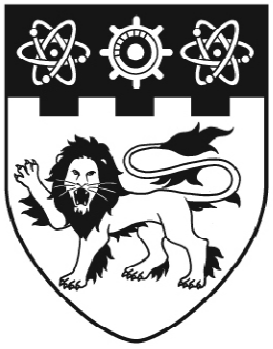


NANYANG TECHNOLOGICAL UNIVERSITY

FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF THE PROTEINS FROM EARLY STEPS OF CALICHEAMICIN BIOSYNTHESIS

KONG RONG

SCHOOL OF BIOLOGICAL SCIENCES

2011

FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF THE PROTEINS FROM EARLY STEPS OF CALICHEAMICIN BIOSYNTHESIS

KONG RONG

SCHOOL OF BIOLOGICAL SCIENCES

NANYANG TECHNOLOGICAL UNIVERSITY

A thesis submitted to the Nanyang Technological University

In partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2011

Acknowledgement

I would like to express my gratitude to my supervisor, Asst. Prof. Liang Zhao-xun, for granting me the opportunity to be a part of such an intriguing project. Thank you for the invaluable guidance and inspiration throughout the project. I would like to show my greatest appreciation to Dr. Liang for taking his precious time to read through and give constructive comments on the writing of this thesis as well.

Apart from my supervisor, I would also like to thank for the spontaneous help and support given by my current and former colleagues, Ela, Rao feng, Ya Ning, Ji Qiang, Dong Wei, Hui Hua, Swathi, Lawrence, Lan Pei, Qin Shi, Mary, Chong Wai, Sun Chi, Siew Lee, Chun Cheong, Thomas, Yee Wei and Xu Min. Thank you all for providing me with such an enjoyable experience in the laboratory.

Moreover, I would like to thank our collaborators Assoc. Prof. Julien Lescar, Asst. Prof. Tang Kai, Assoc. Prof. Yang Daiwen, Asst. Prof. Mu Yuguang, Dr. Masayo, Dr. Li Bin and Mr. Lim Jackwee for their help in structural determination, mass spectrometry and computational simulations. Besides, I would like to show my appreciation to Edward Tan, Dr. Ye Hong and Dr. Anirban for their help in NMR experiments involved.

Lastly but not least, I owe my deepest gratitude to my family members, especially my parents for their ever-lasting support and encouragement throughout my post-graduate studies.

Table of Contents

List of Figures

Abbreviations

1-D	One-Dimensional
2-D	Two-Dimensional
Acetyl-CoA	Acetyl-Coenzyme A
ACN	Acetonitrile
ACP	Acyl Carrier Protein
AML	Acute Myeloid Leukemia
AT	Acyl Transferase
CalE8	Calicheamicin Polyketide Synthase
CalE7	Calicheamicin Thioesterase
CalU15	Calicheamicin Acetylenase
$CDCl_3$	Deuterated Chloroform
CLF	Chain Length Factor
CoA	Coenzyme A
COSY	Correlation Spectroscopy
DEBS	6-Deoxyerythronolide B Synthase
DH	Dehydratase
DTT	Dithiothreitol
DynE8	Dynemicin Polyketide Synthase
DynE7	Dynemicin Thioesterase
EDTA	Ethylenediaminetetraacetic Acid
ER	Enoyl Reductase
FA	Formic Acid
FAS	Fatty Acid Synthase
FPLC	Fast Performance Liquid Chromatography
FTIR	Fourier Transform Infrared Spectroscopy
HCl	Hydrochloric Acid
HPLC	High Performance Liquid Chromatography

HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRE	Internal Reflection Element
kDa	KiloDaltons
KR	Ketoreductase
KS	Keto-acyl Synthase
LB	Luria-Bertani Media
LC-MS	Liquid Chromatography Mass Spectroscopy
Malonyl-CoA	Malonyl-Coenzyme A
*Me*ACP	ACP Domain of CalE8 from *M. echinospora ssp. Calichensis*
MS/MS	Tandem Mass Spectroscopy
MUC-1	Polymorphic Epithelial Mucin
MW	Molecular Weight
m/*z*	Mass to Charge Ratio
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (reduced form)
Ni^{2+}-NTA	Nickel-nitrilotriacetic Acid
NMR	Nuclear Magnetic Resonance Spectroscopy
NOE	Nuclear Overhauser Effect
NRPS	Non-Ribosomal Peptide Synthetase
Octanoyl-CoA	Octanoyl Coenzyme A
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PKS	Polyketide Synthase
PDB	Protein Data Bank
PKSE	Enediyne Polyketide Synthase
PMSF	Phenylmethylsulphonyl Fluoride
ppm	Parts per Million
PPTase	Phosphopantetheinyl Transferase

RPM	Revolutions per Minute
R.M.S.D.	Root Mean Square Deviation
SAX	Small Angle X-ray Diffraction
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SgcE	C-1027 Polyketide Synthase
SgcE10	C-1027 Thioesterase
TE	Thioesterase
TMS	Tetramethylsilane
TS	Transition State
TFA	Trifluoroacetic Acid
TOCSY	Total Correlation Spectroscopy
UV-Vis	Ultraviolet-Visible
WT	Wild Type

Abstract

Polyketides are structurally diverse secondary metabolites produced by bacteria, fungi and plants with a vast variety of biological activities and pharmaceutical properties. Among the polyketide products, naturally-occurring enediynes are some of the most potent antitumor natural products ever discovered. There are two different types of enediyne cores known to date, namely the bicyclo[7.3.0]-dodecadienediyne and bicyclo[7.3.1]-tridecadiynene. Despite some variations in structure, all enediynes consist of three important functional moieties that account for their unprecedented antibiotic and antitumor activities. A central aglycone warhead "attacks" DNA by hydrogen abstraction through the free radicals generated by the Bergman or Myers-Saito cyclization mechanism; a delivery system composed of oligosaccharide fragments guides the molecule to the DNA target and a triggering device activates the molecule to generate the highly reactive free radicals. Due to their astonishing capacity in cleaving DNA, some of the enediynes have already been employed in cancer treatment.

Calicheamicin γ_1^I, which belongs to the 10-membered enediyne family, is one of the earliest enediyne natural products used in the treatment of leukemia. Although the structure and the chemical functionality of Calicheamicin γ_1^I have been fully deciphered since its discovery, the biosynthetic pathway of the bicyclo[7.3.1]-tridecadiynene is still shrouded in mystery. Studies on the biosynthesis of enediynes have recently been made possible with the identification of gene clusters responsible for enediyne biosynthesis in 2002. The identification of the "minimal enediyne PKS gene cassette" in the gene cluster also paved the way for scientists to look into the biosynthetic pathway of the enediyne core. With the genomic information at hand, we set out to explore the biosynthetic origins of the enediyne core of calicheamicin γ_1^I. After successfully cloning and expressing three proteins encoded by the genes from the minimal cassette in an *E.*

coli expression system, a series of biochemical and structural studies have been carried out to examine the structure and function of the three proteins that include the type I iterative polyketide synthase (PKS) CalE8, the thioesterase CalE7, and the putative acetylenase CalU15. The roles of the three proteins in the early steps of calicheamicin biosynthetic pathway were investigated. A series of products of CalE8 have been identified through *in vitro* and *in vivo* biochemical assays to establish CalE8 as the first enzyme in the biosynthetic pathway for the synthesis of the enediyne moiety. The NMR structure of the acyl carrier protein (ACP) domain of CalE8 was determined to yield insight into the structure and function of the shuttling domain. Together with mutagenesis studies, the crystal structures of CalE7 and the homologous DynE7 were determined, revealing a remarkable induced-fit mechanism during substrate binding and a novel catalytic mechanism in releasing the PKS product.

CHAPTER 1 Introduction

1.1 Enediyne Natural Products

1.1.1 Introduction to Enediyne

Polyketides are structurally diverse secondary metabolites produced by bacteria, fungi, plants and even sea urchins [1] with a vast variety of biological activities and pharmaceutical properties. The term polyketide was coined more than a hundred years ago by organic chemist Collie in attempt to classify a type of aromatic molecules of synthetic origin during his studies on pyrones [2]. From the experimental data obtained, Collie proposed that the biosynthesis of phenolic natural products with specific hydroxyl substitutions in living organisms could be from a β-polyketone origin. Moreover, Collie's theory also raised the possibility of polyketones being derived from polymerization of ketenes. Therefore the term polyketide came into existence with reference to both the polyketones and their phenolic derivatives. However, the notion did not receive enough attention from the scientific world until 50 years later. In the mid-1950s, another biochemist named Birch developed Collie's theory further. He suggested that the polyketones might originate from the repetitive condensation of acetate units. With the advancement in radioactive techniques, Birch was able to test out his hypothesis via feeding experiment using isotopically-labeled acetates [3]. After the addition of ^{14}C labeled acetic acid into the medium, isotopically labeled 6-methylsalicylic (6-MSA) acid was obtained from the fungal culture. The discovery of labeled 6-MSA along with many other phenolic natural products further validated his hypothesis. Finally, the scientific community realized the importance of Collie's theory and the term polyketide was universally recognized afterwards. Following the formulation of the

polyketide theory, rigorous research in the field of polyketide natural products ensued. Subsequently, the number and diversity of polyketide structures discovered escalated at an astronomical speed [4-7].

Among the polyketide natural products, the enediynes are some of the most potent antineoplastic agents ever discovered. The first naturally-occurring enediyne with its structure deciphered is neocarzinostatin, which was reported by Edo and his group in 1985 [8, 9]. The structure of neocarzinostatin features an extraordinary bicyclo[7.3.0]-dodecadienediyne core with an epoxide-masked 1, 5-diyne-3-ene unit encompassed in a 9-membered ring. The study on enediyne continued to advance with the discovery of calicheamicins [10, 11] and esperamicins [12, 13] in 1987. Astonishingly, the discovery depicted a different ring structure from that of neocarzinostatin, with a bicyclo[7.3.1]-tridecadiynene core containing a 1, 5-diyne-3-ene unit. Since then, research on enediyne flourished and new members of enediyne family have been discovered and characterized afterwards.

Since there are two different types of enediyne core in existence, namely the bicyclo[7.3.0]-dodecadienediyne and bicyclo[7.3.1]-tridecadiynene, naturally-occurring enediynes are classified into 9-membered and 10-membered families. The structures of the different members for the two families of enediyne are illustrated below (**Figure 1.1**, **Figure 1.2**). Among the 9-membered enediynes discovered, the last two members highlighted in brackets are hypothesized to be the enediyne precursors for sporolides and cyanosporasides, which are the stable forms isolated from the marine bacterium *Salinospora* [14-16].

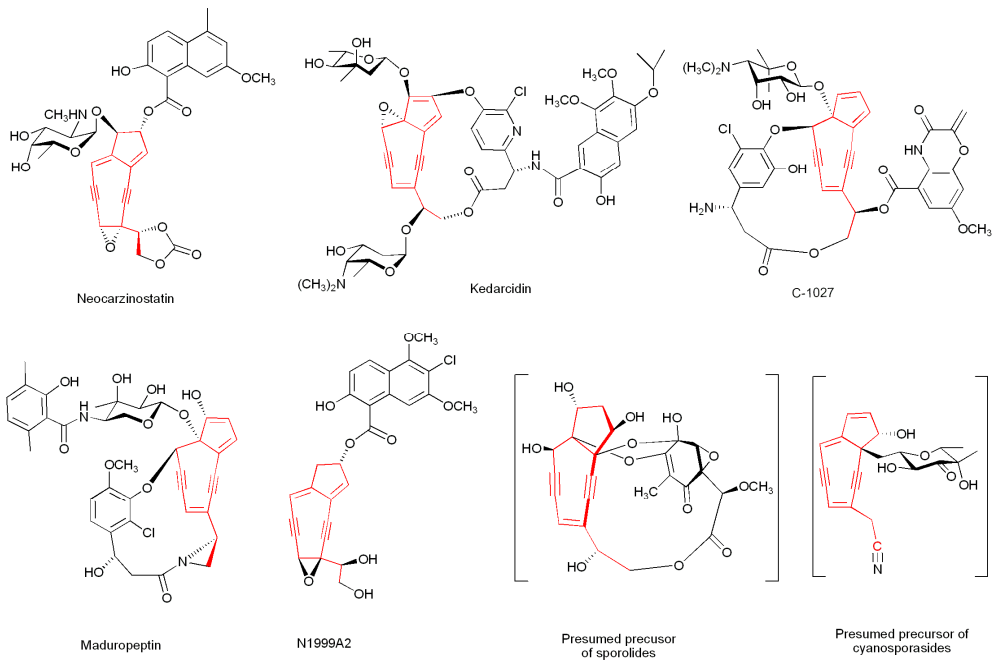


Figure 1.1 9-membered enediynes with warhead structure highlighted.

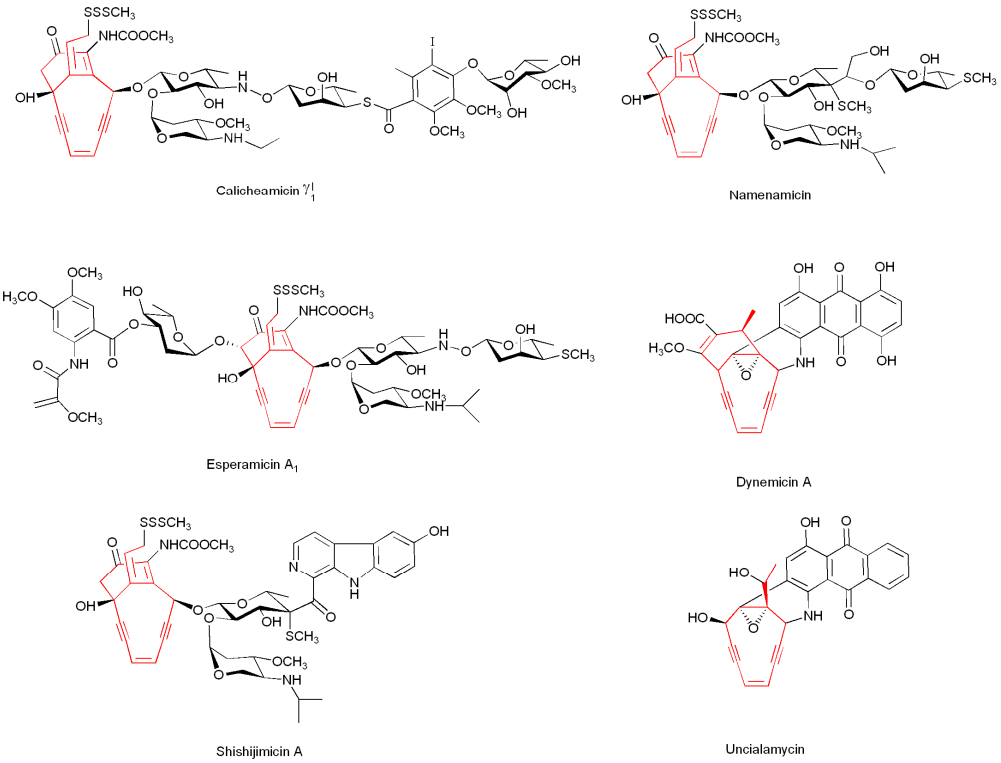


Figure 1.2 10-membered enediynes with warhead structure highlighted.

All 9-membered enediynes require an *apo*-protein for stabilization and transportation within cells except N1999A2, which was discovered without an apo-protein component. Unlike their 9-membered counterparts, the 10-membered enediynes do not require the stabilization or transportation by an *apo*-protein, and they can be further categorized into two sub-families: the calicheamicin-like enediynes and dynemicin-like enediynes, with slight differences lying in both the bicyclic enediyne core and the anthraquinone moiety.

1.1.2 Structure and Functionality of Enediynes

Despite the difference in structure, all enediynes consist of three important functional domains responsible for their unprecedented DNA cleaving activity [17, 18]: 1) an aglycone warhead that attacks DNA after cyclization; 2) a delivery system composed of oligosaccharide fragments that guides the molecule to DNA target [19, 20] and 3) a triggering device that activates the molecule to generate highly reactive radicals [21]. The widely accepted cyclization mechanisms of enediyne core are Bergman cyclization (enediyne) and Myers-Saito (enyne-allene) reaction [22-24], both of which lead to the production of *p*-benzyne diradical as illustrated below (**Figure 1.3**). Upon administration, the enediynes are delivered to the chromosomal DNA where intercalation takes place in a sequence-specific manner. Subsequently, apoptosis is induced when DNA is damaged by the radicals generated [17, 25-28].

Among all the enediynes, diradical intermediates are readily generated upon cycloaromatization via nucleophilic addition on the warhead for C-1027 and calicheamicin-like enediynes; whereas for neocarzinostatin, cycloaromatization takes place spontaneously even at ambient temperature due to the high reactivity of the chromophore. Once enediyne molecules are tightly bound to the minor groove of DNA, abstraction of hydrogen atoms occurs on the DNA molecule in the presence of the diradicals generated which leads to the formation of a carbon-

centered radical on the ribose. As a result, DNA molecule will undergo double- or single-stranded cleavage via oxidative radical pathway with the help of molecular oxygen.

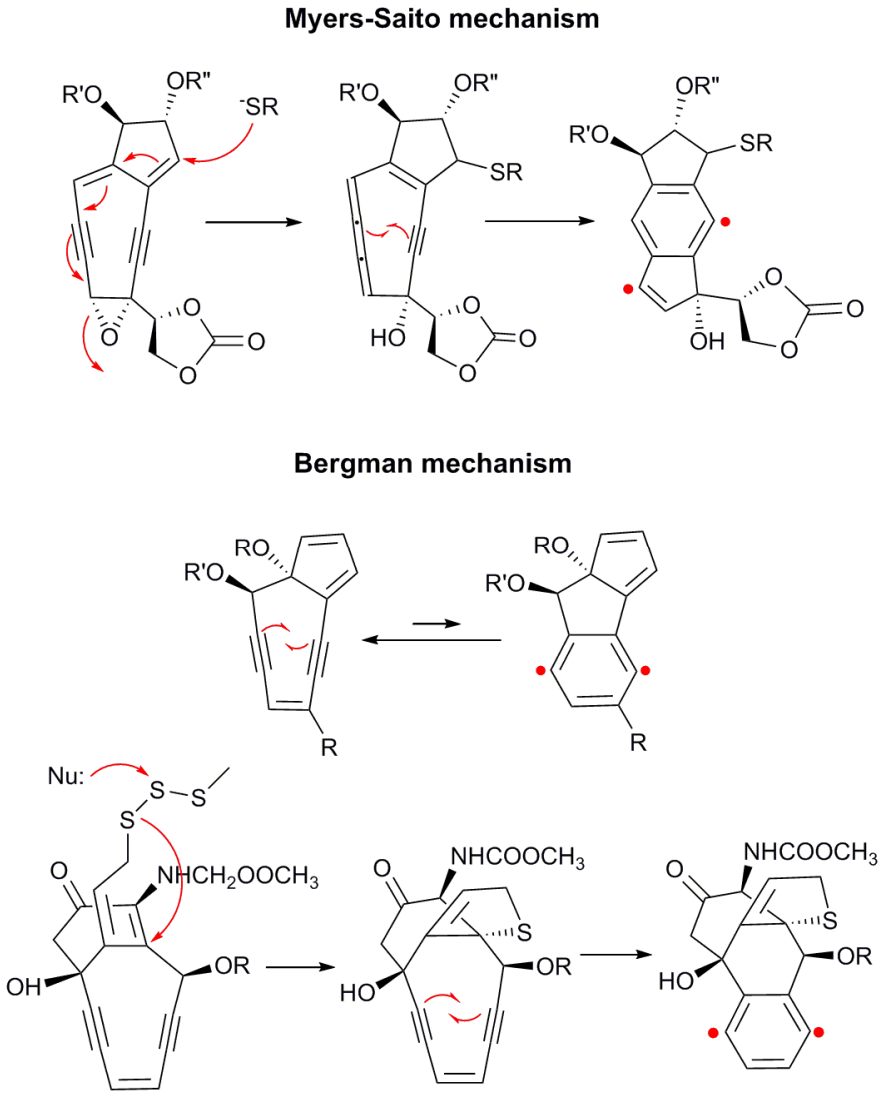


Figure 1.3 Cyclization mechanisms of enediyne core via either Bergman (enediyne) or Myers-Saito (enyne-allene) reaction with the generation of diradicals.

However, due to the inadequate selectivity of the compound, enediynes destroy tumor cells as well as normal cells indiscriminatively upon administration. Therefore, research on harnessing the rigorous DNA-cleavage activity of enediynes is becoming an area of high priority in designing of anti-tumor drugs. Up to date, several different methods are being employed to

overcome the problem of non-specificity. For instance, a derivative of neocarzinostatin was developed in 1994 with a polymer conjugation, which was later known as zinostatin stimalamer (SMANCS). The drug showed an improved specificity towards tumor cells and was being used for the treatment of hepatocellular carcinoma in Japan [29]. Besides, the possibility of developing antibody-conjugated enediynes has also been explored, which has yielded a few successful drugs for cancer treatment as well. Apart from the medical applications, the capability of enediynes to undergo cycloaromatization is also drawn upon in the development of selective nucleases and high performance linear aromatic polymers for microelectronic fabrication. The expansion of enediyne-related applications in chemistry and medicine will be within our reach if the cycloaromatization process can be spatially and temporally controlled.

Given their extraordinary structural complexity and biochemical properties, great interests have been roused in organic chemists worldwide to study the chemical synthesis of enediynes. Attempts in the construction of enediyne entailed enormous efforts and long period of time, which eventually led to the production of almost all the enediynes purely by chemical synthesis [30-40]. The extensive studies on the total synthesis of enediynes unraveled the mystery of their complex structure as well as their mode of action in DNA cleavage. In addition, many research groups have ventured into the field of synthesis of enediynes with modified functional groups, which further provided researchers with the opportunity for enediyne engineering to cater for a broader range of medical applications [41-44]. With the chemical synthesis of the enediynes accomplished, researchers started to ponder the biosynthetic origins of enediyne products. Due to the structural similarities between the 9- and 10-membered enediynes, a hypothesis that features the two types of enediynes stemming from a common precursor was formulated among the researchers (**Figure 1.4**).

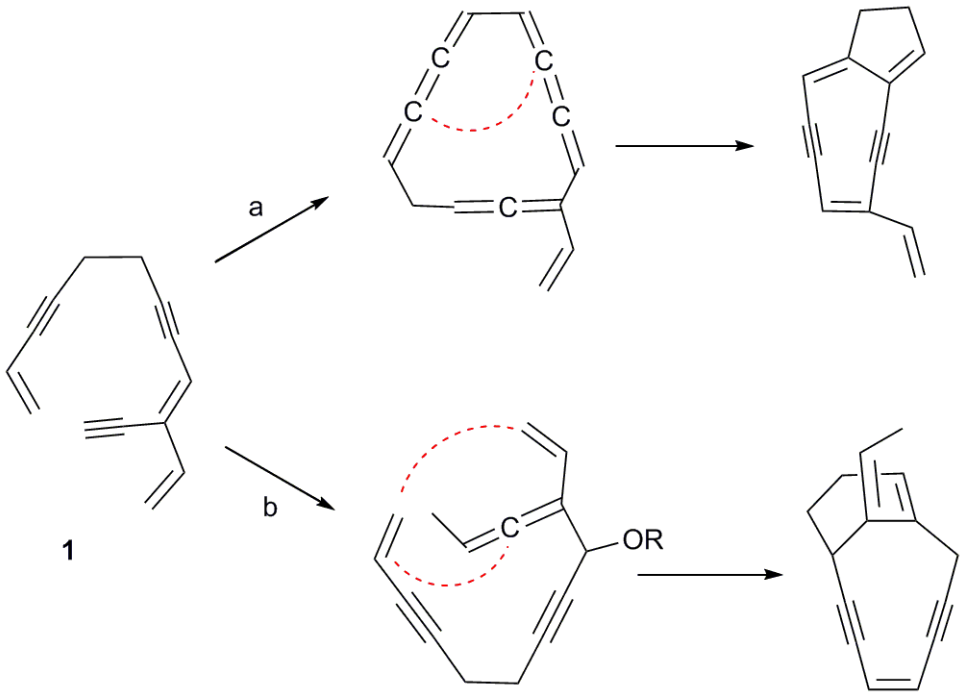


Figure 1.4 An early biosynthetic mechanism postulated for enediynes. The branched molecule **1** shown is thought to be the common precursor for both 9- and 10-membered enediynes.

9-membered enediynes could be obtained from the branched molecule **1** through a series of chemical reactions which involved electrocyclic ring closure, proton transfer and oxygenation (pathway a). On the other hand, the 10-membered enediynes could only be generated by a single carbon addition on the acetylene terminus of the molecule **1** via intramolecular Diels-Alder reaction (pathway b) [45]. However, the notion that the enediynes were derived from branched precursor did not last long. Another group contested the idea through a series of carbon labeling experiments performed on one of the 9-membered enediynes— neocarzinostatin [46]. By feeding the bacteria *Streptomyces carzinostaticus* with labeled acetate, the group established that the enediyne core was entirely derived from acetate units connected in a head-to-tail fashion. According to the data collected, a linear acyl chain consisting of seven acetate units is likely to be the precursor of the bicyclo[7.3.0]-dodecadienediyne moiety. Furthermore, the J-coupling results suggested that the two carbons of the *–yne* group belonged to the same acetate unit

whereas the two terminal carbons were observed as singlets. As a result, the authors proposed that the 9-membered ring was derived more likely from terminal decarboxylation of oleate via the well-known oleate-crepenynate pathway for polyacetylenes than *de novo* synthesis from acetates (**Figure 1.5**).

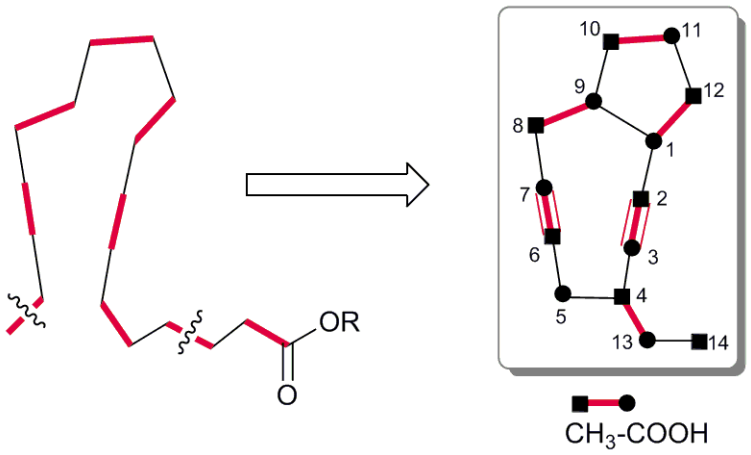


Figure 1.5 Proposed folding pattern for the 9-membered enediyne of neocarzinostatin through the established oleate-crepenynate pathway.

In 1992, Tokiwa's group conducted ^{13}C feeding experiments on the organism *Micromonospora chersina* M956-1 in attempt to obtain the biosynthetic mechanism for a 10-membered C_{14} enediyne— dynemicin A [47]. After extensive NMR studies, the authors concluded that the bicyclic enediyne core was produced from heptaketide precursor consisting of seven head-to-tail coupled acetate units (**Figure 1.6**). Furthermore, the proposed labeling pattern was extended to explain the structure of another class of 10-membered bicyclic ring, namely the C_{15} calicheamicin-like enediynes. In the hypothesis an octaketide was thought to be the precursor for both dynemicin- and calicheamicin-like enediyne cores. The biosynthetic pathway branched out at the decarboxylation stage where a formal loss of two carbons from the carboxylate end or a single decarboxylation step occurred at either end of the octaketide led to the differentiation between dynemicin-like and calicheamicin-like enediynes. The origination of

the hepta- and octaketides were attributed to the oleate-crepenynate pathway as well, although the authors acknowledged that the pathway has been shown to exist only in higher organisms like plants and fungi. Through J-coupling analysis, the two constituent carbons of the –*yne* group were found to be derived from separate acetate units, unlike their 9-membered enediyne counterparts in which the two carbons originated from a single acetate unit.

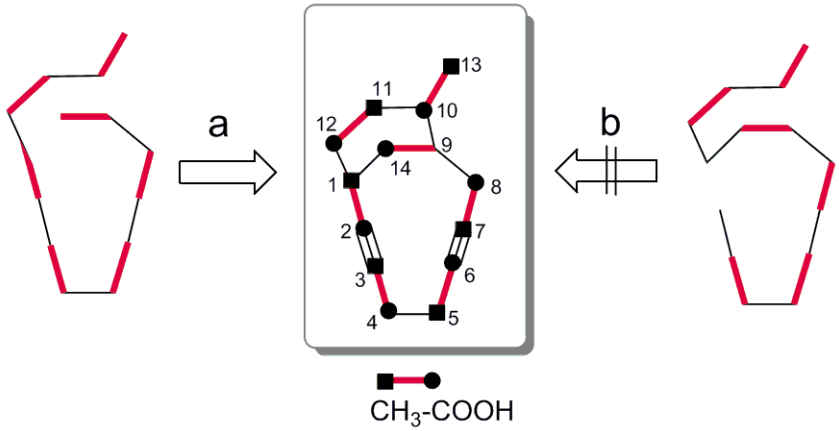


Figure 1.6 Proposed folding pattern for the 10-membered enediyne of dynemicin A. Two possible folding mechanisms were analyzed, pathway a is preferred according to the data from the feeding experiments.

One year later, Lam and coworkers performed another isotopic labeling experiment with esperamicin from the class of calicheamicin-like enediynes [48]. The J-coupling constants depicted a C_{15} acyl chain of seven acetate units linked in a head-to-tail fashion with the last carbon C15 existing as a singlet. Four ways of folding the acyl chain into the enediyne core were proposed as illustrated below (**Figure 1.7**). With respect to the connectivity between the carbon atoms based on the J-coupling constants, folding pattern a and d were not feasible and hence ruled out. Comparison among the labeling intensity of the different carbon atoms revealed that C11 and C12 might be the chain termination unit whereas C15 was probably part of the starter acetate unit. Loss of the carboxylate group from the starter acetate unit would yield a C_{15}

polyketide which could be further folded into the matured product of esperamicin core. Therefore the authors preferred path c rather than b for the folding pattern of esperamicin core.

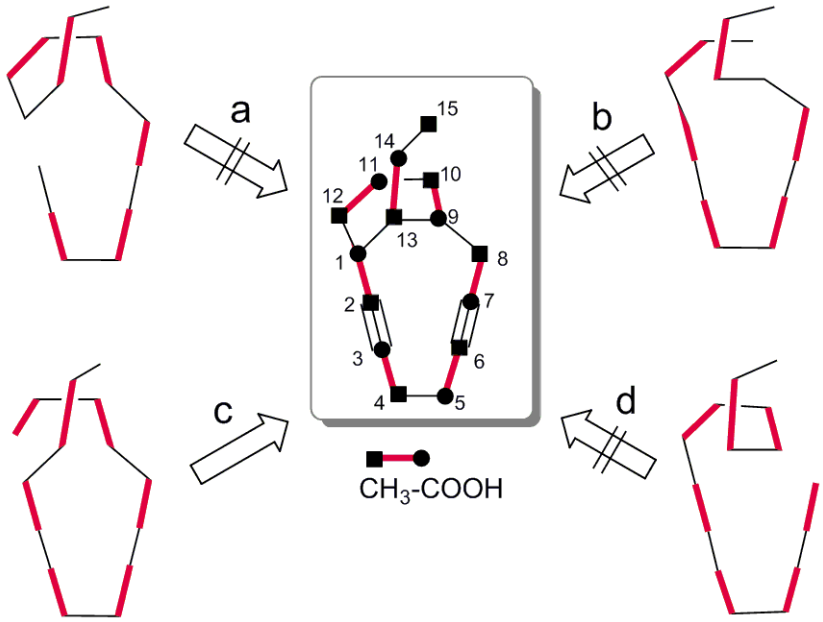


Figure 1.7 Proposed folding patterns for the 10-membered enediyne of esperamicin A_1. Pathway c is more likely according to the values obtained for J-coupling constant.

Apart from the labeling experiment, Lam and coworkers checked on the feasibility of deriving enediynes from oleate-crepenynate biosynthetic pathway as well. In the experiment conducted, cerulenin was added to bacterial culture to inhibit the function of β-ketoacyl-acyl-carrier-protein synthase, which invariably led to the abolishment of esperamicin production. However, subsequent supplement of sodium oleate to the culture medium did not restore the production of esperamicin. As a result, the authors concluded that esperamicin was synthesized *de novo* from acetate units rather than from fatty acid origin.

With evidence shown from the various labeling experiments done on the enediyne moieties, it is rational to ascribe the enediyne core to a linear polyketide origin. Furthermore, 9-membered

and 10-membered enediynes might be produced via different biosynthetic pathways due to the disparity in the labeling patterns.

1.1.3 Studies on Enediyne from a Genomic Approach

Even though more structural insights were gained through chemical synthesis and isotopic labeling of the enediyne products, little was known about the genomic aspect of the enediyne systems. Several methods have been devised in locating the polyketide synthase gene clusters, but invariably ended up in vain. In year 2000, Shen's group accidentally stumbled upon the genetic fragment containing C-1027 polyketide synthase gene cluster with probes designated for gene encoding chromophore-associated apo-protein CagA and other two that were hypothesized to be involved in production of deoxysugar moiety [49]. Meanwhile, *en route* to searching for a gene conferring calicheamicin self-resistance in *M. echinospora*, Thorson's group discovered the gene cluster responsible for calicheamicin production through extensive genomic library screening [50]. Two years later, the two gene clusters were completely sequenced and annotation was done for most of the genes based on homology [51, 52]. With a shot-gun based approach, the calicheamicin locus was found to harbor 74 open reading frames with a size greater than 90 kb. The cluster includes genes involved in polyketide and sugar moiety production, transcriptional regulators, membrane transporters, glycosyltransferases as well as proteins of unknown functions (**Figure 1.8**).

Among all the ORFs residing in the cluster, two separate iterative polyketide synthase (CalE8 and CalO5) genes were discovered. CalO5 bears astonishingly high homology to the polyketide synthase involved in the production of orsellinic acid. Therefore it was assigned to be responsible for the construction of the aryltetrasaccharide moiety. On the other hand, CalE8 is highly homologous to the polyketide synthases implicated in the biosynthesis of polyunsaturated

fatty acids. Disruption of CalE8 totally abolished calicheamicin production, which implied that CalE8 was indeed imperative for warhead construction. Within CalE8, the authors assigned the boundaries for most of the domains, including a keto-acyl synthase (KS), an acyl transferase (AT), a ketoreductase (KR) and a dehydratase (DH). Apart from that, the sequence between AT and KR was speculated to encompass an acyl carrier protein domain (ACP), and the C-terminal domain was not assigned with any function due to low homology which was later found out to be a phosphopantetheinyl transferase (PPTase) in another biochemical study [53].

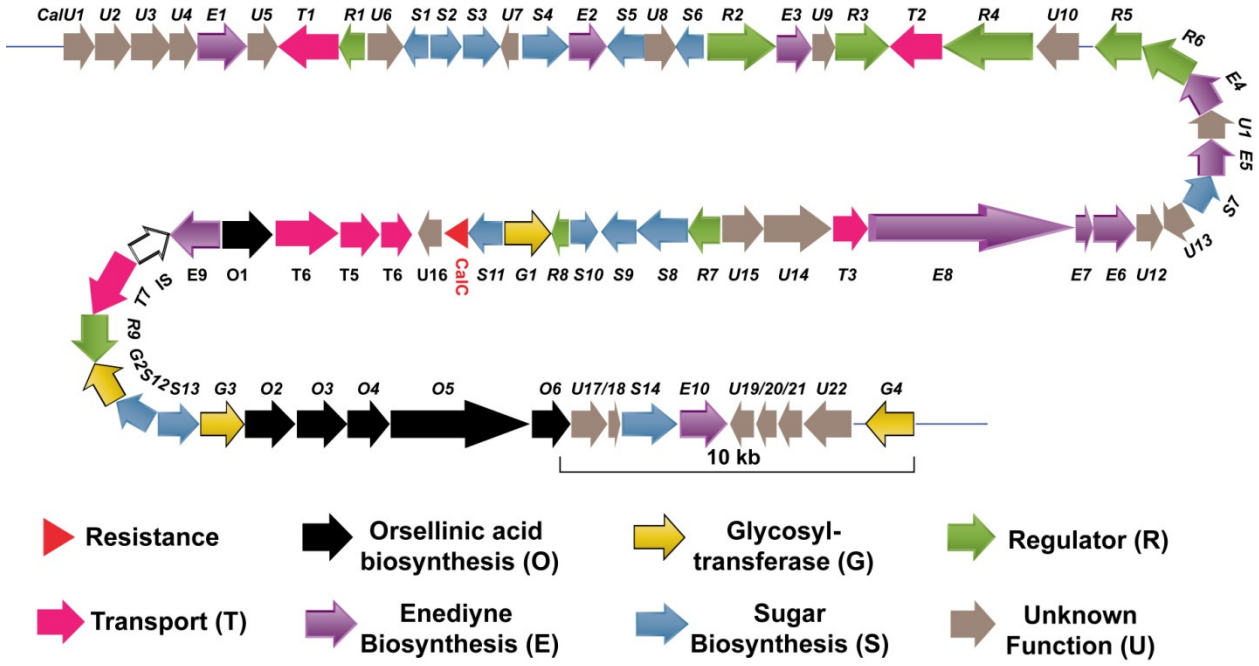


Figure 1.8 The Calicheamicin locus from *Micromonospora echinospora* spp. *Calichensis*.

Another sequencing experiment went on concurrently in Shen's laboratory revealed the gene locus for the production of C-1027. Initially an 85-kb DNA was identified which harbored 67 open reading frames. Subsequently the boundaries of the cluster were determined by gene disruption techniques, which ruled out several ORFs from both ends of the sequence as irrelevant to C-1027 production. The final assignment of the ORFs led to the discovery of genes involved

in the biosynthesis of enediyne, deoxy aminosugar, β-amino acid and benzoxazolinate. Among all the genes discovered, SgcE was hypothesized to be a polyketide synthase responsible for the production of the enediyne core. Mutation in SgcE KS domain abolished the biosynthesis of C-1027 totally, whereas complementation by a plasmid harboring a functional copy of SgcE restored the production to a level comparable to that of the wildtype. More importantly, SgcE exhibits high level of sequence homology and shares the same domain organization in comparison to CalE8. The results strongly suggested a convergent polyketide pathway for the production of both 9- and 10-membered enediyne cores.

1.1.4 Minimal Enediyne PKS Gene Cassette

Among all the ORFs, four more conserved genes beside polyketide synthase were identified in the calicheamicin and C-1027 gene clusters. Due to the fact that the two enediynes do not share any common peripheral moieties, the conserved genes are regarded as the essential components in assembly of the enediyne core. One of the four genes was proposed to encode a thioesterase based on sequence homology, but the rest of the genes apparently did not share similarity with any of the genes with known functions in the database. Collectively, the four conserved genes together with PKS are known as the "minimal enediyne PKS gene cassette" or "warhead gene cassette" [54]. With the foundation laid down for genomic screening of enediynes, research on enediyne biosynthetic pathway marched forward rapidly. Within a few years, the identification of gene clusters for neocarzinostatin, maduropeptin and dynemicin was achieved. Unsurprisingly the presence of minimal gene cassette was confirmed for all the enediyne biosynthesis clusters [55-57]. Moreover, a high throughput genomic scanning method has been devised by Thorson's group which unveiled several gene loci containing the gene cassette in soil-dwelling actinomycetes [54]. Apart from that, the ubiquitous presence of the

minimal enediyne PKS gene cassette in a plethora of organisms was further substantiated when it was discover in marine bacterium *Salinispora tropica*. Among the two minimal gene cassettes identified within the genome of *Salinispora tropica*, one of them was confirmed to play a vital role in the biosynthesis of the enediyne precursor of sporolides, while the other was hypothesized to dictate the production of a 10-membered enediyne based on its phylogenetic relationship with other PKS genes [58, 59].

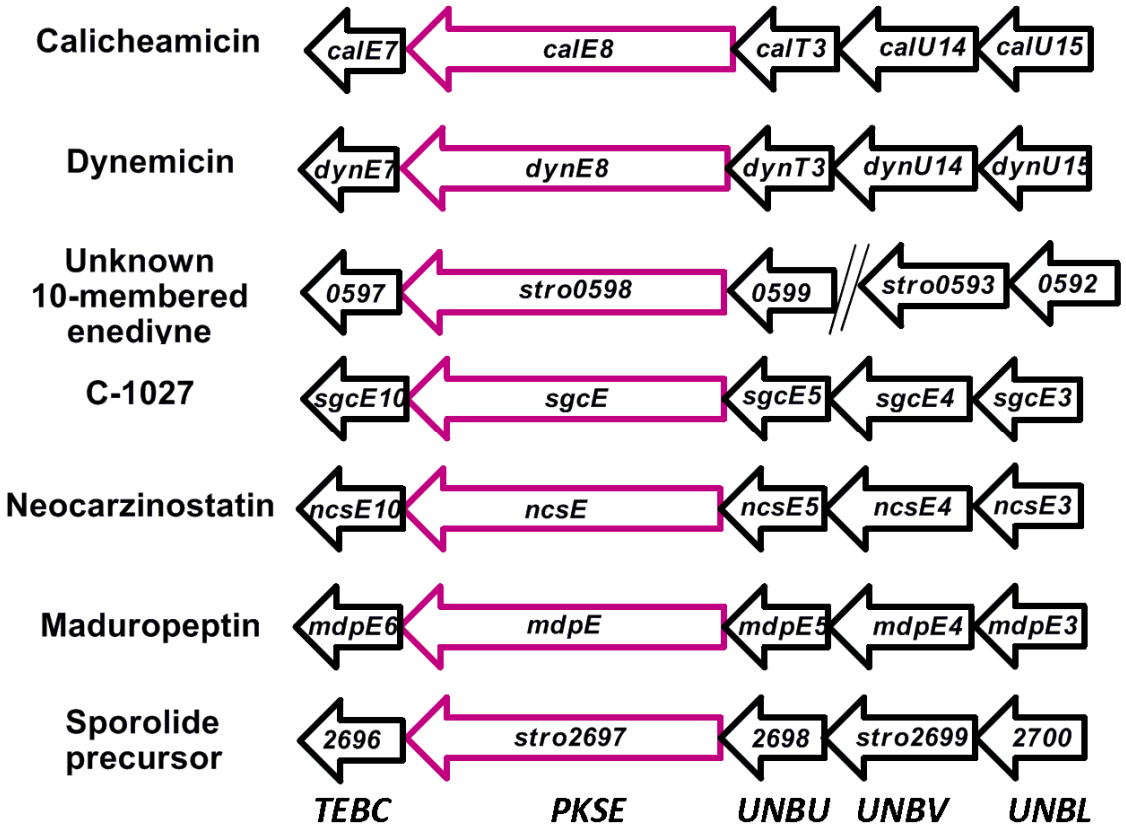


Figure 1.9 Conserved warhead gene cassettes from different enediyne gene clusters. Organization of the highly conserved warhead gene cassettes with the top 3 clusters from the 10-membered enediyne family and the lower 4 from the 9-membered enediyne family.

In the minimal cassette, the five conserved genes were arranged in a hypothetical operon with unidirectional transcription. Moreover, overlap in translational start and stop codons are common among the five genes, implying that the expression of the five genes is highly

coordinated and their protein products should be involved in performing related functions in the bacteria. The five conserved genes were assigned as *PKSE*, *TEBC*, *UNBL*, *UNBV* and *UNBU* (**Figure 1.9**).

As discussed in the section above, *PKSE* encodes PKS which is the major enzyme involved in the production of the polyketide precursor for enediyne core. After comparing with the known proteins in the databases, protein encoded by *TEBC* was found to be homologous to the 4-hydroxybenzoyl-CoA thioesterase of *Pseudomonas* sp. Strain CBS-3. Thus, protein from *TEBC* family was speculated to catalyze the acyl chain release in polyketide production and also have a possible role in the chain cyclization. The other three families of proteins, namely *UNBL*, *UNBV* and *UNBU*, bear no significant homology to any of the proteins with known functions in the database. As a result, those families of proteins are proposed to carry out functions unique to enediyne biosynthesis. From structural perspective, proteins in *UNBV* family are likely to be secreted since they all carry N-terminal signal sequences. On the other hand, proteins from *UNBU* family are speculated to be imbedded in the membranes due to the fact that they contain 7 to 8 putative membrane spanning α-helices. Lastly, the proteins from *UNBL* family bear very low homology to some of the desaturases found in the database, which leads to the hypothesis that they might play a part in the formation of the triple bonds in the enediyne core. However, the speculations are purely based on bioinformatics studies, which require further substantiation from either *in vivo* or *in vitro* biochemical experiments.

The genomic advancement in enediyne biosynthesis opens up a new realm for researchers who are eager to get a glimpse of the biosynthetic mechanism of enediynes. Engineering of the biosynthetic pathway for enediyne production also becomes possible which provides us with opportunities for structural manipulation of enediyne products to cater for different applications.

1.1.5 The Iterative Type I Polyketide Synthase for Enediyne Production

Polyketide synthases (PKSs) are closely related to fatty acid synthases (FASs) and responsible for producing a myriad of secondary metabolites. In recent years, the research on PKSs flourished due to their intriguingly complex structure and functionality as well as the vast commercial value that can be exploited. Conventionally, polyketide synthases have been categorized into type I, II and III PKSs. Among the three types of PKSs, modular, non-iterative PKSs (type I) produce aliphatic products, iterative PKSs (type II) synthesize aromatic compounds and type III PKSs produce monocyclic aromatic polyketide in the absence of acyl carrier protein [60, 61] (**Figure 1.10**). Type I PKSs are comprised of several modules, with each of them responsible for a distinct set of enzymatic functions to complete one cycle of the elongation process. All the modules are generally only used once during the elongation process, although stuttering or skipping of modules does take place in some type I PKSs. DEBS (6-deoxyerythromycin B synthase), is a typical type I PKS that accounts for the biosynthesis of reduced polyketide product erythromycin A [62]. On the other hand, type II PKSs only consist of a single set of enzymatic domains that are used iteratively to produce the polyketides, such as tetracenomycin [63]. Lastly, type III PKSs are also known as chalcone synthase-like PKSs, which catalyze condensation iteratively to produce mono- or bi-cyclic aromatic products independent of ACP. A typical example for type III PKS is RppA synthase which is responsible for the production of flavolin [64]. Despite the differences in the three types of PKSs, the individual catalytic domains are highly similar among them. Besides, all three types of PKSs assemble acyl chains by decarboxylative condensation mechanism from acyl-CoA units.

However, the existence of enediyne PKSs has blurred the boundaries for PKS classification, whereby the novel PKSs display structures and functions well beyond the paradigm of type I,

type II and type III PKSs. Astonishingly, the enediyne PKS is more like a hybrid between the type I and type II PKSs. It acts iteratively but produces aliphatic polyketides instead. As a result, it provides scientists with a perfect opportunity to revise their perception on bacterial PKSs as well as to further appreciate the beauty of PKSs in terms of molecular diversity and biosynthetic complexity.

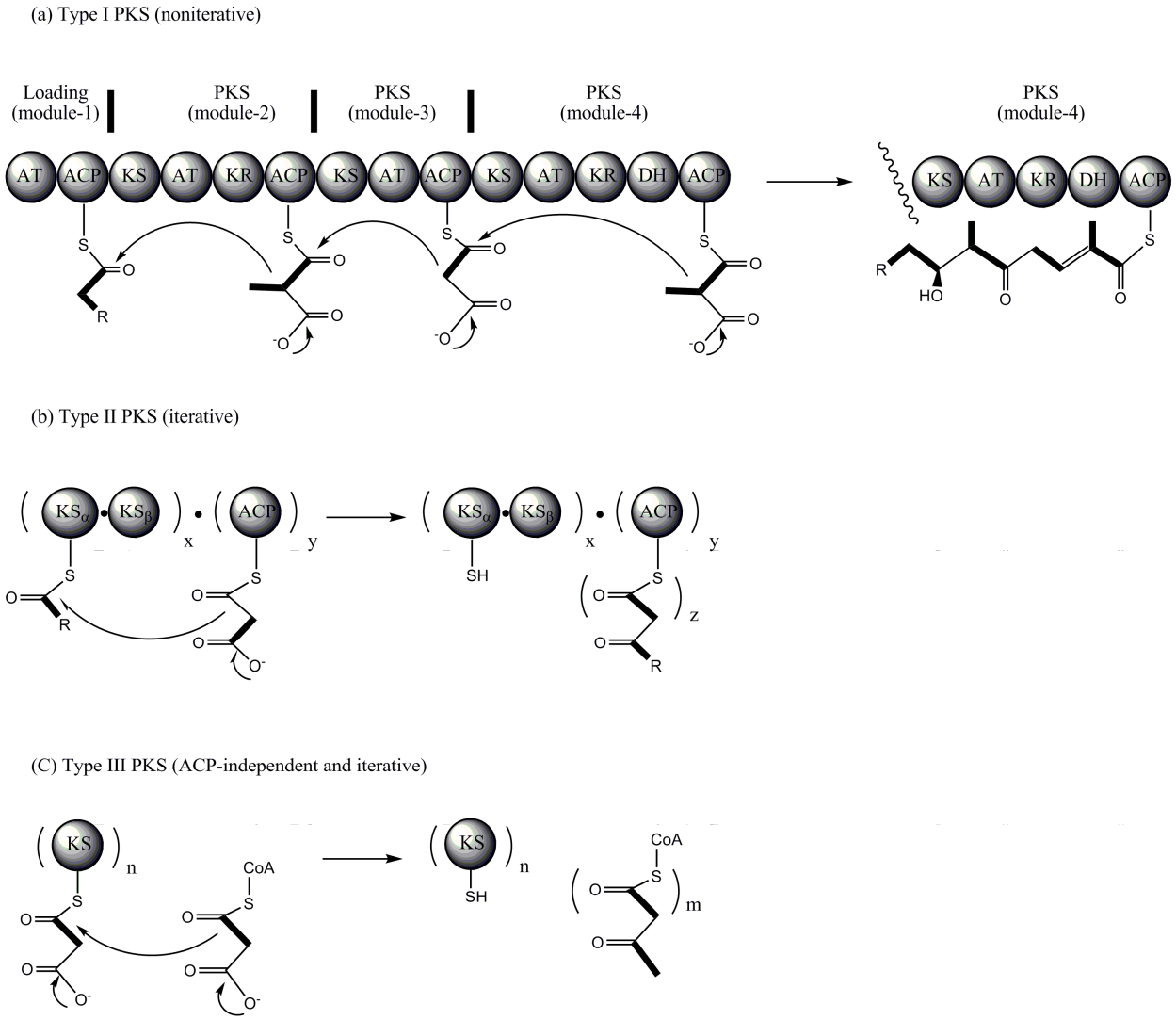


Figure 1.10 A simplified schematic diagram for the three types of PKS. (a) Type I PKS which is modular and non-iterative; (b) Type II iterative PKS; (c) Type III PKS which is iterative and ACP-independent. Figure was adapted from Shen, B., Current Opinion in Chemical Biology, 2003. **7**(2): p. 285-295. [60]

Biosynthesis of polyketide in fungi is marked by the presence of iterative type I PKSs, which lead to the production of aromatic polyketides like 6-methylsalicyclic acid. On the contrary, the aromatic polyketide synthesis is under the regime of iterative type II PKSs in bacteria. In 1997, the first bacterial iterative type I PKS was uncovered from *Streptomyces viridochromgens*. The PKS is termed AviM, which is responsible for the production of the aromatic avilamycin [65]. The domain composition of AviM is typical of type I PKSs, including KS, AT, DH and ACP. Instead of producing aliphatic polyketide, AviM was proposed to be involved in aromatic polyketide synthesis. In a series of heterologous expression experiments, the presence of AviM gene in the transformed hosts in fact led to the production of orsellinic acid, which was the only plausible component of polyketide origin in avilamycin. However, the authors suggested that AviM might be an exception to the type II bacterial PKSs.

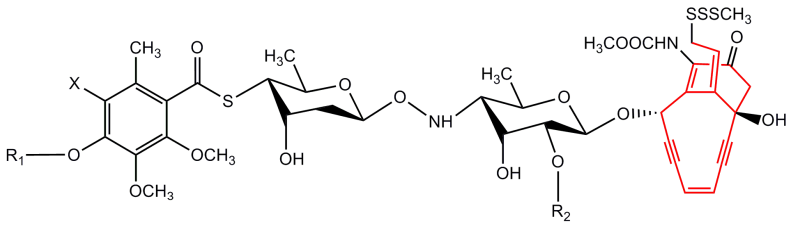
The recent discovery of calicheamicin gene cluster finally toppled the conventional classification of PKSs with the identification of two iterative polyketide synthases. As discussed previously, the two PKS genes, namely CalE8 and CalO5, are responsible for the production of the enediyne moiety and orsellinic acid component of the antibiotic calicheamicin respectively. Both PKSs demonstrate signature characteristics of iterative type I PKSs. More astonishingly, CalO5 displays the same domain organization and bears high sequence homology to that of AviM. The discovery of more iterative type I PKSs serves as manifestation of the vast diversity and complexity of the polyketide synthases which is impossible to be covered by a simplistic classification of only three types of PKSs. Furthermore, function of iterative type I PKS is not only restricted to the synthesis of monocyclic aromatic polyketides. The discovery of neocarzinostatin gene cluster ushered the introduction of another iterative type I PKS— NcsB [55]. NcsB is capable of synthesizing higher-order aromatic polyketides like naphthalinic acid.

Besides, cloning of NcsB gene invariably failed in the earlier days because of the type II PKS probes being used. Even though the horizon for PKS has been widened with the discovery of iterative type I PKSs, the vast versatility of PKSs developed during the course of evolution still awaits to be further explored.

1.2 Calicheamicins

1.2.1 Structure and Functionality of Calicheamicins

Among all the 10-membered enediynes that have been characterized, calicheamicins, produced by *Micromonospora echinospora ssp. calichensis*, are some of the earliest to have their structures deciphered (**Figure 1.11**). The compounds were first identified through an experiment originally meant to screen for products from microbial fermentation that were active in induction assay.



Calicheamicin	X	R_1	R_2	R_3
Calicheamicin β_1^{Br}	Br	Rha	Ami	$CH(CH_3)_2$
Calicheamicin γ_1^{Br}	Br	Rha	Ami	C_2H_5
Calicheamicin α_1^{I}	I	H	Ami	C_2H_5
Calicheamicin α_3^{I}	I	Rha	H	
Calicheamicin β_1^{I}	I	Rha	Ami	$CH(CH_3)_2$
Calicheamicin γ_1^{I}	I	Rha	Ami	C_2H_5
Calicheamicin δ_1^{I}	I	Rha	Ami	CH_3

Figure 1.11 Members of the calicheamicin family.

From the culture broth supplemented with bromide and iodide salts, seven different members of the calicheamicin family were isolated and identified. The iodine derivatives were obtained with much high yield compared to that of bromide-containing calicheamicins [66]. Among all the derivatives, calicheamicin γ_1^I was the most prominent member in the family. The calicheamicin compounds have very high activity in terms of biochemical induction assay even at a concentration as low as 1 pg/ml. Furthermore, it was found to be extraordinarily effective against Gram-positive bacteria and considerably active against Gram-negative bacteria. Apart from its antibacterial properties, calicheamicins themselves carry an intrinsically high efficacy against P338 murine tumors, L1210 leukemia as well as solid neoplasms like B-16 melanoma and colon 26. The optimal dosage was expected to be around 0.15-5 μg/kg [67]. Similar to other enediynes, calicheamicins work by causing sequence-specific double strand cleavage which will in turn induce cell apoptosis [68].

Structurally, calicheamicins consist of two parts— the aglycone constituent and the glycone constituent, with each of them performing distinct and imperative tasks in cleaving double-stranded DNAs. The glycone component features a massive sugar residue consisting of four monosaccharides and a substituted benzene ring joined together in a highly unusual fashion. Functionally the glycone constituent is employed in the delivery of the molecule to its DNA targets. Once in close proximity of DNA duplex, it will bind tightly to the minor groove of the DNA helices with very high specificity towards sequences such as 5'-TCCT-3' and 5'-TTTT-3' through strong hydrophobic interactions [19, 69-72]. In the case of DNA recognition, the oligosaccharide is thought to undergo considerable reorganization to afford a rigid and extended conformation. This change in configuration in turn allows the unconventional hydroxylamine linker to assume a specific torsion angle in the middle of calicheamicin, which is complementary

to the shape of the minor groove. Apart from that, DNA helices undergo conformational changes upon binding as well. The minor groove of the helices is further opened up to accommodate the full binding of calicheamicin molecule via an induced-fit mechanism. Given that the specificity of interaction is dictated by the overall hydrophobicity of the molecule, both the aglycone and glycone are supposed to play a part in the recognition process. Through molecular modeling of calicheamicin γ_1^I, it was suggested that the sequence specificity of the molecule, especially for 5'-TCCT-3', might be attributed to the interaction between polarized iodo-substituent on the hexa-substituted benzene ring and the exocyclic amino groups of the two guanines from the complementary strand of the DNA [73]. Furthermore, the hypothesis of iodine playing a critical role in the DNA recognition was proven in the subsequent years through a series of DNA footprinting experiments with synthetic glycone analogs [74]. Continued study on the sequence specificity of calicheamicins led to the realization that the preference for pyrimidine-rich region could be due to the greater susceptibility for distortion of the DNA sequence. Hence the accommodation of the molecule in the minor groove could be further facilitated [75, 76].

On the other hand, the aglycone part, which is better known as calicheamicinone, is comprised of a bicyclic core with two functionally critical *–yne* groups harbored within the 10-membered ring. The enediyne core will be activated to undergo cycloaromatization under appropriate conditions [77]. The triggering device— an allylic trisulfide, which is part of the aglycone, will result in the formation of a thiol once a nucleophile attacks the central sulphur atom. Subsequently the thiol will react with the adjoining α, β-unsaturated ketone, resulting in the formation of a tetragonal center via hetero-Michael addition. The extensive change in molecular configuration and structural geometry imposed an enormous amount of strain on the 10-membered enediyne core. After that, cycloaromatization of the enediyne core ensues which

eventually relieves the geometrical strain and leaves behind the extremely reactive 1, 4-benzenoid diradical as the product. The diradical then abstracts hydrogen from the DNA helices at C(5') position of the cytidine (the first C from 5'-TCCT-3') and the C(4') position of the nucleotide three base pair away on the 3' end of the complementary strand, resulting in double-stranded cleavage of DNA [78-80] (**Figure 1.12**).

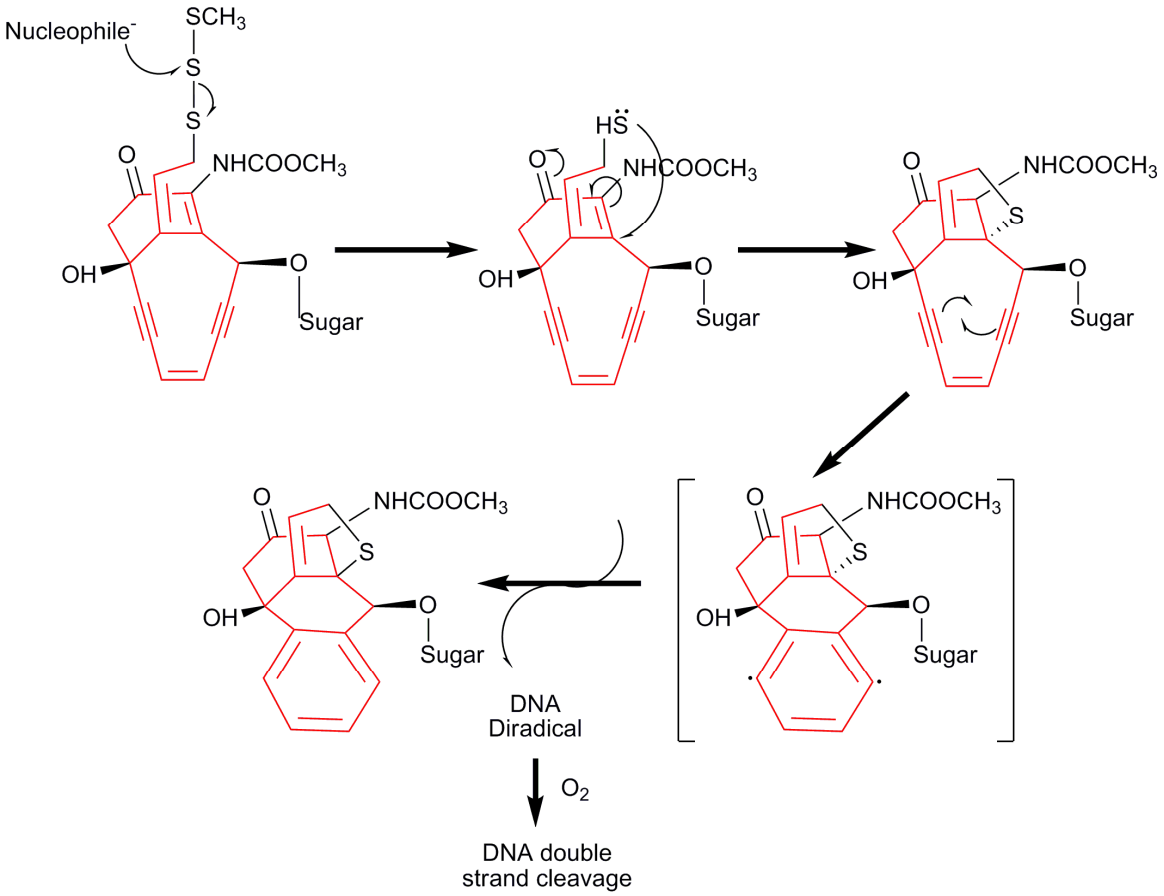


Figure 1.12 DNA-cleaving mechanism of calicheamicin through Bergman cyclization after thiol activation and subsequent hydrogen abstraction on DNA molecule.

Careful kinetics measurement implicated that a series of complex process might be involved in calicheamicin-DNA interaction and finally the cleavage of the double strand. Firstly, calicheamicin interacts and binds to DNA. In the DNA-bound form, the molecule forms adduct with glutathione (**Figure 1.13**). The adduct formed then break off from the DNA strand and

further react with glutathione to give rise to the activated form of calicheamicin. In its activated state, the calicheamicin intermediate re-associates itself with DNA strands and undergoes a series of conformational changes to yield the diradical. Eventually the diradical formed will finish off its designated task by abstracting hydrogen from DNA double helix and trigger the double-stranded cleavage. Up to date, the structural and functional aspects of calicheamicin were well deciphered from NMR data, biochemical experiments and total chemical synthesis.

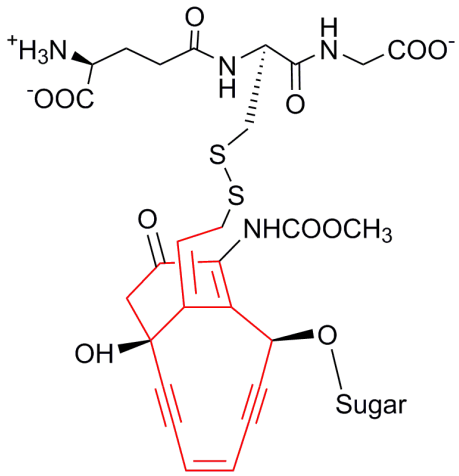


Figure 1.13 Calicheamicin-glutathione adduct.

1.2.2 Self-Resistance Mechanism

Since the discovery of calicheamicin and other enediynes, scientists simply could not stop marveling the architectural and functional ingenuity of the molecule. Through various biochemical experiments, it has been shown that calicheamicin is an extremely potent DNA-cleaving agent. The lethal dosage of calicheamicin for most of organisms falls within femtomolar range. The amazing fact that *Micromonospora echinospora* is capable of surviving the high concentration of calicheamicin produced by itself has roused great interest in the scientific world. With the help of genomic cosmid library screening, a gene responsible for self-resistance

mechanism in *Micromonospora echinospora* was identified. Through heterologous expression of the gene in *E. coli*, a non-heme iron metalloprotein CalC was obtained [50]. Following that, biochemical assay was carried out to test the inhibitory effect of CalC towards the DNA-cleavage activity of calicheamicin. It was found that the addition of CalC reconstituted with Fe^{2+} or Fe^{3+} could block DNA cleavage by reductively activated calicheamicin. With the data from spectrophotometry and biochemical assay collectively, a few hypotheses were proposed to explain the self-resistance mechanism of CalC. For instance, the catalysis could be carried out through radical disproportionation comparable to that of bacterial iron superoxide dismutases. Apart from that, the possibility of interference with DNA-binding, or reductive activation, or cycloaromatization by CalC was also considered in light of the functionality of known mononuclear iron-containing oxygenases. Lastly, CalC might simply sequester the enediyne in a way similar to its 9-membered apo-protein counterparts. However, the dispute over the self-resistance mechanism of CalC did not last long; Thorson's group conducted a series of brilliantly designed experiments a couple of years later that demonstrated the novel mechanism of self-resistance to calicheamicin by CalC protein [81]. A more quantitative molecular break-lights assay was devised which allowed Thorson and his coworkers to monitor DNA-cleavage by calicheamicin in real time. Titration of CalC into the assay solution established that the inhibition of cleavage was concentration-dependent. CalC with equivalent concentration to that of calicheamicin actually abolished DNA scission completely. Furthermore, the inhibitory effect of CalC was specific towards calicheamicin and its close relatives namenamicin and shishijimicin, whereas in the case of esperamicin and dynemicin, no cross-resistance was observed. The disparity in efficiency of inhibition on the 10-membered enediynes by CalC was attributed to the subtle difference in the warhead structures. Besides, the attempt of detecting interaction between

CalC and calicheamicin failed which ruled out the possibility of CalC being a sequestration protein for the enediyne.

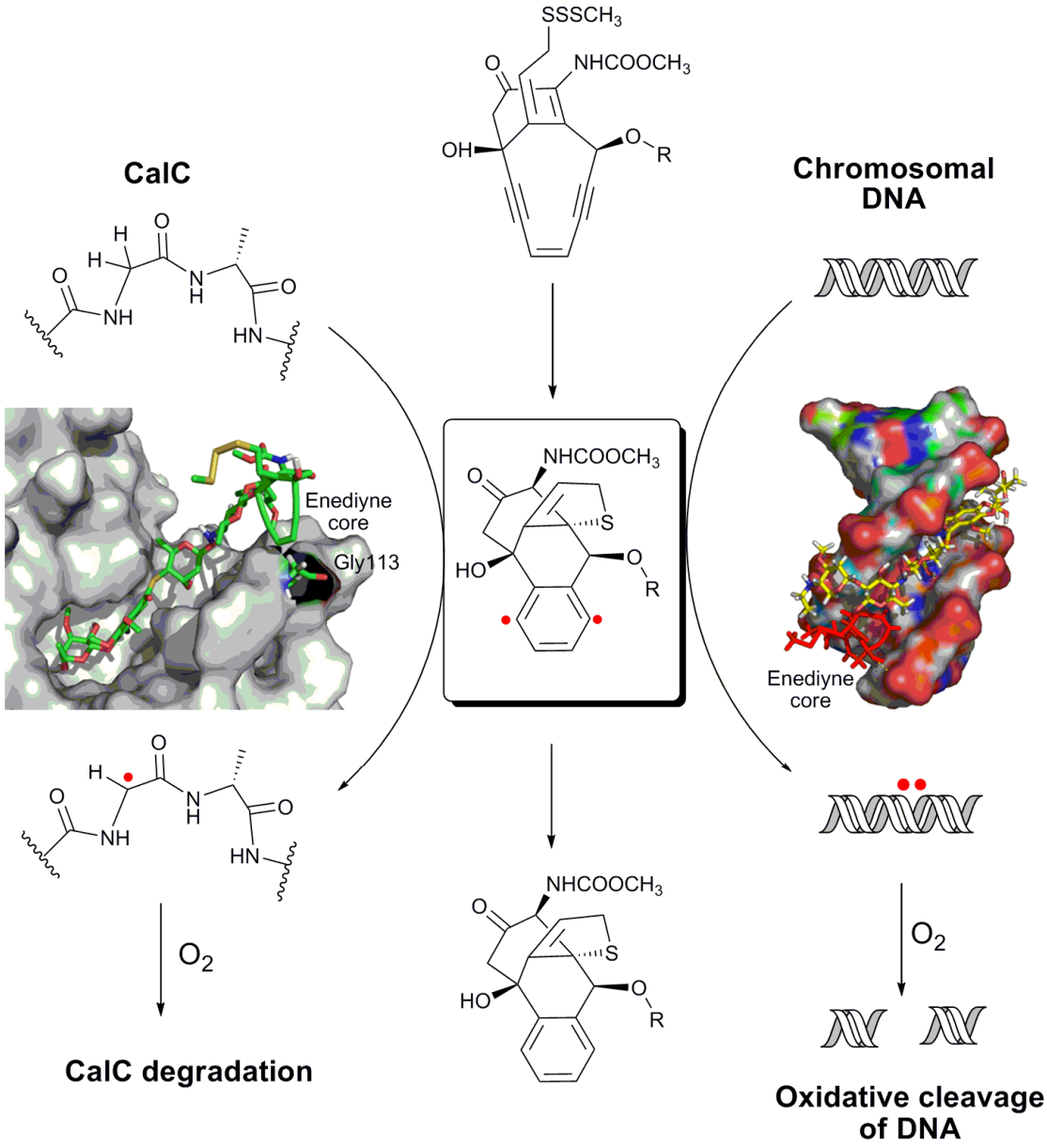


Figure 1.14 The sacrificial protection by CalC. In the presence of oxygen, activated calicheamicin readily abstracts hydrogen from the backbone of gly[113] in CalC. Through a series of radical-mediated reactions, the protein is cleaved into two distinct fragments at the expense of 1, 4-benzenoid diradicals.

However, when reductively activated calicheamicin was used in the assay, CalC was observed to undergo specific proteolysis which gave rise to two peptides of distinct lengths. As a result, it is possible that CalC sacrifices itself in place of DNA for the hydrogen abstraction after reductive activation of calicheamicin. Furthermore, the mutational studies on the residue preceding the cleavage site—specifically glycine[113], concluded that the residue plays a crucial role in hydrogen abstraction process. In the presence of oxygen, activated calicheamicin readily abstracts hydrogen from the backbone of gly[113]. Through a series of radical-mediated reactions, the protein is cleaved into two distinct fragments at the expense of 1, 4-benzenoid diradicals. The capture by CalC is hypothesized to take place in between thiol-calicheamicin adduct formation and hetero-Michael addition during which dissociation of adduct from DNA molecule allows the access of CalC to the reductively activated calicheamicin. In the context of within *Micromonospora echinospora*, CalC is proposed to serve as the safeguard for the stringently controlled enediyne production and transportation pathway. Any of the calicheamicin molecules that accidentally escaped sequestration into the cytoplasm will be readily deactivated by CalC through sacrificial mechanism (**Figure 1.14**). From the author's point of view, the seemingly "wasteful" sacrificial detoxification is absolutely indispensable for the survival of the bacteria given the extreme potency of the enediyne secondary metabolite.

1.2.3 Medical Application of Calicheamicins— MyloTarg™

Due to its rigorous DNA-cleaving attributes, DNAs are naturally the primary target for calicheamicin action inside cells. On the other hand, there is little record for damage inflicted on the other components of the cell by calicheamicin. According to previous studies on human diploid fibroblasts, calicheamicin is capable of dealing serious damage to chromosome at range as low as pg/ml. Other studies on HeLa cells showed that calicheamicin will block 50% of DNA

synthesis at a concentration of 50 pg/ml. Further increase in concentration of calicheamicin to around 15 ng/ml will result in the loss of 50% RNA synthetic activity. Besides, calicheamicin is capable of causing mitochondrial DNA damage as well. Once under the effect of calicheamicin, DNAs are damaged at such an alarming rate which far outpaces the DNA repair capacity of the cell. Consequently, majority of the energy production will be channeled from the maintenance of normal cellular activities into DNA repairing which eventually leads to demise of the cell due to energy deprivation. Moreover, it has been shown in various experiments that calicheamicin and several aryltetrasaccharide derivatives are able to disrupt protein-DNA interactions [82, 83]. For most of the compounds, they might have behaved differently in *in vitro* and *in vivo* experiments. However, a remarkable consistency is observed for calicheamicin in inducing DNA-scission both *in vitro* and *in vivo*, which makes it a perfect candidate for medical engineering in cancer treatment.

Conventional chemotherapy is always accompanied with serious side effects because of the lack of specificity for tumor cells. Tumor and normal cells are wiped out alike after administration since most of the drugs do not make a distinction between the two. Minimizing the side effects of cancer drugs has been one of the top priorities for medical researchers. Due to the extraordinary potency of calicheamicin, it is more desirable to deliver the drug directly to the affected area in the body or conjugate the drug to antibodies/polymers to enhance the specificity of the drug towards tumor cells. Monoclonal antibody is selected as one of the preferred conjugates for calicheamicin ascribed to its high specificity for cell surface antigens. The disulfide chemistry of calicheamicin is employed in the conjugation process whereby monoclonal antibodies are attached [84]. Most of the monoclonal antibodies discovered so far target tumor cell surface antigens which are commonly found on normal cells as well. Thus

administration must be spatially and temporally controlled so that the drugs can be directed to a specific group of cells with a particular antigen expressed for maximal efficacy and minimal side effects.

MyloTarg^{TM}, which is also known as Gemtuzumab Ozogamicin, consists of calicheamicin conjugated to anti-CD33 antibody. The drug is being utilized in the treatment of acute myeloid leukemia (AML). In blood, only matured myeloid cells express distinct surface markers like CD33 unlike the pluripotent hematopoietic stem cells [85-87]. The expression of surface antigen is an important process for blood cell maturation since it signals the further differentiation of the stem cells into progenitors with specific functions. 90% of patients suffering from AML have CD33 expressed on AML blast cells. Hence, CD33 serves as an excellent target for the binding of monoclonal antibodies. It has been shown that selective ablation of cells expressing CD33 from bone marrow aspirates of leukemia patient in fact induced the differentiation of hematopoietic stem cells into normal blood cells [88, 89]. In light of this, clinical trials have been carried out to test the effectiveness of monoclonal anti-CD33 antibodies. Following the administration of the antibodies, the CD33 sites were readily saturated and the internalization of the antigen-antibody complex ensued [90, 91]. In the following phase, calicheamicin was attached to the anti-CD33 antibodies which marked the inception of the drug gemtuzumab ozogamicin. Subsequently phase I clinical trials were carried out at hospitals to study the effect of the drug on leukemia [92]. According to the data collected, the drug worked well for the treatment of AML with only slight side effects like fever and chills, which were commonplace among all the chemotherapeutic drugs.

In the phase II studies, it was established that the drug was effective in treatment of patients suffering from AML relapse. More importantly, the drug did not trigger immune response in any

of the patients, unlike other immune-conjugates which often elicited severe immune response in the body after administration. Since the drug carries with itself very little non-hematologic toxicity, it is generally regarded as a decent replacement for anthracycline in combinatorial chemotherapy as well as preparative medication for hematopoietic stem cell transplant.

1.2.4 Other Medical Applications for Calicheamicins

Mortality rate of ovarian cancer is usually quite high due to the fact that there are no obvious symptoms linked to it at early stage of tumor development. Thus accurate diagnosis at early stage is difficult and by the time symptoms surface, metastasis may have occurred already. The general treatment involves surgical removal of the ovary followed by a combination of radio- and chemotherapy. It has been shown that most of ovarian cancer cases were linked with an overexpression of MUC-1 (polymorphic epithelial mucin) in an abberantly-glycosylated form on the cell surface. Similar to that of CD33, the MUC-1 antigen is capable of being internalized as well upon binding of antibodies [93]. Therefore, MUC-1 was chosen as the primary target for monoclonal antibody recognition. Through various experiments, it was found that one of the antibodies used— hCTMO1, was well capable of delivering the therapeutic agents into the cancer cells after binding of the antibody to surface MUC-1 antigen took place. Immediately following the initial success of identifying the desired antibody, calicheamicin was conjugated to hCTMO1 to form the experimental drug CMB-401. Phase I clinical trials have been conducted with ovarian cancer patients who were refractory to conventional platinum therapy. Most of the patients administered with the drug only suffered from minor side effects like nausea and vomiting. In general the drug was well tolerated in the body of the patients. A proportion of the patients experienced attenuation in severity of the cancer while no clinically significant allergic reaction was detected in any of the patients. Based on the phase I clinical trial results, phase II

studies have been initiated to further characterize the efficacy of the drug in treatment of ovarian cancers.

Besides the compounds mentioned above, CMC-544 (also known as inotuzumab ozogamicin) is another potential antibody-conjugated calicheamicin drug for the treatment of non-Hodgkin's lymphoma [94]. Currently clinical trials are being carried out to test the efficacy of the drug on the patients. Collectively, drug leads designed based on calicheamicin have sparked off a bright future for research on antibody-conjugated enediyne products in treatment of cancers.

1.3 Objective and Organization of the Thesis

Since the discovery of the minimal gene cassettes for enediyne biosynthesis, attention has been drawn towards the biosynthetic pathway of the structurally complex enediyne core. Our main objective is to decipher the biosynthetic pathway of calicheamicin enediyne core as well as characterize the structure and function of the biosynthetic enzymes. CalE8 as the sole polyketide synthase in the minimal gene cassette, its precise role in the production of the polyketide precursor for calicheamicin core is still unknown. Despite all the data from the radio-isotopic labeling experiments that provided us with hints on the folding pattern of the enediyne cores, little was known about the polyketide precursors of the bicyclic rings and how the precursors are actually assembled from the acetate units. In light of this, our laboratory set out to explore the function of CalE8 and its auxiliary proteins through a series of *in vitro* and *in vivo* experiments. Three proteins encoded by the minimal cassette, namely CalE8, CalE7, CalU15 will be expressed by using the heterologous *E. coli* expression system. Subsequently their *in vivo* and *in vitro* enzymatic activities will be examined. Supplied with the cognate substrates under appropriate experimental conditions, CalE8 may produce the polyketide precursor or even the crude enediyne core itself with the help from its accessory proteins. Once identified with techniques like high resolution mass spectrometry and NMR, the polyketide intermediate could bridge the gap between polyketide production and maturation of the polyketide into the enediyne core. Moreover, as an iterative PKS that features a unique domain organization and contains an integrated PPTase domain, CalE8 provides us with a model system to study the structure and function of the type I iterative PKSs. At the same time, the comparative study of the homologous enediyne PKSs (CalE8, SgcE and DynE8) will further help us understand the differences among the homologous PKSs as well as their closely-related evolutionary lineage.

Cloning of the individual domains of CalE8 is also part of our research plan. We hope to examine the individual activity of the KS, AT, ACP, KR, DH and PPTase as isolated domains. By studying the structure and function of the individual domain, the information can be pieced together to provide us with a detailed mechanistic view of the iterative type I PKS.

In addition, we hope that the study on the putative thioesterase CalE7 will shed light on the structural and functional aspects of the novel enzyme as well. It is not known whether CalE7 functions as a type I or type II thioesterase. Due to the low homology of CalE7 to any known thioesterase in the database, it is almost impossible to deduce the function of CalE7 solely based on bioinformatics studies. Thus, structural studies will provide us with a new perspective on the function of CalE7. With the structure of CalE7, it is possible to probe the catalytic mechanism of the thioesterase and eventually its classification in the realm of thioesterases.

This dissertation emphasizes on the structural and functional characterization of enediyne PKS CalE8. In addition, the NMR solution structure of the ACP domain of CalE8 and the crystal structure of the thioesterases CalE7 (and the homolog DynE7) were determined for the purpose of function elucidation. Chapter 1 is an overview on the enediyne products and some of the recent developments in enediyne-related researches. Chapter 2 describes our effort in functional characterization of CalE8 through both *in vitro* and *in vivo* enzymatic assays. The products generated in the enzymatic assays were isolated and identified using various techniques. Subsequently, hypothesis on the biosynthetic pathway of the enediyne was formulated within the frame of knowledge acquired. As a sequel to the study on CalE8, chapter 3 focused on the structural characterization of ACP domain of CalE8. With data acquired from NMR spectroscopic studies, some interesting features of the ACP domain in the polyketide synthesis are discussed. Chapter 4 revolves around the functional and structural characterization of hotdog

fold thioesterase CalE7 and the homolog DynE7. Through mutagenesis of potential catalytically critical residues, important residues in the hydrolytic mechanism are identified. Lastly, chapter 5 concludes with a summary of the findings and also discusses about the future directions of the research project.

CHAPTER 2 Products of the Iterative Polyketide Synthase CalE8 in Biosynthesis of Calicheamicin

2.1 Introduction

With the two papers published by Thorson and Shen in 2002, a milestone was reached with the successful sequencing of the calicheamicin and C-1027 gene clusters [51, 52]. Moreover, the discovery of CalE8 and SgcE expanded the realm of iterative type I PKS further, by manifesting the capability of the two PKSs in producing enediyne cores. The subsequent devise of a high throughput screening method allowed Thorson's group to discover a plethora of gene clusters containing similar enediyne PKSs [54, 58]. Bioinformatics analysis of the enediyne PKS predicted the presence of at least six functional domains within the megasynthase (**Figure 2.1**).

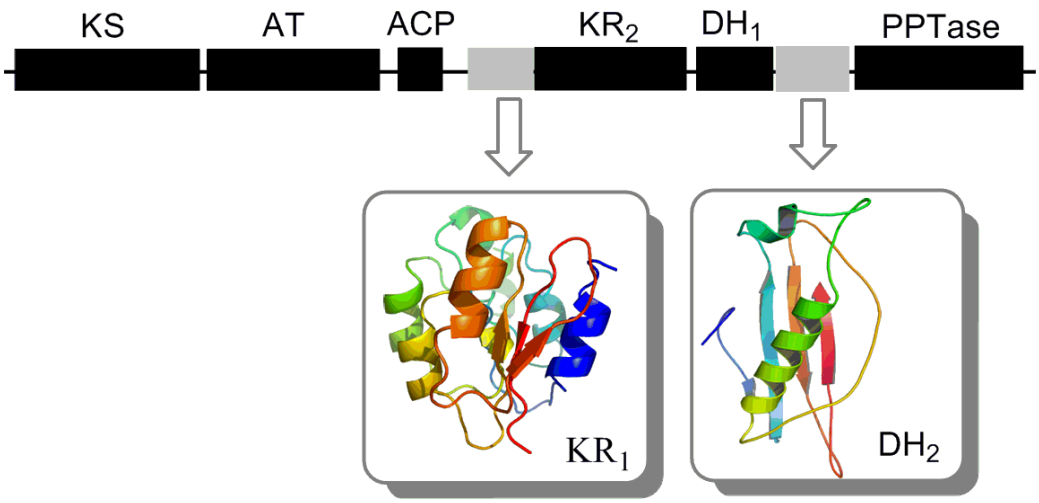


Figure 2.1 Domain composition of enediyne PKS. From N- to C-terminus (left to right): KS (ketosynthase), AT (acyltransferase), ACP (acyl carrier protein), KR (ketoreductase), KR1 is an unknown domain preceding KR domain (KR2), which is supsected to be part of KR, DH (dehydratase), DH2 is an unknown domain following DH domain (DH1), which is suspected to be part of DH, and PPTase (phosphopantetheinyl transferase).

Among the six putative domains within the enediyne PKS, an intrinsic phosphopantetheinyl transferase (PPTase) domain, instead of a stand-alone PPTase protein in other bacteria PKS system, is found to modify the ACP domain by phosphopantetheinylation [53, 54, 95]. After modification, the phosphopantetheine moiety covalently linked to the acyl carrier protein (ACP domain) functions as a flexible arm that carries the extending chain and swings around to access the various enzymatic domains for chain processing. The acyltransferase domain (AT), helps to load the ACP domain with acyl units from either acetyl-CoA or malonyl-CoA pools.

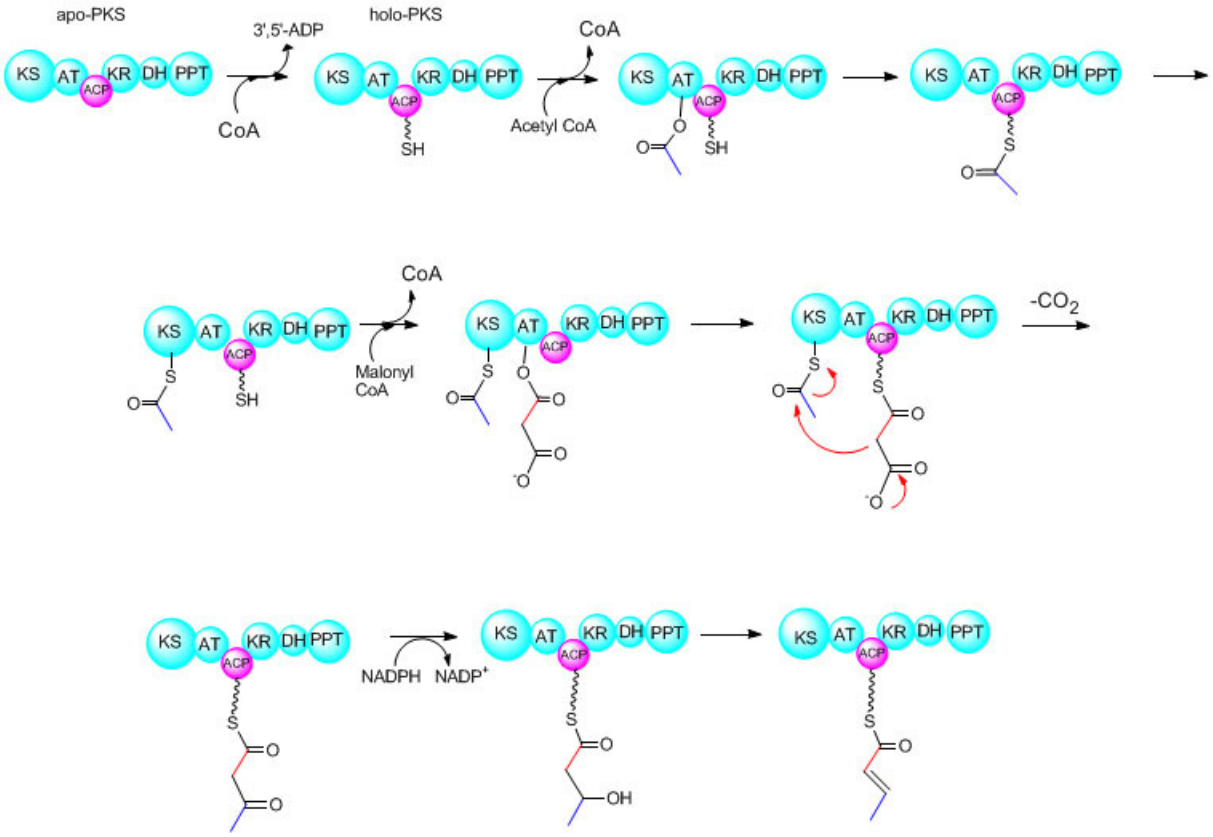


Figure 2.2 Illustration of individual steps of iterative polyketide synthesis. 1). PPTase modifies the ACP domain by phosphopantetheinylation using CoA. 2). Acetyl-CoA starter is loaded onto ACP by AT domain via thioester linkage. 3). Acyl starter is transferred to KS while the AT loads ACP with malonyl-CoA as extender. 4). KS mediates the condensation between the starter and extender molecules. 5). As elongation goes on, KR reduces subsequent keto groups to corresponding hydroxyl groups, which are then removed in desaturation mediated by DH domain to form conjugated double-bond system. 6). The elongation process is stopped when a designated chain length is reached and the final product is released by external thioesterases.

Playing a key role in chain elongation, ketosynthase (KS) domain catalyzes the condensation step to elongate the acyl chain, whereas the ketoreductase (KR) domain is likely to be responsible for the reduction of keto groups to the corresponding hydroxyl groups on the polyketide precursor. Subsequently, the dehydratase (DH) domain removes the hydroxyl group together with the proton on the adjacent carbon to form carbon-carbon double bonds. The whole process bears high resemblance to the synthesis of fatty acid by type I FAS. A simplified version for one cycle of iterative polyketide synthesis is illustrated in **Figure 2.2**.

After the acyl chain reaches a designated length, a series of post-elongation modifications including product cleavage, triple-bond formation, cyclization, oxygenation, and methylation may occur to yield the mature enediyne products. Since the antitumor property of enediyne is largely attributed to the presence of the two chemically reactive triple bonds, the formation of the *–yne* group is of particular interest to us. One of the five genes found in the minimal cassette— *UNBL* is speculated to encode a non-heme di-iron protein [96, 97] that is potentially capable of catalyzing triple bond formation. However, it is still inconclusive whether such an acetylenase gene exists in the enediyne PKS gene cluster. On the other hand, the control over the chain length of the polyketide intermediate and the release of product from PKS normally involves the catalysis from either a type I or type II thioesterase. Type I thioesterases are normally *cis*-acting domains fused to the C-terminus of the most downstream module of PKS, while type II thioesterases are discrete proteins responsible for the *trans* hydrolytic release of the aberrant products [98-100]. Even though another gene in the minimal cassette— *TEBC* has a very high chance to encode a thioesterase involved in product release from PKS, the translated protein shares very low sequence homology to any of the known type I or type II thioesterases. Instead,

the putative thioesterase encoded by *TEBC* shares moderate sequence homology with a family of hotdog fold proteins that are characterized by a long central α-helix packed against five-stranded anti-parallel β-sheet. The hotdog fold proteins include many dehydratases and thioesterases that use acyl-CoA as substrates [101]. Apart from *–yne* formation and cleavage of matured acyl chain, cyclization is also one of the crucial steps in the formation of enediyne core. To date there is no protein candidate for the cyclase, since the products encoded by the other two genes in the minimal cluster bear extremely low homology to any protein with known function in the database. With the completion of the whole series of complex modification process, the enediyne core is presumably furnished and ready to be assembled with orsellinic acid and sugar moieties by specific glycosyltransferases.

Aiming to understand the enediyne biosynthetic mechanism, our project focused on the biosynthesis of 10-membered enediyne calicheamicin γ_1^I, whereby the function of CalE8 (*PKSE*), CalE7 (*TEBC*) and CalU15 (*UNBL*) are investigated using *in vitro* and *in vivo* methods.

2.2 Materials and Methods

Materials

CoA, acetyl-CoA, malonyl-CoA, NADPH and other chemicals were purchased from Sigma-Aldrich and stored at -20 °C. Expression vectors pET-28b(+), pET-26b(+), pCDF-2 and pQE-30 were bought from Novagen and Qiagen, from where BL21(DE3) and M15 competent cells were also obtained respectively. The genes that encode CalE8 (putative enediyne PKS), CalE7 (putative thioesterase) and CalU15 (putative acetylenase) in calicheamicin biosynthesis (*Micromonospora echinospora ssp. calichensis*) were provided by GenScript Corporation (NJ, USA) with sequences optimized for protein expression in *E. coli*. The genes were synthesized based on the reported protein sequences and were provided as pUC57-based plasmids pUC57-CalE8, pUC57-CalE7 and pUC57-CalU15. The sequences for the synthetic genes have been deposited in Genbank under the accession numbers of EU697953 and EU697954.

Protein Cloning, Expression and Purification

The 5757 bp gene that encodes CalE8 was excised from pUC57-CalE8 and cloned into pET-28b(+) between the *Nde*I and *Xho*I restriction sites. Then the plasmid harboring the PKS gene was sequenced and transformed into *E. coli* strain BL21(DE3) for protein expression. After inoculation into LB medium, two liters of culture was grown up to OD_{600} = 0.6 before induction with 0.2 mM IPTG. The culture was kept shaking at 16 °C for 20 hours at 160 RPM before harvesting the cells by centrifugation. The cell pellet was resuspended in lysis buffer (50mM phosphate (pH 6.5), 500 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol) and lysed by sonication. After centrifugation at 22,500 rpm for 30 minutes at 4 °C, the supernatant was filtered and then loaded onto HP Ni^{2+}-NTA column (GE Healthcare). A stepped gradient (0 mM,

20 mM, 50 mM, 100 mM, and 200 mM imidazole) was employed for vigorous washing before elution with 500 mM imidazole-containing elution buffer. Eluent that contains the N-terminal $(His)_6$-tagged CalE8 was applied onto Superdex-200 gel-filtration column (GE Healthcare) for further purification. Afterwards, fractions containing the bright yellow CalE8 protein were pooled and concentrated. The final protein concentration was determined by Bradford method before the concentrated protein solution was flash frozen in liquid nitrogen and stored in -80 °C freezer. The gene encoding the putative TE was excised from pUC57-CalE7 and first cloned into pET-26b(+) plasmid, which gave low yield under various expression conditions. Subsequently, the gene was cloned into pQE-30 and transformed into *E. coli* strain M15 for expression. TE expression and purification were similar to the procedure described above for CalE8. Upon induction with 0.6 mM IPTG at OD_{600} = 0.6, the culture was kept shaking at 160 RPM for 20 hours at 16 °C. The harvested cells were lysed in buffer containing 20 mM Tris (pH 8.5), 200 mM NaCl, 5% glycerol and 5 mM β-mercaptoethanol. Ni^{2+}-NTA purification and gel-filtration were carried out sequentially to yield the colorless tetrameric protein. The final protein concentration was determined by Bradford method before the concentrated protein solution was flash frozen in liquid nitrogen and stored in -80 °C freezer.

The gene encoding the putative acetylenase was excised from the plasmid pUC57-CalU15 and ligated into pET-28b(+) (Novagen) between the *Nde*I and *Xho*I sites. The recombinant plasmid was then transformed into the *E. coli* strain BL21(DE3) for protein expression. Two liters of LB medium were inoculated and subsequently induced with 0.6 mM IPTG at $O.D_{600}$ = 0.6. After overnight incubation with shaking at 16 °C, the cells were harvested by centrifugation and washed with lysis buffer (50 mM Tris (pH 7.4), 300 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF). SDS-gel analysis showed that about half of the recombinant was expressed as

inclusion body. Following lysis by sonication, the lysate was centrifuged at 20,000 rpm for 45 min. The clear supernatant obtained was then incubated with Ni^{2+}-NTA resin for one hour at 4 °C. After washing with a stepped gradient, the $(His)_6$-tagged protein was eluted with a 500 mM imidazole-containing elution buffer. The protein solution was desalted using a PD-10 column (GE Healthcare) and exchanged into the storage buffer (50 mM Tris (pH 7.4), 300 mM NaCl, 10% glycerol, 1 mM DTT). The protein was flash frozen using liquid nitrogen after measurement of final protein concentration by Bradford assay.

Co-Expression of CalE8, E7 and U15

The PCR primers for CalE7 and CalU15 genes were designed according to the instruction manuals for pCDF-2 and LIC Duet Adaptor kit (Novagen). Amplified PCR products were treated with T4 DNA polymerase and ligated into the pCDF-2 vector to yield the plasmids pCDF-2-CalE7 or pCDF-2-CalE7-CalU15 which contains both genes. pCDF-2-CalE7 or pCDF-2-CalE7-CalU15 was co-transformed with pET-CalE8 into BL21(DE3) competent cells. The cells harboring the recombinant plasmids were selected against kanamycin and streptomycin. Successful co-transformation was confirmed by colony PCR and protein expression screening. For the co-expression, 1 liter of LB medium containing both kanamycin (50 μg/mL) and streptomycin (50 μg/mL) was inoculated with 2 mL of culture from small-scale growth. The culture was then grown at 37 °C at 200 RPM to $O.D_{600}$ = 0.6 and induced with 0.4 mM IPTG. Following induction for 20 hours at 16 °C, the cells were harvested and lysed for protein purification using the procedures described above. Because only CalE8 and E7 contain the $(His)_6$-Tag, the purification by Ni^{2+}-NTA yielded a yellow protein solution containing both CalE8 and E7. Through the use of a size-exclusion column, the colorless CalE8 and colored E7 were separated and purified.

In vitro Activity Assay on the Integral KR Domain of CalE8

Enzymatic reactions involving CalE8 KR domain was carried out by monitoring the disappearance of NADPH peak at 340 nm using the spectrophotometer. The reaction mixture contained 40 μL of CalE8 (13.7 mg/mL), 0.5 μL of decalone(100 mM), 2.5 μL of NADPH(10 mM), 4 μL of DTT (40 mM) and 113 μL of reaction buffer (50 mM sodium phosphate at pH 7.5, 300 mM NaCl). The control reaction had the same composition of the reagents without the addition of decalone. Both reaction and control were subject to spectrophotometry scanning before and after 1.5 hours of incubation in 30 °C waterbath.

In vitro Activity Assay by Absorption Spectroscopy

Enzymatic reactions involving CalE8 and E7 were conducted in a semi-micro quartz cuvette and monitored with a Shimazu UV-Vis 1700 spectrophotometer. The sample chamber was kept at 30 °C through an external temperature controller. 22 μL of colorless CalE8 (13.7 mg/mL), 20 μL of colorless CalE7 (1.96 mg/mL), 3 μL of acetyl-CoA (10 mM), 21 μL of NADPH (10 mM) and 213 μL of buffer (100 mM Tris (pH 8.2), 300 mM NaCl and 1 mM DTT) were added into the cuvette with gentle mixing. The reaction was incubated in the sample chamber for 15 minutes prior to the addition of 21 μL malonyl-CoA (10 mM). Subsequently the absorption spectrum of the reaction mixture was taken at various time intervals to monitor the progress of product formation. The reactions were also conducted with the addition of the putative acetylenase under similar conditions. Small molecule electron donors, such as ascorbate, sodium dithionite ($Na_2H_2SO_4$), as well as the ferredoxin/ferredoxin reductase system were tested in order to reduce the di-iron cofactor in the putative acetylenase. Under all the conditions tested, no oxidation of the products was observed.

In vitro Activity Assay and Product Analysis by HPLC

22 μL of colorless CalE8 (13.7 mg/mL) and 20 μL of colorless CalE7 (1.96 mg/mL) were incubated with 200 μM acetyl-CoA and 600 μM NADPH in buffer (100 mM Tris (pH 8.2), 300 mM NaCl and 1 mM DTT) at 30 °C for 15 minutes prior to the addition of 1.2 mM of malonyl-CoA. The 200 μl of enzymatic reaction was carried out in a 30 °C water bath for 3 hours, together with two controls that lack either malonyl-CoA or CalE8. The reaction was then quenched with the addition of 5% TFA. To extract the yellow pigment from CalE7, an equal volume of ethyl acetate was added into the reaction mixture and vigorous vortexing ensued. Subsequently the reaction mixture was centrifuged at 15,000 RPM for 15 minutes. The organic solvent layer was then pipetted out and the reaction mixture was subject to extraction for the second time. The ethyl acetate extracts were pooled and the solvent was evaporated under a gentle flow of nitrogen gas. The yellow pigment was re-dissolved in a small volume of methanol for HPLC analysis using an eclipse XDB RP C18 column (4.6 × 150mm). The gradient employed was from 100% buffer A (HPLC grade water with 0.045% TFA) to 100% buffer B (90% acetonitrile, 10% HPLC grade water with 0.045% TFA) in 60 minutes.

Large Scale Preparation of Yellow Pigment

The colored CalE7 protein purified from the co-expression strain as well as the large volume *in vitro* reaction mixture was pooled together and acidified using 1% TFA, followed by ethyl acetate extraction as described above. The ethyl acetate extract was then blown to dryness under a constant gentle flow of nitrogen gas. The dried yellow pigment was re-dissolved in methanol and loaded onto a semi-preparative eclipse XDB RP C18 column (4.6 × 250mm) for HPLC purification. The wavelength of the UV detector was set at 410nm, with a reference wavelength

of 600nm. The gradient employed was from 70% buffer A (HPLC grade water with 0.045% TFA) + 30% buffer B (acetonitrile with 0.045% TFA) to 100% buffer B in 90 minutes. The HPLC fractions were collected on Agilent 96 well-plates. The yellow fractions were then pooled together and lyophilized for structure determination.

LC-MS

The column used for LC-MS analysis was eclipse XDB C18 column with a dimension of 2.3 × 150mm. The gradient employed in the analysis was 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 120 minutes. The ionization energy was set at 5.0 kV with an ESI 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.

UV-Vis and FTIR Spectroscopy

The absorption spectra of the proteins and small molecules were measured using a semi-micro cuvette on a Shimazu 1700 UV-Vis spectrophotometer. Data collection and sample preparation for FTIR Infrared spectra were recorded on a Nicolet Nexus spectrometer (Madison, WI) equipped with a MCT/A detector. The sample was applied onto a trapezoidal Ge internal reflection element (IRE). A dry nitrogen gas stream flowing through the ATR compartment was used to remove water vapor in the sample chamber throughout the measurement.

Structure Determination of the Carbonyl-Conjugated Polyene

^{1}H-^{1}H COSY and TOCSY proton NMR spectra were collected on a Bruker 700 MHz NMR spectrometer in the School of Biological Sciences (NTU) using $CDCl_3$ as the solvent and TMS as the internal reference. About 1 mg of the orange-colored product was obtained from large-scale *in vitro* reactions and extraction of the co-expressed CalE7. The structure of the yellow pigment was established by HRMS, IR, UV-Vis and ^{1}H NMR data.

Concentration, pH and Temperature Dependence of Enzymatic Reactions Involving CalE8

A typical reaction mixture includes 1 μL of CalE8 (14 mg/mL), 2.5 μL of CalE7 (25.2 mg/mL), 2.5 μL of NADPH (10 mM) and 91.5 μL of buffer [100 mM phosphate (pH 6.5, 7.0, 7.5, 8.0) or Tris (pH 8.5), or CHES (pH 9.0), 300 mM NaCl and 1 mM DTT]. After gentle mixing, the reaction was incubated in the sample chamber of UV-Vis spectrophotometer for 15 minutes prior to the addition of 2.5 μL malonyl-CoA (10 mM). The malonyl-CoA/NADPH concentration dependent reactions were done at constant NADPH/malonyl-CoA concentration, with varying malonyl-COA/NADPH concentration of 100, 250, 500, 1000 and 1500 μM. The reactions were carried out at pH 8.2, 30 °C for 3 hours. At the end of 3 hours, the reaction was then quenched with the addition of HCl to bring the pH to ~2. An equal volume of ethyl acetate was added into the reaction mixture and vigorous vortexing ensued. Subsequently the reaction mixture was centrifuged at 15,000 RPM for 15 minutes. The organic solvent layer was then pipetted out and the reaction mixture was subject to extraction for the second time. The ethyl acetate extracts were pooled and the solvent was evaporated using a Speed-Vac. The yellow pigment was re-dissolved in a small volume of methanol for HPLC analysis using an eclipse XDB RP C18 column (4.6 × 250mm). The gradient employed was 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.

2.3 Results

2.3.1 Purification of Proteins

For all the three over-expressed proteins, namely CalE8, CalE7 and CalU15, immobilized metal ion affinity chromatography (Ni^{2+}-NTA column) and size exclusion chromatography (high load gel filtration Superdex 200 column) were employed sequentially for the purification process. Subsequently, SDS-PAGE was run to check the relative purity of the proteins and the identity of each protein was verified using mass spectrometry.

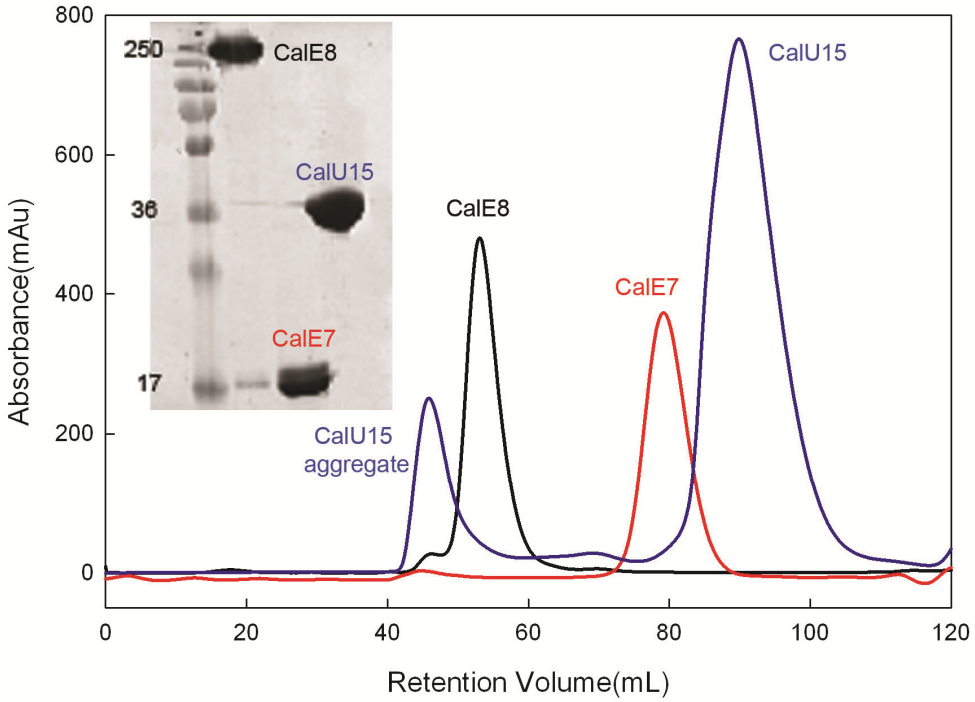


Figure 2.3 Protein purification of CalE8, CalE7 and CalU15. The FPLC chromatograms of CalE8, CalE7 and CalU15 and their respective SDS-PAGE images are shown. CalE8 is proposed to be a dimer based on its retention time in FPLC chromatogram, whereas CalE7 appears to be a tetramer and CalU15 is purified in its monomeric form.

CalE8 protein was purified as dimer with a bright yellow coloration. On the other hand, colorless CalE7 solution was obtained with CalE7 existing in tetrameric form. Similarly, the

solution of the monomeric CalU15 appeared colorless after purification as well (**Figure 2.3**). Extensive dialysis or extraction with organic solvent after protein denaturation could not separate the yellow pigment from CalE8, suggesting that the yellow pigment is covalently linked to the protein.

2.3.2 *In vitro* Assay for the Integral KR Domain of CalE8

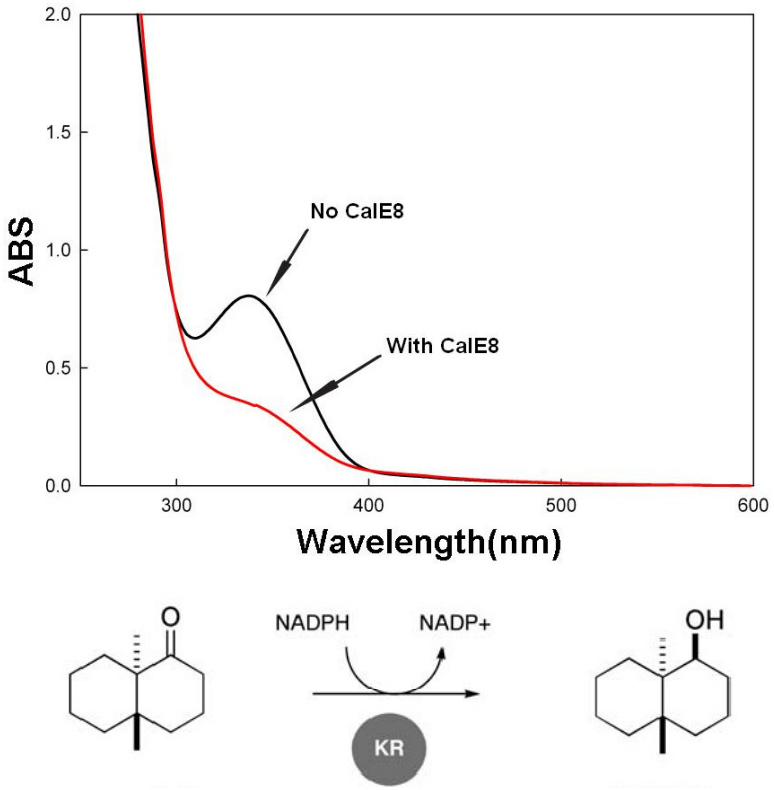


Figure 2.4 Reaction of CalE8 with decalone and NADPH. The keto group is converted to a hydroxyl group catalyzed by the KR domain of CalE8 in the presence of NADPH.

From the spectroscopic study on the KR domain, it is evident that the KR domain of CalE8 is fully functional. In the presence of CalE8 and decalone, the consumption of NADPH implied the conversion of the keto group to the hydroxyl group. On the other hand, when decalone was not added into the reaction mixture a decrease in the level of NADPH was absent (**Figure 2.4**). Later on, in another control reaction performed whereby decalone and NADPH were added in

the presence of CalE7 instead of CalE8, there was no reduction in the NADPH level either. Collectively, the consumption of NADPH in the presence of decalone is solely due to the reductive function of KR domain in CalE8. Since there is no enoyl reductase (ER) domain in CalE8, the products from CalE8 should contain a certain degree of desaturation provided the downstream DH domain being fully active.

2.3.3 *In vitro* Enzymatic Assays

Subsequently, a series of *in vitro* enzymatic assays were carried out with CalE8 and its auxiliary proteins CalE7 and CalU15. It was observed that the yellow pigment could be gradually removed from CalE8 by incubation with purified CalE7. Hence, the yellow pigment is thought to be covalently attached to the phosphopantetheinyl group of the ACP domain by a thioester linkage and likely to be synthesized by CalE8 in *E. coli*. This is most likely due to the fact that CalE8 can be activated by self-phosphopantetheinylation with the integrated PPTase domain at the C-terminus [53].

The activities of the proteins were further investigated by incubating CalE8 and CalE7 with their putative substrates, namely acetyl-CoA, malonyl-CoA and NADPH under the experimental conditions described in the experimental section. Prior to that, extensive screening on experimental conditions regarding buffer pH, salt concentration, reaction temperature and substrate concentration was carried out for optimization purposes.

The reaction mixture readily generated products that absorbed in the range of 350-450 nm upon incubation when both CalE8 and CalE7 were present. CalE8 was shown to be the essential component of the reaction since no products were generated when CalE8 was absent. On the other hand, CalE7 was imperative for the formation of the products at considerable quantities,

indicating its plausible role in product release.

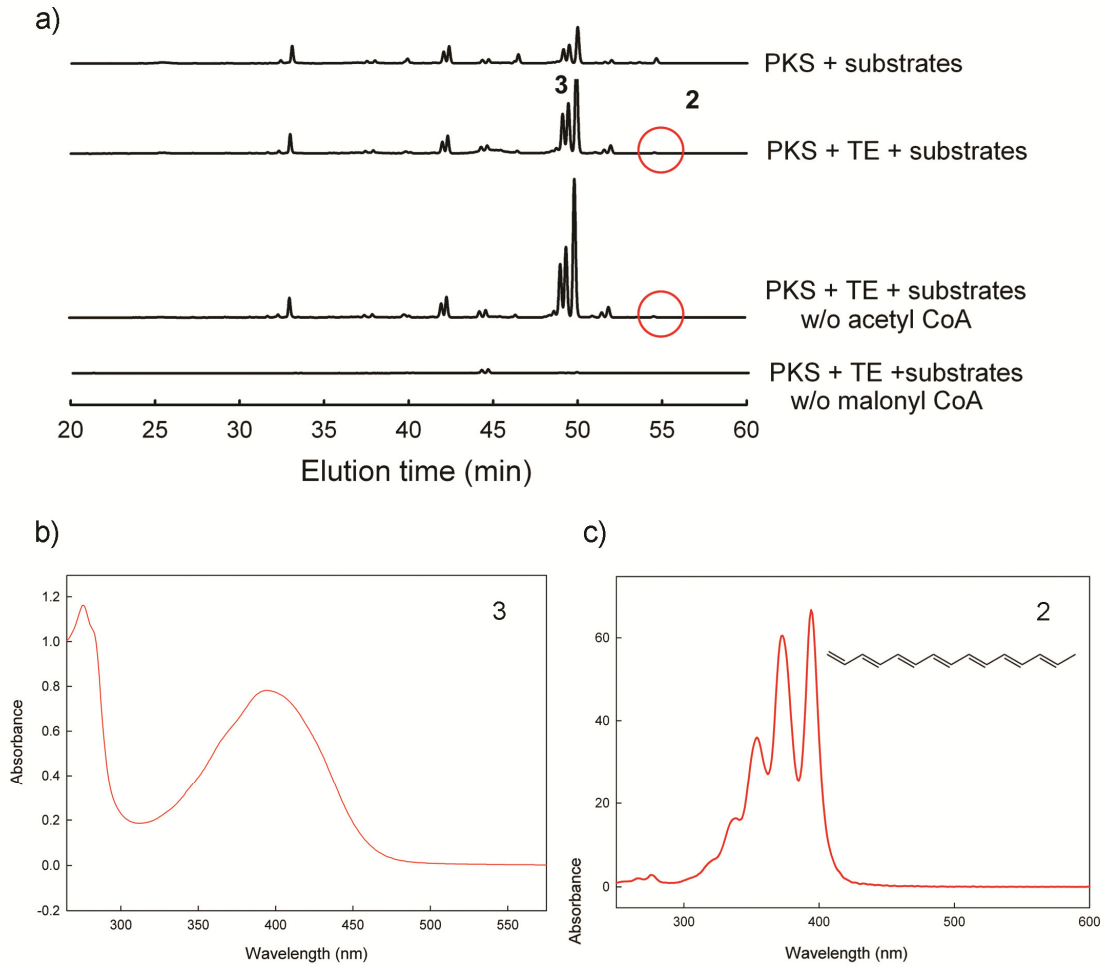


Figure 2.5 *In vitro* enzymatic assay for CalE8 and CalE7. a) *In vitro* enzymatic assay involving CalE8, CalE7, malonyl-CoA, Acetyl-CoA and NADPH. The conjugated polyene 1, 3, 5, 7, 9, 11, 13-pentadecaheptaene (**2**), which was discovered in a parallel *in vivo* study of the 9-membered C-1027 enediyne system, was produced at negligible amount. Instead another compound **3** appeared to be the major product of the reactions. Both CalE8 and CalE7 were imperative to obtain **3** in considerable amount. b) The spectrum of the major product **3**. c) The structure and spectrum of the polyene **2**.

Moreover, the reaction can proceed without the addition of acetyl-CoA, suggesting that CalE8 is capable of generating the "starter" acetyl-CoA from malonyl-CoA by an intrinsic decarboxylation mechanism. The HPLC analysis of the *in vitro* reaction mixture revealed the

formation of several compounds with differential product distribution. The major product was seen as three consecutive peaks eluted out one after another within a matter of seconds with an absorption maximum at 395 nm (**3**). The fractions collected for the three peaks displayed clearly a yellow coloration. In addition, a minute amount of 15-carbon conjugated polyene 1, 3, 5, 7, 9, 11, 13-pentadecaheptaene (**2**), which was identified as the putative precursor for 9-membered enediyne C-1027 in an *in vivo* study conducted by Shen and coworkers [95], was detected in the *in vitro* enzymatic reaction as well. According to Shen and coworkers, the polyene **2** is produced *de novo* by the C-1027 PKS SgcE in the heterologous *E. coli* expression host. With the help of the putative thioesterase SgcE10 in the same co-expression system, the matured polyene **2** is cleaved off from the ACP domain of SgcE and released into the cytoplasm. Due to the high sequence homology shared by SgcE and CalE8, it is not surprising to find **2** as one of the products formed by CalE8 in *in vitro* assays. The structure and spectrum of **2** along with the spectrum of **3** are shown in **Figure 2.5c** and **Figure 2.5b**. Nonetheless, the extremely low quantity of **2** detected resulted in the manifestation of the product as tiny peaks that were hardly discernable on most of the occasions. Since the amount of **3** and **2** detected in HPLC was vastly different, compound **3** was regarded as the only major product for the *in vitro* reaction conducted (**Figure 2.5a**). Apart from the products **3** and **2**, a few more minor products were formed as well during the course of enzymatic reactions involving CalE8 and CalE7.

It was found out that at relatively low pH, e.g. pH 6.6, the reaction mixture turned yellow after 3 hours of incubation at 30 °C in water bath. Subsequent HPLC analysis revealed that three minor products with absorption maxima at 310 nm (**4**), 355 nm (**5**) and 375 nm (**6**) respectively were formed besides **3** and **2**. At higher pH, e.g. pH 8.8, the yield of the major product **3** was

augmented drastically. The amount of polyene **2** extracted after trifluoroacetic acid (TFA) quenching remained low despite the change in pH (**Figure 2.6**).

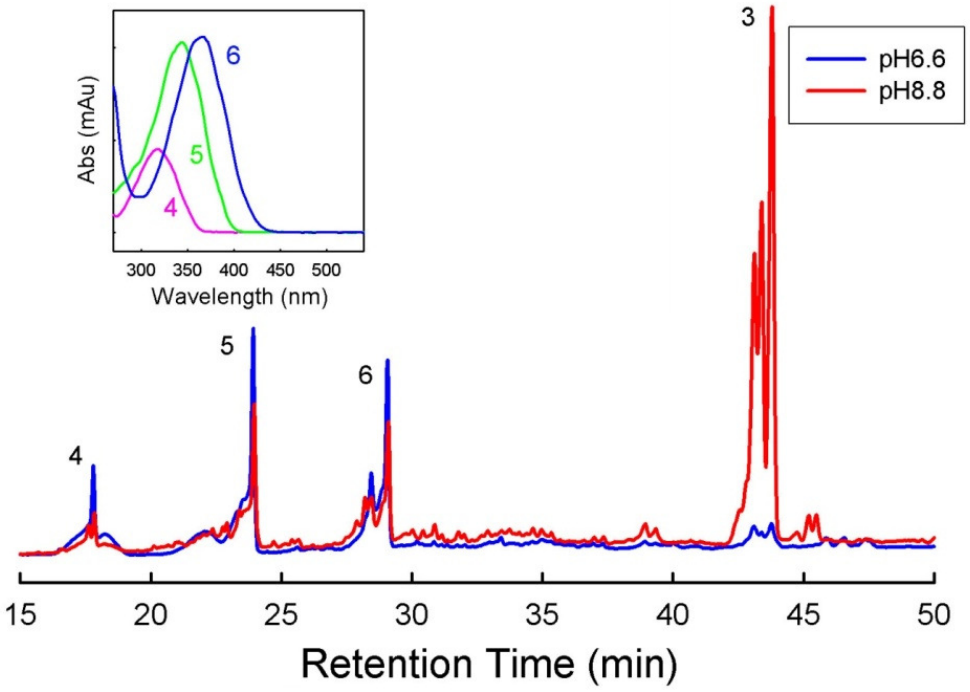


Figure 2.6 HPLC analysis of the products generated by CalE8 and E7 at pH 8.8 and 6.6. HPLC chromatograms are shown along with the absorption spectra of **3**, **4**, **5** and **6**. (Reaction conditions: 14.28 µM CalE8, 8.33 µM CalE7, 25 µM acetyl-CoA, 175 µM NADPH and 175 µM malonyl-CoA in 100 mM Tris with 300 mM NaCl and 1 mM DTT).

Apart from CalE7 and CalE8, the putative acetylenase CalU15 was also tested for activities in the *in vitro* reaction together with the other two proteins. Various reducing agents, such as ascorbic acid, sodium dithionite and ferredoxin—ferredoxin reductase pair, have been employed as electron donors in the oxidative reaction. However, no detectable changes occurred to the products formed compared to reaction involving only CalE8 and CalE7. The inability of CalU15 to carry out the desired desaturation step is probably due to the lack of di-iron cluster or a proper electron donor which is thought to be crucial for the function of CalU15 in the catalysis of triple bond formation.

In addition to the HPLC analysis carried out for the *in vitro* reactions, UV-Vis spectrometry was also employed to study the kinetics of the reaction (**Figure 2.7**). From the UV-Vis data obtained, the amount of NADPH in the reaction declines with increasing concentration of products generated. The presence of an isosbestic point indicates a correlation with simple molar ratio between NADPH consumed and products formed. Thus, a linear stoichiometric relationship is to be expected between the formation of the products and the consumption of the substrates, namely NADPH and malonyl-CoA.

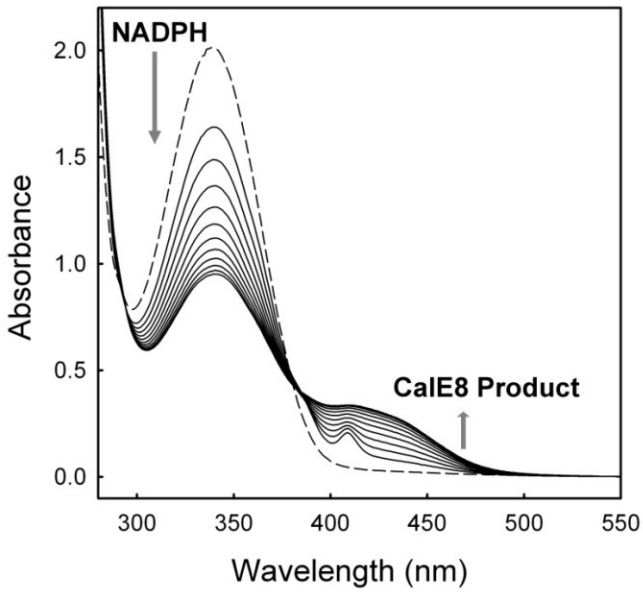


Figure 2.7 UV-Vis spectrum of the CalE8-CalE7 reaction with time. The spectra were scanned at every 5 minutes interval for 2 hours since the start of the reaction.The absorption maxima for NADPH and products of CalE8 are around 340nm and 410nm respectively.

2.3.4 *In vivo* Co-expression Studies

Besides the *in vitro* assays, the results from the *in vivo* experiments also led to the same conclusion that CalE7 is crucial in the release of full-length product from CalE8. Cells containing the recombinant pET28-CalE8 plasmid produced intensely yellow CalE8 proteins, presumably with thioester precursors of **3** and **2** attached. Neither extensive dialysis nor

extraction by organic solvents could remove the yellow coloration from the protein. Meanwhile, the CalE7 expressed alone in *E. coli* was tetrameric and colorless in solution. In contrast, strains carrying both pET28-CalE8 and pCDF-2-CalE7 plasmids gave rise to proteins with distinct color changes from those expressed alone. After Ni^{2+}-NTA purification and size exclusion chromatography, the CalE8 protein obtained after concentration was only slightly yellowish in color, whereas the CalE7 purified was bright yellow (**Figure 2.8**).

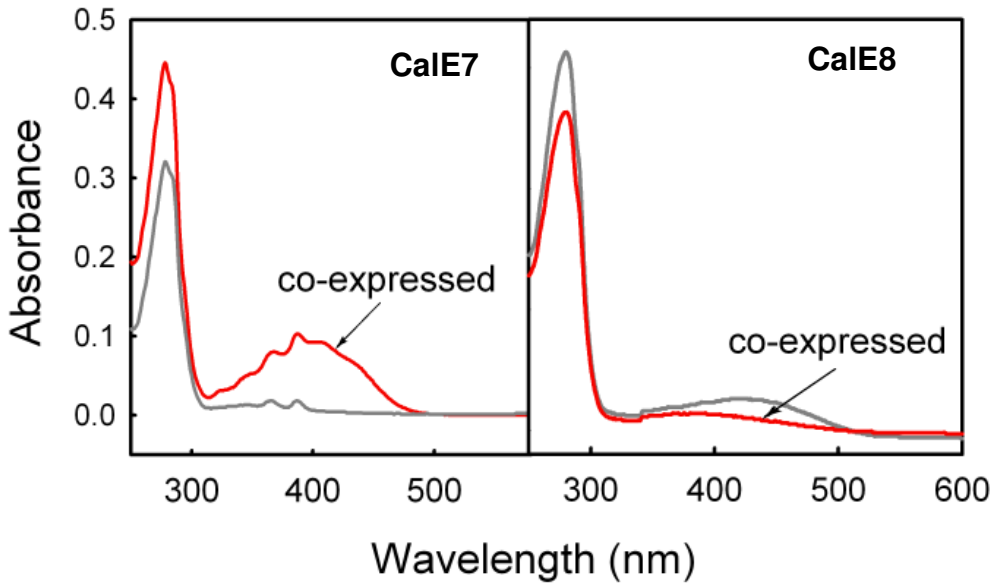


Figure 2.8 The difference in spectra of co-expressed and singly-expressed CalE8 and CalE7. The co-expressed CalE8 has a much lighter shade of yellow compared to the bright yellow coloration of CalE8 expressed alone. The reverse occurred on CalE7, indicating that the yellow pigment is transferred onto CalE7 from CalE8 within the *E. coli* cells.

The observation indicates a possible transfer of the yellow pigment from CalE8 to CalE7 within the cells. Moreover, the yellow pigments **3** and **2** associated with CalE7 could be extracted readily by organic solvent. In light of the *in vitro* and *in vivo* experiment results, it suggests that the products are likely to be produced by and covalently bound to CalE8 protein

given the access to all the necessary substrates within *E. coli* cells after self-phosphopantetheinylation by the cognate PPTase domain. Upon incubation with CalE7, the products are cleaved off from CalE8 through hydrolysis of the thioester bond catalyzed by CalE7 and subsequently become non-covalently associated with CalE7 protein. In order to investigate the possibility of the products being released from CalE7 into the cytoplasm or further secreted into the culture medium, extraction by organic solvent on supernatant (after removal of CalE8 and CalE7) and cell debris from centrifuged cell lysate as well as culture medium for the co-expression strain was carried out. HPLC analysis on the organic solvent extract from the cell debris showed negligible amount of product **3** most probably due to the presence of small amount of CalE7 present in the inclusion bodies after cell lysis. With the help of LCMS, the molecular weight and the fragmentation pattern of the products were determined.

2.3.5 LC-MS Analysis

Since it has been established that the enediyne core is formed by at least seven acetate units connected in a head-to-tail fashion, the results from high-resolution mass spectrometry suggested that **4**, **5** and **6** could only be a series of aberrant or immature products with the molecular formula of $C_8H_8O_3$ (**4**, $[M+1]^+$, *m/z* 152.05), $C_{10}H_{10}O_3$ (**5**, $[M+1]^+$, *m/z* 179.02) and $C_{12}H_{12}O_3$ (**6**, $[M+1]^+$, *m/z* 205.02) (**Appendix figure 2.1** and **2.2**). Although the characterization of the molecular structures of **4**, **5** and **6** by NMR remained to be difficult given the low yield of the three products, the high-resolution molecular formulae and MS/MS spectra implied that they are most likely the lactones formed in solution from the linear tetra-, penta- and hexaketide PKS products [102] (**Figure 2.16**). The observed *m/z* of 215.1432 (MH^+) for **3**, and the two minor components eluted immediately before **3** indicated that they are most likely geometrical isomers sharing the same molecular formula of $C_{15}H_{18}O$ (calcd *m/z* 215.1430 (MH^+)). The MS/MS

spectrum exhibits a series of fragments, suggesting that the molecule is likely to be a linear molecule consisting of at least seven acetate units (**Appendix figure 2.3**).

2.3.6 Structural Determination for Carbonyl-Conjugated Polyene 3

The analysis of the ^{1}H-NMR spectrum of purified **3** revealed that the majority of the protons resonate in the 5.8 - 6.8 ppm region, in agreement with a conjugated polyene structure. No chemical shift between 9 - 10 ppm was observed, ruling out the presence of an aldehyde group. Moreover the ^{1}H-^{1}H COSY and TOCSY NMR experiments established that the olefinic protons in the 5.8 - 6.8 ppm region are coupled to two sets of protons with the chemical shifts of 7.18 and 7.63 ppm and another set of protons with the chemical shift of 1.80 ppm (**Appendix figure 2.4** and **Appendix table 2.1**). The COSY spectrum further established the connectivity among some of the protons, such as the protons at 1.80 and 5.80 ppm.

It is well known that conjugated polyenes exhibit the signature fine structure in their absorption spectrum, and that the fine structure would be absent if the olefinic groups are conjugated to a carbonyl group [103, 104]. Hence, the lack of fine structure in the UV-Vis spectrum for **3** was the first indicator that the olefinic groups are conjugated to a carbonyl group (**Figure 2.5**), in contrast to the polyene **2** identified in both C-1027 [95] and calicheamicin systems. This conclusion was further supported by a prominent band at 1676 cm^{-1} in the IR spectrum (**Appendix figure 2.5**), which resembles the 1652 cm^{-1} and 1680 cm^{-1} bands observed for the carbonyl-conjugated polyene laetiporic acid and flavofungin [105].

In conjunction with the structure of **2**, it is likely that the product is a linear carbonyl-conjugated polyene (3, 5, 7, 9, 11, 13- pentadecen-2-one, **3**). The chemical shifts for H_{13}, H_{14}, H_{15} protons in **3** are identical to those in **2** produced. The NMR data for a compound (candidin

macrolide) that contains a similar carbonyl-conjugated polyene moiety is available [106]. The chemical shifts for the protons on C3 to C14 in **3** and that of the candidin macrolide described in the article are almost identical except for the 7.63 and 6.82 ppm signals arisen from a geometrical isomer.

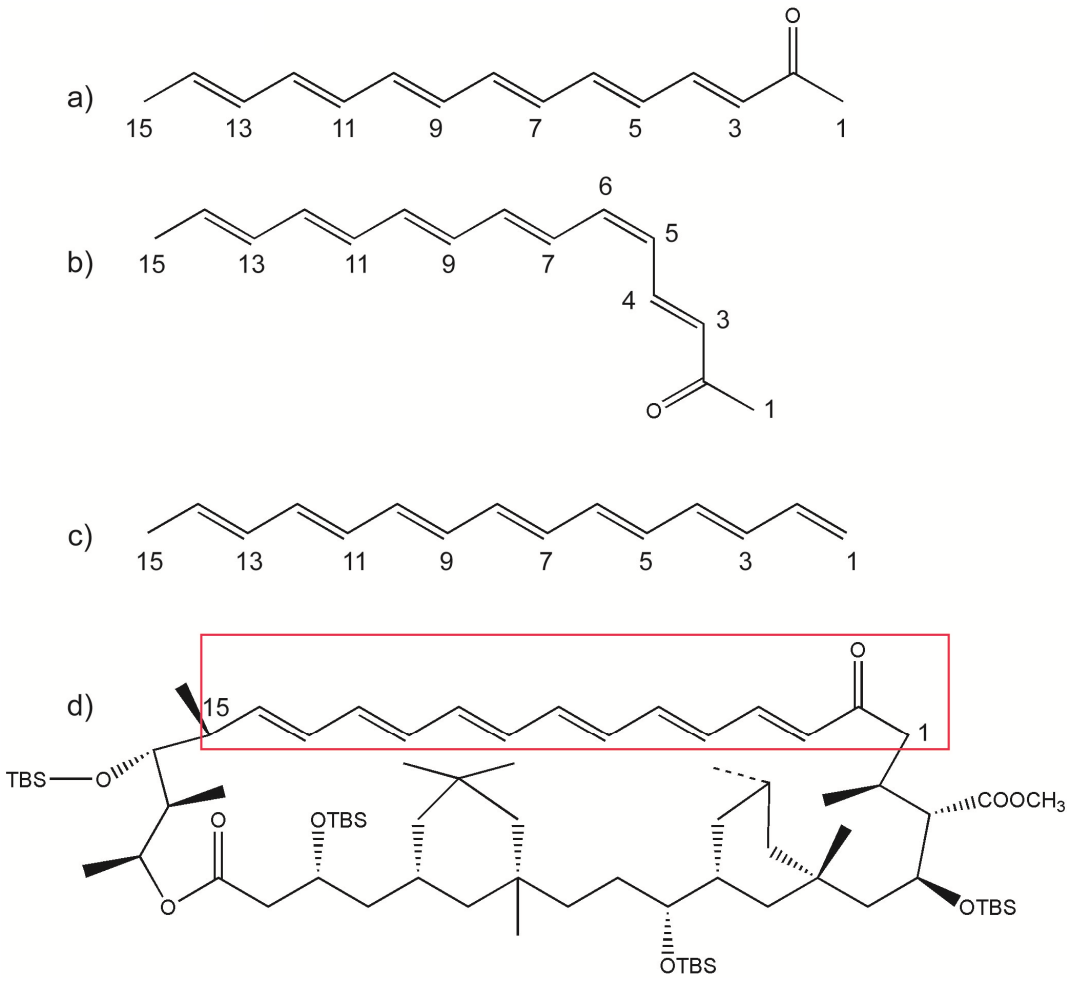


Figure 2.9 Structure of different products of enediyne PKSs and the reference compound candidin macrolide. (a) Structure of major isomer of the yellow product **3**. (b) Structure of minor isomer of the yellow product **3**. (c) Structure of the polyene precursor **2** discovered in both SgcE/SgcE10 and CalE8/CalE7 system. (d) Structure of candidin macrolide with similar carbonyl-conjugated polyene moiety as the yellow product.

Analysis of the two-dimensional NMR spectra also revealed two sets of signals indicating that the HPLC purified sample was actually a mixture of at least two geometric isomers.

Although it was not possible to follow completely the two major correlation patterns in the 2-D NMR spectra due to overcrowding of the signals, the downfield changes in chemical shift for the 7.18 ppm signal (H_4 of major isomer) suggested that the two geometric isomers were likely to differ in the geometry around the double bond C5-C6, with the minor isomer bearing the *cis* configuration [107] (**Figure 2.9**). All the significant cross peaks in the two spectra can be accounted for based on the structures of these two isomers. The isomerization was likely to be attributed to photo-isomerization during the reaction and purification processes, despite precautionary measures taken to minimize light exposure.

2.3.7 Validation of the Structure and Configuration of Double Bonds of 3 by Synthetic Chemistry

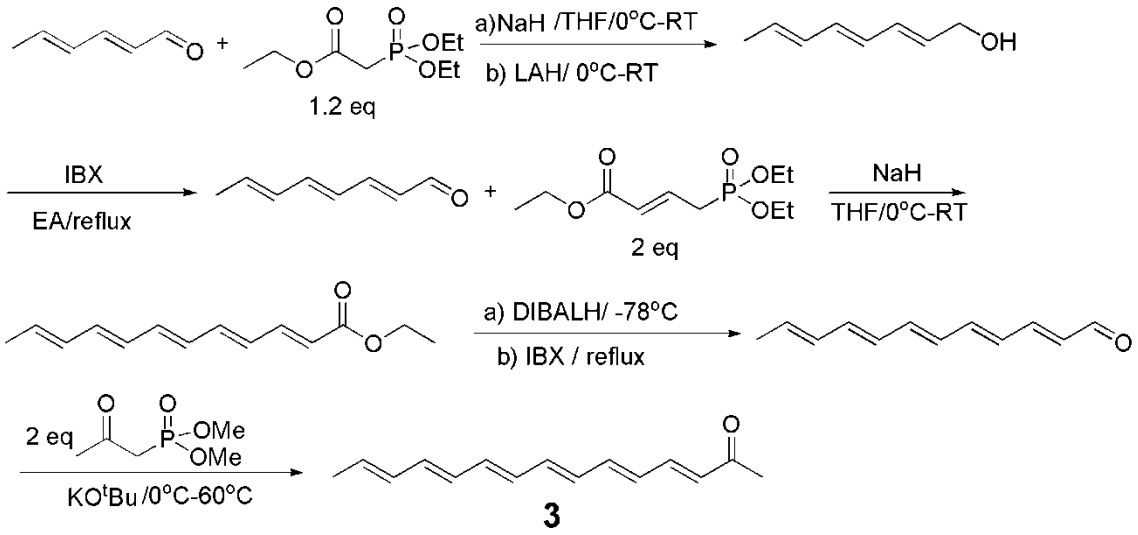


Figure 2.10 Synthesis of (*3E*, *5E, 7E, 9E, 11E, 13E*)-pentadecen-2-one (**3**).

Although the structures of **3** and **2** have been determined by multi-dimensional NMR [95, 108], the absolute configuration of the double bonds of **3** was not established due to the overlap of the resonance signals in the ethylenic proton region. It was speculated that the products of CalE8 may contain different ratio of *cis* and *trans* double bonds based on triplet peak of **3**

observed in the *in vitro* experiment. To validate the presence, or absence, of the *cis* double bond, the all *trans* (*3E*, *5E, 7E, 9E, 11E, 13E*)-pentadecen-2-one was chemically synthesized as standard by our collaborators in the chemistry department (**Figure 2.10**). HPLC analysis revealed that the standard shares identical retention time and absorption spectrum with **3**, suggesting that the products of CalE8 are likely to contain *E*-configured double bonds.

2.3.8 Product Distribution with Enzymatic Assays Quenched by HCl

Initially carbonyl-conjugated polyene **3** was regarded as the only major product from the enzymatic reaction involving CalE8 and CalE7 after quenching by TFA. However, when hydrochloric acid (HCl) was used in place of TFA to quench the reaction, we discovered that addition of TFA into the reaction mixture actually had adverse effect on the composition of the products in the *in vitro* enzymatic assays. Due to the chemical lability of the polyene **2** under variety of conditions including heat, acid and light [109], The likelihood that majority of **2** produced was destroyed by TFA crossed our mind. It was previously observed that the CalE8/CalE7 pair produced carbonyl-conjugated polyene **3** in large quantity but only negligible amount of **2**. Given the instability of **2** in the presence of TFA as observed in the SgcE/SgcE10 assay, the yield of **2** must have been underestimated. Indeed, we could observe both **3** and **2** when HCl was used to quench the CalE8/E7 reaction.

The three previously characterized pyrone derivatives (**4**, **5**, and **6**) were still found to be present as minor products. Enzymatic activity assays further suggested that the product ratio varies considerably with different assay conditions and substrate concentrations. Product **2** was the only major product at mild conditions (pH 7.0, 16-23 °C), even though the overall productivity of the enzymes was compromised. In contrast, CalE8 generated significant amount of **3** at elevated pH and temperature, as well as high malonyl-CoA and low NADPH

concentrations, with carbonyl-conjugated polyene **3** emerging as the major product under some conditions (**Figure 2.11, Figure 2.12, Figure 2.13** and **Figure 2.14**).

In the NADPH concentration-dependent experiment, the amount of NADPH present in the reaction mixture affected the product distribution to a large extent. Elevated by-product formation was observed at low NADPH concentration range. The peaks of product **3** became highly prominent at low NADPH concentration of 100 μM. On the other hand, the amount of **2** produced increased with increasing NADPH concentration in the reaction mixture. Taken together, the amount of by-products formed in the enzymatic reaction is inversely correlated to the concentration of NADPH at the expense of the synthesis of **2**. As a result, much of acetate units are channeled into the production of **3** as well as the pyrone by-products when NADPH is deficient (**Figure 2.11**).

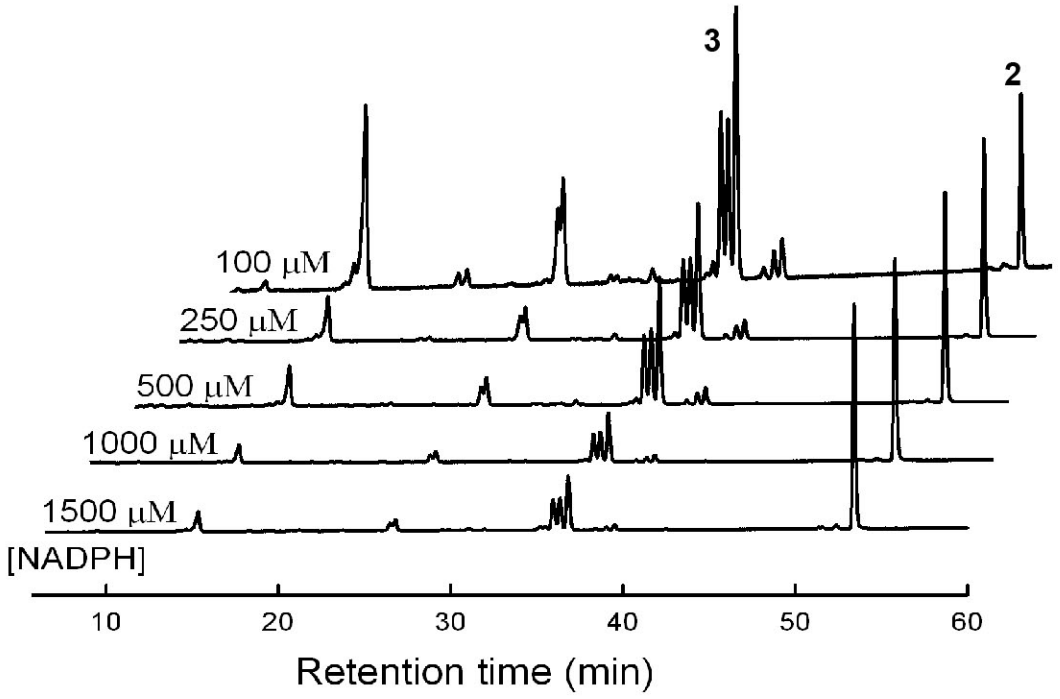


Figure 2.11 HPLC analysis of the NADPH concentration-dependent product formation for CalE8.

With reference to the malonyl-CoA concentration-dependent experiment, an excess of malonyl-CoA enhanced the production of **3** by a smaller margin during the course of the reactions in comparison to that of NADPH concentration-dependent enzymatic assays. Similarly, the formation of the pyrones also remained at relatively low level even at elevated malonyl-CoA concentration. Unlike the NADPH concentration-dependent experiments, the correlation between the amount of **2** produced and malonyl-CoA concentration was not very obvious. The result raised the possibility that the rate of production of **2** is tightly controlled by CalE8 irrespective of the concentration of malonyl-CoA present (**Figure 2.12**).

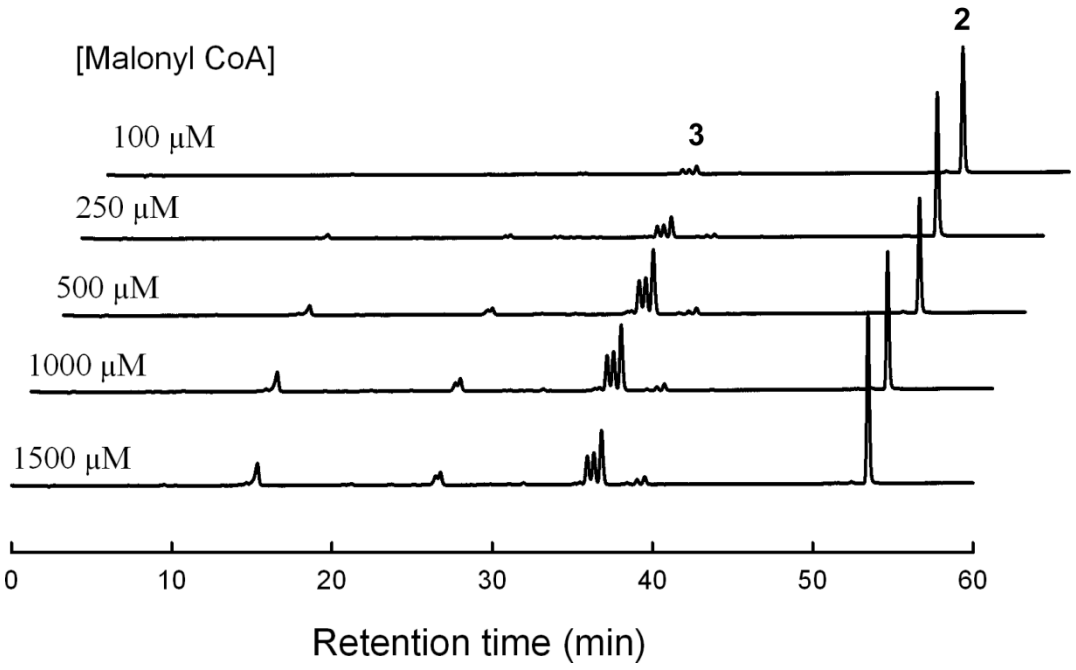


Figure 2.12 HPLC analysis of the malonyl-CoA concentration-dependent product formation for CalE8.

Following the substrate concentration-dependent experiments, the impact of experimental conditions on the product distribution was investigated as well. At slightly acidic or neutral pH, the formation of **3** and the pyrones was either abolished or highly suppressed, while the amount

of **2** produced was compromised to a certain extent as well. The optimal pH for the production of **2** fell within the range of pH 7.5 to pH 8.5. At higher pH, e.g. pH 9.0, the amount of **3** being produced increased drastically. The formation of the pyrones was hardly observed in the pH-dependent reactions, which indicated that the formation of pyrones was highly dependent on the substrate availability rather than external factors such as pH (**Figure 2.13**).

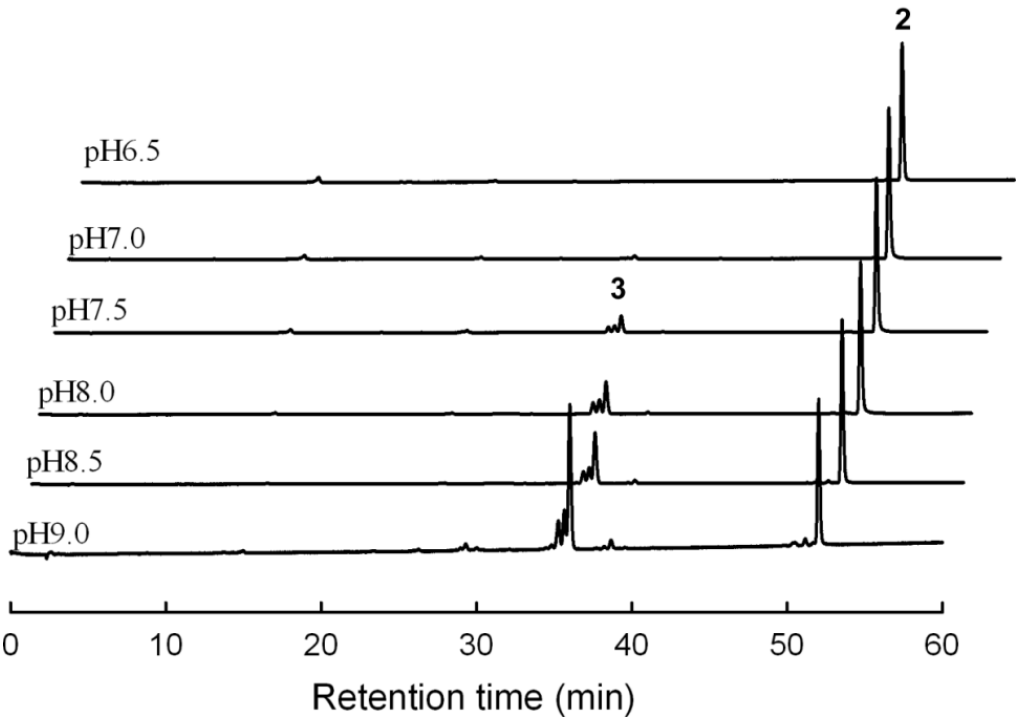


Figure 2.13 HPLC analysis of the pH-dependent product formation for CalE8.

Besides pH, the effect of temperature on the product distribution was examined as well. The formation of **2** remained more or less constant regardless of the temperature at which the reactions were carried out. On the other hand, the effect of temperature on the formation of **3** was highly prominent. At low or ambient temperature, the formation of **3** was suppressed to a large extent. When reaction temperature was increased to 30 °C or 37 °C, the amount of **3** produced escalated correspondingly. The presence of pyrones was almost negligible within the

temperature range investigated. In accordance with the pH-dependent experiment, the result further substantiated the hypothesis that the formation of pyrones was correlated to the substrate concentrations rather than environmental factors (**Figure 2.14**).

In conclusion, the variation of **3** to **2** ratio observed in the series of experiments performed suggested that the last keto-reduction step catalyzed by the KR domain is sensitive to environmental factors and substrate availability. On the other hand, the formation of truncated pyrones can be ascribed to either an excess of malonyl-CoA present or a deficiency in NADPH in the reaction mixture.

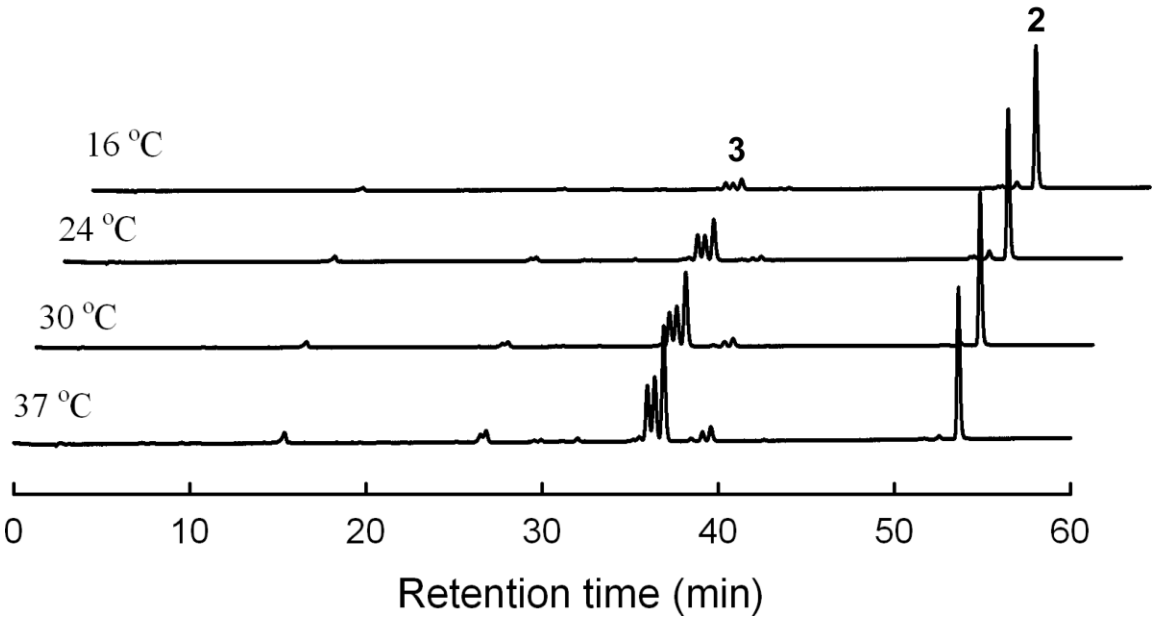


Figure 2.14 HPLC analysis of the temperature-dependent product formation for CalE8.

2.4 Discussion

From the sequence comparison of the three PKSs CalE8, DynE8 and SgcE, it is evident that the three PKSs are highly homologous despite of the difference in the enediyne cores produced. Although CalE8, DynE8 and SgcE belong to the family of calicheamicin-like 10-membered enediyne PKS, dynemicin-like 10-membered enediyne PKS and 9-membered enediyne PKS respectively, they share the same domain organization with a sequential arrangement of KS, AT, ACP, KR, DH and PPTase spanning from N- to C-terminus.

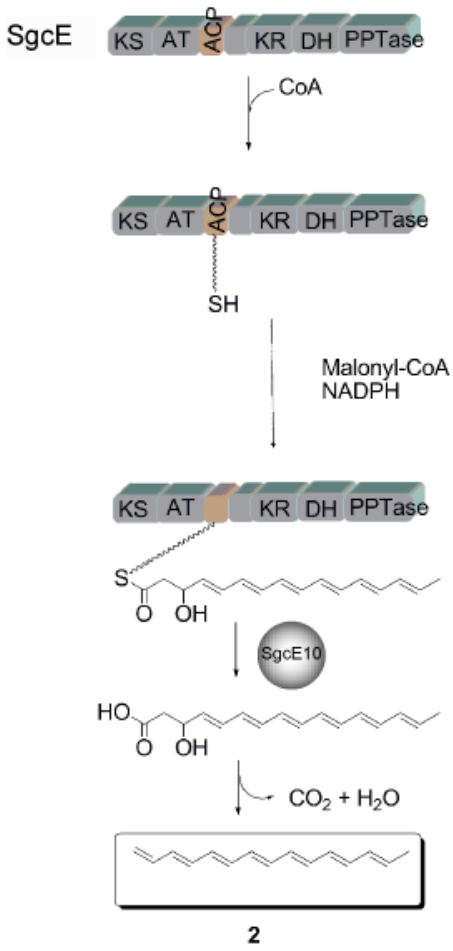


Figure 2.15 The production of 1, 3, 5, 7, 9, 11, 13-pentadecaheptaene (**2**) by SgcE and Sgc10 in *E. coli* heterologous expression system.

In a paper published by Shen's group in 2008, a potential candidate for the polyketide intermediate in the biosynthesis of 9-membered enediyne C-1027 was identified in a heterologous expression system. In the experiment performed, *E. coli* cells harboring both SgcE (encoded by *PKSE*) and SgcE10 (encoded by *TEBC*) produced the 15-carbon polyene 1, 3, 5, 7, 9, 11, 13-pentadecaheptaene (**2**). Shen and coworkers subsequently proposed that the polyene **2** could be the precursor for C-1027 enediyne core. SgcE is likely to produce an octaketide thioester intermediate by joining eight acetate units in a head-to-tail fashion. Subsequently the putative type I thioesterase SgcE10 releases the product from the ACP domain of SgcE. The β-hydroxyl carboxylic acid intermediate formed after thioester hydrolysis then undergoes decarboxylation and dehydration to yield the conjugated polyene **2** (**Figure 2.15**). However, the *in vitro* enzymatic assays performed by Shen and coworkers yielded little success primarily due to the abnormally low activity of SgcE in the reaction conducted.

On the other hand, CalE8 appears to exhibit the highest *in vitro* activity therefore is more suitable for biochemical studies. The *in vitro* and *in vivo* experiments performed have indicated that CalE8 is the major enzyme involved in the biosynthesis of the polyketide precursor for the enediyne core of calicheamicin. Given the high sequence homology between SgcE and CalE8, it is not surprising to find polyene **2** as one of the products in the enzymatic reaction involving CalE8 and the cognate thioesterase CalE7. Theoretically CalE8 functions by joining eight acetate units in a head-to-tail fashion through decarboxylative condensation to yield the octaketide thioester intermediate that is covalently tethered to the phosphopantetheinyl group on the ACP domain. Unlike most of the PKSs that require a starter molecule for the initiation of decarboxylative condensation, the KS domain of CalE8 presumably is capable of generating the starter molecule *de novo* from the malonyl-CoA substrates via decarboxylation. This is shown in

the experiment whereby CalE8 does not require acetyl-CoA to be present in the reaction mixture for the synthesis of full-length products **2** and **3**. More importantly, only negligible amount of product **2** and **3** could be observed in the absence of CalE7 under all conditions, indicating that CalE7 is responsible for the release of the octaketides linear intermediates. The capability of CalE7 to release both full-length products **2** and **3** suggested that the enzyme exhibits relaxed specificity towards the polyketide chains bound by CalE8. Moreover, CalE7 must contain or undergo extensive conformational changes to produce an elongated substrate-binding channel upon binding of the PKS products in order to accommodate the lengthy thioester precursors of **2** and **3**. Along with the production of polyene **2** and carbonyl-conjugated polyene **3** (**Figure 2.17**), the immature products **4**, **5** and **6** are also observed in the CalE8/CalE7 reactions (**Figure 2.16**).

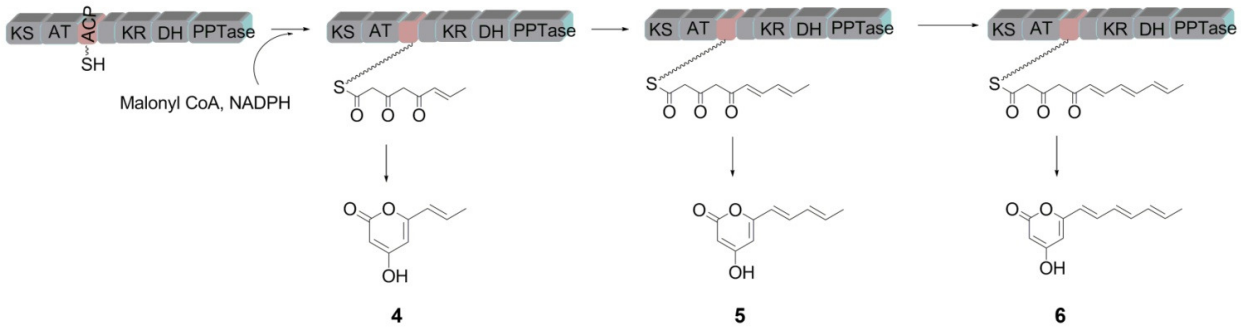


Figure 2.16 the different pyrones generated by CalE8. A schematic representation of CalE8 generating the truncated polyketides of **4**, **5** and **6**.

The formation of the by-product pyrones is highly dependent on the concentration of NADPH present in the reaction mixture [102]. The deficiency in NADPH impedes the function of the KR domain in the conversion of the keto group to hydroxyl group on the growing acyl chain. As a result, 3, 5-diketo thioester intermediates that are highly susceptible to self-cyclization are formed. Due to the high acidity of α-hydrogen in the 3, 5-diketo thioester intermediates, the cyclization process is normally spontaneous and does not require the aid of a

thioesterase for release. Generally, the formation of pyrones as by-products is inevitable in FAS or PKS reactions. However, an ample supply of NADPH in the reaction mixture will minimize the formation of the pyrones so that most of the resources can be channeled into the production of full-length products.

According to the assay results, the synthesis of **2** as the major product under physiologically-relevant conditions seems to suggest polyene **2** as the polyketide intermediate *en route* to the construction of enediyne core. The conjugated polyene **2** becomes the dominant product under physiologically relevant conditions that suppresses the generation of other polyketides like carbonyl-conjugated polyene **3** and pyrones **4**, **5** and **6**. It seems that conjugated polyene **2** could be the authentic biosynthetic intermediate for both 9- and 10- membered enediynes as implicated by the studies on DynE8 and SgcE as well [110]. On the other hand, some intrinsic differences may exist among the three PKSs in 9 and 10-membered enediyne biosynthesis, based on their propensity to generate different polyketide products [110]. The significance of the observed capability of the PKSs to generate more than one full-length product is not clear at this moment. It has been thought that TE plays a critical role in the formation of different products. It is likely that the catalytic cavity of enediyne PKS is in fact easily accessible to TE. The event of thioester cleavage takes place once the acyl chain grows to a designated length. Provided the catalytic rate of TE is faster than that of the KR domain in PKS, the octaketide chain is readily cleaved even before the KR domain can act on the thioester intermediate for the reduction of the β-keto group. In accordance with the pH- and temperature-dependent experiments, high pH is likely to facilitate the cleavage of thioester through increased availability of hydroxide ions in the catalytic cavity which are proposed to act as nucleophiles in the hydrolytic reaction (This part will be further discussed in Chapter 4); while at elevated temperature, it is possible that an

increase in the conformational flexibility of PKS grants easy access of the ACP domain to TE, which results in the increased production of **3** since there is less time allowed for the KR domain to reduce the β-keto group. Hence, a plausible conclusion can be drawn whereby the indiscrimination of TE towards the full length acyl chains leads to the product variations. Considering that iterative PKSs may or may not require ancillary proteins to generate the correct product, further evidence is required to fully establish **2**, or **3**, or even another PKS product as the biosynthetic intermediate.

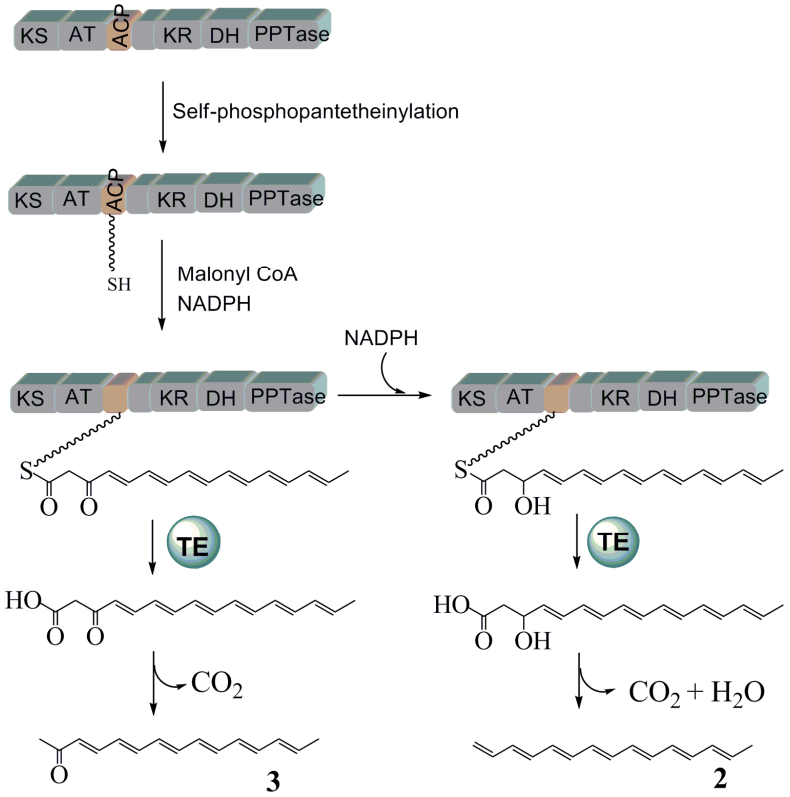


Figure 2.17 The different 15-carbon polyketides generated by CalE8. A schematic representation of the biosynthesis of product **2** and **3** by CalE8 and CalE7.

However, in spite of the fact that both **2** and **3** are formed as the putative full-length linear products (**Figure 2.17**), we still cannot rule out the possibility that both compounds are merely by-products of CalE8, as proposed by Townsend and coworkers [111]. In their opinion, it is

somewhat wasteful to generate two different versions of the full-length products at the expense of limited resources available in the cells. Thus it is likely that only aberrant products are synthesized in enzymatic reactions involving only CalE8 and CalE7. As a result, Townsend's group proposed that the synthesis of the real intermediate entails the participation of a regulatory protein which interacts with CalE8 to steer the biosynthetic process towards the production of indigenous polyketide precursor for enediyne core. Nonetheless, the existence of such a regulatory protein in the gene cluster still awaits to be verified.

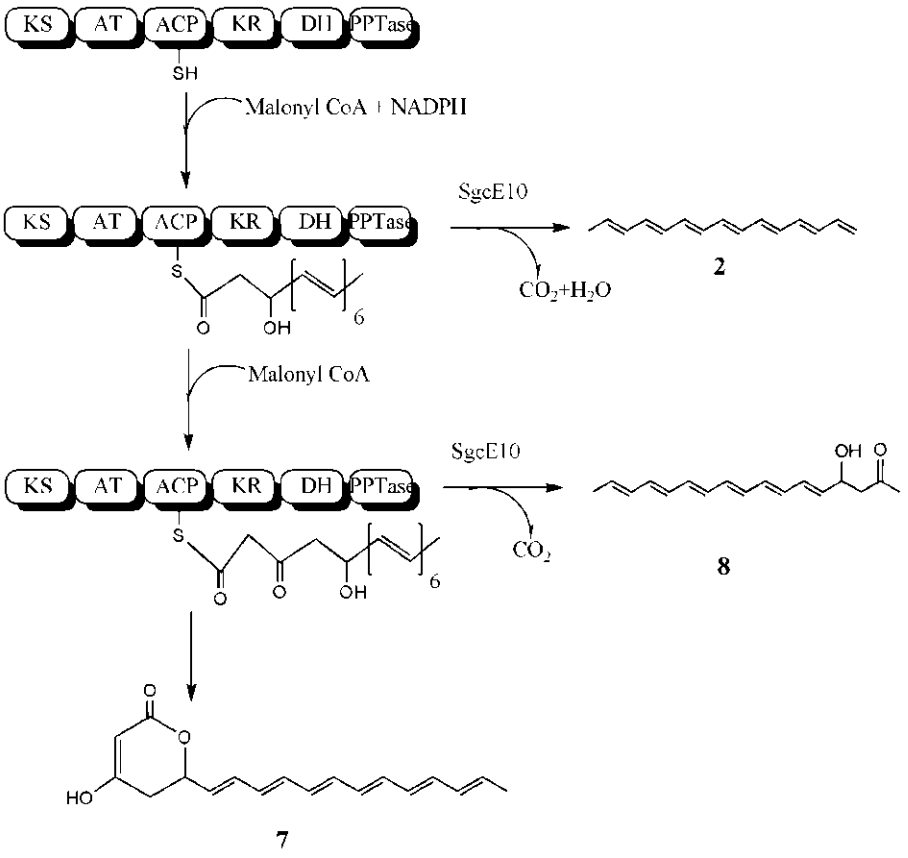


Figure 2.18 A schematic representation of SgcE generating different products in the presence or absence of the TE-SgcE10.

With respect to another paper published recently by Shen's group, polyene **2** was found to be the major product in *E. coli* strains harboring five different sets of enediyne *PKSE-TEBC*

genes respectively. Besides, **2** also emerged as the major product in experiments conducted on the co-expression of different enediyne PKS-TE pairs in *Streptomyces lividans* K4-114, a host more closely related to the indigenous enediyne-producing bacteria. In addition, the extraction of large amount of **2** from several native enediyne-producing bacteria lent further support to the hypothesis of **2** as the authentic intermediate for the synthesis of enediyne core. With aforementioned observations, it is evident that the synthetic pathway of **2** works in an enediyne-PKS-dependent fashion. However, the possibility of **2** as by-product cannot be ruled out since there is no definite experimental proof for the conversion of **2** into the actual enediyne core.

A recent paper published by Guo and coworkers argued that polyene **2** is but only an aberrant product in their *in vitro* studies on the C-1027 system [112]. In the experiment conducted, a nonaketide **7** was produced independent of the thioesterase SgcE10 (**Figure 2.18**). Based on their results, the polyene **2** is merely a truncated version of matured polyketide product. Moreover, SgcE10 is redefined as a type II thioesterase which is responsible for the cleavage of aberrant products. Truncated polyketides such as **2** and noncyclized hydroxyl polyene **8** are removed by SgcE10 in attempt to recycle the active sites of SgcE, whereas matured nonaketide undergoes self-cyclization and detaches from SgcE by itself to form **7.**

Up to now, there is still no consensus on which PKS product is the real intermediate for 9- and 10-membered enediyne core. Therefore, the point at which the evolutionary divergence occurs between the 9- and 10-membered enediyne biosynthetic pathways remains to be established. One school of thought claims that the divergence between the two classes of enediyne takes place after the PKS-TE stage. With evidence from bioinformatics studies, as well as several *in vitro* and *in vivo* experiments, a common precursor is envisioned for the early stage of enediyne biosynthesis, such as the polyene **2** described above. The auxiliary enzymes of the

enediyne system, potentially the cyclase or acetylenase, will dictate the differentiation of the common precursor into the respective core structures of the 9- and 10-membered enediynes. In contrast, another proposed biosynthetic pathway depicted the occurrence of the evolutionary divergence at PKS level, whereby the subtle intrinsic differences in the PKSs resulted in the production of 9- and 10-membered enediynes via distinct polyketide intermediates. However, both models require further substantiation because the crucial link between the polyketide intermediate and the folded enediyne core is still veiled in mystery. Unraveling the subsequent post-PKS steps is the key to fully decipher the biosynthetic pathway of enediynes.

CHAPTER 3 Solution Structures of the Acyl Carrier Protein Domain of CalE8

3.1 Introduction

One of the most distinctive features of the fatty acid, nonribosomal peptide and polyketide biosynthetic pathways is the utilization of the acyl carrier protein (ACP) or peptidyl carrier protein (PCP) for the shuttling of biosynthetic intermediates among various catalytic domains or proteins [113-117]. The small and highly dynamic ACPs or PCPs can be either free-standing proteins or integrated domains within a complex multidomain fatty acid synthase (FAS), polyketide synthase (PKS) or nonribosomal peptide synthase (NRPS). The free-standing ACPs from type II FAS pathways are the best studied ACP systems, with crystallographic and solution NMR studies showing that the type II FAS ACPs adopt a canonical four-helix bundle fold with a binding pocket for sequestering the growing fatty acid chain [118-120]. Meanwhile, although the ACP domains from the multidomain type I FASs adopt a similar overall structure, they do not seem to contain a pocket for binding and protecting the growing fatty acid chain [121-123]. Recent studies on the integrated ACP domain from the modular type I PKS DEBS and the discrete ACPs from type II PKS have revealed the similarity in overall protein structure but salient differences in the binding of acyl chain between type I and type II PKS ACPs [124-129]. Subtle differences between the ACPs in local structure, surface electrostatic potentials and binding mode of biosynthetic intermediates have been also documented. Recent NMR studies further revealed functionally important protein dynamics in PCPs for the modulation of the interaction between the PCPs and the NRPS catalytic domains [117, 130, 131].

As described in chapter 1, CalE8 is an iterative type I PKS that plays a central role in the early stage of the biosynthesis of the naturally-occurring enediyne calicheamicin γ_1^I in *Micromonospora echinospora ssp. calichensis*. Distinct from the modular type I PKSs, CalE8 is composed of a single module consisting of several catalytic domains for the synthesis of the putative precursor of the 10-membered enediyne moiety [51, 108, 110, 132, 133]. The ACP domains of CalE8 and other enediyne PKSs share very low sequence homology with known ACPs [51, 54, 58]. In fact, the initial assignment of the ACP domain in CalE8 was associated with great degree of uncertainty, not just because of the low sequence homology shared with other ACPs, but also because of the absence of the signature GX(H/D)S(L/I) motif conserved in many ACPs. We previously confirmed the identity of the ACP domain (*me*ACP) in CalE8 through *in vitro* modification of the excised *me*ACP by the phosphopantetheinyl transferase (PPTase) Sfp [53, 134]. Shen and coworkers also demonstrated that the ACP domain of SgcE, a homolog of CalE8 from the biosynthetic pathway of the 9-membered enediyne natural product C-1027, can be phosphopantetheinylated at the predicted Ser site by mass spectrometry [95].

According to the degree of reduction of keto groups in the final polyketide product, iterative type I PKS have been classified into the non-reducing (NR), partially reducing (PR) and highly reducing (HR) families [132, 135] (**Figure 3.1**). CalE8 belongs to the HR type I PKS family given that CalE8 and its homologs produce at least two conjugated polyenes (**2** and **3**) under both *in vitro* and *in vivo* conditions [108, 110, 111, 136].

To date, only one structure of the iterative type I PKS ACP domain has been reported [137]. Crump and coworkers determined the solution structure of the ACP domain from the fungal norsolorinic acid synthase (NSAS), which belongs to the iterative NR PKS family [137]. Unlike the polyketide products of the iterative NR and PR PKSs that are cyclized immediately upon

formation, the chemically labile polyenes **2** and **3** produced by CalE8 seem to be stabilized by the protein to such an extent that it can be co-purified with the protein [108, 133, 138]. Thus it is intriguing to speculate whether the ACP domain plays any role in protecting the conjugated polyenes. In this chapter the NMR solution structure of the ACP domain (*me*ACP) of CalE8 is reported, which represents the first structure of ACP domains from an HR type I PKS. Studies on the *apo*-, *holo*- and *acylated-me*ACP reveal that the modifications do not alter the overall protein structure but affect the structure at a local level. Two-dimensional NMR spectra of *me*ACP collected in the presence of thioesterase (CalE7) also provide interesting insight into the interaction between *me*ACP and the cognate protein partner.

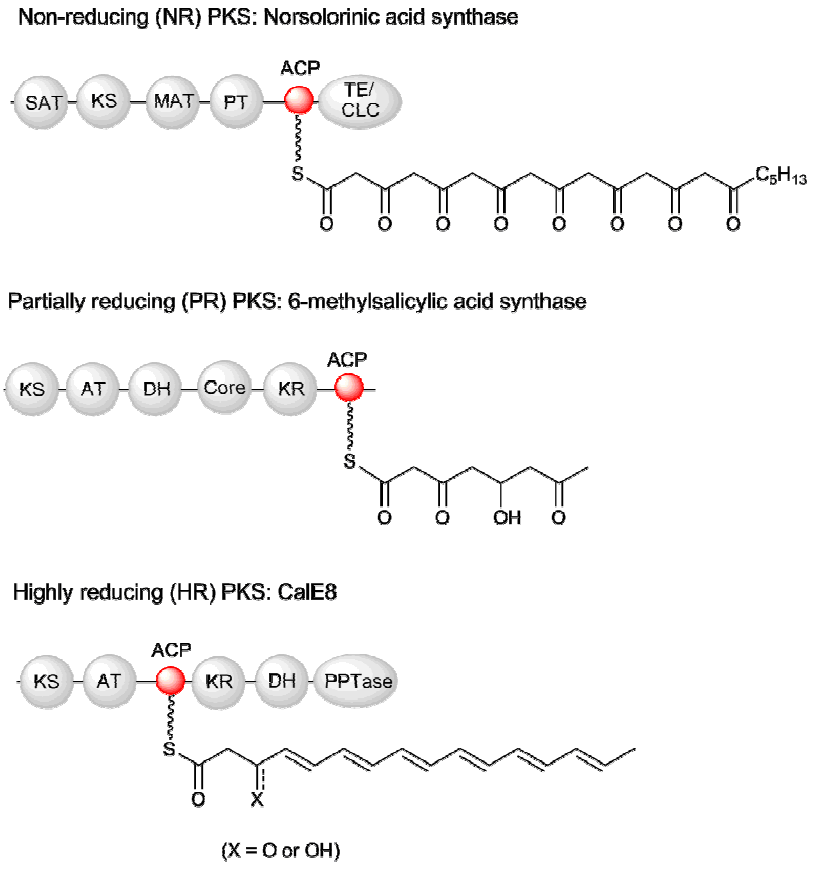


Figure 3.1 Three classes of type I iterative polyketide synthases (PKS) with the proposed PKS products tethered to the ACP domains.

3.2 Material and Methods

Materials

Coenzyme A (CoA), hydroxybutyryl-CoA, octanoyl-CoA and other chemicals were purchased from Sigma-Aldrich and stored at -20 ^{o}C. Expression vectors pET-26b(+), and *E. coli* BL21(DE3) were obtained from Novagen.

Protein Cloning, Expression and Purification

The cloning of *me*ACP and Sfp was done by my colleague Murugan, the detailed procedure was described in the paper published in 2008 [53]. The ACP and catalytic domains of CalE8 were cloned into compatible vectors and transformed into BL21(DE3) cells for expression. Seed culture of BL21(DE3) cells containing the respective plasmids were grown overnight before being inoculated into large scale cultures. The cultures were allowed to grow at 37 °C and 200 RPM until the OD_{600nm} reached 0.6. Induction was done with IPTG at 0.4 mM concentration. Cells were harvested 20 hours later after being left to grow at 16 °C and 160 RPM. The cell pellet was then resuspended in lysis buffer (50 mM HEPES at pH 7.5, 300 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol). Subsequently the cells were lysed by sonication. The supernatant obtained after high speed centrifugation was applied onto Ni^{2+}-NTA column for purification with a stepped gradient. The final eluent containing the recombinant protein was applied onto gel-filtration column S75 or S200 depending on the size of the individual protein concerned for further purification in a buffer suitable for NMR experiments (50mM NaH_2PO_4/ Na_2HPO_4 at pH 6.85, 50 mM NaCl, 1 mM DTT). The fractions containing the protein were pooled and concentrated. After that, the proteins solutions were stowed away in the -80 °C freezer for future NMR analysis. The same expression and purification protocols were used to

prepare ^{15}N, ^{13}C-labeled *me*ACP protein by using an M9 medium containing ^{15}N-isotopic ammonium chloride and ^{13}C isotopic glucose. The labeled protein obtained from gel filtration was further purified with a Mono Q column. A refined gradient was employed using buffer A (50 mM NaH_2PO_4/ Na_2HPO_4, 1 mM DTT) and buffer B (50 mM NaH_2PO_4/ Na_2HPO_4, 1 M NaCl, 1 mM DTT). The salt gradient increased gradually from 0% to 40% of B in 50 ml. Majority of the *me*ACP did not bind to the column and was eluted at early stage of the gradient. The fractions containing *me*ACP were pooled and concentrated for NMR as well as acylation experiments. The thioesterase CalE7 were cloned, expressed and purified as described in chapter 1 [108]. The PPTase Sfp used for modification of meACP was prepared by my colleague Lawrence according to the established procedure [53].

Preparation of *Holo* and *Acylated Me*ACP

The reaction for the production of *holo*- and *hydroxybutyryl-me*ACPs were summarized as follows: 400 µl of ACP (12 mg/ml), 100 µl of Sfp (68mg/ml), 40 µl of CoA or acyl-CoA (50 mM) and 10 µl of $MgCl_2$ (1 M) were mixed together in 1 ml of reaction buffer (100 mM Tris at pH 8.2, 300 mM NaCl, 1 mM DTT). The reaction was carried out at 30 °C for 8 - 12 hours. The progress of the conversion was monitored by HPLC analysis using an eclipse XDB RP C8 column at 2 hour intervals. The gradient employed was from 10% of acetonitrile to 90% acetonitrile in half an hour. Elution of *me*ACP and Sfp were monitored at the wavelength of 220 nm. Upon completion of the reaction, the mixture was desalted with NaH_2PO_4/ Na_2HPO_4 buffer without NaCl to bring down the NaCl concentration to ~50 mM. Aforementioned refined gradient was employed using MonoQ for the separation of *holo*- or *acylated-me*ACP from Sfp. A gel filtration step with Superdex-75 column was added to further improve the purity of the *holo*- and *hydroxybutyryl-me*ACP. For the production of *octanoyl-me*ACP, the reaction buffer

was modified to have the following composition: 50 mM NaH_2PO_4/ Na_2HPO_4 at pH 7.0, 200 mM NaCl, and 1 mM DTT. The duration of the reaction was prolonged to 16 to 20 hours at 25 °C and the conversion rate for the modifications is > 95%.

NMR Spectroscopy, Structural Calculation and Titration of *Me*ACP with CalE7

The NMR spectra were collected and the NMR solution structure was calculated by our collaborator, Mr. Lim Jack Wee from Dr. Yang's laboratory in National University of Singapore. Subsequently the structural and functional analysis was done by both Mr Lim Jack Wee and me. The detailed NMR experimental procedures are as follows:

All NMR experiments were performed on an 800MHz NMR spectrometer (Bruker) unless otherwise indicated. All NMR samples including ACPs and TE were prepared in a buffer with 50 mM NaH_2PO_4/ Na_2HPO_4 at pH 6.85, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% D_2O. Because ACP protein has weak UV absorbance at 280 nm, the protein concentrations were instead determined by the Bradford assay at 595 nm. To determine the structure of *apo-me*ACP, the following NMR spectra were collected on a ~1 mM ^{13}C,^{15}N-labeled protein sample: 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C HSQC, 3D HN(CO)CA, 3D HNCA, 3D MQ-(H)CCH-TOCSY [139] and 4D timeshared ^{13}C/^{15}N edited-NOESY [140]. To measure the heteronuclear NOEs, two spectra without and with proton saturation were measured on a ~1 mM ^{15}N-labeled sample with a saturation delay of 4 s and recycle delay of 4 seconds by the inverse-detected 2D NMR method. Proton saturation was achieved by a train of 120° pulses spaced at 5 ms.

In the TE titration experiments, a concentrated unlabeled CalE7 (TE) protein solution (0.8mM) was added to ^{15}N-apo-*me*ACP (0.3 mM) or ^{15}N-holo-*me*ACP (0.3 mM). The samples were first allowed to incubate for 10 min prior to NMR data collection. TROSY-HSQCs were

recorded at three TE concentrations (0 mM, 0.22 mM and 0.38 mM) in the same ACP buffer till a final molar ratio [CalE7]: [*me*ACP] of 2.5:1. All data were processed with NMRpipe and analyzed with NMRspy (http://yangdw.science.nus.edu.sg/Software&Scripts/XYZ4D/index.htm). NMRspy was recently developed to facilitate resonance assignment with the 4D NOESY-based strategy [141].

NOE restraints were obtained from the timeshared 4D $^{13}C/^{15}N$-editedNOESY (containing ^{13}C, ^{15}N-edited, ^{13}C, ^{13}C-edited and ^{15}N, ^{15}N-edited subspectra) using NMRspy. Backbone torsional angle restraints, Φ and Ψ were obtained from chemical shift data using TALOS [142]. Ambiguous NOEs were assigned with the iterative structure calculation method using Cyana 2.1 [143]. The final 10 lowest energy structures out of 100 calculated were selected. ^{1}H-^{15}N HSQCs of hydroxylbutyryl- and octanoyl-*me*ACP (0.1mM) were acquired at 15°C instead of 25°C on an 800MHz NMR spectrometer. This allows us to observe Ser971 which was predominantly weak at higher temperature > 20°C using 0.1mM of acylated *me*ACPs. The combined chemical shift perturbation is calculated using equation 1:

$$\Delta_{ppm}=[(\Delta\delta_{HN})^2+(\Delta\delta_N/7)^2]^{0.5} \qquad (1)$$

Where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the respective differences of ^{1}H and ^{15}N chemical shifts between different forms of *me*ACP: apo-*me*ACP, holo-*me*ACP or acylated *me*ACP (hydroxylbutyryl- and octanoyl-*me*ACP).

3.3 Results

3.3.1 Overall Structure of *Me*ACP

The ACP domain construct used in this study encompasses a 92 residue long (Ala^{925}-Pro^{1016}) fragment of CalE8. Previous studies have confirmed that the *E. coli* overexpressed *me*ACP is in its *apo* form and can be modified by the PPTase Sfp to generate *holo-me*ACP [53]. 94% of the residues of the recombinant *apo-me*ACP were unambiguously assigned. Unassigned residues were the amino-terminal Met^{924} and Ala^{925} and four Histidine residues of the carboxy-terminal $(His)_{6x}$-tag. The only disordered segments are the unstructured termini that include Ala^{925} to Thr^{936} and Ala^{1015} to Pro^{1016} (**Figure 3.2a**). The structured region of *me*ACP has a backbone R.M.S.D. of 0.41Å and no violation greater than 0.3Å.

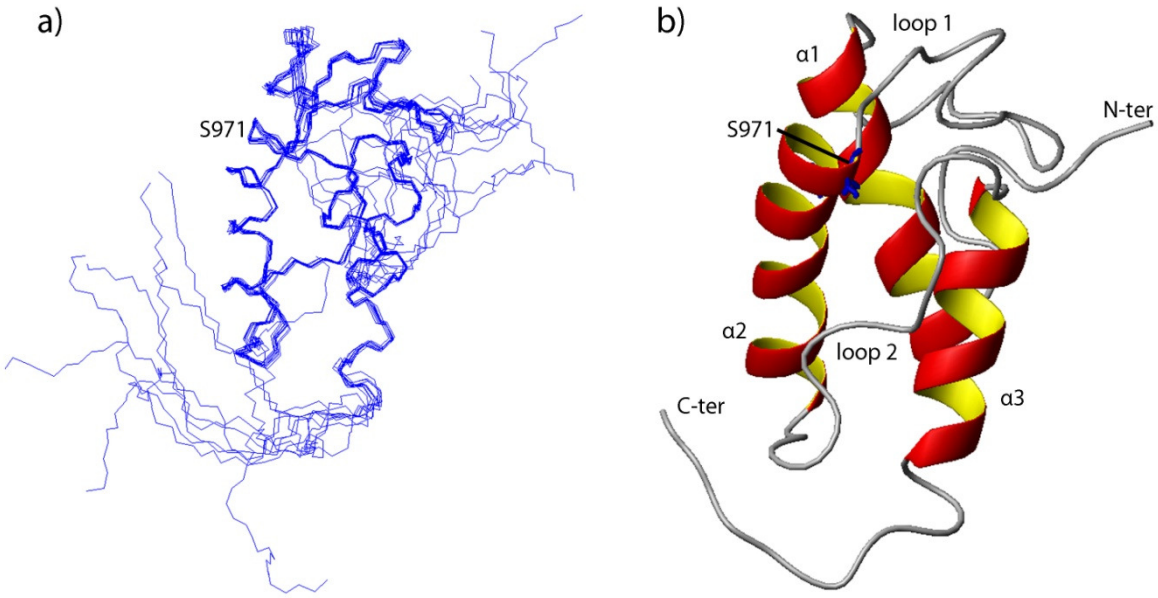


Figure 3.2 NMR structure of *me*ACP. (a) Backbone of an ensemble of the lowest energy conformations shown as line representation. (b) Mean *me*ACP solution structure is shown as ribbon model. S971 is the phosphopantetheinylation site.

Overall, *me*ACP assumes a globular fold of a twisted three-helix bundle (α1 [Ala^{938}-Ala^{950}], α2 [Ser^{971}-Met^{984}] and α3 [Val^{1001}-Glu^{1014}]) and two well-defined long loops (loop-1 [Glu^{951}-Ser^{970}] and loop-2 [Met^{985}-Thr^{1000}]) (**Figure 3.2b**). The domain seems to be stabilized through the packing of the hydrophobic side chains of the residues originated from the three helices as well as the two loops. The estimated helical content of 44.6% based on 41 out of 92 residues, is significantly lower than that of other known ACPs, which usually have more than 50% of helical content [144]. Most characterized ACPs adopts a canonical four-helix bundle structure with helix α1 running almost anti-parallel to helices α2 and α4, as well as a short helix α3 within the loop-2 that connects the helices α2 and α4. In *me*ACP, the region where the short helix should be located seems to be disordered and the lack of the local helical feature contributes to the low helical content of the protein. Some ACPs and PCPs have been reported to adopt two or more inter-converting conformations [130, 145]. We however did not observe a significant sub-population for *me*ACP.

3.3.2 Local Structural Features of *Me*ACP

A comparison with other ACP structures reveals some salient features of *me*ACP. Despite the facts that *me*ACP shares very low sequence homology with other ACPs and it lacks signature GX(H/D)S(L/I) motif conserved in many ACPs (**Figure 3.3a**), the phosphopantetheine attachment site (Ser^{971}) is still located at the beginning of helix α2 (**Figure 3.4**). The GX(H/D)**S**(L/I) motif is replaced by a $H^{968}MS\mathbf{S}^{971}I$ motif in *me*ACP with His^{968} and Ser^{970} substituting the canonical Gly and Asp/His residues. The essential role of Ser^{971} was confirmed by the observation that S971A mutation completely abolished the activity of CalE8. The HMS^{970} triad constitutes part of loop-1 with the solvent-exposed side chain of His^{968}, instead of the typical Gly residues conserved among many characterized ACPs (**Figure 3.4**). In addition, Ser^{970}

replaces an Asp residue commonly found in other type I and II ACPs. This Asp from *B. subtilis* FAS ACP forms a salt bridge with the Arg^{14} of ACP synthase (ACPS) [146]. Thus considering that the residues preceding Ser^{971} have been known to affect protein-protein interaction in some ACPs and given the different physio-chemical property of the residues, a different protein surface of *me*ACP may be involved in the recognition of some of its partners.

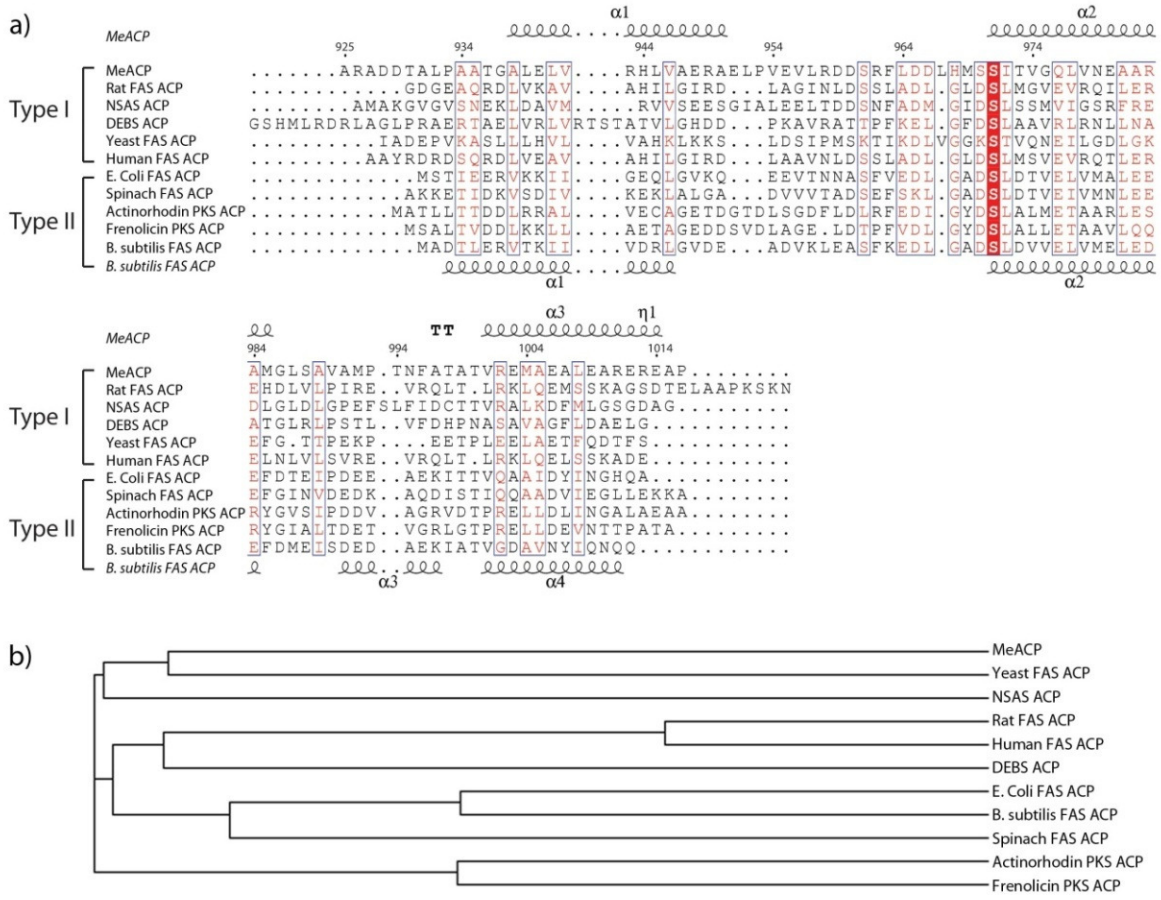


Figure 3.3 Comparison of *me*ACP with other type I and type II ACPs. (a) Multiple sequence alignment of *me*ACP with selected type I and type II ACPs. The secondary structures of *me*ACP (top) and *B. subtilis* FAS ACP (bottom) are shown. The phosphopantetheine attaching serine site is shaded in red. (b) Phylogenetic analysis of the type I and II ACPs from (a).

One of the common features of types II FAS or PKS ACPs is that they contain a helix α2 that is rich in acidic residues. The clustered negative charges on helix α2 are considered to be

critical for the recognition of the free-standing ACPs by their protein partners. Helix α2 is therefore known as the 'recognition helix' in carrier proteins. Several acidic residues in helix α2 of type II ACPs can form salt bridges with positively charged residues on their protein partners [147]. However, in type I FAS ACP domains the helix α2 seems to be less negatively charged, presumably because the specific domain-domain interaction driven by electrostatic interactions is less critical in the multidomain type I FAS system. In accordance with other type I ACPs, the helix α2 of *me*ACP only contains a single acidic residue (Glu^{980}). The resemblance of *me*ACP to type I ACPs is also consistent with the phylogenetic relationship of *me*ACP with other carrier proteins (**Figure 3.3b**).

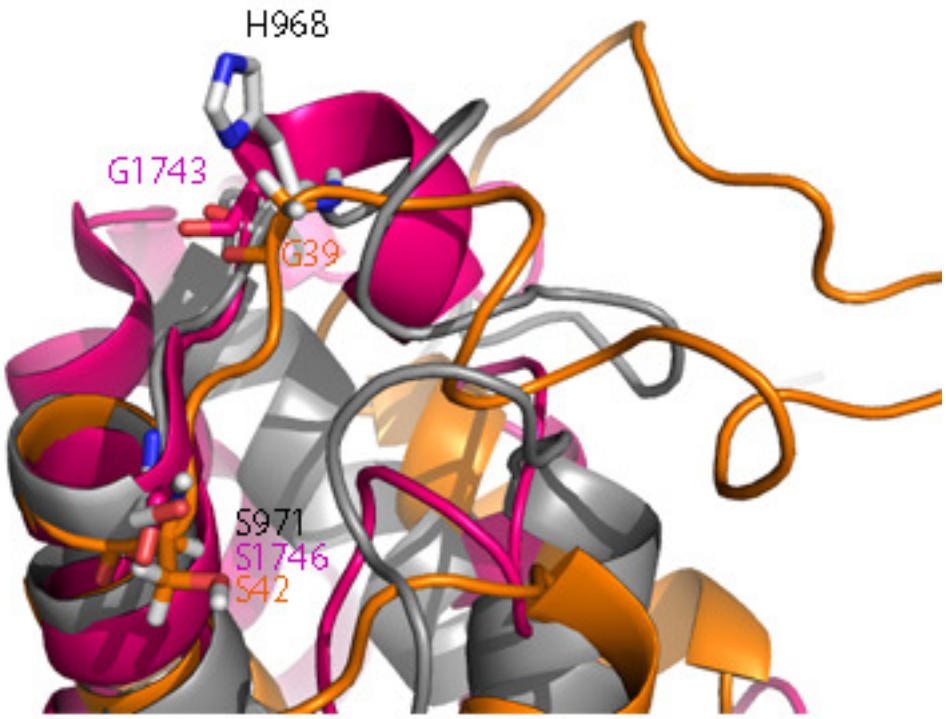


Figure 3.4 Conformation of the motif harboring the phosphopantetheine-attaching serine in *me*ACP and two other ACPs. The GX(H/D)S(L/I) motif in type I NSAS PKS ACP (PDB ID: 2kr5) (pink) and type II actinorhodin PKS ACP (PDB ID: 1af8) (orange) and the HMSSI motif in *me*ACP (grey) are shown as ribbons. The conserved Gly and Ser residues in **G**X(H/D)**S**(L/I) motif are shown as stick. The His and Ser residues of the **H**MS**S**I motif in *me*ACP are also shown as stick.

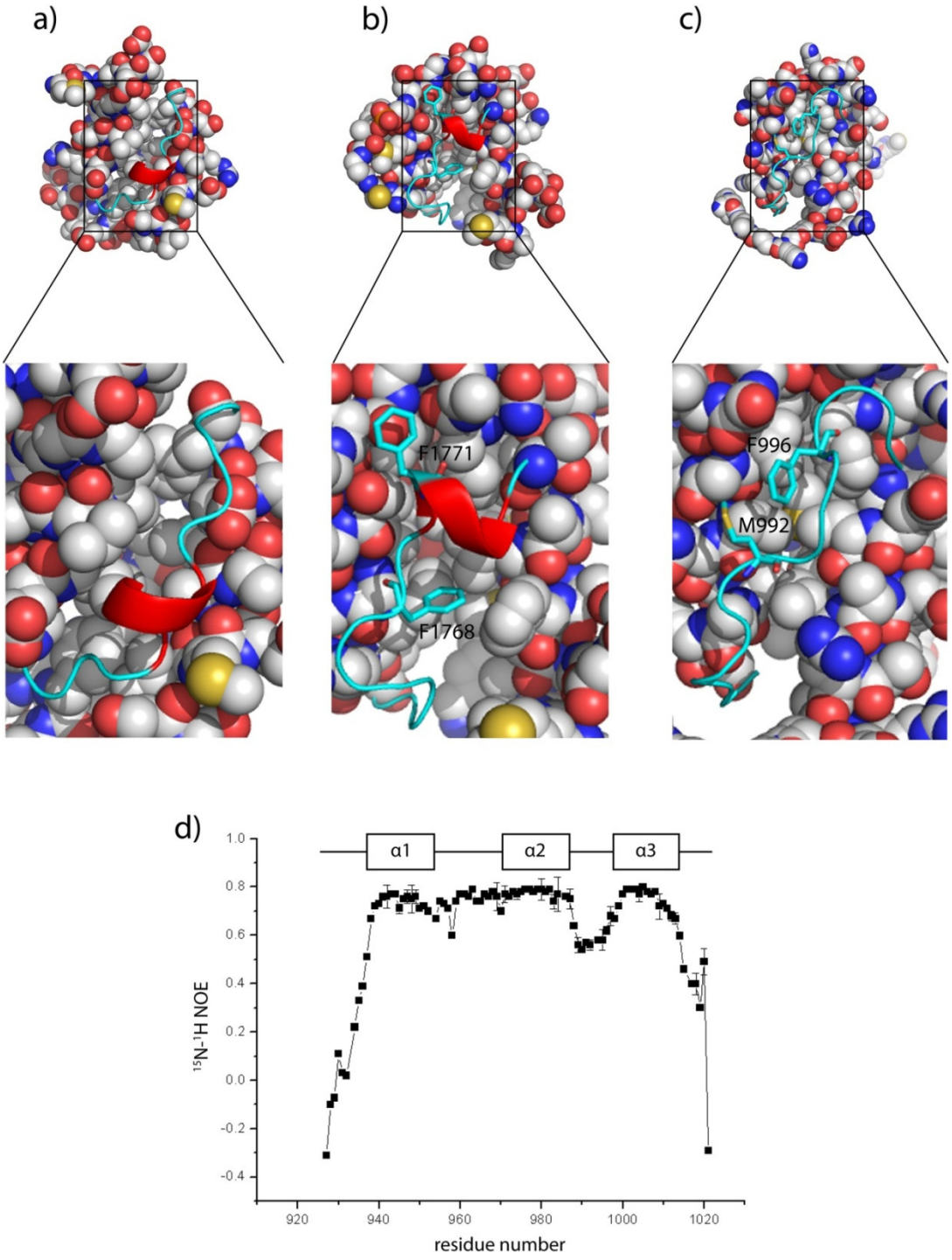


Figure 3.5 Conformation mobility of loop-2 in *me*ACP compared to other ACPs. (a) Type II actinorhodin PKS ACP (PDB ID: 1af8) that contain mainly small residues in loop-2. (b) Type I NSAS PKS ACP (PDB ID: 2kr5) with the two bulky residues (Phe^{66} and Phe^{69}) highlighted (c) *me*ACP with the two bulky residues (Phe^{996} and Met^{992}) packed against the hydrophobic pocket highlighted. (d). ^{1}H-^{15}N NOE plot of *apo-me*ACP acquired at 25 °C. The positions of the helices are indicated.

In contrast to the relatively neutral helix α2, the adjacent loop-1 harbors a large number of highly charged residues including Glu^{951}, Glu^{955}, Arg^{958}, Asp^{959}, Asp^{960}, Arg^{962}, Asp^{965}, Asp^{966} and His^{968}. The enrichment of the charged residues on loop-1 is not uncommon and has also been seen in ACPs, such as the ones from the biosynthetic pathways of frenolicin and norsolorinic acid [137, 148]. It has been suggested that loop-2, which contains a short helix for some ACPs, serves as a conformational switch in sequestering ACP domain-tethered acyl chains in type II FAS systems [117]. With the exceptions of *act* ACP and *B. subtilis* FAS ACP [118, 128], all the characterized ACPs contain a short helix within loop-2. The short helix seems to be absent in *me*ACP as well. In the case of type II ACPs, loop-2 is largely comprised of charged or small non-polar residues that allow the opening of a hydrophobic cleft for binding acyl intermediates (**Figure 3.5a**); Whereas loop-2 of type I ACPs, such as NSAS ACP, usually contain bulky residues that are packed against the protein core to prevent the opening of the binding cleft (**Figure 3.5b**)[137]. In *me*ACP, loop-2 also contains two hydrophobic residues (Met^{992}, Phe^{996}) with their bulky side chains packed against the hydrophobic core of the protein (**Figure 3.5c**). Although the packing of Met^{992} and Phe^{996} may restrict the conformational flexibility of loop-2, ^{1}H-^{15}N NOE plot of *me*ACP indicates that the residues of loop-2 (Gly^{986} -Thr^{1000}) are still significantly more flexible than the residues of loop-1 and helices α1, 2 and 3 (**Figure 3.5d**). The observed conformational flexibly of loop-2 raises the tantalizing possibility that it may undergo a large conformational change to open a protein pocket for binding of the PKS products.

3.3.3 Comparison between *Holo-* and *Apo-Me*ACP

ACP domain only becomes fully functional after conversion to *holo* protein with a phosphopantetheine group attached to the invariant Ser residue. To find out whether the presence of the phosphopantetheine group perturbs the *me*ACP structure and dynamics, the *holo* form of

the protein was prepared by incubating *me*ACP with coenzyme A and the PPTase Sfp. Comparing the HSQC spectra of *apo*- and *holo-me*ACP, we found that the overall structures of the two forms are essentially the same (**Figure 3.6**).

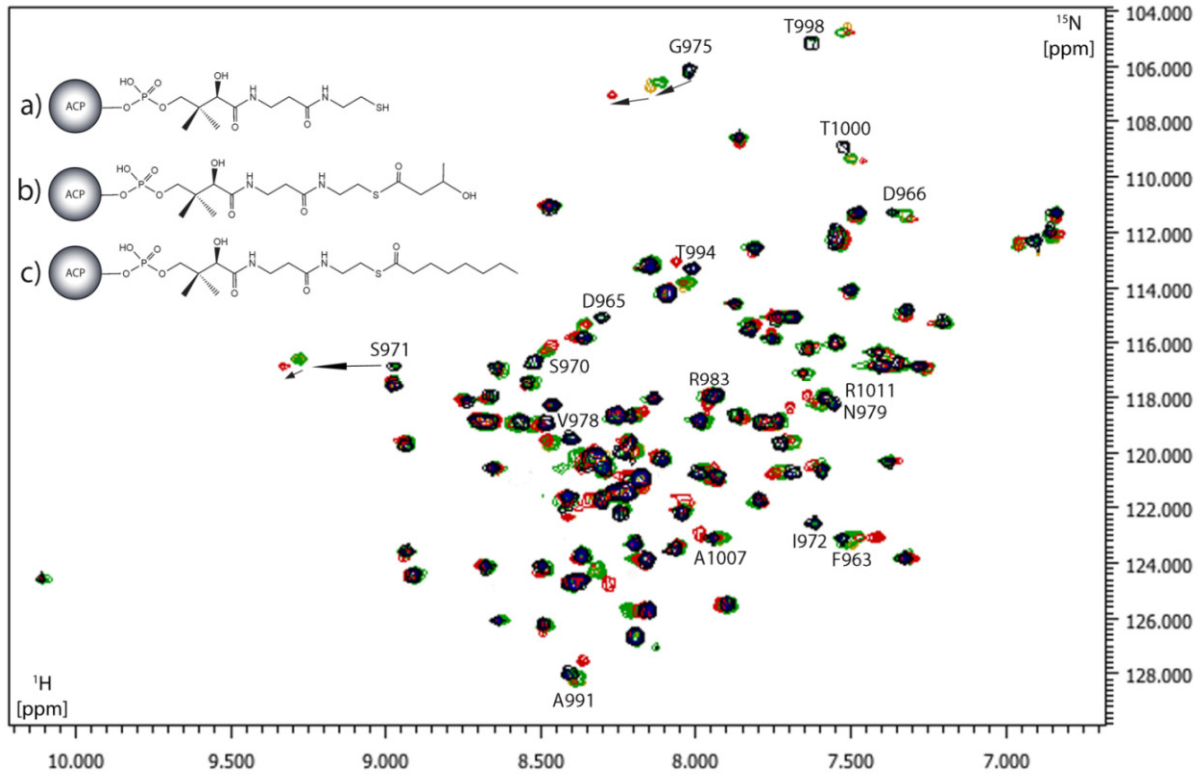


Figure 3.6 Overlaid ^{1}H-^{15}N HSQC spectra of *apo-me*ACP (black), *holo-me*ACP (yellow), *hydroxybutyryl-me*ACP (green) and *octanoyl-me*ACP (red). The change of chemical shifts for the residues Ser^{971} and Gly^{975} from *apo*, *holo*, *hydroxybutyryl* to *octanoyl-meACP* are indicated by the black arrows.

Nonetheless, Ser^{971} exhibits a large shift from the *apo* form (^{1}H: 8.97 ppm, ^{15}N: 116.80 ppm) to the *holo* form (^{1}H: 9.28 ppm, ^{15}N: 116.50 ppm) or a combined chemical shift change of 0.31 ppm. The presence of the phosphopantetheine group also clearly affects backbone amides of the neighboring residues that include Asp^{965} and Ser^{970} from loop-1, Ile^{972}, Gly^{975} and Val^{978} from helix α2 and Thr^{994}, Thr^{998}, Ala^{999} and Thr^{1000} from loop-2 (**Figure 3.7a** and **g**). These perturbed residues are mostly located at protein surface in the vicinity of Ser^{971} (with the exception of

Val^{978} that is probably due to a subtle conformational change in helix α2). The observation indicates that the *me*ACP-attached phosphopantetheinyl group is most likely exposed to solvent and accessible to other catalytic domains.

3.3.4 Interaction of Acylated Moiety with *Me*ACP

One of the most interesting findings about PKS/FAS and NRPS is the presence of a ligand binding pocket on the surface of type II ACPs or PCPs for the binding of the growing intermediates. For example, the sequestering of aliphatic segment of decanoate (10:0) or stearate (18:0) in the binding pocket of spinach ACP induces striking perturbations of HSQC peaks [149, 150]. On the other hand, hexanoyl (6:0) or palmitoyl (16:0) chain attached to type I rat FAS ACP showed little changes of HSQC peaks, suggesting that the rat FAS ACP does not bind the aliphatic chain [121].

Given the observed conformational flexibility of loop-2 in *me*ACP, we asked whether type I *me*ACP can also bind the chemically labile polyene intermediates in the same hydrophobic pocket that accommodates the side chains of Met^{992} and Phe^{996}. Loading *me*ACP with the highly conjugated polyenes **2** or **3** is impractical given the chemical reactivity of the polyenes. Instead, we used hydroxybutyryl- and octanoyl- groups to mimic the aliphatic chains of the polyene products. Overall, only minor perturbations in ^{15}N-HSQC spectrum were observed between the *holo* and *acylated-me*ACPs, suggesting that residues lining the putative binding pocket are not disturbed (**Figure 3.6**). However, although only negligible differences between the *holo*- and *hydroxybutyryl-me*ACP (combined chemical shift differences < 0.05 ppm) were observed, extending the acyl chain from hydroxybutyryl to octanoyl group causes some greater and more extensive chemical shift changes (0.05 - 0.20 ppm) (**Figure 3.7b-f**).

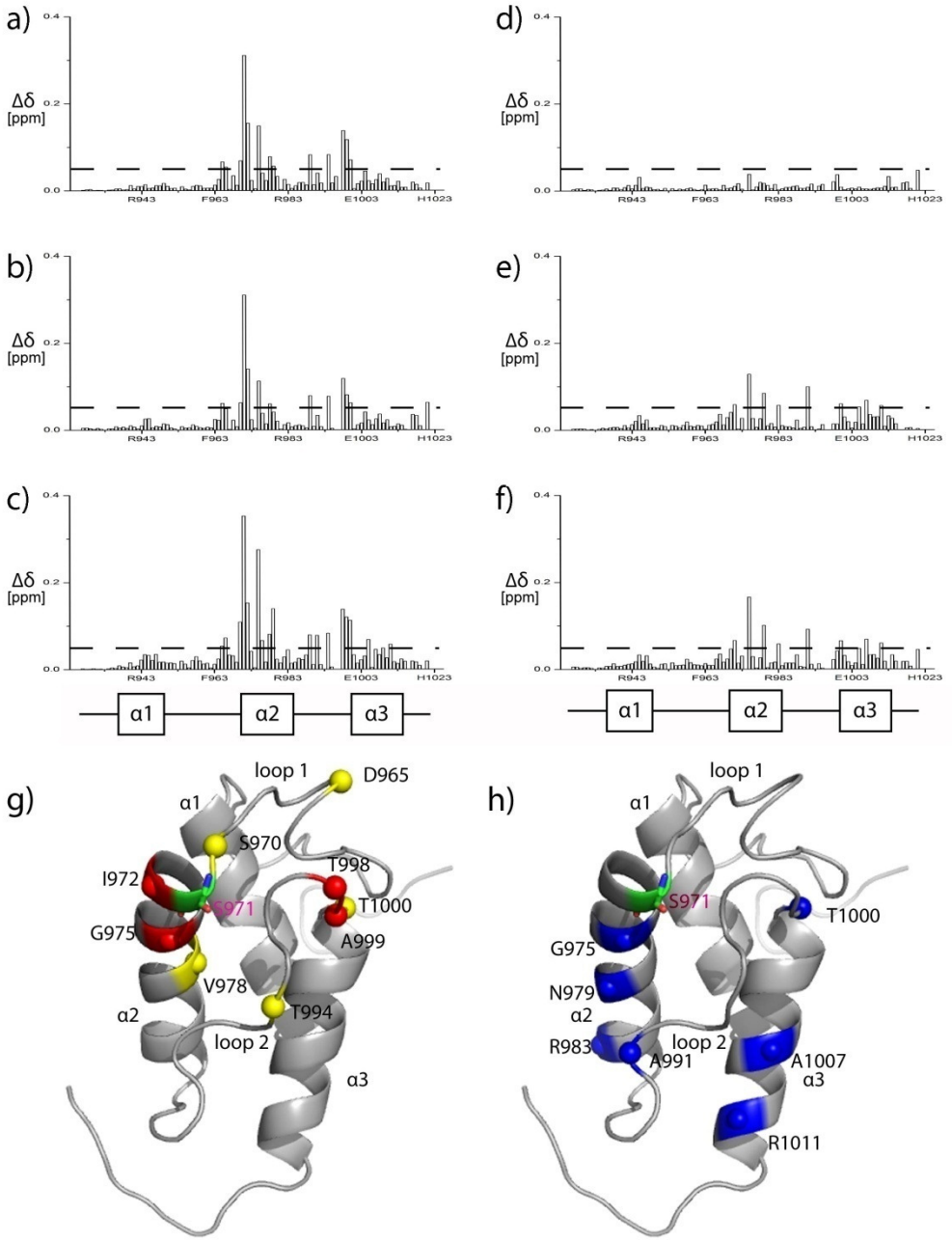


Figure 3.7 Chemical shift perturbations (Δδ) caused by phosphopantetheinylation and acylation. Combined chemical shift change plots of *apo-me*ACP versus (a) *holo-me*ACP, (b) *hydroxybutyryl-me*ACP and (c) *octanoyl-me*ACP. Combined chemical shift change plots of (d) *hydroxybutyryl-me*ACP versus *holo-me*ACP, (e) *octanoyl-me*ACP versus *holo-me*ACP and (f) *octanoyl-me*ACP versus *hydroxybutyryl-me*ACP. A combined chemical shift cut-off of 0.05ppm is shown as dashed line. The residue number and the positions of the helices (α1, α2 and α3) are also indicated. (g) Ribbon representation of *apo-me*ACP with the positions of the perturbed residues upon phosphopantetheinylation are highlighed (combined chemical shift cut-off > 0.05 ppm (yellow sphere) and > 0.10 ppm (red sphere)). (h) Ribbon representation of *apo-me*ACP

with the positions of the perturbed residues based on (f) upon acylation by octanoyl moiety are highlighted (combined chemical shift cut-off > 0.05 ppm (blue sphere)).

With a chemical shift change cut-off of 0.05 ppm, eight residues that include Ser^{971}, Gly^{975}, Asn^{979}, Arg^{983}, Ala^{991}, Thr^{1000}, Ala^{1007} and Arg^{1011} exhibit greater perturbation in *octanoyl-me*ACP than *hydroxybutyryl-me*ACP. Note that the three residues Ser^{971}, Gly^{975} and Thr^{1000} are affected by both phosphopantetheinylation and extension of the acyl chain, indicating that elongation of acyl chain may also alter the position or conformation of the phosphopantetheinyl moiety. These perturbed residues of *octanoyl-m*eACP are mapped on the protein and most of them are located on helix α2, loop-2 and as well as helix α3 (**Figure 3.7h**).

3.3.5 Interaction between *Me*ACP and the Thioesterase CalE7

*Me*ACP domain must interact with the ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase (DH) and PPTase domains of CalE8 for the synthesis of the polyene product as well as the thioesterase CalE7 for hydrolytic release of the product. It would be interesting to know the molecular basis of the interaction between *me*ACP and the catalytic domain of CalE8 as well as CalE7. However, cloning and expression of the stand-alone catalytic domains of CalE8 failed to produce soluble or stable proteins. We were only able to examine the interaction between *me*ACP and the hotdog fold thioesterase CalE7 in an attempt to map out the recognition surface of *me*ACP for CalE7. When we first determined the crystal structure of CalE7, a few charged residues at the entrance of the substrate-binding tunnel were speculated to be involved in binding *me*ACP (the structure of CalE7 will be discussed later in Chapter 4) [138]. However, when the thioesterase CalE7 was titrated into the ^{15}N-labelled *apo-* and *holo-me*ACP protein solution for NMR measurement, no HSQC peak was perturbed in both chemical shifts and intensity even when the ratio of CalE7 to *me*ACP reached 2.5:1. This indicates that the

interaction between CalE7 and *me*ACP is too transient or weak to be detected (**Appendix figure 3.1** and **Appendix figure 3.2**). The transient nature of the CalE7 and *me*ACP interaction is further supported by the observation that none of the charged residues at the substrate-binding channel is conserved in the homologous DynE7 protein [151]. In conjunction with the observation that a catalytic incompetent mutant of CalE7 (but not the wild type CalE7) can form stable 1:1 complex with CalE8 [151], it is reasonable to believe that the interaction between *me*ACP and CalE7 is rather non-specific and the recognition is through the binding of the polyene product in the hydrophobic substrate channel of CalE7.

3.4 Discussion

As the first structure determined for the ACP domain of a highly reducing (HR) iterative type I PKS, *me*ACP shares extremely low sequence homology with other known ACPs, including the ACP domain of the non-reducing (NR) iterative type I PKS NSAS (12% identity). *me*ACP also features an unusual HMS**S**I motif (or H(L/M)(S/T/N)**S**(I/L) for *me*ACP homologs) harboring the Ser^{971} phosphopantetheinylation site, in contrast to the more common GX(H/D)**S**(L/I) motif seen in other ACPs. Despite the low sequence homology and the uncommon HMSSI motif, *me*ACP adopts a helix-bundle structure that is highly similar to other ACPs with the exception of the absence of the short helix within loop-2. As a result of the missing short helix, the structure of *me*ACP is best described as a twisted three-helix bundle instead of the canonical four helix-bundle structure. In addition, the relatively shortage of negative charges on the 'recognition helix' (α2) of *me*ACP is consistent with the observations for other type I FAS and PKS ACPs, presumably resulted from the lack of evolutionary selection pressure on type I ACP domains for specific domain-domain interactions within the multidomain FAS or PKS protein.

It is well known that loop-2 of type II FAS and PKS ACPs generally consists of charged and less bulky residues to confer great conformational mobility to the loop; whereas type I ACPs tend to contain bulky hydrophobic residues on the loop to restrict the mobility of the loop through hydrophobic interaction with the protein core. In light of this, *me*ACP harbors two non-polar residues (Met^{992} and Phe^{996}) with the side chains packed against the protein core. Similar hydrophobic interactions were recently reported for the ACP domains of the rat type I FAS, the non-reducing (NR) type I iterative PKS NSAS and the type I modular PKS DEBS [121, 137]. The protection of the intermediates is considered to be particularly important for type II FASs

and PKSs because the free-standing ACPs need to transport the intermediates from one protein to another in bulk solvent. The conformational mobility of loop-2 has been suggested to be crucial for the opening of a hydrophobic cleft for sequestering and protecting acyl intermediates. Such protection, however, is less critical for type I FASs and PKSs with the integrated ACP domain and intermediates probably already shielded from solvent by the surrounding catalytic domains. The structural observations thus seem to reinforce the view that restricted conformational mobility of loop-2 resulted from interaction of the non-polar residues may be a common feature for the ACP domains of the multidomain type I FAS and type I iterative PKS.

Despite the hydrophobic interaction between Met^{992} and Phe^{996} with the protein core, the ^{1}H-^{15}N NOE plot revealed relatively greater conformational flexible for loop-2 than loop-1 and the three helices (**Figure 3.5d**). This observation first raised the question whether loop-2 of *me*ACP still can undergo conformation change to create a transient protein pocket for binding the highly hydrophobic polyene products. Attachment of 4'-phosphospanteine group and the additional hydroxybutyryl group to the Ser^{971} site has little effect on the overall protein fold, despite minor local perturbations restricted to residues in the vicinity of Ser^{971}. These observations suggest that the phosphopantetheine and hydroxybutyryl moieties are unlikely to be bound in a protein pocket and thus remain in the bulk solvent. In contrast, the NMR spectra of the *octanoyl-me*ACP show more extensive perturbations, with the perturbed residues distributed in a region flanked by helix α2, loop-2 and helix α3 (**Figure 3.7h**). The magnitude of the perturbation is considered to be relatively small compared to that for the type II ACPs that sequester the acyl chain through a significant conformational change [125, 152].

Hence, the results do not seem to support that *me*ACP is able to undergo dramatic conformational change and sequester the PKS product in a well-defined protein pocket. Instead,

examination of the electrostatic potential surface of *me*ACP reveals a distinct hydrophobic patch that encompasses the eight residues that exhibit perturbations (**Figure 3.8a** and **d**). No such distinct hydrophobic patch can be found on the surfaces of other type I ACPs, such as the rat FAS ACP and type I NSAS ACP (**Figure 3.8**) [137]. The observed NMR perturbations and the presence of the distinct hydrophobic surface on *me*ACP suggest that the non-polar octanoyl chain may interact with hydrophobic patch to avoid the unfavorable solvation of the hydrophobic groups. Given the highly hydrophobic nature of the polyenes produced by CalE8 (**Figure 3.1**), it is reasonable to expect that the polyene can interact with the hydrophobic patch as well. Such interaction may play a significant role in the stabilization of the chemically labile polyene. Alternatively, the hydrophobic patch could be conserved for domain-domain interaction. The latter remains as a possibility since this region also encompasses the helix α2 that is most likely to be involved in domain-domain interaction.

In summary, although the biosynthetic mechanism for the 10- and 9-membered enediyne moieties and the precise function of CalE8 and its homologs are still under intensive investigation, the first structure of the HR iterative PKS ACP domain presented here reveals some interesting structure features that reinforce some of the views held for type I FAS and PKS ACPs. The structure also discloses some unique characteristics of the *me*ACP domain. The lack of a strong negatively charged helix α2 and the moderate conformational flexibility of loop-2 are consistent with the role of the *me*ACP domain in shuttling acyl intermediates within the catalytic chamber of the multidomain PKS CalE8. The absence of the short helix within loop-2 is rare but not totally unprecedented. Although the biologically significance of the interaction of the non-polar acyl chain with the hydrophobic patch remains to be established, the interaction is distinct from the well-established binding of acyl chains inside a well-defined protein pocket. Meanwhile,

no significant interaction was observed between the ACP domain and the discrete thioesterase CalE7. The recognition of the *me*ACP-tethered polyene product by CalE7 is likely to take place through the binding of the nascent polyene in the substrate-binding channel of the thioesterase.

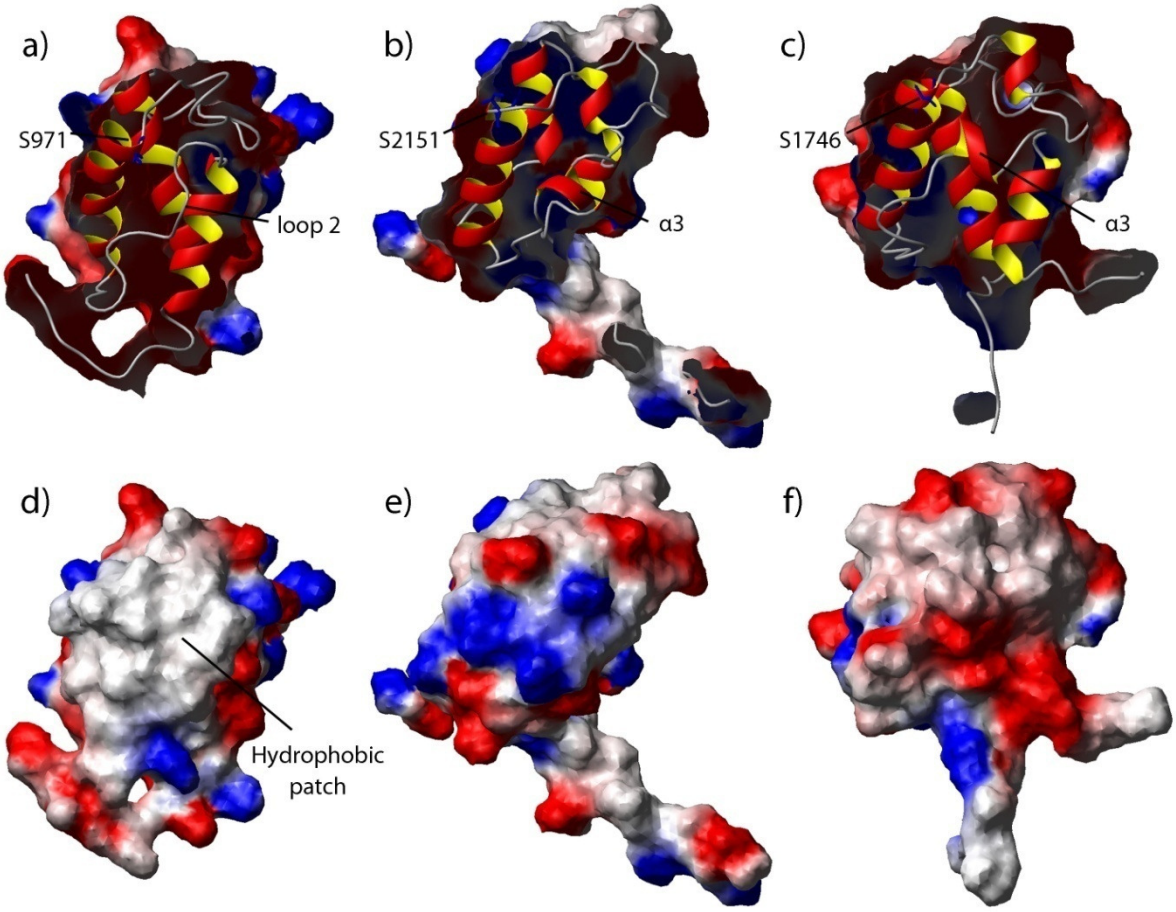


Figure 3.8 Embedded ribbon representation and electrostatic potential surface of *me*ACP. (a and d) rat FAS ACP (b and e, PDB ID: 2png) and NSAS PKS ACP (c and f, PDB ID: 2kr5). The invariant Ser is shown as blue stick and labeled. The protein surfaces are colored as white (hydrophobic), blue (basic) and red (acidic) with the same electrostatic potential scale.

CHAPTER 4 Structure and Catalytic Mechanism of the Thioesterases CalE7 and DynE7 in Enediyne Biosynthesis

4.1 Introduction

Polyketide and nonribosomal peptide synthesis generally involves the so-called type I and type II thioesterases for the release of final product or removal of aberrant products. Type I thioesterases (TE I) are *cis*-acting domains fused to the C-terminus of the most downstream module of PKS or nonribosomal peptide synthase (NRPS) for the release and cyclization of the final product [153, 154]. In contrast, type II thioesterases (TE II) are discrete proteins responsible for the *trans* hydrolytic release of aberrant products [98, 100, 155]. TE II proteins are structurally and evolutionarily related to a family of well-known α/β hydrolase that contain 240 - 260 residues [156]. A common serine esterase motif GXSXG and another downstream motif GXH are conserved in TE II proteins [157, 158]. Based on the sequence information, the stand-alone CalE7 and the homolog DynE7 from the pathway of the 10-membered enediyne natural product dynemicin do not belong to the TE II family because they are neither α/β fold hydrolases nor proteins containing the two conserved motifs for TEII. Instead, CalE7 and DynE7 share moderate sequence homology with a family of hotdog fold proteins characterized by a long central α-helix packed against a five-stranded anti-parallel β-sheet. Such hotdog fold proteins include many characterized and hypothetical thioesterases that use acyl-CoA as substrates [101]. The three-dimensional structure and substrate specificity of several hotdog fold thioesterases have been determined, including YbgC from *Helicobacter pylori* [159], the homotetrameric

medium chain acyl-CoA thioesterase II from *E. coli* [160] , Paal from *Thermos thermophilis* HB8 [161], FcoT from *Mycobacterium tuberculosis* [162], YciA from *Haemophilus influenza* [163], human THEM2 [164] and 4-hydroxylbenzoyl-CoA thioesterases (4-HBT) from *Pseudomonas sp.* Strain CBS and *Arthrobacter sp.* strain SU [165-167]. Despite their diverse specificity towards acyl substrates, almost all known hotdog fold thioesterases catalyze the hydrolysis of thioester bond by using a Glu/Asp residue as nucleophile or general-base catalyst [101, 164].

In this chapter, we present X-ray crystallographic and biochemical data to show that CalE7 and the homolog DynE7 from the dynemicin biosynthetic pathway do not contain catalytically critical acidic residues in the active site. Thus the two hotdog fold proteins are likely to employ a novel catalytic mechanism for thioester hydrolysis, in comparison to the canonical type I and II thioesterases in the biosynthesis of many polyketide natural products. From the data collected, the crystal structure of CalE7 in its closed conformation displays a kinked substrate-binding channel which is too short to accommodate a full length product. On the other hand, despite the overall structural similarity to CalE7, the DynE7 crystal reveals an extended substrate-binding channel with the carbonyl-conjugated polyene (**3**) trapped inside. Thus, the comparison between the two closely-related homologs revealed a remarkable induced-fit mechanism for the substrate-binding of the *TEBC* family proteins and suggested a novel catalytic mechanism for this subfamily of thioesterases.

4.2 Material and Methods

Materials

Coenzyme A (CoA), acetyl-CoA, malonyl-CoA, NADPH and other chemicals were purchased from Sigma-Aldrich and stored at -20 ^{o}C. Expression vectors pET-28b(+), pET-26b(+), pCDF-2, and *E. coli* BL21(DE3) were obtained from Novagen.

Cloning, Protein Expression and Purification for CalE8, Wildtype and Mutants of CalE7

Cloning, expression and purification of CalE8 and CalE7 have been described in details in chapter 2 [108]**.** Site-directed mutagenesis of CalE7 was performed using QuickChange II kit (Stratagene) following the manufacturer's instructions and with the mutations confirmed by DNA sequencing. The mutant proteins were expressed and purified following the same procedure as the wild type protein.

Cloning, Co-Expression and Co-Purification of DynE8 and DynE7

The DNA fragment encoding DynE8 was sub-cloned into pET-28b(+) between *Nde*I and *Xho*I sites and transformed into *E. coli* BL21(DE3). At the same time, gene encoding DynE7 was sub-cloned into pCDF-2 according to the instruction manual for pCDF-2 Ek/LIC kit (Novagen). pET-28b(+)-DynE8 and pCDF-2-DynE7 were co-transformed into *E. coli* BL21(DE3). Cells were grown at 37^{o}C in LB with kanamycin and streptomycin selection and protein expression was induced with 0.5 mM IPTG for 20 h at 16 °C. The pelleted cells co-expressing DynE8 and DynE7 were sonicated in a buffer containing 50 mM HEPES, pH 8.0, 500 mM NaCl, 20 mM imidazole, 15 % glycerol, and 10 mM β–mercaptoethanol. Due to the presence of His_6-tag on both DynE8 and DynE7, the purification step by Ni^{2+}-NTA affinity column yielded a mixture of both proteins. The solution containing the DynE8 protein is colorless whilst that of DynE7 is

yellow. In the latter case, this is due to the presence of the linear conjugated polyene released from DynE8. The two proteins were subsequently separated and purified by ion exchange followed by size-exclusion chromatography.

Site-Directed Mutagenesis on DynE7

Site-directed mutagenesis of DynE7 was performed using the QuickChange II kit (Stratagene) following the manufacturer instructions. Two single mutants, DynE7-R35A and DynE7-E36A, harboured by pCDF-2 Ek/LIC were subsequently confirmed by DNA sequencing. The mutant proteins were separately co-expressed and co-purified with pET-28b(+)-DynE8 following the same procedure as for the wild-type protein. The absorption spectra of each co-purified DynE7-R35A and DynE7-E36A proteins were measured using a semi-micro cuvette on a Shimazu 1700 UV-Vis spectrophotometer.

Activity Assay and Product Analysis by Absorption Spectroscopy and HPLC for CalE7 and Its Respective Mutants

Enzymatic assay was carried out with 7.2 μM CalE8, 27.8 μM CalE7 (or CalE7 mutant), 0.15 mM acetyl-CoA, 1.05 mM NADPH and 1.05 mM malonyl-CoA in 900 μL of pH 8.2 buffer (100 mM Tris, 300mM NaCl, 1mM DTT). The reaction mixture in the absence of malonyl-CoA was equilibrated at 30°C for 15 min in the temperature-controlled sample chamber of Shimazu UV-Vis 1700 spectrometer. Enzymatic reaction was initiated with the addition of malonyl-CoA, and was carried out for 2 hours at 30°C with values for OD_{384} and OD_{409} taken at 5 minute intervals. For the analysis of products by HPLC, the yellow reaction mixture was extracted by adding ethyl acetate and resulting organic extract was evaporated by using a Speed-Vac. The dried sample was re-dissolved in methanol before application onto an analytical eclipse XDB C18 column (4.6 × 250mm) 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 (acetonitrile with 0.045% TFA) in 60 minutes. HPLC UV detector was set at multiple wavelengths, such as 280nm, 375nm, 400nm, 420nm with a reference wavelength of 600nm.

Kinetic Measurement of Reaction Involving DynE8, DynE7 and its Respective Mutants

Steady-state kinetic measurement was carried out with 3.2 µM DynE8, 66.8 µM DynE7-R35A or DynE7-E36A, 1.4 mM malonyl-CoA and 0.15 mM NADPH (saturating concentration) in 200 µL of HEPES buffer (100 mM HEPES (pH 8.5), 300 mM NaCl and 1 mM DTT) at 37 °C using a semi-micro cuvette. Enzymatic reaction was initiated by addition of malonyl-CoA. A full wavelength spectrum scan was performed at every 2 minutes interval throughout the course of the experiment using Shimazu 1700 UV-Vis spectrophotometer.

Analysis of the Product Bound by DynE7 in Solution and Crystal by HPLC

An equal volume of ethyl acetate was added into the yellowish DynE7 protein solution, which was co-purified with DynE8, and vigorous vortexing ensued. Subsequently the mixture was centrifuged at 15,000 rpm for 15 minutes. The organic solvent layer was then pipetted-out and the protein solution was subjected to extraction for a second time. Ethyl acetate extracts were pooled and the solvent was evaporated using a Speedvac centrifuge. The yellow pigment was re-dissolved in a small volume of methanol before application onto an analytical eclipse XDB RP C18 column (4.6 × 250 mm) using an Agilent 1200 HPLC. The gradient employed was from 70% buffer A (HPLC grade water with 0.045% TFA) and 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B in 60 minutes. The diode-array detector was set at 410 nm with a reference wavelength of 600 nm. For liganded DynE7 crystals, the bound product was extracted from dissolved crystals following the procedure described above. The standards, 1,3,5,7,9,11,13-

pentadecaheptanane (**2**) and 3,5,7,9,11,13-pentadecen-2-one (**3**), were produced by the *in vitro* enzymatic reaction with CalE8 and CalE7.

Crystallization, X-ray Diffraction Data Collection, Structure Solution, Refinement and Model Analysis for CalE7 and DynE7

The X-ray related experiments were done by our collaborators in Dr. Julien Lescar's laboratory [138, 151]. Dr. Masayo and Dr. Insaf were in charge of the crystallization and structural calculation of CalE7, while crystallization and structural calculation of DynE7 was chiefly done by Mr. Liew Chong Wai. Subsequently analysis of structure, correlation between structural aspects and biochemical data as well as proposal of catalytic mechanism were mainly done by our lab. A brief description of X-ray diffraction related experimental procedures is given below:

Prior to crystallization, CalE7 purified from co-expression with CalE8 was buffer exchanged and concentrated to ~15 mg/ml in 50 mM potassium phosphate (pH 6.5), 150 mM NaCl using Amicon centrifugal concentrators (Millipore). An automated initial crystallization screen of 672 conditions was performed using the CyBio®-Crystal Creator (Jena Biosciences). A volume of 200 nl of purified CalE7 was added to an equal volume of crystallization solution using the sitting drop vapor diffusion method. Crystals were obtained at 291 K with Index Screen condition 38 (0.1M HEPES, pH 7.0, 30% Jeffamine-M600) (Hampton Research). Crystals up to 0.2 mm × 0.2 mm × 0.1 mm in size could be grown in 0.1M Hepes, pH 7.0, 28% Jeffamine-M600. Before data collection, crystals were transferred to a cryoprotectant of 0.1 M HEPES, pH 7.0, 28% Jeffamine-M600 and 25% glycerol and cooled to 100 K in a gaseous N_2 stream using an Oxford cryosystem. Diffraction intensities were collected at the NSRRC (Hsinchu, Taiwan) to

a resolution of 1.75 Å using one single crystal and were integrated, merged and scaled using the HKL2000 suite.

The structure of CalE7 was solved using the molecular replacement software MrBUMP [168]with the crystal structure of a hypothetical protein (AQ1494) from *Aquifex Aeolicus* (PDB code 2EGI, RIKEN Structural Genomics/Proteomics Initiative) having an overall homology of only 28% of 126 aligned residues with CalE7, as a search probe. Automated model building was carried out with Arp/warp [169]. The structure was refined using molecular dynamics and simulated annealing as implemented in the program CNS [170], with positional and individual temperature factor refinement without NCS restrains. The computer graphics software O was used for model rebuilding between refinement cycles [171]. Analysis of the atomic model was carried out with the CCP4 suite of programs [172]. The refined coordinates and structure factor amplitudes have been deposited in the PDB (PDB code: 2W3X).

DynE7 proteins, from single- and co-expression systems, were purified with a similar procedure compared to that of CalE7. Purified DynE7 was concentrated to ~15 mg/ml in 25 mM HEPES, pH 7.0, 200 mM NaCl, 0.5 mM DTT, 5% glycerol. Initial crystallization screening produced crystals for the *apo*-DynE7 at 291K with PEG Ion Screen II condition 18 (8% tacsimate, pH 8.0, 20% w/v PEG 3350, Hampton Research), while crystals for the ligand-bound DynE7 protein were obtained at 291 K in condition 36 of the PEG Ion Screen I (0.2 M sodium tartrate dibasic dehydrate, 20% w/v PEG3350, Hampton Research). Diffraction intensities were collected at the NSRRC synchrotron (Hsinchu, Taiwan) to a resolution of 2.1 Å for ligand-bound DynE7 and to 2.7 Å for the *apo*-enzyme. Integration, scaling and merging of the intensities were carried out using programs MOSFLM and SCALA from the CCP4 suite [172].

Structures for the ligand-bound and for the *apo*-enzyme DynE7 were solved using the molecular replacement software Molrep from the CCP4 suite with the CalE7 protein from *Micromonospora echinospora spp. calichensis* (PDB: 2W3X) as a search probe. These two proteins have an overall identity of 45% for 150 aligned amino-acid residues. Automated model building was carried out with the program Chainsaw from the CCP4 suite. Refinement cycles carried-out using REFMAC5 without ncs restraints between the eight subunits of the two *apo*-tetramers and the four subunits of the ligandbound tetramer were interspersed with model rebuilding using Coot [173]. The program BUSTER was introduced for the last refinement steps [174]. The quality of the structures was analyzed using PROCHECK.

4.3 Results

4.3.1 Overall Structure of CalE7

We have previously observed the production of colored CalE7 when the protein was co-expressed with CalE8 [108]. The yellow coloration is due to the binding of the PKS product by CalE7. The colored CalE7 was used for crystallization in an effort to obtain the structure of the enzyme-product complex. However, despite the intensely yellow appearance of the protein solution, crystallization in the presence of Jeffamine M-600, a polyether amine, preferentially selected and crystallized the subpopulation of unliganded CalE7. As a result, only the structure of the ligand-free protein was solved with the presence of six molecules per asymmetric unit. The refinement yielded a R_{work} of 0.196 (R_{free} of 0.234) while retaining good stereochemistry. Missing amino-acids located at the N- and C-terminal ends of each monomer could not be traced presumably because of their high mobility in the crystal.

The monomer of CalE7 consists of a five-stranded antiparallel β-sheet and three α-helices. It exhibits a typical α/β hotdog fold with the five-stranded antiparallel β-sheet wrapping around the long α-helix α1 and the two shorter α-helices (α2 and α3) are found at N-terminus of the central helix (**Figure 4.1b**) [101]. The major conformational differences between the six independent monomers are observed in the β5-α3 loop where a displacement of ~ 2.8 Å is detected for one of the monomers (**Figure 4.3**). A search for similar protein structures revealed that CalE7 is more closely related to the members of YbgC-like thioesterase from *E. coli* and *H. pylori* [159] (PDB codes 1S5U and 2PZH respectively) with R.M.S.D. of 1.6 Å for 129 and 127 equivalent Cα atoms respectively upon superposition. As for 4-hydroxybenzoyl-CoA thioesterase of

Pseudomonas sp. strain CBS-3 (PDB code 1LO9, 1LO8 and 1LO7) [167], the superposition on CalE7 yielded an R.M.S.D. of 1.9 Å for 135 equivalent Cα atoms.

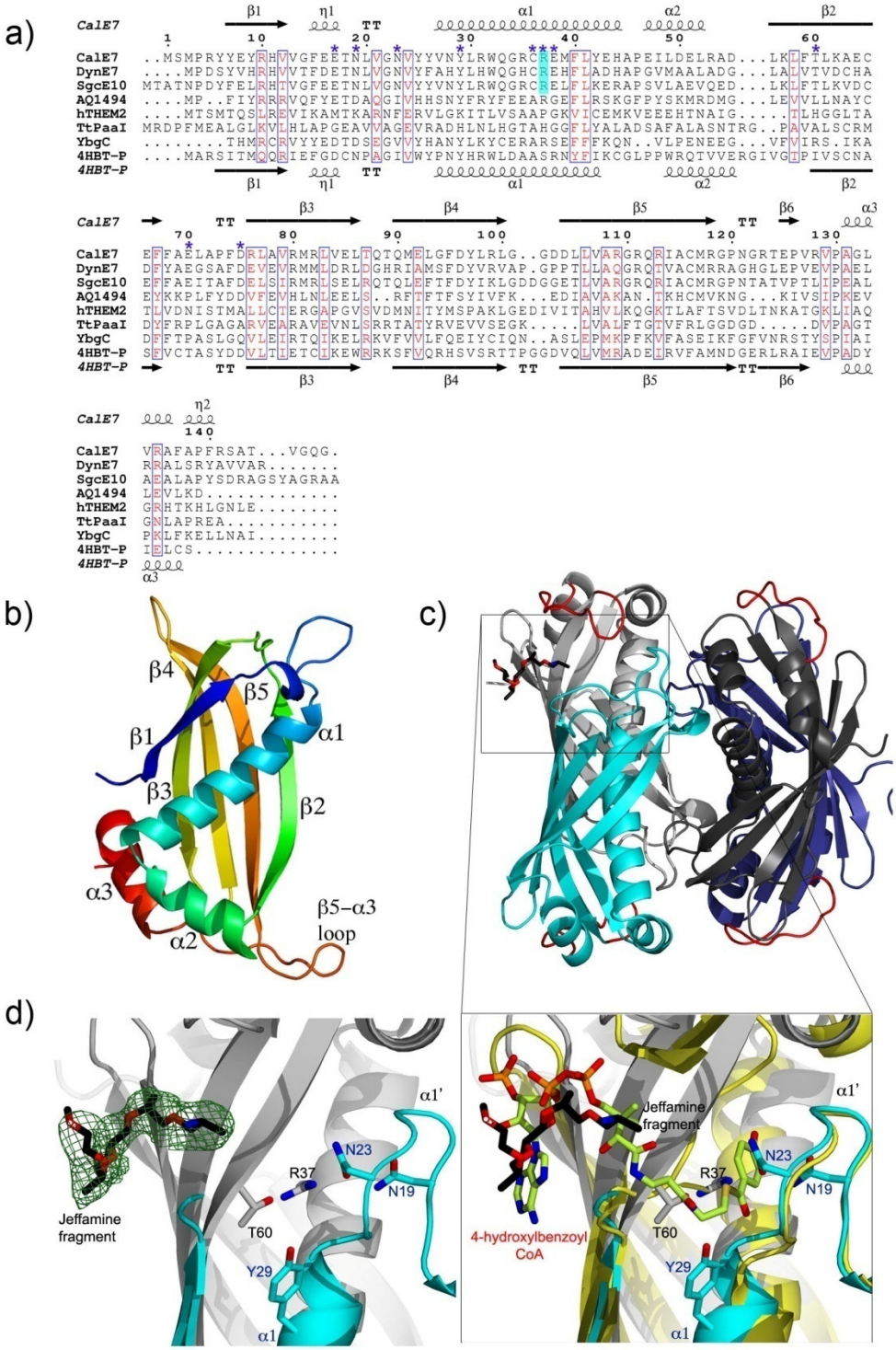


Figure 4.1 Sequence alignment and crystal structure of CalE7. a) Sequence alignment of CalE7 with DynE7, SgcE10 and some other hotdog fold proteins. The critical arginine (colored in cyan) is conserved throughout the *TEBC* family proteins. The residues mutated in the present study are indicated by blue asterisks. The secondary structures for CalE7 and 4HBT-*P* are shown at the top and the bottom of the alignment respectively. b) Representation of the CalE7 monomer. The protein secondary structures elements are labelled and coloured from blue at the N-terminus to red at the C-terminus. c) Cartoon representation of the CalE7 tetramer showing the Jeffamine fragment (sticks) binding site between two subunits. Subunits are coloured grey and cyan for one dimer and dark blue and dark grey for the other. The β5-α3 loops at the active sites are coloured red. The insert shows a close-up view of the Jeffamine binding site superimposed with 4-hydroxybenzoyl-CoA bound 4 BHT (PDB code: 1LO9). d) Simulated annealing difference Fourier map with coefficients $|F_{obs}|-|F_{calc}|$ and phases calculated from the protein model with atoms from the Jeffamine fragment omitted and contoured at a level of 3σ.

Based on size exclusion chromatography, CalE7 is a tetramer in solution. Six subunits were seen in the asymmetric unit: four form a 222 (or D_2) tetramer and two form a 2-fold symmetric dimer. A second complete 222 tetramer can be formed from the latter dimer via the crystallographic dyad. As seen for other hotdog fold thioesterases, dimerization proceeds through the formation of a continuous 10-stranded antiparallel β-sheet with the central helices on the inner side of the β-sheet (**Figure 4.1c**). This buried dimer interface (4540 $Å^2$) is stabilized via main-chain interactions between strands β2 and by residues 25 to 34 of the central helix α1 with their 2-fold counterparts, as well as by hydrophobic interactions between α2 of one subunit and the β1-α1 loop of the other subunit. Tetramer formation leads to the burying of the four central helices while the β-sheets are exposed to the solvent, as observed in other hotdog fold thioesterases with the εγ oligomeric arrangement [159, 165, 167]. A hydrophilic channel approximately 8 Å wide lies at the centre of the tetramer along a two-fold symmetry axis. Pyrophosphate molecules, possibly carried over from bacterial expression, are observed at both ends of this solvent accessible channel.

4.3.2 Substrate-Binding Pocket

An L-shaped putative substrate-binding pocket is visible at each dimer interface, over the β-sheet edge (**Figure 4.2a**). The location of the substrate-binding pocket agrees well with other characterized hotdog fold thioesterases. The inner channel extends inside the other subunit, surrounded by residues from helices α1, α2 and sealed by the short C-terminal helix α3 (**Figure 4.2a** and **b**), in particular the side chains of Leu^{133} and Phe^{137}. The putative substrate-binding channel is lined with hydrophobic residues that project from the two neighbouring subunits, as well as five polar residues that include Arg^{37} and Thr^{60} from one subunit and Asn^{19}, Asn^{23}, Tyr^{29} from the adjacent subunit (**Figure 4.2a** and **b**). An additional ~7 Å deep side pocket capped by the phenolic ring of Tyr^{29} faces the inner segment of the substrate-binding channel with several water molecules making hydrogen-bonds to residues Tyr^{29} (2.9Å), Asn^{19} (2.84Å), Arg^{37} (2.85Å) and Thr^{60} (2.7Å) (**Figure 4.2a**).

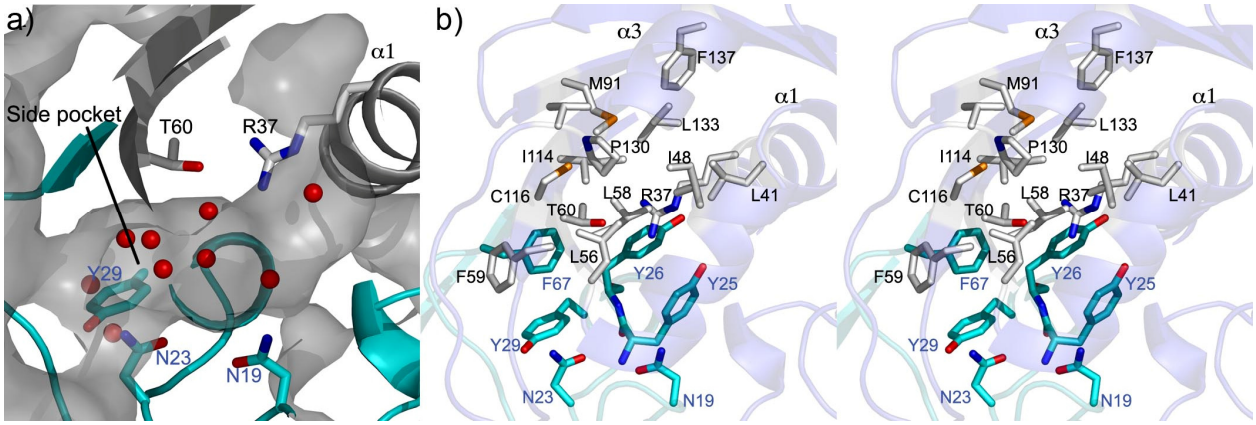


Figure 4.2 Substrate-binding pocket of CalE7. a) Cross-section through the surface of the elongated CalE7 substrate-binding pocket. The two subunits of CalE7 are shown in gray and cyan. The polar residues lining the substrate-binding channel and the water molecules (red dots) in the channel are shown. b) Stereoview of the residues lining the substrate-binding pocket.

As inferred from a superposition with substrate-bound structures of 4-HBT from *Pseudomonas* [167], the segment forming the entrance of the channel at the dimer interface is

likely to house the phosphopantetheinyl moiety of the substrate. Interestingly, only one (subunit E) out of the six substrate-binding pockets adopts an open conformation, with the substrate-binding channels of the other five subunits (A-D, F) shielded from bulk solvent by loop β5-α3. Structural comparison reveals that a movement of 2.8 Å of the β5-α3 loop results in the opening of the substrate-binding pocket to the bulk solvent. This observation indicates that substrate-binding is likely to be accompanied by movement of the flexible β5-α3 loop that acts as a gate (**Figure 4.3**).

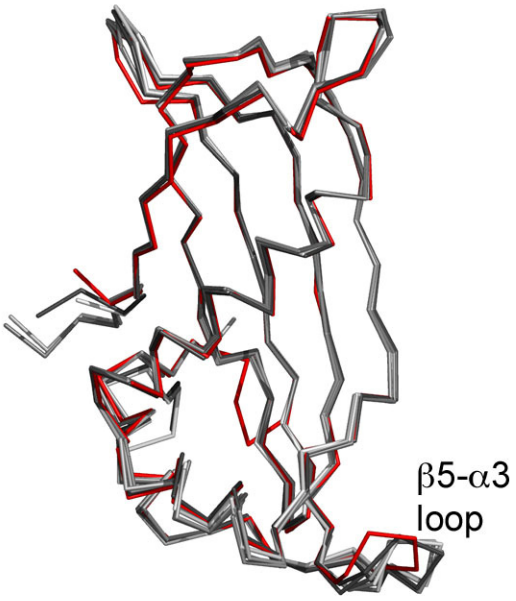


Figure 4.3 Superposition of six subunits of CalE7 in an asymmetric unit. Subunit E which is most diverse is shown in red while the other molecules are shown in shades of gray. Note that the displacement of the β5-α3 loop in subunits other than E blocks the entrance of the substrate-binding channel.

No bound substrate was observed in the substrate-binding pocket. However, at the edge of the dimer interface, close to the entrance of the substrate-binding pocket, a Jeffamine molecule from the crystallization solution is clearly visible in the electron density map (**Figure 4.1d**). The Jeffamine molecule sits between one dimer and the symmetry-related tetramer, stabilized mostly

by hydrophobic contacts. Comparison with 4-HBT of *Pseudomonas* sp. strain CBS-3 (PDB code: 1LO9) reveals that the Jeffamine fragment overlaps with the nucleotide and phosphopantetheinyl moieties of hydroxybenzoyl-CoA (**Figure 4.1c**). The location of the Jeffamine fragment may mimic interactions formed by the exposed region of the phosphopantetheinyl arm tethered to the ACP domain with the CalE7 protein.

4.3.3 Overall Structure of *Apo*-DynE7 and Product-Bound DynE7

Shortly after the elucidation of the crystal structure of CalE7, the structure of the highly homologous DynE7 is also solved at a resolution of 2.1 Å. Both CalE7 and DynE7 belong to the *TEBC* family and perform similar functions in the biosynthesis of enediyne cores. The conformation of DynE7 protein resembles that of CalE7 and other hotdog fold TEs with five-stranded antiparallel β-sheet wrapping around the long α-helix α1 (**Figure 4.4a** and **b**). The overall structure of the DynE7 monomer is highly similar to that of CalE7 protein with superposition of 134 equivalent Cα atoms from the two proteins yielding an R.M.S.D. of 1.46 Å. The channel for ligand-binding is formed at the interface of the dimer with entrance at one of the monomer and the exit at the other (**Figure 4.4c**). Similar to CalE7, the movement of β5-α3 loop leads to the opening of the hydrophilic segment at the beginning of the channel for incoming substrate. The finding confirmed the role of β5-α3 loop as part of the gating mechanism that controls the accessibility of the channel to the ligands. Unlike the truncated channel resulted from steric hindrance posed by side chains of Leu^{133} and Phe^{137} in CalE7, the differential conformation of Leu^{136} and other residues that project from the C-terminal helix α3 in DynE7 lead to the extension of the rear portion thus an increase in the length of the substrate-binding channel.

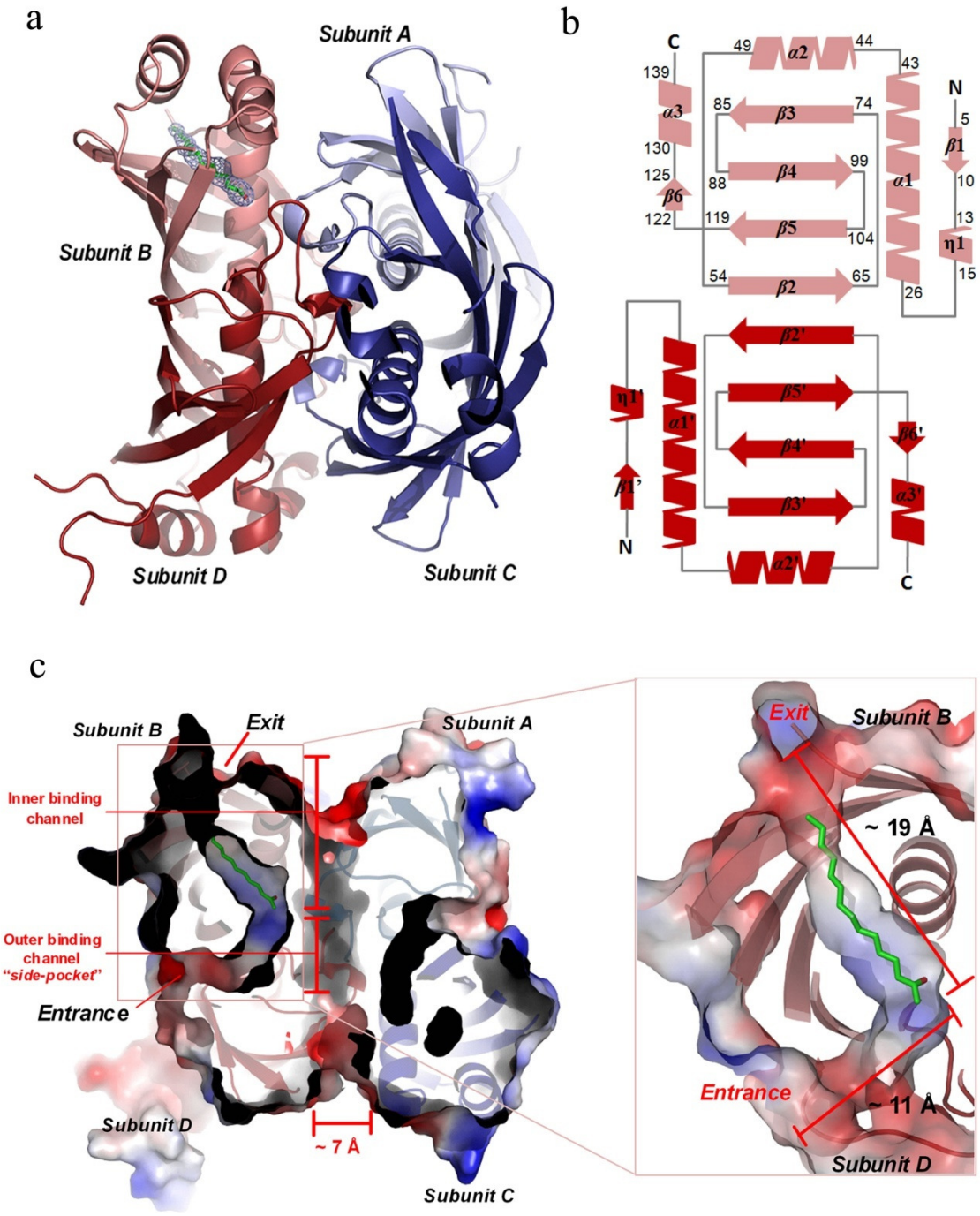


Figure 4.4 Overall structure and topology of DynE7. (a) Schematic representation of the DynE7 tetramer showing the ligand as sticks colored in green at the interface of two subunits. Subunits are colored pink and red for one dimer and dark blue and light blue for the other. (b) Topology diagram of the dimer of dimers. The secondary structure elements are numbered according to the CalE7. (c) The L-shaped ligand binding channel of DynE7. An open conformation is observed at the entrance and the exit of this channel when it contains a ligand. The carbonyl-conjugated polyene **3** is colored in green and is observed in a ~19 Å inner binding channel of subunit B of ligand-bound DynE7.

Each of the substrate channels can be divided into 2 portions: the outer hydrophilic segment near the entrance constitute the phosphopantetheine binding site while the hydrophobic inner channel shielded from the bulk of the solvent is ready to accept the polyene portion of the substrate. The total length of the prolonged channel is estimated to be around 30 Å, which could accommodate either of the two ACP-tethered β-keto and β-hydroxyl thioesters.

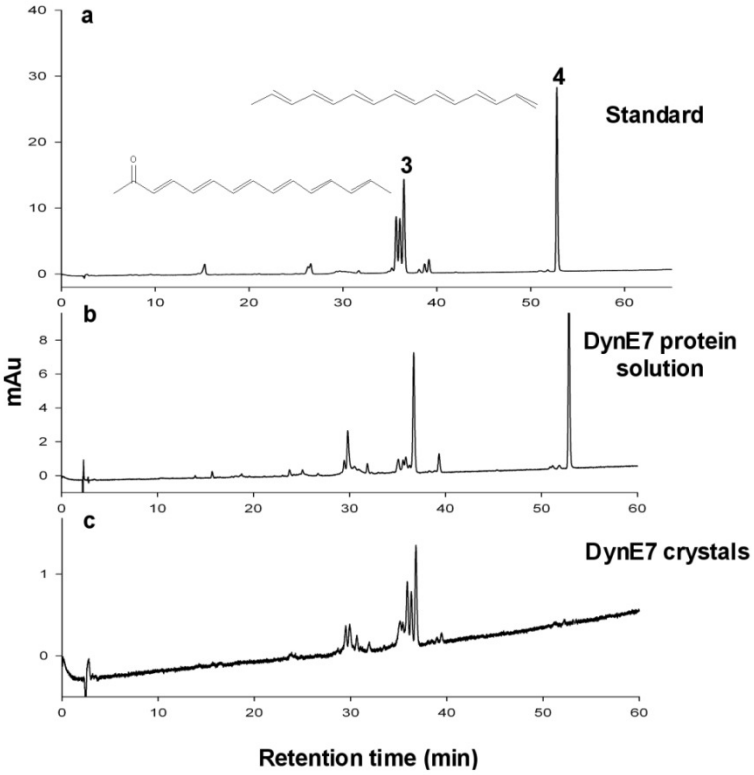


Figure 4.5 HPLC analysis of products generated from DynE8 and DynE7 co-expression recorded at 410 nm. (a) The standards of **2** and **3**. (b) Extract from ligand-bound DynE7 protein solution. Both product **2** and **3** were observed. (c) Extract from dissolved DynE7 crystal—majority of the product found in the crystal is the carbonyl-conjugated polyene **3**.

To our delight, the crystal structure of DynE7 co-expressed with DynE8 in the *E. coli* system was solved in a ligand-bound state. Though careful HPLC analysis of solution containing dissolved DynE7 crystal, the presence of carbonyl-conjugated polyene **3** in the crystal was

confirmed (**Figure 4.5**). However, by docking **3** in the same orientation into the other channels for the rest of the monomers in the tetrameric DynE7, the crystallographic refinements indicated that the presence of ligand in the other monomers were thermodynamically unfavorable. Hence, it is evident that only one carbonyl-conjugated polyene **3** is bound in the channel per DynE7 homotetramer (**Figure 4.4**).

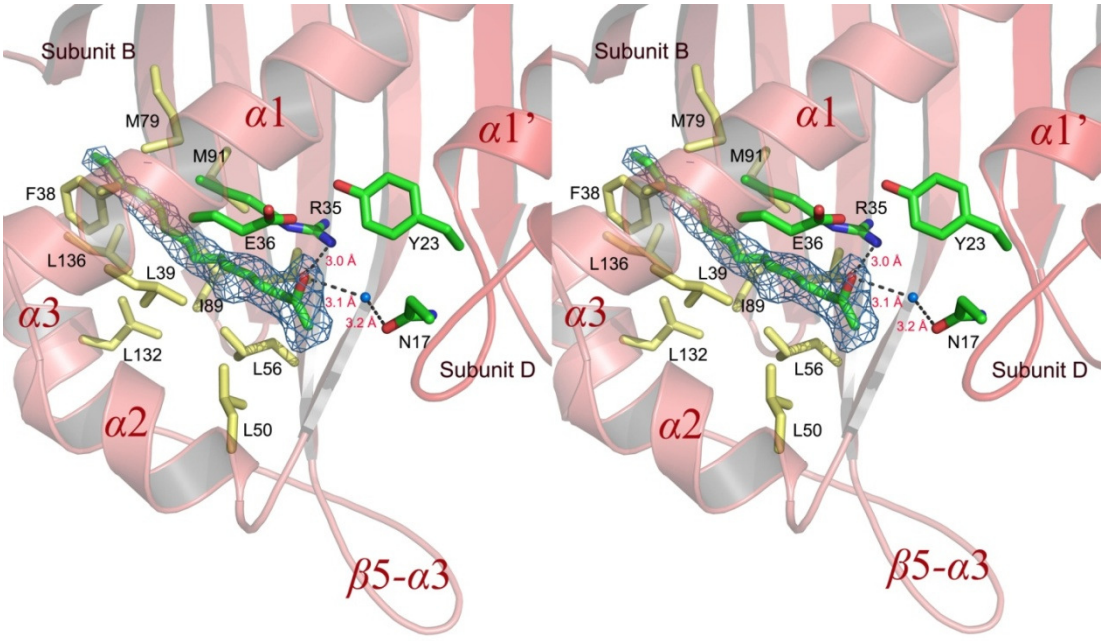


Figure 4.6 A stereo ribbon diagram of the interaction of ligand-protein in the binding channel of DynE7. The carbonyl group of polyene **3** is hydrogen-bonded by Arg^{35} (Arg^{37} in CalE7) from helix α1 and water molecule (blue ball) which hydrogen-bonded by Asn^{17} (Asn^{19} in CalE7) from the adjacent subunit D. The hydrophobic poly-carbon tail of **3** is stabilized by the hydrophobic residues along the channel. Hydrophobic residues are colored in yellow and the conserved polar residues are colored in cyan. Simulated annealing Fourier map with coefficient $|F_{obs}| - |F_{calc}|$ and phases calculated from the protein model with atoms from the carbonyl-conjugated polyene fragment omitted and contoured at a level of 3.0σ.

From the structure of the complex, it was illustrated that multiple interactions between ligand **3** and DynE7 were established. Sequestration of the ligand in the channel was achieved through extensive hydrophobic interactions with the side chains of the residues lining the catalytic cavity. Apart from interactions with hydrophobic residues, the carbonyl group of **3** is

making direct contact with the side chain of Arg^{35} (Arg^{37} in CalE7) via hydrogen bond. Besides, the side chain of Asn^{17} (Asn^{19} in CalE7) also fixes the carbonyl group of **3** in position by a water-mediated hydrogen bond (**Figure 4.6**).

4.3.4 Probing the Catalytic Mechanism by Site-Directed Mutagenesis on CalE7

Given the absence of an acidic residue Glu or Asp in the immediate lining of the binding pocket, CalE7 must catalyze the hydrolysis of the thioester bond using a different mechanism from other hotdog fold thioesterases. The binding pocket mainly consists of hydrophobic residues with the exception of five conserved polar residues, namely Asn^{19}, Asn^{23}, Tyr^{29}, Arg^{37} and Thr^{60}. Five additional polar residues (Glu^{17}, Cys^{36}, Glu^{38}, Glu^{70} and Asp^{75}) in the vicinity of the substrate-binding channel are also well conserved among CalE7 homologs (**Figure 4.7**).

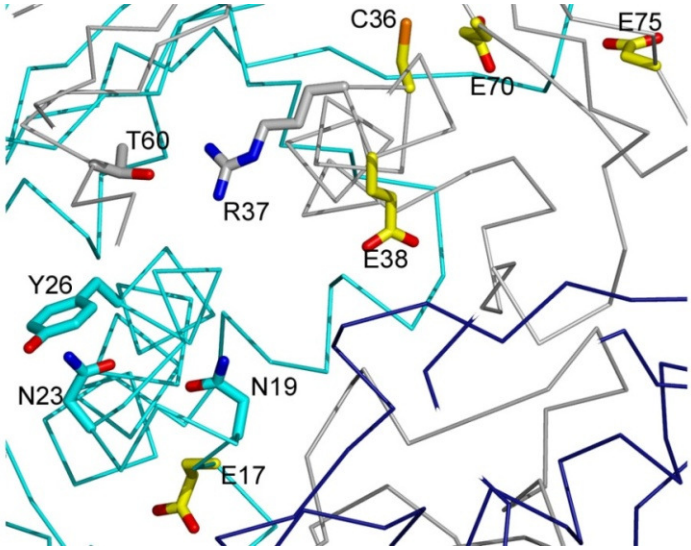


Figure 4.7 The five conserved residues outside the binding pocket. Location of the five conserved residues (shown in yellow) outside of the substrate-binding pocket that were examined by site-directed mutagenesis and kinetic study.

Site-directed mutagenesis and enzymatic assay were carried out to examine the roles of these ten residues in catalysis. Incubation of CalE8 and CalE7 mutants generated various amount of product relative to the wild type CalE7 under different experimental conditions (**Figure 4.8**,

Figure 4.9). For the five residues located in the pocket, the mutation N23A had a negligible effect on product yield at pH 8.2, but increased the yield for both **2** and **3** at pH 7.0. The greatest effect was observed for the mutation R37Q or R37K, which completely abolished the enzymatic activity without any product formation. The mutant T60A favoured the production of **2** rather than **3**, as it decreased the yield for **3** drastically but had negligible effect on the production of **2**. Comparably, mutant Y29F only reduced the production of **3** at pH 8.2 slightly, with minimal effect on the both products at pH 7.0. On the other hand, the mutation N19A seems to facilitate the formation of both **2** and **3** with an increase in the relative yield of more than 300% compared to that of CalE7 wildtype. The enhancement on product formation by N19A was extremely prominent for **3** at pH 7.0 whereby the yield was augmented by almost 17-fold.

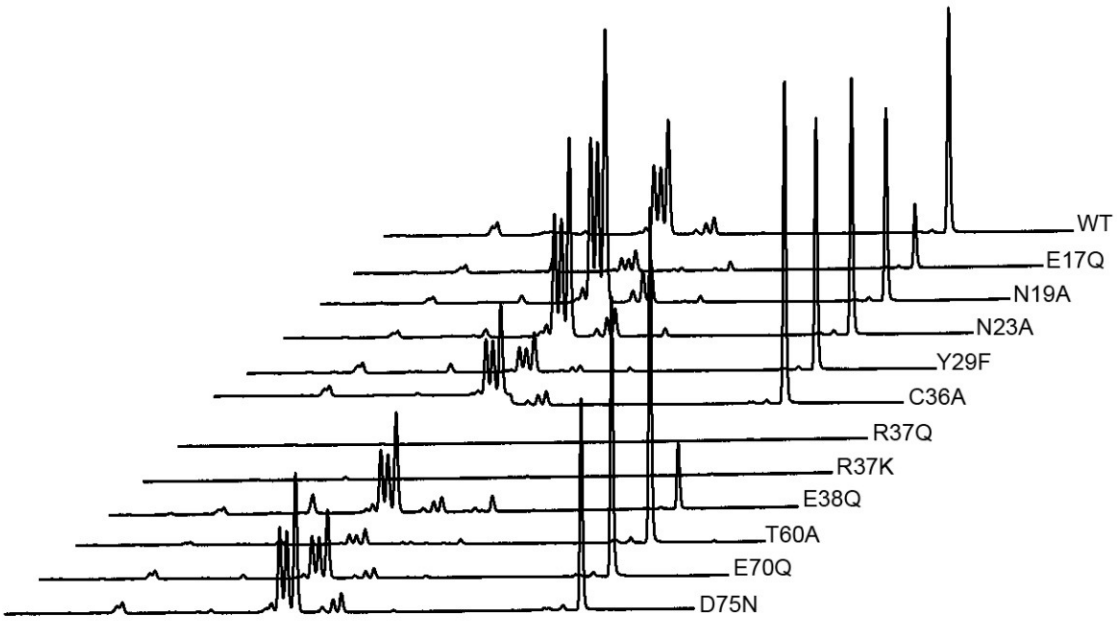


Figure 4.8 HPLC chromatograms of the enzymatic assays carried out with incubation of CalE8 and different CalE7 mutants at pH 8.2.

For the five conserved residues located outside the binding pocket, the mutation E17Q caused a notable decrease in product formation (80%), whereas mutations C36A, E38Q, E70Q

and D75N only caused moderate decreases in product yield. The effect of the E17Q mutation was surprising given that the side chain of the distal Glu^{17} is distant from the substrate-binding pocket (**Figure 4.7**). The E17Q mutant was found to be a dimer in solution by size exclusion chromatography, while all the other mutants remain as tetramer. From the structure, it can be seen that Glu^{17} is located at the tetrameric interface of CalE7 and surrounded by positively charged residues (**Figure 4.10**). Hence, the effect of the E17Q mutation is most likely due to the structural perturbation at the tetramer interface. Generally, the basal level of product formation at pH 7.0 is considerably lower compared to that of pH 8.2, which accounted for the more drastic changes in the relative product yield at neutral pH for most of the CalE7 mutants.

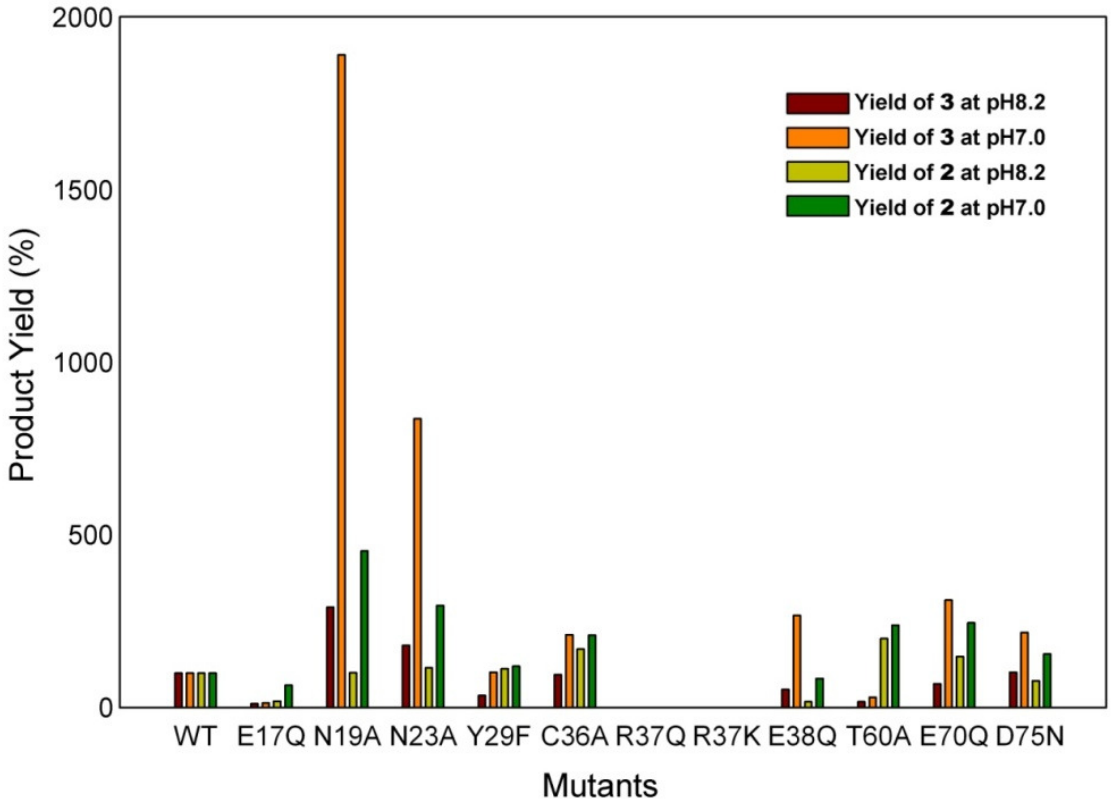


Figure 4.9 Enzymatic activities of the wild type CalE7 and the mutants. Comparison of the relative product yield for CalE7 and its mutants at different pH is shown. Reaction conditions are described in the experimental section and the relative yield was calculated based on the total peak area of the product **2** and **3** obtained from HPLC chromatogram.

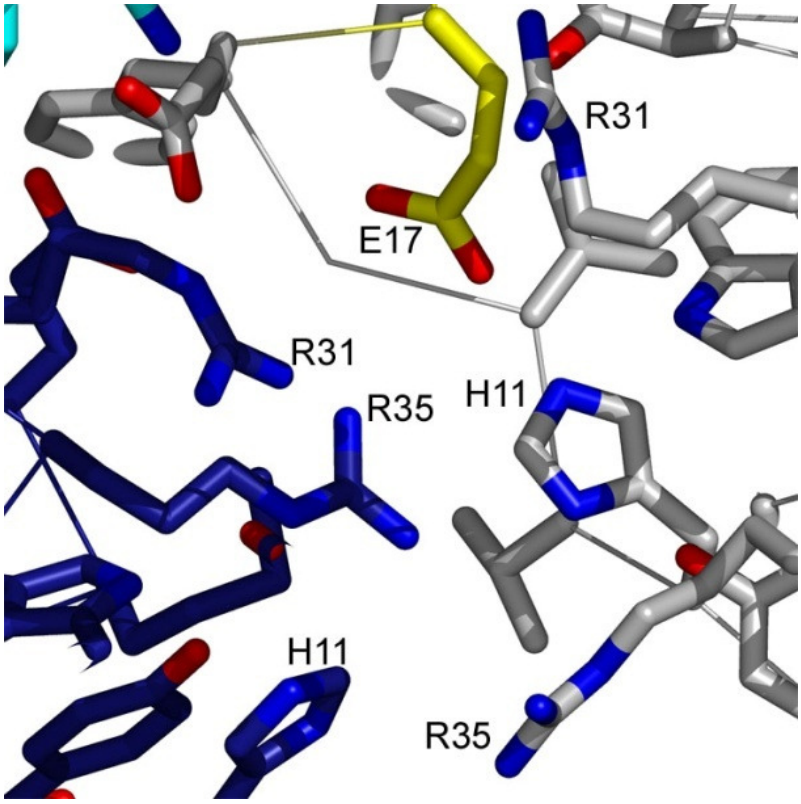


Figure 4.10 Surrounding residues of Glu^{17} at the tetramer interface. Glu^{17} is shown in yellow. Residues from the same subunit as Glu^{17} are shown in grey while the residues from the adjacent dimer are shown in dark blue.

The activities of wild type and the five mutants lining the catalytic cavity were further characterized by steady-state kinetic measurement. The effect of the mutation on enzymatic rate for the five residues in the substrate pocket was examined by following product formation using absorption spectroscopy. The effect of the mutation on Vmax for the mutants generally correlates well with the effect on product yield (**Figure 4.11**). Not surprisingly, the greatest effect was observed for mutation R37Q or R37K, which displays negligible activity even at high enzyme concentration. The T60A mutant exhibited a ~2- fold reduction in V_{max} relative to the wild type enzyme which is probably due to the suppressed formation of **3**; whereas the N19A mutant showed a 1.2-fold greater V_{max} than that of the wildtype CalE7. The mutations Y29F and N23A

caused similar decreases (~1.4-fold) in catalytic rate. Together, the data suggested that Arg^{37} is the only essential residues for catalysis.

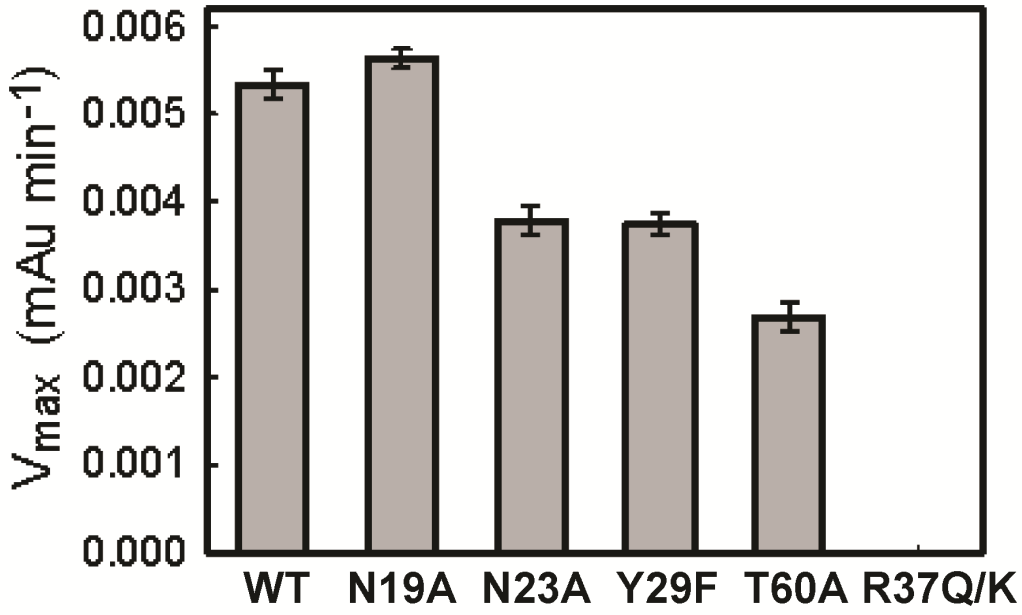


Figure 4.11 Comparison of Vmax for CalE7 and its mutants. The reactions were followed by monitoring the formation of the products using a UV-Vis spectrophotometer. The catalytic activity for the R37Q and R37K mutants was too low to be measured.

4.3.5 Mutagenesis Studies on Arg^{35} and Glu^{36} in DynE7

To further validate the mutagenesis results from CalE8/CalE7 system, the effect of mutations on Arg^{35} and Glu^{36} in DynE7 was also carefully studied. Enzymatic assay was carried out to monitor the rate of consumption of NADPH in reactions involving DynE8 and the two DynE7 mutants. The results obtained were congruent with that of CalE7 mutations. The mutant R35A abolished the catalytic activity completely, leaving NADPH level largely unchanged at the end of the reaction. On the other hand, the E36A mutant displayed a considerable level of NADPH consumption, which implied that the catalytic capability of the mutant was largely retained (**Figure 4.12**).

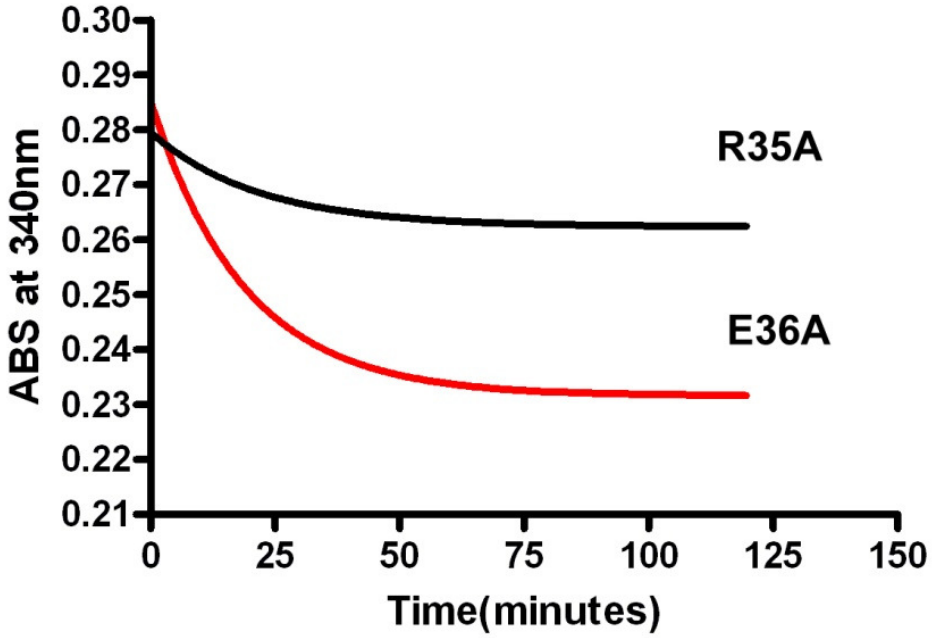


Figure 4.12 Enzymatic assay of DynE8 with DynE7 mutants. R35A mutant shows negligible activity in terms of NADPH consumption, lending further support to its critical role in thioester cleavage. The E36A mutant displays a much higher activity since the rate of NADPH consumption is much faster.

4.4 Discussion

Thioesterases associated with fatty acid synthase (FAS), PKS and non-ribosomal peptide synthase (NRPS) generally belong to the so-called type I and II thioesterases. On the other hand, hotdog fold thioesterases are mainly known to utilize acyl-CoA as substrates, with only a few known examples that act on peptidyl carrier protein (PCP) or ACP-tethered acyl substrates [101, 175, 176], including the EntH protein (or YbdB) involved in the biosynthesis of the nonribosomal peptide-derived enterobactin [175, 177, 178]. Hence, CalE7 and DynE7 appear to represent unconventional examples where hotdog fold thioesterases are recruited for polyketide synthesis.

4.4.1 Substrate Binding in CalE7 and DynE7

The overall structure of CalE7 and DynE7 resemble other hotdog fold thioesterases with the εγ oligomeric arrangement. However, with a sharply kinked substrate-binding channel and an unusually large side pocket, the substrate-binding pocket of CalE7 in the closed conformation differs significantly from other thioesterases despite the common location of the pocket at the subunit interface. According to the crystal structure of CalE7, the channel is a few Angstroms short for the full-length substrate, presumable due to the shortening of the channel by a conformational change involving helix α3. With the *apo*-structure of CalE7 and the structure of DynE7-ligand complex, we propose that the binding of the ACP-tethered substrate requires conformational changes in the β5-α3 loop as well as the α3 helix. The movement of the "gating" β5-α3 loop would open up the entrance, while the movement of the short α3 helix that contains Leu^{133} and Phe^{137} in CalE7 (Leu^{132} and Leu^{136} in DynE7) would extend the channel longitudinally.

The substrate for CalE7/DynE7 is covalently attached to the phosphopantetheinyl group on the ACP domain of CalE8/DynE8. Efficient cleavage of the product may require the recognition of CalE7/DynE7 by CalE8/DynE8. However, no significant complex formation between CalE8/DynE8 and CalE7/DynE7 was detected using size-exclusion chromatography. Besides, our effort in detecting interactions between CalE7 and *me*ACP (isolated ACP domain of CalE8) via NMR titration ended up in vain either. The experimental results indicate that the two proteins may interact with each other with moderate affinity or the interaction is too transient to be detected. In addition, the contact between PKS and TE is unlikely to be mediated by protein-protein interaction alone. Presumably a portion of the phosphopantetheinyl group would be bound inside CalE7/DynE7 while the remainder of the phosphopantetheinyl group would protrude out from the tunnel at the dimer interface of CalE7/DynE7. The ACP domain of CalE8/DynE8 is thus expected to be located at the proximity of the entrance of the active site. Since protein-protein interaction between CalE7 and *holo-me*ACP is absent, the ACP-tethered acyl chain might play an important part in the recognition process. Nonetheless, after recognizing the tethered acyl chain, it is still possible for the positively charged surface consisting of Lys^{62} and Arg^{113} on CalE7 to interact with the acidic patch (loop-1) on the ACP domain in order to enhance the binding and facilitate the cleavage of the product. After insertion of the ACP-tethered thioester intermediates into the catalytic chamber, the thioesterase probably orientates the acyl substrates for hydrolysis largely through interactions between the phosphopantetheinyl moiety and the entrance segment of the binding pocket. This is consistent with the finding that CalE7 was able to generate both conjugated polyene **2** and carbonyl-conjugated polyene **3** indiscriminatively. Given the high structural similarity between the two

thioester intermediates**,** the binding conformation of the phosphopantetheinyl moiety to the hydrophilic segment should be almost identical during the hydrolytic release of both **2** and **3**.

4.4.2 Catalytic Mechanism of CalE7 and its Homologs

A common feature of the hotdog fold thioesterases is the presence of an essential acidic residue Glu/Asp on either side of the binding pocket [101]. Although initial sequence comparison of CalE7 with other thioesterases suggested the conserved Glu^{17} or Glu^{75} could be the catalytic residue, the crystal structure revealed that the two residues are too distant from the substrate-binding channel to play a direct role in catalysis. The absence of an acidic residue as nucleophile or general base was recently reported for the thioesterase FcoT from *Mycobacterium tuberculosis*, an acyl-CoA thioesterase that prefers long-chain fatty acyl-CoA substrates [162]. It was proposed that FcoT activates the thioester bond and stabilizes the oxyanion intermediate via the side chains of Tyr^{33}, Asn^{74} and Tyr^{66}, and that hydrolysis occurs through a direct attack by a hydroxide ion. The superposition of the structures of FcoT and CalE7/DynE7 showed that these three residues are not conserved in CalE7/DynE7, indicating that CalE7/DynE7 utilizes a catalytic mechanism different from FcoT.

With information on catalytic mechanism of conventional thioesterases, the search for the catalytically critical residues began with acidic amino acids located in the vicinity of the substrate binding channel. However, the conserved residue Glu^{38} in CalE7 (Glu^{36} in DynE7) on central helix α1 was not indispensible for the catalysis given the mutagenesis data obtained. On the other hand, another charged residue on the α1— Arg^{37} in CalE7 (Arg^{35} in DynE7) is also strictly conserved throughout the members from *TEBC* family. Mutation of the conserved Arg into Gln or Lys completely abolished the activity of CalE7 and DynE7. Hence it is very likely

that instead of the conventional Glu/Asp, Arg^{37} (Arg^{35} in DynE7) plays a vital role in the cleavage of the thioester bond.

Based on the structural and biochemical data, we propose a catalytic mechanism for CalE7/DynE7 for the thioester hydrolysis. The transient interaction between CalE7/DynE7 and product-tethered CalE8/DynE8 induces a major conformational change in the overall structure of the thioesterase. The shift in the position of β5-α3 loop opened up the substrate binding channel for the incoming acyl chain. Meanwhile, the conformational change in helix α3 results in the extension of the inner channel following the repositioning of Leu^{133} and Phe^{137} (Leu^{132} and Leu^{136} in DynE7). Immediately after insertion into the catalytic chamber, the substrate is anchored within the channel by extensive hydrophobic interactions between the polyene segment and the side chains of the hydrophobic residues lining the cavity. This binding configuration places the carbonyl oxygen of the thioester group within hydrogen-bond distance of Arg^{37} (Arg^{35} in DynE7). As for Asn^{19} (Asn^{17} in DynE7), a water molecule is utilized to mediate the interaction between the carbonyl oxygen of the thioester and the amide group of Asn. In accordance with the ligand-bound structure of DynE7, the keto group of the bound ligand **3** in the crystal indeed interacts with Arg^{35} and Asn^{17}(Arg^{37} and Asn^{19} in CalE7) through hydrogen bonds and water-mediated hydrogen bonds respectively in place of the carbonyl group of the thioester (**Figure 4.6**).

The nucleophilic attack of the thioester carbonyl by a water or hydroxide anion is facilitated with Arg^{37} (Arg^{35} in DynE7) acting as an oxyanion hole to stabilize the transition state. This mechanism is consistent with the presence of several water molecules within the active site pocket (**Figure 4.2a**). Although we cannot totally rule out that Arg^{37} (Arg^{35} in DynE7) functions as a general base for water activation, this alternative mechanism is rather unlikely as it would

require a reduction of the p*Ka* of the Arg by ~4-5 p*Ka* units. Subsequently, the collapse of the transition state would break the thioester bond to generate the 16-carbon β-keto/β-hydroxyl carboxylic intermediate.

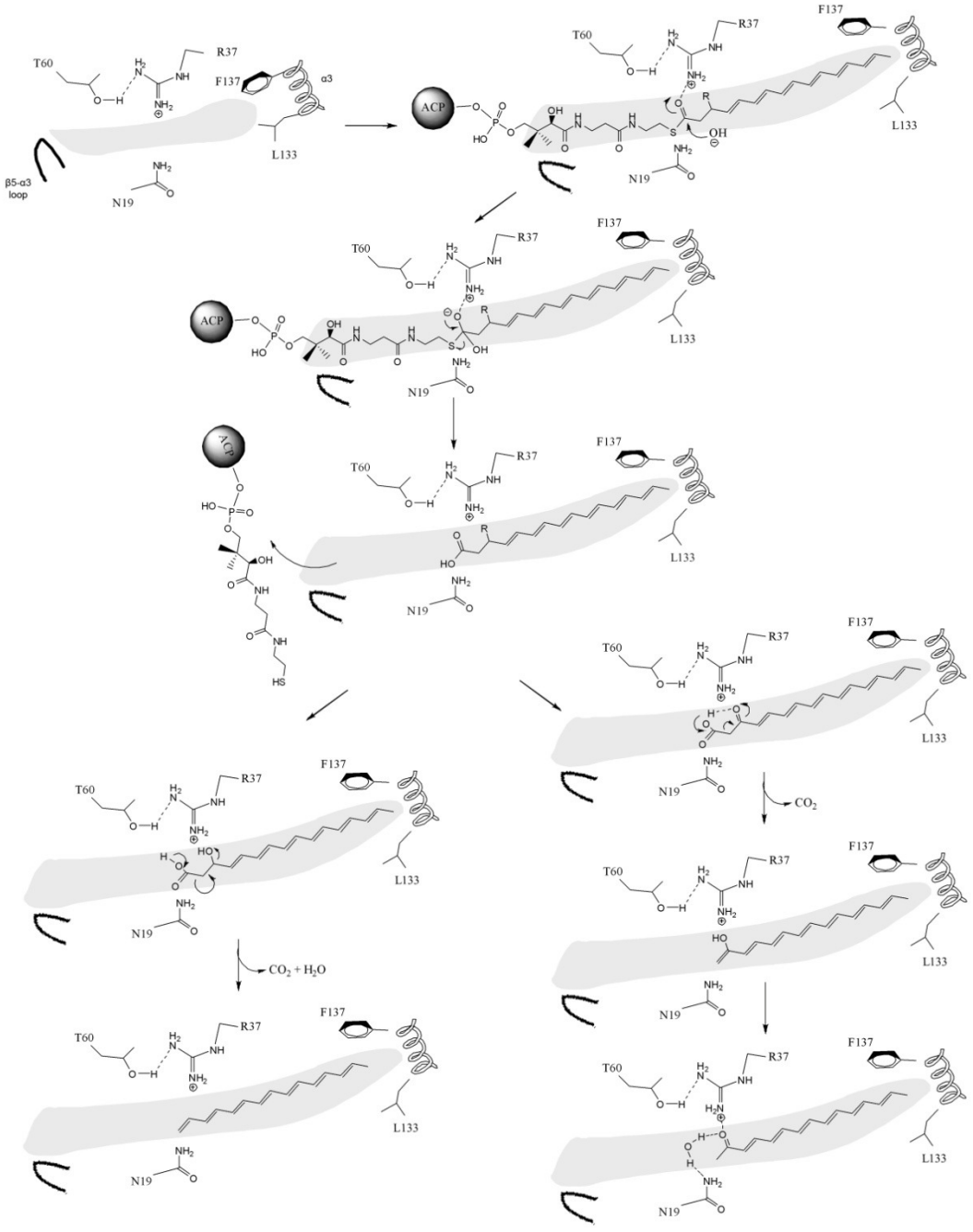


Figure 4.13 Proposed catalytic mechanism based on CalE7/DynE7 structure. The substrate-binding channel is represented by the grey shade. In the hydrolysis step, the essential residue

Arg^{37} (Arg^{35} in DynE7) functions as an oxyanion hole. The step following hydrolysis represents the breakage of the thioester bond. The decarboxylation occurs with an intramolecular proton transfer within the six-centered transition state, and with the tautomerization to form product **3**. As for **2**, an additional step of dehydration occurs to yield the conjugated polyene **2**. Whether the decarboxylation and dehydration steps are assisted by CalE7 is not clear. The last step shows the repositioning of the final product in the channel in accordance with the data obtained from the DynE7-ligand (**3**) complex crystal structure.

In the case of product **3** formation, a facile decarboxylation of the β-keto carboxylic acid could occur with an intramolecular proton transfer between the carbonyl and carboxylic groups within a six-centered transition state. The role of CalE7/DynE7 in the decarboxylation step of the β-keto carboxylic acid is still not clear. A potential proton acceptor might be present in the catalytic channel to assist the decarboxylation of the β-keto carboxylic acid. Eventually, the carbonyl-conjugated polyene **3** is generated after the immediate tautomerization of the enol intermediate with a water molecule in the side pocket acting as proton donor. The formation of **2** is considerably more complicated compared to that of **3**. The mechanism for decarboxylation *en route* to the formation of **2** might be similar to that of **3**. In addition to the decarboxylation, a dehydration step ensues to eliminate the β-hydroxyl group together with a hydrogen atom from the α-carbon of the thioester to form the final double bond in the polyene **2** (**Figure 4.13**). The presence of a proton acceptor might be imperative for either the decarboxylation or the dehydration step. Nonetheless, the mechanism for dehydration is still beyond our comprehension and whether the decarboxylation and dehydration steps are taking place simultaneously or sequentially is not well understood either.

The proposed catalytic mechanism is consistent with several major experimental observations. Firstly, Arg^{37} (Arg^{35} in DynE7) is the only essential catalytic residue as established by mutagenesis study. This is understandable considering the critical role of Arg^{37} (Arg^{35} in DynE7) as oxyanion hole or general base in hydrolysis. Secondly, the intriguing observation that

the mutant N19A in CalE7 exhibited a higher product yield and greater V_{max} can be rationalized by considering the positioning of the ACP-tethered thioester intermediate in the substrate-binding channel prior to hydrolysis. The water-mediated interaction between Asn^{19} in CalE7 and the β-keto/β-hydroxyl group of the substrate upon entry impedes the further insertion of the acyl chain into the catalytic cavity. Hence, the orientation of the carbonyl group of the thioester to be within the catalytic range of Arg^{37} requires the breakage of the water-mediated interaction between β-keto/β-hydroxyl group and Asn^{19}, followed by further insertion of the acyl chain into the cavity. Consequently, with the elimination of the Asn^{19}, the movement of the substrate into the catalytic channel becomes unhindered and the positioning of the thioester at the catalytic site is greatly facilitated.

With respect to the mutagenesis data, the role of Thr^{60} in CalE7 the catalysis presents yet another mystery to be solved. The T60A mutation resulted in the preferential production of **2**, indirectly ruling out the possibility that the hydroxyl side chain functions as the nucleophilic group attacking the thioester. Instead, Thr^{60} is most likely to play a secondary or structural role in catalysis, by positioning the side chain of Arg^{37} through hydrogen-bonding. The mutation of Thr^{60} to alanine probably alters the conformation of the side chain of Arg^{37}, leading to the positioning away of the critical amine groups from the catalytic channel. However, due to the inherent flexibility of the β-hydroxyl intermediate, bond rotation at C2 and C3 positions is still able to place the carbonyl group of the thioester within the vicinity of Arg^{37}. On the other hand, the abstraction of the highly acidic hydrogen on α-carbon may result in the tautomerization of the β-keto intermediate into the corresponding enol form. Therefore, the rotational flexibility around C2 and C3 position is entirely obliterated. The increased distance between Arg^{37} and the

carbonyl group of the thioester intermediate due to keto-enol tautomerism leads to the drastic reduction in the formation of **3** (**Figure 4.14**).

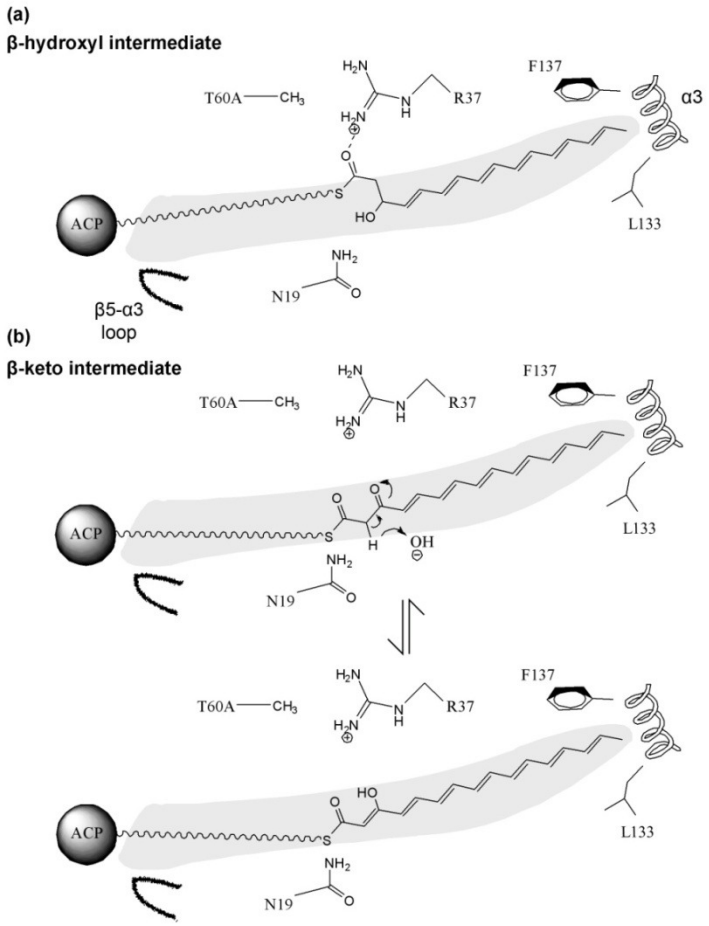


Figure 4.14 Proposed mechanism of mutation T60A on thioester hydrolysis. (a) The conformational changes in β-hydroxyl intermediate still allow the hydrolysis of thioester to occur. (b) The keto-enol tautomerism makes the carbonyl oxygen of the thioester in β-keto intermediate difficult to reach by Arg^{37}.

The exact function of the essential residue Arg^{37} (Arg^{35} in DynE7) still remains to be fully established in the future. Current structural and biochemical data do not distinguish whether the Arg acts as TS stabilizer or general base even though the latter is less likely. We propose that hydrolysis may take place through a direct attack of hydroxide anion, similar to the mechanism

proposed for FcoT and the hydrolytic antibody D2.3 [162, 179]. Since the rate of such nucleophilic reaction depends on the concentration of the hydroxide anion, this is to some extent in line with the observation that high pH favored product formation. Although it is rare for Arg to act as general base in enzymes due to its high pK_a, we could not totally rule out the possibility that the water molecule could be activated by the neutral form of Arg^{37} (Arg^{35} in DynE7) if the pK_a of the Arg is drastically lowered by several pH units in CalE7/DynE7.

In conclusion, the proposed function of CalE7/DynE7 in releasing the linear products for downstream processing is largely based on the assumption that the cleaved acyl chain is the mature product of CalE8/DynE8 and thus the biosynthetic intermediate for enediyne. However, the current structural and biochemical data cannot totally rule out the possibility that only aberrant products are produced in the PKS/TE system or the final products are correctly assembled but not processed immediately. In this case, the function of CalE7/DynE7 is to remove these products from the PKS, resembling the editing roles of the type II TEs. These uncertainties will be eventually borne out with the elucidation of the downstream cyclization and oxidation steps.

CHAPTER 5 Conclusion and Future Directions

5.1 Conclusions

The unique structural and chemical properties of the enediyne warheads, together with a myriad of diverse peripheral moieties attached to them, entail a whole array of novel mechanisms involved in the biosynthetic pathways. Since the discovery of the first naturally-occurring enediyne, our understanding on the mode of action and biosynthesis of the enediynes has been advancing at a steady speed. Among all the advancements, the identification and the complete sequencing of the two enediyne gene clusters for calicheamicin γ_1^I and C-1027, revitalized the field and offered numerous research opportunities to scientists around the world.

With a plethora of novel enzymes involved in the enediyne biosynthesis waiting to be characterized, the genes responsible for the production of the warhead, or the enediyne core, are definitely in the limelight. Debates have been going on and speculations have been made constantly regarding the biosynthetic origin of the enediyne warhead. Identification of genes encoding PKSs within the clusters further substantiated the polyketide origin of the enediyne warheads. Apart from that, the highly homologous enediyne PKSs discovered among various enediyne-producing bacteria clearly indicate that the 9-membered and 10-membered enediynes share similar biosynthetic pathways. From a series of *in vitro* and *in vivo* experiments conducted in our and other laboratories, recombinant enediyne PKSs seem to be able to synthesize a wide range of products of different chain lengths under various experimental conditions. Among all the products formed, the highly conjugated polyene **2** was exclusively produced under certain experimental conditions. Due to the high structural similarity between the conjugated polyene **2** and the carbonyl-conjugated polyene **3**, even with the observation of **3** in the DynE7 crystals

[151], the intermediate *en route* to the biosynthesis of enediyne core remains to be established. The possibility that both **2** and **3** are aberrant products could not be completely ruled out. With the X-ray data obtained for CalE7/DynE7, the role of the sole thioesterase encoded in the gene cluster has been tentatively assigned as the enzyme to release the matured product from PKSs, even though the likelihood of CalE7/DynE7/SgcE10 functioning as a type II thioesterase still exists. By removing the aberrant products stuck on the PKSs, the thioesterases may simply function to regenerate the PKSs for the synthesis of the authentic biosynthetic intermediates. As a result, the physiological product of PKS remains tethered to the ACP domain and waits to be processed by downstream auxiliary proteins in the system.

It has been suspected that putative *trans*-regulating elements may play a role in the biosynthetic pathway. By intervening the critical steps of the polyketide synthesis at the PKS stage, the regulatory factors may function to direct the PKSs to produce the authentic biosynthetic intermediate. From the bioinformatics studies on the gene clusters, it is still not clear which gene may play the role of the regulator for the enediyne biosynthetic pathway. Besides, there are still a few more intriguing questions related to enediyne biosynthesis waiting to be addressed. For instance, the role of CalU15 is speculated to be an acetylenase responsible for the formation of the *–yne* groups. Nonetheless, all the experiments designed to probe the function of CalU15 yielded little results. We suspect that CalU15 lacks the activity of acetylenase probably due to the absence of the cognate electron donor in the assay system. In spite of our continued attempts in reconstitution and crystallization, CalU15 remains reluctant to unveil itself as an imperative component for enediyne synthesis. Whether cyclization precedes the formation of acetylenic bond is yet another piece of puzzle to be solved. Furthermore, the crucial cyclase that dictates the different folding pattern of enediyne precursors presents another missing link in the

pathway. Up to date, only two of the five conserved genes *PKSE* (encoding PKS) and *TEBC* (encoding TE) are characterized, while the elusive *UNBL* (U15) is still under intensive study in our laboratory. The other two genes, namely *UNBU* and *UNBV*, stand as insurmountable obstacles even at the stage of protein expression, not to mention any structural and biochemical characterizations.

The novelty in the structure of proteins encoded by the minimal gene cassette is also manifested in the individual domains of the enediyne PKSs (e.g. CalE8). Functioning as a shuttling domain within the PKS, the ACP domain is the smallest but most versatile domain in the megasynthase. The indispensable ACP domain is crucial as the platform on which the elongating acyl chain is anchored. With a three-helix bundle structure, the ACP domain of CalE8 features tight hydrophobic packing that seems to prevent the binding of the growing acyl chain. The nascent acyl chain is precluded from the central cavity of *me*ACP and can be easily accessed by the catalytic domains of CalE8 for processing at different elongation stages. The distinct hydrophobic patch on the exterior of the *me*ACP instead raises the possibility that *me*ACP might be able to stabilize the labile polyene through extensive hydrophobic interaction with protein surface.

With the determination of the structures of CalE7 and the homolog DynE7, the function and catalytic mechanism of the novel type of thioesterase came to light. Unlike other hotdog fold thioesterases, the absence of canonical Glu/Asp for hydrolysis of thioester bond makes CalE7 and DynE7 unique in thioester hydrolytic catalysis. The structural change upon the binding of the PKS product revealed a remarkable induced-fit mechanism for substrate binding. Despite the production of **2** implies that the proteins encoded by *TEBC* family may function as type I

thioesterase, the precise role of the TEs as type I or type II TEs in enediyne biosynthesis remains to be fully established.

5.2 Future Directions

The precise role of the type I iterative PKSs in enediyne biosynthesis remains to be fully defined. Besides, there is much more to be learned about the structural and mechanistic aspects of the fascinating megasynthase. One of the intriguing aspects of enediyne PKSs is the intrinsic decarboxylation mechanism associated with the generation of the acetate starter unit. From all the experimental data collected, the presence of acetyl-CoA as the starter molecule in the reaction mixture is not required. Hence, it is reasonable that PKS is actually capable of generating acetyl group *de novo* from the malonyl-CoA substrate. The decarboxylation reaction is most likely carried out by the KS domain of PKS. In a parallel type III PKS system, the chain length factor (CLF) within the heterodimeric KS-CLF didomain in fact initiates the biosynthetic process by generating acetyl group from malonyl-CoA [180]. So far decarboxylation of malonyl-CoA to generate acetyl starter has been shown in a few aromatic and modular PKS systems. On the contrary, in mammalian FAS, the same process takes place at a much slower rate compared to the coupled decarboxylation-condensation in the elongation process. Hence, the intrinsic catalytic activity of malonyl-CoA decarboxylation by enediyne PKSs presents us with an interesting prospect in the functional and structural study of isolated KS domain. In theory, tracking the activity of the decarboxylase can be assayed by radioactive-isotopic labeling experiment and steady-state kinetic studies with established protocols.

The mystery shrouding the structure and function of the individual domains of PKS is but only a small part of the entire puzzle. Unraveling the secrets of PKS presumably entails the complete structural and functional study on the megasynthase itself. To date no crystal of any full length PKS has been determined. The total elucidation of the structure for iterative enediyne PKS would represent a major breakthrough in the field of PKS research. The crystallization of

full length CalE8, DynE8 and SgcE is already under way. The expression and purification processes have been optimized and preliminary crystallization screening on the colorless co-expressed CalE8 has yielded small needle-shaped microcrystals. In the future, it is possible to generate a Ser^{971} mutant which will inevitably inactivate the ACP domain of PKS thus maximizes the homogeneity of the PKS protein. On the other hand, co-crystallization of PKS can also be tried out with the cognate TE. Given that the R37Q/K mutants of CalE7 have been proven to be completely inactive in thioester cleavage, co-crystallization of R37Q/K with product-tethered CalE8 can be carried out in the screening process since the acyl chain will be bound inside the substrate-pocket of the CalE7 mutants without being cleaved off. As a result, it is feasible to trap both CalE8 and CalE7-R37Q/K in the crystals because of the constraint imposed by protein-ligand interactions between the two. Moreover, SAX (small angle X-ray diffraction) has been carried out as well to study the overall topology of CalE8. With the existing homologous structures of some of the PKS domains, the structural information of individual domain can be pieced together to generate a structural model for the entire PKS.

The resemblance borne by iterative type I PKSs to FAS suggests that they may have many structural and functional features in common. Large-scale domain movement during the course of chain elongation is implicated in some of the recent structural studies on mammalian FAS [181]. However, the hypothetical internal movements of the megasynthases during chain elongation steps are difficult to detect, if not impossible with crystallization studies. Given the advancement in the H/D-exchange-coupled mass spectrometry, it has become possible to probe the conformational changes and protein dynamics by MS studies on the digested peptide fragments. Comparison of H/D exchange pattern for different states or conformations of the PKS can yield invaluable information on the dynamic aspects of PKS structure.

Biochemical and structural studies on CalU15 will continue to be the focus of our research. Screening of reducing agents that may act as the electron donor will be carried out. At the same time, crystallization screening on CalU15 is currently being performed by our collaborators in an effort to deduce the function of the protein from the structural perspective.

In summary, the chemistry involved in the biosynthesis of enediynes such as chalicheamicin is extremely complex. For instance, the formation of the triple bonds in the enediyne core entails an extraordinarily energy-demanding reaction. By far all the attempts in identifying the mysterious acetylenase ended up with no avail. As a result, exploration into the modification process of the enediynes is still at its infant stage, whereby only the processing of peripheral moieties is being gradually uncovered. Thus, the elucidation of the biosynthetic process of enediyne production will help us undstand and appreciate more on the wonder of the nature in building complex biological molecules. Furthermore, deciphering the biosynthesis of enediynes will shed light on the evolutionary linkage between 9- and 10-membered enediynes as well. With highly homologous enzymes, microorganisms are capable of producing enediyne rings of different structures. The comparison of the fascinating synthetic mechanisms among different enediyne-producing organisms may allow us to catch a glimpse of how evolution works in terms of genetics as well as environmental influence. Besides, understanding the biosynthetic mechanism of enediynes will have great impact on medicinal chemistry as well because manipulation of the synthetic pathway will allow us to modify the anti-cancer enediyne molecules for the treatment of various cancers.

Although the enediyne biosynthetic mechanism still waits to be fully unraveled, the on-going research in our as well as others' laboratory will eventually reveal the secrets locked in the biosynthetic pathway of enediynes.

References

1. Calestani, C., et al., *Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening.* Development, 2003. **130**(19): p. 4587-4596.

2. Collie, N. and Myers, W.S., *VII.-The formation of orcinol and other condensation products from dehydracetic acid.* Journal of the Chemical Society, Transactions, 1893. **63**: p. 122-128.

3. Birch, A.J., et al., *Studies in relation to biosynthesis .7. 2-hydroxy-6-methylbenzoic acid in Penicillium griseofulvum Dierckx.* Australian Journal of Chemistry, 1955. **8**(4): p. 539-544.

4. Staunton, J. and Weissman, K.J., *Polyketide biosynthesis: a millennium review.* Natural Product Reports, 2001. **18**(4): p. 380-416.

5. Hill, A.M., *The biosynthesis, molecular genetics and enzymology of the polyketide-derived metabolites.* Natural Product Reports, 2006. **23**(2): p. 256-320.

6. Van Lanen, S.G., *Biosynthesis of enediyne antitumor antibiotics.* Current Topics in Medicinal Chemistry, 2008. **8**(6): p. 448-459.

7. Wenzel, S.C. and Muller, R., *Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry.* Natural Product Reports, 2007. **24**(6): p. 1211-1224.

8. Edo, K., et al., *Naphthalenecarboxylic acid from neocarzinostatin (NCS).* Journal of Antibiotics, 1980. **33**(3): p. 347-351.

9. Edo, K., et al., *Unstability of neocarzinostatin-chromophore.* Journal of Antibiotics, 1986. **39**(4): p. 535-540.

10. Lee, M.D., et al., *Calichemicins, a novel family of antitumor antibiotics. 2. Chemistry and structure of calichemicin .gamma.1I.* Journal of the American Chemical Society, 1987. **109**(11): p. 3466-3468.

11. Lee, M.D., et al., *Calichemicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calichemicin .gamma.1I.* Journal of the American Chemical Society, 1987. **109**(11): p. 3464-3466.

12. Golik, J., et al., *Esperamicins, a novel class of potent antitumor antibiotics. 2. Structure of esperamicin X.* Journal of the American Chemical Society, 1987. **109**(11): p. 3461-3462.

13. Golik, J., et al., *Esperamicins, a novel class of potent antitumor antibiotics. 3. Structures of esperamicins A1, A2, and A1b.* Journal of the American Chemical Society, 1987. **109**(11): p. 3462-3464.

14. Oh, D.C., et al., *Cyanosporasides A and B, chloro- and cyano-cyclopenta[a]indene glycosides from the marine actinomycete "Salinispora pacifica".* Organic Letters, 2006. **8**(6): p. 1021-1024.

15. Feling, R.H., et al., *Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus Salinospora.* Angewandte Chemie International Edition in English, 2003. **42**(3): p. 355-357.

16. Buchanan, G.O., et al., *Sporolides A and B: Structurally unprecedented halogenated macrolides from the marine actinomycete Salinispora tropica.* Organic Letters, 2005. **7**(13): p. 2731-2734.

17. Smith, A.L. and Nicolaou, K.C., *The enediyne antibiotics.* Journal of Medicinal Chemistry, 1996. **39**(11): p. 2103-2117.

18. Thorson, J.S., et al., *Understanding and exploiting nature's chemical arsenal: The past, present and future of calicheamicin research.* Current Pharmaceutical Design, 2000. **6**(18): p. 1841-1879.

19. Walker, S., et al., *Sugars as DNA binders: a comment on the calicheamicin oligosaccharide.* Journal of the American Chemical Society, 1990. **112**(17): p. 6428-6429.

20. Drak, J., et al., *The carbohydrate domain of calicheamicin gamma I1 determines its sequence specificity for DNA cleavage.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(17): p. 7464-7468.

21. Nicolaou, K.C. and Dai, W.M., *Chemistry and biology of the enediyne anticancer antibiotics.* Angewandte Chemie International Edition in English, 1991. **30**(11): p. 1387-1416.

22. Koga, N. and Morokuma, K., *Comparison of biradical formation between enediyne and enyne-allene. Ab initio CASSCF and MRSDCI study.* Journal of the American Chemical Society, 1991. **113**(6): p. 1907-1911.

23. Jones, R.R. and Bergman, R.G., *p-Benzyne. Generation as an intermediate in a thermal isomerization reaction and trapping evidence for the 1,4-benzenediyl structure.* Journal of the American Chemical Society, 1972. **94**(2): p. 660-661.

24. Myers, A.G., *Proposed structure of the neocarzinostatin chromophore-methyl thioglycolate adduct; a mechanism for the nucleophilic activation of neocarzinostatin.* Tetrahedron Letters, 1987. **28**(39): p. 4493-4496.

25. Zein, N., et al., *Selective proteolytic activity of the antitumor agent kedarcidin.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(17): p. 8009-8012.

26. Biggins, J.B., et al., *A continuous assay for DNA cleavage: The application of 'break lights' to enediynes, iron-dependent and nucleases.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(25): p. 13537-13542.

27. Sugiura, Y. and Matsumoto, T., *Some characteristics of DNA strand scission by macromolecular antitumor antibiotic C-1027 containing a novel enediyne chromophore.* Biochemistry, 1993. **32**(21): p. 5548-5553.

28. Zein, N., et al., *Kedarcidin chromophore: An enediyne that cleaves DNA in a sequence-specific manner.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(7): p. 2822-2826.

29. Maeda, H., *SMANCS and polymer-conjugated macromolecular drugs: Advantages in cancer chemotherapy.* Advanced Drug Delivery Reviews, 2001. **46**(1-3): p. 169-185.

30. Nicolaou, K.C., et al., *Total synthesis of the oligosaccharide fragment of calicheamicin .gamma.1I.* Journal of the American Chemical Society, 1990. **112**(22): p. 8193-8195.

31. Hitchcock, S.A., et al., *A convergent total synthesis of calicheamicin γ1 I.* Angewandte Chemie International Edition in English, 1994. **33**(8): p. 858-862.

32. Hitchcock, S.A., et al., *A remarkable glycosylation reaction: The total synthesis of calicheamicin γ1(I).* Journal of the American Chemical Society, 1995. **117**(21): p. 5750-5756.

33. Shair, M.D., et al., *Total synthesis of (±)-dynemicin A.* Angewandte Chemie International Edition in English, 1995. **34**(16): p. 1721-1723.

34. Shair, M.D., et al., *The total synthesis of dynemicin A leading to development of a fully contained bioreductively activated enediyne prodrug.* Journal of the American Chemical Society, 1996. **118**(40): p. 9509-9525.

35. Ji, N., et al., *Enantioselective synthesis of N1999A2.* Journal of the American Chemical Society, 2006. **128**(46): p. 14825-14827.

36. Kobayashi, S., et al., *Formal total synthesis of neocarzinostatin chromophore.* Journal of Organic Chemistry, 2006. **71**(2): p. 636-644.

37. Nicolaou, K.C., et al., *Total synthesis and stereochemistry of uncialamycin.* Angewandte Chemie International Edition in English, 2007. **46**(25): p. 4704-4707.

38. Kawata, S., et al., *Synthetic study of kedarcidin chromophore: Revised structure.* Journal of the American Chemical Society, 1997. **119**(49): p. 12012-12013.

39. Ren, F., et al., *Kedarcidin chromophore: Synthesis of its proposed structure and evidence for a stereochemical revision.* Journal of the American Chemical Society, 2007. **129**(17): p. 5381-5383.

40. Nicolaou, K.C., et al., *Asymmetric synthesis and biological properties of uncialamycin and 26-epi-uncialamycin.* Angewandte Chemie International Edition in English, 2008. **47**(1): p. 185-189.

41. Jones, G.B., et al., *Synthesis and photochemical activity of designed enediynes.* Journal of the American Chemical Society, 2000. **122**(40): p. 9872-9873.

42. Kraka, E. and Cremer, D., *Computer design of anticancer drugs. A new enediyne warhead.* Journal of the American Chemical Society, 2000. **122**(34): p. 8245-8264.

43. Nuss, J.M., et al., *Transition-metal-catalyzed strategies for the synthesis of neocarzinostatin chromophore and analogues: Intramolecular delivery of palladium controls construction of the biologically relevant dienediyne core.* Journal of the American Chemical Society, 1993. **115**(15): p. 6991-6992.

44. Kennedy, D.R., et al., *Designer enediynes generate DNA breaks, interstrand cross-links, or both, with concomitant changes in the regulation of DNA damage responses.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(45): p. 17632-17637.

45. Schreiber, S.L. and Kiessling, L.L., *Synthesis of the bicyclic core of the esperamicin/calichemicin class of antitumor agents.* Journal of the American Chemical Society, 1988. **110**(2): p. 631-633.

46. Hensens, O.D., et al., *Biosynthesis of NCS Chrom A, the chromophore of the antitumor antibiotic neocarzinostatin.* Journal of the American Chemical Society, 1989. **111**(9): p. 3295-3299.

47. Tokiwa, Y., et al., *Biosynthesis of dynemicin A, a 3-ene-1,5-diyne antitumor antibiotic.* Journal of the American Chemical Society, 1992. **114**(11): p. 4107-4110.

48. Lam, K.S., et al., *Biosynthesis of esperamicin A1, an enediyne antitumor antibiotic.* Journal of the American Chemical Society, 1993. **115**(26): p. 12340-12345.

49. Liu, W. and Shen, B., *Genes for production of the enediyne antitumor antibiotic C-1027 in Streptomyces globisporus are clustered with the cagA gene that encodes the C-1027 apoprotein.* Antimicrobial Agents and Chemotherapy, 2000. **44**(2): p. 382-392.

50. Whitwam, R.E., et al., *The gene CalC encodes for a non-heme iron metalloprotein responsible for calicheamicin self-resistance in Micromonospora.* Journal of the American Chemical Society, 2000. **122**(7): p. 1556-1557.

51. Ahlert, J., et al., *The calicheamicin gene cluster and its iterative type I enediyne PKS.* Science, 2002. **297**(5584): p. 1173-1176.

52. Liu, W., et al., *Biosynthesis of the enediyne antitumor antibiotic C-1027.* Science, 2002. **297**(5584): p. 1170-1173.

53. Murugan, E. and Liang, Z.-X., *Evidence for a novel phosphopantetheinyl transferase domain in the polyketide synthase for enediyne biosynthesis.* FEBS Letters, 2008. **582**(7): p. 1097-1103.

54. Zazopoulos, E., *A genomics-guided approach for discovering and expressing cryptic metabolic pathways.* Nature Biotechnology, 2003. **21**(2): p. 187-190.

55. Liu, W., et al., *The neocarzinostatin biosynthetic gene cluster from Streptomyces carzinostaticus ATCC 15944 involving two iterative type I polyketide synthases.* Chemistry & Biology, 2005. **12**(3): p. 293-302.

56. Gao, Q.J. and Thorson, J.S., *The biosynthetic genes encoding for the production of the dynemicin enediyne core in Micromonospora chersina ATCC53710.* FEMS Microbiology Letters, 2008. **282**(1): p. 105-114.

57. Van Lanen, S.G., *Characterization of the maduropeptin biosynthetic gene cluster from Actinomadura madurae ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis.* Journal of the American Chemical Society, 2007. **129**(43): p. 13082-13094.

58. Liu, W., et al., *Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(21): p. 11959-11963.

59. Udwary, D.W., et al., *Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(25): p. 10376-10381.

60. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Current Opinion in Chemical Biology, 2003. **7**(2): p. 285-295.

61. Wenzel, S.C. and Muller, R., *Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic.* Current Opinion in Chemical Biology, 2005. **9**(5): p. 447-458.

62. Katz, L. and David, A.H., *Chapter 6 The DEBS paradigm for type I modular polyketide synthases and beyond*, in *Methods in Enzymology*. 2009, Academic Press. p. 113-142.

63. Shen, B., *Biosynthesis of aromatic polyketides.* Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids, 2000. **209**: p. 1-51.

64. Moore, B.S. and Hopke, J.N., *Discovery of a new bacterial polyketide biosynthetic pathway.* Chembiochem, 2001. **2**(1): p. 35-38.

65. Gaisser, S., et al., *Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tu57.* Journal of Bacteriology, 1997. **179**(20): p. 6271-6278.

66. Maiese, W.M., et al., *Calicheamicins, a novel family of antitumor antibiotics: taxonomy, fermentation and biological properties.* Journal of Antibiotics 1989. **42**(4): p. 558-563.

67. Zhao, B., et al., *Modulation of nicotinamide adenine dinucleotide and poly(adenosine diphosphoribose) metabolism by calicheamicin gamma 1 in human HL-60 cells.* Cancer Letters, 1990. **50**(2): p. 141-147.

68. Zein, N., et al., *Calicheamicin gamma 1I: an antitumor antibiotic that cleaves double-stranded DNA site specifically.* Science, 1988. **240**(4856): p. 1198-1201.

69. Uesugi, M. and Sugiura, Y., *New insights into sequence recognition process of esperamicin A1 and calicheamicin gamma 1I.: origin of their selectivities and "induced fit" mechanism.* Biochemistry, 1993. **32**(17): p. 4622-4627.

70. Zein, N., et al., *Calicheamicin gamma 1I and DNA: molecular recognition process responsible for site-specificity.* Science, 1989. **244**(4905): p. 697-699.

71. Krishnamurthy, G., et al., *Salt dependence of calicheamicin-DNA site-specific interactions.* Biochemistry, 1995. **34**(3): p. 1001-1010.

72. Walker, S., et al., *Analysis of hydroxylamine glycosidic linkages - Structural consequences of the NO bond in calicheamicin.* Journal of the American Chemical Society, 1994. **116**(8): p. 3197-3206.

73. Hawley, R.C., et al., *Model of the interactions of calichemicin gamma 1 with a DNA fragment from pBR322.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(4): p. 1105-1109.

74. Li, T., et al., *Carbohydrate-minor groove interactions in the binding of calicheamicin .gamma.1I to duplex DNA.* Journal of the American Chemical Society, 1994. **116**(9): p. 3709-3715.

75. Walker, S.L., et al., *NMR characterization of calicheamicin gamma(1)(I) bound to DNA.* Tetrahedron, 1994. **50**(5): p. 1351-1360.

76. Mah, S.C., et al., *Features of DNA recognition for oriented binding and cleavage by calicheamicin.* Tetrahedron, 1994. **50**(5): p. 1361-1378.

77. Bergman, R.G., *Reactive 1,4-dehydroaromatics.* Accounts of Chemical Research, 1973. **6**(1): p. 25-31.

78. Zein, N., et al., *Exclusive abstraction of nonexchangeable hydrogens from DNA by calicheamicin gamma-1.* Journal of the American Chemical Society, 1989. **111**(17): p. 6888-6890.

79. De Voss, J.J., et al., *Site-specific atom transfer from DNA to a bound ligand defines the geometry of a DNA-calicheamicin .gamma.1I complex.* Journal of the American Chemical Society, 1990. **112**(26): p. 9669-9670.

80. Hangeland, J.J., et al., *Specific abstraction of the 5'(S)-deoxyribosyl and 4'-deoxyribosyl hydrogen-atoms from DNA by calicheamicin-gamma-1(I).* Journal of the American Chemical Society, 1992. **114**(23): p. 9200-9202.

81. Biggins, J.B., et al., *Resistance to enediyne antitumor antibiotics by CalC self-sacrifice.* Science, 2003. **301**(5639): p. 1537-1541.

82. Liu, C., et al., *Sequence-selective carbohydrate-DNA interaction: dimeric and monomeric forms of the calicheamicin oligosaccharide interfere with transcription factor function.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(2): p. 940-944.

83. Sissi, C., et al., *Interaction of calicheamicin gamma1(I) and its related carbohydrates with DNA-protein complexes.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(19): p. 10643-10648.

84. Hinman, L.M., et al., *Preparation and characterization of monoclonal antibody conjugates of the calicheamicins: a novel and potent family of antitumor antibiotics.* Cancer Research 1993. **53**(14): p. 3336-3342.

85. Andrews, R.G., et al., *Myeloid-associated differentiation antigens on stem-cells and their progeny identified by monoclonal-antibodies.* Blood, 1983. **62**(1): p. 124-132.

86. Griffin, J.D., et al., *A monoclonal antibody reactive with normal and leukemic human myeloid progenitor cells.* Leukemia Research 1984. **8**(4): p. 521-534.

87. Dinndorf, P.A., et al., *Expression of normal myeloid-associated antigens by acute leukemia cells.* Blood, 1986. **67**(4): p. 1048-1053.

88. Fialkow, P.J., et al., *Clonal development, stem-cell differentiation, and clinical remissions in acute nonlymphocytic leukemia.* New England Journal of Medicine 1987. **317**(8): p. 468-473.

89. Bernstein, I.D., et al., *Differences in the frequency of normal and clonal precursors of colony-forming cells in chronic myelogenous leukemia and acute myelogenous leukemia.* Blood, 1992. **79**(7): p. 1811-1816.

90. Scheinberg, D.A., et al., *A phase I trial of monoclonal antibody M195 in acute myelogenous leukemia: specific bone marrow targeting and internalization of radionuclide.* Journal of Clinical Oncology, 1991. **9**(3): p. 478-490.

91. Appelbaum, F.R., et al., *The use of radiolabeled anti-Cd33 antibody to augment marrow irradiation prior to marrow transplantation for acute myelogenous leukemia.* Transplantation, 1992. **54**(5): p. 829-833.

92. Sievers, E.L., et al., *Selective ablation of acute myeloid leukemia using antibody-targeted chemotherapy: a phase I study of an anti-CD33 calicheamicin immunoconjugate.* Blood, 1999. **93**(11): p. 3678-3684.

93. van Hof, A.C., et al., *Biodistribution of (111)indium-labeled engineered human antibody CTMO1 in ovarian cancer patients: influence of protein dose.* Cancer Research 1996. **56**(22): p. 5179-5185.

94. Takeshita, A., et al., *CMC-544 (inotuzumab ozogamicin), an anti-CD22 immuno-conjugate of calicheamicin, alters the levels of target molecules of malignant B-cells.* Leukemia, 2009. **23**(7): p. 1329-1336.

95. Zhang, J., et al., *A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(5): p. 1460-1465.

96. Fox, B.G., et al., *Stearoyl-acyl carrier protein delta 9 desaturase from Ricinus communis is a diiron-oxo protein.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(6): p. 2486-2490.

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

98. Hu, Z., et al., *A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases.* Microbiology, 2003. **149**(8): p. 2213-2225.

99. Liu, T., et al., *Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin.* Chemistry & Biology, 2008. **15**(5): p. 449-458.

100. Heathcote, M.L., et al., *Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units.* Chemistry & Biology, 2001. **8**(2): p. 207-220.

101. Dillon, S.C. and Bateman, A., *The Hotdog fold: Wrapping up a superfamily of thioesterases and dehydratases.* BMC Bioinformatics, 2004. **5**.

102. Nixon, J.E., et al., *Synthesis of triacetic acid lactone by the pigeon liver fatty acid synthetase complex.* Journal of Biological Chemistry, 1968. **243**(20): p. 5471-5478.

103. Hamilton-Miller, J.M., *Chemistry and biology of the polyene macrolide antibiotics.* Bacteriological Reviews, 1973. **37**(2): p. 166-196.

104. Narasimhachari, N. and Swami, M.B., *Dermostatin: a revised hexaene structure.* Journal of Antibiotics, 1970. **23**(11): p. 566.

105. Davoli, P., et al., *Laetiporic acids, a family of non-carotenoid polyene pigments from fruit-bodies and liquid cultures of Laetiporus sulphureus (Polyporales, Fungi).* Phytochemistry, 2005. **66**(7): p. 817-823.

106. Kadota, I., et al., *A unified approach to polyene macrolides: Synthesis of candidin and nystatin polyols.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(33): p. 11992-11995.

107. Weber, R.W.S., *Laetiporic acid, a new polyene pigment from the wood-rotting basidiomycete Laetiporus sulphureus (Polyporales, Fungi).* Tetrahedron Letters, 2004. **45**(5): p. 1075-1078.

108. Kong, R., et al., *Characterization of a carbonyl-conjugated polyene precursor in 10-membered enediyne biosynthesis.* Journal of the American Chemical Society, 2008. **130**(26): p. 8142-8143.

109. Tee, E.S. and Lim, C.L., *The analysis of carotenoids and retinoids - a review.* Food Chemistry, 1991. **41**(2): p. 147-193.

110. Sun, H., et al., *Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis.* Chemical Communications, 2009(47): p. 7399-7401.

111. Belecki, K., et al., *Production of octaketide polyenes by the calicheamicin polyketide synthase CalE8: Implications for the biosynthesis of enediyne core structures.* Journal of the American Chemical Society, 2009. **131**(35): p. 12564–12566.

112. Chen, X., et al., *Identification of a nonaketide product for the iterative polyketide synthase in biosynthesis of the nine-membered enediyne C-1027.* Angewandte Chemie International Edition in English, 2010. **49**(43): p. 7926-7928.

113. White, S.W., et al., *The structural biology of type II fatty acid biosynthesis.* Annual Review of Biochemistry, 2005. **74**: p. 791-831.

114. Mercer, A.C. and Burkart, M.D., *The ubiquitous carrier protein - a window to metabolite biosynthesis.* Natural Product Reports, 2007. **24**(4): p. 750-773.

115. Smith, S. and Tsai, S.C., *The type I fatty acid and polyketide synthases: a tale of two megasynthases.* Natural Product Reports, 2007. **24**(5): p. 1041-1072.

116. Jenni, S., et al., *Structure of fungal fatty acid synthase and implications for iterative substrate shuttling.* Science, 2007. **316**(5822): p. 254-261.

117. Lai, J.R., et al., *Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis.* Biochemistry, 2006. **45**(50): p. 14869-14879.

118. Xu, G.Y., et al., *Solution structure of B. subtilis acyl carrier protein.* Structure, 2001. **9**(4): p. 277-287.

119. Roujeinikova, A., et al., *Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates.* Journal of Molecular Biology, 2007. **365**(1): p. 135-145.

120. Upadhyay, S.K., et al., *Structural insights into the acyl intermediates of the Plasmodium falciparum fatty acid synthesis pathway: The mechanism of expansion of the acyl carrier protein core.* Journal of Biological Chemistry, 2009. **284**(33): p. 22390-22400.

121. Ploskon, E., et al., *A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains.* Journal of Biological Chemistry, 2008. **283**(1): p. 518-528.

122. Reed, M.A.C., et al., *The type I rat fatty acid synthase ACP shows structural homology and analogous biochemical properties to type II ACPs.* Organic & Biomolecular Chemistry, 2003. **1**(3): p. 463-471.

123. Leibundgut, M., et al., *Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase.* Science, 2007. **316**(5822): p. 288-290.

124. Alekseyev, V.Y., et al., *Solution structure and proposed domain-domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase.* Protein Science, 2007. **16**(10): p. 2093-2107.

125. Evans, S.E., et al., *An ACP structural switch: Conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein.* Chembiochem, 2008. **9**(15): p. 2424-2432.

126. Hadfield, A.T., et al., *The crystal structure of the actIII actinorhodin polyketide reductase: proposed mechanism for ACP and polyketide binding.* Structure, 2004. **12**(10): p. 1865-1875.

127. Findlow, S.C., et al., *Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from Streptomyces rimosus.* Biochemistry, 2003. **42**(28): p. 8423-8433.

128. Crump, M.P., et al., *Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2).* Biochemistry, 1997. **36**(20): p. 6000-6008.

129. Crump, M.P., et al., *Conserved secondary structure in the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2) and the fatty acid synthase acyl carrier protein from Escherichia coli.* FEBS Letters, 1996. **391**(3): p. 302-306.

130. Koglin, A., et al., *Conformational switches modulate protein interactions in peptide antibiotic synthetases.* Science, 2006. **312**(5771): p. 273-276.

131. Weber, T., et al., *Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases.* Structure with Folding & Design, 2000. **8**(4): p. 407-418.

132. Cox, R.J., *Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets.* Organic & Biomolecular Chemistry, 2007. **5**(13): p. 2010-2026.

133. Liang, Z.-X., *Complexity and simplicity in the biosynthesis of enediyne natural products.* Natural Product Reports, 2009. **27**(4): p. 499-528.

134. Murugan, E., et al., *Expression, purification and characterization of the acyl carrier protein phosphodiesterase from Pseudomonas aeruginosa.* Protein Expression and Purification 2010. **71**(2): p. 132-138.

135. Bingle, L.E., et al., *Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes.* Fungal Genetics and Biology 1999. **26**(3): p. 209-223.

136. Horsman, G.P., et al., *Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(25): p. 11331-11335.

137. Wattana-Amorn, P., et al., *Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase.* Biochemistry, 2010. **49**(10): p. 2186-2193.

138. Kotaka, M., et al., *Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis.* Journal of Biological Chemistry, 2009. **284**(23): p. 15739-15749.

139. Yang, D., et al., *Sequence-specific assignments of methyl groups in high-molecular weight proteins.* Journal of the American Chemical Society, 2004. **126**(12): p. 3710-3711.

140. Xu, Y., et al., *Rapid data collection for protein structure determination by NMR spectroscopy.* Journal of the American Chemical Society, 2007. **129**(25): p. 7722-7723.

141. Xu, Y.Q., et al., *A new strategy for structure determination of large proteins in solution without deuteration.* Nature Methods, 2006. **3**(11): p. 931-937.

142. Cornilescu, G., et al., *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* Journal of Biomolecular NMR 1999. **13**(3): p. 289-302.

143. Herrmann, T., et al., *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* Journal of Molecular Biology, 2002. **319**(1): p. 209-227.

144. Kim, Y. and Prestegard, J.H., *Refinement of the NMR structures for acyl carrier protein with scalar coupling data.* Proteins, 1990. **8**(4): p. 377-385.

145. Sharma, A.K., et al., *Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from Plasmodium falciparum provides insight into the mechanism of activation of ACPs.* Biochemistry, 2006. **45**(22): p. 6904-6916.

146. Parris, K.D., et al., *Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites.* Structure, 2000. **8**(8): p. 883-895.

147. Arthur, C.J., et al., *Structure and malonyl CoA-ACP transacylase binding of streptomyces coelicolor fatty acid synthase acyl carrier protein.* ACS Chemical Biology, 2009. **4**(8): p. 625-636.

148. Li, Q., et al., *Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein.* Biochemistry, 2003. **42**(16): p. 4648-4657.

149. Zornetzer, G.A., et al., *Solution structures of spinach acyl carrier protein with decanoate and stearate.* Biochemistry, 2006. **45**(16): p. 5217-5227.

150. Zornetzer, G.A., et al., *The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond.* Biochemistry, 2010. **49**(3): p. 470-477.

151. Liew, C.W., et al., *Induced-fit upon ligand binding revealed by crystal structures of the hot-dog fold thioesterase in dynemicin biosynthesis.* Journal of Molecular Biology 2010. **404**(2): p. 291-306.

152. Ploskon, E., et al., *Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis.* Chemistry & Biology, 2010. **17**(7): p. 776-785.

153. Akey, D.L., et al., *Structural basis for macrolactonization by the pikromycin thioesterase.* Nature Chemical Biology, 2006. **2**(10): p. 537-542.

154. Trauger, J.W., et al., *Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase.* Nature, 2000. **407**(6801): p. 215-218.

155. Kim, B.S., et al., *Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin.* Journal of Biological Chemistry, 2002. **277**(50): p. 48028-48034.

156. Koglin, A., et al., *Structural basis for the selectivity of the external thioesterase of the surfactin synthetase.* Nature, 2008. **454**(7206): p. 907-U968.

157. Linne, U., et al., *Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis.* European Journal of Biochemistry, 2004. **271**(8): p. 1536-1545.

158. Kotowska, M., et al., *Type II thioesterase from Streptomyces coelicolor A3(2).* Microbiology-Sgm, 2002. **148**: p. 1777-1783.

159. Angelini, A., et al., *Structural and enzymatic characterization of HPO496, a YbgC thioesterase from Helicobacter pylori.* Proteins-Structure Function and Bioinformatics, 2008. **72**(4): p. 1212-1221.

160. Li, J., et al., *Crystal structure of the Escherichia coli thioesterase II, a homolog of the human Nef binding enzyme.* Nature Structural Biology, 2000. **7**(7): p. 555-559.

161. Kunishima, N., et al., *A novel induced-fit reaction mechanism of asymmetric hot dog thioesterase PaaI.* Journal of Molecular Biology, 2005. **352**(1): p. 212-228.

162. Wang, F., et al., *Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence.* Chemistry & Biology, 2007. **14**(5): p. 543-551.

163. Willis, M.A., et al., *Structure of YciA from Haemophilus influenzae (HI0827), a hexameric broad specificity acyl-coenzyme a thioesterase.* Biochemistry, 2008. **47**(9): p. 2797-2805.

164. Cheng, Z.J., et al., *Crystal structure of human thioesterase superfamily member 2.* Biochemical and Biophysical Research Communications, 2006. **349**(1): p. 172-177.

165. Benning, M.M., et al., *The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from Pseudomonas sp. strain CBS-3.* Journal of Biological Chemistry, 1998. **273**(50): p. 33572-33579.

166. Thoden, J.B., et al., *The structure of 4-hydroxybenzoyl-CoA thioesterase from Arthrobacter sp strain SU.* Journal of Biological Chemistry, 2003. **278**(44): p. 43709-43716.

167. Thoden, J.B., et al., *X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase.* Journal of Biological Chemistry, 2002. **277**(30): p. 27468-27476.

168. Keegan, R.M. and Winn, M.D., *Automated search-model discovery and preparation for structure solution by molecular replacement.* Acta Crystallographica Section D: Biological Crystallography 2007. **63**(Pt 4): p. 447-457.

169. Morris, R.J., et al., *ARP/wARP and automatic interpretation of protein electron density maps.* Methods Enzymology, 2003. **374**: p. 229-244.

170. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallographica Section D: Biological Crystallography 1998. **54**(Pt 5): p. 905-921.

171. Jones, T.A. and Kjeldgaard, M., *Electron-density map interpretation.* Methods Enzymology, 1997. **277**: p. 173-208.

172. *The CCP4 suite: Programs for protein crystallography.* Acta Crystallographica Section D: Biological Crystallography, 1994. **50**(5): p. 760-763.

173. Emsley, P. and Cowtan, K., *Coot: Model-building tools for molecular graphics.* Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12 I): p. 2126-2132.

174. Bricogne, G., *Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives.* Acta Crystallographica Section D: Biological Crystallography, 1993. **49**(Pt 1): p. 37-60.

175. Leduc, D., et al., *The hotdog thioesterase EntH (YbdB) plays a role in vivo in optimal enterobactin biosynthesis by interacting with the ArCP domain of EntB.* Journal of Bacteriology, 2007. **189**(19): p. 7112-7126.

176. Salas, J.J. and Ohlrogge, J.B., *Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases.* Archives of Biochemistry and Biophysics, 2002. **403**(1): p. 25-34.

177. Guo, Z.-F., et al., *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): p. 1712-1722.

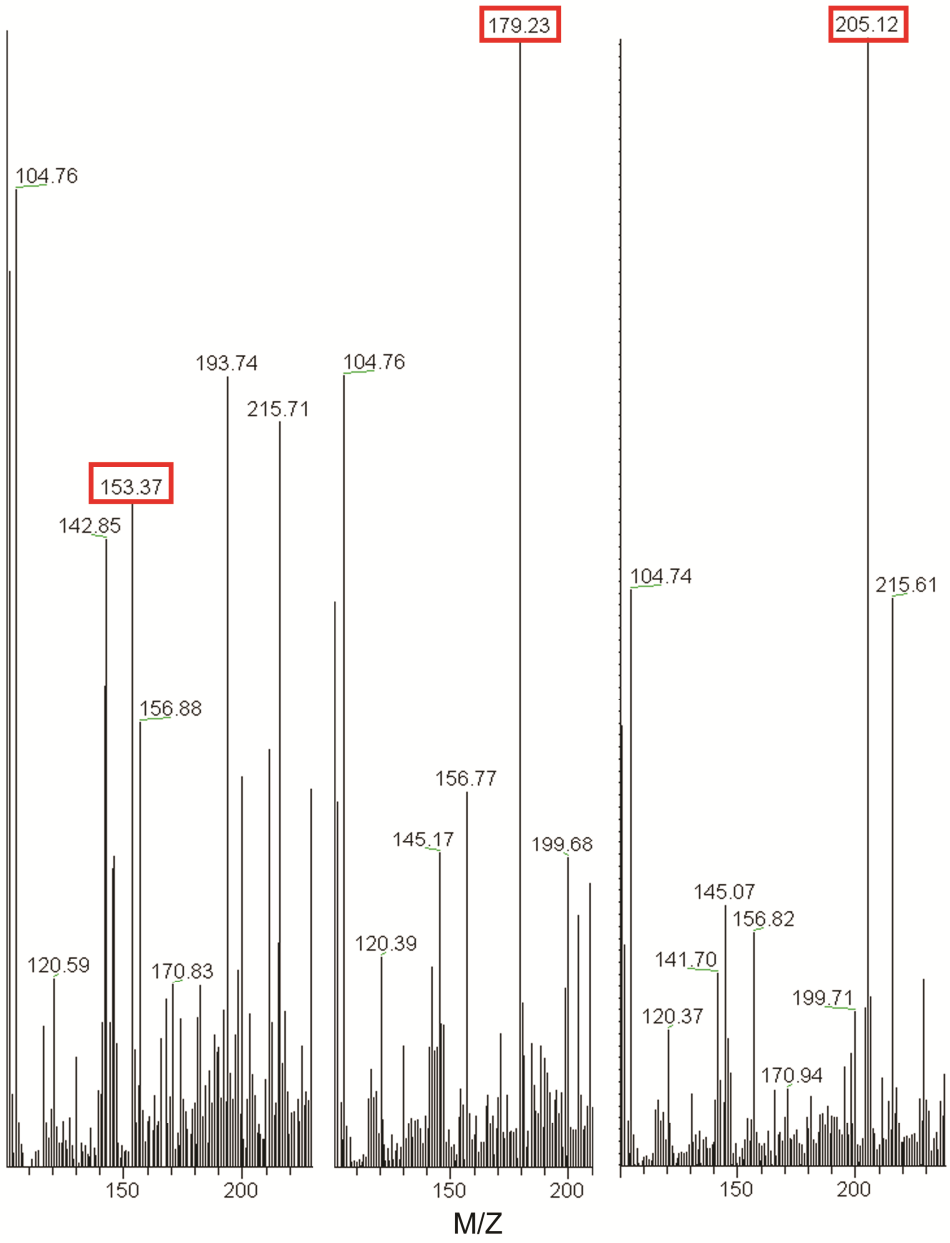
178. Chen, D.Q., et al., *In vitro kinetic analysis of substrate specificity in enterobactin biosynthetic lower pathway enzymes provides insight into the biochemical function of the hot dog-fold thioesterase EntH.* Biochemistry, 2009. **48**(3): p. 511-513.

179. Gigant, B., et al., *X-ray structures of a hydrolytic antibody and of complexes elucidate catalytic pathway from substrate binding and transition state stabilization through water attack and product release.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(15): p. 7857-7861.

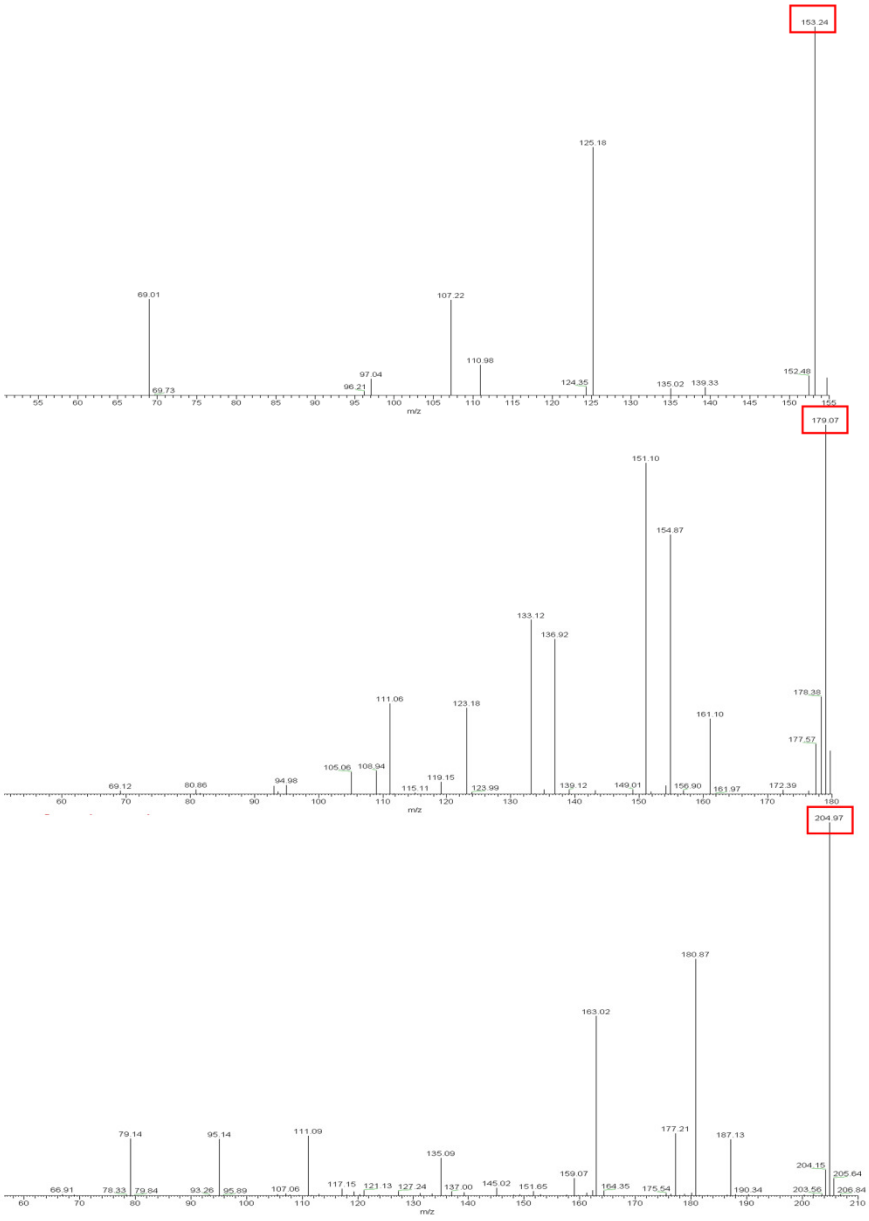
180. Keatinge-Clay, A., et al., *An antibiotic factory caught in action.* Nature Structural & Molecular Biology, 2004. **11**(9): p. 888-893.

181. Maier, T., et al., *The crystal structure of a mammalian fatty acid synthase.* Science, 2008. **321**(5894): p. 1315-1322.

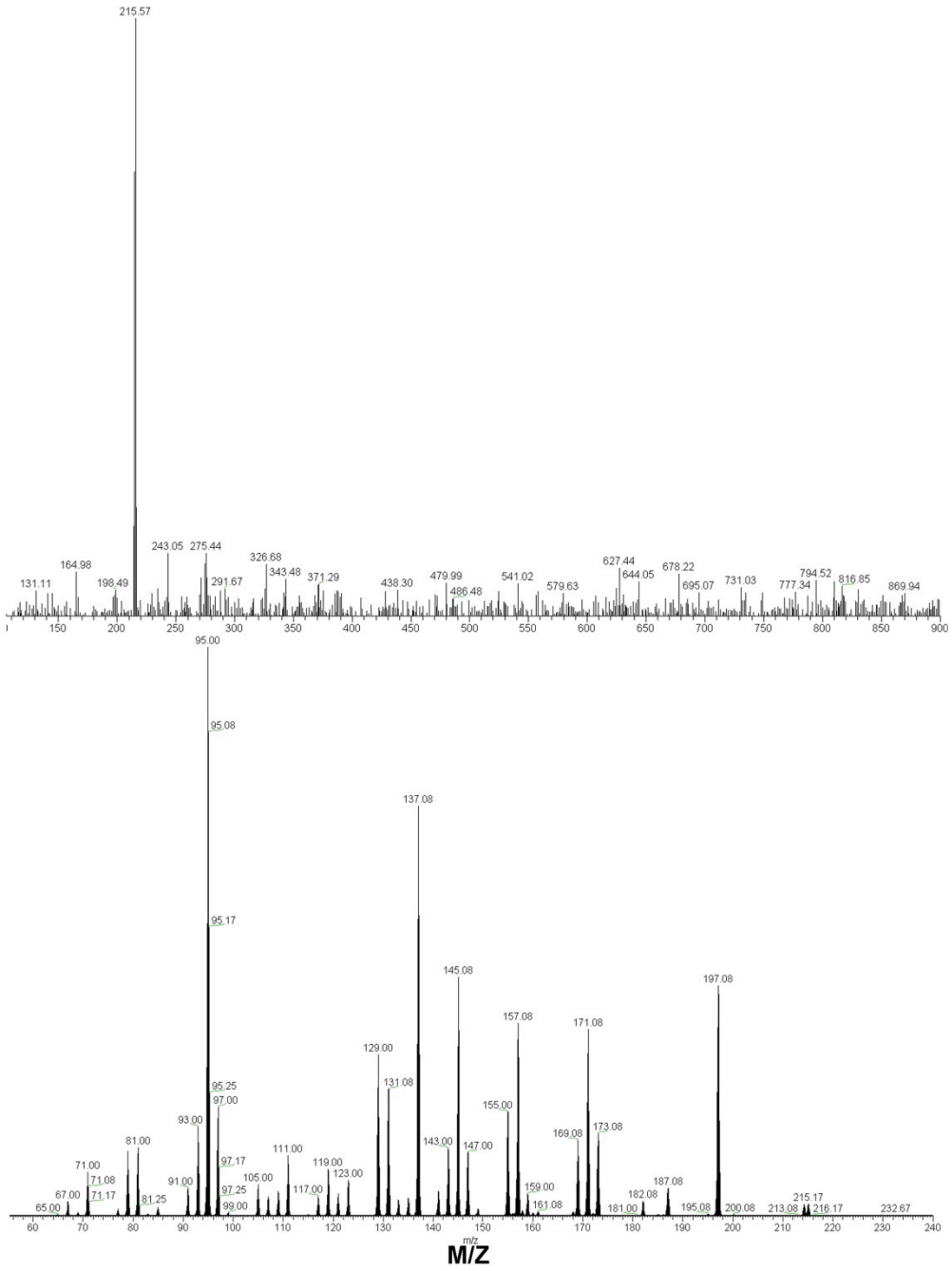
Appendix



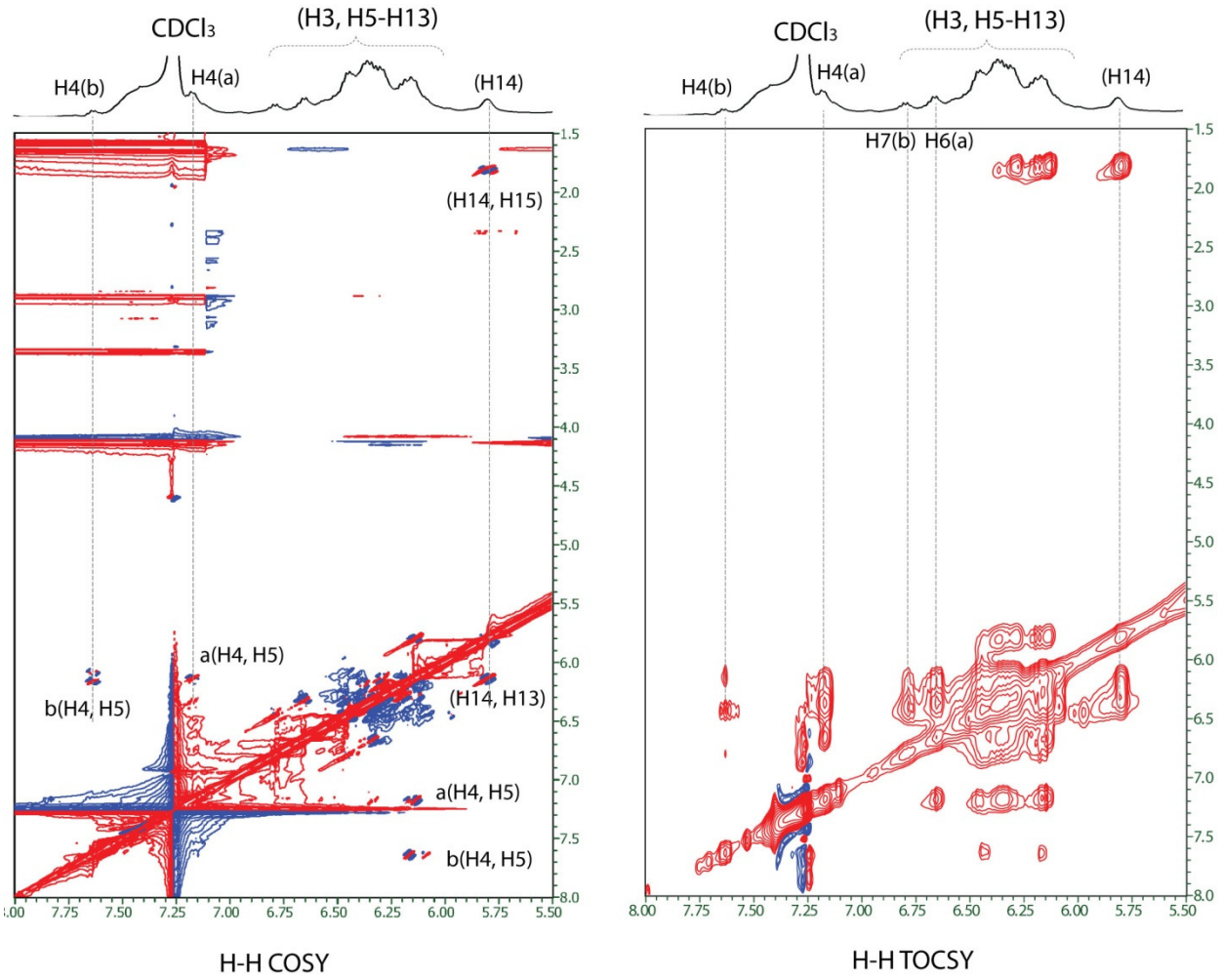
Appendix figure 2.1 Full MS for product **4**, **5** and **6**. Since PKS products are comprised of at least 8 acetate units connected head-to-tail, the masses (circled in red) shown here can only be aberrant products with predicted molecular formula of $C_8H_8O_3$(**4**), $C_{10}H_{10}O_3$ (**5**) and $C_{12}H_{12}O_3$(**6**).



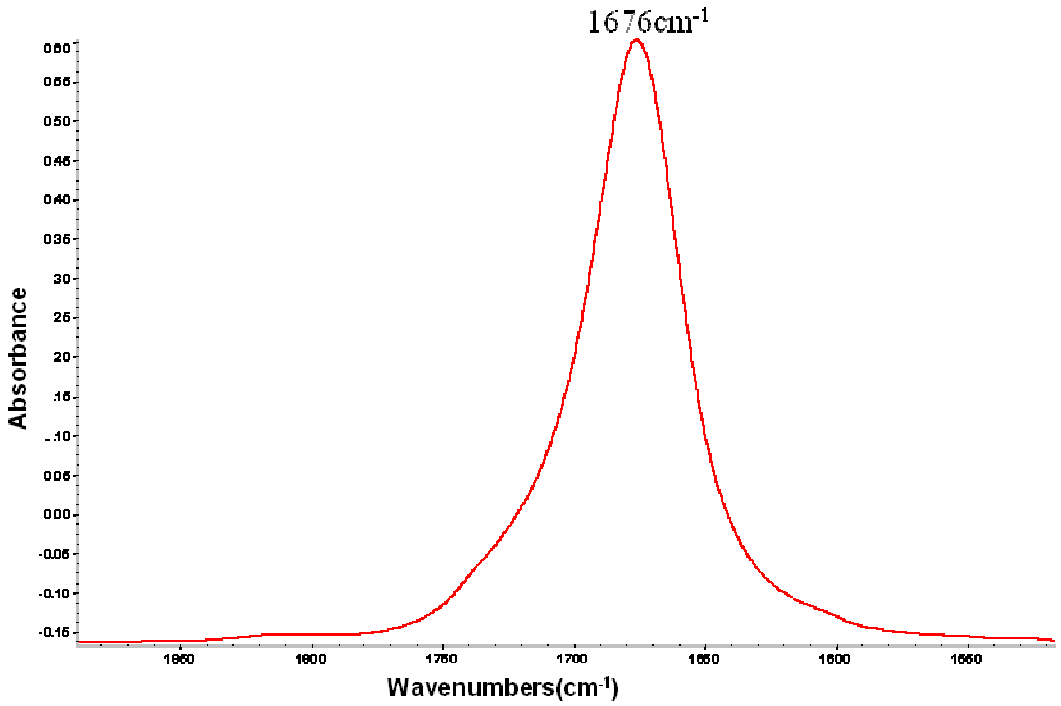
Appendix figure 2.2 Fragmentation pattern for product **4**, **5** and **6** in positive mode MSMS. The intensity for fragment ions are extremely low compared to that of parental ions for all 3 species (figures are not shown in proportion). Therefore, it is likely that all 3 aberrant products are cyclic in nature. The parental species are highlighted in red box.



Appendix figure 2.3 Full MS and MSMS for product **3**. The high resolution mass obtained for protonated **3** (predicted molecular formula of $C_{15}H_{18}O$) is 215.1430. Fragmentation for **3** generates a series of broken-down masses of considerable intensity with respect to parental ion. Given the fact that the long chain acyl product only contains a single oxygen, it is highly plausible that **3** is a linear polyunsaturated polyketide product.



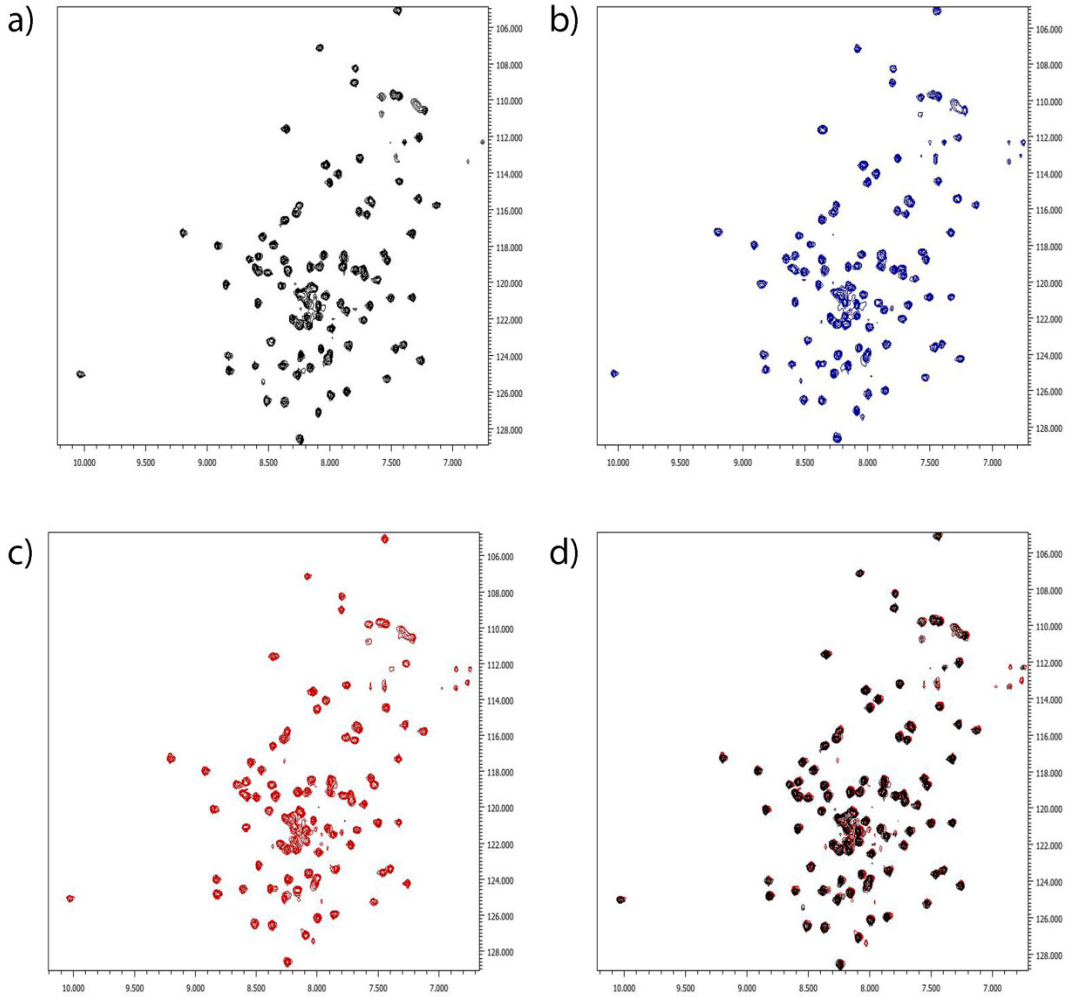
Appendix figure 2.4 H^1-H^1 COSY and TOCSY NMR spectra of product **3** (major and minor isomers). Refer to **Appendix table 2.1** below for the numbering system. Most of the carbon atoms identified in the spectrum are clustered in the polyene region.



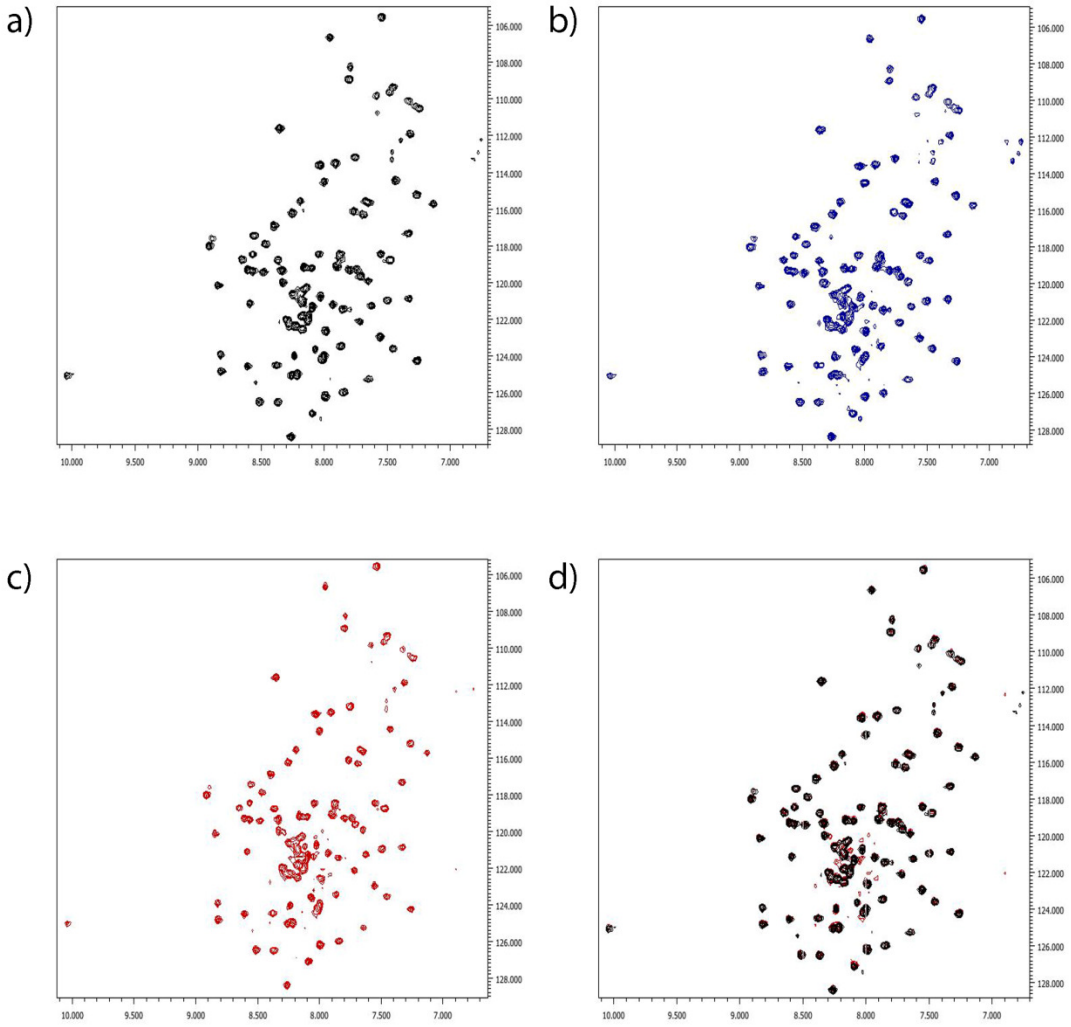
Appendix figure 2.5 FTIR spectrum of the yellow product **3**. The peak at 1676 cm^{-1} is a strong indication for the presence of carbonyl group in product **3**.

	Major	Minor
Proton	**δ_H**	**δ_H**
H1	**2.30**	**2.30**
H3	**6.13**	**6.17**
H4	**7.18**	**7.63**
H5	**6.34**	**6.10**
H6	**6.66**	**6.17**
H7	**6.38**	**6.82**
H8-H12	**6.2-6.5**	**6.2-6.5**
H13	**6.16**	**6.16**
H14	**5.80**	**5.80**
H15	**1.80**	**1.80**

Appendix table 2.1 Proton chemical shift for the major and minor isomers of **3** in $CDCl_3$. Coupling constant (J) cannot be determined for most protons due to signal overlap. Without the coupling constants, the configuration of the olefinic bonds could not be established with the exception of the C5=C6 bond.



Appendix figure 3.1 ^{1}H-^{15}N HSQC of ^{15}N *holo-me*ACP titrated against unlabeled CalE7 at 25 °C in monomeric *holo-me*ACP: CalE7 molar ratio of (a) 1:0 (black) (b) 1:1 (blue) (c) 1:2.5 (red) and (d) an overlaid spectrum between (a) and (c).



Appendix figure 3.2 ^{1}H-^{15}N HSQC of ^{15}N *apo-me*ACP titrated against unlabeled CalE7 at 25 °C in monomeric *apo-me*ACP: CalE7 molar ratio of (a) 1:0 (black) (b) 1:1 (blue) (c) 1:2.5 (red) and (d) an overlaid spectrum between (a) and (c).

Publications

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.* Journal of the American Chemical Society, 2008. **130**(26): p. 8142-8143.

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.* Journal of Biological Chemistry, 2009. **284**(23): p. 15739-15749.

Sun H**, Kong R**, Zhu D, Lu M, Ji Q, Liew CW, Lescar J, Zhong G, Liang ZX., *Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis*. Chemical Communications, 2009. (47): p. 7399-7401.

Murugan E, **Kong R**, Sun H, Rao F, Liang ZX., *Expression, purification and characterization of the acyl carrier protein phosphodiesterase from Pseudomonas aeruginosa*. Protein Expression and Purification, 2010. **71**(2): p. 132-138.

Liew CW, Sharff A, Kotaka M, **Kong R**, Sun H, Qureshi I, Bricogne G, Liang ZX and Lescar J., *Induced-fit upon ligand binding revealed by crystal structures of the hot-dog fold thioesterase in dynemicin biosynthesis.* Journal of Molecular Biology, 2010. **404**(2): p. 291-306.

Lim JW*****, **Kong R***, Murugan E, Ho CL, Liang ZX, Yang DW., *Solution Structures of the Acyl Carrier Protein Domain from a the Highly Reducing Type I Iterative Polyketide Synthase CalE8.* PLoS ONE, (Submitted)

***** Co-first authors.