, and BamHI were sequentially introduced onto the C-terminal of the gene. The plasmid pUC57-SgcE10yeast was digested with NdeI-BamHI to yield the fragment SgcE10yeast. SgcE10-yeast was gel-purified and ligated into the identical sites of YEplac112-C17 and YEplac181-C17 to give YEplac112-C17-SgcE10yeast and YEplac181-C17-SgcE10yeast. The ligation mixture was introduced to *E. coli* Top10 competent cells and cells were plated on LB medium supplemented with 75 μg/ml ampicillin. The colonies were screened by NdeI-XhoI double digestion to confirm the existence of YEplac112-C17-SgcE10yeast and YEplac181-C17-SgcE10yeast. Then the plasmid was amplified in *E. coli* Top10 and purified for sequencing.

The plasmid pUC57-SgcE was digested with NdeI-XhoI to yield the fragment SgcE. The 5.8-kb SgcE fragment was gel-purified and ligated into the identical sites of YEplac112-C17-SgcE10yeast to give YEplac112-C17-SgcE. The ligation mixture was introduced to *E. coli* Top10 competent cells and cells were plated on LB medium supplemented with 75 μg/ml ampicillin. The colonies were screened by NdeI-XhoI double digestion to confirm the existence of YEplac112-C17-SgcE. Then the plasmid

was amplified in *E. coli* Top10 and purified for sequencing.

2.2.3.2 Co-expression of SgcE and SgcE10 in yeast

YEplac181-C17-SgcE10yeast and YEplac112-C17-SgcE were co-transformed into *S. cerevisiae* 1000M/FY1679-08A competent cells. The cells were streaked onto Synthetic Defined (SD) selective plate lacking both Leu and Trp. Four colonies were picked and further maintained on SD plate lacking Leu and Trp and then inoculated on YPD (Yeast extract/Peptone/Dextrose) plate. A single colony was inoculated in 20 ml YPD medium, and incubated overnight at 30°C at 220 rpm. Transfer 7 ml overnight cells to 500 ml YPD medium (final OD_{600} is 0.217) in a 2 L flask, and incubate the culture at 30°C at 230 rpm. When OD_{600} reached 0.746, the cells were spun at 2,000 g for 7 minutes. The cell pellet was resuspended in autoclaved distilled water and spun at 2,000 g for 5 minutes. The cell pellet was resuspended in 500 ml YPG (Yeast extract/Peptone/Galactose) medium, and incubated at 30°C at 220 rpm for 24 hours.

Then cells were harvested and spun at 5,000 rpm. The cell pellet (~25 g) was resuspended in 200 ml lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, 5 mM β-ME and 10% (v/v) glycerol] and lysed by French Press with gauge pressure at 3,000 psi and working pressure at 32,000 psi. After centrifugation at 18,000 rpm for 30 minutes at 4°C, the supernatant was filtered by 0.45 μm membrane and loaded onto HiTrap™ Ni^{2+}-NTA column (GE Healthcare). The column was then washed by lysis buffer and wash buffer containing 40 mM imidazole before eluted with elution buffer containing 500 mM imidazole. The eluted SgcE and SgcE10 were

further purified by gel filtration using a Superdex-200 column (GE Healthcare). Proteins were desalted into Tris buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT and 10% (v/v) glycerol]. Its purity was determined by SDS-PAGE. The protein concentration and profile were examined by UV-Vis spectrometer before the proteins were flash frozen in liquid nitrogen and stored in -80°C freezer.

2.2.4 Alkaline hydrolytic release of PKS products

To release the product covalently attached to PKS expressed alone in *E. coli*, the following hydrolysis procedure was used to cleave the thioester bond between the product and the phosphopantetheine thiol of ACP domain in PKS. A 200 μl concentrated protein solution was mixed with 800 μl NaOH solution [5 mM NaOH (pH 10.8), 100 mM NaCl]. The solution was incubated at 37 °C in water bath for 3 hours. The products released from the proteins were analyzed by HPLC method.

2.2.5 *In vitro* enzymatic assays

For enzymatic assay with SgcE, a typical reaction contains 0.25 mM NADPH, 1.0 mM malonyl CoA, 0.15 mM acetyl CoA and 6.65 μM PKS in the reaction buffer [50 mM Tris (pH 8.5), 150 mM NaCl and 1 mM DTT] in a total volume of 400 μl. The reaction mixture was incubated at 30°C for 2 hours.

For enzymatic assays with SgcE and SgcE10, a typical enzymatic reaction contains 0.075 mM NADPH, 1.4 mM malonyl CoA, 3.1 μM PKS and 62.9 μM TE in the reaction buffer [50 mM Tris (pH 8.5), 150 mM NaCl and 1 mM DTT] in a total

volume of 200 μl. The reaction mixture was incubated at 37°C for 2 hours. The temperature dependent reaction was carried out with the same components at 23°C, 30°C or 37°C. The malonyl-CoA concentration dependent reaction was carried out with the same components, with varying malonyl-CoA concentration of 0.5 mM or 1.4 mM. The NADPH concentration-dependent reaction was carried out with the same components, with varying NADPH concentration of 0.075 mM or 0.35 mM. The pH reaction was carried out with the same components at 7.0 or 8.5.

Experimental conditions for the enzymatic assays in Figure 2.9. (A) The temperature-dependent product formation for SgcE-SgcE10. Assay conditions: 3.1 μM SgcE, 62.9 μM SgcE10, 0.075mM NADPH, 1.4 mM Malonyl-CoA, 50 mM Tris (pH8.5), 300 mM NaCl, 1 mM DTT. (B) The malonyl CoA-dependent product formation for SgcE-SgcE10. Assay conditions: 3.1 μM SgcE, 62.9 μM SgcE10, 0.075 mM NADPH, 50 mM Tris (pH 8.5), 300 mM NaCl, 1 mM DTT, 37°C. (C) The NADPH-dependent product formation for SgcE-SgcE10. Assay conditions: 3.1 μM SgcE, 62.9 μM SgcE10, 1.4 mM Malonyl-CoA, 50 mM Tris (pH8.5), 300 mM NaCl, 1 mM DTT, 37°C.

Experimental conditions for the enzymatic assays in Figure 2.10. (A) SgcE-SgcE10 reaction. Reaction conditions: 2.88 μM SgcE, 64.05 μM SgcE10, 0.25 mM NADPH, 50 mM Tris (pH8.5), 30°C (B) SgcE-SgcE10-SgcE3 reaction. Reaction conditions: 2.88 μM SgcE, 64.05 μM SgcE10, 4.56 μM SgcE3, 0.25 mM NADPH, 50 mM Tris (pH8.5), 30°C.

The reactions were conducted in a semi-micro quartz cuvette and monitored with

a Shimazu UV-Vis 1700 spectrophotometer. The sample chamber was kept at certain degree through an external temperature controller. The absorption spectrum of the reaction mixture was taken at various time intervals to monitor the progress of product formation.

2.2.6 Product extraction

For product extraction from proteins and reactions, an equal volume of ethyl acetate was added into the protein solution or reaction mixture and vigorous vortexing ensued. Subsequently the mixture was centrifuged at 14,500 rpm for 15 minutes. The resulting organic extract was evaporated by using a Speed-Vac. The dried sample was re-dissolved in 20 μl methanol for HPLC analysis.

For product extraction from cell pellet, the insoluble cell debris after sonication was suspended with 5 ml 0.1 M sodium acetate pH 6.0. An equal volume of ethyl acetate was then added into the suspended mixture and vigorous vortexing ensued. Subsequently the mixture was centrifuged at 14,500 rpm for 15 minutes. The resulting organic extract was evaporated by using a Speed-Vac. The dried sample was re-dissolved in 20 μl methanol for HPLC analysis.

2.2.7 HPLC analysis

For SgcE hydrolysis, SgcE reactions and SgcE/E3 co-expression, the HPLC analysis was performed with an analytical eclipse XDB C18 column (4.6 × 150 mm) using an Agilent 1200 HPLC. A full gradient was employed from 100% buffer A

(HPLC grade water with 0.045% TFA) to 100% buffer B (100% acetonitrile with 0.045% TFA) at 1 ml/min in 60 minutes.

For SgcE-SgcE10 reactions, SgE-SgcE10-SgcE3 reaction and SgcE/E10 co-expression, the HPLC analysis was performed with an analytical eclipse XDB C18 column (4.6 × 250 mm) using an Agilent 1200 HPLC. A full gradient was employed from 70% buffer A (HPLC grade water with 0.045% TFA) + 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B at 1 ml/min in 60 minutes.

2.2.8 LC-MS analysis

The column used for LC-MS analysis was eclipse XDB C18 column with a dimension of 1.0 × 150 mm. The gradient employed in the analysis was from 90% buffer A (HPLC grade water with 0.01% FA) + 10% buffer B (100% acetonitrile with 0.01% FA) to 100% buffer B in 60 minutes. The ionization energy was set with ESI or APCI ionization source for the Finnigan LTQ Orbitrap mass spectrometer (Thermo Electron). The result was analyzed with the software Xcalibur for the determination of plausible molecular compositions based on the observed molecular weight and fragmentation pattern.

2.3 Results

2.3.1 Purification and characterization of enzymes

The *E. coli* cells that contained the over-expressed SgcE appeared bright orange after harvest. The pET-28b(+) vector produced an N-terminal $(His)_6$-tagged recombinant protein to facilitate the purification of the protein by Ni^{2+}-NTA affinity chromatography. The eluted protein solution appeared orange, indicating that it is the recombinant protein that contributes to the bright color of the cells. Denaturation of the protein did not release the orange pigment from the protein, suggesting that the pigment was most likely covalently attached to SgcE.

The protein purified after gel filtration was approximately 95% pure as estimated by SDS-PAGE analysis (**Figure 2.4A**). The theoretical molecular weight of SgcE is 203.4 kDa and gel filtration analysis showed that the elution volume of the protein was 54.95 ml with an estimated molecular mass of ~400 kDa, suggesting that SgcE is present as a dimer in solution. The concentrated protein appeared to be stable at -80°C for several months and at 4°C for one month.

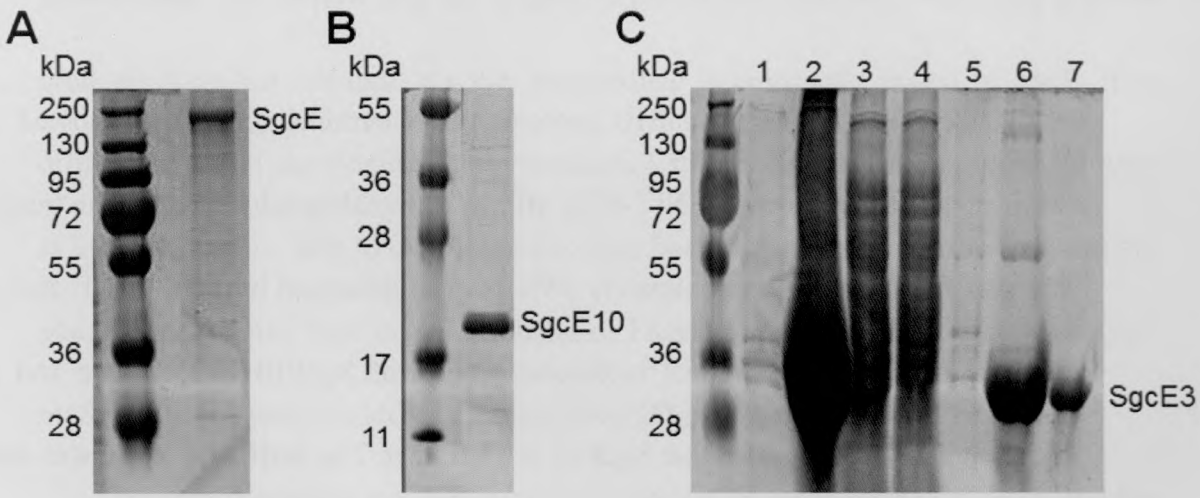


Figure 2.4. SDS-PAGE gel images for SgcE, SgcE10 and SgcE3 after gel filtration. (A) SDS-PAGE for SgcE. (B) SDS-PAGE for SgcE10. (C) SDS-PAGE analysis of SgcE3. Lane 1 – Uninduced, Lane 2 – Insoluble, Lane 3 - Soluble, Lane 4 – Flow through, Lane 5 – Wash, Lane 6 – Elution, Lane 7 – After gel filtration.

The UV-Vis spectrum of SgcE exhibits three peaks at 371, 391 and 413 nm (**Figure 2.5A**). When the predicted key catalytic Cys residue of the KS domain (Cys^{211}) was replaced with Ala, the UV-Vis spectrum of the SgcE mutant lacks the absorbance in the 300-500 nm range (**Figure 2.5B**). This further suggests that the orange pigment is indeed produced by SgcE and likely derived from a polyketide intermediate.

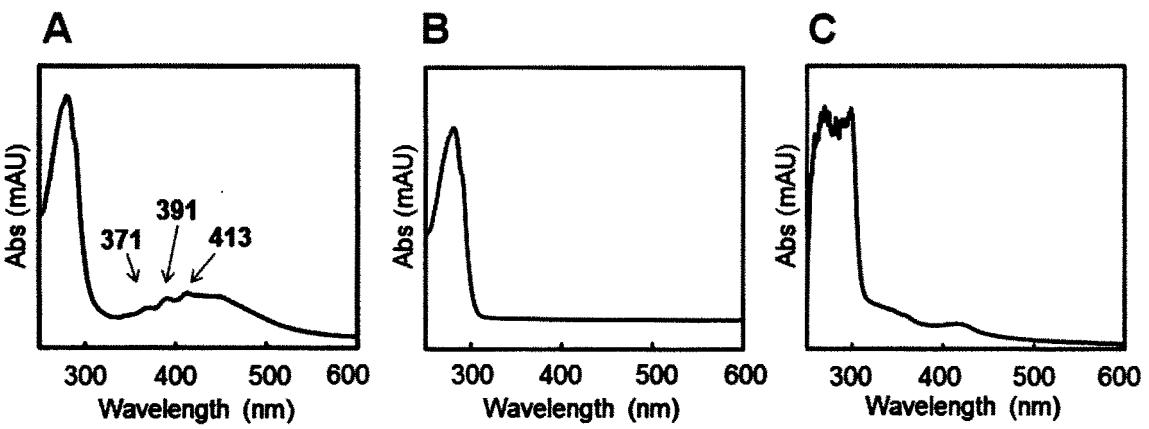


Figure 2.5. (A) UV-Vis spectrum of SgcE. (B) UV-Vis spectrum of the SgcE-C211A mutant. (C) UV-Vis spectrum of concentrated SgcE3.

For the recombinant SgcE10 protein, an N-terminal $(His)_6$-tag allowed the purification of the protein by Ni^{2+}-NTA affinity chromatography. The protein purified after gel filtration was approximately 95% pure as estimated by SDS-PAGE analysis (**Figure 2.4B**). The theoretical molecular weight of SgcE10 is 17.7 kDa and gel filtration analysis suggested that SgcE10 is a tetramer. The purified enzyme was stable at -80°C for several months and at 4°C for one month.

For the recombinant SgcE3 protein, an N-terminal $(His)_6$-tag also allowed the purification of the protein by Ni^{2+}-NTA affinity chromatography. Approximately 50% of the total SgcE3 produced was found in inclusion body. The protein purified after gel filtration was approximately 95% pure as estimated by SDS-PAGE analysis. Progress of the expression and purification was monitored by SDS-PAGE gel (**Figure 2.4C**). The theoretical molecular weight of SgcE3 is 35.5 kDa and gel filtration analysis suggests that SgcE3 is a dimeric protein. The slightly brownish SgcE3 protein appeared to be a rather unstable protein that prone to precipitation at 4°C. The protein exhibited broad absorption bands from 300 to 450 nm (**Figure 2.5C**), reminiscent of the diiron-oxo cluster-containing desaturase despite of the lack of other evidence that it is a diiron protein [92].

2.3.2 Hydrolytic release of PKS products and *in vitro* enzymatic assays

2.3.2.1 Hydrolytic release of the SgcE-tethered products

When the denatured SgcE protein solution was extracted with organic solvent, the protein layer was yellow and the organic layer looked colorless, indicating that the products were not released. As the intermediate is covalently bound to SgcE, it is difficult to reveal the identity of the products. It has been shown that the hydrolysis at high pH may be able to cleave the thioester bond between the intermediate and the phosphopantetheine thiol of ACP domain of PKSs. We performed alkaline hydrolysis on SgcE in an attempt to identify the products (**Figure 2.6A**).

The protein solution was treated with 5 mM NaOH and incubated at 37 °C and the

solution was extracted with ethyl acetate. After extraction, the organic solvent layer looked yellow and the yellow coloration of protein was diminished, indicating that the products were successfully released. HPLC analysis detected compound **1** as the major hydrolyzed product of SgcE, which exhibits absorption maximum at 380 nm (**Figure 2.6C**). Because the product **1** is such a difficult compound to be ionized by MS-ESI ion source that I could not obtain the molecular mass by LC-ESI-MS analysis. When we turned to LC-APCI-MS, the signals were too noisy to confirm the true mass of **1**. So, further structure determination of **1** was hindered by lack of mass data and insufficient compound for NMR analysis.

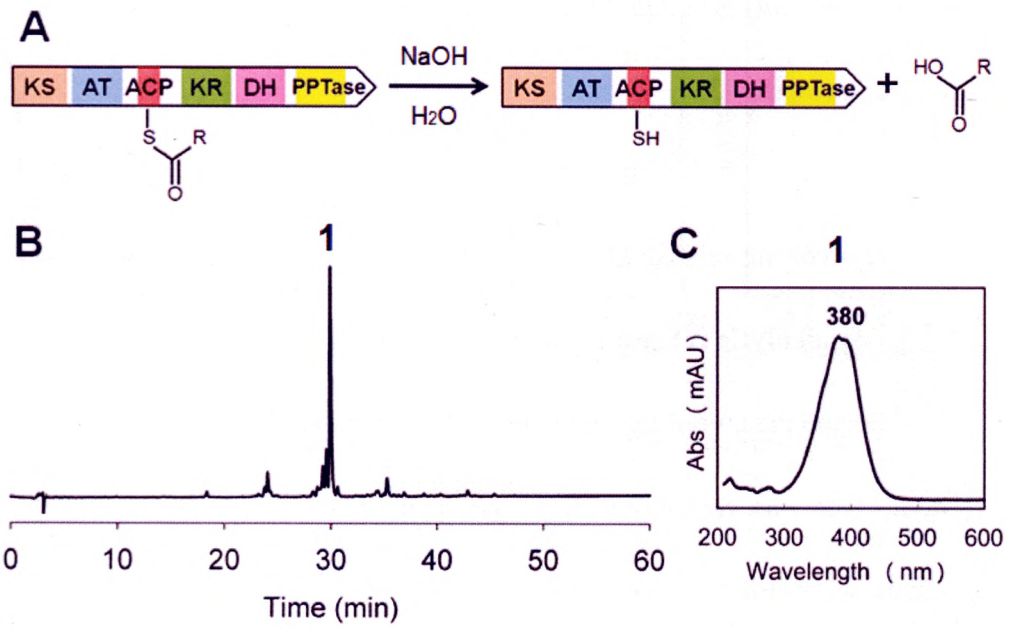


Figure 2.6. (A) Schematic illustration of the hydrolysis of SgcE to release the product. (B) HPLC analysis of hydrolyzed product of single-expressed SgcE at 400 nm. HPLC program: 0-60 min, 100% A (H_2O, 0.045% TFA) to 100% B (CH_3CN, 0.045% TFA) using an eclipse XDB C18 column (4.6 × 150 mm). (C) UV-Vis spectrum of **1**.

2.3.2.2 Enzymatic assays with SgcE

The enzymatic activity of SgcE was examined by *in vitro* activity assay. The

enzymatic assay was performed by incubating SgcE with acetyl CoA, malonyl CoA and NADPH under the experimental conditions described in the method section. Later on, we found that acetyl CoA is not needed for the reaction. The SgcE used in the *in vitro* reaction was the SgcE purified from SgcE/E3 co-expression. The reaction progression was observed by UV-Vis spectroscopy with time. The incubation of reaction mixture without malonyl CoA showed no change in the UV-Vis spectrum. As soon as malonyl CoA was added into the mixture, product peaks in the 400-450-nm range increased, and NADPH peak at 342 nm decreased concurrently (**Figure 2.7**). As the major peaks of the product are within 400-450-nm range, HPLC was performed using UV detection at various wavelengths within that range. Product **2** was detected with the UV detection at 400 nm and it exhibits absorption maximum at 420 nm (**Figure 2.8**). Surprisingly, SgcE consumed NADPH even without acetyl-CoA, indicating that the PKS is able to generate acetyl group from malonyl-CoA through an unknown mechanism.

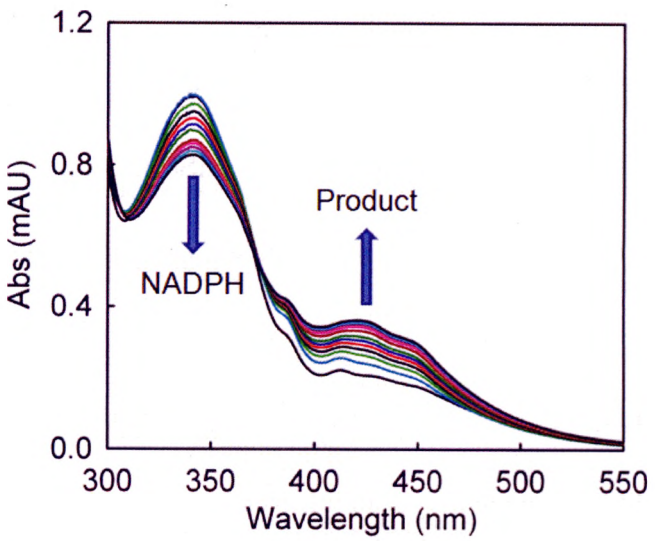


Figure 2.7. Enzymatic assay of SgcE as monitored by absorption spectroscopy. Spectra were taken at 10 minutes interval in a 2-hour time span.

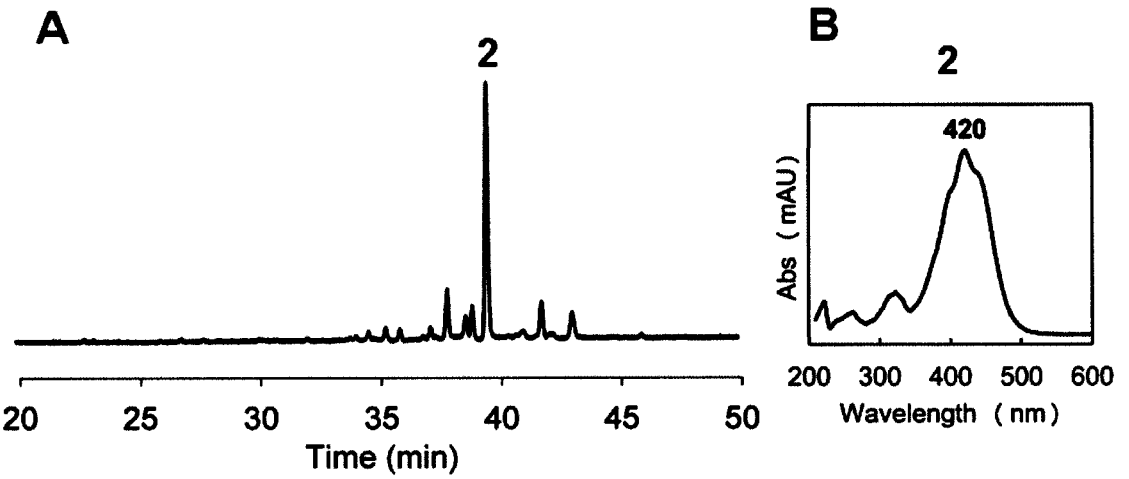


Figure 2.8. (A) HPLC analysis of the product of SgcE reaction with UV detection at 400 nm. HPLC conditions: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm). (B) UV-Vis spectrum of product **2**.

Because **2** turned out to be a difficult compound to be ionized by MS-ESI ion source, we intially could not obtain the molecular mass by LC-ESI-MS analysis. High-resolution LC-APCI-MS was performed to yield a $[M+H]^+$ ion with a *m/z* of 241.1202 for **2**, which suggests a molecular formula of $C_{16}H_{16}O_2$ (calculated *m/z* = 240.1150) with nine degrees of unsaturation. As the linear structure can only afford maximum of eight double bonds for sixteen carbons, we suspect that the compound adopts a cyclic structure. Moreover, the MS/MS spectrum exhibits a series of fragments to suggest that the cyclic compound is likely an aliphatic, but not aromatic. We speculated the **2** to be a highly conjugated molecule with seven double bonds and one ring, which is consistent with the long wavelength absorption at 420 nm resulted from the highly conjugated system. However, the amount of **2** was not sufficient for NMR analysis, so the structure of **2** remains to be determined.

2.3.2.3 Enzymatic assays with SgcE and SgcE10

To probe the function of SgcE10, the enzymatic activity of the thioesterase SgcE10 was examined by *in vitro* activity assays. The enzymatic assays were carried out by incubating the recombinant SgcE and SgcE10 with malonyl-CoA and NADPH under varying conditions described in the method section. The SgcE used in the *in vitro* reaction was the product-free SgcE purified from the SgcE/E10 co-expression Since SgcE seems to be able to generate acetyl group from malonyl-CoA, no acetyl-CoA was added in the reaction mixture..

The enzymatic assays showed that SgcE and SgcE10 produced a series of products that include the previously observed **2** and two new products (**3** and **4**) (**Figure 2.9A-C**). Product **3** exhibits absorbance at 355, 373 and 394 nm and a $[M+H]^+$ ion with *m/z* of 199.15 (**Figure 2.9D**), consistent with the linear conjugated polyene product reported by Shen and co-workers [44]; whereas product **4** exhibits absorbance at 395 nm and a $[M+H]^+$ ion with *m/z* of 215.14, consistent with the carbonyl conjugated polyene structure reported by our group previously for the homologous CalE8 protein [97]. The comparison of **4** with the synthesized all *trans* (3*E*, 5*E*, 7*E*, 9*E*, 11*E*, 13*E*)-pentadecen-2-one established that all double bonds of **4** are *E* configuration as we will discuss in detail in Chapter 3.

The efficient production of **2** observed in the SgcE reaction in the absence of SgcE10 reveals that a discrete PKS product can be released by SgcE under certain assay conditions. This leads us to suspect that the thioesterase SgcE10 may be a type II thioesterase (TE-II), which is not responsible for releasing the final product from PKS

but for the removal of aberrant acyl units from carrier domains. We cannot rule out the possibility that SgcE10 functions as type II TE with relaxed substrate specificity. One of such example is pikromycin TE-II (pikAV), which exhibits no apparent ACP specificity and the high level expression of pikAV in *Streptomyces venezuelae* leads to significant decrease in aglycone production [99].

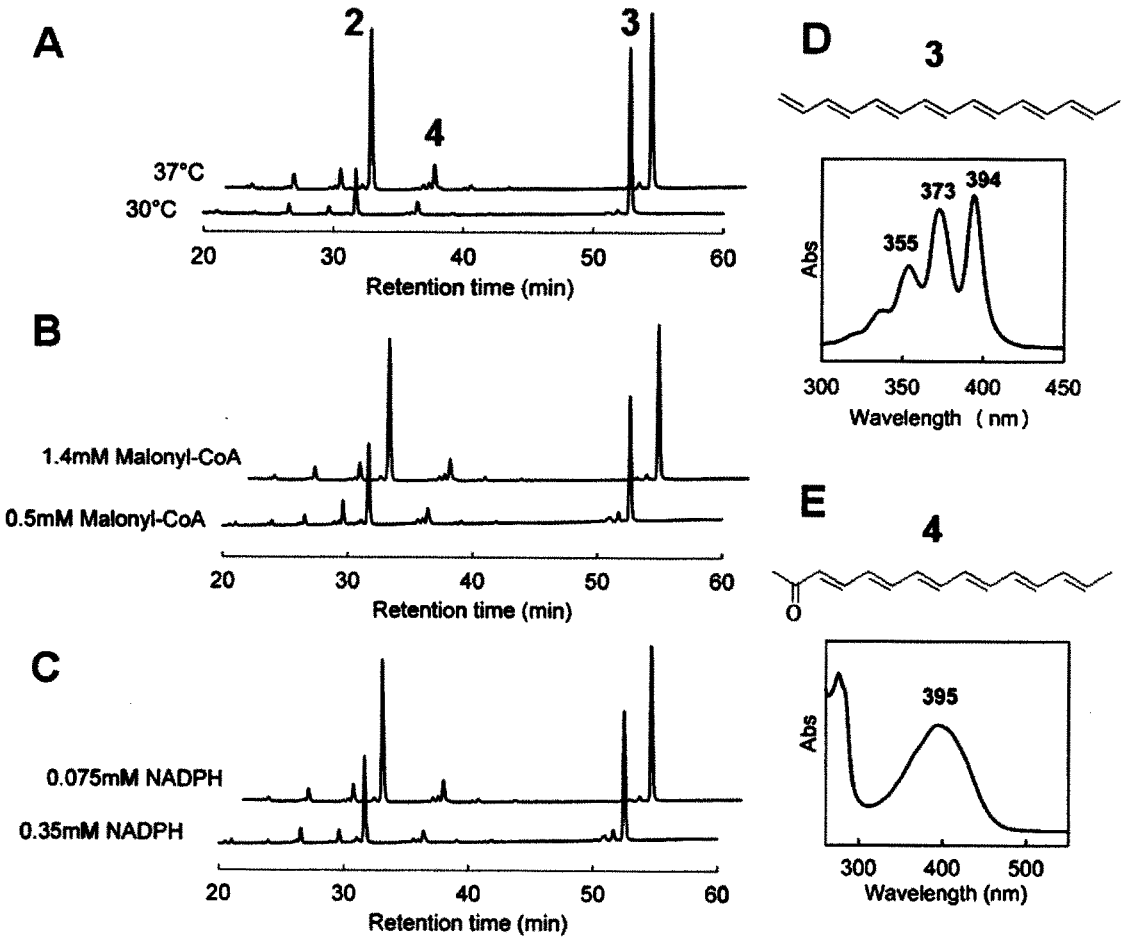


Figure 2.9. (A-C) HPLC analysis of the products of SgcE-SgcE10 reaction with the wavelength set at 400 nm. HPLC program: 70% buffer A (H_2O, 0.045% TFA) + 30% buffer B (CH_3CN, 0.045% TFA) to 100% buffer B at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 250 mm). (A) The temperature-dependent product formation for SgcE-SgcE10. (B) The malonyl CoA-dependent product formation for SgcE-SgcE10. (C) The NADPH-dependent product formation for SgcE-SgcE10. (D) The structure [44,96] and UV-Vis spectrum of **3**. (E) The structure [97] and UV-Vis spectrum of **4**.

The condition-dependent reactions show that SgcE and SgE10 produce higher level of product **2** at higher temperature, higher malonyl CoA concentration and lower NADPH concentration (**Figure 2.9A-C**). It is plausible that when SgcE reaction takes place at a high rate under these conditions, there is less chance for SgcE10 to remove the products from SgeE. In addition, only product **3** was observed in the SgcE-SgcE10 reaction at pH 7.0. We speculate that hydrolysis by SgcE10 and self-hydrolysis are competing processes. When the production of **2** by alkaline hydrolysis is suppressed at lower pH, production of 3 by SgcE10 hydrolysis will dominate.

2.3.2.4 Enzymatic assays with SgcE, SgcE10 and SgcE3

To further explore the biosynthetic mechanism, the putative acetylenase SgcE3 expressed from *E. coli* was incubated together with SgcE and SgcE10. The progression of the enzymatic reaction was observed by absorption spectroscopy at two minutes interval in a two-hour reaction. As soon as the reaction started, absorbance in the 400-450 nm range began to increase and the NADPH absorbance at 342 nm to decrease. The absorption spectra of the SgcE-SgcE10-SgcE3 reaction showed clearly the presence of an isosbestic point (**Figure 2.10A**), suggesting a direct relationship between NADPH consumption and the product formation.

In contrast to the absorption spectrum of the products of the SgcE-SgcE10 reaction that resembles that of the product **3** (**Figure 2.10A**), the absorption spectrum of the products of the SgcE-SgcE10-SgcE3 reaction resembles that of **2** (**Figure 2.10B**). Inspection of the spectra reveals a shift of absorbance of the product of the

SgcE-SgcE10-SgcE3 reaction to longer wavelengths including the shift from 389 to 392 nm and the shift from 409 nm to 417 nm. Although there was no novel product detected from SgcE-SgcE10-SgcE3 reaction after HPLC analysis at various UV detection wavelengths, the yield of **2** is much higher compared to that from SgcE-SgcE10 reaction under similar conditions. These observations indicate that that SgcE3 may play a regulatory role in controlling the formation of the PKS products. Experimental evidence for the interaction between SgcE and SgcE3 will be further discussed in section 2.3.3.3.

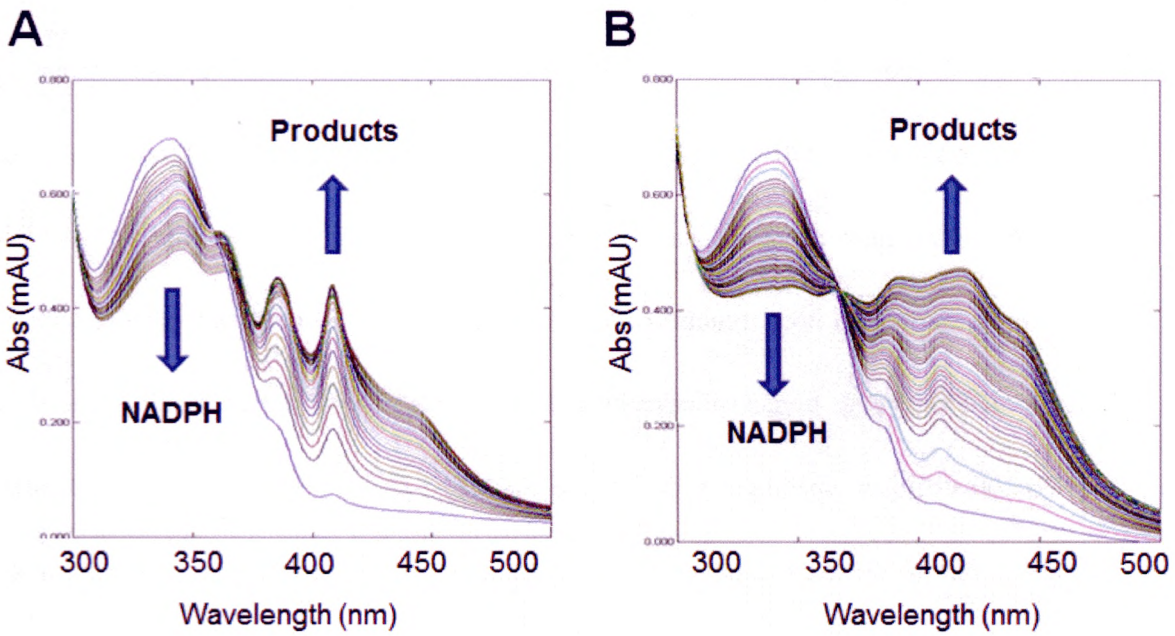


Figure 2.10. (A) UV-Vis spectra of the SgcE-SgcE10 reaction taken at two minutes interval for a one-hour reaction. (B) UV-Vis spectra of the SgcE-SgcE10-SgcE3 reaction taken at two minutes interval for a two-hour reaction.

2.3.3 Characterization of SgcE products by *in vivo* protein expression

2.3.3.1 Co-expression of SgcE and SgcE10

The co-expression of SgcE/E10 in *E. coli* yielded bright yellow cells. Both SgcE and SgcE10 contain $(His)_6$-tag to allow the purification of the proteins by Ni^{2+}-NTA affinity chromatography. The elution fractions containing SgcE and SgcE10 showed an intense yellow coloration. Separation of SgcE and SgcE10 by gel filtration revealed that the co-expression strain produced SgcE10 about four times more than SgcE.

The SgcE protein obtained after gel filtration and concentration was slightly yellowish in color, whereas the SgcE10 purified was bright yellow. The absorption spectra showed that co-expressed SgcE lost the three absorption bands at λ_{max} 371, 391, and 413 nm compared to the single-expressed SgcE. Meanwhile, SgcE10 from SgcE/E10 co-expression exhibited three absorption bands at λ_{max} 348, 366, and 388 nm compared to single-expressed SgcE10 (**Figure 2.11A, B**). This indicates that the PKS product of SgcE has been effectively released and transferred onto SgcE10. The yellow pigment can be released from SgcE10 when the denatured protein was subjected to organic solvent extraction. Hence, investigation of the products non-covalently bound by SgcE10 provides an alternative way to analyze the products of SgcE.

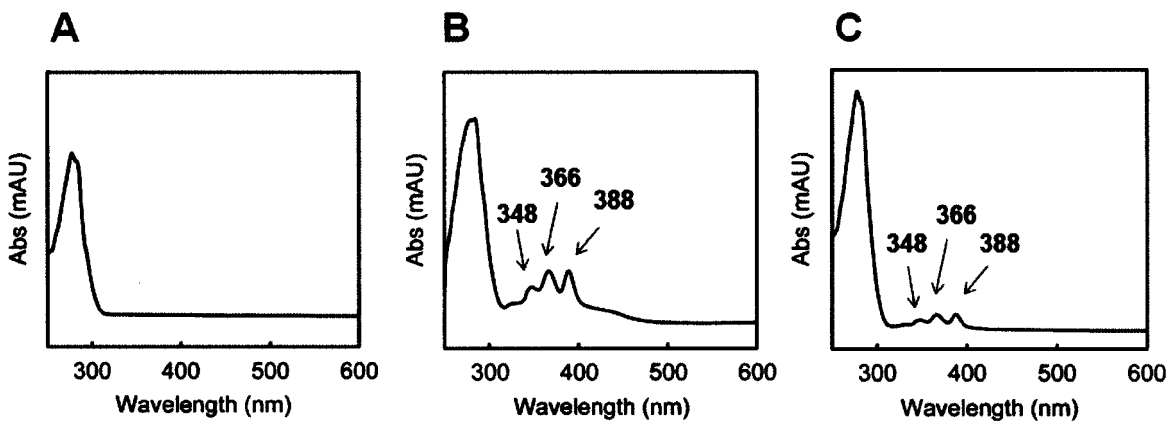


Figure 2.11. UV-Vis spectra of SgcE10. (A) Single-expressed SgcE10 in *E. coli.* (B) SgcE10 from SgcE/E10 co-expression in *E. coli.* (C) SgcE10 from SgcE/E10 co-expression in *S. cerevisiae.*

To ensure that the function of SgcE10 is not host-dependent, co-expression of SgcE and SgcE10 was also performed in *S. cerevisiae.* The expression level of both proteins was much lower than that of *E. coli.* The purified SgcE10 exhibited three absorption peaks at λ_{max} 348, 366, and 388 nm (**Figure 2.11C**), which is identical to that for SgcE10 produced from SgcE/E10 co-expression in *E. coli.* This suggests that SgcE10 releases and bind the SgcE-tethered intermediate in a similar way in these two types of heterologous cells.

The yellow pigment on SgcE10 could be readily extracted by organic solvent, indicating that the product is non-covalently bound by SgcE10. HPLC analysis of the pigment extracted from the colored SgcE10 revealed two major components, which were confirmed to be **1** and **3** by comparison with the HPLC chromatogram for *in vitro* reaction products (**Figure 2.12**). **1** was the hydrolyzed product from single-expressed SgcE and **3** was the dominant product in the SgcE-SgcE10 reaction. Binding of **3** by

SgcE10 implies that SgcE10 can efficiently release the SgcE-tethered **3** in the cell environment. Meanwhile, the observation of **1** suggested that the thioester bond may undergo auto-hydrolysis in the cell, presumably facilitated by the binding of the PKS-tethered pigment by SgcE10.

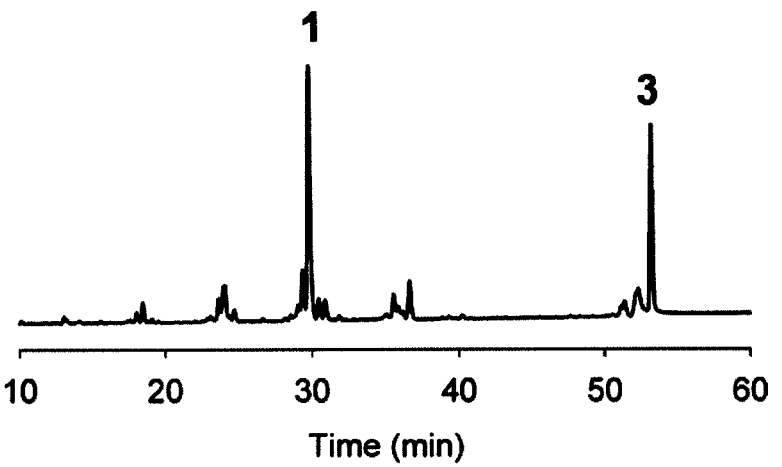


Figure 2.12. HPLC analysis of the products extracted from SgcE10 in SgcE/E10 co-expression at 400 nm. HPLC program: 70% buffer A (H_2O, 0.045% TFA) + 30% buffer B (CH_3CN, 0.045% TFA) to 100% buffer B at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 250 mm).

In addition, **3** was also found in the bright yellow-colored cell pellet, which was first extracted and identified by Shen and co-workers [44]. This means **3** was released after being processed by SgcE10. However, product **2** was not detected in the cell pellet either in *E. coli* or *S. cerevisiae* system. There are two possible explanations for this observation. First, in the *in vivo* co-expression, SgcE10 was cloned into a high copy plasmid and was expressed at high level, which makes the probability for SgcE10 to remove SgcE-tethered product much higher. In a similar case, Kim and coworkers discovered that the pikromycin TEII (pikAV) exhibited no apparent ACP

specificity and the high level expression of pikAV in *Streptomyces venezuelae* led to a significant decrease in aglycone production [99]. Second, the *in vivo* environment is different from *in vitro* buffer solution in pH, temperature and substrate concentration, which may account for the different product distribution.

2.3.3.2 Co-expression of SgcE and SgcE3

To probe the function of SgcE3, SgcE3 was co-expressed with SgcE in *E. coli.* Both protein were expressed as $(His)_6$-tagged proteins to allow the purification of SgcE and SgcE3 by Ni^{2+} affinity chromatography. SgcE and SgcE3 were separated by gel filtration after the elution from the Ni^{2+}-NTA column. When the fractions containing SgcE were examined by SDS-PAGE, we found that SgcE and SgcE3 were eluted out together in fractions A11-B7 (**Figure 2.13B**). This is very surprising and suggests that SgcE3 can directly interact with SgcE to form stable protein complex.

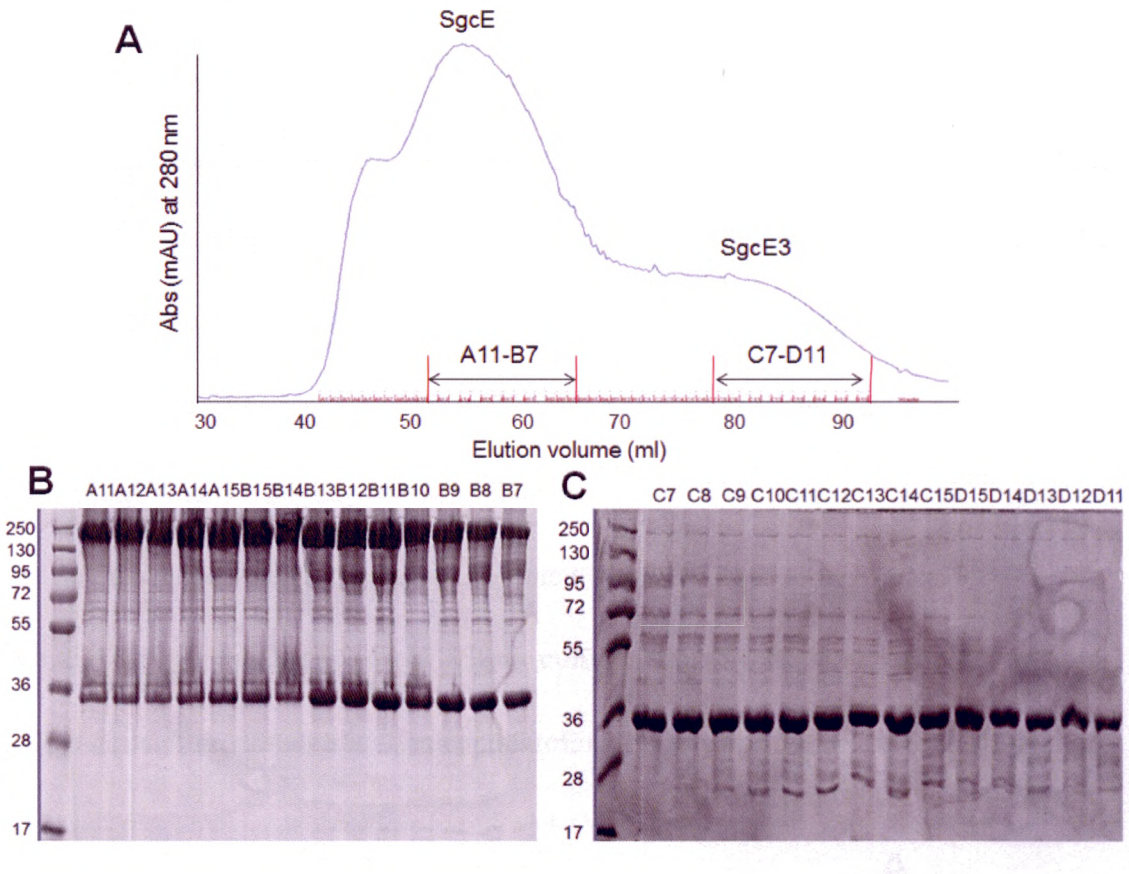


Figure 2.13. (A) FPLC chromatogram of the proteins eluted from Ni^{2+}-NTA column from SgcE/SgcE3 co-expression using a HiLoad™ 16/60 Superdex™ 200 column. (B) SDS-PAGE for fractions collected from A11-B7 containing SgcE. (C) SDS-PAGE for fractions collected from C7-D11 containing SgcE3.

The fractions C7-D11 containing SgeE3 alone (**Figure 2.13C**) were collected and concentrated to characterize SgcE3. The concentrated SgcE3 looked slightly yellow and the yellow pigment on SgcE3 could be readily extracted by organic solvent, indicating that the compounds are non-covalently bound by SgcE3. HPLC analysis of the pigment extracted from the colored SgcE3 showed that it consisted of two major components that were identified to be **1** and **2** (**Figure 2.14A**). As previously discussed, compound **1** is the hydrolytic product from single-expressed SgcE and compound **2** is the major product of SgcE reaction. Binding of **1** by SgcE3 implies that SgcE3

contains a substrate binding pocket for it and presumably facilitate the hydrolysis of SgcE-tethered products. Meanwhile, the observation of **2** suggests that SgcE3 can somehow facilitate the production of the novel product **2**. In addition, SgcE3 from SgcE/E3 co-expression exhibits greater stability than SgcE3 expressed alone, presumably due to the stabilizing effect of the **1** and **3** bound by SgcE3.

To our surprise, **2** was also detected in the insoluble debris of cells co-expressing SgcE/E3, which was not found in the cells expressing SgcE alone (**Figure 2.14B, C**). However, we previously observed that SgcE produced **2** without any other accessory proteins in the *in vitro* reaction. Thus, in the *in vivo* conditions, SgcE3 is somehow needed to interact with the SgcE intermediate and facilitate the production of **2**.

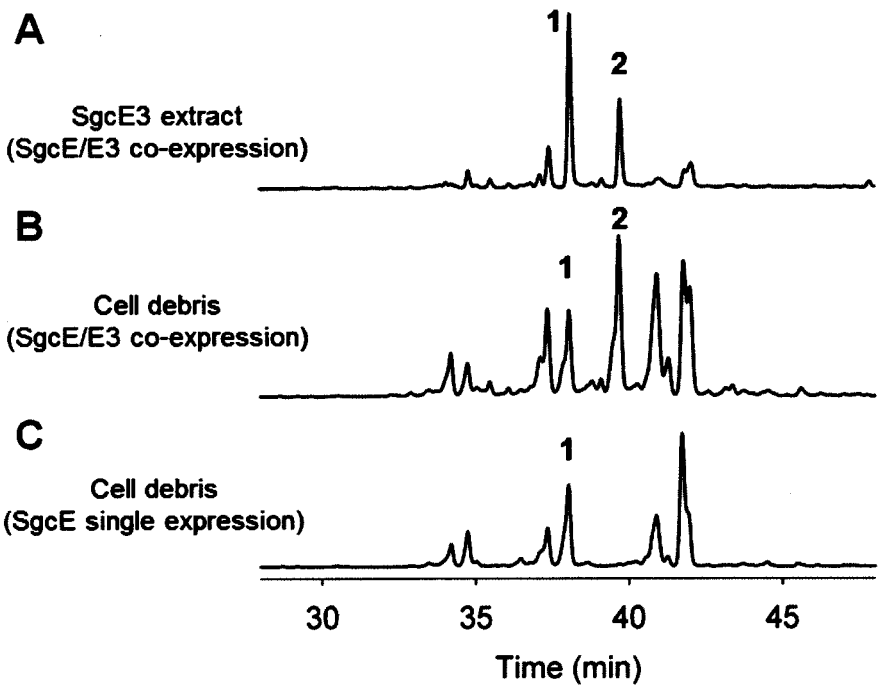


Figure 2.14. (A-C) HPLC analysis of the products extracted from: (A) SgcE3 in SgcE/E3 co-expression; (B) The debris of cells co-expressing SgcE/E3; (C) The debris of cells expressing SgcE alone. HPLC program: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm).

2.4 Discussion

My research work succeeded in expressing and purifying SgcE, SgcE10 and SgcE3 proteins for enzymatic assay and functional studies. In addition to the products identified from the *in vitro* enzymatic assays, we isolated a novel SgcE-tethered product after hydrolyzing SgcE by alkali and the product seems to be different from all the products we identified in *in vitro* assay. Although the identity of this compound remains to be revealed, the alkaline hydrolysis method will be very useful for study of the PKS-tethered intermediates and the biosynthetic mechanism.

From the enzymatic assays, it is surprising that SgcE does not necessarily use acetyl-CoA as a starter unit. Instead, SgcE seems to be able to generate the starter acetate group by decarboxylating malonyl-CoA. We speculate that the KS domain is responsible for the decarboxylation of the the malonate to form acetate as the starter unit for polyketide synthesis. Notably, bacterial type II PKS can also use malonate as a starter unit with the KS_{β} domain catalyzing the decarboxylation prior to chain extension [100]

In the presence of SgcE10, SgcE produces conjugated polyene (**3**) as the major product under all assay conditions. Meanwhile, a high level of **2** and a minor product carbonyl-conjugated polyene (**4**) were produced at elevated pH, temperature and malonyl-CoA concentration. The possible biosynthetic mechanism for the production of **3**, **4** and **2** are shown in **Figure 2.15**. The fifteen carbon-containing **3** and **4** are likely to derive from their corresponding octaketide intermediates, with the hydrolytic release of the products catalyzed by the thioesterase [44,97,101]. The minor

production of **4** is due to inefficient keto-reduction by SgcE in the final round of iterative chain extension [97]. The mechanism has also been proposed for the formation of **4** through thioester hydrolysis and facile decarboxylation of the transient β-keto acid intermediate [98].

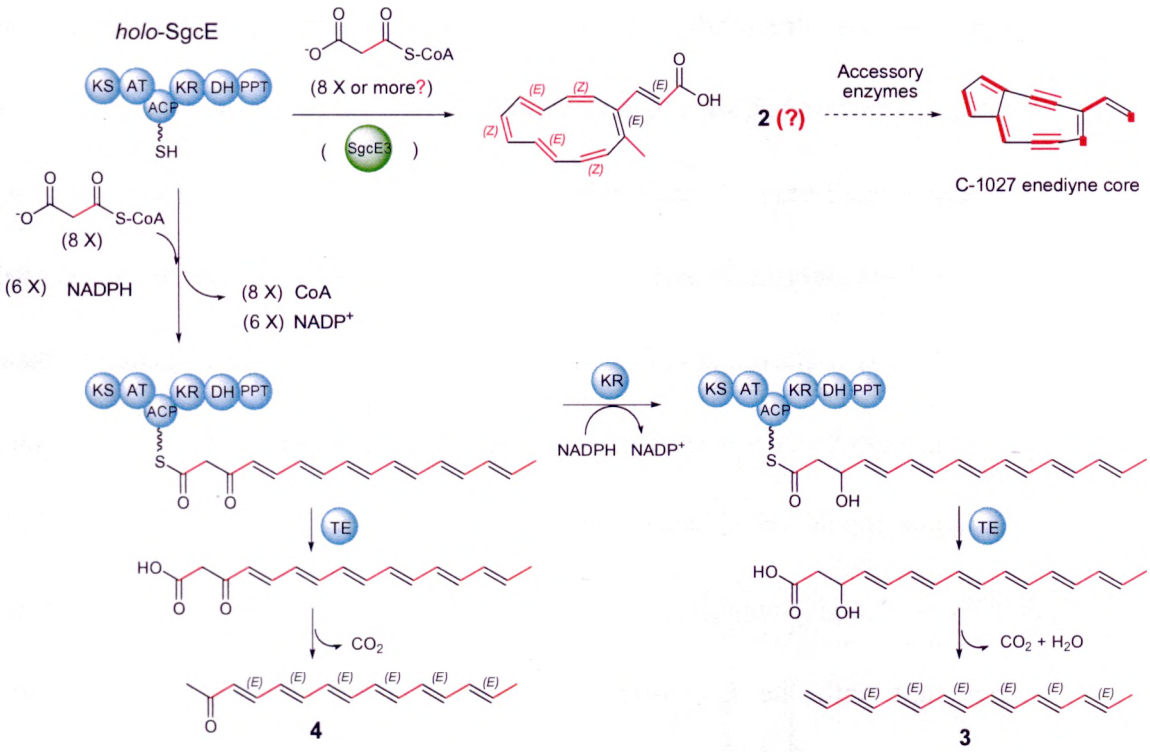


Figure 2.15. The proposed mechanism for the production of **2**, **3** and **4**. The structure of **2** was based on the LC-MS results and need to be further confirmed by NMR spectroscopy. Atoms derived from intact acetate are highlighted in red. AT, acyl transferase; KS, ketosynthase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; PPT, phosphopantetheinyl transferase; TE, thioesterase.

At this moment, it is not known which product is the real biosynthetic intermediate *en route* to the enediyne core. It is possible that neither **3** nor **4** is the real biosynthetic intermediate as recently suggested by Townsend and co-workers from the study of the homologous CalE8/E7 system. Interestingly, the *in vitro* reaction of SgcE

without SgcE10 was carried out and a novel product **2** with unknown structure was identified for the first time. Although the structure of **2** remains to be established, **2** is speculated to be a highly conjugated cyclic aliphatic molecule based on the high resolution MS result. A possible structure of **2** is shown in figure 2.15 with the absolute configuration of the double bonds assigned arbitrarily. If the cyclic structure of **2** is confirmed, it is more reasonable to believe that the cyclic molecule is likely to be the real biosynthetic precursor for the bicyclic enediyne core (**Figure 2.15**). This would also raise the puzzling question of how the PKS catalyzes the cyclization of the linear product. It is noteworthy that the double bonds of **3** or **4** are not derived from intact acetate, while five double bonds of the C-1027 enediyne core are derived from intact acetates. Hence, it is hypothetical that the cyclization of any polyene intermediate into **2** requires the rearrangement of the conjugated double bonds to match the acetate incorporation patterns of the enediyne core. We will focus on the determination of the structure of **2** in the next stage.

SgcE10 may function as type II thioesterase considering that Type II TEs (TE-IIs) are commonly found as a discrete thioesterase within polyketide biosynthetic cluster [99,102]. The fact that SgcE can produce the released product **2** in the absence of SgcE10 indicates that SgcE10 is not responsible for polyketide termination and cyclization. It has been proposed that TE-II may serve as a "housekeeping" enzyme and removes aberrant acyl units from carrier domains. Two models have been proposed for the TEII housekeeping function [103]. In the high specificity model, the TE-II scans the PKS-tethered intermediates and efficiently removes only aberrant

compounds. In the low specificity model, the TE-II removes both correct and incorrect PKS-tethered intermediates at an inefficient rate. SgcE10 is most likely a TE-II with low substrate specificity. It can remove the correct SgcE-tethered intermediate and consequently produce the by-product **3** with its decarboxylation activity. When the SgcE-SgcE10 reaction was carried out at low rate, it provides more opportunity for removal by SgcE10, leading to a dominant production of **3.**

We discovered that the functionally unknown SgcE3 can not only influence the product formation of SgcE but also bind the products. SgcE-SgcE10-SgcE3 reaction produces more **2** than SgcE-SgcE10 reaction, indicating SgcE3 can somehow facilitate the production of the novel product **2**. Moreover, the size-exclusion chromatography seems to suggest that SgcE3 protein can physically interact with SgcE, further supporting a regulatory role of SgcE3 in controlling PKS product formation. The absorption between 300 to 450 nm indicated that SgcE3 could be a di-iron protein that is potentially capable of catalyzing triple bond formation.

In the future work, we would like to determine the structure of the novel product **2** and continue the exploration of the function of SgcE3 as a potential acetylenase that also plays a regulatory role. We will also include other potentially key ancillary enzymes in our enzymatic assay in a hope to reconstitute the biosynthesis of bicyclic enediyne structure.

CHAPTER 3

Comparative study of the iPKSs in 9- and 10-membered Enediyne Biosynthesis

3.1 Introduction

We discussed the biosynthetic mechanism of the 9-membered enediyne C-1027 in the previous chapter. In this chapter, the iPKS from two 10-membered enediynes, calicheamicin and dynemicin will be examined to investigate whether the biosynthetic pathways of the 9- and 10-membered enediynes diverge at the PKS stage. The *in vitro* and *in vivo* products of the PKSs CalE8 and DynE8 from the calicheamicin and dynemicin biosynthetic pathways will be characterized.

As introduced previously, C-1027 is a representative 9-membered enediyne natural product isolated from a culture filtrate of *Streptomyces globisporus* C-1027 [67,78-81]; whereas calicheamicin γ_1^I from *Micromonospora echinospora ssp. calichensis* [65] and dynemicin from *Micromonospora chersina* [66] are the two most well-known 10-membered enediyne natural products. ^{13}C-isotope feeding experiments revealed different isotope incorporation patterns between 9- and 10-membered enediyne cores, suggesting the pathways may involve different precursors or folding patterns (**Figure 3.1A**) [15,84-86]. More intriguingly, the acetate units were

incorporated into different positions in the enediyne moiety for dynemicin and calicheamicin [85,86] **(Figure 3.1B)**.

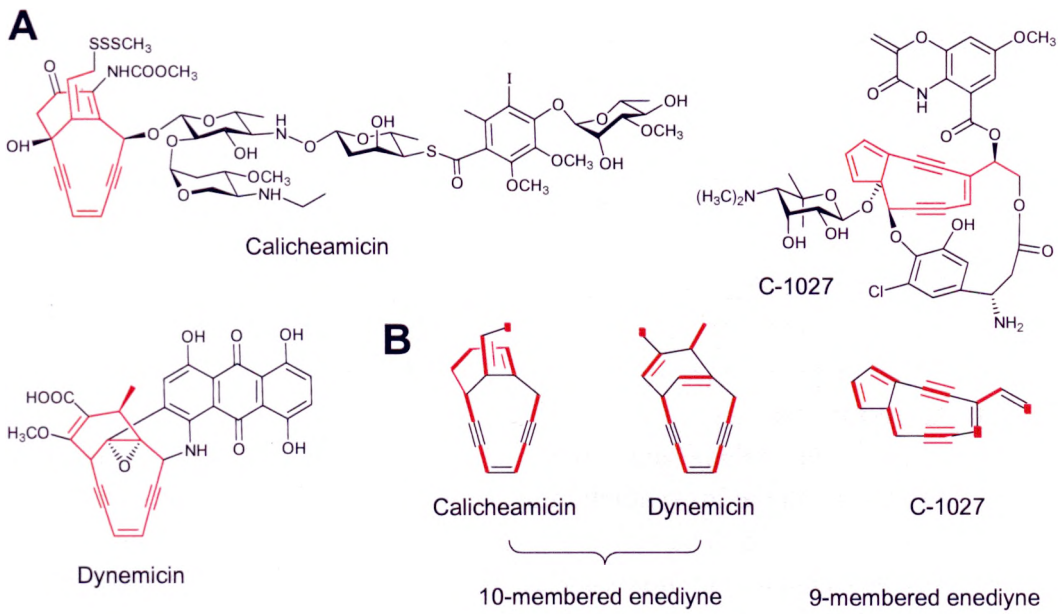


Figure 3.1. Three 9- and 10-membered enediyne natural products. (A) Structures of three enediyne natural products with the 9- or 10-membered enediyne cores highlighted. (B) Acetate incorporation patterns for the enediyne cores with the atoms and bonds derived from intact acetate highlighted in red.

The recent cloning, sequencing, and characterization of the gene clusters responsible for the biosynthesis of three 9-membered enediynes (C-1027 [41], neocarzinostatin [35] and maduropeptin [38]) and two 10-membered enediynes (calicheamicin [42] and dynemicin [43]) suggested that the biosynthesis of the highly reactive enediyne core structure (or "warhead") is initiated by a unique enediyne iPKS (PKSE) [41,42]. Further comparison of the five biosynthetic clusters revealed the minimal warhead cassette of five genes encoding a PKSE, a thioesterase (TE) and three proteins of unknown function highly conserved among the 9- and 10-membered

enediyne families (**Figure 2.2A**) [87].

Although the high sequence homology and identical domain organization among PKSEs suggest that all the enediyne pathways may share a common intermediate that diverge at a late stage to furnish 9- or 10- membered enediyne core, phylogenetic analysis of PKSEs (**Figure 2.2B**) reveals that they are grouped into two distinct genotypic clades according to the enediyne core structure (9- versus 10-membered). Given the increasing understanding of the programmed synthesis of polyketide by iterative PKSs, the bioinformatic analysis raises the intriguing possibility that the 9- and 10-membered enediyne PKSs may be programmed to generate structurally different biosynthetic intermediates.

In 2010, Shen and co-workers reported that co-expression of all five cognate PKSE-TE pairs in *E. coli* commonly produced a conjugated polyene 1,3,5,7,9,11,13-pentadecaheptaene (**3**) extracted from the cell pellets [104]. Meanwhile, studies on SgcE and SgcE10 in our lab have shown the production of other products in addition to **3** by SgcE (see Chapter 2). In this chapter, we performed comparative *in vitro* enzymatic assays with three PKSEs from the C-1027, calicheamicin and dynemicin pathways to gain insight into the function of the PKSEs.

3.2 Materials and Methods

3.2.1 Strains, plasmids, and chemicals

BL21 (DE3) competent cells were from Novagen. Expressing vectors CalE8-pET28, CalU15-pCDF2, DynE8-pET28 and DynE7-pCDF2 were directly obtained from lab colleagues Liew Chongwai, Kong Rong and Ji Qiang. Acetyl-CoA, malonyl-CoA, NADPH and other chemicals were purchased from Sigma-Aldrich and stored at -20 °C.

3.2.2 Protein expression and purification

Expression and purification of CalE8 and DynE8 were carried out following the similar procedures described for SgcE in chapter 2. Expression and purification of CalE7 and DynE7 were carried out following the similar procedures described for SgcE10 in chapter 2. Co-expression of DynE8 and CalE7 was carried out following the similar procedures described for the co-expression of SgcE and SgcE10 described in chapter 2. Co-expression of CalE8 and CalU15 was carried out following the similar procedures described for the co-expression of SgcE and SgcE3 in chapter 2.

3.2.3 *In vitro* enzymatic assays

For enzymatic reaction with PKS, a typical reaction contained 0.25 mM NADPH, 1.0 mM malonyl CoA, 0.15 mM acetyl CoA and 6.65 μM PKS in the reaction buffer [50 mM Tris (pH 8.5), 150 mM NaCl and 1 mM DTT] in a total volume of 400 μl. The reaction mixture was incubated at 30°C for 2 hours.

For the enzymatic reaction with PKS and TE, a typical reaction contained 0.075 mM NADPH, 1.4 mM malonyl CoA, 0.15 mM acetyl CoA, 3.1 μM PKS and 62.9 μM TE in the reaction buffer [50 mM Tris (pH 8.5), 150 mM NaCl and 1 mM DTT] in a total volume of 200 μl. The reaction mixture was incubated at 37°C for 2 hours. To test pH and temperature effects on the enzyme activities, the reactions were carried out with the same components at pH 7.0 and 23°C.

Experimental conditions for the enzymatic assays in Figure 3.5. (A) SgcE-SgcE10 reaction. Assay conditions: 3.1 μM SgcE, 62.9 μM SgcE10, 0.25 mM NADPH, 0.25 mM malonyl-CoA, 50 mM Tris (pH 7.0), 23°C. (B) SgcE-SgcE10 reaction. Assay conditions: 3.1 μM SgcE, 62.9 μM SgcE10, 0.075 mM NADPH, 1.4 mM malonyl-CoA, 50 mM Tris (pH 8.5), 23°C. (C) CalE8-CalE7 reaction. Assay conditions: 0.33 μM CalE8, 8.7 μM CalE7, 0.25 mM NADPH, 0.25 mM malonyl-CoA, 50 mM Tris (pH 7.0), 23°C. (D) CalE8-CalE7 reaction. Assay conditions: 0.33 μM CalE8, 8.7 μM CalE7, 0.25 mM NADPH, 2.4 mM malonyl-CoA, 50 mM Tris (pH 8.5) 37°C. (E) DynE8-DynE7 reaction. Assay conditions: 3.2 μM DynE8, 66.8 μM DynE7, 0.35 mM NADPH, 0.25 mM malonyl-CoA, 50 mM Tris (pH 7.0), 23°C. (F) DynE8-DynE7 reaction. Assay conditions: 3.2 μM DynE8, 66.8 μM DynE7, 0.15 mM NADPH, 1.4 mM malonyl-CoA, 50 mM Tris (pH 8.5), 37°C.

3.2.4 Product extraction and HPLC analysis

An equal volume of ethyl acetate was added into the protein solution or reaction mixture and vigorous vortexing ensued. Subsequently the mixture was centrifuged at

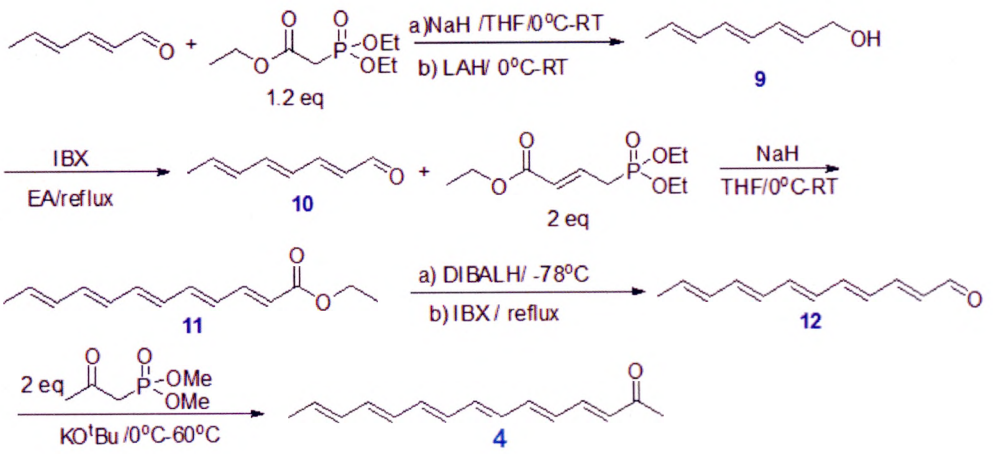
14,500 rpm for 15 minutes. The resulting organic extract was evaporated by using a Speed-Vac. The dried sample was re-dissolved in 20 μl methanol for HPLC analysis.

For PKS reactions and PKS/U15 co-expression, the HPLC analysis was performed with an analytical eclipse XDB C18 column (4.6 × 150 mm) using an Agilent 1200 HPLC. A full gradient was employed from 100% Buffer A (HPLC grade water with 0.045% TFA) to 100% Buffer B (100% acetonitrile with 0.045% TFA) in 60 minutes.

For PKS hydrolysis, PKS-TE reactions and PKS/TE co-expression, the HPLC analysis was performed with an analytical eclipse XDB C18 column (4.6 × 250 mm) using an Agilent 1200 HPLC. A full gradient was employed from 70% buffer A (HPLC grade water with 0.045% TFA) + 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B in 60 minutes.

3.2.5 Synthesis of the standard for product 4

The all *trans* (3*E*,5*E*,7*E*,9*E*,11*E*,13*E*)-pentadeca-3,5,7,9,11,13-hexaen-2-one was synthesized as a standard for product **4**. The synthetic experiments were done by our collaborator Dr. Zhong Guofu's lab from the Division of Chemistry and Biological Chemistry in the School of Physical and Mathematical Sciences in Nanyang Technological University. The chemical synthesis was designed based on the procedures reproted for similar compounds in the literature [105-108].



(2E,4E,6E)-octa-2,4,6-trien-1-ol (9) [107]. To a suspension of NaH (60% in mineral oil, 400 mg, 10 mmol) in THF (50 ml) at 0 °C was slowly added the phosphonate (2.24 g, 10 mmol). After the addition, the suspension turned clear. The aldehyde (808 mg, 8 mmol) in THF (15 ml) was added. The reaction mixture was stirred at 0 °C for 2 h before being quenched by saturated NH_4Cl (50 ml). The product was extracted with ethyl acetate (3 x 70 ml). The combined extracts were washed with brine, dried and concentrated. The white solid (1.32 g) was pure enough and used in the next step. To a suspension of $LiAlH_4$ (608 mg ,16 mmol) in anhydrous THF (100 ml) at 0 °C, the ester (1.32 g , 8 mmol) in 50 ml of THF was added dropwise. After the addition, the mixture was allowed to stirred at room temperature for 2 hours, monitored with TLC, then quenched by 20 ml of 1N NaOH, The mixture was stirred for 1 h at room temperature, then the white precipitate was removed by filtration through a celite pad and the pad was rinsed with 3 x 30 ml of ethyl acetate (EA). The combined organic filtrates were washed with 100 ml of brine, and the aqueous layer was extracted with diethyl ether (3 × 30 ml). The combined organic phases were dried

with anhydrous sodium sulfate and concentrated, the residue was purified by silica gel column chromatography to afford 910 mg white solid, yield 91%. 1H NMR ($CDCl_3$, 400 MHz). δ1.35 (s, 1H), 1.77 (d, J = 6.8 Hz, 3H), 4.18 (d, J = 6.8 Hz, 2H), 5.69-5.83 (m, 2H), 6.05-6.28 (m, 4H) 13C NMR ($CDCl_3$, 100 MHz) δ 18.3, 63.5, 129.3, 130.5, 131.0, 131.5, 131.9, 133.6.

(2E,4E,6E)-octa-2,4,6-trienal (10) [106]. To a solution of alcohol 9 (500 mg, 4.0 mmol) in 50 ml of EA was added IBX (3.42 g, 12 mmol) in one portion, the mixture was refluxed for 3 hours and then cooled to room temperature, then the white precipitate was removed by filtration through a celite pad and the pad was rinsed with 3 x 20 ml of EA, The combined organic filtrates were concentrated, the residue was purified by silica gel column chromatography to afford 418 mg solid, yield 86%. 1H NMR ($CDCl_3$, 400 MHz). δ1.86 (d, J = 6.8 Hz, 3H), 6.01-6.17 (m, 3H), 6.30-6.36 (m, 1H), 6.60-6.64 (m, 1H), 7.08-7.26 (m, 1H) , 9.54 (d, J = 8 Hz, 1H). 13C NMR ($CDCl_3$, 100 MHz) δ18.7, 127.6, 130.7, 131.2, 137.2, 143.1, 152.5, 193.6.

(2E,4E,6E,8E,10E)-ethyl dodeca-2,4,6,8,10-pentaenoate (11) [108]. To a suspension of NaH (60% in mineral oil, 160 mg, 4 mmol) in THF (30 ml) at 0 °C was slowly added the phosphonate (1.0 g, 4.0 mmol). After the addition, the suspension turned clear. The aldehyde **10** (244 mg, 2.0 mmol) in THF (15 ml) was added dropwise. The reaction mixture was stirred at 0 °C for 2 hours before being quenched by saturated NH_4Cl (50 ml). The product was extracted with ethyl acetate (3 x 70 ml). The combined extracts were washed with brine, dried and concentrated. the residue was purified by silica gel column chromatography to afford 418 mg yellowish solid

313 mg, yield 72% . ^{1}H NMR ($CDCl_3$, 400 MHz). δ1.29 (t, J = 7.2 Hz, 3H), 1.83 (dd, J_1 = 6.8 Hz, J_2 = 15.2 Hz, 3H), 4.20 (q, J = 7.2 Hz, 2H), 5.78-5.86 (m, 1H), 6.09-6.44 (m, 7H) , 6.55-6.59 (m, 1H) 6.60-6.63 (m, 1H), 7.26-7.34 (m, 1H). ^{13}C NMR ($CDCl_3$, 100 MHz) δ14.3, 18.5, 60.2, 120.3, 129.4, 130.0, 131.0, 131.8, 132.1, 135.9, 137.5, 140.9, 144.5, 167.2.

(2*E*,4*E*,6*E*,8*E*,10*E*)-dodeca-2,4,6,8,10-pentaenal (12) [105]. To a solution of the ester **11** (218 mg, 1 mmol) in hexanes (10 ml) at -78 °C was added Dibal-H solution (1.0 M in hexanes, 2 ml). The reaction mixture was stirred for 30 minutes at -78 °C before being quenched with 5 ml of saturated Rochelle's salt solution. The resultant mixture was stirred at room temperature for 1.5 hours before diethyl ether (20 ml) was added. The organic layer was separated, dried and concentrated and used directly in the next step. To a solution of alcohol (80 mg, 0.5 mmol) in 20 ml of EA was added IBX (427 mg, 1.5 mmol) in one portion, the mixture was refluxed for 1 hour and then cooled to room temperature, then the white precipitate was removed by filtration through a celite pad and the pad was rinsed with t3 x 20 ml of EA, The combined organic filtrates were concentrated, the residue was purified by silica gel column chromatography to afford 68 mg yellow solid, yield 88%. ^{1}H NMR ($CDCl_3$, 400 MHz). δ1.82 (d, J = 6.8 Hz, 3H), 5.83-5.88 (m, 1H) ,6.12-6.53 (m, 7H), 6.68-6.74 (m, 1H), 7.11-7.26 (m, 1H) ,9.56 (d, J = 8 Hz, 1H). ^{13}C NMR ($CDCl_3$, 100 MHz) δ18.6, 129.3, 129.8, 130.7, 130.7, 131.7,133.0, 137.0 139.3, 143.0, 152.0, 193.5.

(3*E*,5*E*,7*E*,9*E*,11*E*,13*E*)-pentadeca-3,5,7,9,11,13-hexaen-2-one (4). To a solution of KO^{t}Bu (56 mg, 0.5 mmol) in THF (10 ml) at 0 °C was slowly added the

phosphonate (83 mg, 0.5 mmol). After the addition, the suspension turned clear. The aldehyde **12** (43 mg, 0.25 mmol) in THF (1 ml) was added via syringe. The reaction mixture was stirred at 0 °C for 15 minutes then heated to 60°C, monitored by TLC, the mixture was concentrated and purified by silica gel column chromatography directly, afford 35 mg yellowish solid, yield 66 % . IR 1676 cm^{-1} (C=O), HR-MS (ESI) *m/e* = 215.1432 (MH^+), ^1H NMR ($CDCl_3$, 300 MHz), δ1.82 (d, *J* = 6.8 Hz, 3H), 2.29 (s, 3H), 5.78-5.85 (m, 1H), 6.11-6.72 (m, 10H), 7.15-7.28 (m, 1H). ^{13}C NMR ($CDCl_3$, 75 MHz), δ18.5, 27.4, 129.6, 130.1, 130.2, 131.4, 131.6, 131.8, 131.9, 135.3, 136.3, 137.9, 141.8, 143.4, 198.4.

3.3 Results

3.3.1 Hydrolytic release of the common PKS-tethered product 1

In the previous chapter, the single-expressed SgcE with bright orange color was hydrolyzed by NaOH to release the PKS-tethered product. The released product **1** exhibited absorption maximum at 380 nm. The PKSE CalE8 from the biosynthetic pathway of the 10-membered enediyne calicheamicin was expressed and purified previously in our lab. Since the single-expressed CalE8 showed a bright yellow color, slightly different from the bright orange color of SgcE, I was interested to find out whether CalE8 releases the same or different product when treated with alkaline.

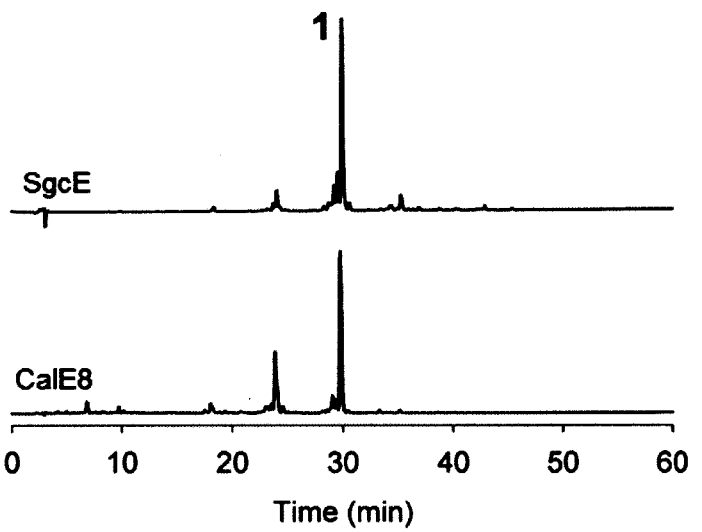


Figure 3.2. HPLC analysis of hydrolyzed product of single-expressed SgcE and CalE8 at 375 nm. HPLC program: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm).

Similar to SgcE, extraction of the denatured CalE8 protein by organic solvent failed to isolate the yellow pigment, indicating that the pigment was covalently attached to CalE8. Then the alkaline hydrolysis method was employed to cleave the

thioester bond to release the product. Surprisingly, HPLC analysis shows that CalE8 and SgcE share the same hydrolyzed compound **1** (**Figure 3.2**). This observation implies that 9- and 10-membered enediyne PKSs generate at least one common product that is covalently attached to the ACP domain in the absence of the thioesterase.

3.3.2 Common products of enediyne PKS and TE

To compare the products of PKSE and TE, the enzymatic reactions of SgcE, CalE8 and DynE8 with their paired TEs were carried out under various conditions. Once again, SgcE-SgcE10, CalE8-CalE7 and DynE8-DynE7 share the dominant product **3** under all conditions (**Figure 3.3**), suggesting that the conjugated polyene **3** is a common product for all three enediyne PKSs.

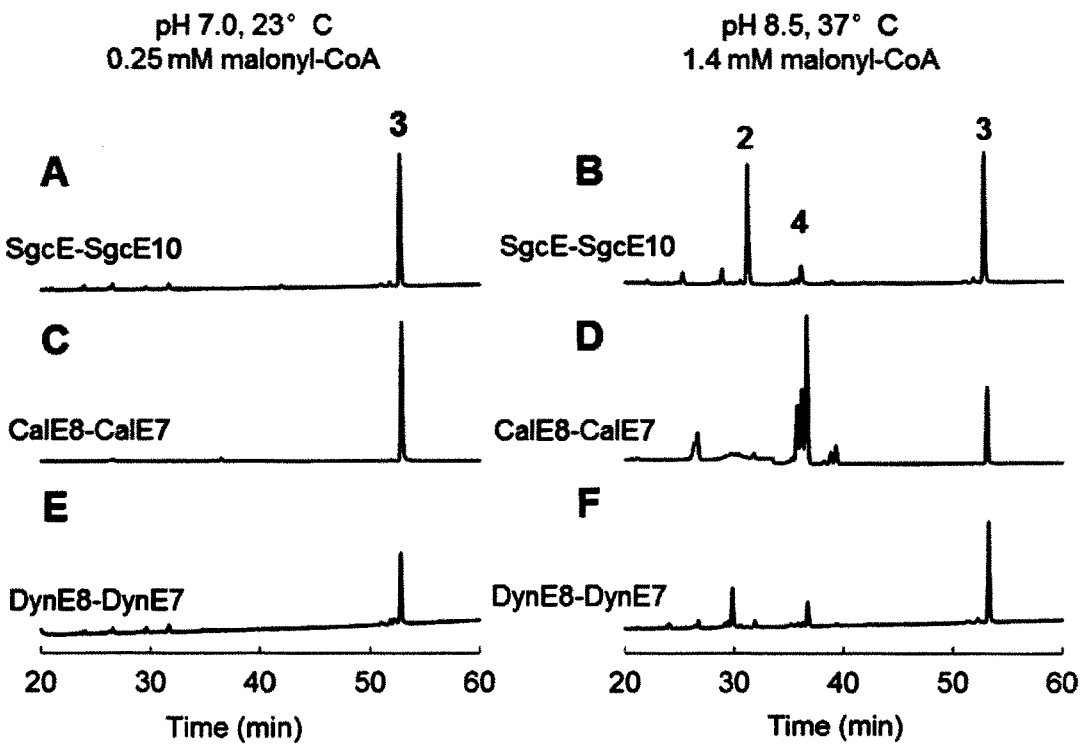


Figure 3.3. (A-F) HPLC analysis of the products of PKS-TE reactions at 400 nm. HPLC program: 70% buffer A (H_2O, 0.045% TFA) + 30% buffer B (CH_3CN, 0.045% TFA) to 100% buffer B at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 250 mm).

SgcE and SgcE10 are the PKS and TE for the enediyne core biosynthesis of the 9-membered enediyne natural product C1027 from *Streptomyces globispus* C-1027 [41]. Shen and co-workers reported the product **3** can be extracted from the insoluble debris of cells co-expressing SgcE and SgcE10 [44]. To examine the *in vitro* products of SgcE and SgcE10, we carried out the enzymatic assays with SgcE and SgcE10 under various conditions. The results showed that the enzymes produced a sets of products, including **2** (MH^+, m/z = 241.12), **3** (MH^+, m/z = 199.15), and **4** (MH^+, m/z = 215.14). Previously, only a small amount of **3** was observed when trifluoric acid (TFA) was used to quench the reaction prior to production extraction. A significant amount of **3** could be observed when the reaction was directly extracted by ethyl acetate or HCl was used for precipitating the protein. The condition-dependent assays reveal that **3** is the single dominant product under the mild assay conditions (pH 7.0, 23 °C, 0.25 mM malonyl-CoA) (**Figure 3.3A**). When the reactions were carried out at higher pH, temperature and malonyl-CoA concentration (**Figure 3.3B**), a comparable amount of **2** was produced together with **3**. Interestingly, only a small amount of carbonyl-conjugated polyene (**4**) was produced by SgcE and SgcE10 under all the reaction conditions.

CalE8 and CalE7 are the PKSE and TE for the 10-membered enediyne natural product calicheamicin biosynthesis in *Micromonospora echinospora ssp. calichensis*

[42]. Previously, our group observed only negligible amount of **3**, when TFA was used to quench the reaction prior to production extraction [97]. Given the instability of **3** in the presence of TFA as observed in the CalE8-CalE7 assays, the yield of **3** must have been underestimated. We carried out the enzymatic assays with CalE8 and CalE7 under various conditions and quenched the reaction with HCl. The condition-dependent assays reveal that **3** is the single dominating product under the mild assay conditions (pH 7.0, 23 °C, 0.25 mM malonyl-CoA) (**Figure 3.3C**). However, **4** was produced as a major product at high catalytic rate with elevated pH temperature and malonyl-CoA concentration (**Figure 3.3D**). A very recent study from Townsend and co-workers also reported the generation of **3** and **4** by CalE8 and CalE7, with the intermediate detected by LC-MS [98]. Production of **4** is due to the "skipping" of the keto-reduction step by CalE8 in the final round of iterative chain extension [97].

DynE8 and DynE7 are the PKS and TE for the 10-membered enediyne natural product dynemicin biosynthesis in *Micromonospora chersina* [43] and have not been characterized previously. We carried out the enzymatic assays with DynE8 and DynE7 under various conditions and quenched the reaction with HCl. The condition-dependent assays reveal that DynE8 and DynE7 also generated **3** as the single dominating product at moderate catalytic rate (pH 7.0, 23 °C, 0.25 mM malonyl-CoA) (**Figure 3.3E**). When the reactions were carried out at higher pH, temperature and malonyl-CoA concentration (**Figure 3.3F**), DynE8 and DynE7 produced **4** as minor product.

Although the structure of **3** and **4** have been initially determined by NMR and MS

[44,97], the geometry of the double bonds was not confirmed due to the significant overlapping of the NMR signals in the ethylenic proton region. We previously speculated that **4** may contain a *cis* double bond based on the kinked substrate-binding channel in the crystal structure of CalE7 [101]. To validate the configuration of the double bonds, we have synthesized all-*trans* (3*E*, 5*E*, 7*E*, 9*E*, 11*E*, 13*E*)-pentadeca-3,5,7,9,11,13-hexaen-2-one as standard. HPLC analysis showed that the standard shares identical retention time and absorption spectrum with **4** (**Figure 3.4**), indicating that all the double bonds of **4** are *E* configuration. The kinked channel in *apo*-CalE7 and the confirmation of the all-*trans* product suggest that CalE7 may undergo significant distortion or conformational change during substrate binding. More recently, Guo et al. confirmed that the six double bonds in the conjugated polyene (**3**) are all in the *trans* configuration as well [96].

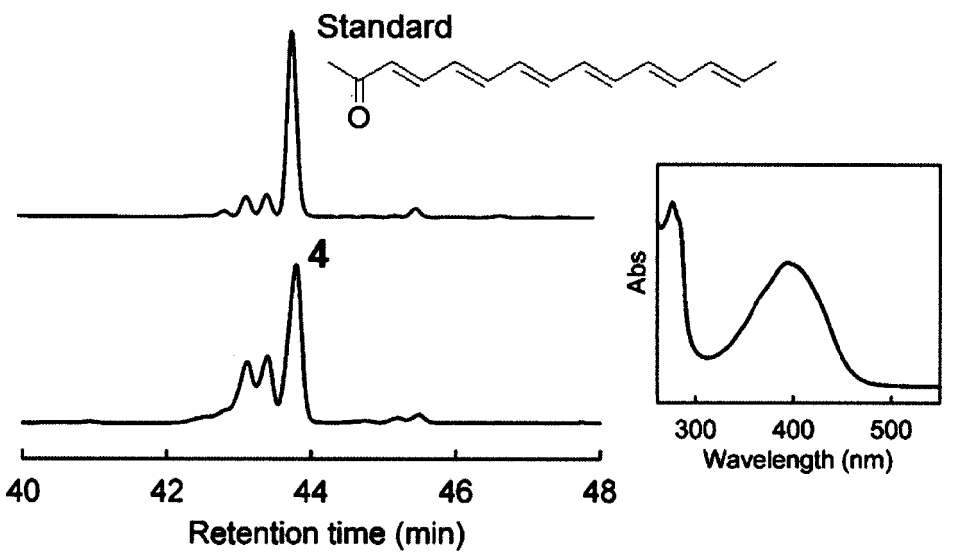


Figure 3.4. Comparison of the retention time of **4** and the all-*trans* standard. HPLC program: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm).

In addition to *in vitro* reactions, co-expression of SgcE/E10 and DynE8/E7 were conducted to further examine whether **3** is commonly produced by enediyne PKS and TE *in vivo*. SgcE10 was purified from SgcE/E10 co-expression and DynE7 from DynE8/E7 co-expression. The yellow pigment on SgcE10 and DynE7 could be extracted by organic solvent, indicating that the compounds are non-covalently bound by the TEs. HPLC analysis of the pigment extracted from the colored SgcE10 and DynE7 shows that they share the two major products, **1** and **3** (**Figure 3.5**). This further supports that same PKS-tethered intermediate is shared among the enediyne PKSs.

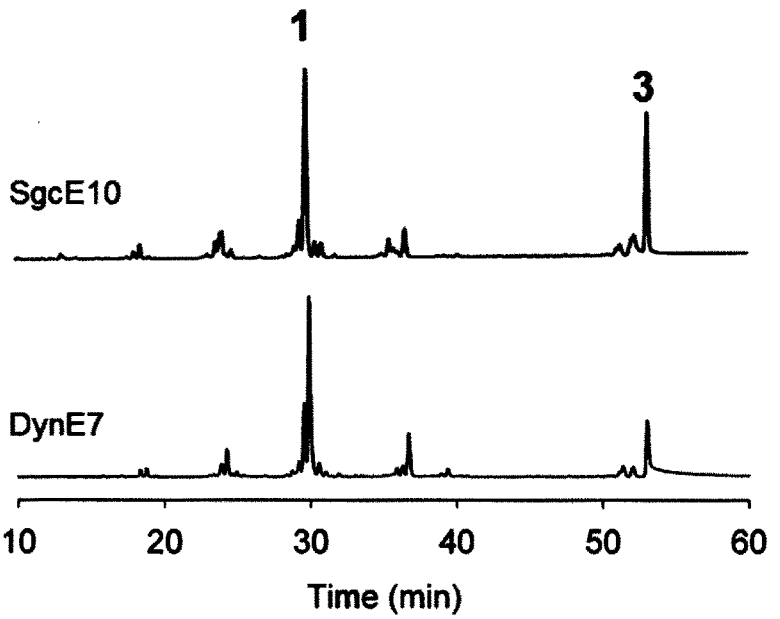


Figure 3.5. HPLC analysis of the products extracted from SgcE10 in SgcE/E10 co-expression and DynE7 in DynE8/E7 co-expression at 400 nm. HPLC program: 70% buffer A (H_2O, 0.045% TFA) + 30% buffer B (CH_3CN, 0.045% TFA) to 100% buffer B at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 250 mm).

3.3.3 Exclusive production of 2 by SgcE

To compare the products of enediyne PKSs, the enzymatic reactions of SgcE, CalE8 and DynE8 in the absence of the TEs were performed under identical conditions. In contrast to SgcE, CalE8 and DynE8 reactions did not consume NADPH in the absence of the TEs according to the UV-Vis spectra. HPLC analysis shows that CalE8 and DynE8 did not produce any detectable products (**Figure 3.6**). This observation implies that CalE8 and Dyne8 are not able to release the PKS-tethered intermediate and consequently cannot load the new substrates onto the thiol group of ACP domain. We speculated that while 9-membered enediyne PKSs are able to off-load and probably cyclize the PKS-tethered intermediate to form **2**, 10-membered enediyne PKSs may need the assistance of other accessory proteins to produce a similar product.

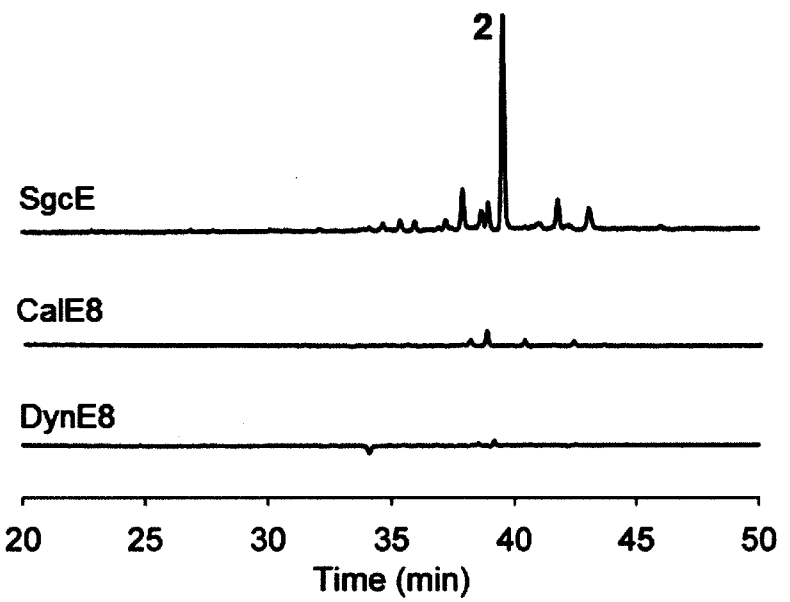


Figure 3.6. HPLC analysis of the product of SgcE reaction with UV detection at 400 nm. HPLC program: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm).

Co-expression of CalE8 and CalU15 was also conducted to see whether **2** is produced by the enediyne PKSs in the cellular environment. It was shown in chapter 2 that co-expression of SgcE and SgcE3 also produced **2.** CalU15 is a protein that shares sequence homology with SgcE3 (identity 48%, homology 61%), the putative acetylenase. Townsend et al. suggest that a *trans*-acting enzyme may be needed for proper functioning of the PKS [98]. It is of great interest to know whether CalU15 is the accessory enzyme that acts in concert with CalE8 to generate the real biosynthetic intermediate.

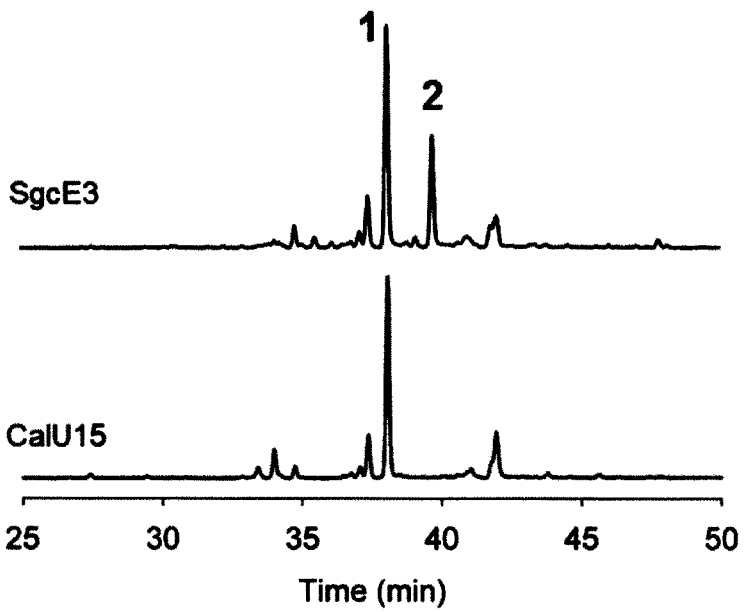


Figure 3.7. HPLC analysis of the products extracted from SgcE10 in SgcE/E3 co-expression and CalU15 in CalE8/U15 co-expression at 400 nm. HPLC program: 100% buffer A (H_2O, 0.045% TFA) to 100% buffer B (CH_3CN, 0.045% TFA) at 1 ml/min in 60 min using an eclipse XDB C18 column (4.6 × 150 mm).

During the purification of the co-expressed CalE8 and CalU15 by gel filtration, a large amount of CalU15 protein was observed to elute out with CalE8, indicating the

possible interaction between CalU15 and CalE8-tethered intermediate. The purified CalU15 looked bright yellow, similar to SgcE3 from SgcE/E3 co-expression. The yellow pigment on CalU15 could be readily extracted by organic solvent, indicating that the compounds are non-covalently bound by CalU15. HPLC analysis of the pigment extracted from the yellow colored CalU15 showed that **1** was bound by CalU15 (**Figure 3.7**). On the contrary, **2** was not detected for CalU15, indicating that CalE8 and CalU15 did not produce **2** *in vivo*.

3.4 Discussion

We were hoping the comparison of the PKSs from the biosynthetic pathways of three enediyne natural products would unveil the origin of biosynthetic divergence between 9- and 10-membered enediynes. All PKSs generate a common intermediate linked to the phosphopantetheine thiol of ACP domtain through thioester bond. The PKSE-tethered intermediate is easily cleaved by the TE to yield the conjugated polyene **3**. However, only SgcE for 9-membered enediyne core biosynthesis is able to process the intermediate and release product **2**.

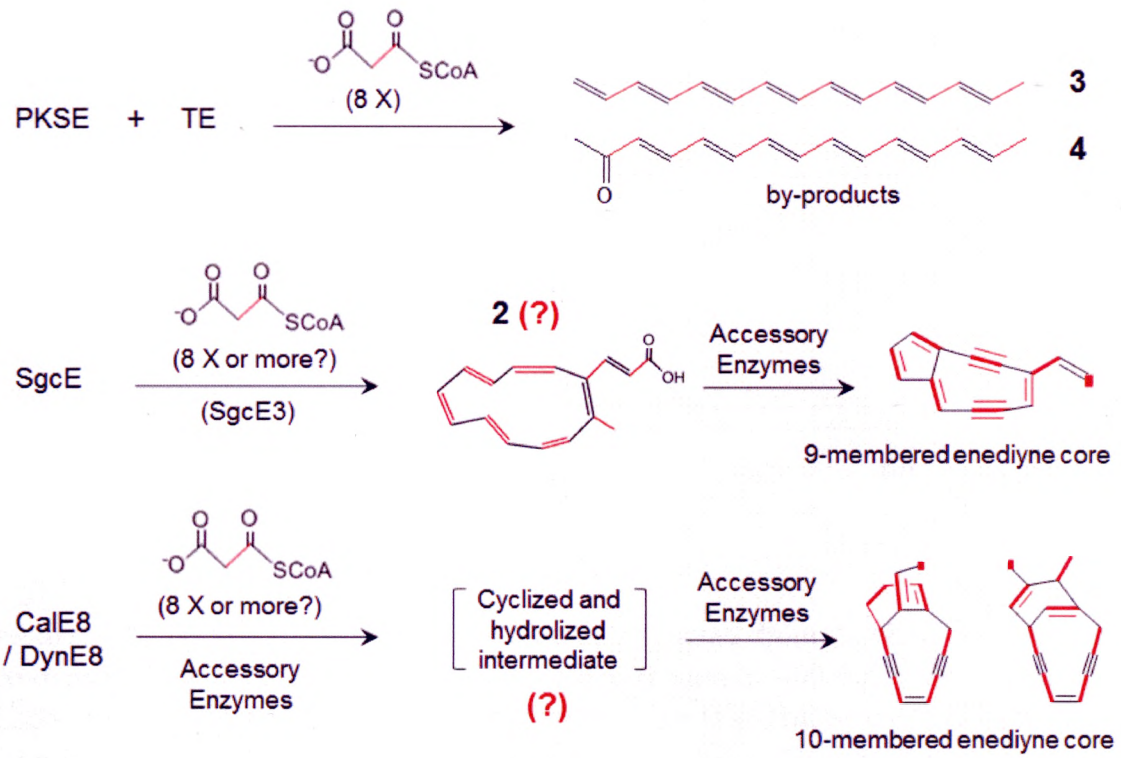


Figure 3.8. Products of enediyne PKSs. The structure of **2** is only based on the LC-MS results and needs to be further confirmed. The carbon atoms of the acetate units are highlighted in red.

Although the production of **3** by SgcE was first reported by Shen and co-workers in their *in vivo* studies [44,104], we found the SgcE can also produce additional products (**2** and **4**). More importantly, we showed that SgcE can generate **2** even in the absence of the TE SgcE10. On the contrary, CalE8 and DynE8 were not able to release any detectable products without TE. Neither CalE8 nor DynE8 produces **2** under any experimental conditions. While SgcE and SgcE3 produced **2** *in vivo*, co-expression of CalE8 and CalU15 did not produce **2**, suggesting that CalE8 may need other accessory proteins specific for 10-membered enediyne biosynthesis to process and release the PKS intermediate. The possible role of CalU15 and SgcE3 as regulatory proteins also needs to be fully established. Put together, comparison of the *in vitro* and *in vivo* products of the three PKSs without reliance on TEs seems to point towards an intrinsic catalytic difference between the two PKS families.

The enzymatic assays of PKS and TE described in this chapter established the conjugated polyene **3** as the major *in vitro* product for the three iterative PKSs from enediyne biosynthetic pathways. Townsend and co-workers recently proposed that neither **3** nor **4** is the biosynthetic intermediate based on the observation that CalE8 produced multiple products under *in vitro* conditions. The observation that **3** becomes nearly the sole product at moderate catalytic rate with physiologically relevant substrate concentrations, however, seems to suggest that **3** could be biologically relevant. Nonetheless, considering that iterative PKSs may or may not need the assistance of ancillary proteins to generate correct products, the final establishment of **3** or **4** as the biosynthetic intermediate still awaits the demonstration of the processing of the putative

intermediate. The comparison of the three PKSs revealed that CalE8 is the only PKS that produces **4** as a major product under the conditions that maintain a high catalytic rate. This strongly suggests that CalE8 has a higher propensity to skip the final reduction step compared to the other two PKSs. Whether the intrinsic differences among the PKSs can be amplified for producing different biosynthetic intermediates with the assistance of accessory proteins remains an intriguing possibility.

CHAPTER 4

Biosynthesis of (R)-Mellein by the iPKS SACE5532 from *Saccharopolyspora erythraea*

4.1 Introduction

Saccharopolyspora erythraea is a mycelium-forming *actinomycete* that has been used for the industrial-scale production of the clinically important macrolide antibiotic erythromycin A. The gene clusters for erythromycin biosynthesis (*ery*) and for a second modular PKS of unknown function (*pke*) have been previously analyzed [109,110]. Recently, Leadlay et al. reported the complete genome sequencing of *Saccharopolyspora erythraea* NRRL23338 and revealed further nine uncharacterized PKS gene clusters for polyketide biosynthesis [111]. None of the hypothetical products of these PKS gene clusters has previously been detected, even extensive fermentation experiments using 50 different solid and liquid media were conducted [110]. One of the uncharacterized orphan PKS is SACE5532 (or PKS8), which is predicted to be a single-modular PKS enzyme. Bioinformatic studies show that SACE5532 exhibits a head-to-tail homology to several characterized bacterial iterative type I PKSs (iPKSs) for aromatic polyketide biosynthesis (**Table 4.1** and **Figure 4.1**).

Table 4.1. Sequence alignment of SACE5532 and bacterial iPKSs.

Gene	Ref	Access no.	Organism	Identity%/ Homology%
NcsB	[31,35]	AAM77986	*Streptomyces carzinostaticus subsp. neocarzinostaticus*	51/66
AziB	[39]	ABY83164	*Streptomyces sahachiroi*	48/64
PokM1	[32]	ACN64831	*Streptomyces diastatochromogenes*	49/63
ChlB1	[34,37]	AAZ77673	*Streptomyces antibioticus*	46/61
MdpB	[38]	ABY66019	*Actinomadura madurae*	47/60
CalO5	[112]	AAM70355	*Micromonospora echinospora*	48/62
AviM	[30,113]	AAK83194	*Streptomyces viridochromogenes*	48/61

Although aromatic polyketide biosynthesis in fungi is catalyzed by type I iPKS [114], aromatic polyketide biosynthesis in bacteria is usually catalyzed by type II PKS. The discovery of the bacterial iPKSs for aromatic polyketide biosynthesis listed in Table 4.1 has revealed bacterial iPKSs that bear great resemblance to fungal iPKSs. AviM [30,113] and CalO5 [112] catalyze the biosynthesis of the monoyclic orsellinic acid (OSA) moiety for avilamycin (AVI) and calicheamicin (CAL). ChlB1 [34,37], MdpB [38] and PokM1 [32] catalyze the biosynthesis of 6-methylsalicyclic acid (6-MSA) moiety for chlorothricin (CHL), maduropeptin (MDP), polyketomyces (POK) and pactamycin (PTM). NcsB [31,35] catalyzes the biosynthesis of bicyclic 2-hydroxyl-5-methyl-naphthoic acid (NPA) for neocarzinostatin (NCS); and AziB[39] catalyzes the biosynthesis of bicyclic 5-methyl-NPA for Azinomycin B. One of the

biggest puzzles about these iPKSs is how they are programmed to perform selective reduction of the polyketide intermediates to produce different products. Although SACE5532 shares sequence homology with these iPKSs, it is not known whether it exhibits different reduction pattern and produces a different product.

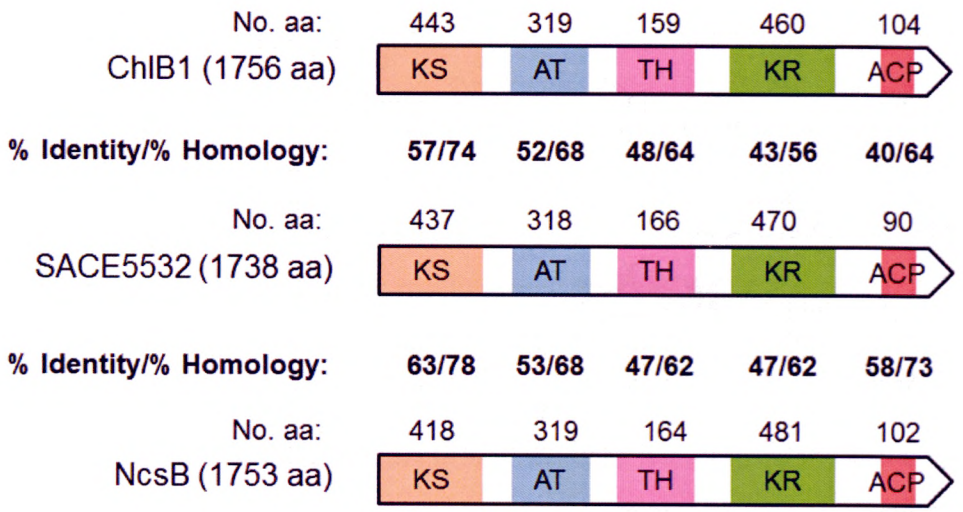


Figure 4.1. Domain organization of SACE5532 and the sequence identity/homology of the domains shared with ChlB1 and NcsB. KS, ketosynthase; AT, acyltransferase; TH, thioester hydrolase; KR, ketoreductase; ACP, acyl carrier protein.

SACE5532 has at least five predicted protein domains based on sequence homology: a ketosynthase (KS), an acyltransferase (AT), a ketoreductase (KR), a dehydratase (DH), and an acyl carrier protein (ACP) domain (**Figure 4.1**). Recently, Fujii and co-workers revealed that the previously assigned DH domain of ATX, a 6-MSA synthase (6-MSAS) from *Aspergillus terreus*, is not a dehydratase domain but a thioester hydrolase (TH) domain that catalyzes the hydrolysis of thioester bond to release the product 6-MSA from the ACP domain [115]. The finding provides an explanation on how 6-MSAS releases the final product and why OSA synthase needs a

DH-like domain without catalyzing a dehydration step. Accordingly, the DH-like domain of SACE5532 may also be a TH domain instead for the release of the final product from the ACP domain.

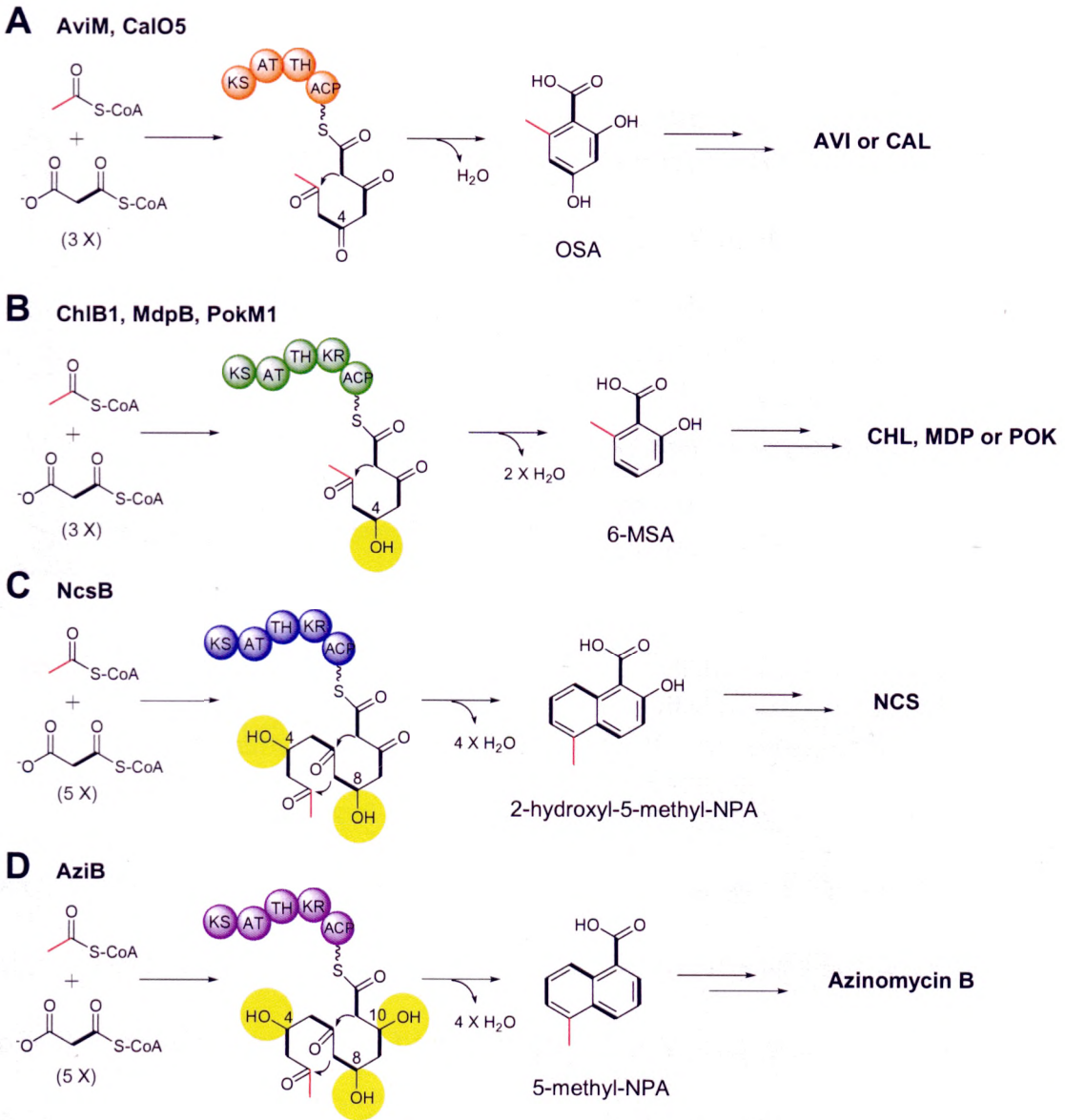


Figure 4.2. Bacterial iPKSs for aromatic polyketide biosynthesis [39]. (A) AviM and CalO5 for orsellinic acid (OSA) biosynthesis. (B) ChlB1, MdpB and PokM1 for 6-methylsalicyclic acid (6-MSA) biosynthesis. (C) NcsB for 2-hydroxyl-5-methyl-naphthoic acid (NPA) biosynthesis. (D) AziB for 5-methyl-NPA biosynthesis. KS, ketosynthase; AT, acyltransferase; TH, thioester hydrolase; KR, ketoreductase; ACP, acyl carrier protein.

Bacterial iPKSs exhibit distinct selective reduction patterns in aromatic polyketide biosynthesis, which is likely to be governed by the KR domain. For example, the KR domain in AziB selectively reduces the keto groups at C4, C8, and C10 positions, which is distinct from the actions of the KR domain of NcsB that reduces the keto groups at C4 and C8 positions. How KR domain selectively reduces certain keto groups at certain positions of the polyketide intermediates remains unknown. How the final length of the product is controlled for these iPKSs is also an unanswered question. We will use SACE5532 as a model system to investigate the intriguing programmed biosynthetic mechanism of the iPKSs.

We envision two stages of investigation for the SACE5532 project. First, we wish to identify the product of SACE5532 to see whether it produces any novel product. This can be achieved by the identification of the product produced through enzymatic reactions and structure determination. Second, we wish to elucidate which domain(s) controls the chain length and reduction pattern of the biosynthetic product. This can be achieved by study of stand-alone KR domains and domain swapping between SACE5532 and NcsB. We were also hoping that the collaboration with Prof. Xiong Yong's structural group at Yale University can help us determine the crystal structure of the full-length PKS protein or stand-alone domains (e.g. KR domain). The structures would gain major insight into the biosynthetic mechanism of these fascinating multi-functional iPKSs.

4.2 Materials and Methods

4.2.1 Strains, plasmids, and chemicals

E. coli TOP10 was used as a general host for sub-cloning. *E. coli* BL-21(DE3) (Novagen) was used as the heterologous host for protein expression. Vectors pET-28b(+) and pCDF-2 Ek/LIC were obtained from Novagen. The genes that encode SACE5532, MatB, NcsBACP, NcsBTH-KR were provided by GenScript Corporation (NJ, USA) with sequences optimized for protein expression in *E. coli*. The reported protein sequences for SACE5532, MatB and NcsB have been deposited in Genbank under the accession numbers of YP_001107644, NP_626687 and AAM77986. The genes were synthesized based on the reported protein sequences and were provided as pUC57-based plasmids. XhoI and NdeI restriction sites are added to the C-terminus and N-terminus respectively. The plasmid that contains the Sfp-encoding gene was obtained from Christopher Walsh's lab at Harvard Medical School. Coenzyme A, acetyl-CoA, malonyl-CoA, NADPH, ATP, malonic acid and other chemicals were purchased from Sigma-Aldrich and stored at -20°C. ^{13}C-labeled malonic acid (1, 2, 3-^{13}C, 99%) was purchased from Cambridge Isotope.

4.2.2 Cloning, mutagenesis and domain swapping

The plasmid pUC57- SACE5532 was digested with NdeI-XhoI to yield the SACE5532. The 5.2-kb SACE5532 fragment was gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-SACE5532.

To generate the SACE5532-C184A mutant, the PCR reaction was performed

using pET28- SACE5532 as a template with the following primers: C184A (forward) 5'- CCTGACCATTGATACTGCTGCCGCGGGCAGC -3' / (reverse) 5'- GCTGCCCGCGGCAGCAGTATCAATGGTCAGG -3'. Successful reaction mixtures consists of 100 ng of template DNA, 300 nM each primer, 300 mM dNTPs, , 1x KAPAHiFi™ Fidelity Buffer, and 0.5 U of KAPAHiFi™ DNA Polymerase in a final volume of 25 μl. The PCR program was as following: initial denaturing at 96°C for 5 minutes, followed by 18 cycles at 96°C for 50 seconds, 54°C for 50 seconds, and 68°C for 6 minutes, and completed by an additional 7 minutes at 68°C. Upon completion, 1 μl (10 U) of DpnI was added directly to the PCR mixture and digested at 37°C for 2 hours. An aliquot (5 μl) of the mixture was directly transformed into *E. coli* Top10 competent cells and plated on LB supplemented with 50 μg/ml kanamycin. The mutant constructs were confirmed by sequencing to give pET28- SACE5532-C184A.

To replace the ACP domain of SACE5532 (1649 - 1738 aa) with that of NcsB (1652-1753 aa), *SACE5532*(ΔACP:*ncsB*-ACP) was generated. The plasmid pUC57-NcsB-ACP and pUC57-SACE5532 was digested with ClaI-XhoI. The *ncsB*-ACP fragment was gel-purified and ligated into the identical sites of pUC57-SACE5532 to give pUC57- SACE5532 (ΔACP:NcsB-ACP). The plasmid was sequenced using pUC57 reverse sequence primer. Then the plasmid pUC57-SACE5532(ΔACP:NcsB-ACP) was digested with NdeI-XhoI to yield SACE5532(ΔACP:*ncsB*-ACP). The 5.2-kb SACE5532(ΔACP:*ncsB*-ACP) fragment was gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-pUC57-SACE5532(ΔACP:NcsB-ACP).

To replace the KR domain of SACE5532 (1170 – 1638 aa) with that of NcsB (1162 – 1641 aa), *SACE5532*(ΔKR:*ncsB*-KR) was generated. The synthesized gene *ncsB*-KR was swapped into KR domain of *pks8* by GenScript Corporation using seamless cloneEZ® cloning technology, leading to the pET-28b(+) plasmid harboring *pks8*(ΔKR:*ncsB*-KR).

To replace the TH-KR di-domain of SACE5532 (900 – 1638 aa) with that of NcsB (886 – 1641 aa), *SACE5532*(ΔTH-KR:*ncsB*-TH-KR) was generated. The synthesized gene *ncsB*-TH-KR was swapped into TH-KR di-domain of *pks8* by GenScript Corporation using seamless cloneEZ® cloning technology, leading to the pET-28b(+) plasmid harboring *pks8*(ΔTH-KR:*ncsB*-TH-KR).

To clone the stand-alone KR domain of SACE5532 (1170 - 1638 aa), the PCR reaction was performed using pUC57-SACE5532 as a template with the following primers: SACE5532-KR (forward) 5'-TTTCATATGCGCGATCTGGCGTATGAAATCATTTGG -3' / (reverse) 5'-TTTCTCGAGGCCGGTATCGCCGCTCGCGGTCAGTTC -3'. Successful reaction mixtures consisted of 52 ng of template DNA, 300 nM each primer, 300 mM dNTPs, , 1× KAPAHiFi™ Fidelity Buffer, and 1.0 U of KAPAHiFi™ DNA Polymerase in a final volume of 50 μl. The PCR program was as follows: initial denaturing at 95°C for 5 minutes; 6 cycles at 95°C for 30 seconds, 45°C for 30 seconds, and 72°C for 40 seconds; 30 cycles at 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 40 seconds; and an additional 7 minutes at 72°C. The 1.5-kb PCR product was gel-purified and digested with NdeI-XhoI. The digested SACE5532-KR fragment was

gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-SACE5532KR. Then the plasmid harboring the SACE5532-KR was sequenced and transformed into *E. coli* strain BL21(DE3) for protein expression.

To clone the stand-alone KR domain of NcsB (1162 – 1641 aa), the PCR reaction was performed using pUC57-NcsB-TH-KR as a template with the following primers: NcsB-KR (forward) 5'- TTTCATATGAGCGAACTGGTTCACGAAATCGTCTGG -3' / (reverse) 5'- TTTCTCGAGGCCATCCGTTTCGCCAGACACCGGCAG -3'. Successful reaction mixtures consisted of 100 ng of template DNA, 300 nM each primer, 300 mM dNTPs, , 1× KAPAHiFi™ Fidelity Buffer, and 1.0 U of KAPAHiFi™ DNA Polymerase in a final volume of 50 μl. The PCR program was as follows: initial denaturing at 95°C for 5 minutes; 6 cycles at 95°C for 30 seconds, 45°C for 30 seconds, and 72°C for 40 seconds; 30 cycles at 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 40 seconds; and an additional 7 minutes at 72°C. The 1.5-kb PCR product was gel-purified and digested with NdeI-XhoI. The digested SACE5532-KR fragment was gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-NcsB-KR. Then the plasmid harboring the NcsB-KR was sequenced and transformed into E. coli strain BL21(DE3) for protein expression.

The gene encoding Sfp was from Christopher Walsh's lab at Harvard Medical School and cloned into pCDF-2 plasmid to give pCDF-Sfp. The plasmid pUC57-MatB was digested with NdeI-XhoI to yield the *matB*. The 1.5-kb *matB* fragment was gel-purified and ligated into the identical sites of pET-28b(+) to give pET28-MatB. Then the plasmid harboring the *matB* was sequenced and transformed into *E. coli*

strain BL21(DE3) for protein expression.

4.2.3 Protein expression and purification

Co-expression of SACE5532 and Sfp. pET28-SACE5532 and pCDF-Sfp were co-transformed into *E. coli* BL21(DE3) competent cells. The cells were plated on LB medium supplemented with 50 μg/ml kanamycin and 50 μg/ml streptomycin. The colonies were screened by PCR to confirm the existence of both pET28-SACE5532 and pCDF-Sfp. A single colony was used to inoculate 20 ml of LB medium containing both kanamycin (50 μg/ml) and streptomycin (50 μg/ml), and incubated overnight at 37°C at 200 rpm. A 5 ml aliquot was transferred to 500 ml of LB medium (added with 10% glycerol) containing both kanamycin (50 μg/ml) and streptomycin (50 μg/ml), and grown at 37°C at 200 rpm. When OD_{600} reached ~0.5 (~4 hours), the culture was cooled down to 16°C, induced with 0.8 mM IPTG. After incubation at 16°C for an additional ~20 hours at 130 rpm, cells were harvested and spun at 8,000 rpm. The cell pellet was re-suspended in lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, 5 mM β-ME and 10% (v/v) glycerol] and lysed by sonication. After centrifugation at 20,000 rpm for 30 minutes at 4°C, the supernatant was filtered by 0.45 μm membrane and loaded onto HiTrap™ Ni^{2+}-NTA column (GE Healthcare). The column was then washed by lysis buffer and wash buffer containing 40 mM imidazole before eluted with elution buffer containing 500 mM imidazole. The eluted SACE5532 and Sfp were further purified by gel filtration using a HiLoad™ 16/60 Superdex™ 200 column (GE Healthcare). Proteins were desalted into Tris buffer [50 mM Tris-HCl (pH

8.0), 150 mM NaCl, 1 mM DTT and 10% (v/v) glycerol]. The purity was determined by SDS-PAGE. The protein concentration and profile were examined by UV-Vis spectrometer before the proteins were flash frozen in liquid nitrogen and stored in -80°C freezer.

Expression of SACE5532(ΔACP:NcsB-ACP), SACE5532(ΔKR:NcsB-KR) and SACE5532(ΔTH-KR:NcsB-TH-KR). SACE5532(ΔACP:NcsB-ACP), SACE5532(ΔKR:NcsB-KR) and SACE5532(ΔTH-KR:NcsB-TH-KR) were co-expressed with Sfp. The co-expression was carried out following the similar procedure described for the co-expression of SACE5532 and Sfp.

Expression and purification of Sfp. pCDF-Sfp was transformed into *E. coli* BL21(DE3) competent cells. The cells were plated on LB medium supplemented with 50 μg/ml streptomycin. A single colony was used to inoculate 20 ml of LB medium supplemented with 50 μg/ml streptomycin, and incubated overnight at 37°C at 200 rpm. A 5 ml aliquot was transferred to 500 ml of LB medium supplemented with 50 μg/ml streptomycin, and grown at 37°C at 200 rpm. When OD_{600} reached ~0.6 (~3 hours), the culture was cooled down to 16°C and induced with 0.2 mM IPTG. After incubation at 16°C for an additional ~20 hours at 130 rpm, cells were harvested and spun at 8,000 rpm. The cell pellet was re-suspended in lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, 5 mM β-ME and 10% (v/v) glycerol] and lysed by sonication. After centrifugation at 20,000 rpm for 30 minutes at 4°C, the supernatant was filtered by 0.45 μm membrane and loaded onto HiTrap™ Ni^{2+}-NTA column. The column was then washed by lysis buffer and wash buffer containing 40

mM imidazole before eluted with elution buffer containing 500 mM imidazole. The eluted protein was further purified by gel filtration using a HiLoad™ 16/60 Superdex™ 200 column. Proteins were desalted into Tris buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT and 10% (v/v) glycerol]. Its purity was determined to be >90 % by SDS-PAGE. The protein was concentrated, flash frozen in liquid nitrogen, and stored in -80°C freezer.

Expression and purification of MatB, $KR_{SACE5532}$ and KR_{NcsB}. The expression and purification of MatB, $KR_{SACE5532}$ and KR_{NcsB} were similar to the procedure described above for Sfp, except that cells were grown in LB media supplemented with 50 μg/ml kanamycin.

4.2.4 *In vitro* enzymatic assays and HPLC analysis

For the enzymatic assays of SACE5532, SACE5532-C184A, SACE5532(ΔACP:NcsB-ACP), SACE5532(ΔKR:NcsB-KR) and SACE5532(ΔTH-KR:NcsB-TH-KR), a typical enzymatic reaction contained 3 μl of $MgCl_2$ (1 M), 8 μl of CoA (50 mM), 70 μl of SACE5532 (11.3 mg/ml), 20 μl of Sfp (27.9 mg/ml) and 93 μl of reaction buffer [50 mM Tris (pH 8.5), 150 mM NaCl and 1 mM DTT]. After incubation at 30°C for 20 minutes, the reaction mixture was added with 5 μl of NADPH (10 mM), 1μl of acetyl CoA (100 mM) and 2 μl of malonyl CoA (100 mM) and incubated at 30°C for 3 hours. The reaction was quenched with 5 μl of 6 mM HCl and vortexed to totally precipitate the enzymes. Then the mixture was spun at 14,800 rpm for 10 min and the supernatant was loaded for HPLC analysis.

For the kinetic analysis of SACE5532 reaction, a total volume of 200 μl of the reaction was carried out under the same conditions as above. 20 μl aliquots were taken from the reaction mixture at 0, 15, 30, 45, 60, 75, 90 and 120 minute. The aliquot at each time point was quenched with 5 μl of 6 mM HCl and vortexed to totally precipitate the enzymes. Then the mixture was spun at 14,800 rpm for 10 minutes and the supernatant was loaded for HPLC analysis.

HPLC analysis was performed with an analytical eclipse XDB C18 column (4.6 × 150 mm) using an Agilent 1200 HPLC. A full gradient was employed from 100% buffer A (HPLC grade water with 0.045% TFA) to 40% buffer A + 60% Buffer B (100% acetonitrile with 0.045% TFA) at 1 ml/min in 60 minutes.

The *in vitro* assays of the KR domain activity were conducted by using a semi-micro quartz cuvette and a Shimazu UV-Vis 1700 spectrophotometer. When *trans*-1-decalone was used as the substrate, a typical enzymatic reaction contained 0.82 mg/ml $KR_{SACE5532}$ or KR_{NcsB} protein, 0.25 mM NADPH and 10 mM *trans*-1-decalone in 100 mM Tris buffer (pH 8.0) in a total volume of 200 μl. The reaction was incubated at 37°C within the sample chamber through an external temperature controller. When S-Ethyl acetothioacetate was used as the substrate, a typical enzymatic reaction contains 0.02 mg/ml $KR_{SACE5532}$ or KR_{NcsB}, 0.25 mM NADPH and 3.7 mM S-Ethyl acetothioacetate in 100 mM HEPES buffer (pH 8.5) in a total volume of 200 μl. The reaction was incubated at 20°C within the sample chamber through an external temperature controller. When acetoacetyl-SNAC was used as the substrate, a typical enzymatic reaction contains 1.46 mg/ml $KR_{SACE5532}$ or KR_{NcsB},

0.25 mM NADPH and 3.3 mM acetoacetyl-SNAC in 100 mM HEPES buffer (pH 8.5) in a total volume of 200 μl. The reaction was incubated at 30°C within the sample chamber through an external temperature controller. The reaction progress was monitored continuously by recording the NADPH absorbance at 340 nm.

4.2.5 *In vivo* production of SACE5532 product

To produce SACE5532 product *in vivo*, SACE5532 and Sfp were co-expressed in *E. coli* BL21(DE3) strain under the similar conditions described above for protein expression, except that the cells were added with 10 mM $MgCl_2$ after induction.

For SACE5532 product isolation, the cells were centrifuged at 4 °C and 8,000 rpm for 10 min. After removal of the cell pellets, the resulting supernatant was extracted twice with an equal volume of ethyl acetate. The combined organic extract was immediately dried over anhydrous magnesium sulfate, concentrated in vacuum, and resolved in methanol for HPLC analysis.

4.2.6 Large scale preparation of SACE5532 product

For the scaled-up reaction, malonyl-CoA was synthesized *in vitro* by using the malonyl-CoA synthase MatB. The MatB reaction was carried out in 60 ml of 100 mM HEPES buffer (pH 8.5), containing 20 mM malonic acid, 10 mM $MgCl_2$, 5 mM ATP, 1 mM CoA and 30 mg MatB. The reaction was incubated at 23°C overnight and analyzed by HPLC to ensure formation of malonyl-CoA. The SACE5532 reaction was set up by adding 400 μl of 100 mM acetyl-CoA, 1.5 ml of 10 mM NADPH and 3 ml of

SACE5532 and Sfp mixture purified together after PD-10 column (~ 60 mg). After incubation of the reaction in 30°C water bath for 2 hours, additional 1.5 ml of 10 mM NADPH was added into the reaction mixture. The reaction was further incubated in 30°C water bath for 2 hours. The final reaction mixture was extracted twice with ethyl acetate (2 × 60 ml). The combined organic extract was immediately dried over anhydrous magnesium sulfate and concentrated in vacuum. The compound was purified by preparative thin layer chromatography (TLC) with hexane: ethyl acetate (5: 1). The pure fractions collected from preparative TLC were collected to be analyzed by LC-MS and NMR spectroscopy for structure determination.

The preparation of ^{13}C-labeled SACE5532 product followed essentially the same protocol described above except that ^{13}C-malonic acid was used to replace the normal malonic acid. TLC purification of the labeled product was performed the same way described above.

4.2.7 LC-MS and NMR spectroscopy

High-resolution LC-MS was carried out on a Michrom Rp18 column (0.1 × 50 mm). The gradient employed in the analysis was from 99% buffer A (HPLC grade water with 0.1% FA) + 1% buffer B (100% acetonitrile with 0.1% FA) to 40% buffer A + 60% buffer B in 30 minutes. The flow rate was set at 1μl/min and the UV detector was set at 314 nm. The ionization energy was set with Nanospray Ionization source for the Finnigan LTQ Orbitrap mass spectrometer (Thermo Electron). The results were analyzed with the software Xcalibur for the determination of plausible molecular

compositions based on the observed *m/z*.

1D ^{1}H NMR, 1D ^{13}C NMR and 2D NOESY spectra were collected on a Bruker 400 MHz NMR spectrometer (Bruker DPX 400) using $CDCl_3$ as the solvent and TMS as the internal reference. About 0.5 mg of the SACE5532 product was obtained from large-scale *in vitro* reactions for NMR analysis.

4.2.8 Synthesis of acetoacetyl-SNAC.

The compound acetoacetyl-SNAC was prepared according to literature reported procedure [116]. Analytical thin layer chromatography (TLC) was performed using pre-coated silica gel plate. Visualization was achieved by UV light (254 nm) and/or $KMnO_4$ stain. Flash chromatography was performed using silica gel and a gradient solvent system (ethyl acetate: hexane as eluent). NMR spectra were recorded at room temperature on Bruker DPX 400 spectrometers with $CDCl_3$ as the solvent and TMS as the internal reference. ^{1}H NMR spectrum [$CDCl_3$, 400 MHz]: δ 5.95 (1H, s, NH), δ 2.27 (3H, s, H-1), δ 3.71 (2H, m, H-3), δ 3.10 (2H, m, H-5), δ 3.46 (2H, m, H-6), δ 1.97 (3H, s, H-8). ^{13}C NMR spectrum [$CDCl_3$, 400 MHz]: 199.86, 192.29, 170.47 58.03, 39.17, 30.29, 29.25, 23.17.

4.3 Results

4.3.1 Expression and purification of SACE5532

Activation of SACE5532 by phosphopantetheinyl transferase (PPTase) is a prerequisite for the function of the PKS. A significant amount of soluble SACE5532 could be expressed when the protein was co-expressed with the PPTase Sfp in *E. coli*. As found by another graduate student (Lawrence Ho) in our lab, growing the cells in the LB medium supplemented with 10% glycerol further improved the protein yield. The two proteins eluted from Ni^{2+}-NTA column was approximately 85% pure as estimated by SDS-PAGE analysis (**Figure 4.3A**). SACE5532 was further separated from Sfp by size-exclusion gel filtration.

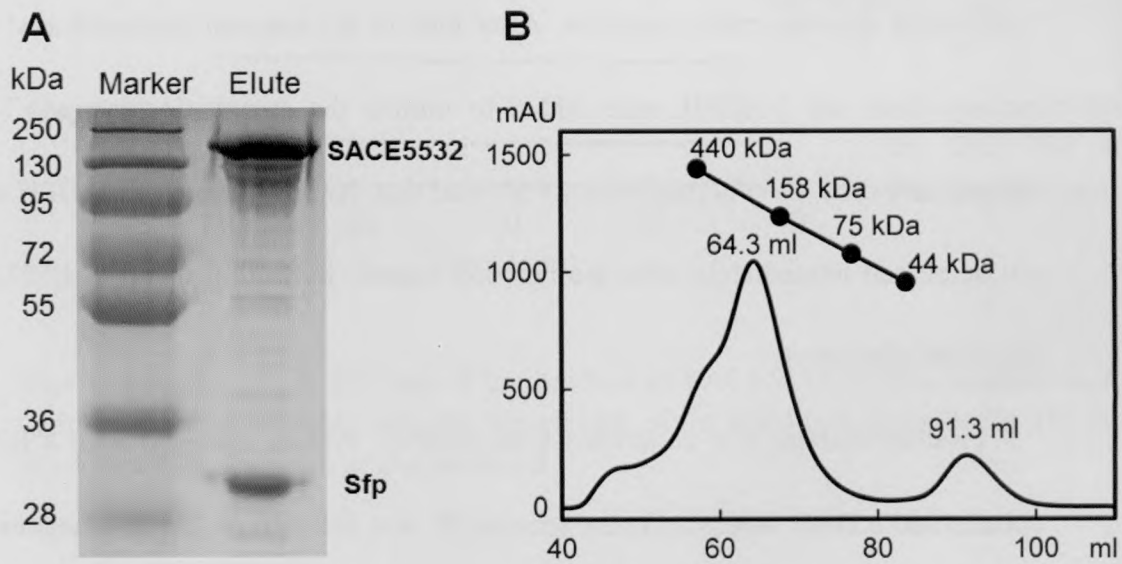


Figure 4.3. (A) SDS-PAGE for SACE5532 and Sfp eluted from Ni^{2+}-NTA column. (B) Gel filtration of SACE5532 and Sfp using a HiLoad™ 16/60 Superdex™ 200 column.

The theoretical molecular weight of SACE5532 is 188.8 kDa. Gel filtration analysis showed the elution volume of the protein was 64.3 ml (**Figure 4.3B**) with an estimated molecular weight of ~190 kDa, suggesting that SACE5532 is present mainly

as a monomer in solution. The concentrated proteins can be stored at -80°C for at least one week before the loss of enzymatic activity.

4.3.2 *In vitro* enzymatic assays and kinetic analysis

To identify the product of SACE5532, SACE5532 was first examined by *in vitro* activity assay. As mentioned earlier, PPTase is needed to transfer the phosphopantetheine component of CoA onto a conserved Ser of ACP domain to activate the SACE5532. Although SACE5532 is co-expressed with the PPTase Sfp, we found a large portion of SACE5532 is still not fully modified *in vivo*. To increase the percentage of activated SACE5532, the protein was further incubated with Sfp and CoA before any enzymatic reactions. After that, to the reaction mixture acetyl-CoA, malonyl-CoA and NADPH were added to initiate the enzymatic reaction. To our delight, real-time UV-Vis spectroscopy showed that the absorbance of NADPH at 342 nm decreased immediately after the reaction started, indicating that SACE5532 was enzymatically active.

Product analysis was conducted with a HPLC system equipped with a RP-C18 column and a DAD detector. A new species (**5**) was detected on the chromagram with absorbance at 246 and 314 nm (**Figure 4.4**). When acetyl-CoA was not included in the reaction mixture, NADPH was not consumed, indicating that SACE5532 uses acetyl-CoA as a starter unit.

To confirm that **5** is indeed produced by SACE5532, three control experiments were conducted. The first two control reactions were carried out without the extender

substrate malonyl-CoA or without modifying protein Sfp. The third control reaction replaced the wild type SACE5532 with the mutant SACE5532-C184A, in which the predicted essential residue Cys^{184} in the KS domain was replaced by Ala. According to the HPLC analysis, no **5** was detected for any of these control reactions (**Figure 4.4**). These experiments strongly suggest that **5** is enzymatically generated by SACE5532 by using the substrates acetyl-CoA and malonyl-CoA.

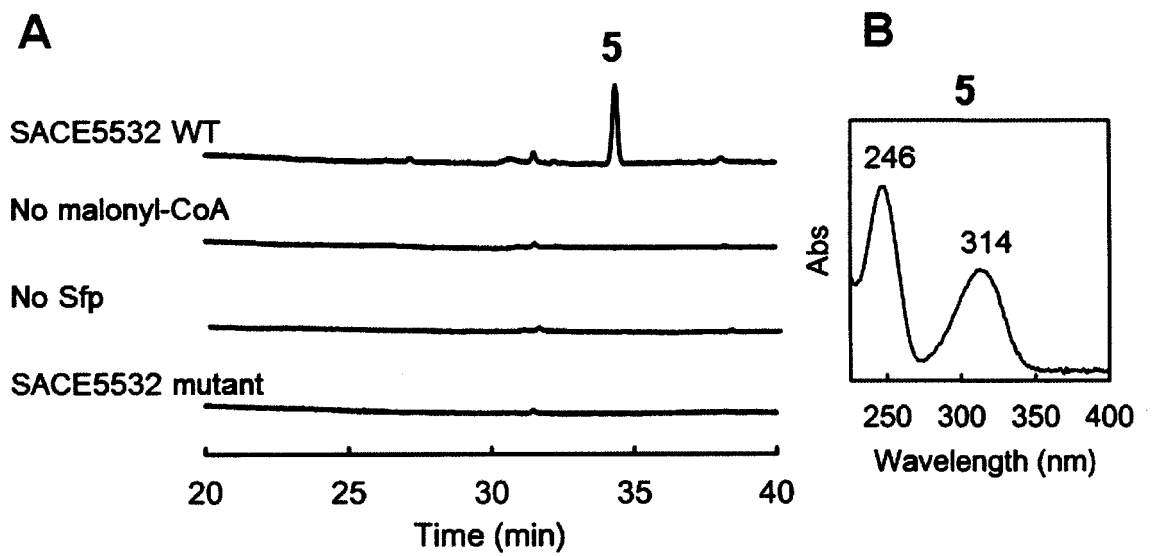


Figure 4.4. (A) HPLC analysis of the product of SACE5532 *in vitro* reaction along with the negative controls with the wavelength of the UV-Vis detector set at 314 nm. (B) Absorption spectrum of the product **5**.

Kinetic experiment was conducted to examine the time-dependent formation of **5**. Equal volumes of reaction mixtures incubated for 0, 15, 30, 45, 60, 75, 90, and 120 minute were quenched by HCl and examined by HPLC. The peak area of **5** at 314 nm increased steadily during the first 60 minutes and leveled off after that due to substrates exhaustion (**Figure 4.5**). This is fully consistent with a typical enzymatic

reaction with the rate of product formation. The kinetic experiment further established that the recombinant SACE5532 is catalytically active by producing a single enzymatic product **5**.

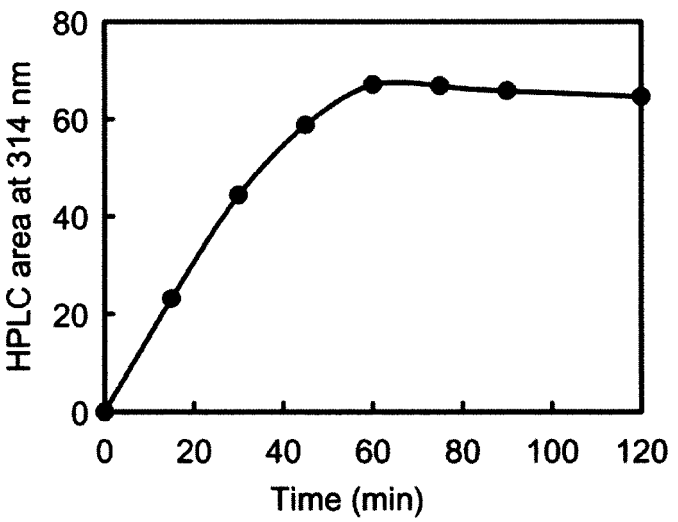


Figure 4.5. Time-dependent synthesis of **5** by SACE5532 by using acetyl-CoA, malonyl-CoA and NAPDH as substrates.

4.3.3 *In vivo* generation of SACE5532 product

To find out whether SACE5532 generates the same product in cellular environment, the medium of the *E. coli* cells co-expressing SACE5532 and Sfp was extracted with organic solvent. To facilitate the modification of SACE5532 by Sfp, the cell culture was supplemented with 10 mM $MgCl_2$ after induction. HPLC analysis of the medium extract using HPLC showed that **5** was also produced by SACE5532 in *E. coli* cells (**Figure 4.6**).

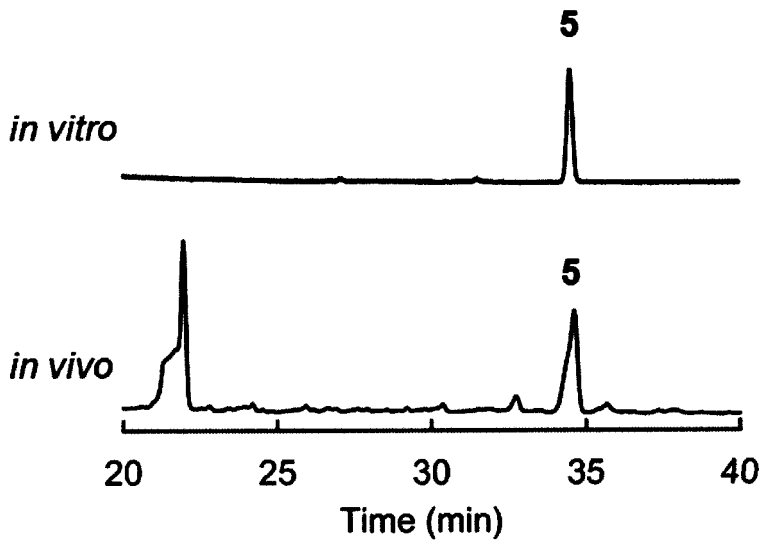


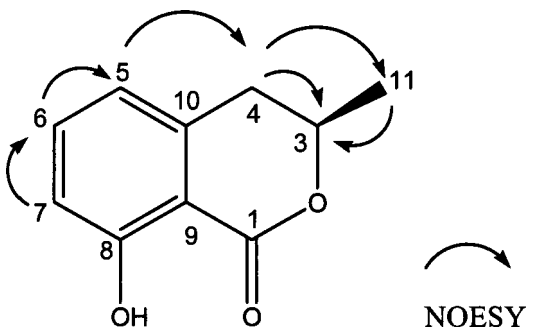
Figure 4.6. HPLC analysis of the product of SACE5532 produced by *in vitro* reaction and by *in vivo* co-expression of SACE5532 and Sfp in *E. coli*.

4.3.4 Structure determination of SACE5532 product

High-resolution LC-MS was performed to yield a $[M+H]^+$ ion with a *m/z* of 179.0723 for **5**, which suggests a molecular formula of $C_{10}H_{10}O_3$ (calculated *m/z* = 178.0630) with six degrees of unsaturation. Moreover, the ^{13}C-labeled **5** prepared from ^{13}C-labeled malonyl-CoA and unlabeled acetyl-CoA exhibits a $[M+H]^+$ ion with a *m/z* of 187.0992 (**Figure A2**). The difference of 8 a.u. suggests that the eight carbon atoms of ^{13}C-labeled **5** are from ^{13}C-labeled extender unit malonyl-CoA, while the other two are from the unlabeled starter unit acetyl-CoA.

The absorption spectrum and molecular weight of the product indicate that the product of SACE5532 is different from these of the known bacterial iPKSs (Table 4.1). To determine the structure of **5**, about 0.5 mg **5** purified by preparative TLC was dissolved in $CDCl_3$ and characterized by 1D ^1H NMR, 1D ^{13}C NMR and 2D ^1H, ^1H-NOESY NMR. The NMR data together with the MS results allowed us to deduce

the structure of **5** (8-hydroxy-3-methyl-3,4-dihydroisocoumarin or mellein) (**Figure 4.7**). The chemical shifts and coupling constants for the one-dimensional ^{1}H NMR spectrum [$CDCl_3$, 400 MHz, appendix figure A3] are: δ 4.73 (1H, m, H-3), δ 2.93 (2H, d, J = 7.2 Hz, H-4), δ 6.69 (1H, J = 7.2 Hz, H-5), δ 7.41 (1H, t, J = 7.8 Hz, H-6), δ 6.89 (1H, d, J = 8.4 Hz, H-7), δ 1.53 (3H, d, J = 6.4 Hz, H-11), δ 11.03 (1H, s, OH-C8). The chemical shifts for the one-dimensional ^{13}C NMR spectrum [$CDCl_3$, 400 MHz, appendix figure A4] are: δ 169.94 (C-1), δ 76.09 (C-3), δ 34.63 (C-4), δ 117.88 (C-5), δ 136.13 (C-6), δ 116.27 (C-7), δ 162.23 (C-8), δ 108.32 (C-9), δ 139.38 (C-10), δ 20.76 (C-11). 2D 1H, 1H-NOESY NMR [$CDCl_3$, 400 MHz, appendix figure A5]: (H-7, H-6), (H-6, H-5), (H-5, H-4), (H-4, H-3), (H-4, H-11), (H-11, H-3). The 1D ^{1}H NMR and 1D ^{13}C NMR are also in excellent agreement with those obtained for a synthesized mellein [117]. The optical rotation of the **5** was measured by using a polarimeter as $[\alpha]_D^{22}$ = -123° (c 0.43, $CHCl_3$), which is similar to the specific ration measured for the synthetic (R)-(-) -mellein ($[\alpha]_D^{22}$ = -102° (c 0.53, $CHCl_3$) [117].



(R)-mellein (**5**)
(8-hydroxy-3-methyl-3,4-dihydroisocoumarin)

Figure 4.7. Determined structure of product **5** with the arrows indicating the NOE effect (see spectrum in appendix figure A5).

4.3.5 Domain swapping

Sequence alignments suggest that SACE5532 shares the highest sequence homology with NcsB among the known iPKSs. The products of SACE5532 and NcsB differ strikingly in chain length and reduction pattern. NcsB produces 2-hydroxyl-5-methyl-NPA by the cyclization of the hexaketide intermediate, while SACE5532 generates (R)-mellein (**5**) via a putative two-step cyclization of the pentaketide intermeidate. To find out which domain determines chain length and the programmed keto-reduction, we set out to replace ACP domain, KR domain or TH-KR di-domain of SACE5532 with the corresponding domain of NcsB.

To replace the ACP domain of SACE5532 (1649 - 1738 aa) with that of NcsB (1652-1753 aa), the construct *SACE5532*(ΔACP:*ncsB*-ACP) was generated. The resulted SACE5532(ΔACP:NcsB-ACP) was co-expressed with Sfp in *E. coli* and purified for enzymatic assay. The enzymatic reaction was carried out under the same conditions as wild-type SACE5532. Interestingly, HPLC analysis showed that SACE5532(ΔACP:NcsB-ACP) produces the same product **5** as wild-type SACE5532 (**Figure 4.8A**). The observation that replacement of the ACP domain does not affect the formation of **5** indicates that the ACP domain plays little role influencing the chain length or reduction pattern.

To replace the KR domain of SACE5532 (1170 – 1638 aa) with that of NcsB (1162 – 1641 aa), the construct *SACE5532*(ΔKR:*ncsB*-KR) was generated. The resulted SACE5532(ΔKR:NcsB-KR) was co-expressed with Sfp in *E. coli* and purified for enzymatic assay. The enzymatic reaction was carried out under the same conditions

as wild-type SACE5532. The SACE5532(ΔKR:NcsB-KR) protein construct was enzymatically inactive and did not produce any product (**Figure 4.8A**). It is plausible that KR domain swapping may have disrupted the domain-domain interface between the TH and KR domains, which could be critical for maintaining the overall architecture of the quaternary PKS structure.

Instead of replacing the KR domain alone, I next replaced the TH-KR di-domain of SACE5532 to avoid disrupting the TH-KR domain interface. To replace the TH-KR di-domain of SACE5532 (900 – 1638 aa) with that of NcsB (886 – 1641 aa), *SACE5532*(ΔTH-KR:*ncsB*-TH-KR) was generated. The resulted SACE5532(ΔTH-KR:NcsB-TH-KR) was co-expressed with Sfp in *E. coli* and purified for enzymatic assay. The enzymatic reaction was carried out under the same conditions as wild-type SACE5532. HPLC analysis showed that SACE5532(ΔACP:NcsB-ACP) produces a new product **6**, instead of **5** by wild-type SACE5532 (**Figure 4.8A**). Intriguingly, the new product **6** shares the same retention time and UV-Vis spectrum with 6-MSA (**Figure 4.8B**). 6-MSA and mellein differs in both reduction pattern and chain length, suggesting that TH-KR di-domain plays a role in the programming of the iterative PKSs.

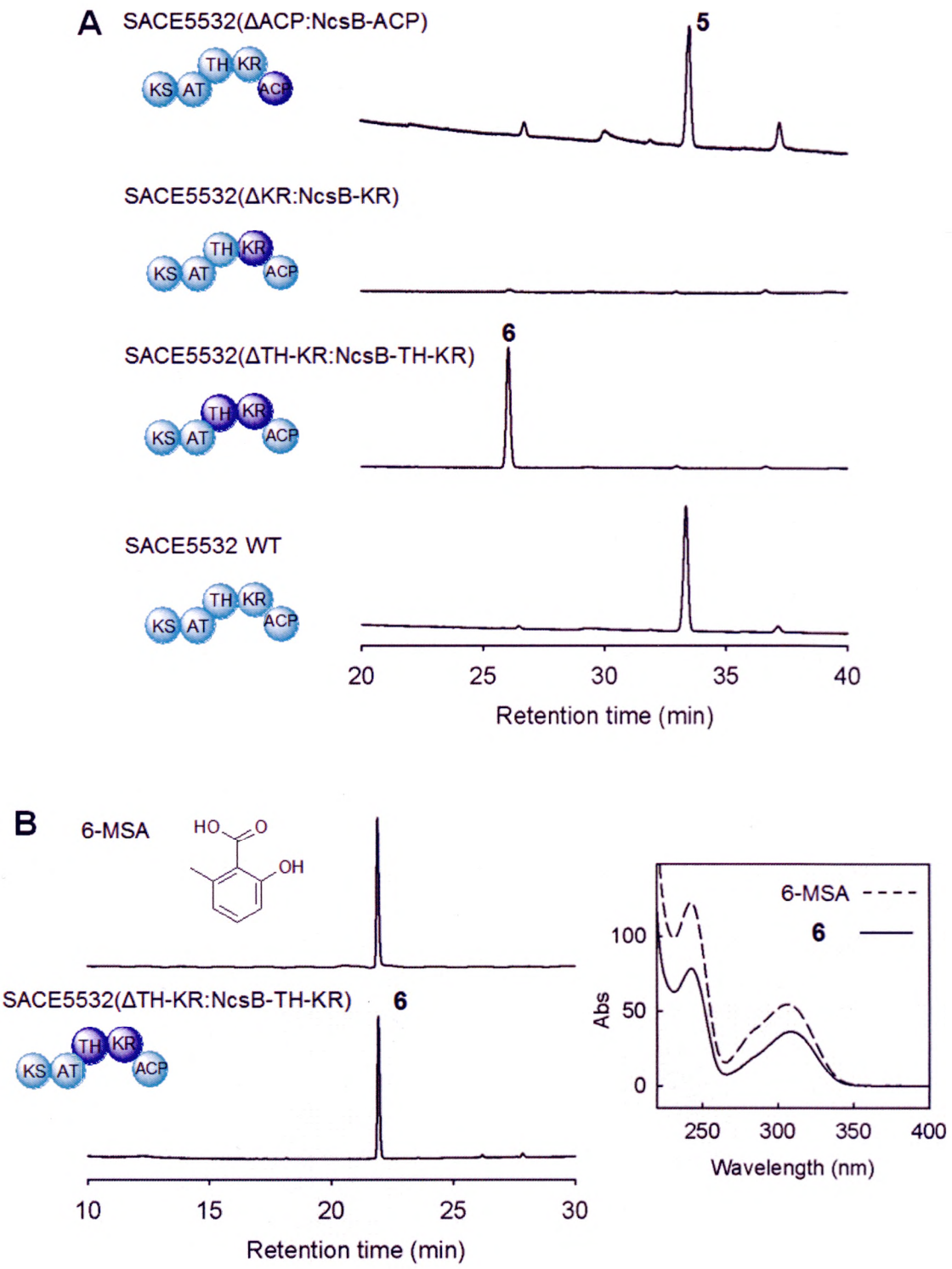


Figure 4.8. (A) HPLC analysis of the product of SACE5532(ΔACP:NcsB-ACP), SACE5532(ΔKR:NcsB-KR), SACE5532(ΔTH-KR:NcsB-TH-KR) and wild-type SACE5532 with UV detection at 314 nm. HPLC program: 0-60 min, 100% buffer A (H_2O, 0.045% TFA) to 60% buffer B (CH_3CN, 0.045% TFA). (B) Comparison of the product 6 with 6-MSA standard. HPLC program: 0-5 min, 10% B; 5-25 min, a linear gradient from 10% B to 60% B; 25-30 min, a linear gradient from 60% B to 90% B.

4.3.6 Determination of reduction pattern by the KR domain

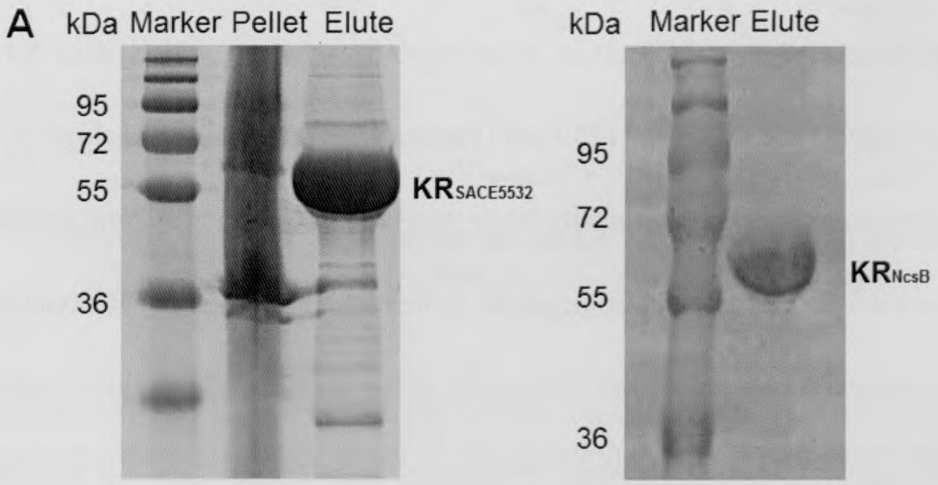
Although SACE5532 shares the same domain composition and head-to-tail sequence homology with ChlB1, MdpB1, PokM1, NcsB and AziB, the mechanism depicted in **Figure 4.11** indicates that SACE5532 exhibits a completely different keto-reduction pattern from the other iPKSs. As illustrated in **Figure 4.12**, while the other homologous iPKSs reduce the keto group at C4, C8 and C10 positions, SACE5532 reduces the keto groups at C2 and C6 positions. In other words, while ChlB1, NcsB and AziB reduce the keto group in the tri-, penta- and hexaketide intermediates, SACE5532 reduces the keto group in the diketide and tetraketide intermediates. To support the proposed mechanism and diketide intermediates in mellein synthesis, we show that SACE5532 is able to accept acetoacetyl-CoA or β-hydroxybutyryl-CoA as starter unit to initiate the synthesis of mellein, albeit at lower rates (**Figure 4.10A**). Direct loading of the two acyl groups presumably enables the enzyme to bypass the first condensation step. This also provides support for the reduction of the diketide intermediate during chain extension.

According to the mechanisms illustrated in **Figure 4.11** and **Figure 4.12**, the keto group of the diketide intermediate is reduced by SACE5532, but not by ChlB1, NcsB or AziB (**Figure 4.10B**). How does the mellein synthase achieve such a completely different keto-reduction pattern? To test whether the KR domain is solely responsible for determining the keto-reduction pattern, we cloned and expressed the stand-alone KR domains from SACE5532 (1170 1638 aa) and NcsB (1162 – 1641 aa).

As with the stand-alone KR domains of modular PKSs (e.g. EryKR1 and TylKR1

[118,119]), $KR_{SACE5532}$ and K_{RNcsB} are composed of two sub-domains. While the assumed TH-KR inter-domain linker forms a structural sub-domain, the "KR" region and the adjacent ~70 residues form the catalytic sub-domain for NADPH binding and substrate reduction. The $KR_{SACE5532}$ and KR_{NcsB} domain proteins were expressed as soluble monomeric proteins (**Figure 4.9A**).

Multiple sequences alignment of the KR domains of SACE5532, AziB, NcsB, MdpB and ChlB1 shows that the KR domains all contain the critical catalytic residues of Lys, Ser, Gln, Tyr and Asn [120-122] (**Figure 4.9B**). Based on previous studies [118,122], similar to the KR domains of AziB, NcsB, MdpB and ChlB1, the KR domain of SACE5532 lacks the His residue at the active site groove and the characteristic L-D-D motif. This would suggest that the KR domain of SACE5532 belongs to the so-called A1 type KR family that acts on the keto group to yield L-configured hydroxyl exclusively [118].



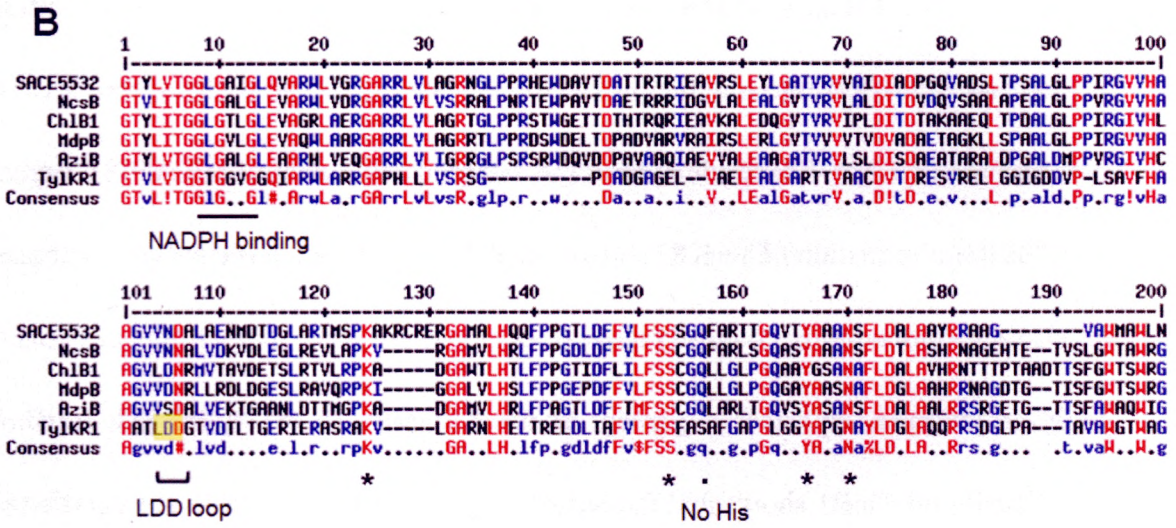


Figure 4.9. (A) SDS-PAGE for the stand-alone KR domains. (B) Multiple sequences alignment of the catalytic KR subdomains of iterative PKSs. The PKSs included SACE5532, NcsB, ChlB1, MdpB and AziB, with TrylKR1 (PDB ID: 2FR0) from the modular type I PKS tylosin PKS. The residues with high consensus value (90%) are in red, and the residues with low consensus value (50%) are in blue. The catalytic residues K, S, Y and N are labeled with asterisks. The conserved motif (GXGXXG) for NADPH binding is underlined. The KR domains of the iterative type I PKSs lack the His residue at the active site groove (highlighted with dot) and characteristic L-D-D motif (highlighted with yellow shadow).

Both $KR_{SACE5532}$ and KR_{NcsB} are enzymatically active as demonstrated by the reduction of *trans*-1-decalone, a non-specific substrate widely used for assaying the activity of KR domains [123,124] (**Figure 4.10C**). Since the analogs of the tri-, tetra- and pentaketide are chemically labile due to their tendency to form lactone [125], only two diketide analogs were prepared to test whether the two KR domains exhibit any substrate preference. When the analog of the diketide intermediate acetoacetyl-SNAC (N-acetylcysteamine thioeter) was used as substrate, $KR_{SACE553}$ is capable of reducing the keto group as evidenced by the depletion of NADPH. However, no reduction was observed by for KR_{NcsB}. Remarkably, $KR_{SACE5532}$ also exhibits enzymatic activity

towards one of the simplest diketide analogs – S-ethyl acetothioacetate, indicating that the KR domains can differentiate the small acetoacetyl moiety. These results strongly suggest that the KR domain alone is able to discriminate the polyketide intermediates, leading to the programmed keto-reduction in these iterative PKSs.

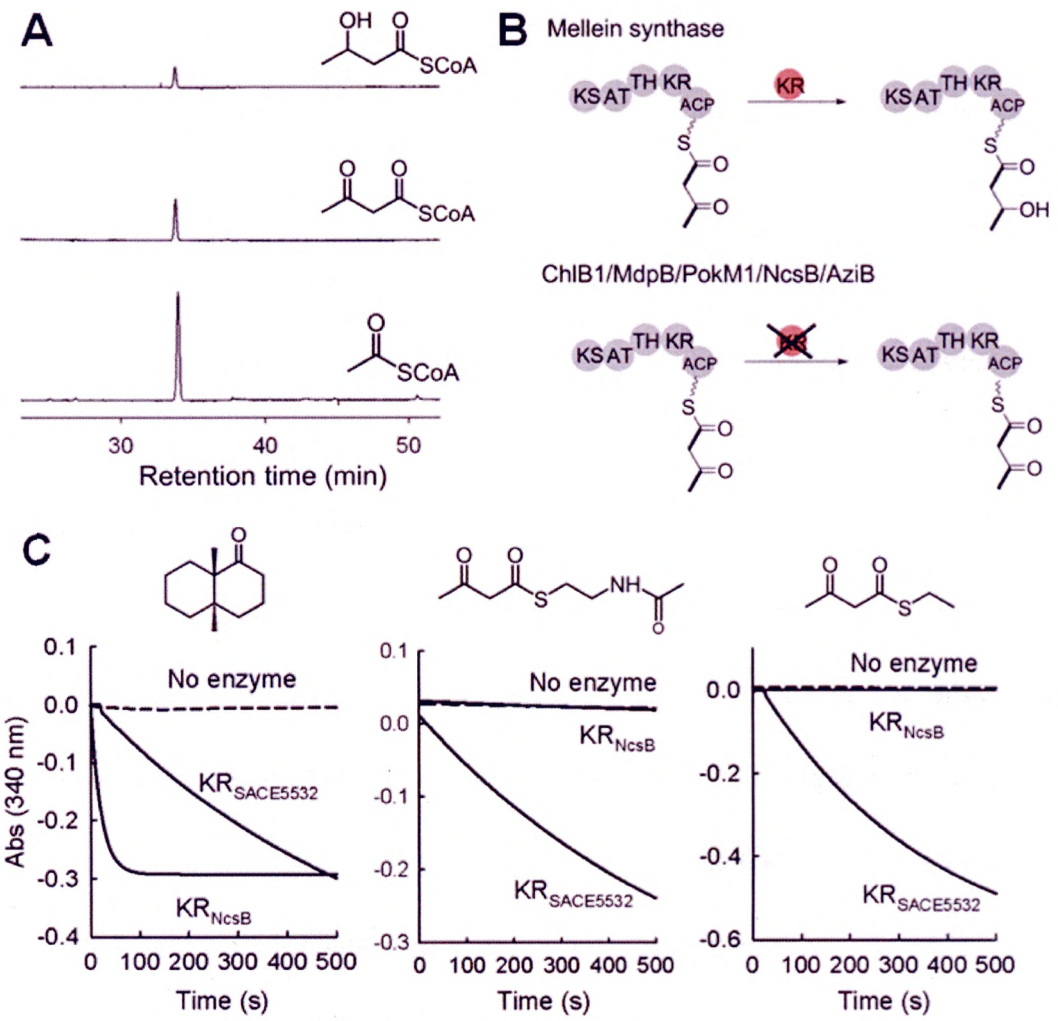


Figure 4.10. (A) HPLC analysis of the production of mellein by SACE5532 with different starter units. (B) Comparison of the reduction of the diketide intermediate by SACE5532 and other iterative PKSs. (C) Enzymatic activity of the stand-alone KR domains towards *trans*-1-decalone and two diketide analogs.

4.4 Discussion

Mellein and its derivatives are isocoumarin natural products widely distributed in bacteria, fungi and higher organisms [126-130]. The isolation of mellein as a trail pheromone from ants raised strong interest in the biological activity of mellein [131]. Many of the naturally occurring mellein or mellein derivatives exhibit a variety of biological activities, such as fungicidal and antibacterial activities [130,132]. The (-)-mellein isolated from the fungus *Aspergillus ochraceus* was reported as a potent inhibitor of the hepatitis C virus (HCV) NS3 protease [127]. Despite the fact that the naturally occurring mellein and related dihydroisocoumarins have been known for many decades, the enzymes responsible for the biosynthesis of these compounds remain to be identified. The unknown iPKS SACE5532 from the soil bacterium *Saccharopolyspora erythraea* shares sequence homology with known bacterial iPKSs for aromatic polyketide biosynthesis. My study of SACE5532 revealed that it synthesizes (R)-mellein (**5**), a product that is different from those of other known bacterial or fugnal iPKSs. Our results also imply that other dihydroisocoumarins are likely to share a polyketide origin.

The biosynthetic mechanism of mellein by SACE5532 is proposed in **Figure 4.11** with the assumption that the protein contains a TH domain, rather than a DH domain. To activate the PKS, SACE5532 needs a phosphopantetheinyl transferase (PPTase) to transfer the 4'-phosphopantetheine (4'-PP) component of CoA onto ACP domain. The starter acetate unit originated from acetyl CoA is loaded onto the ACP domain as catalyzed by the AT domain. Then the KS, AT, KR and ACP domains will be used

iteratively during chain elongation through the following steps: (a) The polyketide chain is handed over from the ACP domain to the KS domain, catalyzed by the KS domain; (b) The extender unit, originated from malonyl-CoA, is loaded onto the ACP domain as catalyzed by the AT domain. (c) The ACP-attached malonyl group undergoes decarboxylative Claisen condensation with the KS-tethered intermediate to elongate the chain. (d) Remarkably, the KR domain selectively reduces the ketone group to the hydroxyl group only on the first and third cycles. Chain extension is repeated four times to generate a pentaketide intermediate covalently tethered to the ACP domain. The hydroxyl group could be retained until the final cyclization step and then eliminated to drive the aromatization [115]. An aldol cyclization takes palces to furnish the aromatic structure, eliminating the hydroxyl group at C5 position. Finally, the TH domain hydrolyzes the thioester bond between the aromatic pentaketide intermediate and the phosphopantetheinyl arm of ACP to release the product **5**.

In the light of the synthesis of mellein by SACE5532, other mellein derivatives could be enzymatically synthesized through rational enzyme engineering of the AT, KS or KR domains. Our preliminary assays implied that SACE5532 is able to use acetoacetyl-CoA and β-hydroxybutyryl-CoA as starter units but in a lower catalytic rate (**Figure 4.10A**), indicating that the AT domain could be further engineered to expand its substrate specificity.

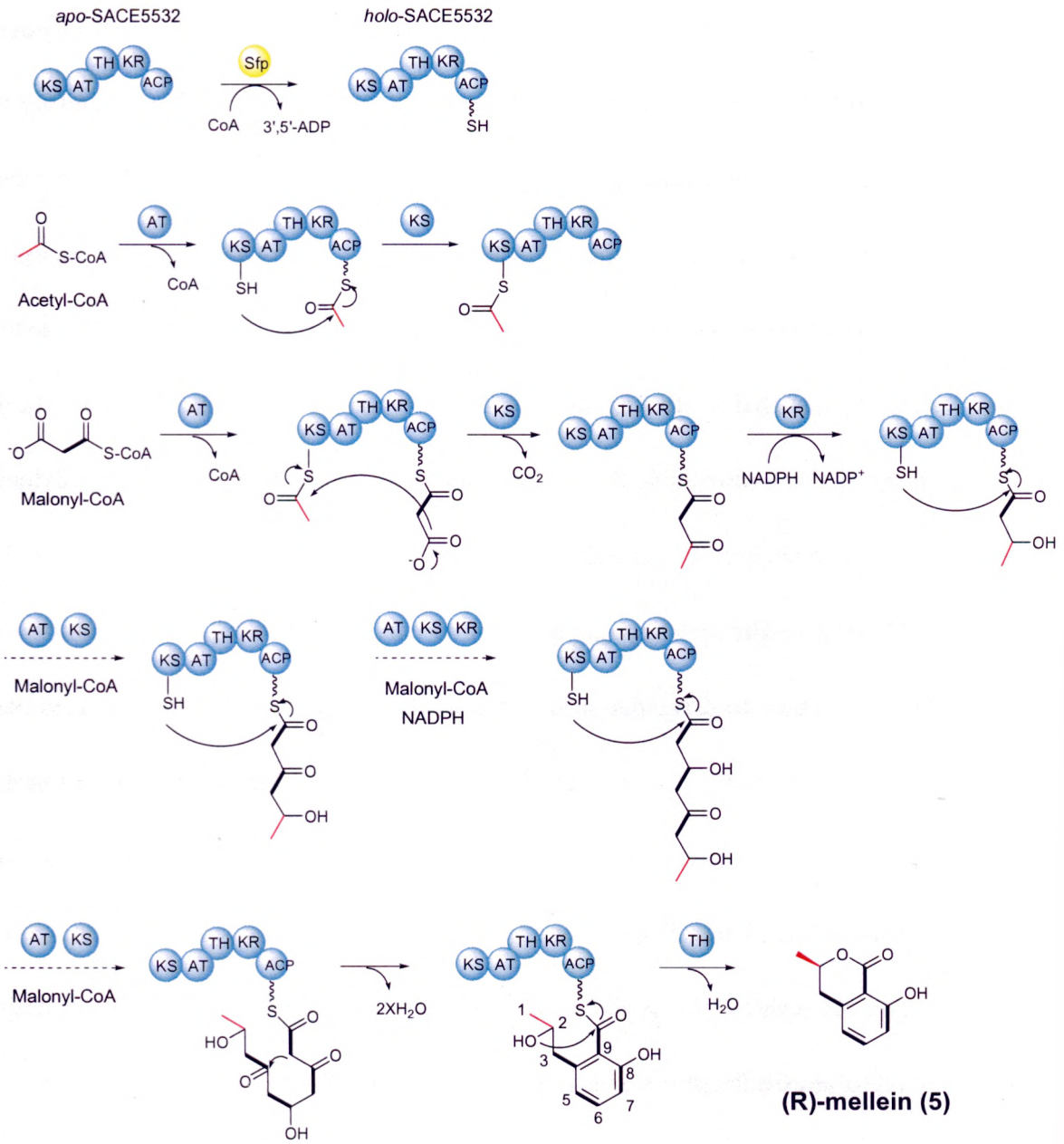


Figure 4.11. Proposed biosynthetic mechanism of **5** by SACE5532.

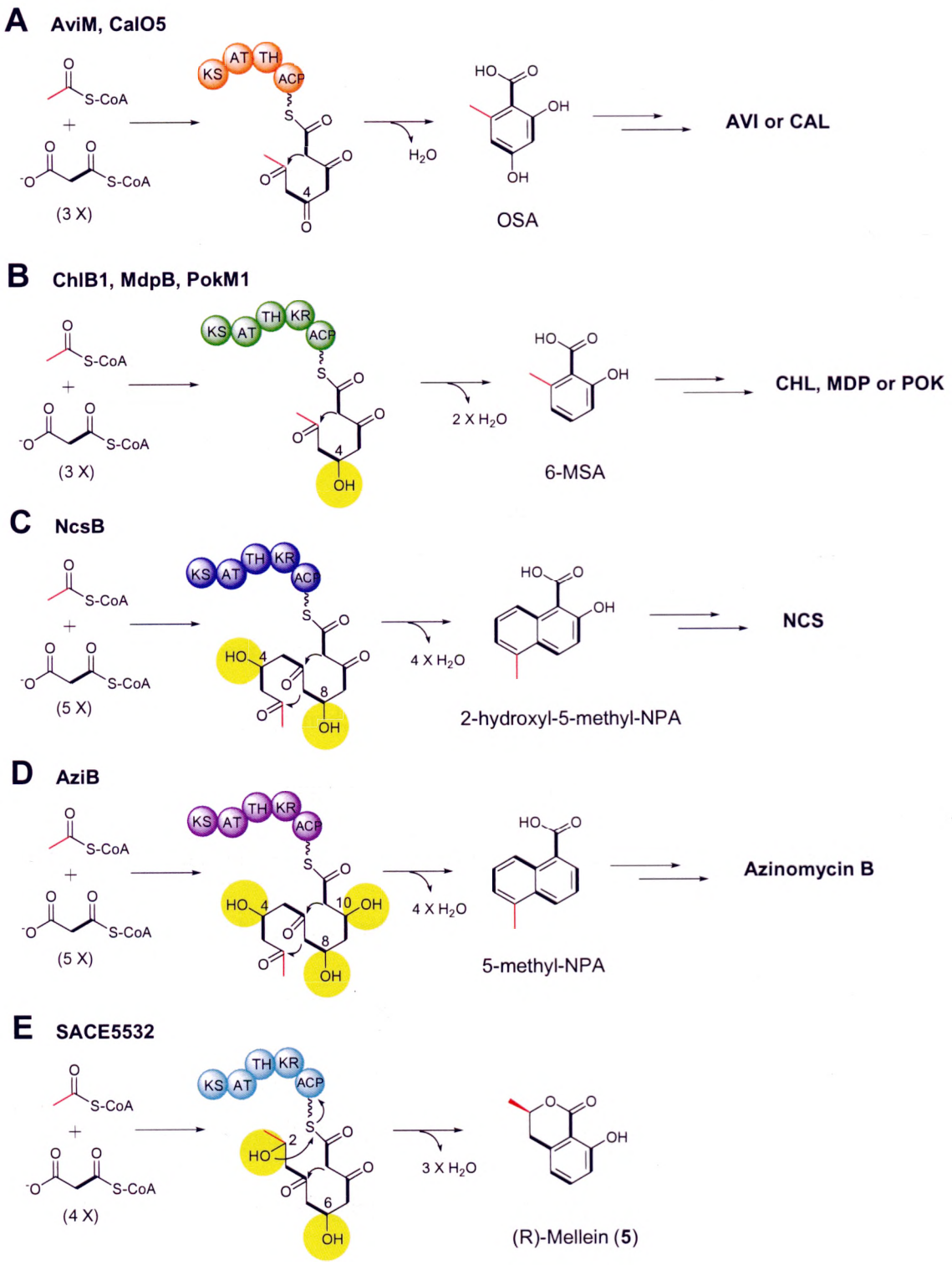


Figure 4.12. Bacterial iPKSs for aromatic polyketide biosynthesis. Reduction patterns are highlighted in yellow. (A) AviM and CalO5 for orsellinic acid (OSA) biosynthesis. (B) ChlB1, MdpB and PokM1 for 6-methylsalicyclic acid (6-MSA) biosynthesis. (C) NcsB for 2-hydroxyl-5-methyl-naphthoic acid (NPA) biosynthesis. (D) AziB for 5-methyl-NPA biosynthesis. (E) SACE5532 for (R)-mellein (**5**) biosynthesis.

SACE5532 shares head-to-tail sequence homology with the known bacterial iPKSs, including ChlB1, MdpB, PokM1, NcsB and AziB. However, their products differ in chain length and reduction pattern (**Figure 4.12**). A pentaketide is produced by SACE5532, which is apparently different from the tetraketides generated by ChlB1, MdpB and PokM1, and the hexaketides generated by NcsB and AziB. The KR domain of SACE5532 selectively reduces the keto groups at C2 and C6 positions, which is also different from the reduction patterns of the KR domains of ChlB1, MdpB and PokM1 (reduction at C4 position), the KR domain of NcsB (reduction at C4 and C10 positions), and the KR domain of AziB (reduction at C4, C8 and C10 positions). How iPKSs achieve the programmed polyketide in an iterative manner is one the most interesting questions in enzymology and biochemistry. Cox's group studied the programming in fungal iterative HR-PKSs by domain swapping and suggested that the KR domain plays a key role in programming. They also showed that the AT, KS, and ACP domains are the basic units for chain elongation and the three domains play no role in programming [48].

SACE5532 shares the highest sequence homology to NcsB, but their products are strikingly different in the reduction pattern and chain length. We previously speculated that the ACP domain plays a role in controlling chain length, assuming that ACP domain may dock differently onto KS or KR domain when the ACP-tethered intermediate reaches the final length. This mechanism is apparently not operative because SACE5532 with the ACP domain replaced by the ACP of NcsB still produces a pentaketide but not a hexaketide product (**Fig. 4.8A**). Interestingly, TH-KR

di-domain swapping between SACE5532 and NcsB indicates that TH-KR di-domain plays a critical role in dictating the keto-reduciton pattern of the iterative PKSs. Comparison of the stand-alone KR domains of SACE5532 and NcsB further reveals that the KR domain alone is able to recognize and differentiate the polyketide intermediates, which provides a mechanistic explanation for the programmed keto-reduction in the iterative PKSs.

And lastly, we suggest that the final chain length of the polyketide product could be controlled by the KS domain by recognizing the length of the growing polyketide. In the case of iterative type II [133] and type III PKSs [134] , crystal structures of the KS components revealed the presence of substrate-binding tunnels within the proteins, which have been suggested to control the length of the polyketide. Recently, Khosla and co-workers reported the crystal structures of the KS-AT fragment of module 3 [135] of the multi-modular PKS 6-deoxyerthronolide B synthase (DEBS). A detailed comparison between the active sites of DEBS KS3 and KS5 [136] revealed a set of residues that are thought to control specificity for their respective triketide and pentaketide substrates. In the future, we would like to examine the KS domain of SACE5532 to identify critical residues or motifs that control the substrate specificity through structure determination and mutagenesis. If we can identify the structural determining factors of the chain length and programmed reduction, it will open the door for the generation of novel dihydroisocoumarins by protein engineering.

In summary, the results established the polyketide origin of mellein and discovered that mellein is synthesized by a bacterial partially reducing iterative PKS

(PR-PKS). Swapping domains between SACE5532 and another iPKS NcsB revealed that the ACP domain does not affect product formation, while TH-KR di-domain plays a role in controlling the keto-reduction pattern. The substrate specificity exhibited by the stand-alone KR domains towards the diketide analogs strongly suggests that the KR domain alone is able to differentiate and selectively reduce the polyketide intermediates, which sheds light on the mechanism of the programmed keto-reduction for the iterative PKSs.

References

1. Staunton J, Weissman KJ: **Polyketide biosynthesis: A millennium review.** *Nat Prod Rep* (2001) **18**(4):380-416.

2. Hopwood DA: **Genetic contributions to understanding polyketide synthases.** *Chem Rev* (1997) **97**(7):2465-2498.

3. Hopwood DA, Sherman DH: **Molecular genetics of polyketides and its comparison to fatty acid biosynthesis.** *Annu Rev Genet* (1990) **24**(37-66.

4. Katz L, Donadio S: **Polyketide synthesis: Prospects for hybrid antibiotics.** *Annu Rev Microbiol* (1993) **47**(875-912.

5. Birch AJ, Donovan FW, Moewus F: **Biogenesis of flavonoids in chlamydomonas eugametos.** *Nature* (1953) **172**(4385):902-904.

6. Maier T, Leibundgut M, Ban N: **The crystal structure of a mammalian fatty acid synthase.** *Science* (2008) **321**(5894):1315-1322.

7. Smith S, Tsai SC: **The type i fatty acid and polyketide synthases: A tale of two megasynthases.** *Nat Prod Rep* (2007) **24**(5):1041-1072.

8. Das A, Khosla C: **Biosynthesis of aromatic polyketides in bacteria.** *Acc Chem Res* (2009) **42**(5):631-639.

9. Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A: **Type ii polyketide synthases: Gaining a deeper insight into enzymatic teamwork.** *Nat Prod Rep* (2007) **24**(1):162-190.

10. Zhang W, Tang Y: **In vitro analysis of type ii polyketide synthase.** *Methods Enzymol* (2009) **459**(367-393.

11. Austin MB, Izumikawa M, Bowman ME, Udwary DW, Ferrer JL, Moore BS, Noel JP: **Crystal structure of a bacterial type iii polyketide synthase and enzymatic control of reactive polyketide intermediates.** *J Biol Chem* (2004) **279**(43):45162-45174.

12. Austin MB, Noel JP: **The chalcone synthase superfamily of type iii polyketide synthases.** *Nat Prod Rep* (2003) **20**(1):79-110.

13. Funa N, Awakawa T, Horinouchi S: **Pentaketide resorcylic acid synthesis by type iii polyketide synthase from neurospora crassa.** *J Biol Chem* (2007)

282(19):14476-14481.

14. Fujii I: **Heterologous expression systems for polyketide synthases.** *Nat Prod Rep* (2009) **26**(2):155-169.

15. Shen B: **Polyketide biosynthesis beyond the type i, ii and iii polyketide synthase paradigms.** *Curr Opin Chem Biol* (2003) **7**(2):285-295.

16. Weissman KJ: **Introduction to polyketide biosynthesis.** *Methods Enzymol* (2009) **459**(3-16.

17. Gokhale RS, Sankaranarayanan R, Mohanty D: **Versatility of polyketide synthases in generating metabolic diversity.** *Curr Opin Struct Biol* (2007) **17**(6):736-743.

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

19. Crawford JM, Thomas PM, Scheerer JR, Vagstad AL, Kelleher NL, Townsend CA: **Deconstruction of iterative multidomain polyketide synthase function.** *Science* (2008) **320**(5873):243-246.

20. Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E: **The multifunctional 6-methylsalicylic acid synthase gene of penicillium patulum. Its gene structure relative to that of other polyketide synthases.** *European Journal of Biochemistry* (1990) **192**(2):487-498.

21. Campbell CD, Vederas JC: **Biosynthesis of lovastatin and related metabolites formed by fungal iterative pks enzymes.** *Biopolymers* (2010) **93**(9):755-763.

22. Crawford JM, Dancy BC, Hill EA, Udwary DW, Townsend CA: **Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type i polyketide synthase.** *Proc Natl Acad Sci U S A* (2006) **103**(45):16728-16733.

23. Crawford JM, Korman TP, Labonte JW, Vagstad AL, Hill EA, Kamari-Bidkorpeh O, Tsai SC, Townsend CA: **Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization.** *Nature* (2009) **461**(7267):1139-1143.

24. Fujii I, Ono Y, Tada H, Gomi K, Ebizuka Y, Sankawa U: **Cloning of the polyketide synthase gene atx from aspergillus terreus and its identification**

as the 6-methylsalicylic acid synthase gene by heterologous expression. *Molecular & General Genetics* (1996) **253**(1-2):1-10.

25. Lu P, Zhang A, Dennis LM, Dahl-Roshak AM, Xia YQ, Arison B, An Z, Tkacz JS: **A gene (pks2) encoding a putative 6-methylsalicylic acid synthase from glarea lozoyensis.** *Mol Genet Genomics* (2005) **273**(2):207-216.

26. Crawford JM, Townsend CA: **New insights into the formation of fungal aromatic polyketides.** *Nat Rev Microbiol* (2010) **8**(12):879-889.

27. Sutherland A, Auclair K, Vederas JC: **Recent advances in the biosynthetic studies of lovastatin.** *Curr Opin Drug Discov Devel* (2001) **4**(2):229-236.

28. Walsh CT: **Polyketide and nonribosomal peptide antibiotics: Modularity and versatility.** *Science* (2004) **303**(5665):1805-1810.

29. Wenzel SC, Muller R: **Formation of novel secondary metabolites by bacterial multimodular assembly lines: Deviations from textbook biosynthetic logic.** *Curr Opin Chem Biol* (2005) **9**(5):447-458.

30. Gaisser S, Trefzer A, Stockert S, Kirschning A, Bechthold A: **Cloning of an avilamycin biosynthetic gene cluster from streptomyces viridochromogenes tu57.** *J Bacteriol* (1997) **179**(20):6271-6278.

31. Sthapit B, Oh T-J, Lamichhane R, Liou K, Lee HC, Kim C-G, Sohng JK: **Neocarzinostatin naphthoate synthase: An unique iterative type i pks from neocarzinostatin producer streptomyces carzinostaticus.** *FEBS Letters* (2004) **566**(1-3):201-206.

32. Daum M, Peintner I, Linnenbrink A, Frerich A, Weber M, Paululat T, Bechthold A: **Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin.** *Chembiochem* (2009) **10**(6):1073-1083.

33. Hang VTT, Oh TJ, Yamaguchi T, Sohng JK: **In vivo characterization of ncsb3 to establish the complete biosynthesis of the naphthoic acid moiety of the neocarzinostatin chromophore.** *FEMS Microbiology Letters* (2010) **311**(2):119-125.

34. Jia X-Y, Tian Z-H, Shao L, Qu X-D, Zhao Q-F, Tang J, Tang G-L, Liu W: **Genetic characterization of the chlorothricin gene cluster as a model for spirotetronate antibiotic biosynthesis.** *Chem Biol* (2006) **13**(6):575-585.

35. Liu W, Nonaka K, Nie L, Zhang J, Christenson SD, Bae J, Van Lanen SG,

Zazopoulos E, Farnet CM, Yang CF, Shen B: **The neocarzinostatin biosynthetic gene cluster from streptomyces carzinostaticus atcc 15944 involving two iterative type i polyketide synthases.** *Chem Biol* (2005) **12**(3):293-302.

36. Penn K, Jenkins C, Nett M, Udwary DW, Gontang EA, McGlinchey RP, Foster B, Lapidus A, Podell S, Allen EE, Moore BS *et al*: **Genomic islands link secondary metabolism to functional adaptation in marine actinobacteria.** *Isme J* (2009) **3**(10):1193-1203.

37. Shao L, Qu XD, Jia XY, Zhao QF, Tian ZH, Wang M, Tang GL, Liu W: **Cloning and characterization of a bacterial iterative type i polyketide synthase gene encoding the 6-methylsalicyclic acid synthase.** *Biochem Biophys Res Commun* (2006) **345**(1):133-139.

38. Van Lanen SG, Oh T-j, Liu W, Wendt-Pienkowski E, Shen B: **Characterization of the maduropeptin biosynthetic gene cluster from actinomadura madurae atcc 39144 supporting a unifying paradigm for enediyne biosynthesis.** *J Am Chem Soc* (2007) **129**(43):13082-13094.

39. Zhao Q, He Q, Ding W, Tang M, Kang Q, Yu Y, Deng W, Zhang Q, Fang J, Tang G, Liu W: **Characterization of the azinomycin b biosynthetic gene cluster revealing a different iterative type i polyketide synthase for naphthoate biosynthesis.** *Chem Biol* (2008) **15**(7):693-705.

40. Ito T, Roongsawang N, Shirasaka N, Lu W, Flatt PM, Kasanah N, Miranda C, Mahmud T: **Deciphering pactamycin biosynthesis and engineered production of new pactamycin analogues.** *Chembiochem* (2009) **10**(13):2253-2265.

41. Liu W, Christenson SD, Standage S, Shen B: **Biosynthesis of the enediyne antitumor antibiotic c-1027.** *Science* (2002) **297**(1170-1173.

42. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM *et al*: **The calicheamicin gene cluster and its iterative type i enediyne pks.** *Science* (2002) **297**(5584):1173-1176.

43. Gao Q, Thorson JS: **The biosynthetic genes encoding for the production of the dynemicin enediyne core in *micromonospora chersina* atcc53710.** *FEMS Microbiology Letters* (2008) **282**(1):105-114.

44. Zhang J, Van Lanen SG, Ju J, Liu W, Dorrestein PC, Li W, Kelleher NL, Shen B: **A phosphopantetheinylating polyketide synthase producing a linear**

polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proc Natl Acad Sci U S A* (2008) **105**(5):1460-1465.

45. Liu T, Chiang YM, Somoza AD, Oakley BR, Wang CC: **Engineering of an "unnatural" natural product by swapping polyketide synthase domains in aspergillus nidulans.** *J Am Chem Soc* (2011) **133**(34):13314-13316.

46. Zhu X, Yu F, Bojja RS, Zaleta-Rivera K, Du L: **Functional replacement of the ketosynthase domain of fum1 for the biosynthesis of fumonisins, a group of fungal reduced polyketides.** *J Ind Microbiol Biotechnol* (2006) **33**(10):859-868.

47. Zhu X, Yu F, Li X-C, Du L: **Production of dihydroisocoumarins in fusarium verticillioides by swapping ketosynthase domain of the fungal iterative polyketide synthase fum1p with that of lovastatin diketide synthase.** *J Am Chem Soc* (2006) **129**(1):36-37.

48. Fisch KM, Bakeer W, Yakasai AA, Song Z, Pedrick J, Wasil Z, Bailey AM, Lazarus CM, Simpson TJ, Cox RJ: **Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite.** *J Am Chem Soc* (2011) **133**(41):16635-16641.

49. Nicolaou KC, Dai WM: **Chemistry and biology of the enediyne anticancer antibiotics.** *Angewandte Chemie International Edition in English* (1991) **30**(11):1387-1416.

50. Nicolaou KC, Smith AL, Yue EW: **Chemistry and biology of natural and designed enediynes.** *Proc Natl Acad Sci USA* (1993) **90**(5881-5888.

51. Doyle TW, Border DB (Eds): *Enediyne antibiotics as antitumor agents*. Marcel-Dekker, New York (1995).

52. Smith AL, Nicolaou KC: **The enediyne antibiotics.** *Journal of Medicinal Chemistry* (1996) **39**(11):2103-2117.

53. Maeda H, Edo K, Ishida N (Eds): *Neocarzinostatin: The past, present, and future of an anticancer drug*. Springer- Verlag, New York (1997).

54. Thorson JS, Sievers EL, Ahlert J, Shepard E, Whitwam RE, Onwueme KC, Ruppen M: **Understanding and exploiting nature's chemical arsenal: The past, present and future of calicheamicin research.** *Current Pharmaceutical Design* (2000) **6**(18):1841-1879.

55. Jones GB, Fouad FS: **Designed enediyne antitumor agents.** *Current Pharmaceutical Design* (2002) **8**(27):2415-2440.

56. Wang X, Xie H: **C-1027.** *Drugs of the Future* (1999) **24**(847-852.

57. Sievers EL, Appelbaum FR, Spielberger RT, Forman SJ, Flowers D, Smith FO, Shannon-Dorcy K, Berger MS, Bernstein ID: **Selective ablation of acute myeloid leukemia using antibody-targeted chemotherapy: A phase i study of an anti-cd33 calicheamicin immunoconjugate.** *Blood* (1999) **93**(11):3678-3684.

58. Brukner I: **C-1027 taiho pharmaceutical co. Ltd.** *Curr Opinion Oncologic, Endocrine & Met Invest Drugs* (2000) **2**(344-352.

59. Stassinopoulos A, Ji J, Gao X, Goldberg IH: **Solution structure of a two-base DNA bulge complexed with an enediyne cleaving analog.** *Science* (1996) **272**(5270):1943-1946.

60. N Ikemoto, R A Kumar, T T Ling, G A Ellestad, S J Danishefsky, Patel DJ: **Calicheamicin-DNA complexes: Warhead alignment and saccharide recognition of the minor groove.** *PNAS* (1995) **92**(10506-10510.

61. Dedon PC, Goldberg IH: **Free-radical mechanisms involved in the formation of sequence-dependent bistranded DNA lesions by the antitumor antibiotics bleomycin, neocarzinostatin, and calicheamicin.** *Chemical Research in Toxicology* (1992) **5**(3):311-332.

62. Bergman RG: **Reactive 1,4-dehydroaromatics.** *Acc Chem Res* (1973) **6**(1):25-31.

63. Edo K, Mizugaki M, Koide Y, Seto H, Furihata K, Otake N, Ishida N: **The structure of neocarzinostatin chromophore possessing a novel bicyclo-[7,3,0]dodecadiyne system.** *Tetrahedron Letters* (1985) **26**(3):331-334.

64. Konishi M, Ohkuma H, Saitoh K-I, Kawaguchi H, Golik J, Dubay G, Groenewold G, Krishnan B, Doyle TW: **Esperamicins, a novel class of potent antitumor antibiotics. I. Physico-chemical data and partial structure.** *Journal of Antibiotics* (1985) **38**(11):1605-1609.

65. Lee MD, Dunne TS, Siegel MM, Chang CC, Morton GO, Borders DB: **Calichemicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calichemicin .Gamma.1i.** *J Am Chem Soc* (1987) **109**(11):3464-3466.

66. Konishi M, Ohkuma H, Tsuno T, Oki T, VanDuyne GD, Clardy J: **Crystal and molecular structure of dynemicin a: A novel 1,5-diyn-3-ene antitumor antibiotic.** *J Am Chem Soc* (1990) **112**(9):3715-3716.

67. Otani T, Minami Y, Sakawa K, Yoshida K-I: **Isolation and characterization of non-protein chromophore and its degradation product from antibiotic c-1027.** *Journal of Antibiotics* (1991) **44**(5):564-568.

68. Schroeder DR, Colson KL, Klohr SE, Zein N, Langley DR, Lee MS, Matson JA, Doyle TW: **Isolation, structure determination, and proposed mechanism of action for artifacts of maduropeptin chromophore.** *J Am Chem Soc* (1994) **116**(20):9351-9352.

69. Kawata S, Ashizawa S, Hirama M: **Synthetic study of kedarcidin chromophore: Revised structure.** *J Am Chem Soc* (1997) **119**(49):12012-12013.

70. Ando T, Ishii M, Kajiura T, Kameyama T, Miwa K, Sugiura Y: **A new non-protein enediyne antibiotic n1999a2: Unique enediyne chromophore similar to neocarzinostatin and DNA cleavage feature.** *Tetrahedron Letters* (1998) **39**(36):6495-6498.

71. Oku N, Matsunaga S, Fusetani N: **Shishijimicins a-c, novel enediyne antitumor antibiotics from the ascidian *didemnum proliferum*.** *J Am Chem Soc* (2003) **125**(8):2044-2045.

72. Davies J, Wang H, Taylor T, Warabi K, Huang X-H, Andersen RJ: **Uncialamycin, a new enediyne antibiotic.** *Organic Letters* (2005) **7**(23):5233-5236.

73. McDonald LA, Capson TL, Krishnamurthy G, Ding W-D, Ellestad GA, Bernan VS, Maiese WM, Lassota P, Kramer RA, Ireland CM: **Namenamicin, a new enediyne antitumor antibiotic from the marine ascidian polysyncraton lithostrotum.** *J Am Chem Soc* (1996) **118**(44):10898-10899.

74. Chimura H, Ishizuka M, Hamada M, Hori S, Kimura K: **A new antibiotic, macromomycin, exhibiting antitumor and antimicrobial activity.** *J Antibiot (Tokyo)* (1968) **21**(1):44-49.

75. Yamaguchi T, Furumai T, Sato M, Okuda T, Ishida N: **Studies on a new antitumor antibiotic, largomycin. I. Taxonomy of the largomycin-producing strain and production of the antibiotic.** *J Antibiot (Tokyo)* (1970) **23**(8):369-372.

76. Khokhlov AS, Reshetov PD, Chupova LA, Cherches BZ, Zhigis LS, Stoyachemko IA: **Chemical studies on actinoxanthin.** *J Antibiot (Tokyo)* (1976) **29**(10):1026-1034.

77. Komiyama K, Umezawa I: **Tissue pharmacokinetics and inhibition of DNA synthesis in mice treated with sporamycin.** *J Antibiot (Tokyo)* (1978) **31**(5):473-476.

78. Hu J, Xue Y-C, Xie M-Y, Zhang R, Otani T, Minami Y, Yamada Y, Marunaka T: **A new macromolecular antitumor antibiotic, c-1027. I. Discovery, taxonomy of producing organism, fermentation and biological activity.** *Journal of Antibiotics* (1988) **41**(11):1575-1579.

79. Otani T, Minami Y, Marunaka T, Zhang R, Xie M-Y: **A new macromolecular antitumor antibiotic, c-1027. Ii. Isolation and physico-chemical properties.** *Journal of Antibiotics* (1988) **41**(11):1580-1585.

80. Otani T: **Conformation studies on and assessment by spectral analysis of the protein-chromophore interaction of the macromolecular antitumor antibiotic c-1027.** *Journal of Antibiotics* (1993) **46**(5):791-802.

81. Matsumoto T, Okuno Y, Sugiura Y: **Specific interaction between a novel enediyne chromophore and apoprotein in macromolecular antitumor antibiotic c-1027.** *Biochem Biophys Res Commun* (1993) **195**(2):659-666.

82. Thorson JS, Shen B, Whitwam RE, Liu W, Li Y, Ahlert J: **Enediyne biosynthesis and self-resistance: A progress report.** *Bioorganic Chemistry* (1999) **27**(2):172-188.

83. Shen B, Liu W, Nonaka K: **Enediyne natural products: Biosynthesis and prospect towards engineering novel antitumor agents.** *Current Medicinal Chemistry* (2003) **10**(2317-2325.

84. Hensens OD, Giner JL, Goldberg IH: **Biosynthesis of ncs chrom a, the chromophore of the antitumor antibiotic neocarzinostatin.** *J Am Chem Soc* (1989) **111**(9):3295-3299.

85. Tokiwa Y, Miyoshi-Saitoh M, Kobayashi H, Sunaga R, Konishi M, Oki T, Iwasaki S: **Biosynthesis of dynemicin a, a 3-ene-1,5-diyne antitumor antibiotic.** *J Am Chem Soc* (1992) **114**(11):4107-4110.

86. Lam KS, Veitch JA, Golik J, Krishnan B, Klohr SE, Volk KJ, Forenza S, Doyle TW: **Biosynthesis of esperamicin a1, an enediyne antitumor antibiotic.** *J Am Chem Soc* (1993) **115**(26):12340-12345.

87. Zazopoulos E: **A genomics-guided approach for discovering and expressing cryptic metabolic pathways.** *Nat Biotechnol* (2003) **21**(2):187-190.

88. Thoden JB, Holden HM, Zhuang Z, Dunaway-Mariano D: **X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-coa thioesterase.** *J Biol Chem* (2002) **277**(30):27468-27476.

89. Thoden JB, Zhuang Z, Dunaway-Mariano D, Holden HM: **The structure of 4-hydroxybenzoyl-coa thioesterase from *arthrobacter* sp. Strain su.** *J Biol Chem* (2003) **278**(44):43709-43716.

90. Kotaka M, Kong R, Qureshi I, Ho QS, Sun H, Liew CW, Goh LP, Cheung P, Mu Y, Lescar J, Liang Z-X: **Structure and catalytic mechanism of the thioesterase cale7 in enediyne biosynthesis.** *J Biol Chem* (2009) M809669200.

91. Lee M: **Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation.** *Science* (1998) **280**(5365):915-918.

92. Fox BG, Shanklin J, Somerville C, Münck E: **Stearoyl-acyl carrier protein δ^9 desaturase from *ricinus communis* is a diiron-oxo protein.** *Proc Natl Acad Sci U S A* (1993) **90**(6):2486-2490.

93. Lindqvist Y, Huang W, Schneider G, Shanklin J: **Crystal structure of δ^9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins.** *Embo J* (1996) **15**(16):4081–4092.

94. Dyer DH, Lyle KS, Rayment I, Fox BG: **X-ray structure of putative acyl-acp desaturase desa2 from mycobacterium tuberculosis h37rv.** *Protein Sci* (2005) **14**(6):1508-1517.

95. Guy JE, Whittle E, Kumaran D, Lindqvist Y, Shanklin J: **The crystal structure of the ivy delta4-16:0-acp desaturase reveals structural details of the oxidized active site and potential determinants of regioselectivity.** *J Biol Chem* (2007) **282**(27):19863-19871.

96. Chen X, Guo ZF, Lai PM, Sze KH, Guo Z: **Identification of a nonaketide product for the iterative polyketide synthase in biosynthesis of the nine-membered enediyne c-1027.** *Angew Chem Int Ed Engl* (2010) **49**(43):7926-7928.

97. Kong R, Goh LP, Liew CW, Ho QS, Murugan E, Li B, Tang K, Liang ZX: **Characterization of a carbonyl-conjugated polyene precursor in**

10-membered enediyne biosynthesis. *J Am Chem Soc* (2008) **130**(26):8142-8143.

98. Belecki K, Crawford JM, Townsend CA: **Production of octaketide polyenes by the calicheamicin polyketide synthase cale8: Implications for the biosynthesis of enediyne core structures.** *J Am Chem Soc* (2009) **131**(35):12564-12566.

99. Kim BS, Cropp TA, Beck BJ, Sherman DH, Reynolds KA: **Biochemical evidence for an editing role of thioesterase ii in the biosynthesis of the polyketide pikromycin.** *J Biol Chem* (2002) **277**(50):48028-48034.

100. Bisang C, Long PF, Cortes J, Westcott J, Crosby J, Matharu A-L, Cox RJ, Simpson TJ, Staunton J, Leadlay PF: **A chain initiation factor common to both modular and aromatic polyketide synthases.** *Nature* (1999) **401**(6752):502-505.

101. Kotaka M, Kong R, Qureshi I, Ho QS, Sun H, Liew CW, Goh LP, Cheung P, Mu Y, Lescar J, Liang ZX: **Structure and catalytic mechanism of the thioesterase cale7 in enediyne biosynthesis.** *J Biol Chem* (2009) **284**(23):15739-15749.

102. Guo ZF, Sun Y, Zheng S, Guo Z: **Preferential hydrolysis of aberrant intermediates by the type ii thioesterase in escherichia coli nonribosomal enterobactin synthesis: Substrate specificities and mutagenic studies on the active-site residues.** *Biochemistry* (2009) **48**(8):1712-1722.

103. Heathcote ML, Staunton J, Leadlay PF: **Role of type li thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units.** *Chem Biol* (2001) **8**(2):207-220.

104. Horsman GP, Chen YH, Thorson JS, Shen B: **Polyketide synthase chemistry does not direct biosynthetic divergence between 9-and 10-membered enediynes.** *Proc Natl Acad Sci U S A* (2010) **107**(25):11331-11335.

105. Blatz PE, Tompkins JA: **Thermodynamic studies of hydrogen bonding and proton transfer between weak acids and bases in nonaqueous solvents as a model for acid-base reactions in proteins.** *J Am Chem Soc* (1992) **114**(3951-3956.

106. Ley SV, Smith SC, Woodward PR: **Further reactions of tert-butyl 3-oxobutanthioate and tert-butyl 4-diethyl-phosphono-3-oxobutanthioate - carbonyl coupling reactions, amination, use in the preparation of 3-acyltetramic acids and application to the total synthesis of fuligorubin-a.**

Tetrahedron (1992) **48**(6):1145-1174.

107. Novak L, Gy. P, Kovacs P, Kolonits P, Szantay C: **Rearragement of allyl aryl ethers i. Reaction of hydroquinone with conjugated dien-ols and trien-ol.** *Tetrahedron* (1995) **51**(9367-9374.

108. Pini E, Bertacche V, Molinari F, Romano D, R. G: **Direct conversion of polyconjugated compounds into their corresponding carboxylic acids by acetobacter aceti.** *Tetrahedron* (2008) **64**(8638-8641.

109. Staunton J, Wilkinson B: **Biosynthesis of erythromycin and rapamycin.** *Chem Rev* (1997) **97**(7):2611-2630.

110. Boakes S, Oliynyk M, Cortes J, Bohm I, Rudd BA, Revill WP, Staunton J, Leadlay PF: **A new modular polyketide synthase in the erythromycin producer saccharopolyspora erythraea.** *J Mol Microbiol Biotechnol* (2004) **8**(2):73-80.

111. Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF: **Complete genome sequence of the erythromycin-producing bacterium saccharopolyspora erythraea nrrl23338.** *Nat Biotechnol* (2007) **25**(4):447-453.

112. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM *et al*: **The calicheamicin gene cluster and its iterative type i enediyne pks.** *Science* (2002) **297**(5584):1173-1176.

113. Weitnauer G, Muhlenweg A, Trefzer A, Hoffmeister D, Sussmuth RD, Jung G, Welzel K, Vente A, Girreser U, Bechthold A: **Biosynthesis of the orthosomycin antibiotic avilamycin a: Deductions from the molecular analysis of the avi biosynthetic gene cluster of streptomyces viridochromogenes tu57 and production of new antibiotics.** *Chem Biol* (2001) **8**(6):569-581.

114. Shen B: **Biosynthesis of aromatic polyketides biosynthesis.** In: 209. Leeper F, Vederas J (Eds), Springer Berlin / Heidelberg, (2000):1-51.

115. Moriguchi T, Kezuka Y, Nonaka T, Ebizuka Y, Fujii I: **Hidden function of catalytic domain in 6-methylsalicylic acid synthase for product release.** *J Biol Chem* (2010) **285**(20):15637-15643.

116. Gilbert IH, Ginty M, Oneill JA, Simpson TJ, Staunton J, Willis CL: **Synthesis of beta-keto and alpha,beta-unsaturated n-acetylcysteamine thioesters.**

Bioorg Med Chem Lett (1995) **5**(15):1587-1590.

117. Islam MS, Ishigami K, Watanabe H: **Synthesis of (-)-mellein, (+)-ramulosin, and related natural products.** *Tetrahedron* (2007) **63**(5):1074-1079.

118. Keatinge-Clay AT: **A tylosin ketoreductase reveals how chirality is determined in polyketides.** *Chem Biol* (2007) **14**(8):898-908.

119. Keatinge-Clay AT, Stroud RM: **The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases.** *Structure* (2006) **14**(4):737-748.

120. Caffrey P: **Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases.** *Chembiochem* (2003) **4**(7):654-657.

121. Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R: **A model of structure and catalysis for ketoreductase domains in modular polyketide synthases.** *Biochemistry* (2003) **42**(1):72-79.

122. Wu J, Zaleski TJ, Valenzano C, Khosla C, Cane DE: **Polyketide double bond biosynthesis. Mechanistic analysis of the dehydratase-containing module 2 of the picromycin/methymycin polyketide synthase.** *J Am Chem Soc* (2005) **127**(49):17393-17404.

123. O'Hare HM, Baerga-Ortiz A, Popovic B, Spencer JB, Leadlay PF: **High-throughput mutagenesis to evaluate models of stereochemical control in ketoreductase domains from the erythromycin polyketide synthase.** *Chem Biol* (2006) **13**(3):287-296.

124. Ostergaard LH, Kellenberger L, Cortes J, Roddis MP, Deacon M, Staunton J, Leadlay PF: **Stereochemistry of catalysis by the ketoreductase activity in the first extension module of the erythromycin polyketide synthase.** *Biochemistry* (2002) **41**(8):2719-2726.

125. Castonguay R, He W, Chen AY, Khosla C, Cane DE: **Stereospecificity of ketoreductase domains of the 6-deoxyerythronollde b synthase.** *J Am Chem Soc* (2007) **129**(44):13758-13769.

126. Cole RJ, Cox RH: *Handbook of toxic fungal metabolites.* Academic Press, New York (1981).

127. Dai J, Carte BK, Sidebottom PJ, Sek Yew AL, Ng S, Huang Y, Butler MS:

Circumdatin g, a new alkaloid from the fungus aspergillus ochraceus. *J Nat Prod* (2001) **64**(1):125-126.

128. Devys M, Barbier M, Bousquet JF, Kollmann A: **Isolation of the new (-)-(3r,4s)-4-hydroxymellein from the fungus septoria-nodorum berk.** *Z Naturforsch C* (1992) **47**(9-10):779-781.

129. Devys M, Bousquet JF, Kollmann A, Barbier M: **Dihydroisocoumarins and mycophenolic-acid of culture-medium of a phytopathogenic fungus, septoria-nodorum.** *Phytochemistry* (1980) **19**(10):2221-2222.

130. Krohn K, Bahramsari R, Florke U, Ludewig K, KlicheSpory C, Michel A, Aust HJ, Draeger S, Schulz B, Antus S: **Dihydroisocoumarins from fungi: Isolation, structure elucidation, circular dichroism and biological activity.** *Phytochemistry* (1997) **45**(2):313-320.

131. Brand JM, Fales HM, Sokoloski EA, MacConnell JG, Blum MS, Duffield RM: **Identification of mellein in the mandibular gland secretions of carpenter ants.** *Life Sci* (1973) **13**(3):201-211.

132. Holler U, Konig GM, Wright AD: **Three new metabolites from marine-derived fungi of the genera coniothyrium and microsphaeropsis.** *J Nat Prod* (1999) **62**(1):114-118.

133. Keatinge-Clay AT, Maltby DA, Medzihradszky KF, Khosla C, Stroud RM: **An antibiotic factory caught in action.** *Nat Struct Mol Biol* (2004) **11**(9):888-893.

134. Morita H, Kondo S, Kato R, Wanibuchi K, Noguchi H, Sugio S, Abe I, Kohno T: **Crystallization and preliminary crystallographic analysis of an octaketide-producing plant type iii polyketide synthase.** *Acta Crystallogr Sect F Struct Biol Cryst Commun* (2007) **63**(Pt 11):947-949.

135. Tang Y, Chen AY, Kim CY, Cane DE, Khosla C: **Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide b synthase.** *Chem Biol* (2007) **14**(8):931-943.

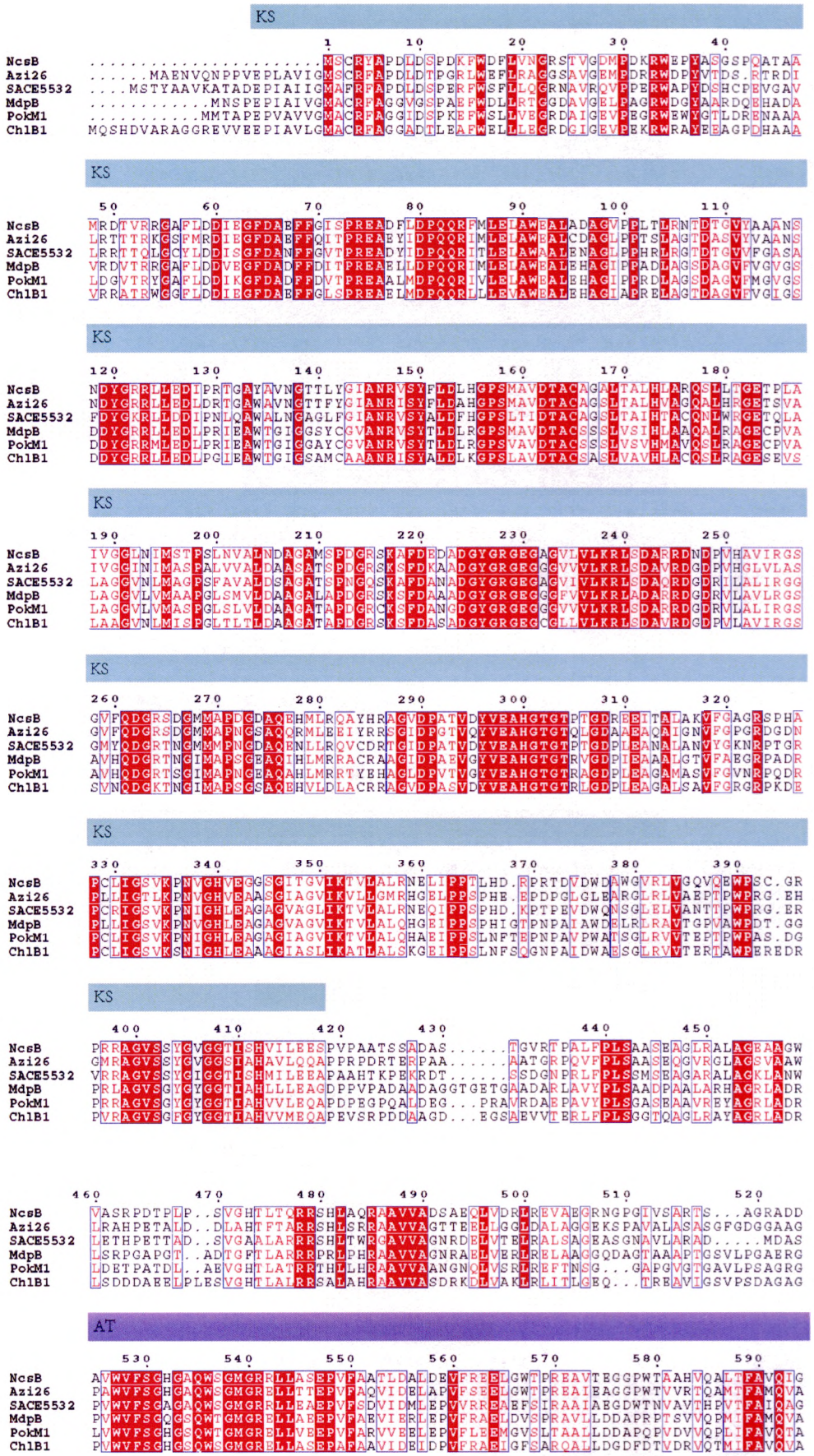
136. Tang Y, Kim CY, Mathews, II, Cane DE, Khosla C: **The 2.7-angstrom crystal structure of a 194-kda homodimeric fragment of the 6-deoxyerythronolide b synthase.** *Proc Natl Acad Sci U S A* (2006) **103**(30):11124-11129.

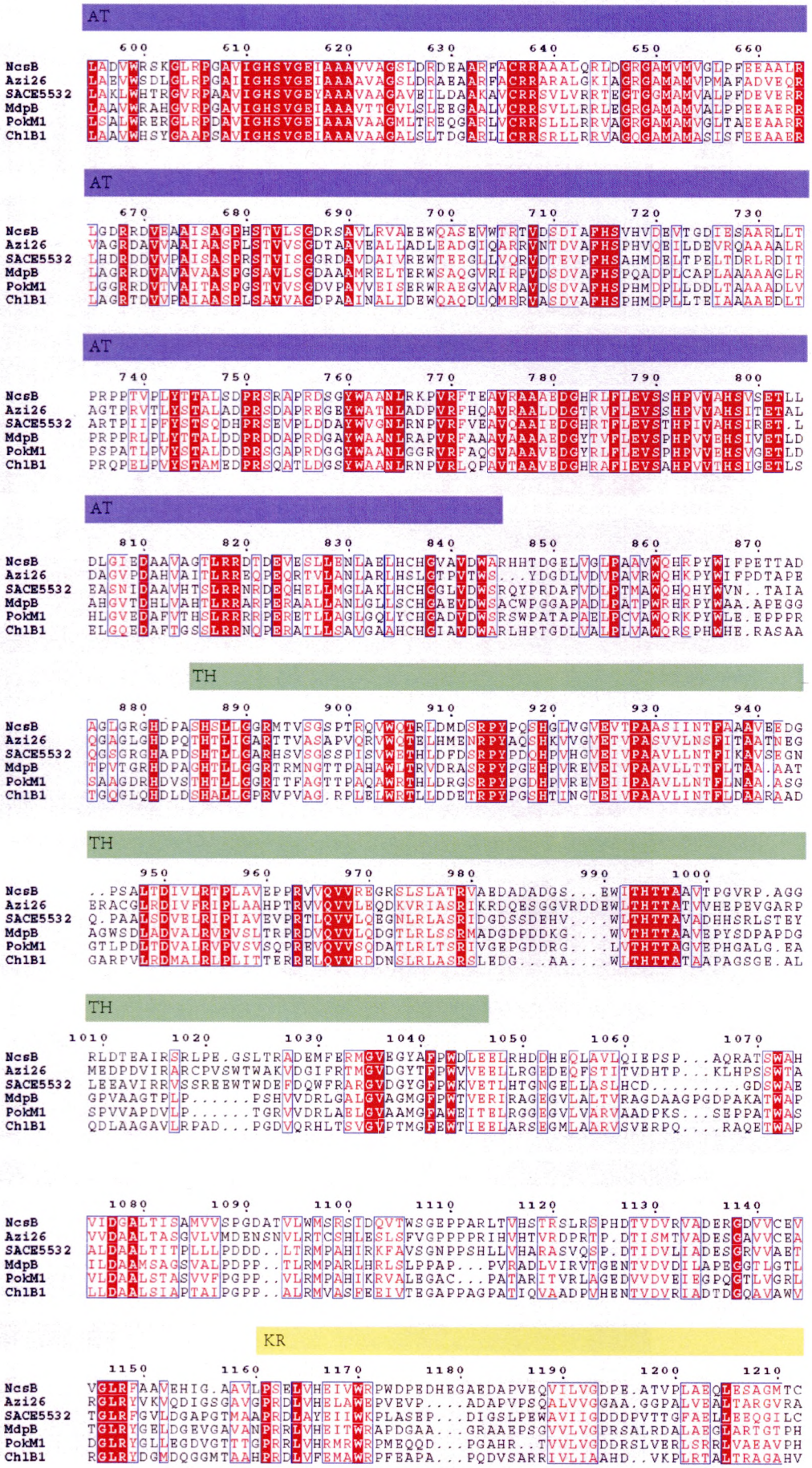
Appendix

SACE5532 protein sequence:

```
1   mstyaavkat adepiaiigm afrfapdlds perfwsfllq grnavrqvpp erwapydshc
61  pevgavlrrt tqlgcylddi sgfdanffgv tpreadyidp qqritlelaw aalenaglpp
121 hrlrgtdtgv vfgasafdyg krllddipnl qawalngagl fgianrvsya ldfhgpslti
181 dtacagslta ihtacqnlwr getqlalagg vnlmagpsfa valdsagats pngqskafda  KS
241 nadgygrgeg agvivlkrls daqrdgdril alirgggmyq dgrtngmmmp ngdaqenllr
301 qvcdrtgidp atvdyveahg tgtptgdple analanvygk nrptgrpcri gsvkpnighl
361 eagagvagli ktvlalrneq ippsphdkpt pevdwqnsgl elvanttpwp rgervrragv
421 ssygiggtis hmileeapaa htkpekrdts sdgnprlfpl ssmseagara lagklanwle
481 thpettadsv gaalarrrsh ltwrgavvag nrdelvtelr alsageasgn avlaradmda
541 spvwvfsgag aqwsgmgrrl leaepvfsdv idmlepvvrr eaefsiraai aegdwtnvav
601 thpvtfaiqa glaklwhtrg vrpaavighs vgeyaaavaa gaveildaak avcrrsvlvr
661 rtegtggmam valpfdever rlhdrddvvp aisasprstv isggrdavda ivrewteegl  AT
721 lvqrvdtevp fhsahmdelt peltdrlrdi tartpiipfy stsqdhprse vplddaywvg
781 nlrnpvrfve avqaaiedgh rtflevsthp ivahsiretl easnidaavh tslrrnrdeq
841 hellmglakl hchgglvdws rqyprdafvd lptmawqhqh ywvntaiaqg sgrghapdsr
901 tllgarhsvs gsspisvwet hldfdsrpyp dqhpvhgvei vpaavllntf ikavsegnqp  TH
961 aalsdvelri piavevprtl qvvlqegnlr lasridqdss dehvwlthtt avadhhsrls
1021 teyleeavir rvssreewtw defdqwfrar gvdgygfpwk vetlhtgnge llaslhcdgd
1081 swaealdaal titplllpdd dltrmpahir kfavsgnpps hllvharasv qspdtidvli
1141 adesgrvvae ttglrfgvld gapgtmaapr dlayeiiwkp lasepdigsl pewaviigdd
1201 dpvttgfael leeqgilcrr istpeavpep spgdsalivi apgvvthded pvdqaeratw
1261 sliratqrlp rtatenvrlw citrgvragn drtalahrpl wgaariisge rpdlwgglid
1321 vesgvrsqer llklvlgtpl dgediiavgd dgdyvermvp vtreparsel ecrpdgtylv
1381 tgglgaiglq varwlvgrga rrlvlagrng lpprhewdav tdattrtrie avrsleylga
1441 tvrvvaidia dpgqvadslt psalglppir gvvhaagvvn dalaenmdtd glartmspka  KR
1501 rgamalhqqf ppgtldffvl fssgqfart tgqvtyaaan sfldalaayr raagchetis
1561 vawmawlntg maasiessma earangidgf tpeeafqawq fvsrfdlpya vvlrsipipg
1621 nvprpavlse ltasgdtg                                          ACP
                                                                kakvsle
```

Multiple sequence alignment of SACE5532, NcsB, Azi26, MdpB, PokM1 and ChlB1:





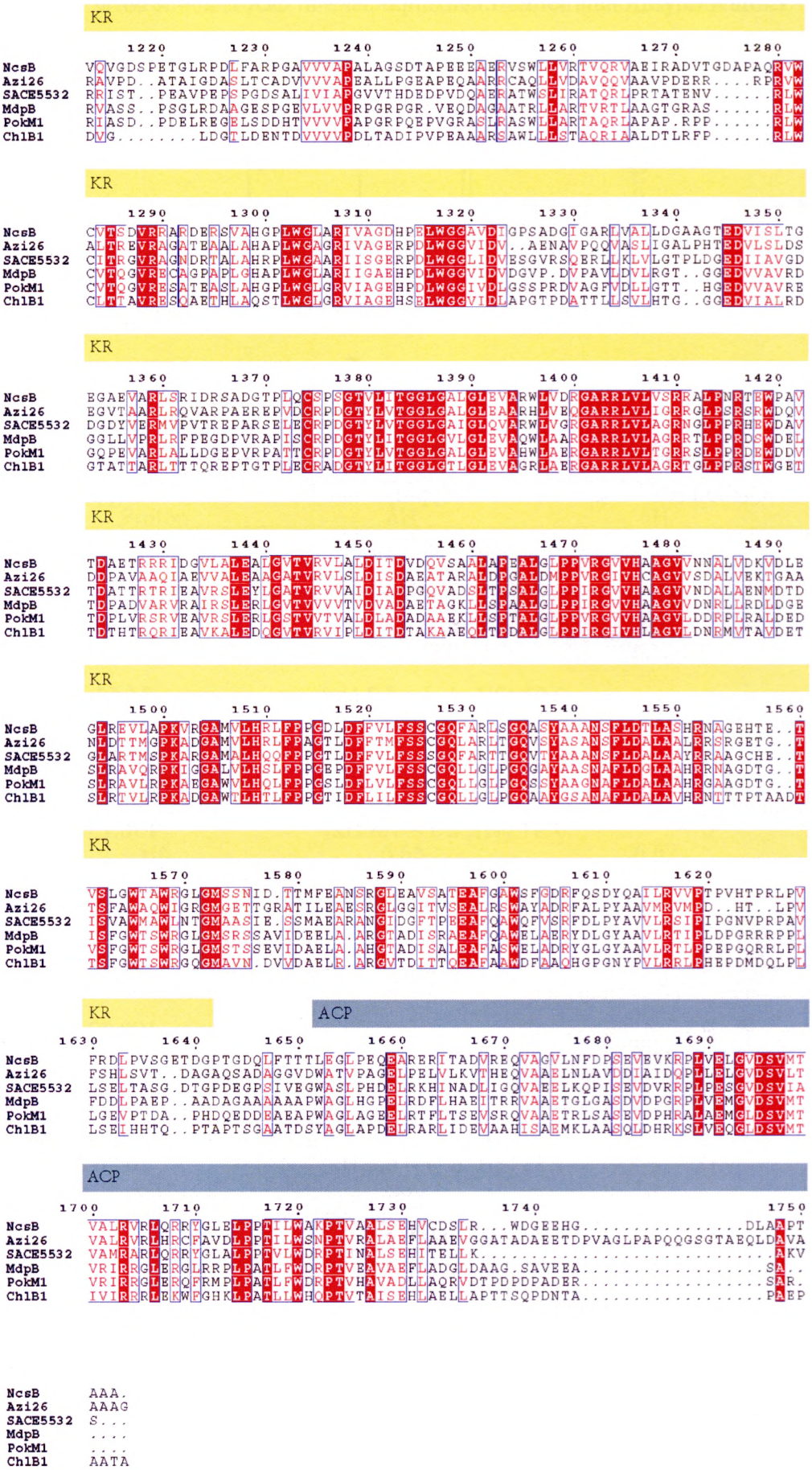
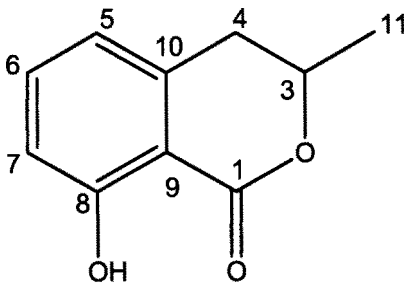


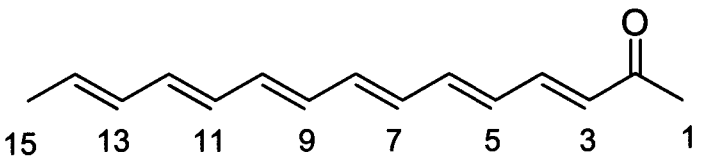
Table A1. NMR data for normal compound **5** in $CDCl_3$ at 400 MHz.



5

Carbon No.	δ ^{1}H (ppm)	δ ^{13}C (ppm)
1		169.94
3	4.73 (1H, m)	76.09
4	2.93 (2H, d, J = 7.2 Hz)	34.63
5	6.69 (1H, d, J = 7.2 Hz)	117.88
6	7.41 (1H, t, J = 7.8 Hz)	136.13
7	6.89 (1H, d, J = 8.4 Hz)	116.27
8		162.23
9		108.32
10		139.38
11	1.53 (3H, d, J = 6.4 Hz)	20.76
8-OH	11.03 (1H, s)	

Table A2. Proton chemical shifts for the isolated and synthetic all-*trans* pentadeca-3,5,7,9,11,13-hexaen-2-one.



	Isolated	Synthetic
Proton	δH	δH
H1	2.30	2.29
H3	6.13	6.13
H4	7.18	7.18
H5	6.34	6.35
H6	6.66	6.67
H7	6.38	6.36
H8-H12	6.2-6.5	6.2-6.5
H13	6.16	6.16
H14	5.80	5.82
H15	1.80	1.82

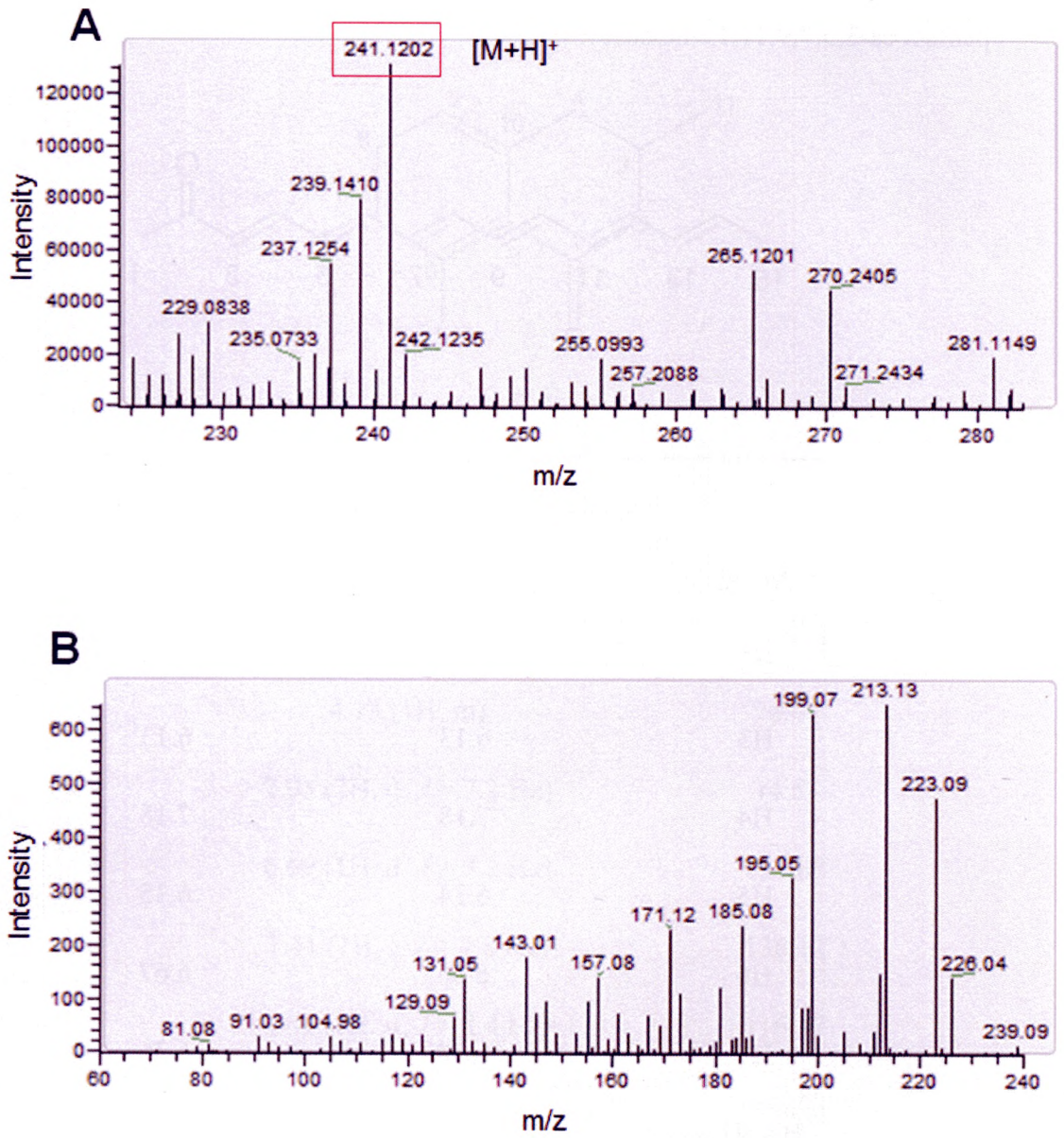


Figure A1. (A) Full mass for **2**. (B) MS/MS spectrum for **2**.

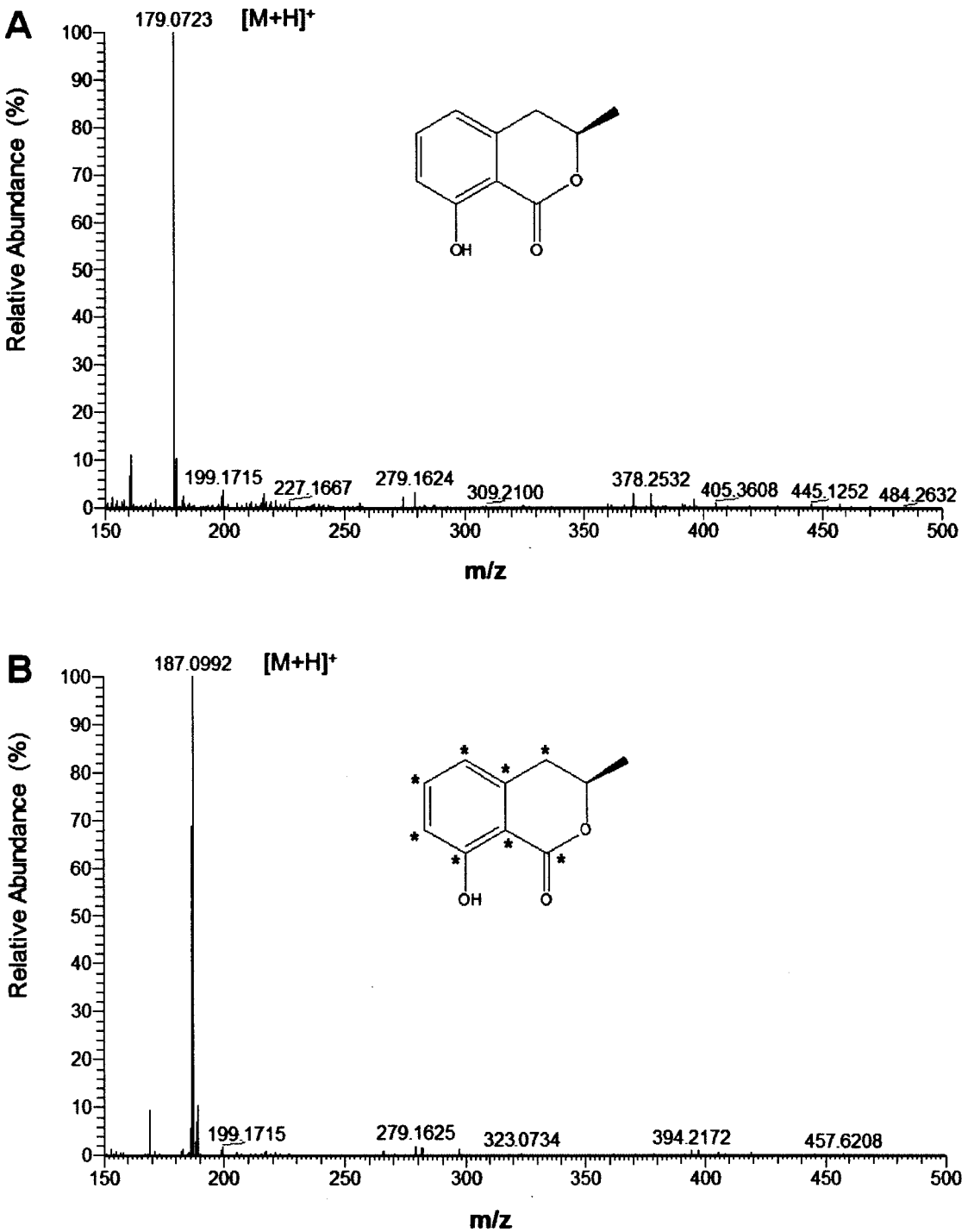


Figure A2. (A) Full mass for **5** produced by unlabeled acetyl-CoA and malonyl-CoA.

(B) Full mass for **5** produced by unlabeled acetyl-CoA and ^{13}C-labled malonyl-CoA.

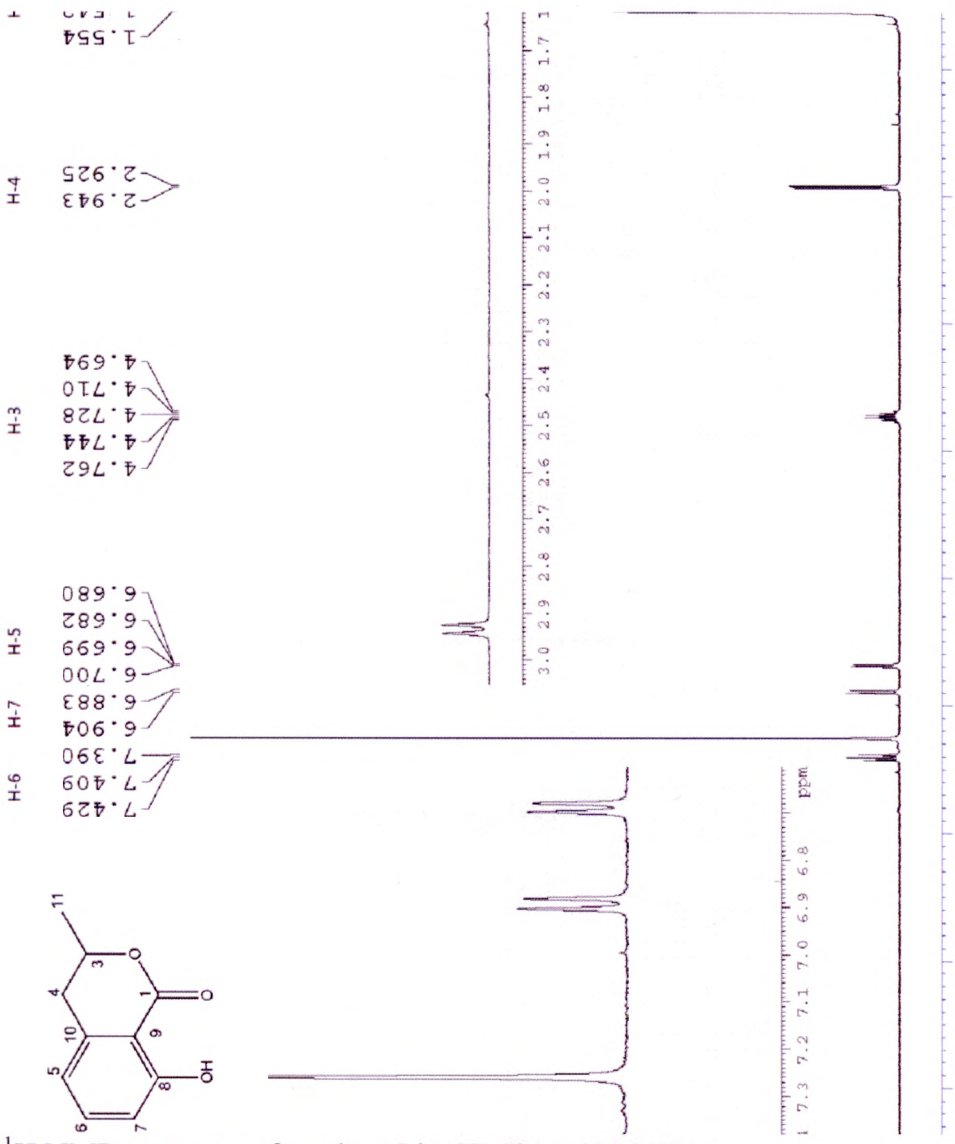


Figure A3. 1D ^{1}H NMR spectrum of product **5** in $CDCl_3$ at 400 MHz.

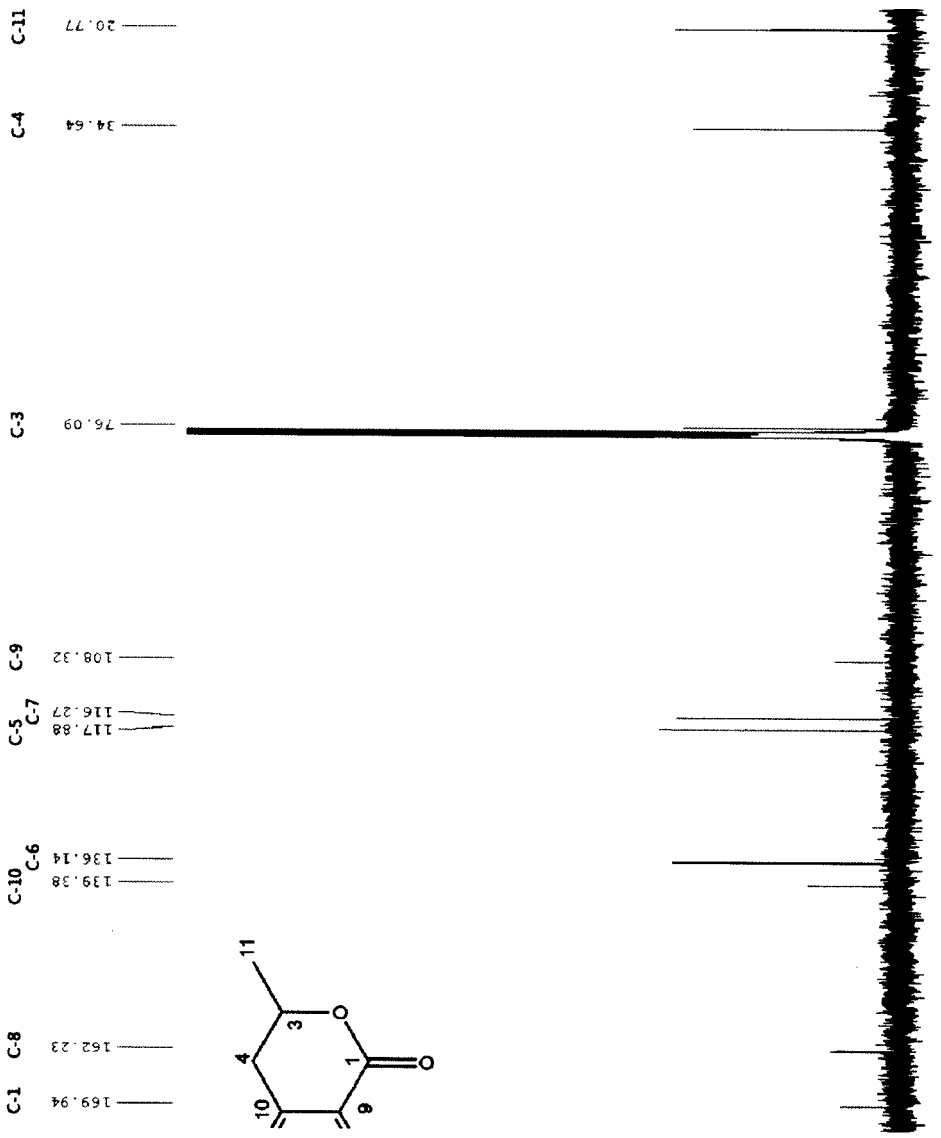


Figure A4. 1D ^{13}C NMR spectrum of product **5** in $CDCl_3$ at 400 MHz.

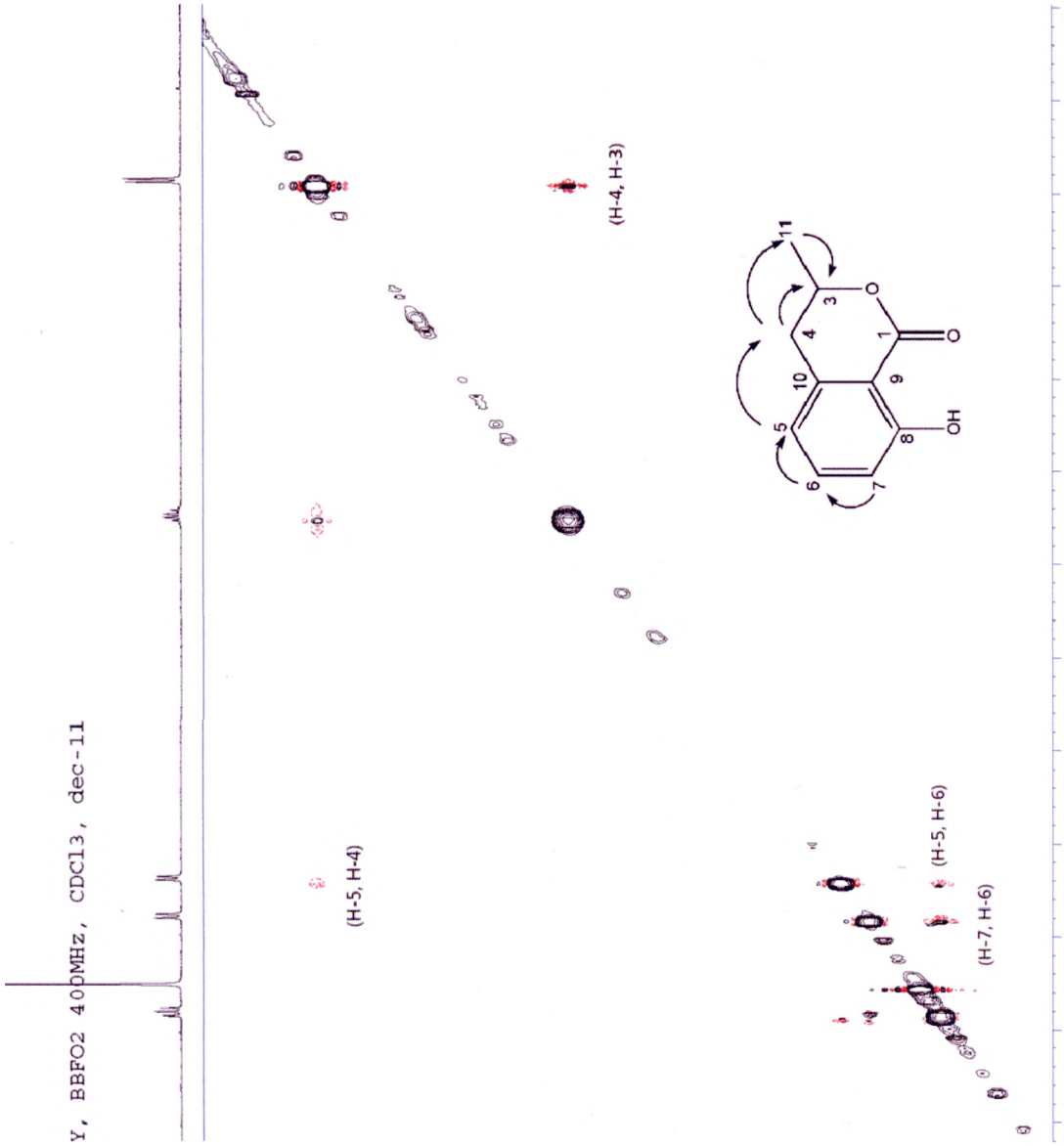


Figure A5. 2D ^{1}H, ^{1}H-NOESY NMR spectrum of product **5** in $CDCl_3$ at 400 MHz.

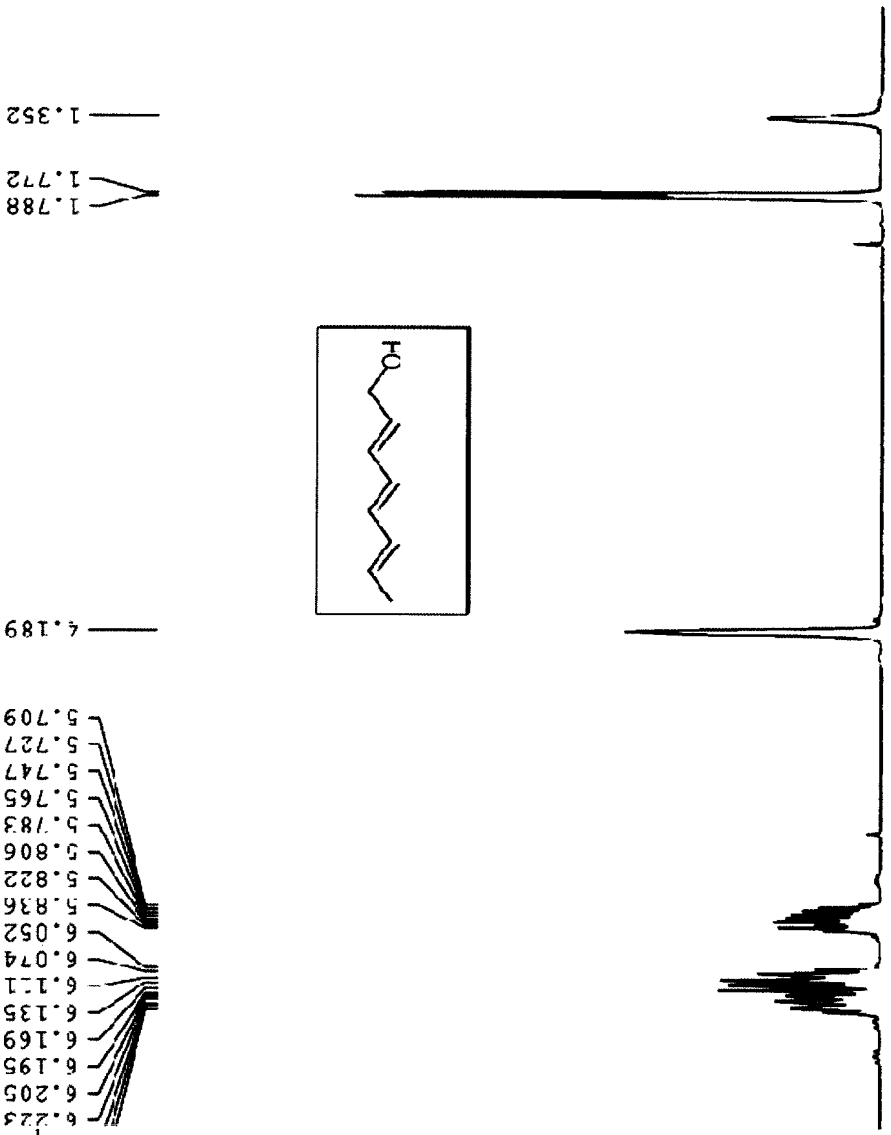


Figure A6. 1D ^{1}H spectrum of **9** in $CDCl_3$ at 400 MHz.

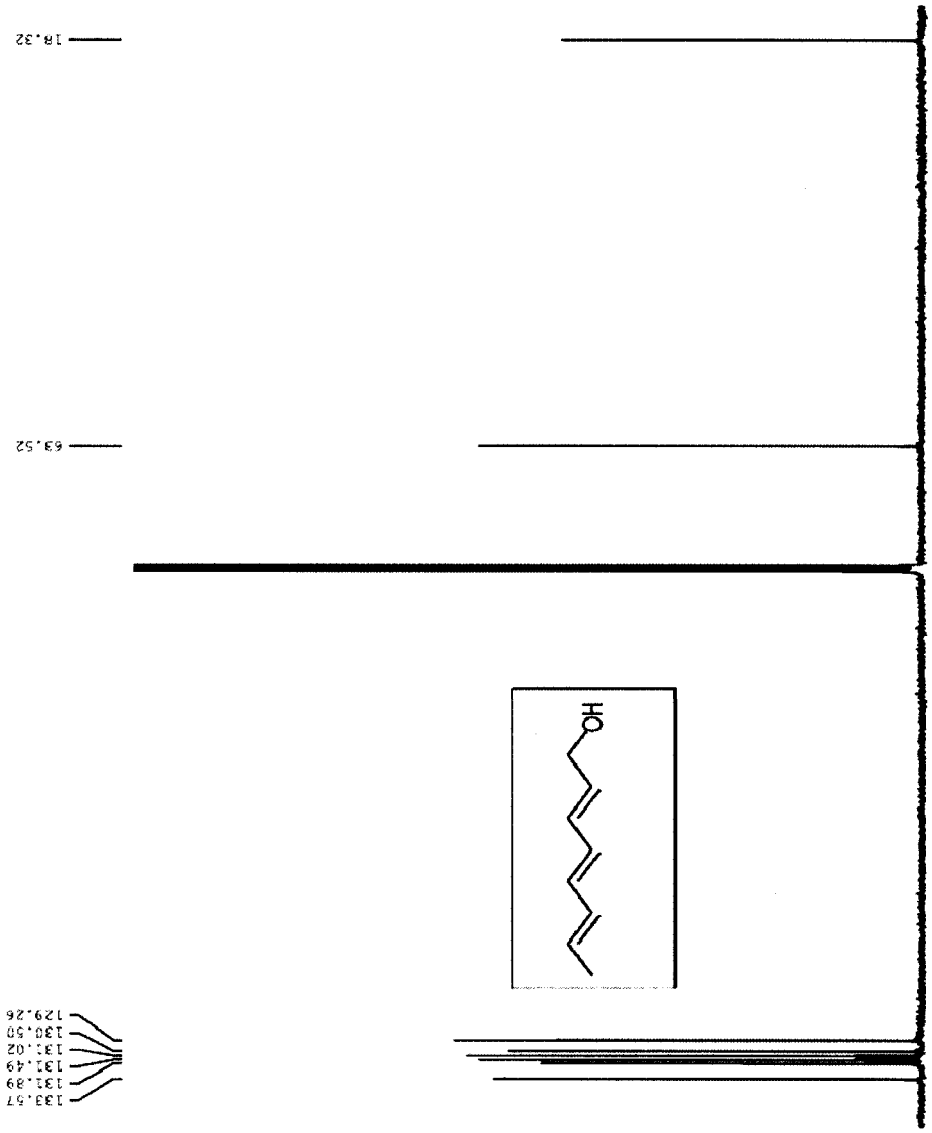


Figure A7. 1D ^{13}C spectrum of **9** in $CDCl_3$ at 100 MHz.

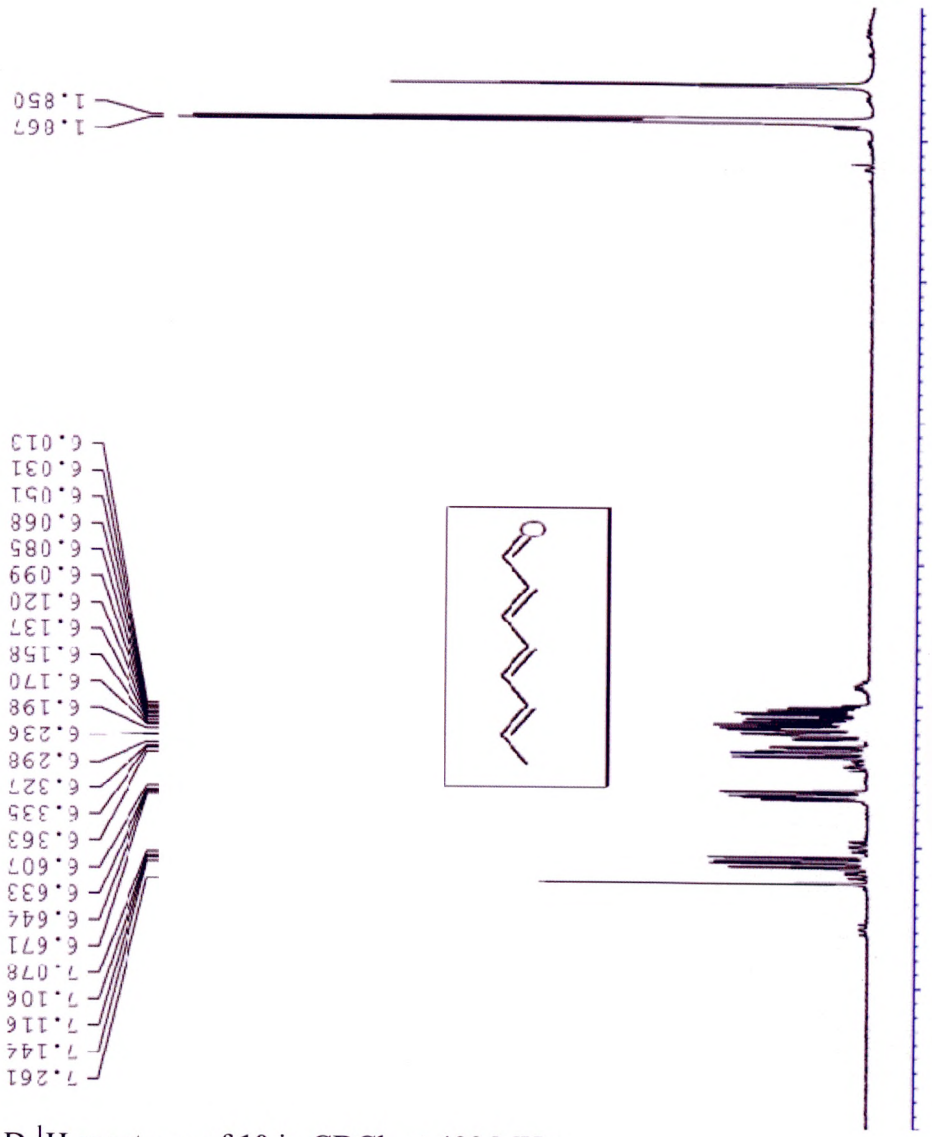


Figure A8. 1D ^{1}H spectrum of **10** in $CDCl_3$ at 400 MHz.

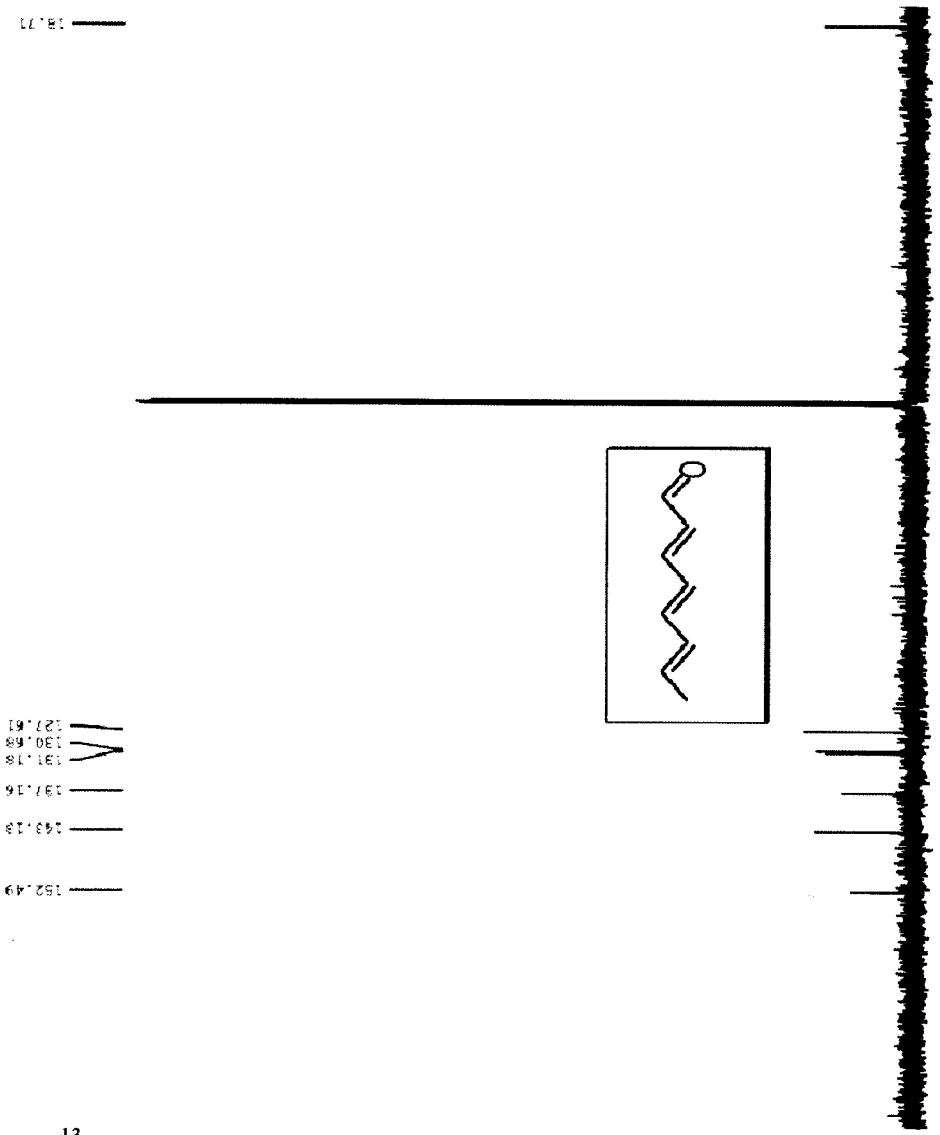


Figure A9. 1D ^{13}C spectrum of **10** in $CDCl_3$ at 100 MHz.

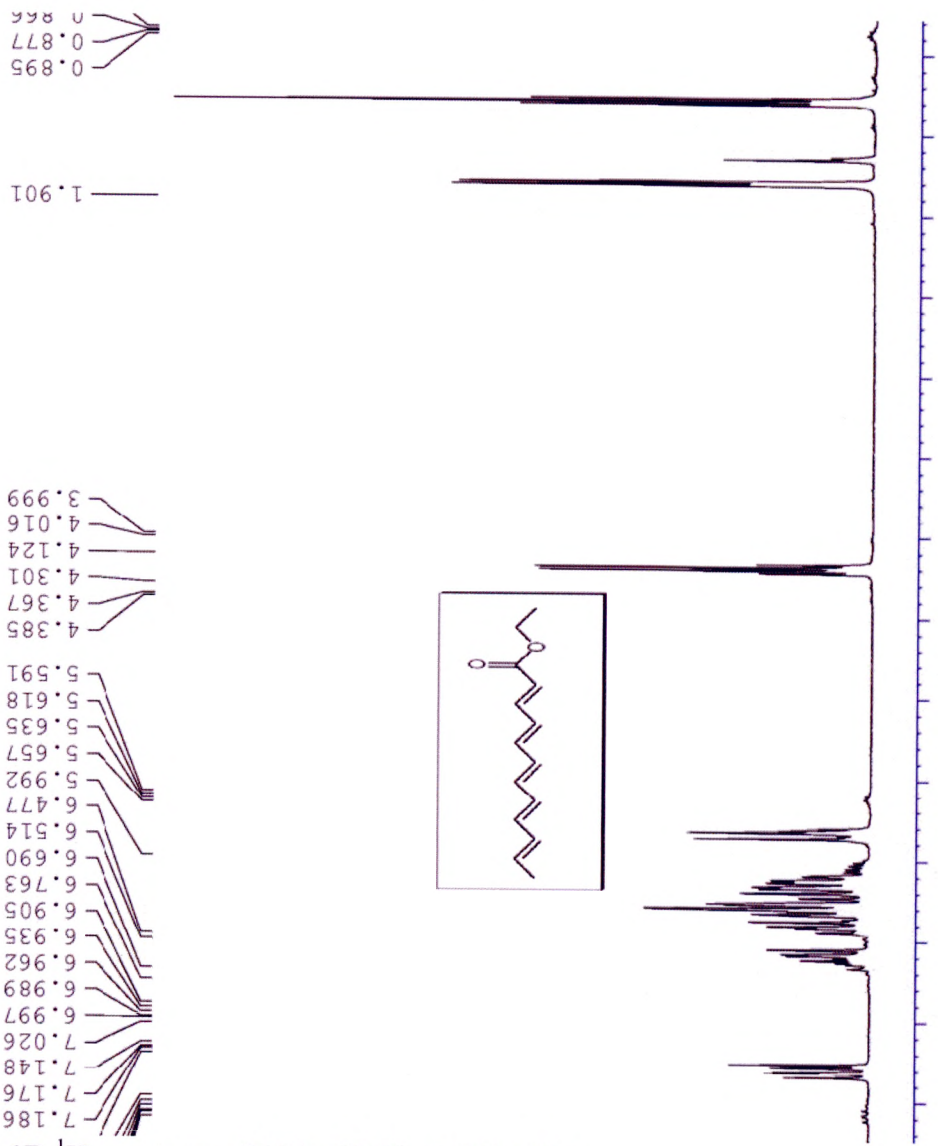


Figure A10. 1D ^{1}H spectrum of **11** in $CDCl_3$ at 400 MHz.

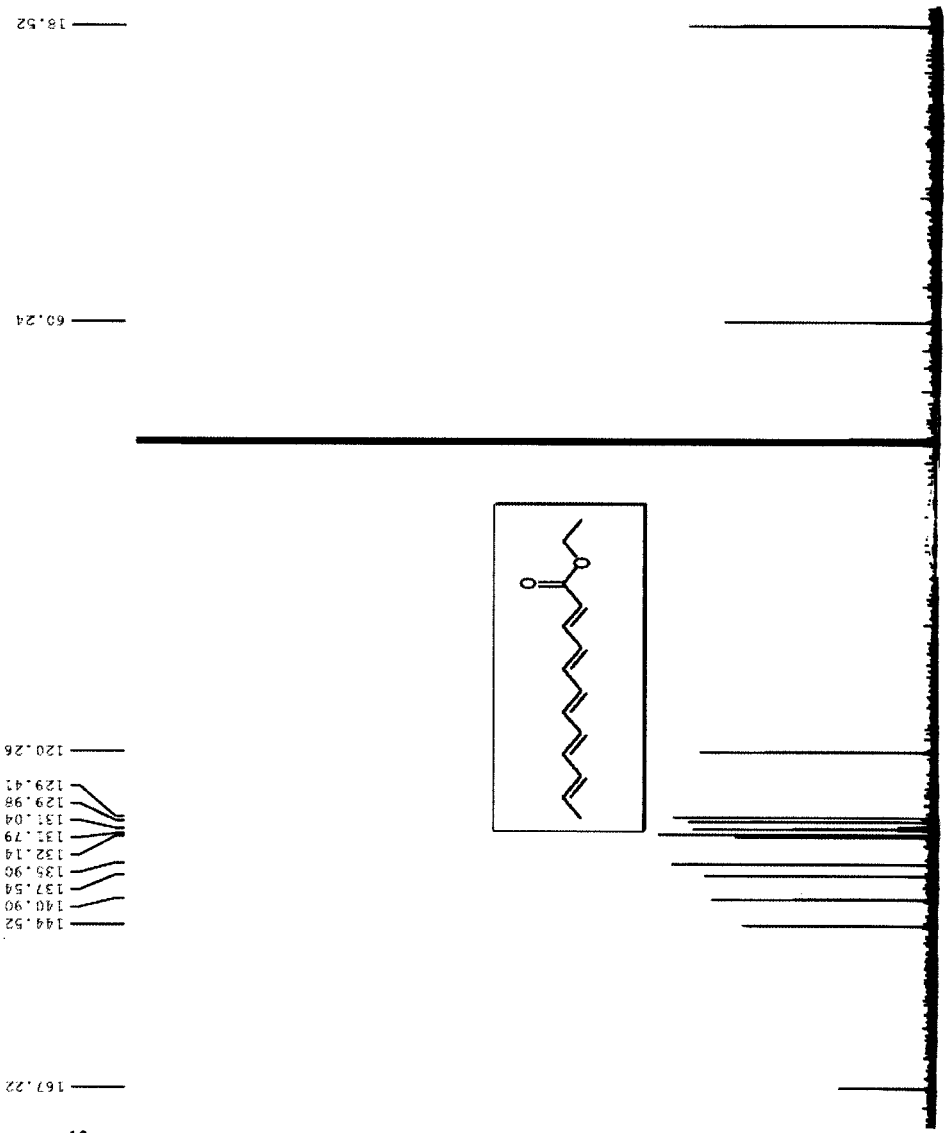


Figure A11. 1D ^{13}C spectrum of **11** in $CDCl_3$ at 100 MHz.

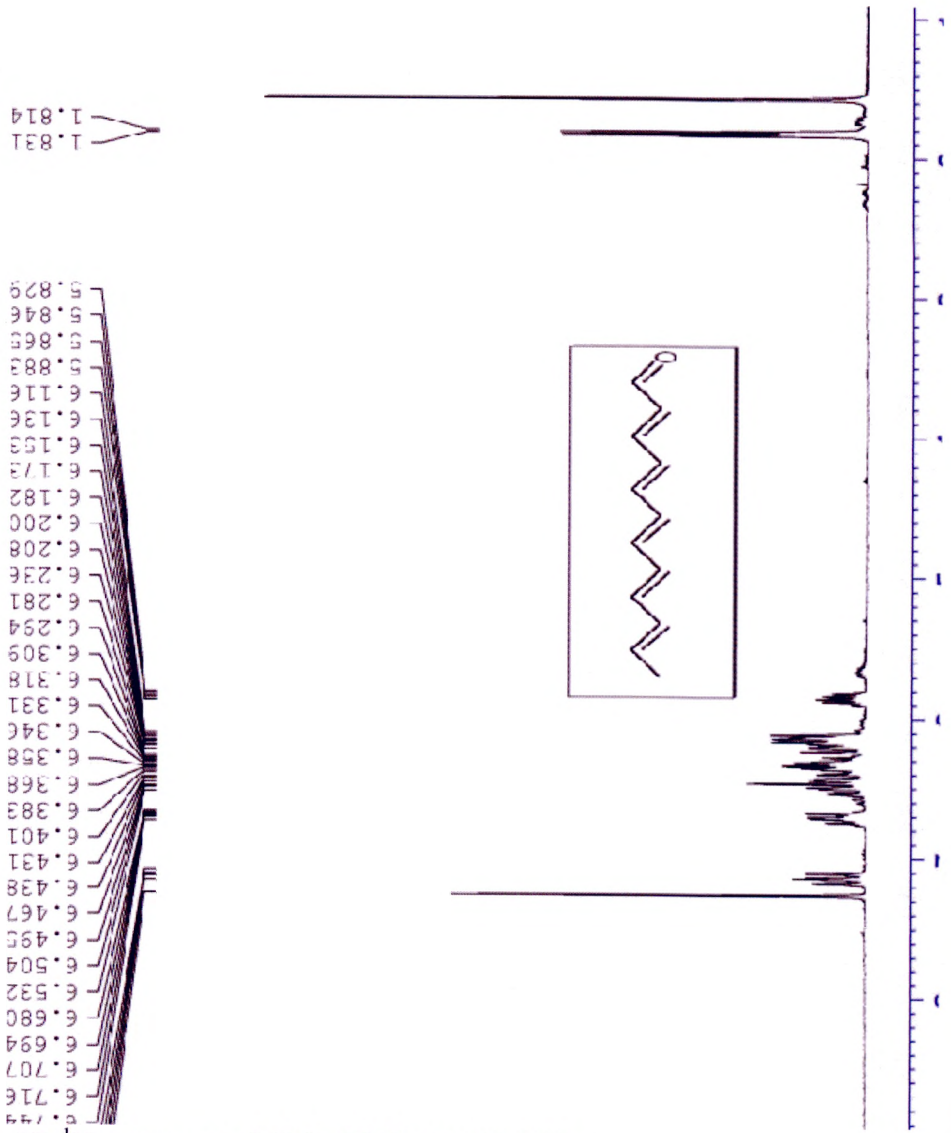


Figure A12. 1D ^{1}H spectrum of **12** in $CDCl_3$ at 400 MHz.

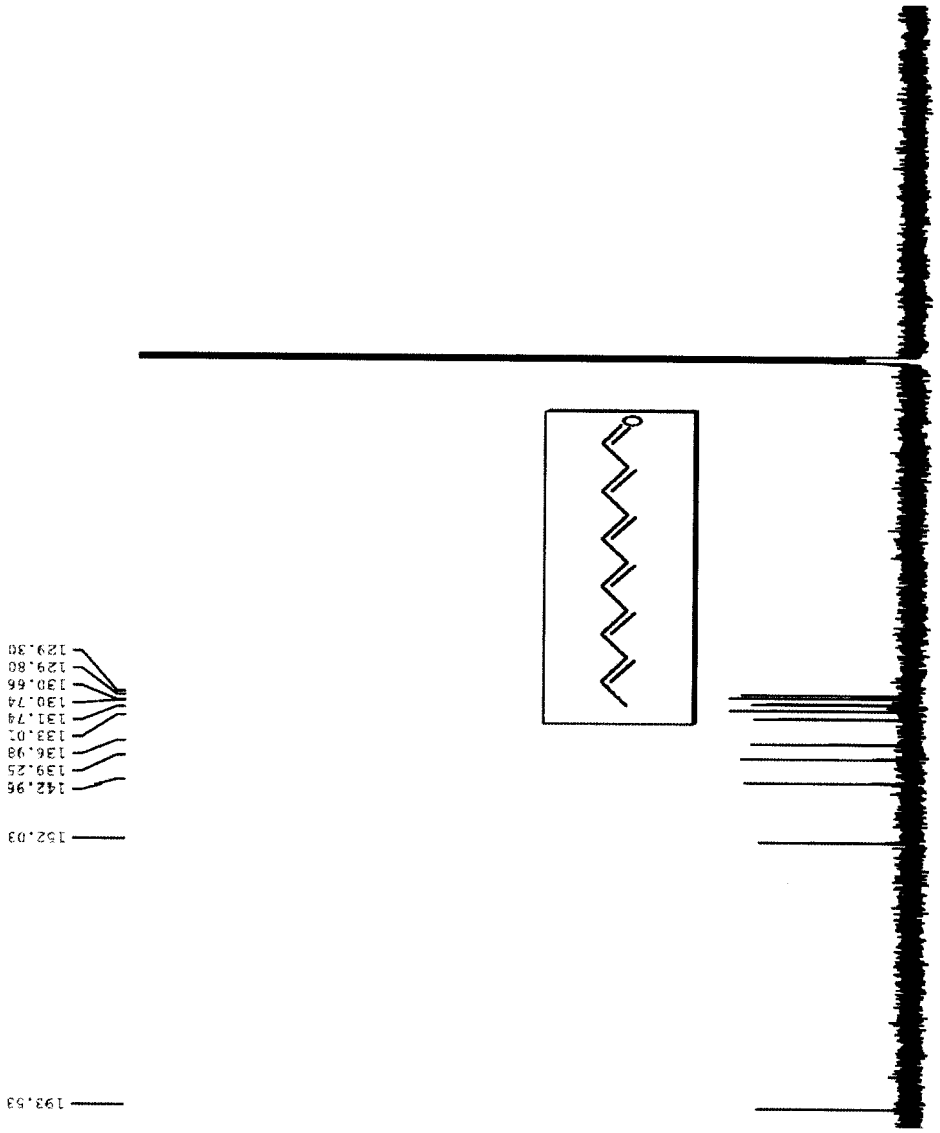


Figure A13. 1D ^{13}C spectrum of **12** in $CDCl_3$ at 100 MHz.

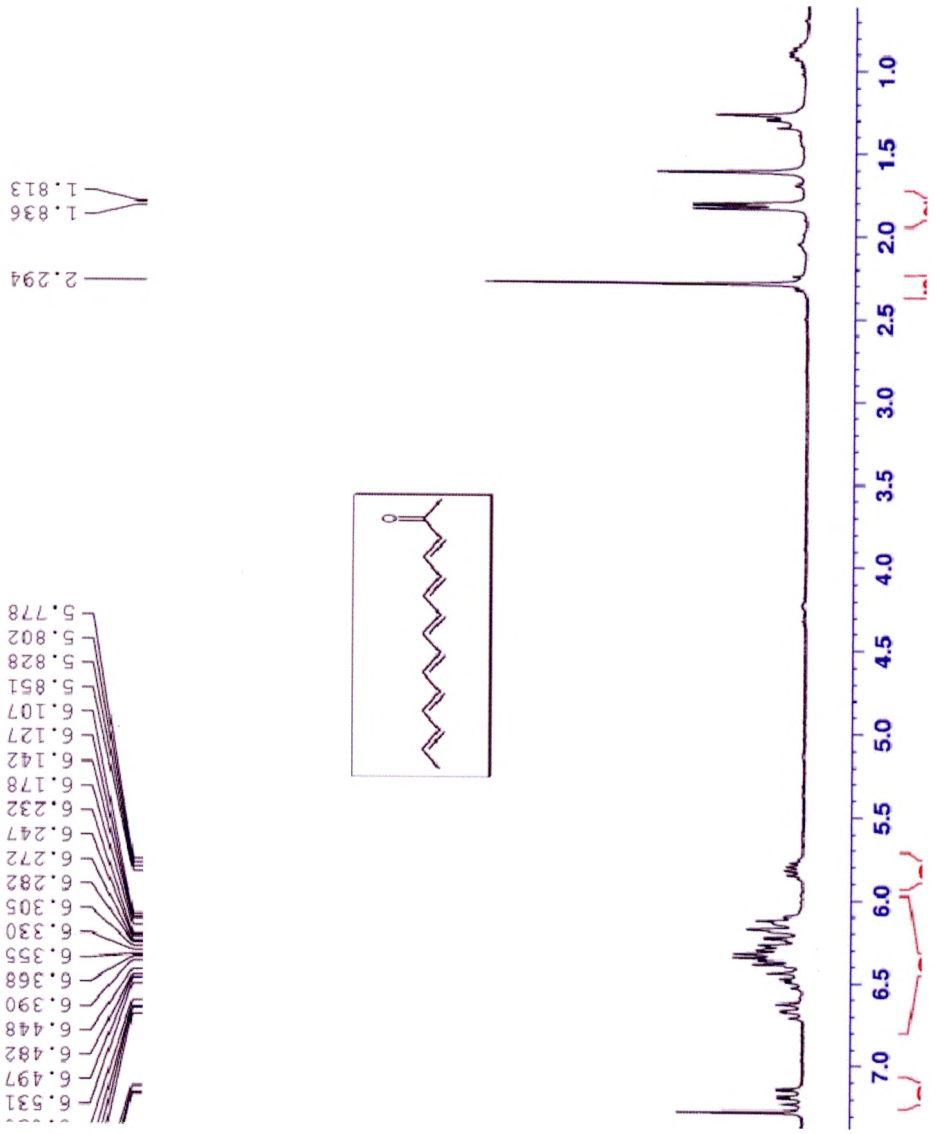


Figure A14. 1D ^{1}H spectrum of **3** in $CDCl_3$ at 400 MHz.

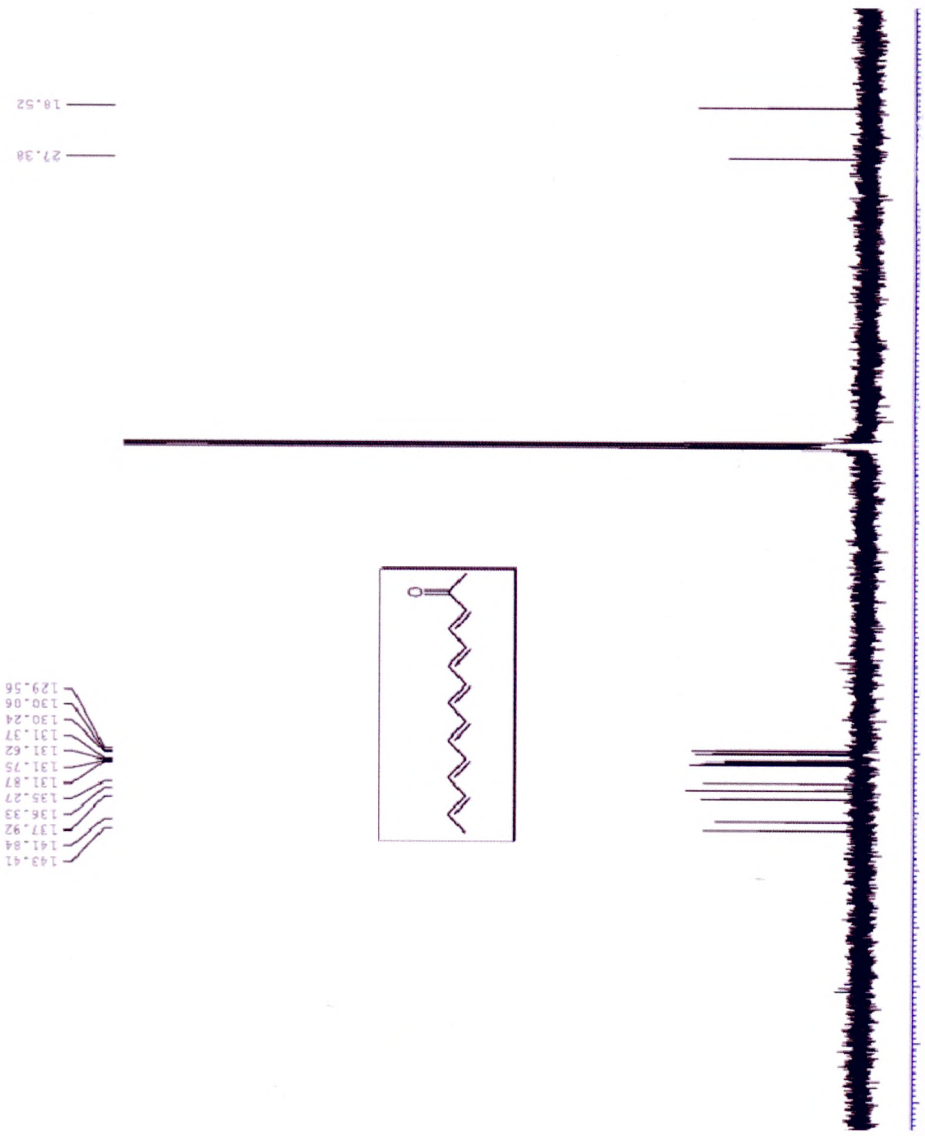


Figure A15. 1D ^{13}C spectrum of **3** in $CDCl_3$ at 100 MHz.