A study of the functional domains of the type I iterative polyketide synthase CalE8 in calicheamicin biosynthesis

Murugan, Elavazhagan

2012

Murugan, E. (2012). A study of the functional domains of the type I iterative polyketide synthase CalE8 in calicheamicin biosynthesis. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/50212

https://doi.org/10.32657/10356/50212

A STUDY OF THE FUNCTIONAL DOMAINS OF THE TYPE I

ITERATIVE POLYKETIDE SYNTHASE CALE8 IN

CALICHEAMICIN BIOSYNTHESIS.

ELAVAZHAGAN MURUGAN

SCHOOL OF BIOLOGICAL SCIENCES

NANYANG TECHNOLOGICAL UNIVERSITY

2012

A STUDY OF THE FUNCTIONAL DOMAINS OF THE TYPE I ITERATIVE POLYKETIDE SYNTHASE CALE8 IN CALICHEAMICIN BIOSYNTHESIS.

ELAVAZHAGAN MURUGAN

SCHOOL OF BIOLOGICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY

A thesis submitted to the Nanyang Technological University

In fulfillment of the requirement for the degree of

Doctor of Philosophy

2012

Acknowledgement

First and foremost, I submit my gratitude to A/P. Liang Zhao-Xun. I started this PhD at a junction where I had to ask myself the question of whether to pursue science and research. It was my association with him and his research penchant that realigned my pursuit for excellence in research. His patience and understanding has really helped me during the lag phases of my research. I thank my guide A/P. Liang Zhang-Xun for his teaching and guidance. His regular assessments and continuous discussions have been instrumental in the shaping of this thesis. His sharing of wisdom, patience and understanding have made me from student of science to person who has decided to pursue a lifelong career in science. Thanks for instilling the love of science in me, Prof.

I thank my colleagues Kong Rong, Yaning, Rao Feng, Swathi, Mary, Lawrence, Hui Hua and all others who shared the science along with the fun and pressure. I couldn't have made it without you all. I thank my friends and peers in NTU. Thanks Krupakar, for your continuous and unyielding presence and support from the start. Thanks Malathy, Rathi, Gayathri, Priya-Rishi, Mano Ravi, Kareem, Uma, Manisha-Raj and all my fellow peers and juniors for making this process a memorable and less stressful one.

Thanks Ashok, Inthrani, Sara and all my friends from NUS and NUS TLS – Praba, Sajitha, Sasi, Kiran, Jraj-Durkesh, Chanthira, Tahira-Waseem, Kumar, Ruby and all others for having been there throughout. Special thanks to Darshini-Selva and family. Thanks to my former housemates and cricket gang including Chan, Leo, Siva and all my theatre colleagues for all the fun and support that kept me going even during the toughest of times.

Thanks Babu, Shobana and Arun for always being there.

I dedicate this thesis to my parents, my sister, my wife and my son, for their relentless patience and love.

Elavazhagan Murugan

Table of Contents

Table of Contents

List of Figures

List of Tables

Abbreviations

Acetyl-CoA	Acetyl-Coenzyme A
ACN	Acetonitrile
ACP	Acyl carrier protein
actKR	actinorhodin KR
AT	Acyl Transferase
CalE8	Calicheamicin Polyketide Synthase
CalE7	Calicheamicin Thioesterase
CoA	Coenzyme A
CP	Carrier Protein
DACP	DEBS ACP
DEBS	6-Deoxyerythronolide B Synthase
DH	Dehydratase
DHD	Double hot-dog
DTT	Dithiothreitol
DynE8	Dynemicin Polyketide Synthase
EACP	*Escherichia coli* ACP
*Ec*AcpH	*Escherichia coli* phosphodiesterase
EDTA	Ethylenediaminetetraacetic Acid
ER	Enoyl Reductase
FAS	Fatty Acid Synthase
FPLC	Fast Performance Liquid Chromatography
GPCP	GrsA peptidyl carrier protein
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	KiloDaltons
KR	Ketoreductase
KS	Keto-acyl Synthase
LB	Luria-Bertani Media
LC-MS	Liquid Chromatography Mass Spectroscopy
Malonyl-CoA	Malonyl-Coenzyme A
MAT	Human FAS malonyl acyl transferase
*me*ACP	ACP Domain of CalE8 from *M. echinospora ssp. Calichensis*
MS/MS	Tandem Mass Spectroscopy
MT	Human mitochondrial malonyl CoA transferase
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
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
pNpp	para nitro phenol phosphate
PPTase	Phosphopantetheinyl Transferase
rpm	Revolutions per Minute
SDR	Short chain dehydrogenase/reductase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SgcE	C-1027 Polyketide Synthase
TE	Thioesterase
TFA	Trifluoroacetic Acid
tylKR	Tylosin KR
UV-Vis	Ultraviolet-Visible
WT	Wild Type

Abstract

Title: A study of the functional domains of the type I iterative polyketide synthase CalE8 in calicheamicin biosynthesis.

Naturally occurring enediynes are potent antibiotics produced by soil and marine microorganisms. Their robust antitumor activities and unique mode of action make them a significant topic of study. The synthesis of the enediyne products is initiated by a type I iterative polyketide synthase (PKS). In this project, we examine the structures and functions of the domains of CalE8, to demarcate the domains of the the iterative PKS from the biosynthetic pathway of the 10-membered enediyne calicheamicin in *Micromonospora echinospora spp*.

The 212 KDa CalE8 contains several domains including the predicted ketoacyl synthase (KS), acyl transferase (AT), ketoreductase (KR) and dehydratase (DH) domains. In addition, CalE8 also contains a postulated acyl carrier protein (ACP) domain and a C-terminal domain with unknown function. The first step of enediyne biosynthesis involves a post-translational modification of the ACP domain by 4'- phosphopantetheinylation. The ACP domain and C-terminal portion of CalE8 were first cloned and expressed as stand-alone proteins to study their functions. The identity of the ACP domain was established by in vitro phosphopantetheinylation using the surfactin PPTase (Sfp) from *Bacillus subtilis*. The NMR solution of the ACP domain was solved to show that the ACP exhibits some rather distinct structure feature from other ACPs. Furthermore, we found that the C-terminal domain exhibits PPTase activity towards various carrier proteins. Sequence analysis and modeling studies suggest the C-terminal domain is an unusual Sfp-like PPTase domain integrated into CalE8.

Finally, the AT, KS, DH and KR domains of the CalE8 PKS were examined thoroughly by bioinformatics tools such as structural modeling to define the domain boundaries and catalytic residues. The individual domains were cloned and expressed in *E. coli* for structure determination. Although the KS, DH and KR domain proteins were found to be insoluble, the AT domain was soluble and purified for further studies. The access to the soluble AT domain will be valuable for studying the substrate specificity of the AT domain in accepting malonyl-CoA, but not acetyl-CoA as substrate.

For characterizing the covalently-attached products of CalE8, an ACP phosphodiesterase capable of cleaving the growing polyketides was examined. This family of phosphodiesterases was found to be highly unstable with the propensity to form precipitate in solution. We have identified a phosphodiesterase (*Pa*AcpH) from *Pseudomonas aeruginosa* that can be expressed and purified as a soluble protein. The function and substrate specificity of *Pa*AcpH was first validated and examined with several carrier proteins from different pathways. We demonstrate that *Pa*AcpH is indeed able to catalyze the removal of the phosphopantetheinyl moiety and the tethered-intermediates from CalE8. The capability of releasing the polyketide intermediates by *Pa*AcpH is valuable in the study of PKS mechanisms.

Chapter 1 Introduction

1.1 Natural products as drugs

Natural products have always been the crux of most of the discoveries of the scientific world. Medical research through the ages has always sought to mimic the systems existing in nature. The major breakthroughs of medical science, especially pertaining to drug discovery, have mostly exploited natural systems and products. In the modern scientific universe, almost 20,000 proposed bioactive microbial secondary metabolites are known. They can adopt many different roles as antibiotics, antitumor and anticancer agents, toxins, ionophores, bioregulators, and intra- and interspecific signaling molecules. One of the most sought-after secondary metabolite molecules in the field of drug discovery is the polyketide family of natural products.

1.2 Polyketides

Polyketides are a family of secondary metabolites that are ubiquitous in distribution produced by bacteria, fungi, plants, insects, dinoflagellates, molluscs and sponges. Research on polyketides dates back to 1893, when Collie and Myers isolated the first polyketide, orcinol (*1*). In the process of classifying synthetic aromatic molecules during his studies on pyrones, organic chemist, J.Collie, first employed the term, "polyketides" (*1*). He suggested that the β-polyketides could form the starter molecules for the biosynthesis of hydroxyl-substituted natural phenolic compounds found in living organisms. The term polyketide coined by Collie encompassed both polyketones and their phenolic derivatives and suggested that polyketides can be derived from polymerization of ketenes. His proposition was dormant for almost fifty years before another biochemist, Birch, developed this theory and postulated that polyketones might originate from the repetitive condensation of acetate units. He tested his hypothesis via feeding experiments with the advent of radioactive techniques (*2*). Following the addition of ^{14}C labeled acetic acid

into the fungal culture, isotopically labeled 6-methysalicylic acid (6-MSA) was obtained. Polyketides were universally recognized after this validation and Collie's theory gained acceptance. Intensive research in the field of polyketide products followed with objectives of discovering more natural products with unique mechanisms of synthesis and functionalities.

Polyketides are broadly divided into three classes: type I polyketides, mainly macrolides produced by multimodular megasynthases; type II polyketides, aromatic molecules produced by the iterative action of dissociated enzymes; and type III polyketides, small aromatic molecules produced by fungal species. Polyketides are a very diverse family of natural products structurally and functionally with unique biological activities and pharmacological properties. They display an array of antimicrobial, antifungal, antiparasitic, antitumor and agrochemical activities. This abundance of functionalities of polyketides makes them economically, clinically and industrially imperative molecules. Some of the most popular polyketide products such as Erythromycin A, a broad spectrum macrolide antibiotic, the antihelminthic agent avermectin or the immune suppressants FK506 and rapamycin are potent commercially available drugs. Oleandomycin, rifamycin, lovastatin, oxytetracycline and reserveratrol are a few others of the thousands of polyketides discovered so far.

1.3 Enediynes

One of the prominent members of the polyketides is the enediyne class of antibiotics. Enediyne polyketides are a class of natural toxins that demonstrate excellent antibacterial and anticancer properties. In the search for potent natural anticancer agents in medical science research, these secondary metabolites have acquired great significance in the recent years. Enediynes are produced mainly by soil and marine microorganisms. They exhibit remarkable

antitumor properties (*3, 4*) with cytotoxicities analogous to microbial metabolites. Their distinctive structural features combined with their cytotoxicities and unique modes of action have made them an important subject of interest and research.

The first member of the enediyne class of antibiotics, the neocarzinostatin chromophore, was described in 1985 by Edo and co-workers (*5, 6*). The molecular structure of the antibiotic presented a highly unique bicyclo (7.3.0) dodecadienediyne core that holds an epoxide-masked 1, 5-diyne-3-ene (or ene-di-yne) unit embedded inside a 9-membered ring. Two years later, the first members of the calicheamicin (*7-9*) and esperamicin (*10, 11*) families were isolated, followed by the reporting of dynemicin characterization in 1989 (*12*). The structures of the calicheamicins and esperamicins divulged a different bicyclo (7.3.1)-tridecadiynene core that contains the 1, 5-diyne-3-ene unit within a 10-membered ring (*7-14*). Thenceforth, several other subfamilies of enediyne natural products that contain either the bicyclo (7.3.0) dodecadienediyne or bicyclo (7.3.1)-tridecadiynene core have been characterized.

The naturally occurring enediyne chromophores were categorized based on the presence of the bicyclo (7.3.0) dodecadienediyne or bicyclo (7.3.1)-tridecadiynene cores into 9-membered (**Table 1.1**) or 10-membered families (**Table 1.2**). The members of the 9-membered ring family (**Fig. 1.1**) currently includes neocarzinostatin (**1**), kedarcidin (**2**), C-1027 (**3**), maduropeptin (**4**), N1999A2 (**5**). The 10-membered family **(Fig. 2)** includes several subfamilies. The structures of the representative calicheamicin γ_1^I (**8**), esperamicin A_1 (**9**), dynemicin A (**12**) and shishijimicin A (**10**), as well as namenamicin (**11**) and uncialamycin (**13**) are shown in **Fig. 1.2**. The 10-membered enediynes are further divided into the calicheamicin-like enediynes (**8**, **9**, **10** and **11**)

and dynemicin-like enediynes (**12**, **13**), which are characterized by a slightly different bicyclic enediyne core and an anthraquinone moiety.

Table 1.1 Enediyne family. 9-membered ring family of Enediynes and their source organisms

	Name of the enediyne	Organism
Nine-membered ring	Auromycin	*Streptomyces macromomyceticus*
	Largomycin	*Streptomyces pluricolorescens*
	Actinoxanthin	*Actinomyces globisporus*
	Sporamycin	*Streptosporangium pseudovulgare*
	Neocarzinostatin	*Streptomyces carzinostaticus*
	C-1027	*Streptomyces globisporus*
	Maduroprotein	*Actinomadura madurea*
	Kedarcidin	*Actinomycete L585-6*
	N1999A2	*Streptomyces sp. AJ9493*
	Sporolides A and B	*Salinispora tropica*
	Cyanosporasides A and B	*Salinispora pacifica*

Figure 1.1 Enediyne chromophores. Representative molecules of the 9-membered enediyne chromophores with their bicyclo (7.3.0) dodecadienediyne core highlighted in red.
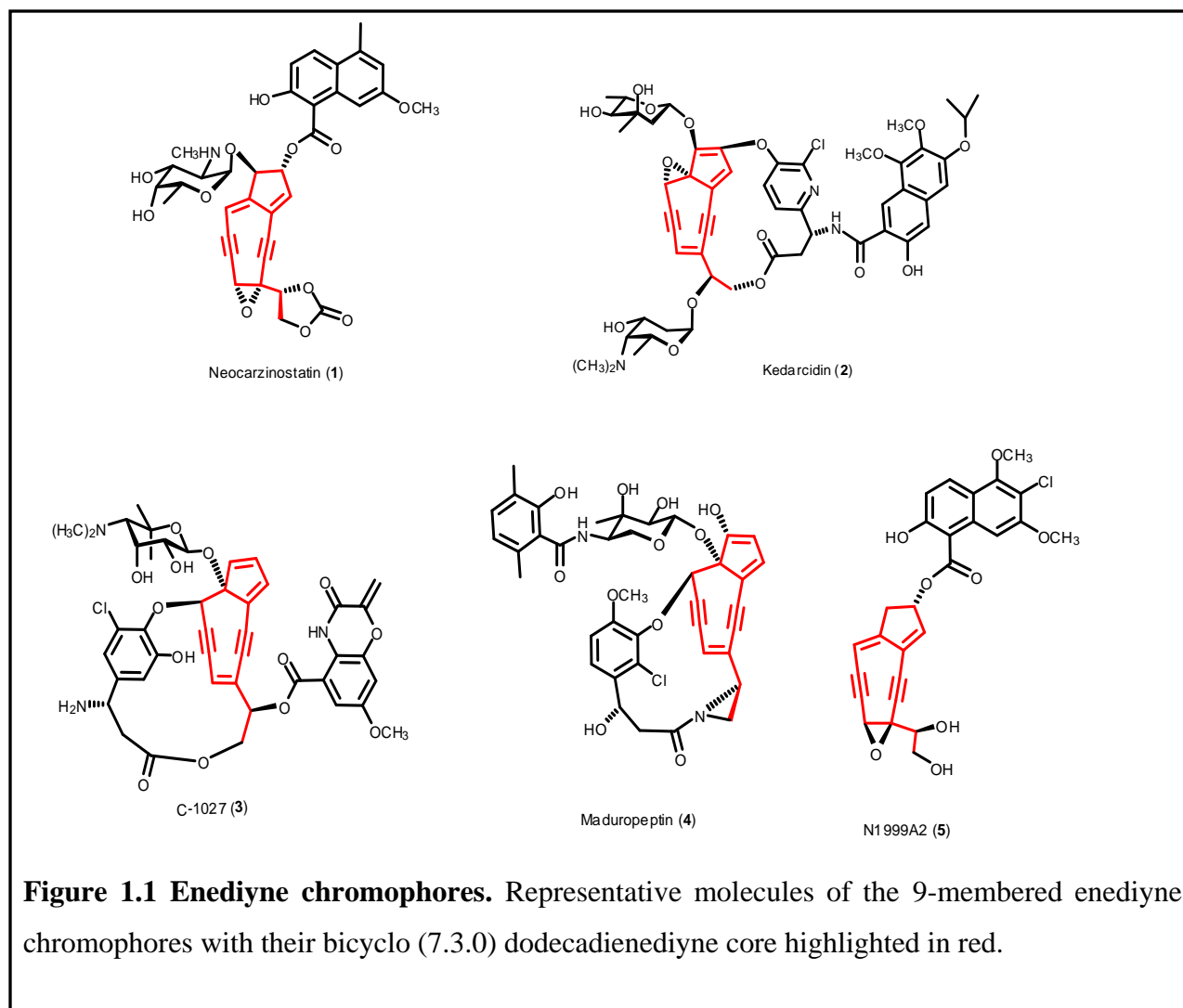
Table 1.2 Enediyne family 10-membered ring family of Enediynes and their source organisms

	Name of the enediyne	Organism
Ten-membered ring	Esperamicin	*Actinomadura verrucosospora*
	Calicheamicin	*Micromonospora echinospora ssp. calichensis*
	Dynemicin	*Micromonospora chersina*
	Namenamicin	*Polysyncraton lithostrotum*
	Shishijimicin	*Didemnum proliferum*
	Uncialamycin	Unknown

Figure 1.2 Enediyne chromophores. Representative molecules of the 10-membered enediyne chromophores with their bicyclo (7.3.1) dodecadienediyne core highlighted in red.
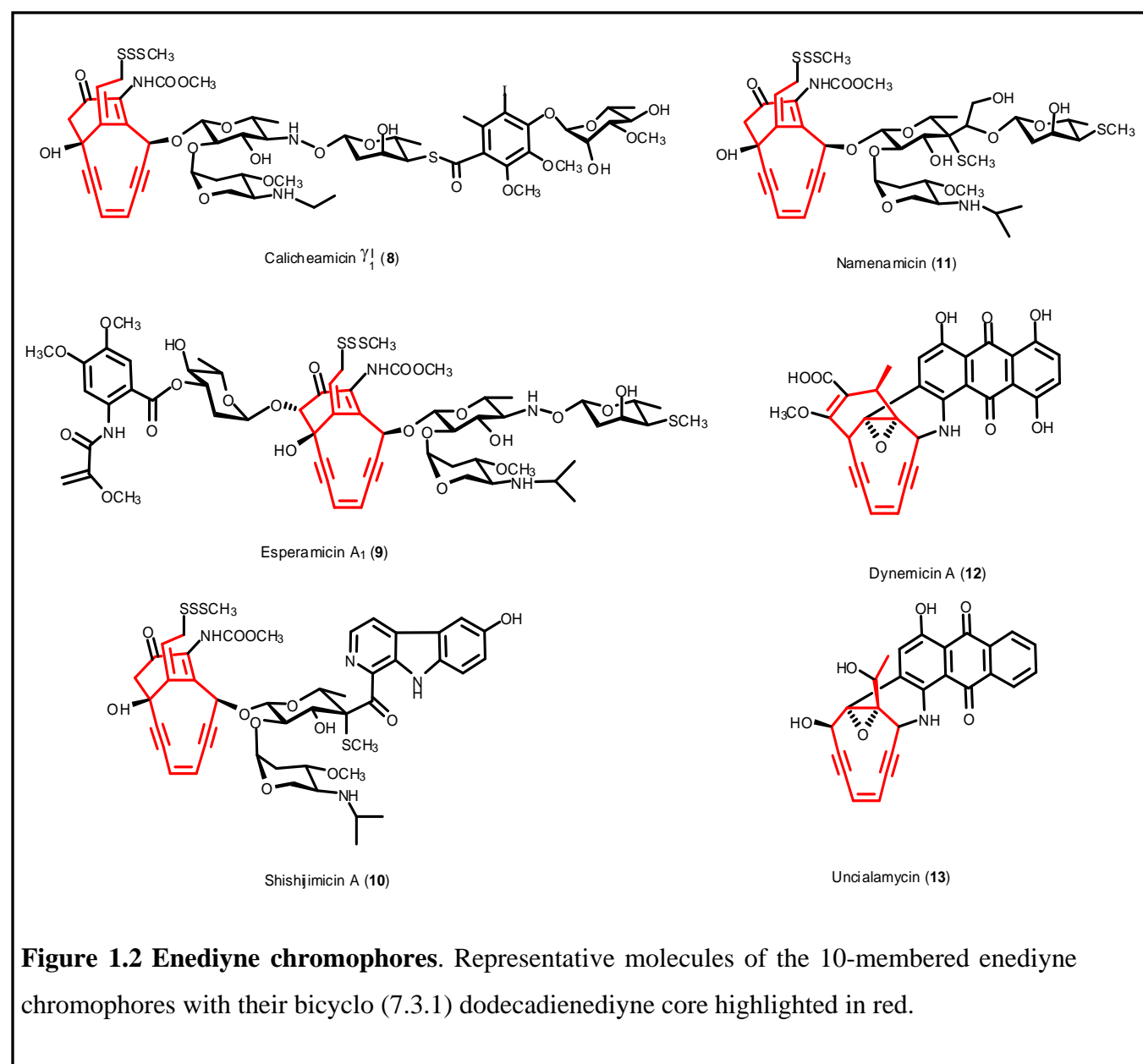
1.3.1 Enediyne – Mechanism of activity

The mode of action adopted by the enediynes in inducing tumor cell killing and cell *apo*ptosis was thoroughly examined. It was found that enediynes intercalated into chromosomal DNA through a sequence-specific procedure and exerted their antibiotic and antitumor activities largely by damaging chromosomal DNA *(4, 16-22).* Synthetic and naturally occurring enediynes have been studied and their mechanism of action, which involves the formation of a reactive enediyne diradical through the Bergman or Myers-Saito cyclization mechanism *(23)* (**Fig. 1.3),** has been established.

Figure 1.3 Mechanisms of diradical formation for 9- and 10- membered enediynes.
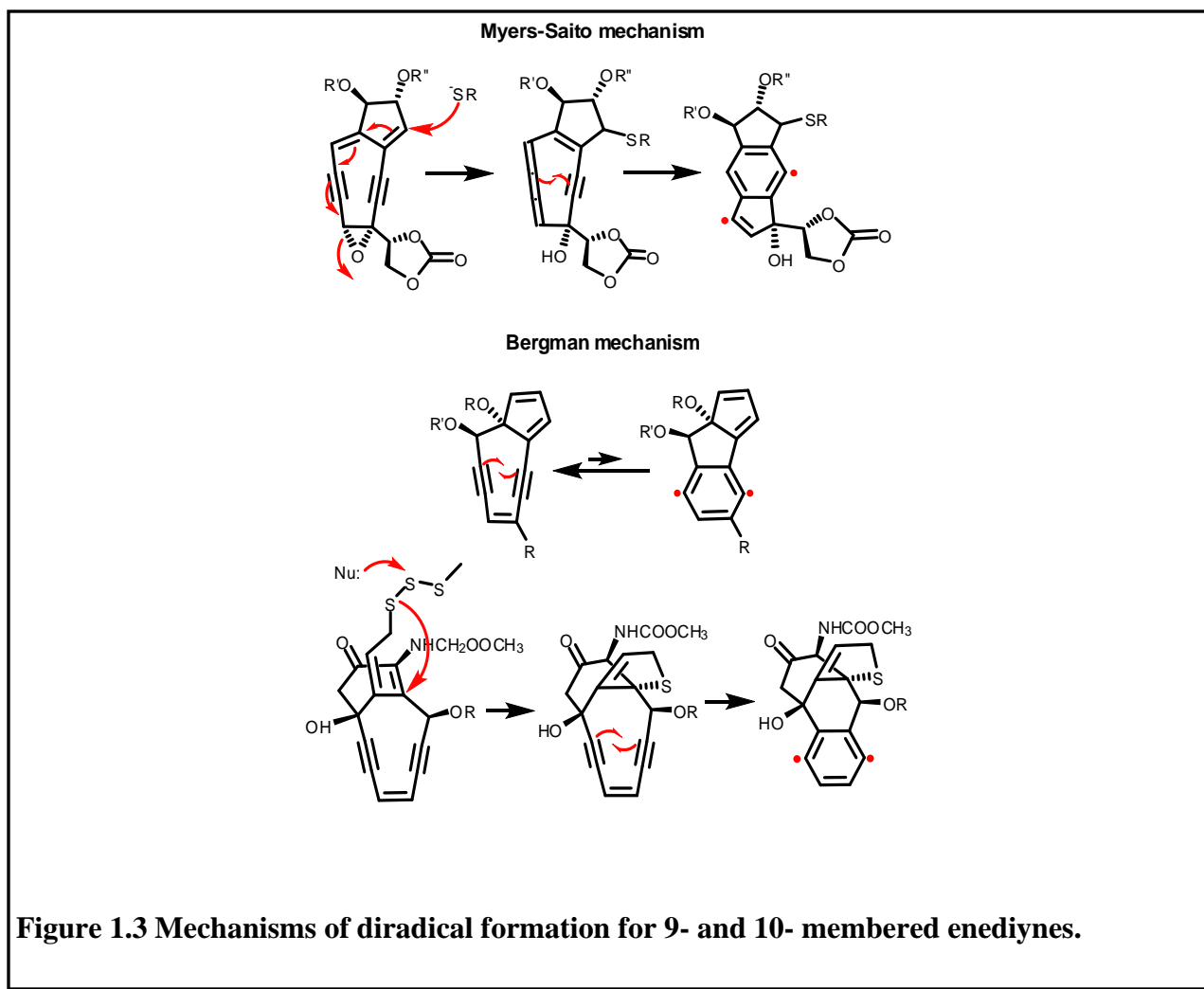
The first detailed study of the unique behaviour of acyclic enediynes upon thermal activation was described in 1972 *(24)*. A simple enediyne A (**Fig. 1.4**), when heated in solution, gave rise to benzene-related products C, depending on the solvent used. Bergman *(24)* recognized the importance of the highly unstable 1, 4-didehydro radical B, responsible for the abstraction of two atoms from the accessible donor such as the solvent, leading to the stable benzene structure C (**Fig. 1.4**).

Figure 1.4 Bergman cycloaromatization of acyclic enediynes (Adapted from Matija, 2007, *25)*)
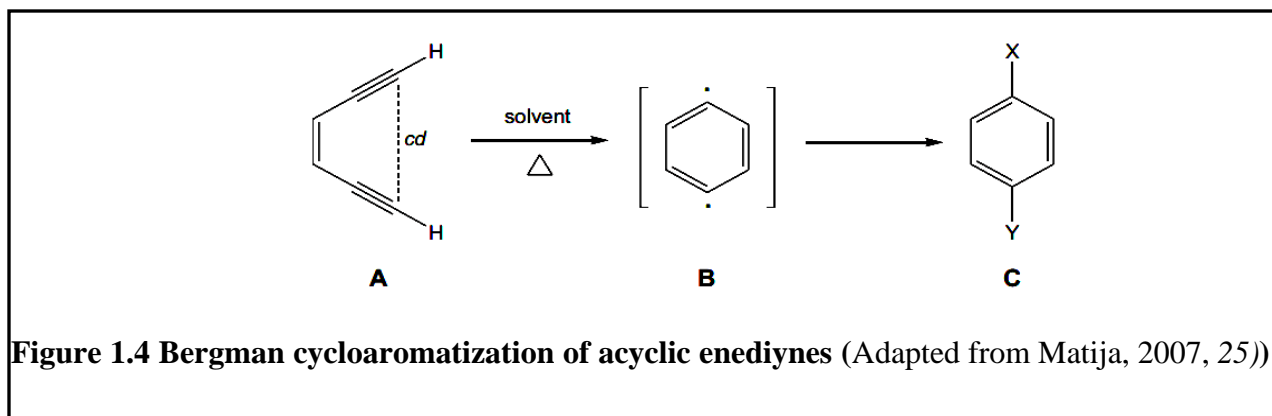
Though Bergman cycloaromatization *(24)* induced a lot of interest among physical, organic and theoretic chemists, it was with the discovery of enediyne anticancer antibiotics, it experienced a revival. The process of cycloaromatization executed by the enediyne moiety, the cause of the DNA cleavage, made enediyne compounds unique in the group of chemotherapeutic agents. Combinatorial experimental and computational methods were employed to determine the factors that affected the reactivity of enediynes in the Bergman reaction *(26)*. These factors include, 1) the *cd* distance; distance between the two terminal acetylenic carbons **(Fig. 1.4),** 2) difference in strain energy between the enediyne and the transition state, 3) the concentration of the trapping agents, and 4) substituent effects.

It has been seen that the cyclic enediynes undergo the Bergman reaction at considerably lower temperatures than the acyclic ones. This could be because of the fact that in the absence of other factors, the interatomic *cd* distance was the factor that determined the rate of the Bergman reaction. Shortening the *cd* value was a way to overcome the energy gap necessary for the cyclization. This was easily achieved by complexation with transition metals allowing smooth control of the *cd* value by means of ionic radii. Also, it was found that substituents at the terminal acetylenic carbons had a significant influence on the energy of the Bergman cyclization, while those attached to the vinyl carbons generally did not have any influence. Compared to cyclic enediynes, the main factor determining the rate of the Bergman cyclization of acyclic enediynes was the difference in the strain energy between the enediyne and the transition state. Along with this the solvent used and the concentration of hydrogen donors are also to be taken into account when comparing cyclization rates of different enediyne compounds.

The mechanism of action of enediyne antibiotics was first elucidated in detail in one of the prominent families of enediynes, the Calicheamicins. More details on this family of enediynes are provided later in this chapter. The extended sugar part is made up of the distinctively linked monosaccharide units. The substituted benzene ring serves as a delivery unit and elicits the binding of the molecule to the minor groove of the double-helical DNA. The nucleophilic addition to the bicyclic enediyne "warhead" triggers the cycloaromatization and concomitant formation of the diradical intermediate. The nucleophilic attack on the allylic trisulphide moiety triggers the device activation **(Fig. 1.5)**. A thiol which is formed in the first step then initiates a cascade of reactions that instigate a significant change in the structural geometry of the molecule. The enediyne moiety trapped within the 10-membered ring is significantly modified by these changes. The modified molecular geometry leads to Bergman cycloaromatization. The enediyne

moiety forms a highly reactive benzene diradical that strips the hydrogen atoms from the sugar phosphate backbone of DNA strands. This stripping leads to the scission of the DNA double helix *(26)*.
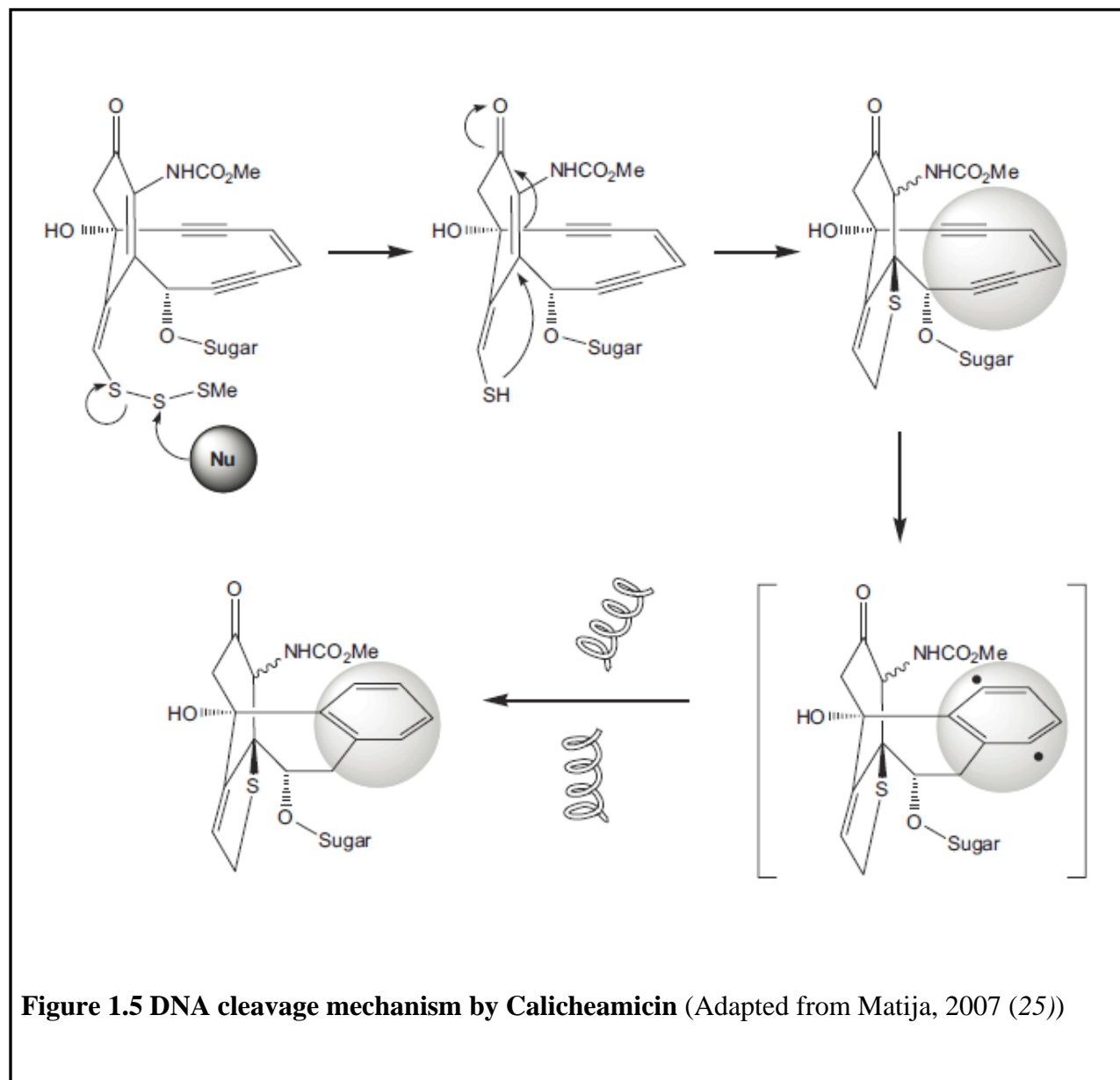
Figure 1.5 DNA cleavage mechanism by Calicheamicin (Adapted from Matija, 2007 (*25*))

In comparison to calicheamicin, neocarzinostatin undergoes spontaneous cycloaromatization at ambient temperature by virtue of the free-standing highly labile chromophore. The diradical gets locked in the minor groove of DNA and readily abstracts hydrogen atoms from DNA. In this process it generates a carbon-centered radical on the ribose. The DNA will then undergo facile double or singled-stranded cleavage in the presence of O_2, through an oxidative radical mechanism.

1.3.2 Process of biosynthesis

The unique mode of biosynthesis adopted by the enediyne family of products has been a subject of great fascination for researchers. Complete synthesis of enediyne chromophores was actively pursued in the 80s and 90s by the organic chemists inspired by the unprecedented structure and biological activity of these natural products. The total synthesis of almost every enediyne subfamily, including the daunting calicheamicin γ_1^I and dynemicin A, has been accomplished *(28-42).* In addition, a large number of artificial or designed enediynes have been synthesized *(35, 36, 43-50).* These elaborate synthetic studies not only assisted the elucidation of the complex structures, but also shed light on the mechanism of action in DNA binding and cleavage.

The atypical molecular architecture and reactivity of the enediynes have induced great interest in the biosynthetic origin of the enediyne cores. With the concise synthetic routes devised for the enediyne cores of calicheamicins and neocarzinostatin in 1988, Schreiber and Kiessling demonstrated a series of steps that include electrocyclic ring closure, proton transfer and oxidation could transform **14** into a 9-membered enediyne (**Pathway a, Fig. 1.6**); whereas an intra-molecular Diels-Alder reaction following the addition of one carbon unit at the acetylene

terminus could convert **14** into a 10-membered enediyne (**Pathway b**, **Fig. 1.6**). The authors postulated that the nonlinear molecule **14** might be the common biosynthetic precursor for both the 9- and 10-membered enediynes (*51*).

Figure 1.6 Proposed biosynthetic mechanism for enediynes
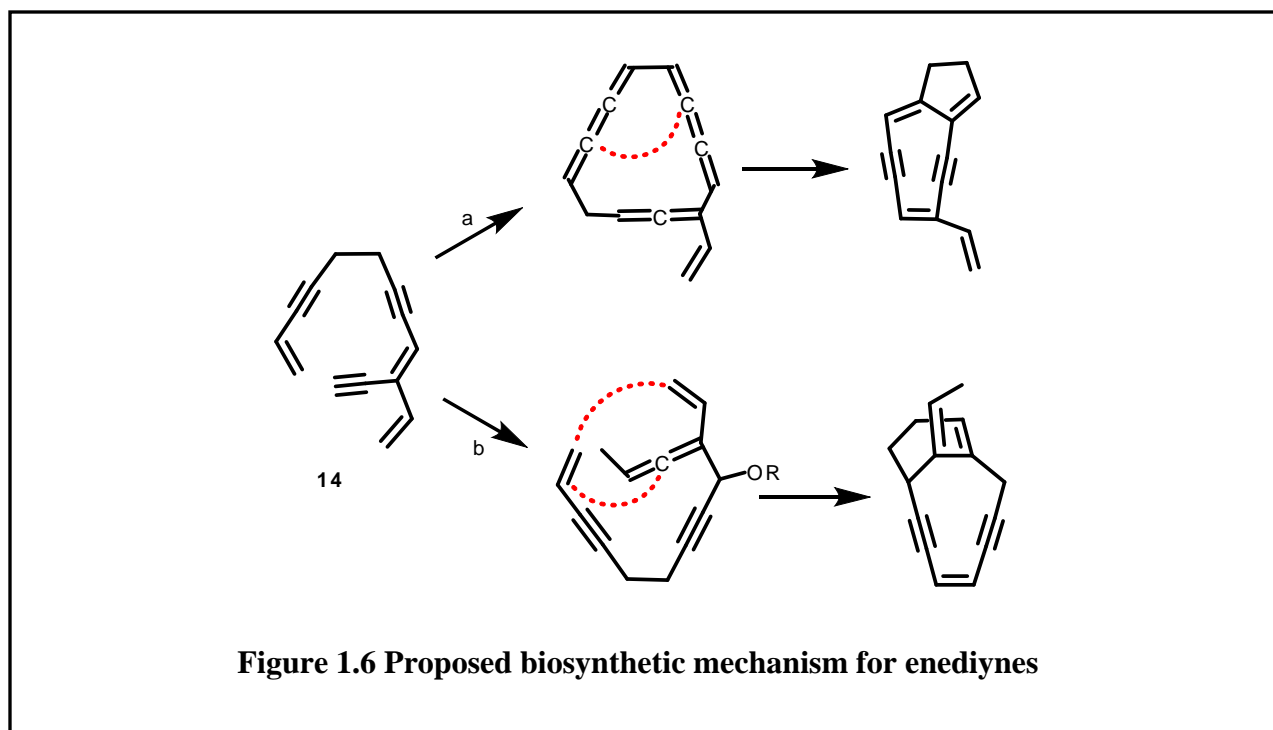
Hensens and coworkers produced ^{13}C-enriched neocarzinostatin in 1989 by feeding ^{13}C-labelled acetate to *Streptomyces carzinostaticus* (*52*). ^{13}C NMR studies confirmed the structures proposed by Edo and Myers and also yielded the first clue about the biosynthetic origins of the molecular building block. The experiment with singly labeled acetate suggested that the bicyclo (7.3.0) dodecadienediyne moiety is likely to be derived from a linear precursor that consists of seven acetate units (**Fig. 1.7**).

Figure 1.7 Folding pattern for the 9-membered enediyne of neocarzinostatin
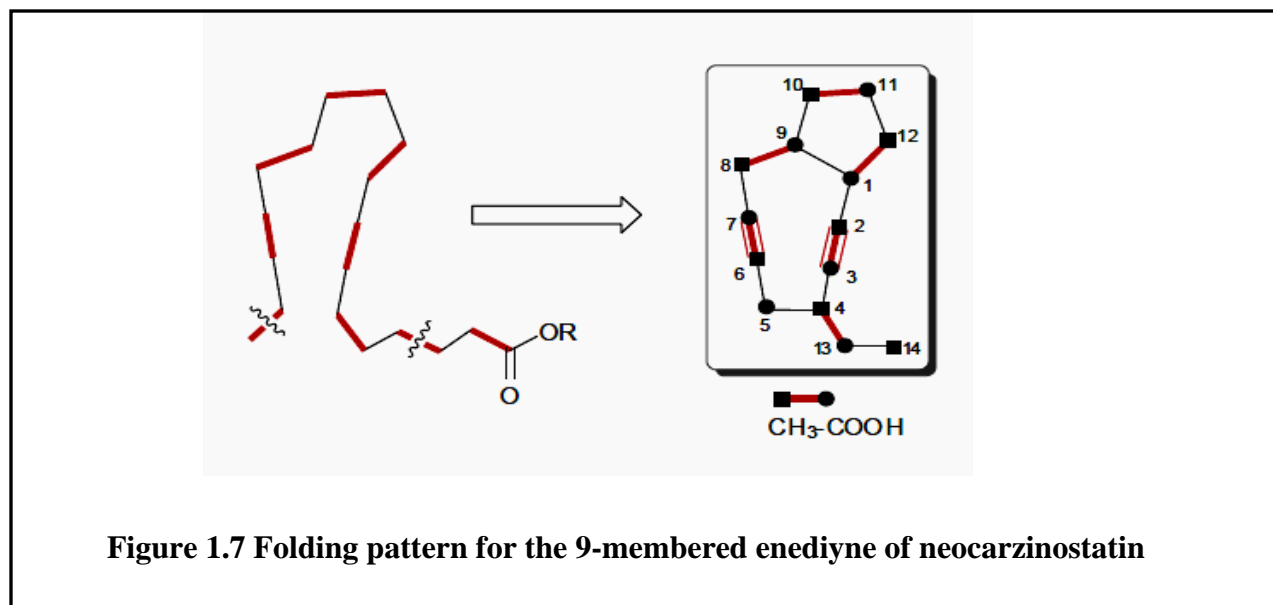
Culturing with doubly- and mixed-labeled acetate further indicated that the acetates were assembled in a head-to-tail fashion. Importantly, the two carbons of the −yne group seem to originate from the same acetate unit, which marks a striking difference between the 9- and 10- membered enediynes. The linear precursor has probably been derived from oleate or crepenynate, and truncated on both ends prior to cyclization. Subsequent oxidation, cyclization and oxygenation would furnish the fully functionalized 14 carbon-containing enediyne core.

Another isotopic labeling study conducted by Tokiwa and coworkers, a few years later, revealed a very different ^{13}C-incoporation pattern for the bicycle (7,3,1)-tridecadiynene core of dynemicin A from *Micromonospora chersina M956-1* (*53*). The authors suggested that the bicyclic enediyne core and the anthraquinone moiety of dynemicin A are derived from two heptaketide chains that consist of two sets of seven acetate units. There are two possible pathways for the linear heptaketide to fold into the final bicyclic structure (**Fig. 1.8**). Identification of the incorporated acetate units from the (1, 2-^{13}C$_2$) acetate feeding experiment suggested that the

cyclization follows pathway a rather than b. From the results, it is also evident that the two carbons comprising the –yne group in dynemicin are derived from different acetate units, in sharp contrast to the early finding that the two carbons of the -yne group in neocarzinostatin originate from the same acetate unit (*52*). Although uncialamycin (**13**), the other member of the dynemicin-type enediyne subfamily, appears to contain a different enediyne core, it was postulated that **13** is derived from a dynemicin-like precursor *(54)*

Figure 1.8 Folding pattern for the 10-membered enediyne of dynemicin A
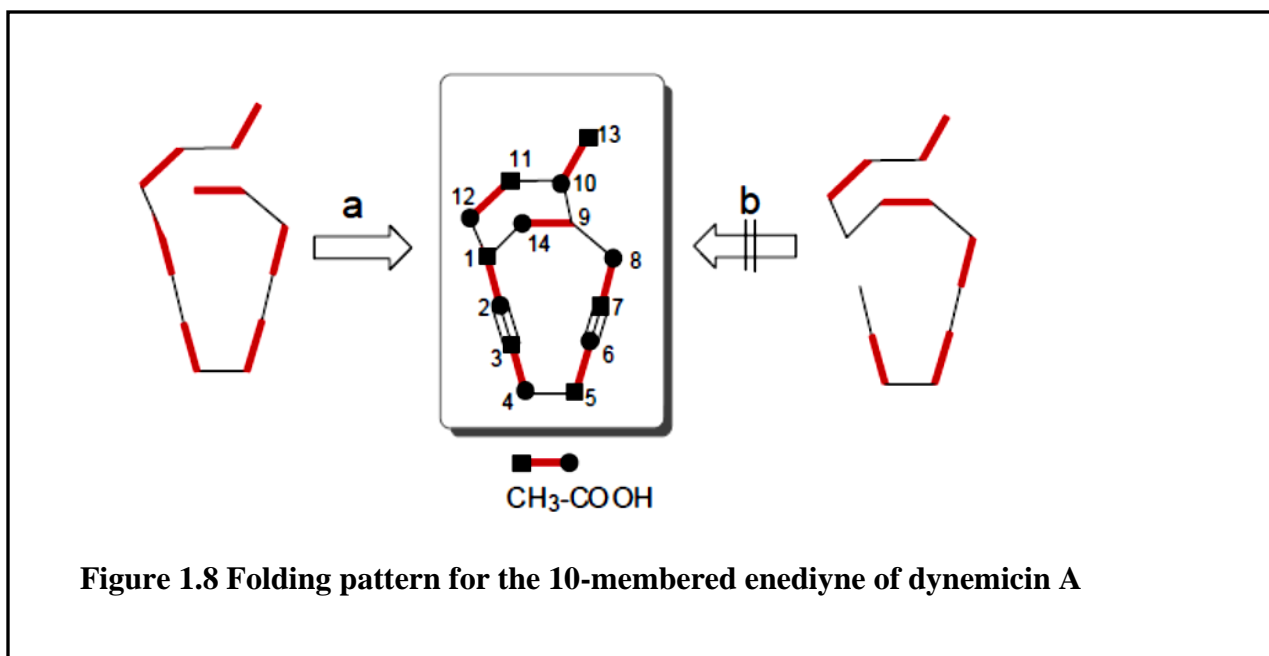
An isotopic labeling experiment similar to the former with the esperamicin-producing *Actinomadura verrucosospora* revealed that the enediyne of esperamicin A_1 is also assembled from seven acetate units in a head-to-tail fashion (*55*). In this scenario, there are four possible pathways for the linear heptaketide to fold into the bicyclic structure (**Fig. 1.9**).
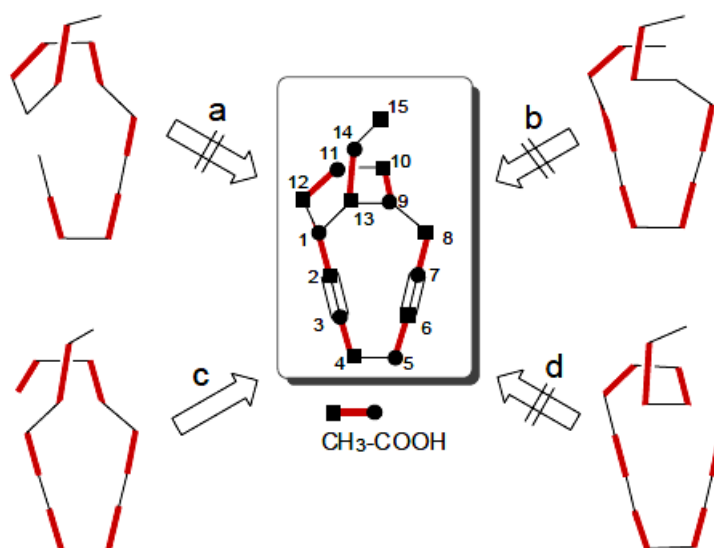
Figure 1.9 Folding pattern for the 10-membered enediyne of esperamicin A

Identification of the incorporated acetate units, particularly the initiating acetate unit C11-C12, ruled out the pathways a, b and d. The synthesis of esperamicin A_1 by *A. verrucosospora* was significantly reduced by the administration of cerulenin, an inhibitor targeting the β-ketosynthase (KS) domain of both fatty acid synthase (FAS) and polyketide synthase (PKS); however supplementation of the culture with the fatty acid oleate did not restore the biosynthesis. With these observations, it was deduced that the enediyne warhead is probably derived from a polyketide precursor, rather than a fatty acid one.

The three isotope labeling experiments together postulated that the enediyne cores are derived from linear, probably polyketide, precursors that comprise head-to-tail coupled acetate units. The dissimilarities observed in the isotope incorporation patterns between neocarzinostatin, calicheamicin and dynemicin, especially in the origins of the –yne carbons, also revealed intriguing differences among the biosynthetic pathways.

1.3.3 Enediyne gene clusters

Little or almost no new insight was gained on the biosynthesis of the enediynes in the next several years following the isotopic labeling experiments. DNA hybridization experiments were done on the genomic DNA of the enediyne producers to locate the putative type I or type II PKS gene. Since these attempts were unsuccessful other genome mining methods were employed. Shen and coworkers searched for the *cagA* gene encoding the chromophore-associated *apo* protein CagA and two putatively conserved deoxysugar-synthesizing genes and succeeded in locating the DNA fragment that encompasses the enediyne gene clusters for C-1027 (*56*). Almost simultaneously, the gene locus for calicheamicin biosynthesis was identified by Thorson and co-workers who screened a genomic library for the cosmids that confer calicheamicin resistance to *M. echinospora* (*3, 57*). Both these groups announced the completion of the sequencing and partial annotation of the C-1027 and calicheamicin gene clusters in 2002 (*58, 59*). The annotated genetic information on calicheamicin provided the researchers with the genetic blueprints and prompted the study of enediyne biosynthesis into the genomic era.

Both the gene clusters were of massive sizes, ~80 kb, and contained 70 to 80 genes that were predicted to encode a multitude of enzymes, transporters, transcriptional regulators and proteins with unknown functions (*58, 59*). The most important aspect of the two gene clusters is that they contain a pair of PKS genes (*calE8* and *sgcE*) which encode two single-module PKSs sharing high sequence homologies and display identical domain organizations. When either of the PKS genes was disrupted, the production of the metabolite was completely abolished. Nonetheless, gene complementation restored the biosynthesis in the mutant strains. Apart from the PKS genes themselves, four more genes were found to be conserved in the vicinity of the PKS gene,

between the two gene clusters. These five conserved genes are most likely involved in the assembly of the enediyne warhead since C-1027 and calicheamicin γ_1^I do not share any common peripheral moiety. The PKS gene and the four conserved genes were suggested to constitute the so-called 'minimal enediyne PKS gene cassette' or 'warhead gene cassette' (**Fig. 1.10**) (*59, 60*). Also, based on sequence homology, the functions of the gene products for the biosynthesis of peripheral building blocks, self-resistance, regulation and uptake were assigned tentatively.
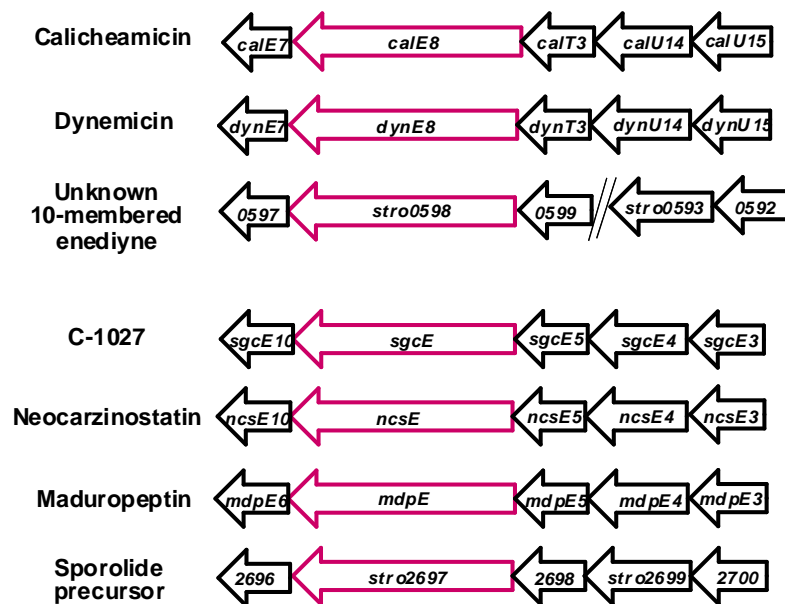
Figure 1.10 Organization of the highly conserved warhead gene cassette

The conservation of the warhead gene cassette was confirmed in the following years by the sequencing of the gene clusters for neocarzinostatin, maduropeptin and dynemicin (**Fig. 1.10**) (*61, 62*). A dozen gene loci containing the gene cassette in soil-dwelling *Actinomycetes* strains

were discovered by Thorson and coworkers using a high-throughput genome scanning method (*61*). This detection revealed a wider distribution of enediynes than realized earlier. Furthermore, the biosynthetic gene clusters of a large number of secondary metabolites were unveiled by the genomic sequencing of the marine bacterium *Salinospora tropica* (*63-65*). Among these gene clusters, two gene clusters were found to contain the minimal enediyne PKS gene cassettes. One of them was confirmed to be responsible for the biosynthesis of the enediyne precursor of sporolides which are macrocyclic metabolites isolated from *Salinispora tropica* by Fellica in 2005. The other gene cluster has been predicted to encode a 10-membered enediyne based on its phylogenetic relationship with other minimal PKS cassette genes (*63-66*).

The sequencing of the enediyne gene clusters has established the polyketide origin of the enediyne core. This has marked the beginning of a new era for comprehensive elucidation of the biosynthetic pathways and mechanisms. And it has also aroused the possibility of employing metabolic engineering techniques to generate enediyne analogues.

1.3.4 Naturally occurring enediynes

1.3.4.1 Neocarzinostatin

Neocarzinostatin (NCS) is one of the most potent antibiotics occurring in nature and was first reported in 1965 (*67*), when it was isolated from *Streptomyces carzinostaticus* Var. F-41. It was a complex consisting of a 1:1 non-covalently associated mixture of NCS *apo* protein and the NCS chromophoric component **(Fig. 1.11).** Further research on this compound revealed that the chromophoric moiety is responsible for the biological activity while the NCS *apo* protein stabilizes and transports the drug. The NCS *apo* protein is made up of a polypeptide chain 113

amino acid in length. This polypeptide binds firmly and specifically to the chromophore and is also responsible for the delivery of the chromophore to the site of action, the DNA molecule, by controlled release. It was also reported (*19*) that NCS *apo* protein activated by proteolysis, possesses cytotoxic activity.
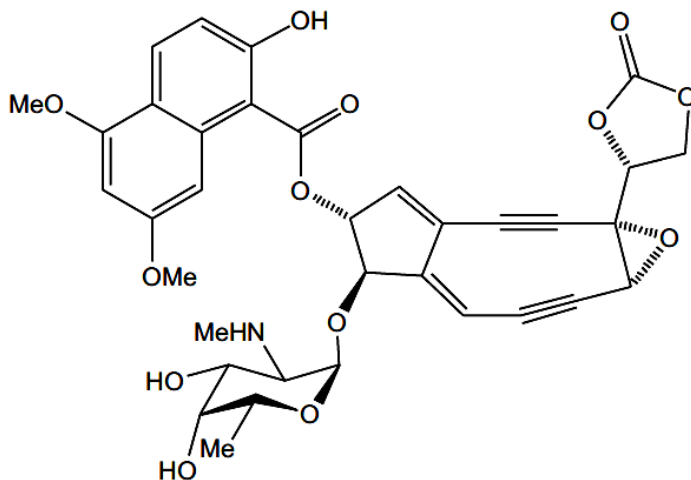
Figure 1.11 Neocarzinostatin chromophore

The effective DNA cleavage induced by NCS dictates its antitumor and antibacterial properties. This involves NCS chromophore intercalation into the DNA with its naphthoate side chain that positions the rest of the molecule within the minor groove. The free and biologically-active NCS chromophore moiety cleaves single-strands of the DNA. The DNA cleavage is followed by an oxygen-dependent reaction. The DNA cleavage by NCS chromophore is greatly enhanced by the presence of thiols and UV radiation (*35, 36*).

The DNA cleavage mechanism of the NCS chromophore (**Fig. 1.12**) was initially proposed in 1987 (*4*). The series of reactions that leads to the DNA damage starts with the stereospecific

nucleophilic attack at Carbon-12, as seen in thiols. Rearrangement of the ring skeleton and epoxide opening follows this initiation event thereby leading to the formation of cumulene structure **2**. This is a highly strained and an extremely reactive intermediate. This intermediate undergoes cycloaromatization instantly to form the diradical **3**. This diradical then attacks the DNA with abstraction of the two hydrogen atoms from the DNA producing **4**.

The three-dimensional solution structure of NCS provides clear information on the role of the NCS *apo* protein. The aminomethyl group attached to the sugar pendant is sterically aligned close to C-12 caused by a salt bridge with asparagine 33 in the polypeptide chain. This nucleophilic attack is expected to be aided by basic nitrogen group of asparagine and sterical interferences at C-12 of the amino acids serine 98, asparagine 33 and phenylalanine 52 in a polypeptide chain. These structural alignments position the epoxide in a hydrophobic pocket away from acid catalysts. The diradical **3** formation induces DNA damage at two abstraction points in the deoxyribose sugar component. The C(5') atom of adenine and thymine residues is the first abstraction point and 80% of DNA cleavages occur at this junction. The abstraction of hydrogen atom from C(1') and C(4') results in less than 20% of strand breaks. The abstraction of the C(5') hydrogen by C(6) in the diradical **3** induces single-strand DNA cuts. Double strand DNA lesions are the result of additional C(1') or C(4') abstraction by the C(2) radical.

Thiol-independent DNA cleavage is possible when a DNA molecule contains regions where double-stranded structures are generated intramolecularly and thus, appear as single-strands. If that is the case, DNA would be a participant in its own destruction (4). NCS requires sulfhydryl activation for the activity. This results in lower selectivity and cytotoxic activity. Preliminary clinical trials were deterred by anaphylactic responses caused by its non-covalently bound

20

protein component. In order to overcome this allergic response, NCS protein has been coupled to a maleic acid-based polymer (SMANCS) to render it immunologically inert. This complex is clinically employed in the treatment of hepatocellular carcinoma (*68*).

1.3.4.2 Calicheamicins

The calicheamicin family of enediynes are some of the most potent natural products ever discovered. It was discovered out of a rock collected at Waco, Texas, in 1987 by a touring scientist from Scripps Research Institute. It was rightly believed that the rock could harbour some unique bacteria or fungi. Cultures were grown from the rock and it was inside this culture the potent calicheamicin was first discovered.

An interesting segment of information on calicheamicin also validates its historical significance and extreme toxicity. Recent scientific analysis has postulated that the Greek warrior king, Alexander, the Great, could have been poisoned by calicheamicin, from the aquatic bacteria that were residents of the river Styx. This assumption is based on the fact that there are abundant populations of the bacteria, *Micromonospora* in river Styx. This theory, though open to contention, is built on the characteristic high toxicity of the calicheamicin producing bacterial species.

The calicheamicin family of enediyne antibiotics were isolated from the microorganism, *Micromonospora echinospora* spp. *calichensis*. They were identified in 1986 through a research program that identified microbial fermentation products active in the biochemical induction assay (*7-9*). They are profoundly potent sequence specific, double strand DNA-cleaving agents. Calicheamicin γ_1^I is the most potent member of this class of antibiotics. They exhibit extreme

potency against Gram-positive bacteria. They are also active against Gram-negative bacteria. They exhibit extraordinary activity against murine tumours, such as leukemia and solid neoplasms (*69*).

The chemical structure of calicheamicin γ_1^I (**Fig. 1.12**) can be studied as two distinct structural components, containing three essential moieties. The larger component consists of an oligosaccharide sugar part made up of four monosaccharide units and one hexa-substituted benzene ring. These sugar units are connected together through a highly unique series of glycosidic, thioester and hydroxylamine linkages. The second structural component is the aglycone also called as calicheamicinone. It has a highly functionalized bicyclic core that contains within itself, a strained enediyne unit contained within a bridging 10-membered ring. It also contains a trisulfide moiety that acts as a triggering device.

Figure 1.12 The chemical structure of calicheamicin γ_1^I
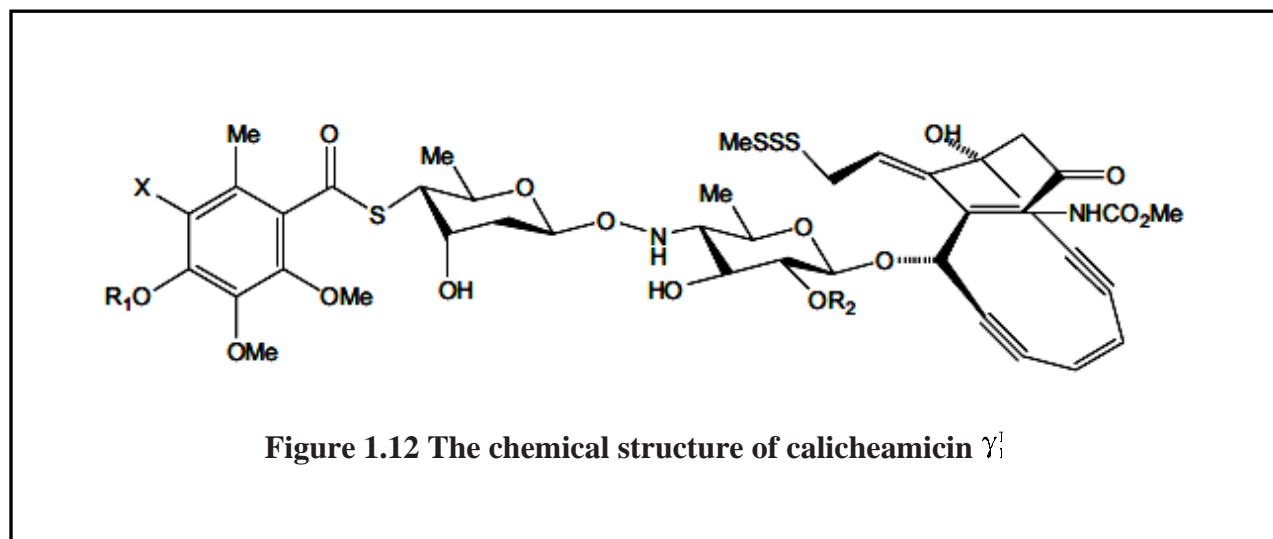
The aryltetrasaccharide component is involved in the delivery of the drug to its target and binds tightly into the minor groove of double-helical DNA. It displays a high specificity for sequences 5'-TCCT-3' and 5'-TTTT-3' through significant hydrophobic interactions (*70, 71*). The binding

of the aryltetrasaccharide component to target sequences is facilitated by transformation of the oligosaccharide component into an inflexible, extended conformation. The distinctive torsion angles provided by this transformation in the central region of the molecule, imparts the necessary complementarity to bind to the minor groove of the DNA. The DNA molecule seems to distort itself upon binding, to accommodate the drug binding. Apart from the sequence recognition by an aglycone and aryl oligosaccharide part, large and polarisable iodine in the hexa-substituted benzene ring also plays a part in the recognition process. The aglycone moiety is a rigid bicyclic core, a 10-membered ring containing the enediyne moiety awaiting activation to undergo cycloaromatization. It also contains an allylic trisulfide, which serves as a trigger. With the arrival of the drug into the vicinity of the DNA, a nucleophile attacks the central sulphur atom in the allylic trisulfide. This results in the formation of a thiol, that is added intramolecularly to the neighbouring *a, b*-unsaturated ketone incorporated within the bicyclic core (**Fig. 1.5**). This conformational change of the aglycone leads to considerable strain of the 10-membered enediyne ring. Subsequently Bergman cycloaromatization occurs, forming a highly reactive 1, 4-benzenoid diradical. The hydrogen atoms from both DNA strands at C 5' position of the cytidine and at C 4' of the three base pairs removed on the 3' side of the complementary strand are abstracted by the calicheamicin diradical. This results in the double stranded DNA cleavage.

Calicheamicins occurring in nature are a thousand times more potent than conventional cytotoxic chemotherapeutics. The most potent member of the calicheamicin family, calicheamicin γ_1^I coupled with antibodies to tumour-specific antigens, has been proposed as a targeted therapy for breast cancer (*72*). Gentuzumab ozogamicin, the first clinically validated cytotoxic immunoconjugate, has a humanized anti--CD33 antibody linked to a calicheamicin derivative. It

has been suggested for the treatment of elderly patients with relapsed acute myeloid leukaemia. Another conjugate from the calicheamicin family, inotuzumab ozogamicin, is presently in Phase I clinical trials in patients with non-Hodgkin's lymphoma. Numerous tumour-targeted immunoconjugates of calicheamicin are employed in preclinical research.

1.3.5 Fabricated DNA-cleavers – Modifications of enediynes

Clinical usage of enediyne natural products, despite the remarkable antitumor activity, was initially hampered by the lack of tumor cell specificity and delayed cytotoxicity. Enediyne antibiotics exhibit excellent anticancer and antitumor activities and are extremely potent DNA cleavers. But their high toxicity hinders their widespread use to treat cancers and tumours. Hence, there is an urgent requirement for new enediyne-related structures that express maximum activity towards tumour cells and minimum toxicity towards normal cells. The structural complexity of natural enediynes also necessitates the advent of synthetically more accessible analogs. There are two main essentialities of efficient artificial DNA cleavers, one being the molecular recognition measured in selectivity and the second, DNA scission scaled in activity (*73*). To elude the cell-specific problem, polymer-based and antibody-conjugated derivatives of enediyne natural products have been developed (*74-76*). Neocarzinostatin, conjugated with a polymer, was developed in 1994 for the treatment of hepatocellular carcinoma in Japan under the name of zinostatin stimalamer (SMANCS) (*76*). Calicheamicins, conjugated with antibodies, have also been developed to target tumor cells through specific antibody-antigen interaction. A conjugate that has been approved by the US FDA for the treatment of refractory acute myeloid leukemia is gemtuzumab ozogamicin (Mylotarg®), which consists of a chemically modified calicheamicin γ_1^I and a recombinant humanized IgG4 κ antibody.

Keeping in mind the three main structural and functional domains of enediyne compounds, modifications are directed toward, a) design of new enediyne "warheads" with better control over the cycloaromatization mechanism, b) selective recognition and target-directed delivery system, and c) triggering device sensitive to chemical or photo-induced activation (*76*). The first generation of designed enediynes (**I** in **Fig. 1.13**) was synthesized by Nicolaou and co-workers (*33, 34*) in order to define the simplest model still capable to cleave the DNA molecule. It was revealed that the derivative with a 10-membered ring ($n = 2$) undergoes the Bergman cyclization at room temperature while those with larger rings ($n = 3–8$) were stable. However, Nicolaou and co-workers were not able to synthesize a derivative with a 9-membered ring ($n = 1$). Subsequent to these findings, the water-soluble derivative **II** (**Fig. 1.13**) was prepared and it was found that it undergoes Bergman cycloaromatization at 37 °C and cleaves double-stranded DNA.
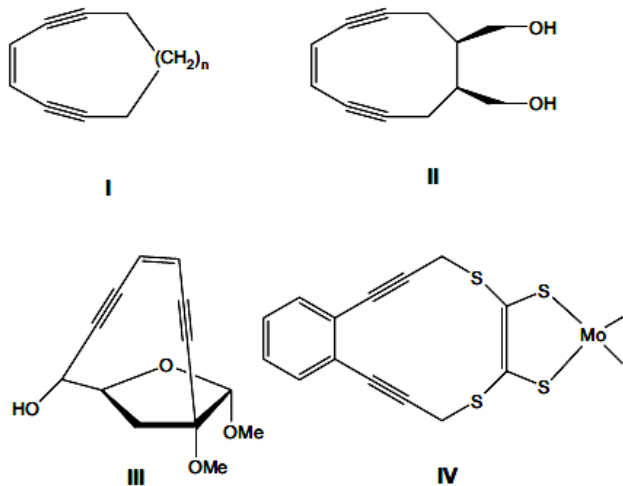
Figure 1.13 Examples of some fabricated enediynes (Adapted from Matija, 2007)

Almost at the same time, Maier and co-workers prepared the bicyclic derivative **III** (**Fig. 1.13**), that was stable at room temperature. But at elevated temperatures there was cyclization. It was

discovered that metal ions facilitate thermally induced Bergman cyclization of stable enediynes through coordination or complexation. Hence the molybdenum derivative of enediyne **IV** (**Fig. 1.13**) was far more .susceptible to cyclization than the free ligand (*77*).

For proper delivery to the site of action, the enediynes were attached to numerous molecules known to interact with the DNA. These hybrid molecules composed of simple cyclic arene enediyne and a synthetic minor-groove binder structure were found to cleave single-stranded DNA at 40 °C (*78*). When a simple cyclic enediyne was coupled with an analogue of netropsin, a minor-groove DNA binder, a hybrid molecule was obtained which exhibited a 2000-fold increased DNA cleaving efficiency compared to an enediyne alone (*79*).

Under controlled conditions, Bergman cycloaromatization is of great utility. Along the lines of this phenomenon, extensive search for adequate triggering mechanisms is in progress. In the presence of an acid catalyst, a simple dynemicin model compound **V** (**Fig. 1.14**) gave a product of Bergman cyclization upon epoxide opening. Under neutral conditions Dynemicin model **VI** (**Fig. 1.14**) can be activated photochemically, while the presence of a base triggers cycloaromatization of **VII** (**Fig. 1.14**) (*26*).

Figure 1.14 Dynemicin models with different triggers (Adapted from Matija, 2007)
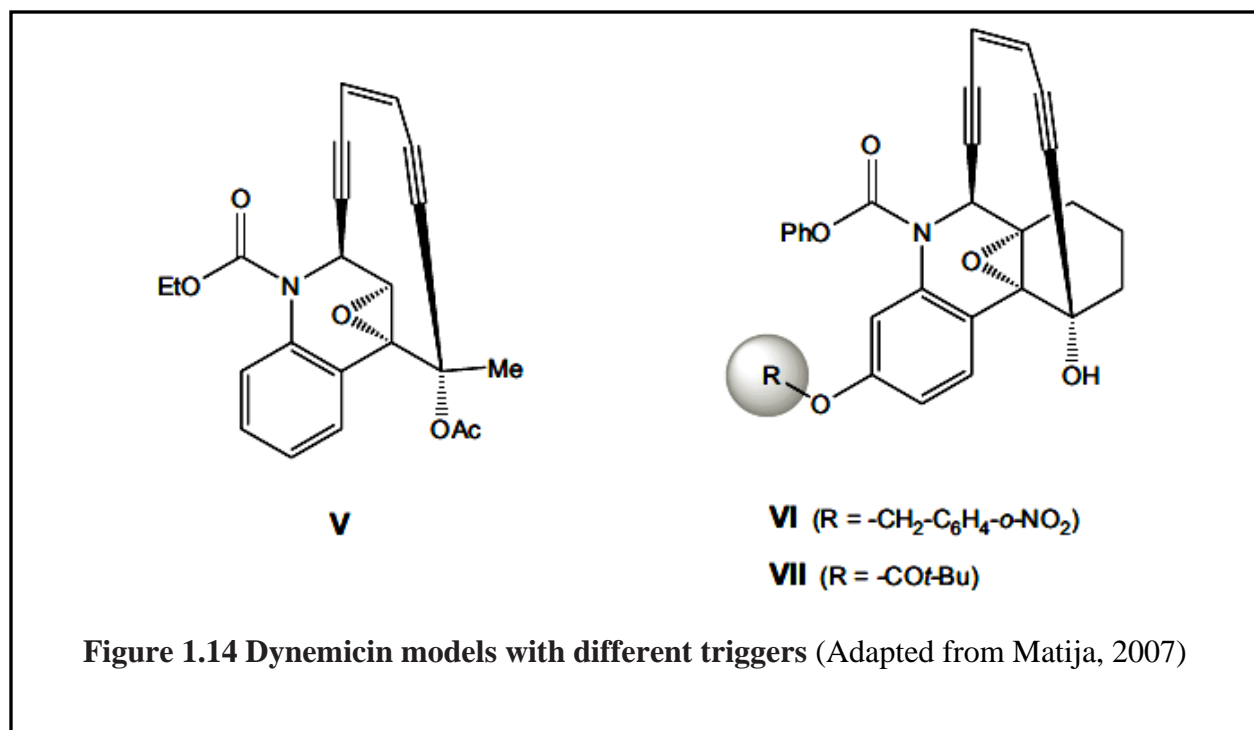
Considerable effort has been invested in the recent years into understanding the photochemical aspect of Bergman cycloaromatization and hence into the development of enediyne compounds equipped with suitable photo-labile groups. When irradiated with visible or UV light without metal ions or reducing agents, these compounds are expected to form reactive diradicals capable of DNA cleavage.

1.4 Polyketide synthases

Polyketide synthases (PKSs) are a family of enzymes from prokaryotic and eukaryotic systems that are involved in the synthesis of polyketides. They catalyze the fusion of short carbon chain monomers into long polymers through successive rounds of Claisen condensation reactions. PKSs are structurally and functionally related to fatty acid synthases. Both these enzyme classes catalyze the condensation of activated primary metabolites, acetyl-CoA and malonyl-CoA, to form β-ketoacetyl polymers linked to the enzyme by thioester bonds.

The difference between fatty acid synthesis and polyketide synthesis lies predominantly in the steps following condensation such as reduction and dehydration. In the synthesis of fatty acids, the condensation step is followed by β -ketoreduction, dehydration and enoyl reduction to yield the final completely reduced, saturated fatty acid. The synthesis of polyketides involves a controlled partial or complete omission of these reduction steps resulting in highly diverse polyketide chains with respect to the occurrence of β -ketone, β -hydroxyl and alkyl groups.

There are several different classes of PKSs, differing in their domain architecture and mode of synthesis involved in catalyzing different reactions. They are classified based on the mode of synthesis as linear or iterative or the number of subunits as single or multiple (**Table 1.3**).

Table 1.3 The differences between the three types of PKSs. The differences between the three types of PKSs with respect to structure, synthesis mechanism, evolutionary relation and distribution (*80*).

Type	Protein Architecture	Homology	Source organism	Mechanism of synthesis
Type I (modular)	Single protein with multiple modules.	NA	Bacteria	Linear (assembly-line style) in which each active site is used only once.
Type I (iterative)	Single protein with one module.	Vertebrate FAS	Fungi	Iterative, in which the active sites are reused repeatedly.
Type II	Multiple proteins, each with a single mono-functional active site.	Bacterial FAS	Bacteria	Iterative, in which active sites may be used only once or repeatedly.
Type III	Single protein with multiple modules	NA	Plants and Bacteria	Iterative, in which the active sites are reused repeatedly.

The protein domains that are in charge of adding the single ketide unit to the growing polyketide and responsible for the modifications, are denoted as modules. Almost eleven different such catalytic domains are generally found in PKSs (**Table 1.4**). The simplest form of a functional PKS comprises of a KS, an AT, an ACP and a TE domain.

Table 1.4 The different types of domains found in PKSs. The eleven different domains can be divided into three groups based on which part of the synthesis they participate in (C - condensation reaction, R - reduction of β-ketone and M - other post-condensation modifications). (*80*).

Domain	Function and Type of reaction catalyzed
Starter Acyltransferase (SAT)	Loading of starter units (C)
Acyltransferase (AT)	Loading of starter, extender and intermediate acyl units. (C)
Acyl Carrier protein (ACP)	Holds the growing polyketide chain as a thiol ester (KS-S-polyketide). (C)
β-ketoacyl synthase (KS)	Condensation reaction between starter/intermediate and extender units. (C)
β -keto reductase (KR)	Reduces β -ketone groups to hydroxyl groups. (R)
Dehydratase (DH)	Reduces hydroxyl groups to enoyl groups (unsaturated). (R)
Enoyl reductase (ER)	Reduces enoyl groups to alkyl groups (saturated). (R)
Thioesterase (TE)	Facilitates the release of the final product from the enzyme. (C)
Methyltransferase (MT)	Transfers methyl groups to the growing polyketide. (M)
Product template domain	Determines the folding pattern of the polyketide backbone in non-reducing iPKSs (C)
Claisen cyclase (CYC)	Facilitates ring formation by a Claisen-type cyclization reaction. (M)
Condensation (CON)	Facilitates the condensation of the synthesized polyketide with other polyketides. (M)

Some of the important domains of the Enediyne PKSs are explained in details in Section 1.5.

1.4.1 Type I modular PKS

The type I PKSs are further subdivided into modular and iterative PKSs. The type I modular PKS is the best characterized class of PKSs. However the functional information derived from these typically also apply to the other classes of PKSs. Though they catalyze similar reactions, there are several different classes of PKSs, differing in their mode of synthesis and domain architecture. Type I PKS are multizymes with a single polypeptide chain holding multiple different active sites capable of catalyzing different reactions. They possess all the necessary enzymatic domains for the formation of a polyketide within one enzyme complex. In type II PKS the catalytic domains required for the synthesis are located on individual proteins which interact to form a functional PKS enzyme complex. The type III PKS (chalcone synthase-like) differs from the type I and type II PKSs by not harboring intrinsic acyl carrier protein domains (*81*).

In the modular type I PKS, there are multiple copies of each active site domain, organized into modules. The modules are responsible for the addition and modification of a single ketide unit. The starter unit is loaded onto the enzyme at the N-terminal loading domain. The growing polyketide chain is then passed from module to module until it reaches the C-terminal end of the enzyme where its release is catalyzed by a TE domain. This mode of synthesis makes it possible to estimate the final polyketide length and the types of modifications the individual ketides units will harbor. The prediction is made possible by knowing the number of modules and deciphering the order of catalytic sites found along the PKS (*81*). As seen before, there are multiple copies of each active site domain in the type I modular PKS that are organized into modules responsible for the addition and modification of a single ketide unit. The starter unit is loaded onto the

enzyme at the N-terminal loading domain. The growing polyketide chain is then passed from module to module until it reaches the C-terminal end of the enzyme where its release is catalyzed by a TE domain. This mode of synthesis makes it possible to estimate the final polyketide length and the types of modifications the individual ketides units will harbor. The prediction is made possible by knowing the number of modules and deciphering the order of catalytic sites found along the PKS (*81*). The final number of ketide units in polyketides synthesized by type I modular PKS is equal to the number of modules that make the PKS. This is a result of the linear synthesis mode adopted by these PKSs, where the growing intermediate is passed along the PKS from module to module (**Fig. 1.15**). The release of the final polyketide is mediated by the TE domain.

Type I modular PKSs are predominantly found in bacteria and are responsible for synthesis of clinical and economical important macrolide polyketides, such as the erythromycin A and rifamycin.

Figure 1.15 The modular structure of the three Type I (modular) DEBS PKSs. The modular structure of the three Type I (modular) DEBS PKSs responsible for the synthesis of erythromycin, with the growing polyketide chain shown, the newest ketide group in each step is highlighted with red. The thick lines at the top denote the extent of the individual module; note that erythromycin is synthesized by three separate enzymes consisting of two modules each. LD = loading module, M1-M6 modules and TE = Thioesterase domain.

1.4.2 Iterative type I PKS

Iterative type I PKSs are megasynthases that use a single set of functional domains to assemble complex metabolites. The iPKS harbors only a single copy of each catalytic domain. But these individual domains can be deployed repeatedly during synthesis of a single polyketide molecule (**Fig. 1.16**). Type I iterative PKSs are further subdivided based on the modifications they introduce into the growing polyketide chain during synthesis. Nevertheless, the activities of the core domains remain similar in all subclasses. These are mostly found in fungi and bacteria

and their functioning mechanism is similar to that of the fatty acid synthases. The metabolites produced by the PKSs are compounds with a wide range of biological activities. The fungal polyketides are one of the first classes of secondary metabolites to be studied extensively in biosynthesis. However the understanding of the related enzyme mechanism is still in preliminary stages.
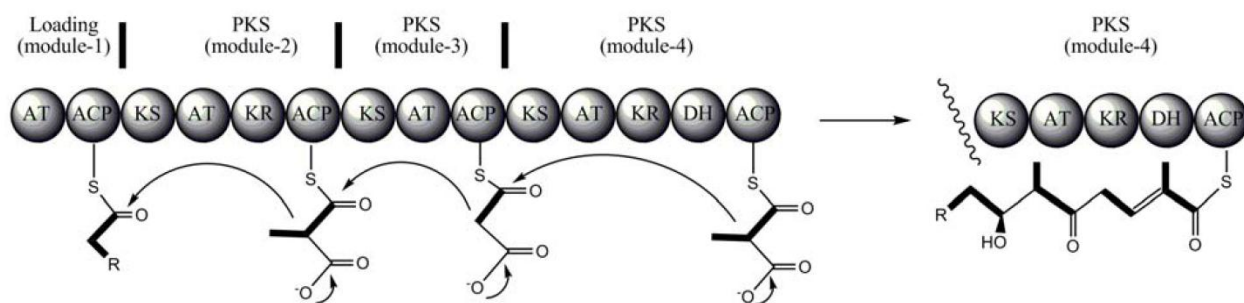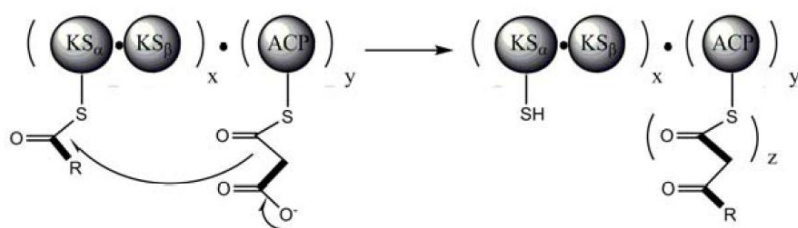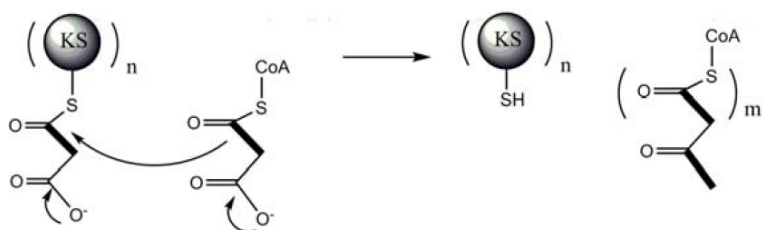
Figure 1.16 A comparative scheme of biosynthesis of the three types of PKS. (a) Type I PKS which is modular and non-iterative; (b) Type II iterative PKS; (c) Type III PKS which is iterative and ACP-independent.

Based on their modifying capabilities and the related domains, fungal iterative PKSs have been classified into three - non-reducing iPKS (NR iPKS), partial reducing iPKS (PR iPKS) and highly reducing iPKS (HR iPKS). The NR iPKS possess only the core set of iPKS domains (AT, KS, ACP and TE/CYC) apart from the MAT and PT domains. The PR iPKS contains a KR domain and possibly also a DH, in addition to the core set of domains. These domains together synthesize products with hydroxyl and enoyl groups. The HR iPKS has all the domains needed to reduce the β-ketone group to an alkyl group. This suggests that they have all the three KR, DH and ER domains in addition to the core domains also. The presence of a modifying domain does not necessarily mean that it is employed in all the iterations of synthesis. Apart from these, there are also systems where two iPKSs interact to form a common product as seen in zearalenone synthesis, or where an iPKS and a NRPS have been fused to form a single enzyme as exemplified by fusarin C synthesis, or systems where iPKS products are modified by NRPS like enzymes as seen in fumonisin synthesis.

1.4.3 Enediyne PKS

Most of the known bacterial PKSs belong to either type I modular PKSs or type II PKSs. In contrast, the enediyne PKSs from the biosynthetic pathways of calicheamicin γ_1^I and C-1027 resemble a family of fungal iterative type I PKSs that use a single set of catalytic domains for polyketide synthesis (*82*). The presence of enediyne PKSs has redefined the boundaries for PKS classification. The novel enediyne PKSs reveal structures and functions well beyond the paradigm of type I, type II and type III PKSs. Surprisingly, the enediyne PKS is more like a hybrid between the type I and type II PKSs. Though it acts iteratively, it produces aliphatic

polyketides. With these novel features, it provides us a revised perception on bacterial PKSs and overall molecular diversity and biosynthetic complexity.

The enediyne PKSs encompass several domains that were identified as KS, AT, KR and DH domains based on sequence homology (*58, 59*). An acyl carrier protein (ACP) domain and a phosphopantetheinyl transferase (PPTase) domain were also postulated initially despite low sequence homology shared with other ACP and PPTase proteins. The significant sequence identity and similarity shared by CalE8 and SgcE, the two PKSs for calicheamicin and C-1027 synthesis, led to the postulation that the 9- and 10-membered enediyne cores may share a common PKS progenitor (*59*).

1.5 Domains within PKS

1.5.1 Acyl carrier protein

A defining characteristic of fatty acid, non-ribosomal peptide and polyketide biosynthetic pathways is the utilization of the ACP or peptidyl carrier protein (PCP) for the shuttling of biosynthetic intermediates between various catalytic domains or proteins (*83-86*). ACPs are some of the most abundant proteins found in the cell. These are essential proteins that play significant roles in various metabolic pathways involving acyl transfer. The small and highly dynamic ACPs or PCPs are either free-standing proteins or integrated domains within a complex multidomain FAS, PKS or nonribosomal peptide synthase (NRPS). ACPs were initially discovered as small proteins aiding in fatty acid synthesis (*87*). Type I FASs are multidomain proteins that exist as dimers with the growing fatty acid chain linked to the phosphopantetheinyl arm of the ACP. These systems are found in metazoans and some microorganisms. In these

systems, ACP occurs as a distinct domain within the large polypeptide comprising the modules necessary for fatty acid synthesis.

ACPs also exist as small discrete soluble proteins in the type II systems predominantly found in bacteria, mitochondria, plastids and apicomplexan protozoans (*88*). The free-standing ACPs from type II FAS pathways are the best studied ACP systems. The fatty acid intermediates are tethered to the acidic CPs and are shuttled to the individual component enzymes. One of the classic examples of this type of proteins is the *E. coli* ACP, whose structure has been solved in the butyrylated form. Their crystallographic and solution NMR studies reveals that the canonical four-helix bundle fold with a binding pocket that sequesters the growing fatty acid chain (*89, 90*). 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 (*91-93*). Apart from the two FAS systems, they were also found to be associated with twelve different proteins involved in the primary metabolic pathway of synthesis of the fatty acid (*94*). Later it was discovered that these CPs also played a significant role in many other primary and secondary metabolic pathways that include biosynthesis of phospholipids, glycolipids, endotoxins, metabolic cofactors and signaling molecules. They are also indispensable in the production of polyketides and non-ribosomal peptides. The CP existing within the 155kDa Lys2 polypeptide, modified by the PPTase like Lys5 enzyme is involved in fungal lysine biosynthesis (*95*). The CP in β-alanyl-dopamine producing multimodular protein, Ebony and Dcp carrier protein involved in D-alanine incorporation system in *B. subtilis* fall under the peptide carrier proteins (PCPs). ACPs are also found in the modular complexes of type I PKSs and type II PKSs. (*96*)

Recent investigations 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 (97-*102*). Subtle differences between the ACPs in local structure, surface electrostatic potentials and binding mode of biosynthetic intermediates have also been documented. Although the sequence similarity between the carrier proteins is not high, they all possess a conserved all α-helix tertiary structural pattern (**Fig. 1.17**). They contain four α-helices harboring a conserved serine residue at the beginning of helix II.
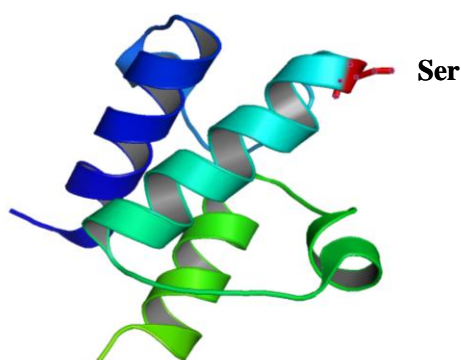
Figure 1.17 Structure of *E. coli* ACP. The 4 - α-helical tertiary structure of *E. coli* ACP (PDB: 2fae), a discrete soluble ACP found in the type II systems. Helix I is colored dark blue, helix II or the "recognition helix" is colored light blue with the conserved serine residue highlighted in red. The smaller helix and C-terminal helix are colored in green.

The growing chains in fatty acid synthesis and secondary metabolite synthesis are attached as thioesters to the serine residue of the ACPs. Most of the CPs have a signature GX (H/D)S(L/I) motif harboring the Ser residue for Ppant attachment. The ACPs share a significant structural homology with a common three-helix bundle with a long loop connecting helix I and II. The

three helices enclose a structural moiety rich in hydrophobic residues. This hydrophobic binding pocket aids in positioning and binding the fatty acyl chain which is attached through a thioester linkage to the 4'-phosphopantetheiene prosthetic group at serine-36 of *holo*-ACP (**Fig. 1.17**). The helix II of CPs like *E. coli* ACP and *V. harveyi* ACP contains a large proportion of acidic residues at the ends. These acidic residues stabilize the binding to divalent cations like Mg^{2+} and thereby stabilize the entire structure. The modification of CPs with the flexible phosphopantetheinyl group acts as a nucleophile to activate acyl groups by a thioester linkage to the terminal thiol and also provides a flexible arm (~20Å) for shuttling of the acyl intermediates between proteins or neighboring domains. The modified *holo*-carrier proteins function as essential components of various enzymatic biosynthesis. 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 (*86*).

1.5.2 Phosphopantetheinyl transferase

The phosphopantetheinylation on the conserved serine residue of carrier proteins is a post translational modification catalyzed by a family of enzymes termed the phosphopantetheinyl transferases. PPTases convert the carrier proteins from its inactive *apo* form to the functionally active *holo* form. The conversion process comprises a nucleophilic attack of the hydroxyl side chain of the highly conserved serine residue of the carrier protein on the 5'-β-phosphate of CoA releasing 3', 5' – ADP in the process.

Based on their sequence homologies and substrate specificities, PPTases have been classified into three types. The 120 aa long AcpS type named after the *E. coli* AcpS. These are involved in modification of the ACP in fatty acid synthesis in almost all microorganisms. They also have

been shown to act on type II PKS systems. The second group of enzymes found associated mostly with the non-ribosomal peptide synthesis, are classified under Sfp like type of PPTases. These are 240 aa in length and exhibit broad substrate specificities ranging from PCPs of NRPS as well as ACPs of FAS and PKS. Human PPTase (329 aa) and Lys5 (309 aa), with extended polypeptide chains are also included in type II PPTase family (*95, 103, 104*). The genes that encode most of the PPTases of type I and II are found associated with the gene clusters encoding the multifunctional enzymes. For example, *mta4* is placed next to the gene encoding myxothiazol synthase, and *Sfp* is found 4kb downstream of the gene cluster encoding surfactin synthase. The third group of PPTases is those found integrated as definite domains within large multifunctional FAS2, as found in *Saccharomyces cerevisiae* (*105*).

More than one PPTase can be expressed in the same organism that adopts multiple Ppant dependent pathways. For example, *E. coli* expresses three PPTases - AcpS involved in Fatty acid synthesis, EntD involved in siderophore enterobactin synthesis and the gene product of yhhU, with unknown physiological function (*106, 107*). Although the PPTases expressed in the same organism exhibit overlapping substrate specificities, they are predominantly recruited to modify carrier specific CPs.

1.5.3 Ketosynthase (KS) domain

The KS domain is one of the core domains of the iterative PKSs, and is expected to catalyze the key decarboxylative Claisen reaction in the formation of the carbon-carbon bond between an acyl thioester and a malonyl thioester. It is also expected to be involved in determining the chain length of the growing polyketide. The chain length of the synthesized polyketides and the substrate specificity exhibited by the PKSs play an important role in the

commercial and medical significance of the polyketide family of products. While the KS domain is known to influence the chain length, it is also believed to co-operate with other domains in determining the substrate specificity.

A comparative analysis of various known crystal structures of KS has revealed that all KS crystal structures from FASs and PKSs exhibit a similar thiolase fold (*108*). This thiolase fold is represented by two copies of α-β-α-β-α folds forming a five-layered core arranged in a 2α-5β-2α-5β-2α pattern with three layers of α-helices interspersed by two layers of β-sheet and held together with extensive connecting loops. There also exists a pseudo twofold axis parallel to the dimer axis between Nα3 and Cα3. The extensive dimer interface and the location of the active-site residues are conserved among the various KS subfamilies. The active-site cysteine which is involved in the covalent attachment of substrates and intermediates is also absolutely conserved. Nevertheless, there are major structural differences among the subfamilies with regards to the extent and structure of the loops on either side of the substrate-binding pocket. The arrangement of the loops affects the position and identity of key catalytic residues and the different substrate chain-length specificities for CoA-linked or ACP-linked thioesters, apart from the universally conserved Cys (*108*). In the modular PKSs and other subfamilies, the KSs are dimers with their extensive dimer interfaces stabilized by pairs of hydrogen-bonded, antiparallel β –strands. These antiparallel β –strands generate a 14-stranded β -sheet spanning both the monomers. In the DEBS KS5 structure, there is a long helix at the N-terminal that could probably facilitate KS dimerization and also to serve as a docking point for the upstream C-terminal docking domain of the previous module 4 (*109*). KS deletion experiments in the type I FAS and PKS demonstrate that the dimeric nature of the KS domains is integral in facilitating the dimerization of the full megasynthase (*110, 111*).

Structural analysis have shown that the active-site, Cys-His-His triad, and the oxyanion well of DEBS KS3 and KS5, act KS/CLF and porcine FAS KS domains can be super-imposed perfectly. This could probably suggest that these structurally homologous domains may adopt a similar catalytic mechanism (*111, 112*) initiated by the docking of acyl-ACP to an electropositive patch of KS. This proposed mechanism in these extending KSs is similar to those observed in the priming KS of ZhuH, except that the catalytic triad of a ZhuH KS consists of His-Asn-Cys. Moreover, in ZhuH it is acyl-CoA, rather than acyl-ACP, that binds first to ZhuH. The PPT-binding region stretches from the enzyme surface to the active-site Cys and this region is relatively well conserved, reflecting its universal role in binding the PPT moiety though the acyl binding region varies significantly. In FAS KS domains, the binding pockets are hydrophobic and hence they are expected to promote fatty acyl binding (*113-116*). The FAS KS domains have a hydrophobic, narrow pocket of a suitable size to specifically accommodate their corresponding fatty acyl substrates (*113-116*). On the contrary, the acyl-binding pockets of PKS KS domains (such as DEBS KS3 and KS5, act KS/CLF and ZhuH) are much wider, amphipathic and hence allow hydrogen-bonding interactions with the carbonyl groups of the growing polyketide chain (*109, 117-119*). This is in perfect consistence with the substrate tolerance reported for type I modular PKS KS domains. Conversely, both the priming and extending KSs of the type II PKS are more substrate-specific. This is clearly seen by the narrower acyl-binding pocket, similar to those in FAS KS domains. Mutations of four residues that define the bottom of the acyl pocket in actKS and tcmKS/CLF lead to a significant confirmation that the pocket size and shape indeed control polyketide-product chain length, with mutations in CLF being sufficient to alter chain length.

1.5.4 Acyltransferase (AT) domain

The AT domain is one of the core domains of the iterative PKSs. The role of the AT domains is to load the starter and the extender units to the ACP domain. The AT domains contribute to the transfer of the α-carboxylic nucleophilic extender units from corresponding acyl CoA molecules to the phosphopantetheine arms extending from the ACPs. Following this transfer step, the interaction between the acyl-KS and acyl-ACP occurs to form the C-C bond, which leads to the formation of a β-ketoacyl-ACP intermediate extended by two backbone carbon atoms. While the acyl-CoA specificity of AT domains has been extensively investigated, little information is available on the ACP specificity.

The proposed structure of the AT domain based on the resolved crystal structures of DEBS AT3 and AT5, *S. coelicolor* MAT, *E. coli* MAT (FabD) and the porcine FAS AT domain, is made up of two important components: the larger core subdomain, with an α/β hydrolase fold and the smaller subdomain, an 130-200 residue insertion with a ferredoxin fold. While the α/β hydrolase fold is made up of a parallel β-sheet flanked on two sides by α-helices, the ferredoxin fold is made up of a four-stranded antiparallel β-sheet capped by two helices. The active site made up of the Cys-His-His triad resides inside the cleft that is formed between the two subdomains. In DEBS AT3 and AT5, the longer C-terminal helix (residues 857 to 867) signifies the structural difference between the type I modular PKS AT domains and other MAT domains. This structural moiety may be important for protein–protein interactions between AT and the KS-AT linker (*109, 119*). The substrate binding pocket of all the proposed AT domains are hydrophobic in nature. This suggests that the substrate-binding pocket would most likely discourage water binding in the cleft. The AT domains of type I modular PKSs, exhibits promiscuity and can accept acetyl-, propionyl-, isopropionyl-, isobutyryl-, crotonyl-, phenylacetyl-, hydroxybutyryl-,

and isopentyl-CoA both *in vivo* and *in vitro* (*120-124*). On the contrary, the AT domains in the extender modules such as DEBS AT modules, are highly specific toward (2S)-methylmalonyl-CoA and the (2S)-methylmalonyl thioester of N-acetylcysteamine (NAC) (*122-125*). Hence it is clear that the extending AT in DEBS serves as an important gatekeeper in macrolide biosynthesis (*126*). Extensive biochemical and structural work on these domains have identified four motifs to explain the observed AT specificity: (1) the ''RVDVVQ'' motif that resides 30 residues upstream of the active-site Ser (*127, 128*); (2) the GHSXG motif around the catalytic Ser (*127, 128*); (3) the YASH motif that resides 100 residues downstream of the active-site Ser (*127, 129*), which based on a systematic mutational analysis of DEBS AT domains is the dominant substrate specificity motif in type I modular PKSs among motifs 1, 2, and 3 (*129*); and (4) the C-terminal region that has been shown to be important for substrate specificity from domain swapping experiments (*122*). The structure of the porcine FAS elucidates that the hydrophobic cavity formed by F682 (part of Motif 1) and F553 (part of Motif 3) could allow M499 to flip in and out thereby allowing both methylmalonyl and malonyl moieties. The preference of malonyl-CoA as a substrate against methylmalonyl-CoA (or methylmalonyl- vs. propionyl-CoA) was thought to be influenced by a limited number of residues. However from the structural studies, it has been proposed to be the result of combinatorial interaction of different structural elements throughout the entire protein fold.

1.5.5 Ketoreductase (KR) domain

The role of the KR domain in polyketide biosynthesis is the NADPH assisted reduction of the β-keto group to a hydroxyl group on the growing polyketide. Amongst the KR domains from various families, the crystal structures of type I modular PKSs (DEBS KR) (*130*) and KR1

of tylosin PKS (tylKR) (*131*) and for type II PKS, the crystal structure of actinorhodin KR (actKR) (*99, 132, 133*) have been solved and reported. The actKR and DEBS KR and tylKR differ in terms of their substrate. Still, the ketoreductions catalyzed by all the three KR domains are chemically identical to the corresponding fatty acid ketoreduction, differing only in regio-specificities (*134*). The type I KR domain in FASs and PKSs is made up of catalytic subdomain and a truncated, non-catalytic structural subdomain. Both these domains exhibit the same fold. While DEBS KR and tylKR are monomeric in solution (*130, 131*), the type II KR exists as a tetramer (*99, 132*), with each monomer consisting of a single domain. The structures of all the KR domains have a global short-chain dehydrogenase/reductase (SDR) fold made up of a highly conserved Rossman fold and the core region. The Rossman fold is made up of two right-handed α-β-α-β-α motifs connected by α 3, while the core region consists of a seven-stranded β -sheet flanked by α -helices. The junction of the two α-β-α-β-α motifs forms a highly conserved characteristic groove of the Rossman fold and binds the cofactor NADPH (*135*). The large cleft formed by helices α6- α7 and the loops between α4 and α5 harbours the polyketide substrate-binding pocket. The typical SDR motifs, such as the TGxxxGxG motif (residues 2 to 19), the D63 and NNAG motifs (residues 89 to 92), the active-site tetrad Asn–Ser–Lys–Tyr, and the PG motif (residues187 to 188) (*135*) are clearly seen in the catalytic subdomains of DEBS KR, tylKR and actKR. The type I and II PKS KRs exhibit different substrate specificities and this could be caused by the presence of a long insertion between helices 6 and 7 for act KR. In DEBS KR and tylKR, the cofactor-binding motif and the substrate-binding portion are absent in the structural subdomains thereby resulting in an inactive structural subdomain (*130, 131*). The active-site tetrads of DEBS KR and tylKR are made up of K1776–S1800–Y1813–N1817 whereas in actKR there is a swap in the positions of the Asn and Lys residues, N114–S144–

Y157–K161 (*130, 131*). These residues are positioned close to the nicotinamide ring of NADPH hydrogen bonds being formed by Tyr and Lys with the NADPH ribose and nicotinamide ring. It has been showed that type II PKS KRs exhibit substrate specificities from that of modular type I PKS KRs (*136*), both of which can reduce linear and monocyclic ketones and FAS (*137, 138*). The actKR, a type II KR, shows a strong preference towards bicyclic polyketides leading to a suggestion that its natural substrate could be a cyclic polyketide. Studies on actKR have shown that they elicit a highly specific C9-reduction. This is brought about by the cyclization of the first ring of the substrate before ketoreduction, thereby sterically constraining the ketoreduction (*133*). The determinants of stereospecificity for the modular PKS KRs were the proposed ''LDD'' and PxxxN motifs (*139, 140*). The presence or absence of these motifs influences the formation of 3R or 3S stereomer, respectively. The DEBS KR crystal structure has that the 93–95 ''LDD'' motif lying in a loop next to the active site (*130*).

1.5.6 Dehydratase (DH) domain

The DH domain is the central domain in mammalian FAS and most PKSs. Though the existence of this domain was discovered long ago and has been studied for many years, very little information is available on the structure and mechanism adopted by this domain within these large proteins. Only recently, few aspects on the structure and mechanism of DH have been reported based on structural and mutational studies carried out on animal FAS (*141*). Recent studies have discovered that the DH domain adopts a double-hotdog fold structure. Analysis of this structure in conjunction with its proposed function highlights a possible mechanism wherein the DH domain establishes a homodimeric contact across the 2-fold axis of the PKS or the FAS and interacts with the KR domain. The DH domain of fungal FAS adopts a triple hotdog fold and

catalyzes the elimination of water from hydroxyacyl-ACP. The pseudotrimeric fold resembles the eukaryotic hydratase 2 enzymes and is placed in between the pseudosymmetric core subunits. The catalytic histidine in fungal FAS DH, H1591 is located at the C-terminus at the mouth of a long hydrophobic tunnel running into the N-terminal subdomain (*85*). In the modular PKS, DEBS, the DH4 domain catalyzes dehydration of a 2R-methyl-3R-OH pentaketide to afford a *trans*-double bond (*142*). The crystal structure of the DEBS DH4 crystal structure, reveals that the DH domain in type I PKSs is made of two subdomains of limited sequence homology, that exhibit a ''hot-dog-in-a-bun'' fold (*143*). This double hot-dog (DHD) fold elicited by PKS DHs is also seen in type I FAS and resembles the dimeric type II bacterial DHs (such as E. coli FabA and FabZ (*144, 145*) that consist of one hot-dog fold per monomer. The long central helix of the fold, the hot-dog, is placed in front of a seven-stranded antiparallel b-sheet that forms the bun. The catalytic H44 of DEBS DH4 resides within the HXXXGXXXXP motif of the first subdomain (*138*). The conserved glycine in the sequence motif is responsible for the turn that enables van der Waals interactions between the residues H44 and P53. The catalytic D206 lies in the sequence motif DXXX(Q/H) found in the second subdomain and is involved in hydrogen-bonding with the side chain of Q210. Similar organization and catalytic-residue-interaction is also seen in the porcine FAS DH. Apart from these, both the ''GYXYGPXF'' and ''LPFXW'' motifs are also highly conserved in FAS DH. These sequence motifs are associated in defining the substrate pocket. In a proposed catalytic mechanism for DEBS DH4, the H44 residue acts as the active-site base for the deprotonation of the α-carbon, whilst the β-hydroxyl group may be polarized by the helix-1 dipole, thereby effecting the elimination of water.

1.6 Calicheamicin PKS (CalE8)

The sequencing of calicheamicin gene cluster has refined the conventional classification of PKSs with the identification of two iterative polyketide synthases. The two PKS genes, *CalE8* and *CalO5*, are involved 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. The discovery of more iterative type I PKSs serves to demonstrate 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 iterative type I PKSs has opened the doors to a still unexplored world of PKSs. The enormous versatility of PKSs developed during the course of evolution remains to be further investigated.

The PKS enzyme involved in the production of Calicheamicin (CalE8) is produced from a large gene segment 5.8 kb in size. The protein product of the *CalE8* gene is a large protein of molecular weight of approximately 200 kDa. The CalE8 protein is multidomain PKS, with the domains conforming to a unique arrangement to perform the stepwise biosynthesis of the polyketide. The domain architecture of iterative PKSs comprises of a KS, an AT, an ACP, a KR, a DH and a phosphopantetheinyl transferase domains **(Fig 1.18)**. The KS domain is believed to determine the chain length and catalyze the decarboxylation reaction. The acyltransferase domain determines the selectivity of the substrates to be attached on to the phosphopantetheinyl moiety on the acyl carrier protein. The ACP provides the site of attachment for the starter and extender molecules. The phosphopantetheinyl arm of ACP provides structural flexibility for the growing polyketide chain to shuttle between the different domains. The role of the KR domain in

48

polyketide biosynthesis is the reduction of the β-keto group to a hydroxyl group. The DH domain catalyzes the elimination of water from hydroxyacyl-ACP. The phosphopantetheinyl transferase domain is involved in modifying the acyl carrier protein by attaching a phosphopantetheinyl arm on ACP. This reaction converts the inactive *apo*-ACP to active *holo*-ACP and marks the initiation of the polyketide synthesis.

Figure 1.18 Proposed arrangement of the individual domains within CalE8
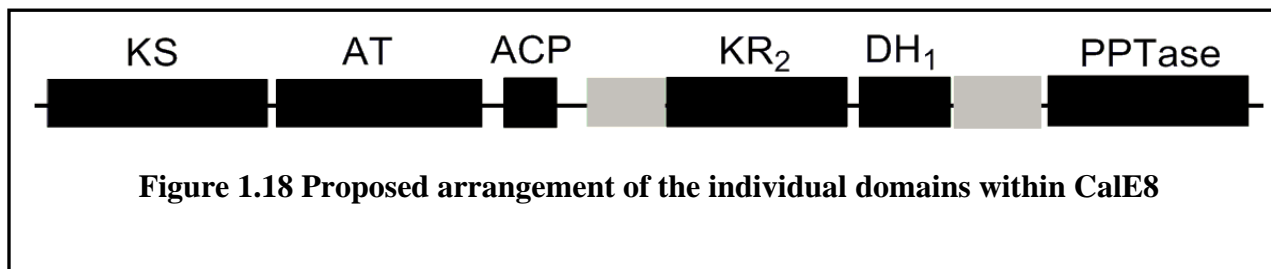
Among the six putative domains within the enediyne PKS, an intrinsic PPTase domain, instead of a stand-alone PPTase protein in other bacteria PKS system, is found to modify the ACP domain by phosphopantetheinylation. After modification, the phosphopantetheine moiety covalently linked to the ACP domain functions as a flexible arm that carries the extending chain and swings around to access the various enzymatic domains for chain processing. The AT helps to load the ACP domain with acyl units from either acetyl-CoA or malonyl-CoA pools. Playing a key role in chain elongation, KS domain catalyzes the condensation step to elongate the acyl chain, whereas the KR domain is likely to be responsible for the reduction of keto groups to the corresponding hydroxyl groups on the polyketide precursor. Subsequently, the 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 summarized scheme of one cycle of iterative polyketide synthesis is illustrated (**Fig. 1.19**). 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.
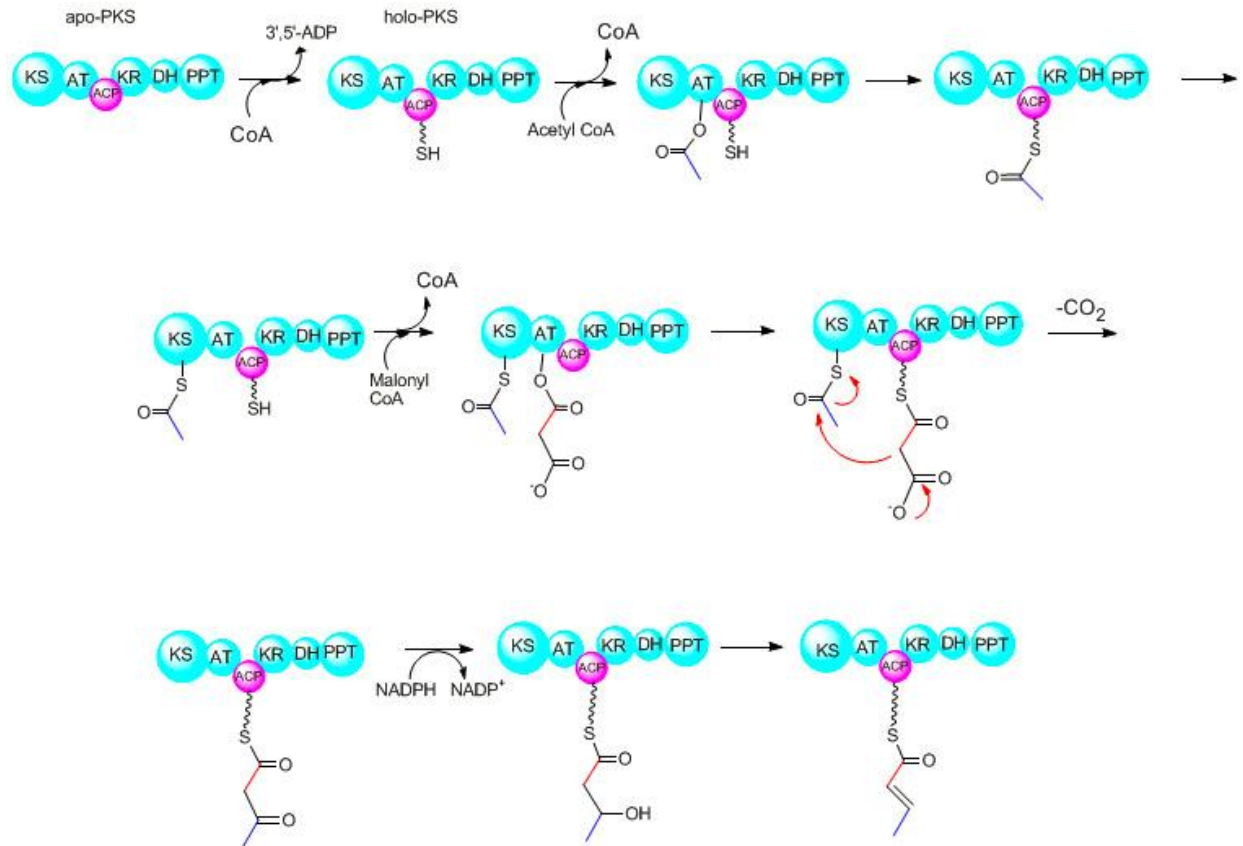
Figure 1.19 Proposed mechanism of iterative polyketide synthesis. 1) Phospho-pantetheinylation of ACP domain by PPTase using CoA. 2) Loading of acetyl-CoA starter onto ACP by AT domain via thioester linkage. 3) Transfer of the acyl starter to KS as the AT loads ACP with the extender, malonyl-CoA 4) Condensation between the starter and extender molecules mediated by KS 5). Reduction of subsequent keto groups to corresponding hydroxyl groups by KR simultaneously with the elongation, 6) Desaturation mediated by DH domain to form conjugated double-bond system 7) Termination of elongation process once the designated chain length is reached and the final product is released by external TEs.

1.7 Objective and Organization of the Thesis

CalE8 is a distinctive PKS with a potentially novel mode of biosynthesis. Our main objective is to identify and characterize the discrete domains within the multidomain CalE8. The whole protein will be examined through bioinformatics tools to define the boundaries of the individual domains. Each domain will be assigned a function based on the primary, secondary, tertiary and quaternary structures adopted by the folded polypeptide. This assignment of functions will be carried out *in silico* based on sequence and structural homology of existing and characterized proteins and domains. Attempts have been made in our lab to express the whole protein as a soluble pure protein for structural and functional studies. The individual domains will be expressed as stand-alone proteins for structural and functional characterization. Following the successful expression and purification of the individual domains, the predicted domain boundaries and assigned functionalities will be further validated. With the obtained structural and functional data, the full length CalE8 will be delineated into domains with defined structure and function.

The dissertation emphasizes on the structural and functional characterization of the domains of the enediyne PKS CalE8. Chapter I is an outline of the enediyne polyketides with special emphasis on the history, classification, modes of synthesis and significance. The chapter introduces the calicheamicin PKS, CalE8 and explains the organization of the thesis. Chapter II explains the characterization of the ACP and PPTase proteins that are the main players in the initiation step of the polyketide synthesis. The identification of these domains within CalE8 and validation by bioinformatics techniques, the cloning and expression of the domains as soluble proteins and the related challenges, and functional characterization of the individual domains have been discussed in detail in this chapter. The chapter also describes the cloning and

expression of CPs from representative PKS families and Sfp, a special PPTase with broad substrate specificity. Sfp was used to modify the carrier proteins and hence examine the functional integrity of ACP and other carrier proteins. The modification of the CPs by Sfp was also employed in studying the activity of the phosphodiesterase, *Pa*AcpH, as explained in Chapter III. We also attempted to study the activity of the AT domain of CalE8 on CPs modified using Sfp, as explained in Chapter IV.

Chapter III introduces a novel phosphodiesterase protein, *Pa*AcpH, identified from *P. aeruginosa* through bioinformatics analysis. This protein family was known to be capable of cleaving the phosphopantetheinyl moiety from a modified carrier protein. Only one protein from this rare non-essential phosphodiesterase family, the AcpH from *E. coli,* was characterized, but as an insoluble protein. The capability of the protein to modify the phosphopantetheinylated carrier protein could be extended to release the growing tethered intermediates from PKS for mechanistic studies. For the first time, a protein from this family was purified in the soluble form. The activity of the protein was examined against a family of homologous and heterologous carrier proteins. The ability of the protein to modify the ACP and its ability to release the products from PKS were examined. In Chapter IV we describe the cloning and expression of the KS, AT, KR and DH domains. The AT domain was expressed and purified as a soluble protein and structurally validated. Functional characterization of the Cal AT domain was tried against carrier proteins with acetyl CoA and malonyl CoA substrates. KS, KR and DH domains were examined through *in silico* techniques. Modeling studies for all the domains were performed in order to validate the discrete domains. The folding pattern and structural integrity were analyzed and the boundaries were assigned in accordance. Chapter V summarizes the findings and concludes the dissertation with emphasis on the future directions of the project.

Chapter 2. Cloning, expression, purification and functional characterization of the acyl carrier protein and phosphopantetheinyl transferase of CalE8

2.1 Introduction

The addition of a 4'-phosphopantetheine (Ppant) moiety to the acyl carrier protein (ACP) domain within the PKS by a phosphopantetheinyl transferase (PPTase) is the primary step in the polyketide synthesis (**Fig 2.1**). The PPTases attach a 4' Ppant prosthetic group from coenzyme A to the carrier proteins (CPs) in their inactive *apo* form to convert them to their active *holo* form. In the process of iterative biosynthesis, selected substrates are added onto the shuttling Ppant arm by a combined activity of the different domains of the PKS.

Figure 2.1 Phosphopantetheinylation of ACP by PPTase. Conversion of *apo*-ACP to *holo*-ACP is catalyzed by the PPTase in the presence of CoA and Mg^{2+}. The 4'Ppant prosthetic group is attached to the serine residue of the CP and 3', 5'- ADP is released.
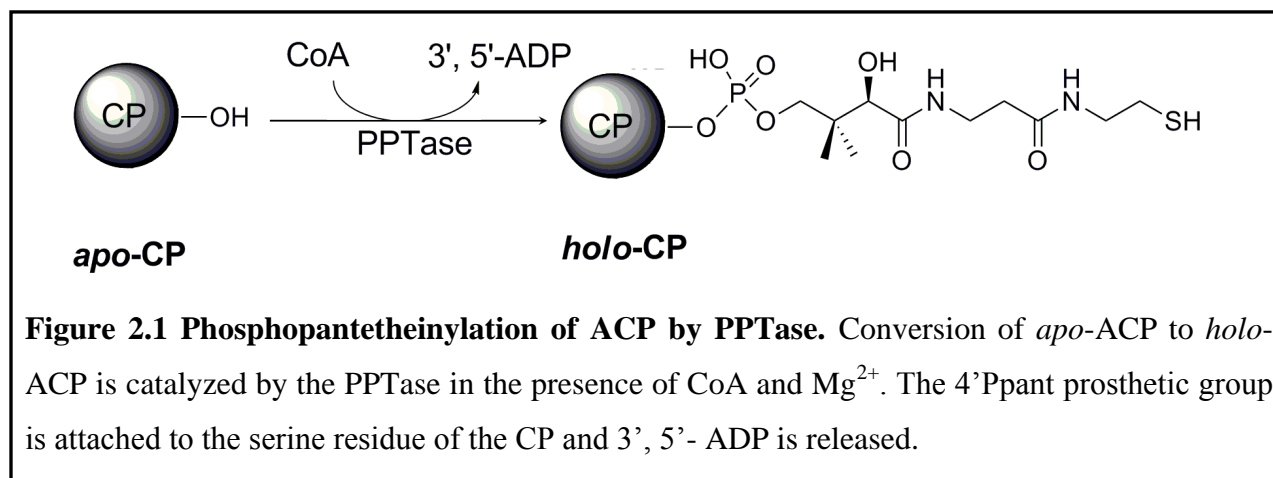
The finding that polyketide chain intermediates were released from CalE8 in solution (*146*) clearly demonstrates that an ACP domain within CalE8 has been modified with the 4'-Ppant shuttling arm to hold the growing intermediates and this modification has been effected by an inherent PPTase. The unique self-phosphopantetheinylation mechanism adopted by CalE8 led us to speculate that the PKS possesses inherent ACP and PPTase domains. The objective of this study is to identify the regions within CalE8 that code for these functional domains and examine them by employing cloning, expression, purification and functional characterization.

The region spanning residues 925 to 1016 within the CalE8 was identified to be a CP based on sequence analysis and comparison with known CP sequences. The ACP domain in CalE8 (**Fig. 2.2**) is an acidic protein spanning 92 residues with no significant sequence homology with other CPs and lacks a signature motif conserved in many CPs.

Figure 2.2 Domain boundaries of ACP within the proposed domain architecture of *CalE8* – The suggested domain boundaries defining ACP based on which the domain was cloned and expressed.
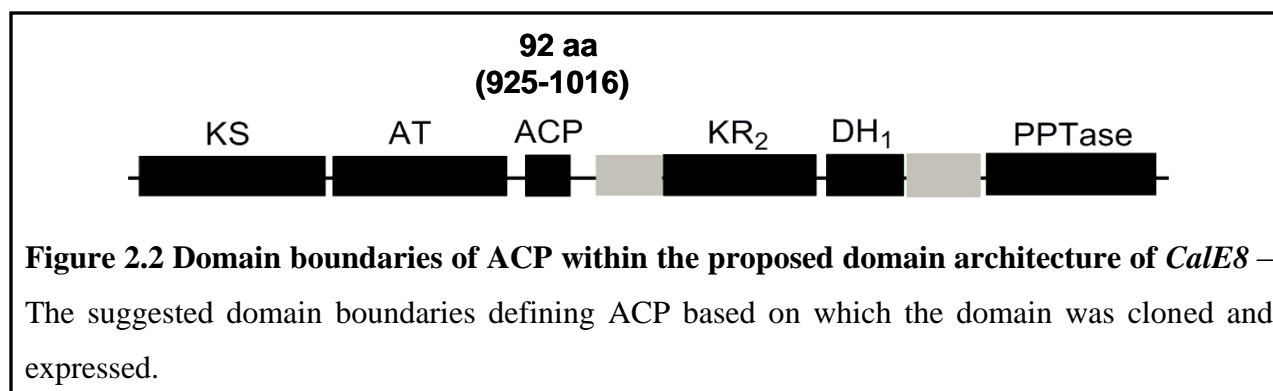
The domain boundaries suggested in literature (*60*) were revised based on secondary structural examination to generate a functional ACP. - The solution structure of the ACP protein has also been solved using NMR (*147*). Structural analysis of ACP has revealed interesting aspects on the probable functioning of ACP and hence it's phylogeny. The activity of ACP was confirmed by treating it with the surfactin PPTase (Sfp) from *B. subtilis*. Sfp exhibits broad substrate specificities and can modify PCPs from NRPS, ACPs of FASs and PKSs (*148, 149*). This attribute of Sfp has been employed in our study to modify the ACP and the other CPs for comparison against activity of the PPTase domain of CalE8.

The C-terminal region of the CalE8 spanning 212 residues has been identified to code for the PPTase. Two more constructs were cloned along with this region that included the whole and partially terminated regions of the 2nd domain of unknown function (**Fig. 2.3**).
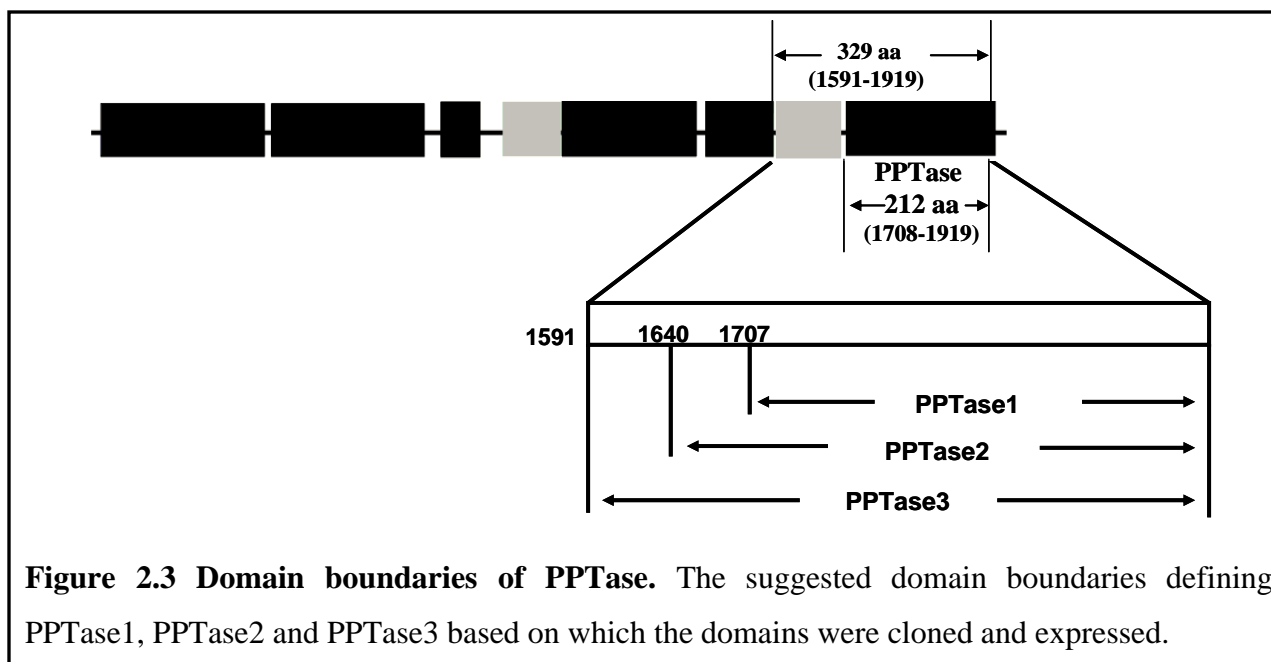
Figure 2.3 Domain boundaries of PPTase. The suggested domain boundaries defining PPTase1, PPTase2 and PPTase3 based on which the domains were cloned and expressed.

The different constructs were expressed and solubilities were examined. The activity of the PPTase was tested against a family of CPs from a NRPS (GPCP), PKS (DACP) and FAS (EACP). These CPs were expressed as soluble proteins. Their activities were tested using the broad spectrum Sfp PPTase. The ability of the PPTase domains to modify the *me*ACP was investigated. The functional capability of the PPTase domains was validated by treating with the CPs from different pathways.

This chapter reports in detail the cloning and expression of the ACP and PPTase domains of CalE8. In this study, we have characterized and validated the proposed ACP and PPTase domain regions within the CalE8. The similarities and dissimilarities at structural and functional levels, between the *me*ACP and representative CPs from other families have been discussed. The revised boundaries of the PPTase domain and its efficiency in modifying other CPs in comparison to *me*ACP have been investigated in detail and reported.

2.2 Materials and methods

2.2.1 Materials

Coenzyme A and other chemicals were purchased from Sigma. Restriction enzymes were purchased from New England BioLabs and Fermentas. Custom synthesized PKS, EACP, DACP and GPCP genes were obtained from GenScript Corporation. HPLC columns were bought from Agilent technologies.

2.2.2 CalE8 cloning, expression and purification

The polyketide synthase gene in pUC57 was digested using NdeI and XhoI restriction enzymes. pET28b+ vector was also simultaneously digested with NdeI and XhoI. The vector and insert were mixed in a ratio of 3:2 and ligated in a ligation mix that contained 2X rapid ligation buffer and T4 DNA ligase (Promega). It was transformed into NovaBlue competent cells (Novagen). Plasmid was extracted using Bio-Rad Plasmid mini-prep kit. The plasmid was validated by restriction and sequencing. It was transformed into BL21DE3 and plated in LB-Agar plates containing kanamycin. Colonies were picked and grown in LB media with kanamycin. The overnight culture was scaled up to 4-5litres and induced with 0.5 mM IPTG once the OD_{600nm} reached 0.6. The induced culture was grown at 16°C for 16-18 hours.

The culture grown overnight was centrifuged using a JA-10 rotor at 8000 rpm for 10 min. The pellet was washed with wash buffer (50 mM Tris, 500 mM NaCl, 5% glycerol, pH 8.0). The resuspended pellet was then sonicated at 40% amplitude, 5 seconds pulse with 2 seconds pause for 5-10 min and sheared using a microfluidizer. The mixture was centrifuged in a JA-25.5 rotor for 45 min at 20000 rpm. The clear lysate was mixed with Ni-NTA resin (GE Healthcare) and the resin-lysate mixture was incubated for an hour. It then was added onto a column and washed

with 20 times the column volume of wash buffer containing 10 mM imidazole. Stringent washes with 10-15 column volumes of 20, 40 mM imidazole was done and batch elutions with 75 mM imidazole, 100 mM imidazole and 200 mM imidazole was performed twice each.

The eluted fractions were run on an SDS-PAGE gel to confirm the presence of the protein and its purity. The protein was concentrated in an Amicon Centriprep centrifuge concentrator at 4000 rpm to a final volume of 150-200 μl. New buffer with 50 mM Tris, 100 mM NaCl, and 10% glycerol at pH 8.0 was added to the concentrated protein and centrifuged again to remove the imidazole by buffer exchange. The protein concentration was estimated by Bradford method and stored in a buffer containing 50 mM Tris, 100 mM NaCl, and 20% glycerol at pH 8.0.

2.2.3 Bioinformatics Analysis of ACP

The ACP domain boundaries were defined as predicted (*68*). The low expression levels led us to examine the ACP sequence to redefine the boundaries. The secondary structure of CalE8 protein sequence was examined using bioinformatics tools. The domain boundaries were redefined to include helices that were intercepted in the previous boundaries.

2.2.4 ACP Cloning, expression and purification

The ACP gene fragment, with revised domain boundaries, *me*ACP (2773–3051 bp) was amplified from the CalE8 gene in pUC57 (5'**GCTCATATG**GCTCGTGCTGACGAC3', 5'-GCCTCGAG TTAGTGCGGAGCTTCAC-3'). The PCR reactions were conducted at two annealing temperatures starting at 42 ℃ for first 10 cycles and 60 ℃ for the next 20 cycles. The initial denaturation was carried at 95 ℃ for 3 min and the amplicons were allowed an extension span of 30 seconds per cycle and a final extension span of 10 min at 72 ℃.

The PCR product was digested with NdeI and XhoI enzymes. pET28b+ vector was also simultaneously digested with NdeI and XhoI. The vector and insert were mixed in a ratio of 3:2 and ligated in a ligation mix that contained 2X rapid ligation buffer and T4 DNA ligase (Promega). It was transformed into NovaBlue competent cells (Novagen). Plasmid was extracted using Bio-Rad Plasmid mini-prep kit. The plasmid was validated by restriction and sequencing. It was transformed into BL21DE3 and plated in LB-Agar plates containing kanamycin. Colonies were picked and grown in LB media with kanamycin. The overnight culture was scaled up to 4-5litres and induced with 0.5 mM IPTG once the OD_{600nm} reached 0.6. The induced culture was grown at 16°C for 16-18 hours.

The culture grown overnight was centrifuged using a JA-10 rotor at 8000 rpm for 10 min. The pellet was washed with wash buffer (50 mM Tris, 500 mM NaCl, 5% glycerol, pH 8.0). The resuspended pellet was then sonicated at 40% amplitude, 5 seconds pulse with 2 seconds pause for 5-10 min and lyses using a microfluidizer. The mixture was centrifuged in a JA-25.5 rotor for 45 min at 20000 rpm. The clear lysate was mixed with Ni-NTA resin (GE Healthcare) and the resin-lysate mixture was incubated for an hour. It then was added onto a column and washed with 20 times the column volume of wash buffer containing 10 mM imidazole. Stringent washes with 10-15 column volumes of 20, 40 mM imidazole was done and batch elutions with 75mM imidazole, 100 mM imidazole and 200 mM imidazole was performed twice each.

Elutes from affinity chromatography were run on an SDS-PAGE gel to confirm the presence of the protein and its purity. The protein was concentrated in an Amicon Centriprep centrifuge concentrator at 4000 rpm to a final volume of 150-200 µl. New buffer with 50 mM Tris, 100 mM NaCl, and 10% glycerol at pH 8.0 was added to the concentrated protein and centrifuged again to remove the imidazole by buffer exchange. The protein concentration was estimated by Bradford

method and stored in a buffer containing 50 mM Tris, 100 mM NaCl, and 20% glycerol at pH 8.0.

2.2.5 PPTase cloning, expression and purification

PPTase1 (5119–5757 bp), PPTase2 (4918–5757 bp) and PPTase3 (4771–5757 bp) were amplified from the synthesized PKS gene based on the sequence of CalE8 in *M. echinospora* (*147*). PCR amplification was carried out using DynazymeEXT (Finnzymes) (Primers: PPTase1-PPT DOM Cl F NdeI- 5'GACCATATGGCTACCGGTGCTGTTATCG 3',

PPT-PKSDomClXhoIR STOP 5' CAACTCGAG**TTA**ACGACCAGCACCAGACAG-3',

PPTase2: PPTExt XhoI R pET28b+ 5' CAACTCGAGTTA**ACGACCAGCACCAG** 3',

PPTExt NdeI F pET28b+ 5' GACCAT**ATGGGTGACCCGGGTGCTC** 3',

PPTase3: PPT+LD dom ClF NdeI 5' GACCATATGGCTCGTGACCTGTACGGTG 3',

PPT-PKSDomClXhoIR   5'-CAACTCGAGACGACCAGCACCAGACAG-3') and the PCR products were ligated into pET-28b(+) (Novagen) between the NdeI and XhoI sites. The plasmids were transformed into the *E. coli* strain BL21 (DE3) for protein expression after the validation of DNA sequences. PPTase1 and PPTase3 were further purified by size-exclusion chromatography with a superset- 75 column and buffer that contains 50 mM Tris (pH 8.0), 5% glycerol and NaCl of various concentrations (50, 100, 200 or 500 mM). Final protein concentrations were measured using the Bradford method.

2.2.6 Sfp cloning, expression and purification

The Sfp gene in pET26b+ obtained from Dr. Christopher Walsh's lab was transformed into BL21DE3 and expressed. Sfp was expressed and purified by Ni-NTA affinity chromatography under the same conditions as described for ACP and PPTase.

2.2.7 DACP, EACP, GPCP- cloning, expression and purification

The genes encoding EACP (from *E. coli*), DACP (from *Saccharopolyspora erythraea*) and GPCP (also GrsA-PCP, from *Bacillus brevis*) were custom synthesized by GenScript Corporation. They were cloned into pET-28b+ at the NdeI and XhoI sites and transformed into BL21DE3. The proteins were purified by Ni-NTA affinity chromatography under the same conditions as described for ACP and PPTase.

2.2.8 PPTase activity assay

The activity assays for PPTases (Sfp, PPTase1, and PPTase3) were performed against different carrier proteins (*me*ACP, DACP, EACP and GPCP) using a HPLC system (LC1200, Agilent). The reaction mixture contains 50 μM CP, 1 mM CoA, 15 mM $MgCl_2$ and 10 μM PPTase in the assay buffer (50 mM Tris, 100 mM NaCl, 5% glycerol). For each reaction, a control reaction without CoA was carried out in parallel. The reactions for Sfp were carried out at pH 7.0 and reactions for PPTase1 and PPTase3 were carried out at their optimal pH of 8.8 as determined by pH screening. The reactions were incubated at 37°C ranging from 15 min to 120 min. After stopping the reaction, 20 μl of the reaction mixture was loaded onto an analytical C8-300SB Zorbax column (Agilent) equilibrated with 0.05% TFA in HPLC-grade H_2O. By monitoring 220 nm absorbance, the *apo-* and *holo*-CPs were eluted with a linear CH_3CN gradient (20–90%) at the flow rate of 1 ml/min.

2.2.9 MALDI mass spectrometry.

The *apo* and the *holo* fractions from HPLC were collected and submitted for MALDI mass spectrometry. The molecular weight of the *apo-me*ACP was determined by MALDI mass spectrometry in Prof. Tang Kai's lab, Nanyang Technological University, Singapore.

2.2.10 NMR

The ACP domain structure was solved using NMR in collaboration with Prof. Yang Daiwen's group at National University of Singapore, Singapore (*147*). 13C and 15N-labelled *me*ACP was prepared for NMR spectroscopic experiment and assignment.

2.2.11 Bioinformatics analysis

The NMR structure of ACP was examined using various bioninformatics tools. The sequences of representative ACP domains with solved crystal structures were selected and retrieved from the Protein Data Bank. Sequence and structural features were investigated. The structures of interest were visualized and examined using molecular visualization software.

2.3 Results and Discussion

2.3.1 Heterologous expression of CalE8

Cloning, expression and purification of the 212 kDa CalE8 enzyme was done to study the whole protein in conjunction with the domains. The *CalE8* gene was cloned into pET28b+ vector and was transformed into BL21(DE3) cells. CalE8 was expressed as a soluble protein at high levels and was purified (**Fig. 2.4**). The enzymatic catalytic activity of the protein has been confirmed by *in vitro* assays (*146*).
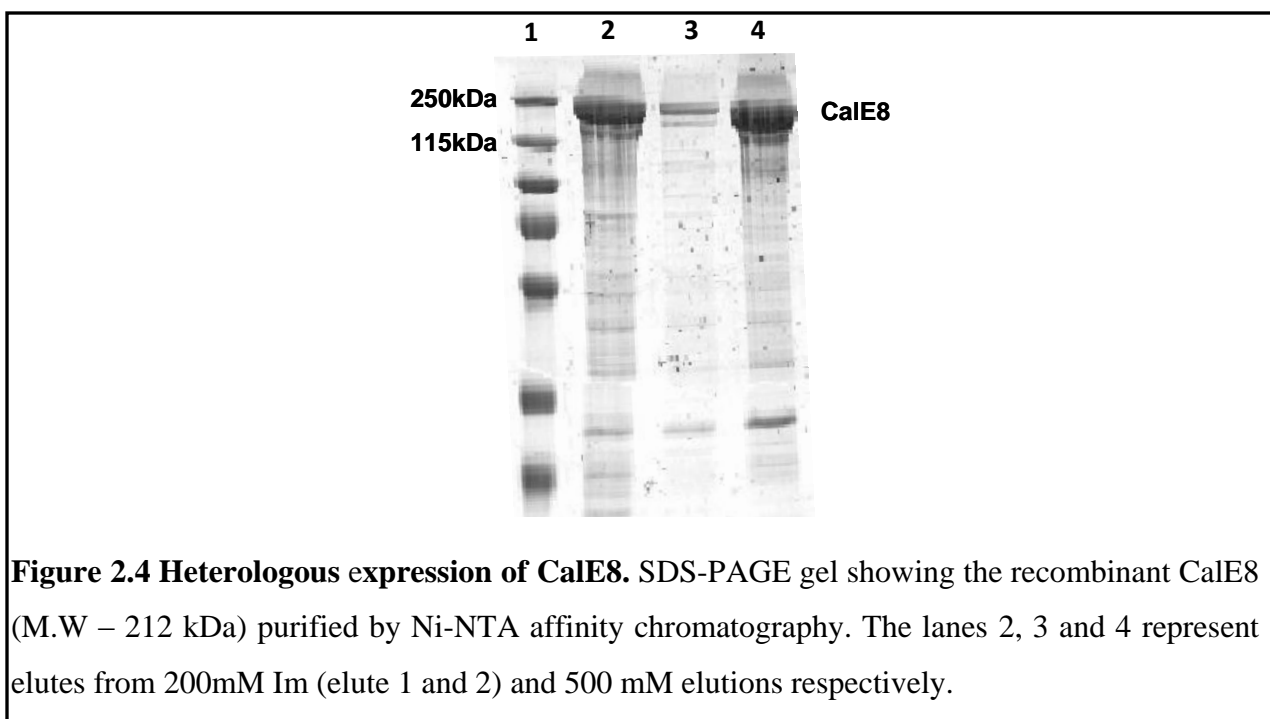
Figure 2.4 Heterologous expression of CalE8. SDS-PAGE gel showing the recombinant CalE8 (M.W − 212 kDa) purified by Ni-NTA affinity chromatography. The lanes 2, 3 and 4 represent elutes from 200mM Im (elute 1 and 2) and 500 mM elutions respectively.

2.3.2 Characterization of *me*ACP

2.3.2.1 Heterologous expression of *me*ACP

To study the activity of the individual domains, the ACP domain was the first domain to be investigated. The *ACP* gene was amplified by PCR using the synthesized *CalE8* gene as template. The gene fragment (216bp) generated, based on the boundaries of ACP as defined

earlier in literature (*60*) was cloned into pET26b+ vector and transformed into BL21(DE3). There was very little or no expression with this construct. The domain boundaries for ACP were then redefined based on secondary structure so as to include an α-helix which was previously partially terminated. The new construct (276bp) was simultaneously cloned into pET26b+ and pET28b+ vectors. Better expression was observed for the ACP domain in *E. coli* (**Fig. 2.5**). The ACP domain (*me*ACP) with revised boundaries contained 92 residues (Ala925-Pro1016). The protein purified using affinity chromatography remained stable in solution.

Figure 2.5 Heterologous expression of *me*ACP. SDS gel showing the purified *me*ACP protein (M.W ~12kDa) from Ni-sepharose affinity chromatography. *me*ACP gene in pET28b+ plasmid was expressed in BL21(DE3) expression host.
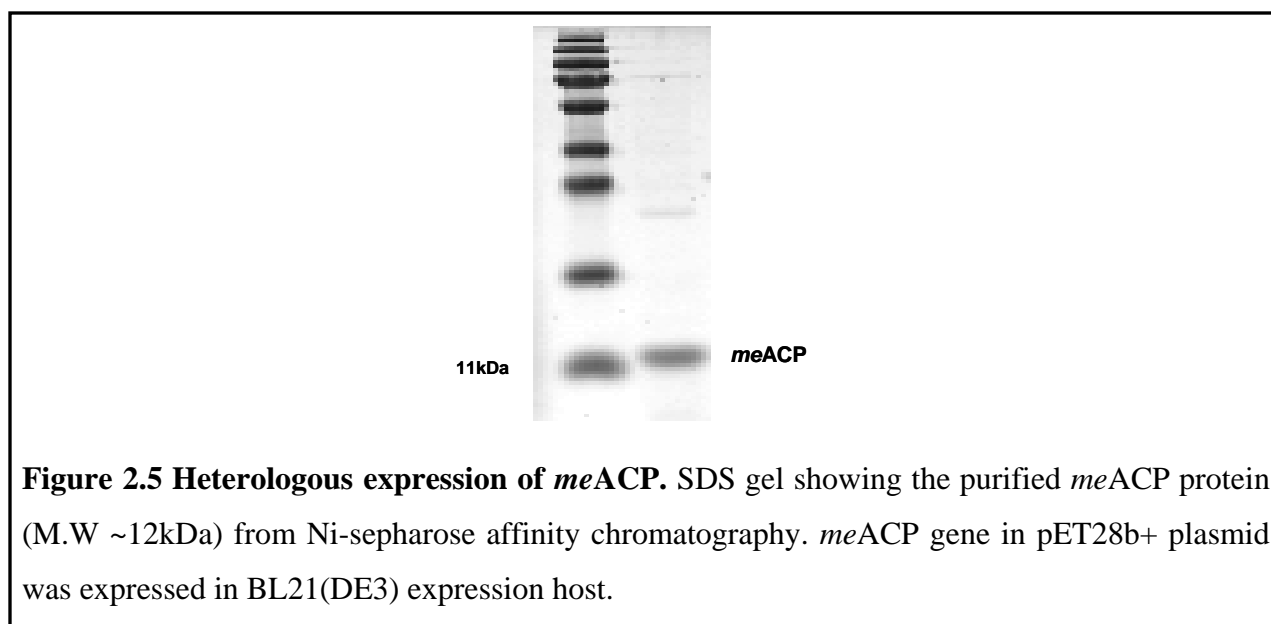
2.3.2.2 Test for activity of *me*ACP

The addition of the 4'-Ppant moiety onto the *apo*-ACP converts the ACP to its active *holo* form. Since this modification falls below the detection limits of normal protein analysis techniques like SDS-PAGE, a more sensitive technique like HPLC was employed to assess the phosphopantetheinylation. HPLC analysis of the purified *me*ACP showed that it was not phosphopantetheinylated by *E. coli* PPTases *in vivo*. The *apo*-form (11451.85 Da) was validated by mass spectrometry.

In order to test the phosphopantetheinylate ACP under *in vitro* conditions, the surfactin PPTase, Sfp was employed. Sfp is a type II PPTase from *B. subtilis* that exhibits broad substrate specificity. The *Sfp* gene obtained from Dr. Christopher Walsh's lab was transformed in BL21(DE3). The protein was expressed and purified in high concentrations (**Fig. 2.6**).
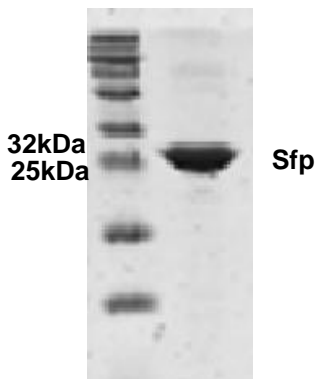
Figure 2.6 Heterologous expression of Sfp. SDS gel showing the Sfp protein purified from Ni-NTA affinity chromatography. Sfp gene in pET26b+ plasmid was expressed in BL21DE3 expression host.

The Sfp protein with a broad substrate spectrum was able to phosphopantetheinylate *me*ACP. The formation of *holo-me*ACP after the incubation with Sfp and CoA for 1 h (**Fig. 2.7 a**) was examined by HPLC. The modification was seen clearly as the peaks corresponding to the *apo* (14.8 min) and *holo* (14.3 min) forms were distinctly apart (**Fig 2.7a**). The overall turnover increased slightly with extended incubation time. The results confirmed the identity of the putative ACP domain despite its low sequence homology shared with other CPs. Three different CPs from different families of CPs, including EACP from the fatty acid synthesis pathway in *E. coli*, GPCP (or GrsA-PCP) from an NRPS in *B. brevis*, and DACP from the modular PKS DEBS (module 2) in *S. erythraea* (*97, 150*) were cloned, expressed and purified. The three *apo*-CPs were phosphopantetheinylated by Sfp, though with varying efficiencies (**Fig. 2.7 b**). The relative

efficiency for the modification of DACP, *me*ACP and GPCP varied slightly with assay conditions such as pH and incubation time. The efficiency for *me*ACP modification is comparable to that of DACP and slightly higher than that of GPCP. EACP modification by Sfp was always rapid and was almost complete within 30 min. The results suggest that *me*ACP resembles DACP and GPCP more than the EACP protein from primary metabolic pathway.
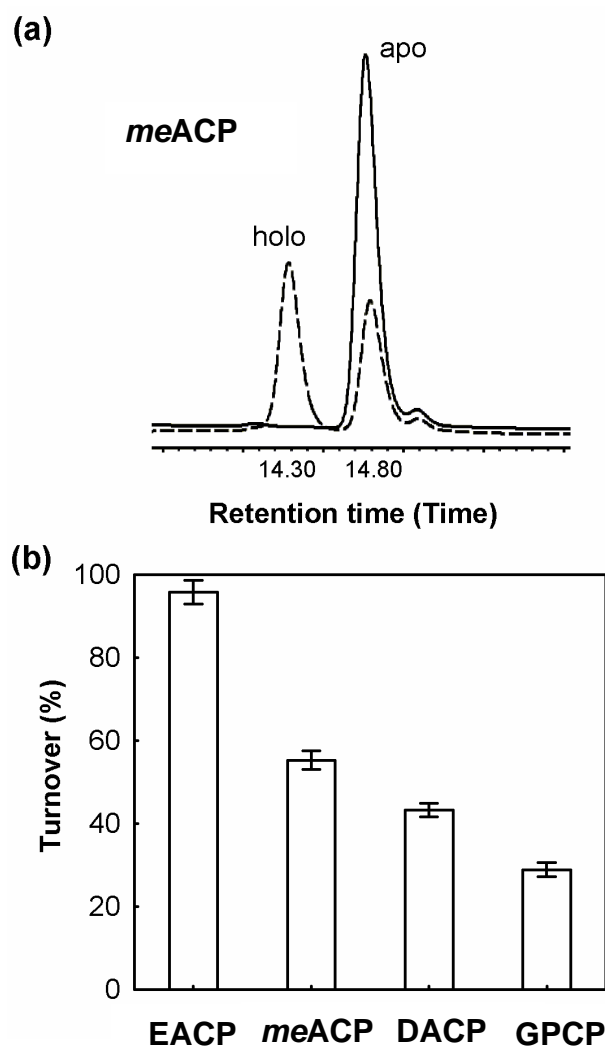
Figure 2.7 Modification of carrier proteins by Sfp. a) HPLC chromatograms of *me*ACP after the incubation with Sfp in the presence (dash line) and absence (solid line) of CoA, b) Relative turnover of *apo*-CPs to *holo*-CPs as catalyzed by Sfp. (Reaction conditions: 50 mM Tris–Cl, pH 7.0, 100 mM NaCl, 5% glycerol, 15 mM Mg^{2+}, 1 mM CoA, 60 min) (*151*).

2.3.3 Characterization of PPTase

2.3.3.1 Heterologous expression of PPTase

The fact that CalE8 was able to release growing intermediates into solution had confirmed that there must be an ACP domain which was phosphopantetheinylated. This means that there must be an inherent PPTase domain in CalE8. The modification of *me*ACP by its cognate partner, the putative C-terminal PPTase domain, had to be examined. After confirming the identity of the putative *me*ACP, we proceeded to test the activity against the PPTase domain that exists within the same multidomain CalE8. The domain boundaries (1707– 1919 aa) for the PPTase domain were initially determined by sequence alignment with type II PPTases (*152*). The purified PPTase domain protein contains 213 residues and is referred as PPTase1. Size-exclusion chromatography revealed that PPTase1 mainly exists as a high-molecular-weight oligomer in 50 mM Tris (pH 8.0) with 100 mM NaCl and 5% glycerol (**Fig. 2.8**). The incubation of PPTase1 and the four CPs separately in the presence of Mg^{2+} and CoA did not yield *holo*-CPs under the same conditions for Sfp activity assay. Extensive screening of buffer conditions with high protein concentration and extended incubation time still failed to generate any detectable *holo* forms. Given the observation that PPTase1 mainly exists as a high molecular-weight oligomer, we reasoned that the inactivity could be caused by the oligomerization; similar to what was observed for the free-standing C-terminal type III PPTase domain from yeast fatty acid synthase (*153*). The assignment of the PPTase domain boundaries based on sequence comparison left 110 residues between the DH domain and the putative PPTase domain (PPTase1) with unknown function. In an effort to obtain active PPTase, we cloned two more proteins with one covering the sequence 1640–1919 aa (referred as PPTase2) and the other covering the sequence 1591– 1919 aa (referred as PPTase3) (**Fig. 2.3**). PPTase2 contains the putative PPTase domain and half

of the upstream linker region, while PPTase3 contains the putative PPTase domain and the whole

linker region. This linker region was later discovered by bioinformatics analysis to be part of the

intact DH domain. The (His) 6-tagged PPTase2 (31.0 kDa) and PPTase3 (35.7 kDa) were both

over-expressed in *E. coli* in high yield (**Fig. 2.8**).
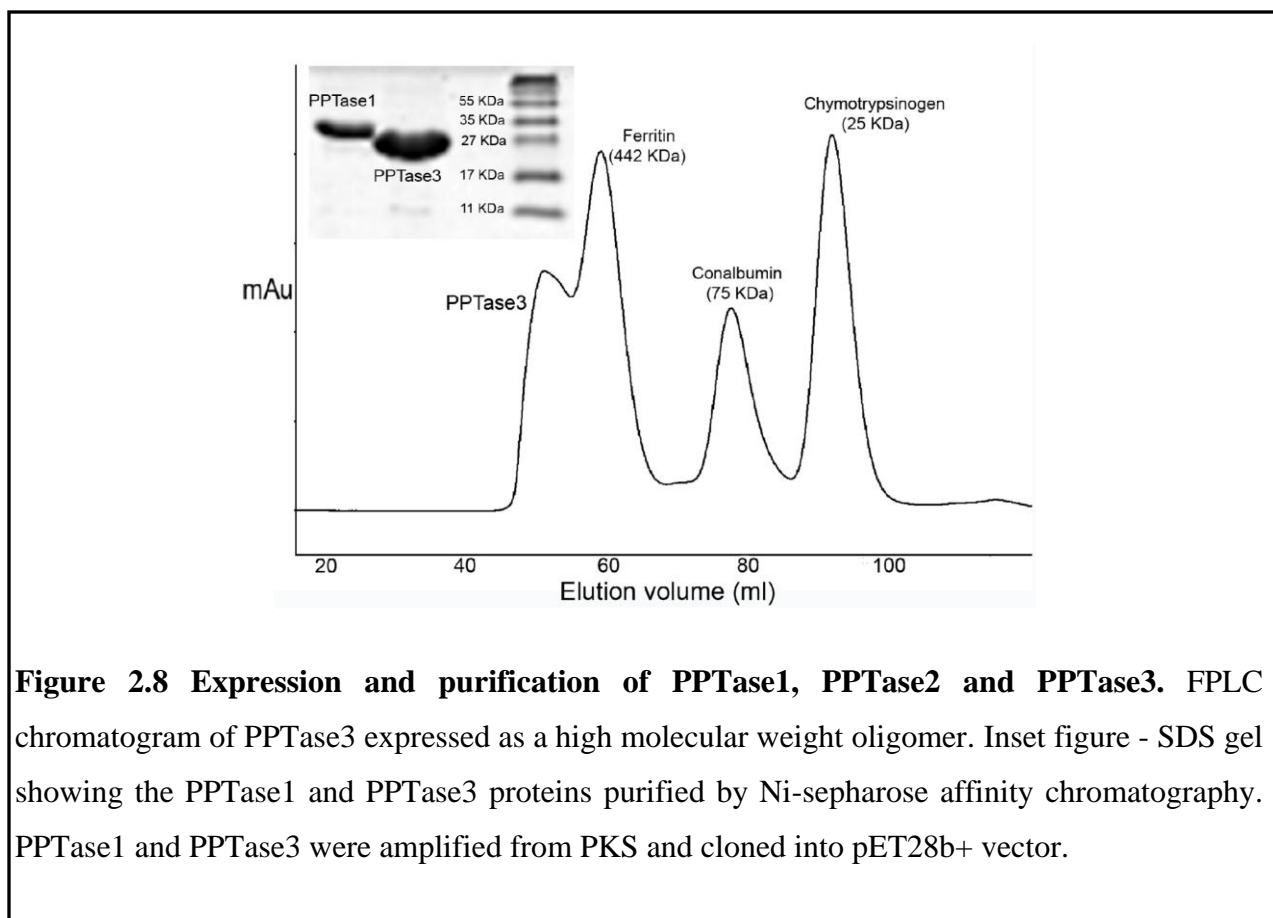
Figure 2.8 Expression and purification of PPTase1, PPTase2 and PPTase3. FPLC chromatogram of PPTase3 expressed as a high molecular weight oligomer. Inset figure - SDS gel showing the PPTase1 and PPTase3 proteins purified by Ni-sepharose affinity chromatography. PPTase1 and PPTase3 were amplified from PKS and cloned into pET28b+ vector.

While PPTase3 was found in soluble fraction and readily purified by Ni-NTA affinity

chromatography, PPTase2 was expressed mainly as insoluble inclusion body. This hindered

further examination of PPTase2. However, PPTase3 was subjected to thorough examination.

Size-exclusion chromatography showed that the majority of PPTase3 also exists as a high-

molecular-weight oligomer (**Fig. 2.8**). The addition of 0.1% Tween and 20% glycerol in lysis

and purification buffers did not change the oligomeric state. The increase of salt concentration up

to 500 mM gave rise to a mixture of oligomers with different molecular weights (data not shown). No monomeric form of PPTase3 was observed for all the buffer conditions we tested.

2.3.3.2 Test for PPTase enzymatic activity

The activity assays showed that PPTase3 is an active PPTase that exhibits high and modest efficiencies in the modification of EACP and GPCP, respectively despite the oligomerization, (**Fig. 2.9**). In contrast, PPTase3 seemed not able to modify *me*ACP and DACP under various assay conditions. The incapability of PPTase3 in modifying *me*ACP and DACP is surprising considering that *me*ACP is its cognate ACP partner. The modification of *me*ACP by PPTase3 was further examined under various conditions. Even with very high enzyme concentrations and longer incubation times, only a negligible amount of *holo-me*ACP was observed, suggesting extreme low catalytic efficiency of PPTase3 for *me*ACP.

Figure 2.9 *In vitro* modification of CPs by PPTase3. HPLC chromatograms of CPs after incubation with PPTase3 in the presence (dash line) and absence (solid line) of CoA. (Reaction conditions: 50 mM Tris–Cl, pH 7.0, 100 mM NaCl, 5% glycerol, 15 mM Mg^{2+}, 1mM CoA, 60 min) (*151*). While EACP and GPCP are modified by PPTase3, there is no modification observed with the cognate partner, *me*ACP and DACP

A comparison with other ACP structures reveals some salient features of *me*ACP (**Fig 2.10, 2.11**).

Figure 2.10 Comparison of *me*ACP with other carrier proteins. (a) Partial sequence alignment of *me*ACP and representative CPs whose structures have been solved by X-ray crystallography b) The conserved Ser971 and related residues in various acyl carrier proteins. (ACP:PKS1 (DACP) and ACP:PKS2 are from the module 2 and 5 of DEBS; ACP:FAS1 (EACP) and ACP:FAS2 are from the fatty acid synthesis pathways of *E. coli* and *B. subtilis*; ACP:PKS3 and ACP:PKS4 are from two aromatic PKSs in *S. coelicolor* and *S. griseus*; PCP1 (GPCP) and PCP2 are from NRPSs of *B. subtilis* and *B. brevis*. ACP:lys2 is from the lysine synthesizing pathway of *S. cerevisiae*).

Figure 2.11 Phylogenetic analysis of *me*ACP. Phylogenetic examination of *me*ACP revealed its evolutionary standing against a) representative CPs (with X-ray crystallographic structure solved) and b) specifically from NRPS, FAS, aromatic and modular PKSs.

2.3.4 Structural characterization of *me*ACP

Analysis of the structure of *me*ACP, solved by NMR (*147*), provided significant information on the structural features of *me*ACP. The *me*ACP protein shares very low sequence homology with other ACPs and lacks the signature GX(H/D)S(L/I) motif conserved in many ACPs (**Fig. 2.11 a, b**). But the attachment site (Ser971) is still located at the beginning of helix α2 (**Fig. 2.12**). The GX(H/D)**S**(L/I) motif is replaced by a H968MS**S**971I motif in *me*ACP with His968 and Ser970 substituting the canonical Gly and Asp/His residues. The essential role of Ser971 was confirmed (*147*) when it was observed that S971A mutation completely abolished the activity of CalE8 (data not shown). The HMS970 triad constitutes part of loop-1 with the solvent-exposed side chains His968 strictly conserved among the *me*ACP homologs (**Fig. 2.12**). Considering that the residues preceding Ser971 have been known to affect protein-protein interaction in some ACPs and given the different physico-chemical property of the residues, a different protein surface of *me*ACP may be involved in the recognition of some of its partners.

Figure 2.12 Solution structure of *me*ACP. a) The solution structure of *me*ACP solved by NMR. Helix I is colored blue. The "recognition helix", helix II is colored green with the conserved Serine residue highlighted in red. The helix III is colored orange. b) Structure of the motif harboring the 4-Ppant-attaching serine in *me*ACP and two other ACPs. The **G**X(H/D)**S**(L/I) motif in type I NSAS PKS ACP (PDB: 2kr5) (blue) and type II actinorhodin PKS ACP (PDB: 1af8) (yellow) and the HMSSI motif in *me*ACP (red) are shown as ribbons. The conserved Gly (G1743) 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. c) Superimposed structures of *me*ACP and 2kgc (actinorhodin PKS ACP from *S. coelicolor*). (Adapted from Lim et al. 2011 *(147))*

The striking resemblance between the types II FAS or PKS ACPs is that they contain a α2 helix that is rich in acidic residues. The negative charges clustered on helix α2 are considered to dictate the recognition of the free-standing ACPs by their protein partners. Hence, *α*2 helix is known as the 'recognition helix' in carrier proteins. In comparison, *α*2 helix in type I FAS ACP domains seems to be less negatively charged. This might be because the specific domain-domain interaction is less critical in the multidomain type I FAS system. On the other hand, the *α2* helix of *me*ACP contains only a single acidic residue (Glu980) which is involved on the formation of a salt bridge with Arg949 from *α*1 helix. This clearly suggests that *me*ACP resembles the ACP domains from type I FASs more than the free-standing type II ACPs. This similarity of *me*ACP

to type I ACPs is consistent with the phylogenetic relationship of *me*ACP with other carrier proteins (**Fig. 2.11 b**). The loop-1 harbors a large number of highly charged residues that include Glu951, Glu955, Arg958, Asp959, Asp960, Arg962, Asp965, Asp966 and His968, compared to the relatively neutral *α*2 helix. The high propensity of the charged residues on loop-1 is not uncommon and has been seen in ACPs such as the ones from the biosynthetic pathways of frenolicin and norsolorinic acid (*155, 156*).

The loop-2, containing a short helix for some ACPs, has been suggested to serve as a conformational switch in sequestering ACP domain-tethered acyl chains in type II FAS systems (*86*). Except for *act* ACP and *B. subtilis* FAS ACP (*102*), all the characterized ACPs contain a short helix with loop-2. This short helix seems to be absent in *me*ACP as well. In type II ACPs, the loop-2 is largely made up of charged or small non-polar residues that facilitate the opening of a hydrophobic cleft for binding acyl intermediates (**Fig. 2.13 a**); however, 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 (**Fig. 2.13 b**) (*155*). The loop-2 in *me*ACP also contains two hydrophobic residues (Met992, Phe996) with their bulky side chains packed against the hydrophobic core of the protein (**Fig. 2.13 c**).

a)   b)   c)
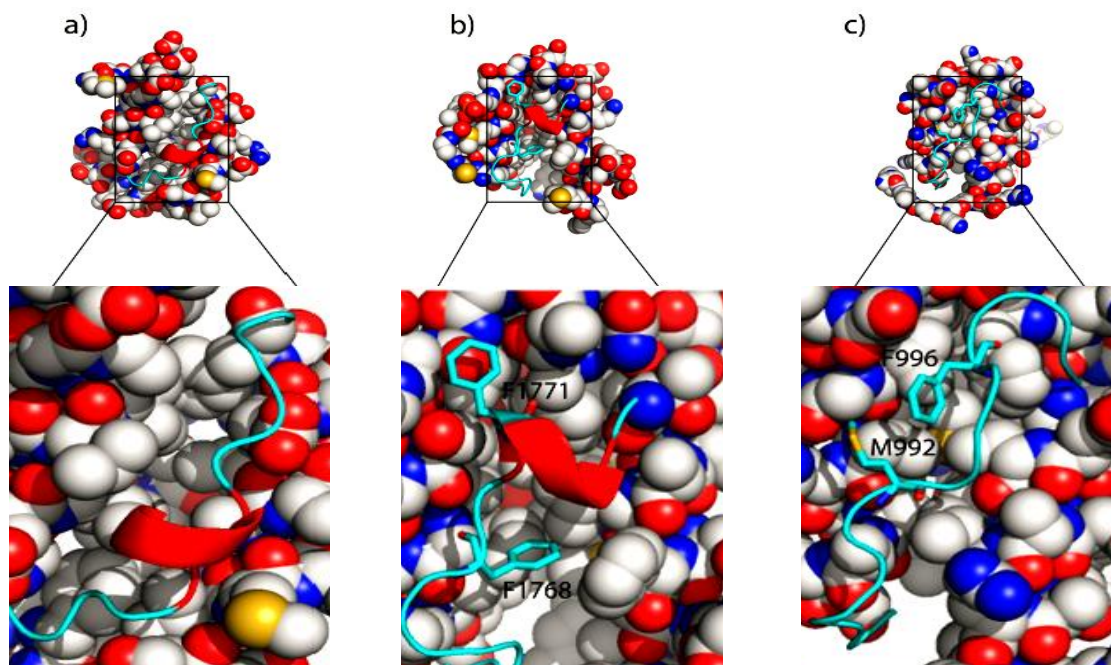
F996

M992
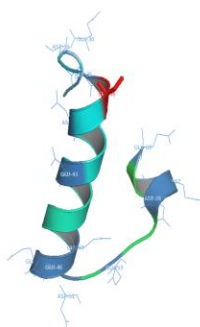
Figure 2.13 Conformation mobility of loop-2. (a) Type II actinorhodin PKS ACP (PDB: 1af8) that contain mainly small residues in loop-2; Actinorhodin PKS ACP (PDB: 2kgc), from *S. coelicolor*, recognition helix with acidic residues marked in blue. (b) Type I NSAS (continued) (continued) PKS ACP (PDB: 2kr5) with the two bulky residues (Phe66 and Phe69, labelled as F1771 and F1768) highlighted; *E. coli* ACP (PDB: 2fae) recognition helix with a high percentage of acidic residues, marked in blue. (c) *me*ACP with the two bulky residues (Phe996 and Met992, labelled as F996 and M992) packed against the hydrophobic pocket highlighted; *me*ACP recognition helix (helix II) with the acidic residues marked in blue. (Adapted from Lim et al. *(147))*
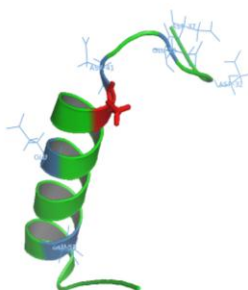
NMR studies of *me*ACP (*147*) reveal that the residues of loop 2 (Gly986 -Thr1000) are still associated with greater flexibility relative to the residues of loop 1 and helices α1, 2 and 3 and raises the interesting possibility that it may undergo a large conformational change to open a protein pocket for binding the PKS products.

The activity of the putative ACP domain was validated by using Sfp, which exhibits broad substrate specificity. The incubation of *me*ACP and three representative CPs with Sfp in the presence of CoA and Mg^{2+} led to the formation of *holo*-CPs. The modification efficiency for *me*ACP is comparable to those of DACP and GPCP but substantially lower than that of EACP. The relative efficiency suggests that *me*ACP resembles the ACP or PCP domains from secondary metabolic pathways. The resemblance of *me*ACP to PKS ACPs is further supported by phylogenetic analysis, which revealed that *me*ACP is evolutionarily close to modular PKS ACPs, rather than FAS-ACPs (**Fig. 2.11 b**). Though Sfp has been suggested to exhibit substrate preference towards the CPs from secondary metabolic pathways (*157*), the higher modification efficiency for EACP in comparison to the three CPs from secondary metabolic pathways probably reflects the variation among different FAS-ACPs. Sfp also has varying efficiencies towards PCPs from different species, including tycc3-PCP from *B. subtilis* ($kcat = 96$ min^{-1}, $kcat/Km = 22.6$ min^{-1} μM^{-1}) and GPCP (GrsA-PCP) from *B. brevis* ($kcat = 10.3$ min^{-1}, $kcat/Km = 2.5$ min^{-1} μM^{-1}) (*158, 159*).

2.3.5 Characterization of PPTase

PPTase1, PPTase2 and PPTase3 showed good expression yield. PPTase1 shares 18% sequence identity and 31% similarity with Sfp, and contains most of the catalytic residues for Mg^{2+} and CoA binding for type II PPTases (**Fig. 2.13 c**). However, PPTase1 is catalytically

inactive and unable to modify the four CPs. In contrast, PPTase3, the variant of PPTase1 that contains the entire upstream linker region (another half of the intact DH domain) exhibits PPTase activity. The formation of high-molecular-weight oligomer for PPTase1 and PPTase3 suggests that the differences in activity between the two are probably not due to differences in oligomeric state. The lack of activity for PPTase1 indicates that the linker region is critical for the function of the PPTase domain. Although it is unclear whether the linker region is directly involved in substrate binding and catalysis, it probably plays a structural rather than catalytic role considering that most putative catalytic residues are contained in the PPTase1 region based on sequence alignment with Sfp. This view is further supported by the observation that PPTase2, the protein that contains only about half of the linker region, is not able to fold properly into a soluble protein. Unexpectedly, when we aligned the sequences of PPTase3 and type II PPTases, the linker region (116 aa) always aligned with the first half of the type II PPTases. The linker region shares sequence homology with type I AcpS from *B. subtilis* (17% identity; 39% similarity) and *S. coelicolor* (28% identity; 38% similarity) (**Fig. 2.13 a**). Also, important catalytic residues in type I AcpS-like proteins, including the residues for Mg^{2+} binding, subunit interaction and 30, 50-ADP stabilization (*160*), also seem to be conserved in the linker region.
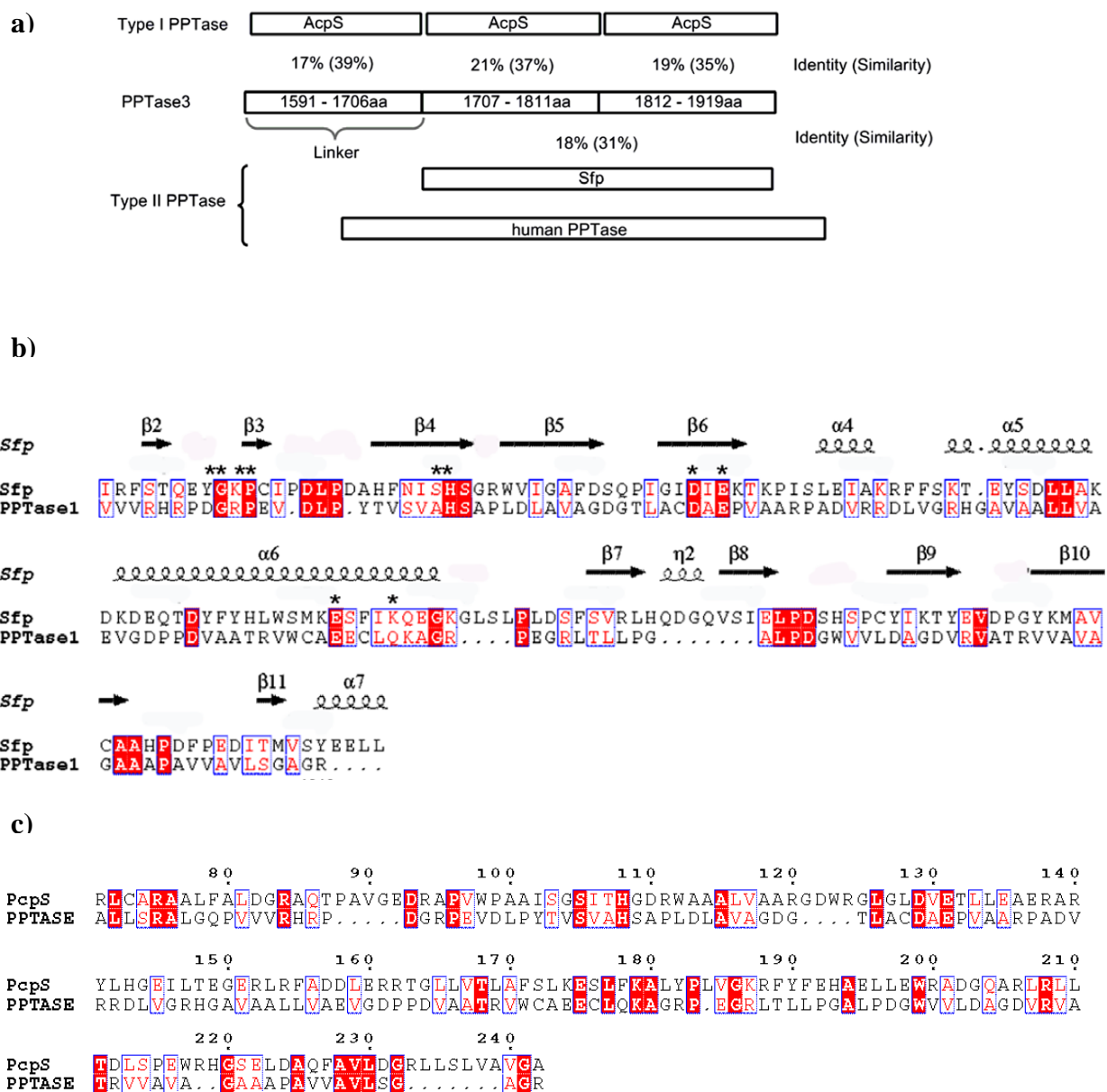
Figure 2.14 Sequence analysis of PPTase. Comparison between PPTase3 and type I (AcpS from *B. subtilis*) and type II PPTases (Sfp from *B. subtilis* and human PPTase). (a) Schematic illustration of the pseudo-trimeric structure of PPTase3 and the pseudo-dimeric structure of Sfp and human PPTase. (b) Sequence alignment of Sfp and the 1707–1919 aa region. The conserved catalytic residues in type II PPTases are indicated by asterisk (*112*). (c) Sequence alignment of PPTase3 and PcpS.

These observations led us to propose that the intact PPTase domain spans the C-terminal region (212 aa) and consists of two subdomains similar to Sfp PPTases. The identification of the conserved catalytic residues for type II PPTases from sequence alignment indicates that the PPTase domain is likely to function like Sfp with only one active site for substrate binding and catalysis. Investigations done on DH domain based on bioinformatics have extended the role of the linker domain. The linker domain that plays a structural role in PPTase may actually be the other half of the intact DH domain. More details on the domain and its dual role have been explained in Chapter IV. The extremely low efficiency in the modification of DACP and the cognate partner *me*ACP by PPTase3 was surprising. Given the substrate preference toward EACP and GPCP, it seems that the PPTase domain resembles PcpS, a type II PPTase that prefers FAS-ACPs and NRPS-PCPs (*157*). Indeed, PPTase3 shares slightly higher sequence homology with PcpS (20% identity; 32% similarity) than Sfp (18% identity; 31% similarity), though the determinants of the substrate specificity are not obvious from sequence comparison. The substrate specificity of PPTase3 suggests that the PKS probably acquired the PPTase domain from a primary metabolic pathway during evolution. The extremely low efficiency in *me*ACP modification also indicates that the PPTase domain has not evolved to be an efficient PPTase for modifying its cognate ACP domain. This is probably because there was little selective pressure for the PPTase domain to improve its substrate specificity towards the *me*ACP domain, due to the low kinetic barrier of self-phosphopantetheinylation within the PKS. Although the *in vitro* activity assay did not clearly demonstrate the modification of *me*ACP by PPTase3, in the study of the full-length recombinant PKS for calicheamicin biosynthesis, we found that the recombinant PKS is able to generate a 15-carbon polyene product in the presence of acetyl CoA,

malonyl CoA and NADPH. The production of the polyene confirms the occurrence of *in vivo* self-phosphopantetheinylation as well as the function of the PPTase domain.

2.3.6 Conclusions

The region spanning 925-1016 in CalE8 protein sequence is the functional acyl carrier protein within the CalE8 protein. *me*ACP shares extremely low sequence homology with other known ACPs, including the ACP domain of the partially reducing (PR) 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 Ser971 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 relative deficiency 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 (Met992 and Phe996) 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 partially reducing (PR) types I iterative PKS NSAS and the

type I modular PKS DEBS. 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 multi-domain type I FAS and type I iterative PKS. Phylogenetic analysis shows that this ACP is similar to ACPs from the secondary metabolic pathways like DEBS ACP and PCP. The observation that *me*ACP was modified by Sfp and not its cognate partner, could be partially based on the unique structural features adopted by *me*ACP.

The C-terminal region (212 aa) of the iterative PKS in enediyne biosynthesis is the proposed PPTase domain. This PPTase domain when expressed as a stand-alone protein however requires the upstream linker region (one-half of the DH domain) to fold properly and to execute its activity. The C-terminal placement of the PPTase domain in enediyne PKSs bears a resemblance to the type III PPTase domains in yeast/fungal FASs (*161*). But identification of catalytic residues conserved in type II PPTases suggested that the PPTase domain is most likely a type II PPTase (*60*). The extremely low catalytic efficiency towards its own ACP partner and DACP indicates that the PPTase domain has not evolved to efficiently phosphopantetheinylate the ACP domains from secondary metabolic pathways.

Chapter 3. Expression, purification and characterization of the acyl carrier protein phosphodiesterase from *Pseudomonas aeruginosa*.

3.1 Introduction

3.1.1 Phosphodiesterases

Complex polyketides that are attached to the PKS are regarded as some of extremely useful antibiotics and antitumor molecules. The examination of the released products tethered to the 4'-Ppant moiety of a CP within the PKS will aid in the understanding mechanism of multidomain complexes and biosynthesis of medicinally important products. The addition of the 4'-Ppant moiety to the CPs by PPTases has been studied and explained to a larger extent compared to the removal of the 4'-Ppant moiety from CP, catalyzed by a lesser known enzyme, the ACP phosphodiesterase (AcpH), unique member in the family of phosphodiesterases. The AcpH catalyzes a divalent cation mediated hydrolysis of the phosphodiester bond (*162, 163*) that attaches the Ppant arm to the serine residue on the CP (**Fig. 3.1**). This process converts the *holo* CPs to their inactive *apo* forms, which could be regarded as a physiological reclamation of the CPs. The aim of this project was to identify and characterize such a phosphodiesterase that was capable of modifying the *holo-me*ACP as a stand-alone protein and as a domain within CalE8, and hence release the growing intermediates from CalE8. Functional characterization of the enzyme will encompass identification, cloning, expression, purification and functional and kinetic assays.

Figure 3.1 Modification of CP by AcpH. Schematic illustration of the reversible modification of CP by PPTase and acyl carrier protein phosphodiesterase (AcpH).
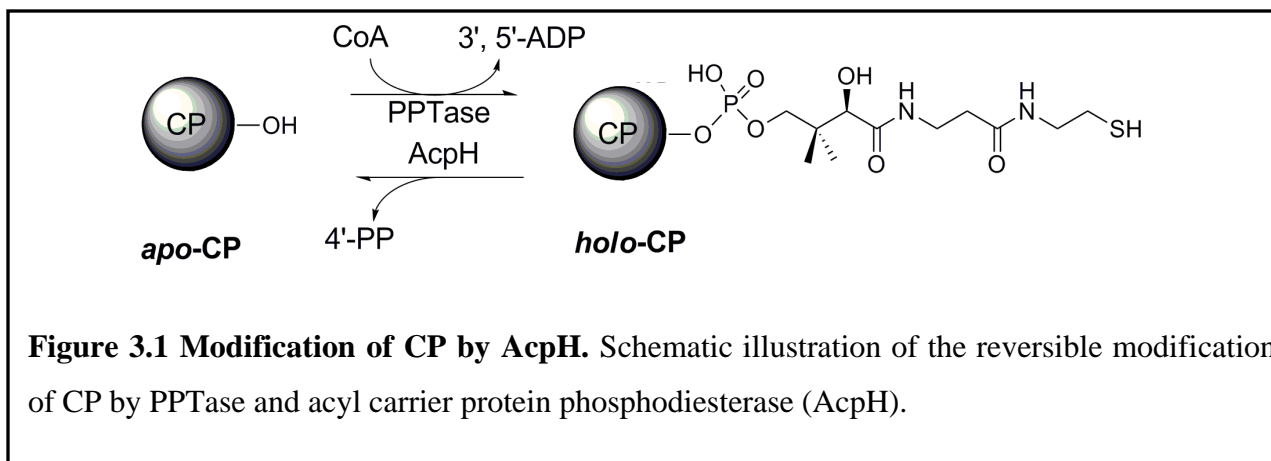
3.1.2 *Escherichia coli* AcpH (*Ec*AcpH)

In 1960s, an enzyme that was capable of removing the 4'-PP moiety from *holo*-ACP was identified by Vagelos and his group in 1967 in *E. coli*. The gene encoding the enzyme, *yajB,* was situated in between two genes unrelated to lipid metabolism – *queA* that encodes S-adenosylmethionine: tRNA ribosyltransferase-isomerase and *malZ* that encodes maltodextrin glucosidase. The gene was later renamed AcpH by Thomas and Cronan (*163*). The enzyme product was initially called ACP Hydrolase (AcpH) and was later renamed as ACP phosphodiesterase. The AcpH was found in *E. coli,* a Gram-negative bacterium, as a non-essential protein. Though its function was demonstrated almost 40 years ago, the cloning and expression of *E. coli* AcpH (*Ec*AcpH) was carried out recently (*163*). *Ec*AcpH is the only enzyme known to cleave a protein bound phosphodiester (*162*) linkage. Based on sequence analysis, it was seen that the AcpH proteins share very little sequence homology with any characterized protein and might be classified under a new fold (*162*). The *Ec*AcpH protein is a member of the unique class of enzymes classified under the HD phosphatase/phosphodiesterase family (*162*). It is to a lesser extent similar to the hydrolase domain of the bifunctional bacterial protein, SpoT, which exhibits phosphodiesterase activity. Though the *Ec*AcpH lacks the

significant histidine motif, a characteristic of this family, it hydrolyzes the phosphodiester link between the 4'-Ppant moiety and the serine-36 residue of the ACP. The working mechanism of *Ec*AcpH remains largely unknown. Studies and experiments have shown that the reactivity of this enzyme involves divalent cations, with Mn^{2+} being an efficient activator (*164, 165*). The enzyme exhibited no activity on the peptide fragments of ACP which suggested that it required ACP to be in its folded form. Interestingly, the removal of the prosthetic group from *holo*-ACPs by the *Ec*AcpH to generate *apo* ACPs has been attributed to the decrease in CoA levels (*166, 167*). It is significant to note here that CoA is the donor of the 4' Ppant group in the phosphopantetheinylation of ACPs by PPTases. A decrease in CoA levels could therefore trigger a possible feedback control in the system. The recombinant *Ec*AcpH remains unstable in solution and tends to aggregate upon overexpression and at high protein concentration. This led us in the pursuit of identification of a homolog with improved solubility and stability that would facilitate the elucidation of the structure, catalytic mechanism, and physiological function of this novel family of enzymes (*162, 163*).

3.1.3 *Pseudomonas aeruginosa* **AcpH (***Pa***AcpH)**

The unique capability of the AcpH to convert *holo*-ACPs to *apo*-ACPs by the phosphodiester cleavage could be employed to release the growing polyketide from a polyketide synthase. This triggered a search for a representative ACP hydrolase in the Calicheamicin producing *Micromonospora echinospora ssp calichensis*. A sequence based search against its genome revealed that a similar protein does not exist in *M. echinospora*. Hence, a BLAST search was done to identify a similar protein in related organisms based on sequence homology. The protein product of the gene *PA4353* from *P. aeruginosa* (PAO1)*, Pa*AcpH, was characterized as a protein of unknown function. It displayed significant sequence similarity (38% sequence

identity and 58% sequence similarity) with EcAcpH. The predicted Mn^{2+} binding residues, His6, Asp24 and Asp78 in EcAcpH were also conserved in PaAcpH. This led us to classify PaAcpH as the phosphodiesterase in *P. aeruginosa*. *P. aeruginosa* is a Gram negative bacteria like *E. coli*. Initial attempts to express the protein in *E. coli* BL21 (DE3) strain gave little expression with a very low fraction of soluble protein. After optimization of expression and purification conditions, we were able to express the recombinant PaAcpH as a soluble and functional protein. The activity of PaAcpH was tested against small molecular phosphodiester substrates such as para-nitrophenyl phosphate (pNpp) and bis-para nitrophenyl phosphate (bis pNpp). PaAcpH enzyme activity was also tested on several CPs from *P. aeruginosa* and other heterologous CPs including a major ACP (PA2966 or AcpP1) from the primary fatty acid synthesis pathway and several ACPs and PCPs with putative functions in other primary and secondary metabolic pathways (*168*). While the phosphodiesterase activity on CPs has been researched and reported (*162, 163*), the activity of ACP phosphodiesterases on ACPs within PKSs has not been studied. The ability of PaAcpH to release the products from PKS was tested with the two iterative PKSs, CalE8 and SgcE. The results obtained from the comprehensive studies carried out on PaAcpH revealed significant features of this novel phosphodiesterase. This chapter details the identification, cloning, expression, purification and functional and kinetic characterization of a novel phosphodiesterase capable of modifying and releasing the tethered intermediates from the CPs within PKSs and as stand-alone proteins.

3.2 Materials and Methods

3.2.1 Bioinformatics Analysis of PA4353

The *E. coli* AcpH (*Ec*AcpH) protein sequence was used as the template for a local search in the non-redundant database of protein sequences employing the BLAST algorithm hosted by the National Centre for Biological Information. The *Ec*AcpH sequence and PA4353 sequences were aligned using the ClustalW program hosted by European Bioinformatics Institute.

3.2.2 PCR amplification of the PA4353 gene

The PA4353 gene (573bp in length) in *Pseudomonas aeruginosa* (PAO1) genome (genomic location - 4881900 - 4881328 (-)) encodes the *Pa*AcpH protein. The primers designed to amplify the gene from the genomic DNA carried BamHI, XhoI restriction sites for cloning into pET28b+ vector (5'-CGAGGATCCATGAACTACCTCGC-3' and

5'-GTACTCGAGTCAGCGCTGGCTCAG-3') and Ek/LIC flanking sites for cloning into pCDF2 vector (5'-GACGACGACAAGATGAACTACCTCGCGC-3' and

5'-GAGGAGAAGCCCGGTCAGCGCTGGCTCAG-3'). PCR amplification was carried out using DynazymeEXT (Finnzymes) and Expand High Fidelity system (Roche). Both the PCR reactions were conducted at two annealing temperatures starting at 42 ℃ for first 10 cycles and 60 ℃ for the next 20 cycles. The initial denaturation was carried at 95 ℃ for 5 min to facilitate maximum relaxation of the genomic DNA template. The amplicons were allowed an extension span of 45 seconds per cycle and a final extension span of 10 min at 72 ℃. The PCR products were confirmed by agarose gel electrophoresis and visualized under Trans UV using GeneSnap (BioRad).

3.2.3 Cloning of *Pa*AcpH and CPs

The *Pa*AcpH gene with BamHI and XhoI flanking sites was subcloned into pGEMT and grown overnight. The *Pa*AcpH in pGEMT plasmid was extracted and restricted with BamHI and XhoI enzymes for an hour at 37 ℃ and purified by Gel extraction with GeneAid PCR and Gel extraction kit. The purified product was then ligated into pET28b+ restricted with BamHI and XhoI enzymes. The ligated product was then transformed into Novablue super-competent cells by heat shock method. Plasmid was extracted from the culture grown in LB medium overnight. The gene of interest in the plasmid was verified by restriction with BamHI and XhoI enzymes. The plasmids were also submitted for sequencing. The verified plasmid was transformed into BL21(DE3) competent cells by heat shock method.

For Ek/LIC cloning, the PCR product was purified by gel extraction and treated with T4 DNA polymerase in T4 DNA polymerase buffer containing 2.5 mM dATP, 5 mM DTT. The mixture was incubated at room temperature for 30 min followed by incubation at 75 ℃ for 20 min. 5 µl of the reaction mixture was then incubated with pCDF2 vector and 5 mM EDTA for 5min at room temperature. 3 µl of the pCDF2-*Pa*AcpH mix was transformed into NovaBlue super-competent cells by heat shock method. Plasmid was extracted from the overnight culture and verified by restriction with BamHI and XhoI. The plasmids were also submitted for sequencing. The verified plasmid was then transformed into BL21(DE3) competent cells by heat shock method. The genes for the expression of PA2966 (encoded by gene PA2966), PA4266 (1067–1140 amino acids of pyoverdine synthetase pvdD, PA4266), PA1869 (encoded by PA1869) and PA3334 (encoded by PA3334) were obtained from GenScript Corporation (NJ, USA). The genes were restricted and subcloned into pET-28b(+) at the NdeI, XhoI restriction sites. For Ek/LIC cloning, the flanking sites were added onto the genes by PCR. The products were purified and

cloned into pCDF$_2$ by Ek/LIC cloning as described previously. They were then transformed into BL21 (DE3) for protein expression. The cloning of EACP, GPCP, *me*ACP and DACP has been described previously (*151*).

For coexpression of *Pa*AcpH with the PKSs, 2 μl of *Pa*AcpH in pCDF2 vector and 2 μl of CalE8 in pET28b+ or SgcE in pET28b+ were co-transformed into BL21(DE3) by heat shock method.

3.2.4 Expression and Purification of *Pa*AcpH and CPs

Freshly transformed colonies carrying the *Pa*AcpH gene in pET28b+ and pCDF2 vectors were inoculated into 5 ml of LB and grown overnight. It was scaled up to 4 litres of LB with 30 μg/ml kanamycin for *Pa*AcpH-pET28b+ transformed BL21DE3 and 50 μg/ml of Streptomycin for *Pa*AcpH-pCDF2 transformed BL21DE3. After few rounds of optimization, the culture was allowed to grow to an OD of 0.8-1.0 before inducing with 0.3-0.5 mM IPTG. After shaking overnight at 16 ℃ and 160rpm for 16 hours, the cells were harvested by centrifugation and washed with lysis buffer (50 mM Tris (pH 8.0), 500 mM NaCl, 10% glycerol, 0.1 mM DTT). The pellet was re-suspended in the lysis buffer and passed through a microfluidizer twice. The lysate was centrifuged at 20,000 rpm for 45 min and the clear supernatant was incubated with Ni^{2+}-NTA resin for one hour at 4 ℃. After washing, the (His)6-tagged protein was eluted out with an imidazole gradient (50-500 mM). The proteins were further purified by using size-exclusion chromatography with a superdex – 75 column (GE Healthcare). The purified proteins were concentrated using Amicon Centriprep and dialyzed into the buffer that contains 50 mM Tris (pH 8.0), 100 mM NaCl and 10% glycerol for storage.

For the expression and purification of PA2966, PA4226, PA1869, PA3334, two or four liters of LB medium were inoculated and induced with 0.5 mM IPTG at OD 0.6. Cell harvesting, lysis,

affinity chromatography and Size-exclusion chromatography were performed and stored as described for *Pa*AcpH.

For the expression and purification of *Pa*AcpH coexpressed with CalE8 and SgcE, four litres of the culture was induced at ODs varying between 0.6-1.0 with 0.5 mM IPTG. Cell harvesting, lysis, affinity chromatography were performed as described for *Pa*AcpH. The two proteins were further purified and separated by using size-exclusion chromatography with a superdex − 200 column (GE Healthcare).

The recombinant protein expression was monitored by SDS–PAGE carried out using a Mini Protean chamber system (Bio-Rad). Final protein concentrations were measured using the Bradford method.

3.2.5 Circular dichroism spectroscopy analysis

Hemoglobin standard and *Pa*AcpH samples for CD were prepared in a buffer at pH 8.0 containing 50 mM Tris, 250 mM NaCl and 10% glycerol. The CD spectra were recorded using the Chirascan CD spectrometer (Applied Photophysics) between the wavelengths 200 and 300 nm in a quartz cuvette with a path length of 0.1 cm at 25 °C. The spectra were analyzed *(169)* and smoothed using the software Pro-Data Viewer (Applied Photophysics).

3.2.6 Enzymatic activity assay of *Pa*AcpH against 2', 3'-cAMP and bis-p-Npp

The non-specific substrate- bis-p-Npp was treated with *Pa*AcpH in a buffer containing 50 mM Tris, 100 mM NaCl, 10% glycerol, 15 mM $MgCl_2$ and 1 mM $MnCl_2$. The hydrolysis was followed using the Kinetics module in Shimadzu UV-1700 spectrophotometer at 405 nm. The cyclic nucleotide 2', 3' -cAMP hydrolysis by *Pa*AcpH was performed in a buffer containing

100 mM Tris, 20 mM KCl, 25 mM $MgCl_2$ and 0.5 mM $MnCl_2$. A HPLC method developed for the analysis of c-di-GMP phosphodiesterases was adopted for product quantification (*170, 171*). The kinetic parameters were deduced by fitting the initial velocities at various substrate concentrations to the Michaelis–Menten equation using the Prism (GraphPad) software.

3.2.7 Enzymatic activity assay of *Pa*AcpH with carrier protein substrates.

For the preparation of *holo*-CPs, 50–100 μM of the carrier proteins were treated with 10 μM Sfp for 90 min in the presence of 1 mM CoA, 15 mM $MgCl_2$ in assay buffer (50 mM Tris, 100 mM NaCl, 10% glycerol). After the Sfp incubation, 200 μl of the *holo* carrier protein mixture was diluted with buffer (50 mM Tris, 100 mM NaCl, pH 8.0, 10% glycerol) to 1 ml. This mixture was concentrated using Amicon Centriprep concentrator to remove remaining CoA and Mg2+. The procedure was repeated to reach a final volume of 150–200 μl. The *holo*-carrier proteins were then treated with *Pa*AcpH (15 uM) in the reaction buffer that contains 50 mM Tris, 100 mM NaCl, 10% glycerol, 15 mM $MgCl_2$ and 1 mM $MnCl_2$. After incubation at room temperature for 90 min, the reaction mixture, along with the *apo*- and *holo*-CPs obtained from Sfp treatment, were loaded separately onto the Agilent LC1200 HPLC system equipped with a reverse phase HPLC 300SB-C8 Zorbax column (250 x 4.6 mm) for analysis. The mobile phase used was a linear gradient (10–90% ACN) with TFA (0.1%) at the flow rate of 1 ml/min. Time-dependent measurement for PA3334 and PA2966 were performed under the same conditions as described above and samples from various time points were loaded onto the Agilent LC1200 HPLC system equipped with a reverse phase HPLC 300SB-C8 Zorbax column (250 x 4.6 mm) for product quantification.

3.2.8 Release of the products of CalE8 by *Pa*AcpH

The CalE8 and SgcE proteins purified using size-exclusion were examined by spectrophotometry with individually expressed CalE8 and SgcE proteins as controls. The protein samples and controls were prepared in the storage buffer (50 mM Tris, 100 mM NaCl, 10% glycerol) and analyzed by a scanning from 600 nm to 200 nm using the spectrum module in the Shimadzu UV-1700 spectrophotometer.

For the in vitro reaction with *Pa*AcpH, CalE8 was incubated with *Pa*AcpH in the buffer containing 50 mM Tris, 100 mM NaCl, 10% glycerol, 15 mM $MgCl_2$ and 1 mM $MnCl_2$, 0.15 mM acetyl CoA, 1 mM malonyl CoA and 1 mM NADPH. Reactions without *Pa*AcpH or $MnCl_2$ were also included as controls. The reaction was incubated at room temperature for 90 min and stopped by adding TFA. The reaction mixture was extracted by using equal volume of ethyl acetate. The organic phase was dried using a gentle flow of nitrogen gas and re-dissolved in a small volume of methanol. The samples were analyzed using a reverse phase HPLC Eclipse XDB C-18 column and eluted with a linear ACN gradient (0.1% TFA)

The samples prepared in methanol were submitted for analysis using LC-MS. LC-MS was performed by Dr. Li Bin in Asst. Prof. Tang Kai's lab, Nanyang Technological University, Singapore.

3.2.9 Activity of *Pa*AcpH on SgcE

The protein SgcE, the PKS from *Streptomyces globisporus*, was treated with *Pa*AcpH in a buffer containing 50 mM Tris, 100 mM NaCl, 10% glycerol, 15 mM $MgCl_2$ and 1 mM $MnCl_2$. The reaction was incubated at room temperature for 90 min. The samples were analyzed using a

UV spectrophotometer scanning from 200-800 nm. The samples were run in a reverse phase HPLC EclipseXDB C-18 column and eluted with a linear gradient of 100% buffer A (HPLC grade water acidified with 0.05% TFA) to 100% buffer B (100% ACN acidified with 0.05% TFA) in 60 min.

3.3 Results and Discussion

The removal of the 4'-Ppant moiety from CP is catalyzed by the enzyme phosphodiesterase in a divalent cation mediated hydrolysis. This process converts the *holo* CPs to their inactive *apo* forms, which could be regarded as a physiological reclamation of the CPs. The identification, cloning, expression, purification and characterization of the novel phosphodiesterase, *Pa*AcpH, capable of releasing tethered intermediates from stand-alone CPs and CPs within PKSs is described in the following sections.

3.3.1 *Ec*AcpH and *Pa*AcpH

The search for a phosphodiesterase in *M. echinospora* yielded no results. Hence the search was extended to organism of similar origin. Sequence searches displayed that the *PA4353* gene from *P. aeruginosa* could be the potential construct for analysis. The *PA4353* gene of *P. aeruginosa* genome encodes a protein (*Pa*AcpH) homologous to *Ec*AcpH (38% sequence identity and 58% sequence similarity). The predicted Mn^{2+} binding residues, His6, Asp24 and Asp78 along with the other catalytic residues (*162*) of *Ec*AcpH are conserved in *Pa*AcpH (**Fig. 3.2**) thereby confirming it as a phosphodiesterase.
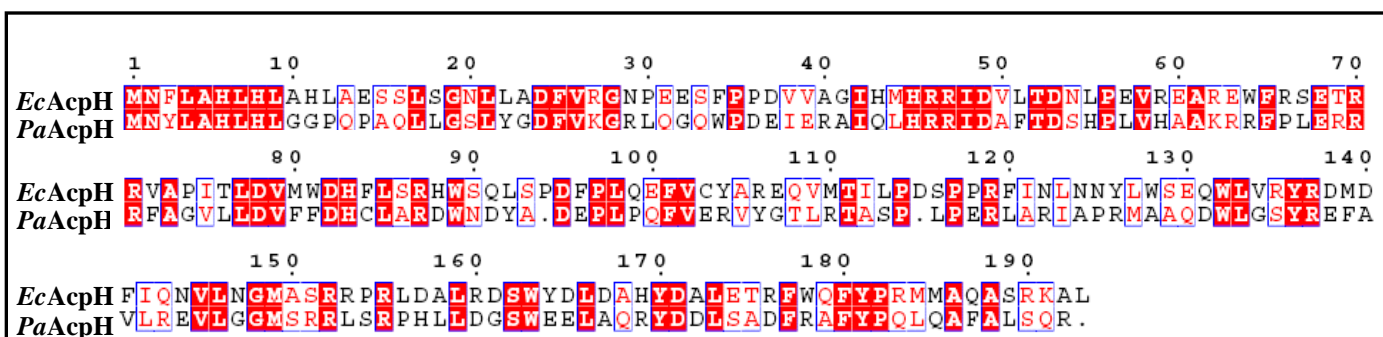
Figure 3.2 Comparison of *Ec*AcpH and *Pa*AcpH proteins. Pairwise sequence alignment of *Ec*AcpH and *Pa*AcpH. The conserved residues are blocked in red. The Mn^{2+} binding residues are marked with red circles.

3.3.2 Cloning, expression, optimization of solubility and purification of *Pa*AcpH

The *PA4353* gene encoding the *Pa*AcpH protein was amplified from the PAO1 genomic DNA by PCR. The PCR product was cloned into pET28b+ and pCDF2 vectors and transformed into BL21(DE3) competent cells. Initial expression of *Pa*AcpH showed a low but significant level of soluble protein. Optimization of expression conditions to improve solubility included 1) Fresh transformation of the BL21(DE3) competent cells with plasmids for expression, 2) Induction at a higher OD (~1), 3) Shaking at a lower rpm (160) and temperature (16°C) after induction, 4) Addition of glycerol in the lysis, washing and elution buffers (10%). After optimization, a significant fraction of the expressed protein was soluble (**Table 3.1, Fig. 3.3**). This was in contrast to *Ec*AcpH which was found in the insoluble fraction. The cells expressing the *Pa*AcpH-pCDF2 plasmid expressed higher levels of soluble protein compared to those expressing *Pa*AcpH-pET28b+ plasmid. Hence the *Pa*AcpH protein from pCDF2 was employed for all the investigations.

It was interesting to note that the solubility of *Pa*AcpH was enhanced when coexpressed with CalE8 or SgcE. It has been reported that solubility of insoluble proteins and less-soluble proteins is enhanced by addition of solubility tags, chaperones and cofactors. CalE8 and SgcE are massive proteins with molecular weights around 200 kDa. When expressed along with either of these proteins, *Pa*AcpH solubility is improved significantly. This could be explained in three ways. 1) The PKS protein with its massive size reduces the probability of interaction between individual *Pa*AcpH molecules thereby reducing the chances of aggregation and precipitation. 2) PKS has a pI of 5.7 at pH 8.0 whereas *Pa*AcpH has a pI of ~7. This relative negativity of PKS also induces electrostatic attraction between the PKS molecules and *Pa*AcpH molecules further reducing interactions between *Pa*AcpH molecules. 3) the *Pa*AcpH could directly bind to CalE8,

and this binding of *Pa*AcpH could prevent it from interacting and aggregating with other *Pa*AcpH molecules.

The soluble *Pa*AcpH protein was purified using Ni-sepharose affinity chromatography followed by size-exclusion chromatography in a Superdex-75 column. The purified protein was stable under the experimental and storage conditions as compared to the refolded *Ec*AcpH which exhibited high levels of aggregation (*163*). Size exclusion chromatography also suggested that a major portion of the soluble *Pa*AcpH was a monomer in solution (**Fig. 3.3 b**). *Pa*AcpH remained stable at concentrations of up to 2 mg/ml as compared to *Ec*AcpH which remained soluble only at low protein concentrations [<10 μg/ml].
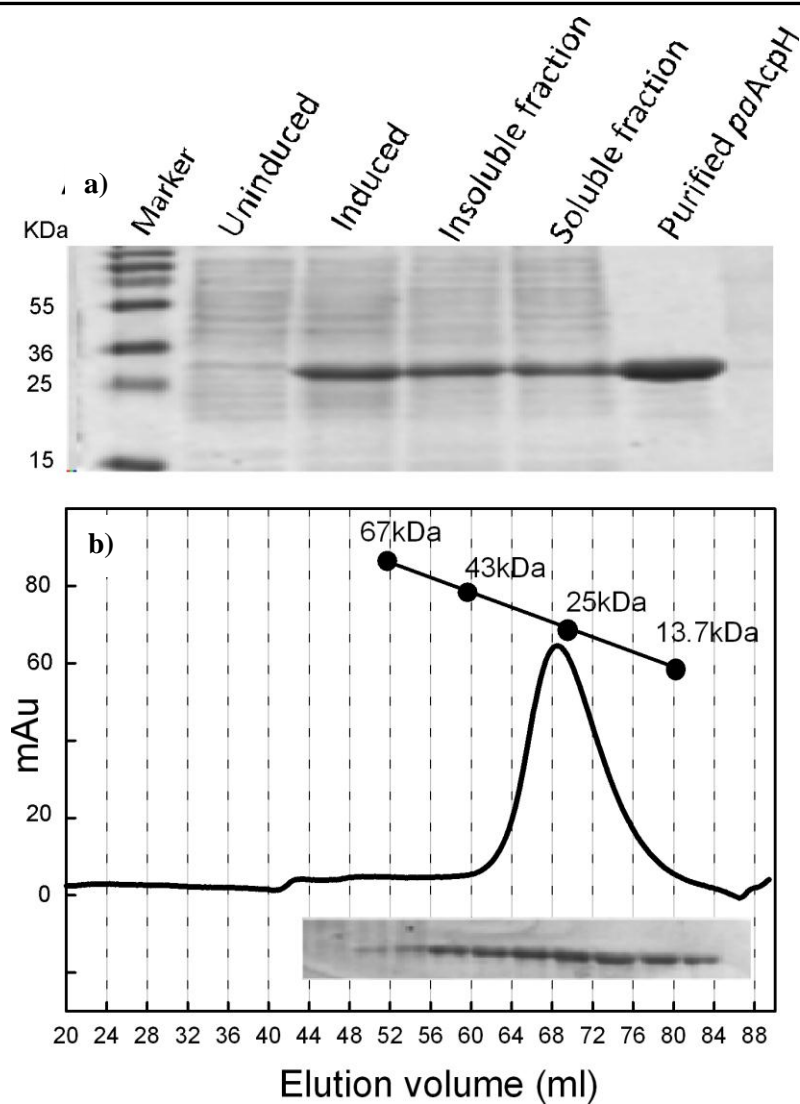
Figure 3.3 Expression, purification and characterization of *Pa*AcpH. a) SDS-PAGE gel analysis of protein expression and purification. b) Size-exclusion chromatography of *Pa*AcpH. Inside panel - SDS-PAGE gel showing *Pa*AcpH fractions purified using size-exclusion chromatography.

Table 3.1 Expression yield of *Pa*AcpH from *E. coli*

Expression and purification of *pa*AcpH from *E. coli*.

Purification step	Total protein (mg)[a]	Target protein (mg)[b]	Yield (%)[d]	Approximate Purity (%)
Soluble cell extract[c]	178	38	100	21
Insoluble protein in cell pellet[b]	160	32	-	20
After Ni$^+$-NTA column	23	19.5	51.2	85
After gel filtration	9.4	8.7	22.8[e]	93

[a] Total protein concentration was measured by Bradford method.

[b] The amount of protein of interest in total protein was determined by quantification of the bands by gel densitometry.

[c] The starting material was 18g of *E.coli* BL21 (DE3) bacterial pellets (wet weight) from a four liter culture.

[d] The purification yield was calculated based on the amount of protein of interest.

e The yield is lowered due to aggregation during protein concentration.

3.3.3 Structural Characterization of *Pa*AcpH

The *Ec*AcpH was proposed to be a helical protein based on the secondary structure prediction and homology modeling on SpoT (*162*). The secondary structure and folding of *Pa*AcpH was verified by circular dichroism (CD) spectroscopy, which showed that it is predominantly a helical protein similar to *Ec*AcpH (**Fig. 3.4**). There was absorption in the buffer at around 200-210 nm, and hence heavy smoothening of the curves was necessary in this region. Hence the coil and strands percentage is not clear.
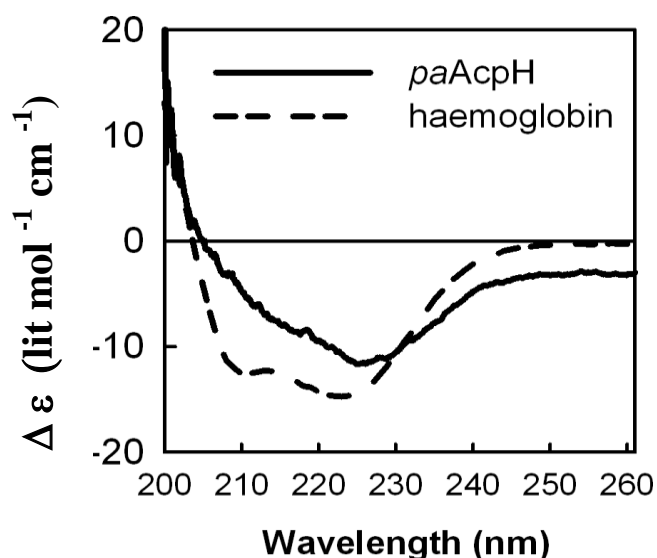
Figure 3.4 Structural integrity of *Pa*AcpH. Circular dichroism (CD) spectra of *Pa*AcpH. Haemoglobin, an all-helical protein was used as a standard for comparison

3.3.4 Confirmation of activity against small molecular substrates

The activity of soluble *Pa*AcpH was first tested with several phosphodiester containing small-molecules. *Pa*AcpH efficiently hydrolyzes the two substrates bis-p-nitrophenyl ester (Bis-

p-NPP) and thymidine-5-monophosphate-p-nitrophenyl ester (thymidine-p-NPP). In addition, it also hydrolyzes the cyclic nucleotide 2, 3'-cAMP to generate 3'-AMP (**Fig. 3.5**).

Figure 3.5 Confirmation of *Pa*AcpH activity. a. HPLC analysis of the hydrolysis of 2', 3'-cAMP by *Pa*AcpH. Trace 1 and 2 were for the standards. Trace 3 was from the incubation of 2', 3'-cAMP with *Pa*AcpH and Mn^{2+}. b) Kinetic analysis of the hydrolysis of 1. bis-pNPP and 2. 2', 3'-cAMP by *Pa*AcpH.
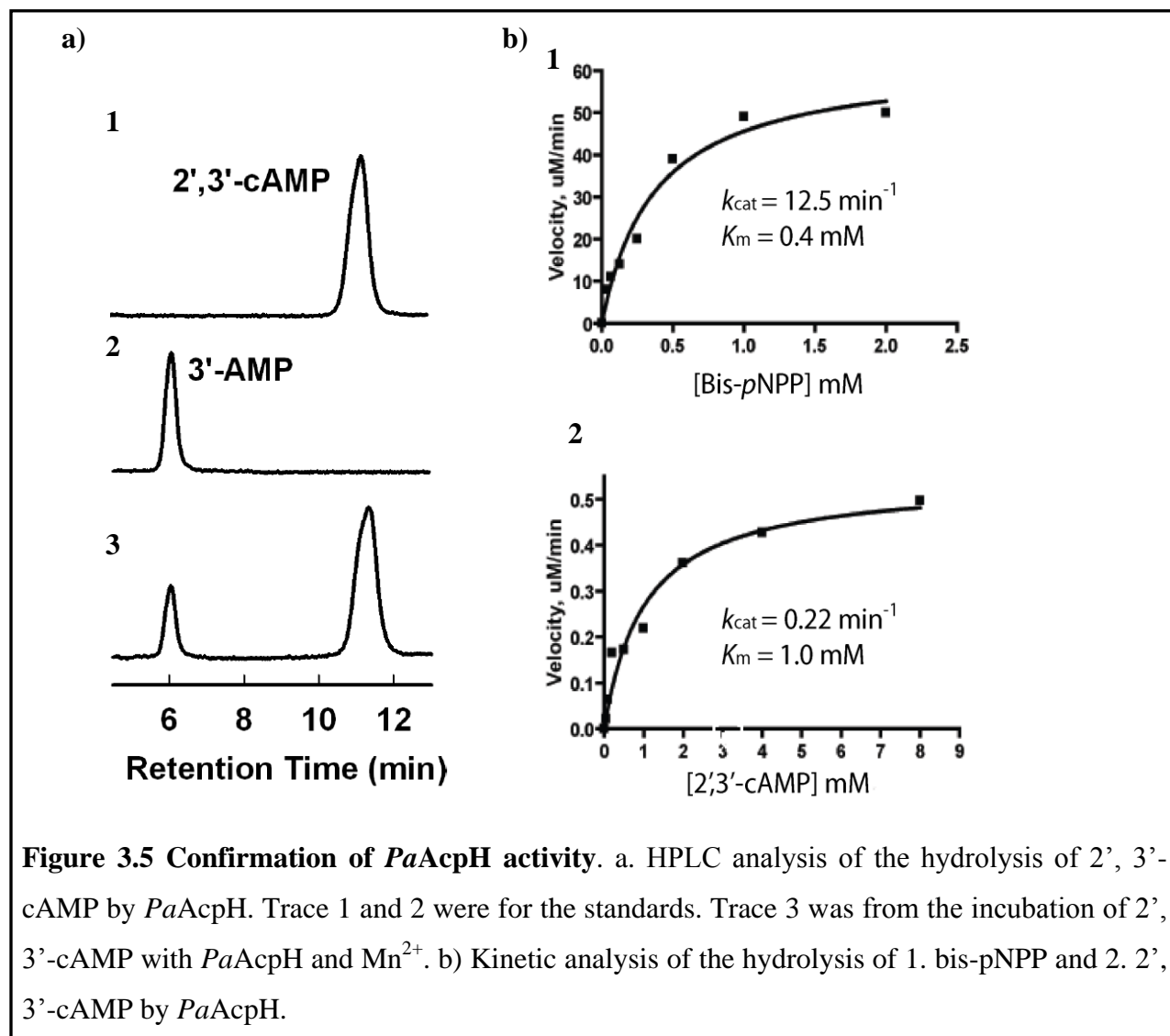
3.3.5 Substrate specificity of *Pa*AcpH towards acyl carrier proteins from *P. aeruginosa*.

It has been shown that *Ec*AcpH can hydrolyze the 4'-phosphopantetheinyl group from the ACPs from the fatty acid synthesis pathways of *E. coli, B. subtilis* and *A. arolicus* (*162, 163*). But whether the AcpH can discriminate the CP of fatty acid synthesis pathway from the CPs of secondary metabolic pathways remains to be validated. To test this, we cloned and expressed

101

four potential cognate ACP substrates of *Pa*AcpH from *P. aeruginosa* PAO-1. The four ACPs include two ACP proteins (PA2966 and PA1869 (also AcpP2)) from the fatty acid or lipid biosynthetic pathway, and two putative PCP domain proteins (PA4226 and PA3334) from the biosynthetic pathways of the pyocheline and another unknown secondary metabolite (*172, 173*). The four CPs were expressed and purified as His-tagged proteins. In order to distinguish the *apo* and *holo* forms of the CPs the modification (~358Da) has to be examined and hence the HPLC method described in Chapter 2 was utilized. The four *apo* CPs were first converted to *holo*-CPs by treating with Sfp, CoA and Mg^{2+}. The Sfp modified proteins were analyzed using HPLC and it revealed that Sfp modified the four CPs with different efficiencies. PA1869 and PA4226 were completely modified whereas PA2966 and PA3334 were only partially modified (**Fig. 3.6**). Following Sfp modification the reaction mixtures were incubated with *Pa*AcpH and Mn^{2+}.

Figure 3.6 Enzymatic activity assays with the endogenous carrier proteins (CPs) from
P. aeruginosa. The HPLC traces of the *apo*-CPs are shown as dashed lines. The HPLC traces of the CPs after Sfp treatment are shown as blue lines; whereas the HPLC traces of the CPs after Sfp and *Pa*AcpH treatment are shown as red lines. PA2966, PA4226, PA1869, PA3334 are from *P. aeruginosa*.
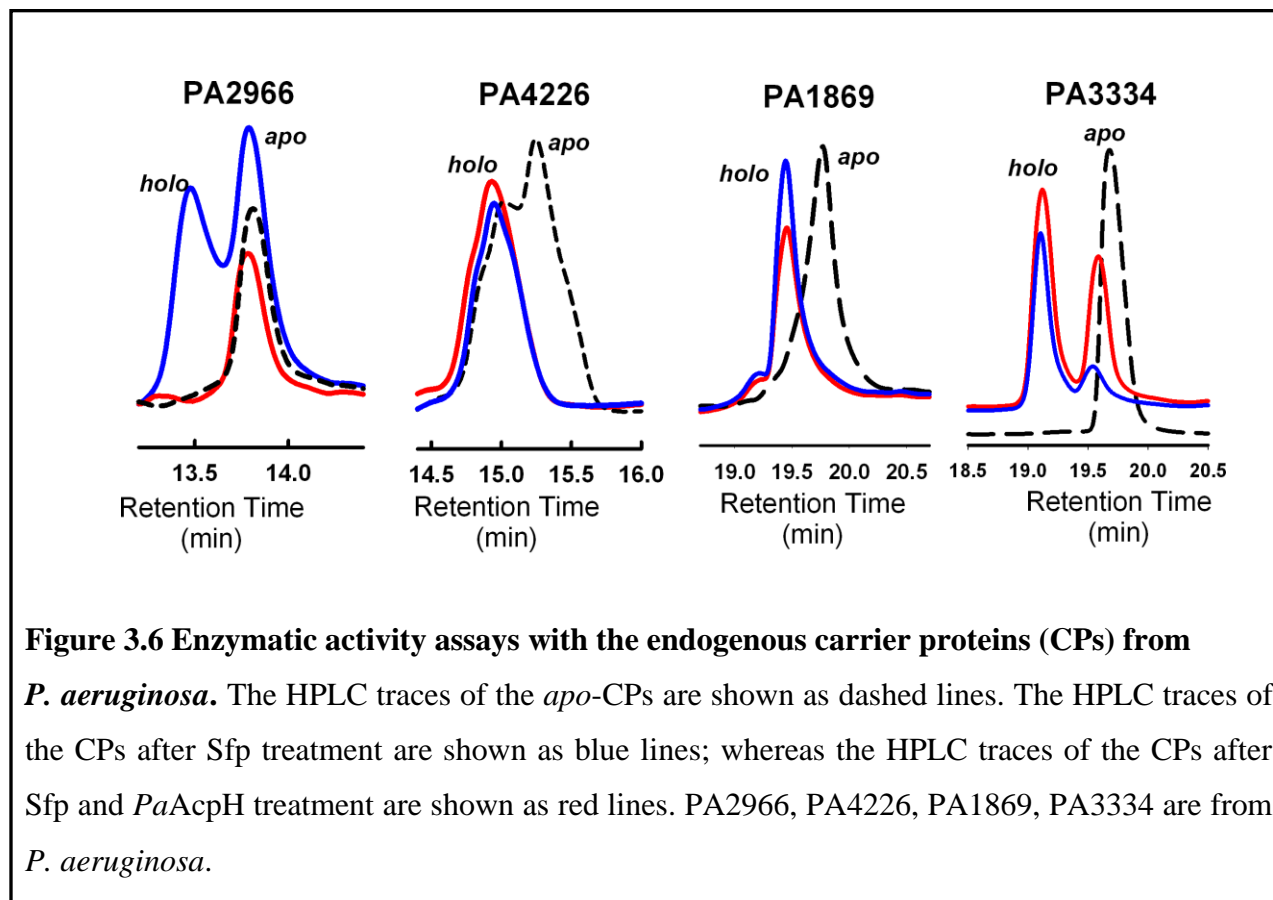
Distinctive differences between the CPs were observed. The *holo*-PA2966 was converted completely back to its *apo*-form within 20 min, whereas PA1869 and PA4226 remained in their *holo* forms even after prolonged incubation. Partial conversion was observed for PA3334 with both *apo* and *holo* forms present after the incubation. The conversion between the *apo* and *holo*-CPs was confirmed by MALDI mass spectrometry analysis. The difference in the observed mass for the *apo* (m/e 10,585.8) and *holo* (m/e: 10944.2) is 358.4, corresponding to the mass of the phosphopantetheinyl group (358.1 Da).

Kinetic measurement further revealed the reaction with PA2966 is significantly faster than that of PA3334, indicating that *Pa*AcpH prefers the ACP protein (PA2966) from the main fatty acid synthesis pathway **(Fig. 3.7).** The results suggested that *Pa*AcpH prefers the ACP protein (PA2966) from the main fatty acid synthesis pathway. Although PA1869 shares significant homology (53% identity and 72% similarity) with PA2966 **(Fig. 3.7)**, it remains to be a poor substrate for *Pa*AcpH. The function of PA1869 remains unknown even though it is a putative ACP involved in fatty acid or lipid biosynthesis based on its gene context.
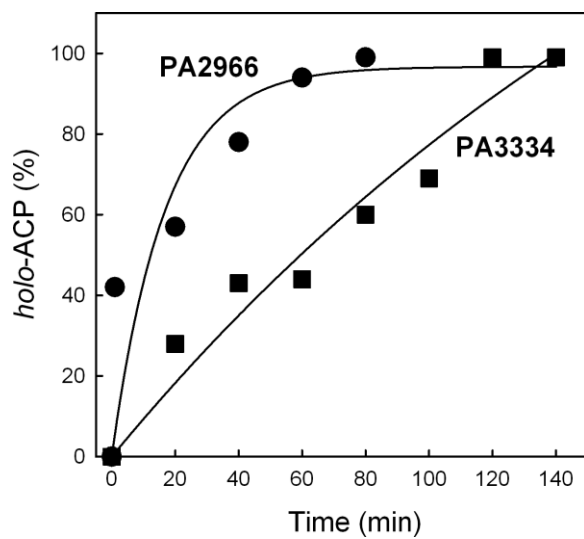
Figure 3.7 Activity assay – Kinetics. Kinetic analysis of the de-phosphopantetheinylation of *holo*-PA2966 and *holo*-PA3334 by *Pa*AcpH.

3.3.6 Substrate specificity of *Pa*AcpH towards heterologous acyl carrier proteins

We examined the substrate specificity of *Pa*AcpH towards various CPs that included four heterologous CPs from the primary and secondary metabolic pathways of other microorganisms. The four CPs include an ACP protein (EACP) from the fatty acid synthesis pathway of *E. coli* and the CPs from three secondary metabolic pathways, including the ACP domain of the type I PKS DEBS (DACP) from *S. erythraea* (*97*), the ACP domain of the iterative PKS CalE8 in calicheamicin biosynthesis from *M. echinospora ssp.* (*me*ACP) and the PCP domain of NRPS GrsA from *B. brevis* (GPCP) (*89, 97, 159*). After the treatment of the CPs with Sfp, complete conversion of *holo*-CPs was observed for all the CPs with the exception of *me*ACP (**Fig. 3.8**). Only a portion of *me*ACP was modified, which indicated a low modification efficiency of *me*ACP by Sfp (*151*). The subsequent incubation of the *holo*-CPs with *Pa*AcpH and Mn^{2+} revealed that complete conversion of *holo* to *apo*-CP for EACP and GPCP, but only partial

conversion for *me*ACP as judged by the *holo* to *apo* ratio. DACP remained in its *holo*-form after prolonged incubation, indicating DACP is a poor substrate for *Pa*AcpH. The preference for the *holo*-EACP is not surprising given the high sequence identity (88%) between EACP and PA2966 (**Fig. 3.9**). However, the rapid conversion of *holo*-GPCP to *apo*-GPCP was unexpected, considering that GPCP is from a secondary metabolic pathway. However, a structural phylogenetic analysis of the carrier proteins revealed that the GPCP from *B. brevis* was phylogenetically closer to the *me*ACP protein. This explains why GPCP, though from a secondary metabolite pathway, was rapidly modified by *Pa*AcpH.

Figure 3.8 Enzymatic activity assays with the four heterologous CPs. The HPLC traces of the *apo*-CPs are shown as dashed lines. The HPLC traces of the CPs after Sfp treatment are shown as blue lines; whereas the HPLC traces of the CPs after Sfp and *Pa*AcpH treatment are shown as red lines. DACP is from the module 2 of the modular PKS DEBS in *S. erythrea*; EACP is from the fatty acid synthesis pathway of *E. coli*; GPCP from the NRPS for gramicidin A biosynthesis of *B. brevis*; *me*ACP from the iterative PKS CalE8 in calicheamicin biosynthesis.

Together, these observations indicated that although the cognate substrate of *Pa*AcpH in *P. aeruginosa* is likely to be the ACP (PA2966) from the fatty acid synthesis pathway, it exhibits relaxed substrate specificity towards some heterologous CP domains. Given the lack of structural information for *Pa*AcpH, the molecular basis for the substrate specificity remains to be elucidated
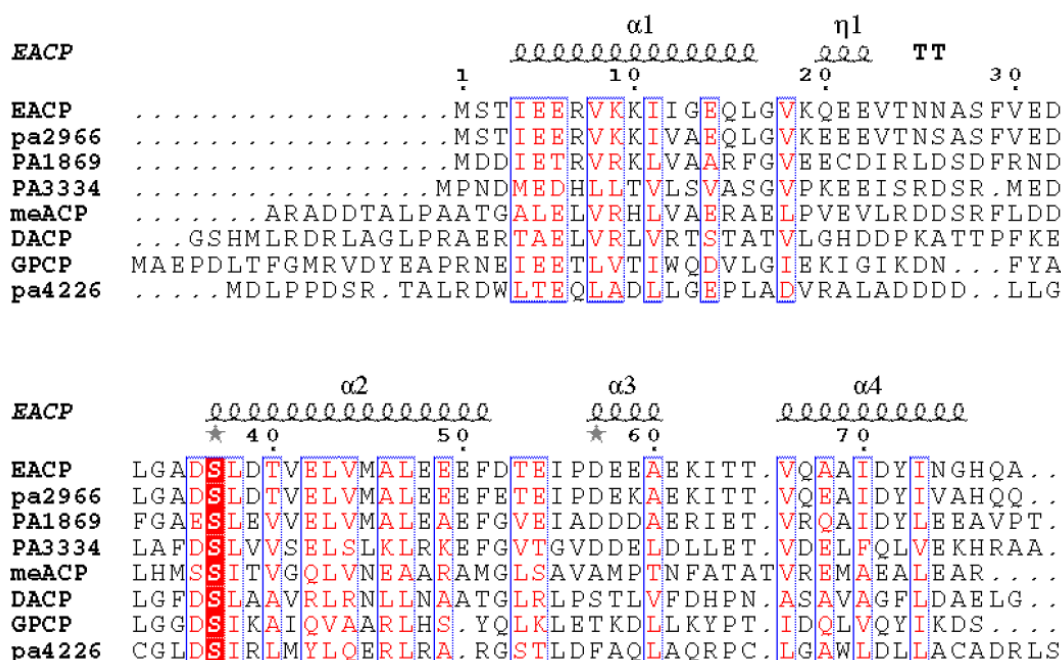
Figure 3.9 Sequence alignment of the eight carrier proteins (CPs). The modification site (serine) and conserved residues were shaded. (PA2966, PA4226, PA1869, PA3334 from *P. aeruginosa*. DACP from the module 2 of the modular PKS DEBS in *S. erythrea*; EACP from the fatty acid synthesis pathway of *E. coli*; GPCP from the NRPS for gramicidin A biosynthesis of *B. brevis*)

3.3.7 Release of the 4'PP-linked products from the polyketide synthase CalE8 by *Pa*AcpH.

PKS and NRPS produced structurally diverse products by forming a series of covalently tethered intermediates during the chain-extension processes (*174*). The observed substrate specificity towards *me*ACP for *Pa*AcpH prompted us to test whether *Pa*AcpH can act on the *me*ACP domain within the PKS complex. The *me*ACP is part of the type I iterative PKS (CalE8) in calicheamicin biosynthesis (*59*). Recent biochemical studies have suggested that CalE8 is the protein that initiates the biosynthesis of 10-membered enediyne (*149, 175*). With an intrinsic PPTase domain, the multi-domain CalE8 produces sixteen carbon-long polyene that is covalently tethered to the *me*ACP domain (*60, 149, 154*). *Pa*AcpH was expected to release the sixteen carbon-long polyene and other premature (or errant) products along with the phosphopantetheinyl group. Indeed, incubation of CalE8, *Pa*AcpH, Mn^{2+}, Mg^{2+}, CoA and the substrates malonyl CoA and NADPH produced a series of products (1-5) with absorbance in the range of 350-450 nm (**Fig. 3.10**). The control reactions without either *Pa*AcpH or Mn^{2+} did not produce these products. Importantly, most of the products generated from the incubation exhibited shorter retention time relative to that of the product produced from the CalE8/CalE7 reaction. The absorption spectrum and retention times suggested that they are most likely phosphopantetheinyl group-tethered products with various chain lengths. The observed multiple products indicated that *Pa*AcpH stalled the chain extension process at various stages by hydrolyzing the phosphodiester bond.
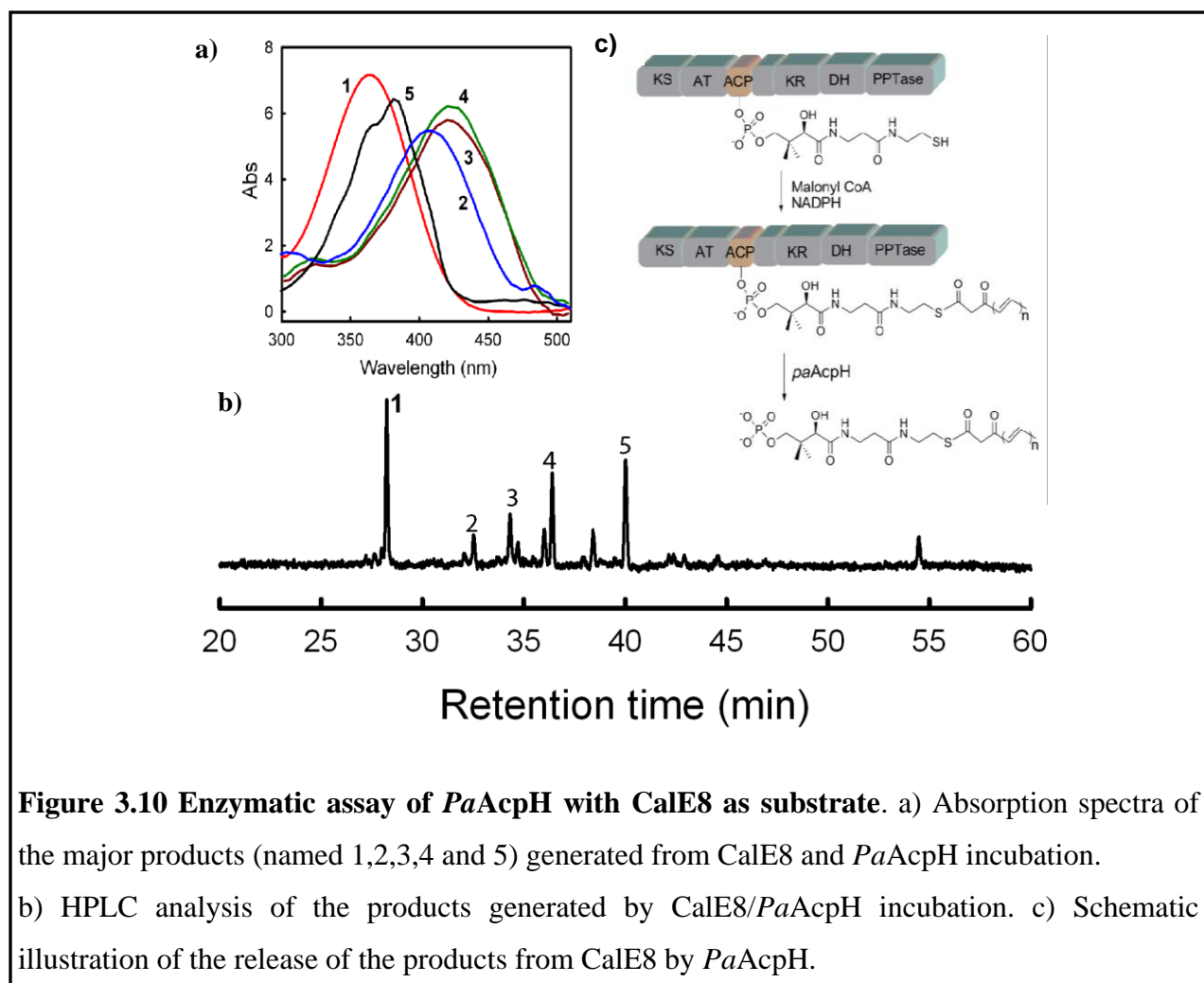
Figure 3.10 Enzymatic assay of *Pa*AcpH with CalE8 as substrate. a) Absorption spectra of the major products (named 1,2,3,4 and 5) generated from CalE8 and *Pa*AcpH incubation.
b) HPLC analysis of the products generated by CalE8/*Pa*AcpH incubation. c) Schematic illustration of the release of the products from CalE8 by *Pa*AcpH.

Another member of the PKS family, SgcE treated with *Pa*AcpH also showed interesting results. 50 μM SgcE treated with 50 μM *Pa*AcpH showed a new peak at 42 min at 400 nm (**Fig. 3.11**). The peak was however not observed when PKS concentration was increased without changing the concentration of *Pa*AcpH. This confirms that the peak was released as a result *Pa*AcpH activity on SgcE. Also, the absorption spectra of SgcE and SgcE treated with *Pa*AcpH show a significance difference at 400 nm. The growing chain intermediate on SgcE has an absorbance at 400 nm, and the observed differences in the spectra, suggests that *Pa*AcpH was able to cleave the growing intermediate from SgcE.
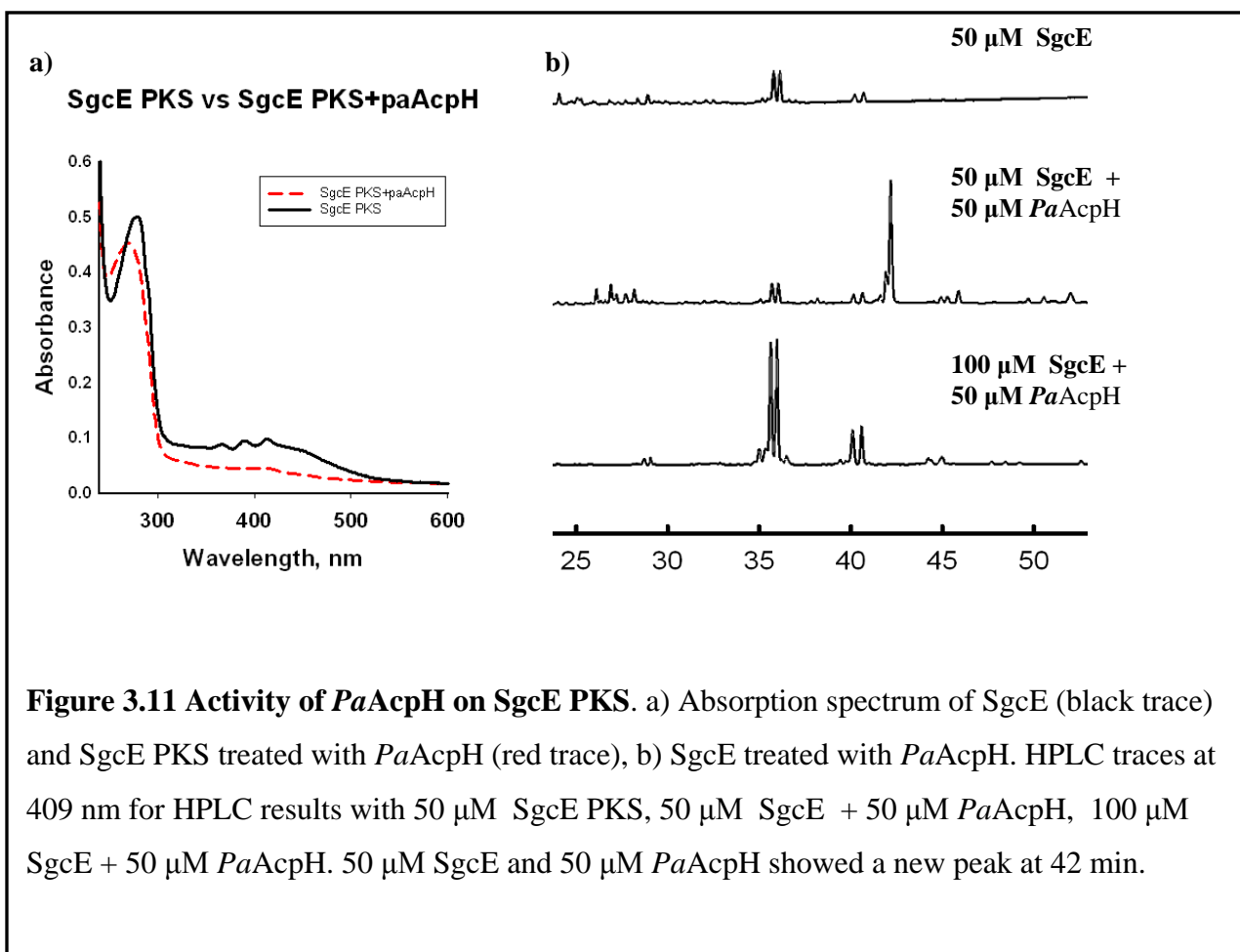
Figure 3.11 Activity of *Pa*AcpH on SgcE PKS. a) Absorption spectrum of SgcE (black trace) and SgcE PKS treated with *Pa*AcpH (red trace), b) SgcE treated with *Pa*AcpH. HPLC traces at 409 nm for HPLC results with 50 μM SgcE PKS, 50 μM SgcE + 50 μM *Pa*AcpH, 100 μM SgcE + 50 μM *Pa*AcpH. 50 μM SgcE and 50 μM *Pa*AcpH showed a new peak at 42 min.

Thomas and coworkers have also showed that the *Ec*AcpH could hydrolyze the stand-alone acyl-ACP with various fatty acyl chain lengths (C-6 to C-16) (*162, 163*). The observation that *Pa*AcpH exhibited substrate preference towards the ACP (PA2966) from the fatty acid synthesis pathway suggests that the enzyme may play a role regulating the dynamic fatty acid and phospholipids biosynthesis in *P. aeruginosa*. Moreover, *P. aeruginosa* is known to utilize auto-inducers such as N-acyl homoserine lactone (AHL) for cell-cell communication. The two major AHL-producing systems requires ACPs for the production of the auto-inducers N-butyryl homoserine lactone (C4-HSL) and N-(3- oxo)-dodecanoyl homoserine lactone (3-oxo-C12-HSL).

Thus, it is also conceivable that *Pa*AcpH may play a regulatory role in the production of the auto-inducers for quorum sensing.

3.3.8 Conclusions

In this project, the *Pa*AcpH protein was purified and expressed as a soluble protein that remained stable without aggregations at working concentrations (2 mg/ml). Cleavage of the phosphopantetheine moiety from the CP within the PKS by PaAcpH further means that a growing polyketide on the ACP within a PKS could also be released by the phosphodiesterase activity. The ability of *Pa*AcpH to release the tethered intermediates from acylated CPs within multidomain PKSs was also confirmed. The growing polyketide intermediates tethered on the carrier protein within a polyketide synthase form the templates of commercially and medicinally important antibiotics like erythromycin and calicheamicin and also physiologically important non-ribosomal peptides. Release of these intermediates from the polyketide synthase would be extremely useful in examination of the products, studying the mechanism, and to release commercially viable products physiologically. Released as phosphopantetheine linked products after *Pa*AcpH cleavage, these products could be purified easily. *Pa*AcpH could further be optimized to physiologically release these products in vivo through co-expression. These results suggest that the *Pa*AcpH protein, after optimization, can be efficiently used to release medicinally and physiologically significant products.

Chapter 4. Characterization of the core and modifying domains of CalE8

4.1 Introduction

The CalE8 is a multidomain iterative PKS, with the domains conforming to a unique arrangement to perform the stepwise polyketide biosynthesis. Each domain is associated with a unique function, and a cooperative functioning of all the domains is essential for the polyketide synthase to be fully functional. Of the six co-operative domains in PKS, the KS, AT and ACP domains are known as the 'core domains'. The AT domain determines the selectivity of the substrates to be attached to the 4'-Ppant moiety on the ACP domain. The KS domain is believed to determine the chain length and catalyze the decarboxylation reaction. While the PPTase domain is involved in modifying the ACP, the KR and DH domains are involved in modifying the growing polyketide. Hence the PPTase, KR and DH domains are known as the 'modifying domains'. In CalE8, apart from the core and modifying domains, there are two more domains which have no assigned function (**Fig. 4.1**). The first domain is placed between ACP and KR domains and the second is placed between the DH and the PPTase domains. The second unknown or linker domain was found to aid in the solubility of PPTase (*151*). Demarcation of the boundaries and characterization of these domains is imperative to understand the architecture and functional mechanism of CalE8.

Figure 4.1 Domain boundaries of ACP within the proposed domain architecture of *CalE8*. The suggested domain boundaries defining ACP based on which the domain was cloned and expressed.
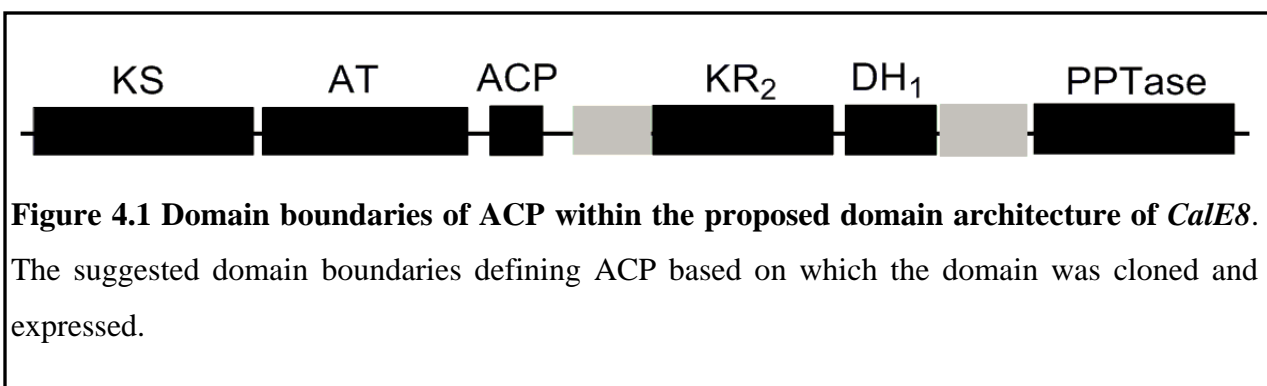
The *me*ACP and the PPTase domains were expressed as soluble proteins, and were purified and functionally characterized in this project, as discussed in Chapter 2. This chapter discusses the cloning, expression and investigation of the function of the other four domains. While the AT$_{CalE8}$ domain was cloned and expressed as a functional domain, the other domains could not be expressed as soluble proteins. All the domains were structurally modelled based on homology and examined at the primary, secondary and tertiary structural levels. These *in silico* structural and functional characterizations of the domains have aided in validating previous hypotheses and have also revealed novel information on these domains.

The region spanning 460 amino acids (1-459) in the N-terminus of the CalE8 PKS has been proposed to be the KS domain (KS$_{CALE8}$). Attempts were made to express the KS domain as a stand-alone protein. The attributes of KS$_{CALE8}$ domain was examined using sequence and phylogenetic analyses. Structural modeling was done based on homology and the structures were examined. The results unveiled information on the KS$_{CALE8}$ domain and the similarity it shares with KS domains from modular PKSs.

The 424 residue region (482-905) in CalE8 located 22 residues downstream of the KS$_{CALE8}$ domain has been proposed to be the AT domain (AT$_{CalE8}$). Following successful expression of the domain as a stand-alone protein, and structural validity of the domain was analyzed using CD spectroscopy. The AT$_{CalE8}$ domain was examined for enzyme activity primarily to study its preference on acetyl CoA and malonyl CoA substrates. Following sequence and phylogenetic studies, structural modeling was done based on homology and the structures were examined. The observations provided crucial information on the AT$_{CALE8}$ domain, its resemblance to the malonyl CoA transferases and its substrate preference.

The region in CalE8, that lies 137 residues downstream of the *me*ACP domain, spanning 261 residues, has been proposed to be the KR domain (KR$_{CALE8}$). The KR$_{CALE8}$ domain was subjected to sequence and phylogenetic examinations. Structural modeling was done based on homology and the structures were examined. The results provided data on the KR$_{CALE8}$ domain and its limited similarity to KR domains from modular DEBS PKS. The 121 residue region between *me*ACP and KR$_{CALE8}$, the first linker domain of unknown function was investigated employing *in silico* techniques of sequence and phylogenetic analyses and structural modeling based on homology. Structural analysis provided considerable evidence on the linker domain upstream of the KR$_{CALE8}$ towards assigning this domain as the extension of the KR domain.

The 144 residue region downstream of the KR$_{CALE8}$ domain of the CalE8 has been proposed to be the DH domain (DH$_{CALE8}$). The domain was subjected to sequence and phylogenetic investigations and characterized using *in silico* modeling studies. The 111 residue region between DH$_{CALE8}$ and CalE8-PPTase domain or the second linker domain of unknown function was examined. It was previously found that this domain was necessary to express CalE8-PPTase as a soluble protein. However, results obtained from structural modeling and related analyses revealed vital information to analyze this domain as an extension of the DH domain.

This chapter describes the cloning and expression of the KS, AT, KR and DH domains. *In silico* examinations using sequence and phylogenetic analysis and structural characterization by homology modeling of the above mentioned domains, along with the two domains of unknown function, were carried out to assign the respective domain boundaries within CalE8.

4.2 Materials and methods

4.2.1 Materials

Coenzyme A and other chemicals were purchased from Sigma. Restriction enzymes and ligation kits were purchased from New England BioLabs, Promega and Fermentas. PCR reagents were bought from Roche.

4.2.2 Generation of the plasmids encoding AT$_{CALE8}$, KS$_{CALE8}$, KR$_{CALE8}$, DH$_{CALE8}$ domains by PCR

The boundaries flanking the gene segments coding for the AT$_{CALE8}$, KS$_{CALE8}$, KR$_{CALE8}$, DH$_{CALE8}$ domains were obtained from literature. The protein products of these gene segments were examined *in silico*. A new set of gene segments coding for the above mentioned domains were constructed by altering their boundaries to preserve the integrity of secondary structural elements which were disrupted in the domain segments proposed in the previously published work (*68*). Two sets of primers were designed for each domain, one set carrying NdeI and XhoI restriction sites on the forward and reverse primers respectively for cloning into pET28b+ vector and another set carrying the Ek/LIC cloning sequence for cloning into pCDF2 vector. The *CalE8* gene in pUC57 vector was used as template and the amplicons carrying the gene fragments coding the entire domains were generated by PCR using Expand High Fidelity PCR system (Roche). The amplicons were examined using agarose gel electrophoresis and the integrity of the gene segments were confirmed by nucleotide sequencing.

4.2.3 Cloning and preliminary examination of expression

For cloning into the pET28b+ vector, the PCR products were digested using NdeI and XhoI restriction enzymes. The pET28b+ vector was also simultaneously digested with NdeI and XhoI. After purification to get rid of the enzymes and other additives, the vector and insert were mixed in a molar ratio of 3:2 and ligated in a ligation mix that contained 2X rapid ligation buffer and T4 DNA ligase (Promega). The PCR products with flanking Ek/LIC cloning sequences were cloned into pCDF2 vector by Ek/LIC cloning. Both the plasmids were transformed into NovaBlue sub-cloning competent cells (Novagen) and selected based on their resistance to kanamycin for genes in pET28b+ vectors and resistance to streptomycin for genes in pCDF2 vectors. Following selection, the colonies from the plates were inoculated into 5 ml LB media with corresponding antibiotics. Plasmids were extracted from the cultures that were grown overnight, using Bio-Rad Plasmid mini-prep kit. The integrities of the plasmids were validated by restriction screening and nucleotide sequencing. The plasmids were then transformed into BL21 (DE3) competent cells and plated in LB-Agar plates containing corresponding antibiotics. Colonies were picked and grown in LB media with kanamycin or streptomycin. From the overnight cultures, 5 ml was used to inoculate 50 ml of LB media containing kanamycin or streptomycin. After an OD_{600nm} of 0.6 was reached, the cultures were induced with 0.5 mM IPTG. Samples were collected at 30 min, 1, 2, 4 and 16 hours. These samples were screened for the expression of the recombinant protein.

4.2.4 Purification of AT_{CalE8}

The clone of AT_{CALE8} domain that showed good expression was selected and grown overnight. This culture was then scaled up to 4-5 liters and induced with 0.5 mM IPTG once the OD_{600nm} reached 0.6. The induced culture was grown at 16°C for 16-18 hours. The culture was

centrifuged using a JA-10 rotor at 8000 rpm for 10 min. The pellet was washed with wash buffer (50 mM Tris, 500 mM NaCl, 5% glycerol, pH 8.0) and resuspended in the same buffer. The resuspended pellet was first lysed by sonication at 35% amplitude, 5 seconds pulse with 2 seconds pause for 5-10 min. They were further lysed using a microfluidizer. The lysed cells were centrifuged for 45 min at 20000 rpm using a JA-25.5 rotor. The supernatant was run on an SDS-PAGE to examine the expression and solubility of the AT_{CALE8} domain. The lysate-supernatant was mixed with Ni-NTA resin (GE Healthcare) and the mixture was incubated for an hour. It was then added in streamline into a column and washed with 20 times the column volume using the wash buffer containing 10 mM imidazole. Stringent washes with 10-15 column volumes of 20, 40 mM imidazole were done and batch elutions with 75mM imidazole, 100 mM imidazole and 200 mM imidazole were performed twice each. The eluted fractions were run on an SDS-PAGE gel to assess the amount and purity of the protein.

4.2.5 FPLC purification

The AT_{CALE8} domain obtained from affinity purification was further purified using a size-exclusion using a Superdex-75 column in a buffer containing 50 mM Tris, 100 mM NaCl, and 10% glycerol at pH 8.0. The eluted fractions from FPLC were concentrated using Amicon Centriprep concentrators by centrifuging thrice at 2000 rpm for 5 min. The protein concentration was estimated by Bradford method and the protein was stored in a buffer containing 50 mM Tris, 100 mM NaCl, and 20% glycerol at pH 8.0.

4.2.6 Circular dichroism spectroscopy

Circular dichroism spectroscopy was performed with the purified protein in a buffer containing 50 mM Tris, 100 mM NaCl, 5% glycerol at pH 8.0. A myoglobin standard was simultaneously run as a control.

4.2.7 Modification of *me*ACP by Sfp

The *apo* ACP was converted to *holo* ACP by treating with 2-4 uM of Sfp in Tris buffer containing 50 mM Tris, 100 mM NaCl and 5% glycerol at pH 8.0. The reaction was carried out in the presence of 15mM $MgCl_2$ with and without 1mM CoA. *Holo* ACP was purified from 500ul of the reaction mixture containing Sfp and CoA by FPLC (Superdex-75 column). HPLC was run with the sample to validate complete conversion.

4.2.8 Modification of *holo-me*ACP by AT_{CALE8}

The Sfp-modified ACP was treated with 0.5 mM acetyl CoA (and malonyl CoA) and 5uM AT_{CALE8}, and the samples were incubated at 30 C for 60 min. The mixture was run on HPLC – C8-300SB Zorbax column.

4.2.9 Bioinformatics analysis and Modeling

Sequence and secondary structural analysis for AT_{CALE8}, KS_{CALE8} and KR_{CALE8} and DH_{CALE8} domains were done using CLUSTALW and Multalin. Homology models for the domains were built using the Swiss-modeller server.

4.3 Results and Discussion

4.3.1 Expression and examination of solubility of the domains

In order to examine the individual domains of CalE8, attempts were made to express the domains as individual proteins. The gene fragments encoding the proposed KS_{CALE8}, AT_{CALE8}, KR_{CALE8} and DH_{CALE8} domains were generated by PCR. These amplicons were cloned into pET28b+ and pCDF2 vectors and transformed into *E. coli* BL21 (DE3) cells for protein expression. While AT_{CALE8} was expressed at relatively high levels, KS_{CALE8}, KR_{CALE8} and DH_{CALE8} domains were expressed in smaller amounts. In addition, the AT_{CALE8} domain was expressed as a soluble protein, but the KS_{CALE8}, KR_{CALE8} and DH_{CALE8} domains were found in the insoluble fraction. Though various methods of optimization of expression and induction conditions were screened, the solubility of the KS_{CALE8}, KR_{CALE8} and DH_{CALE8} domains showed little or no improvement.

4.3.2 Purification of the stand-alone AT domain

The AT_{CALE8} domain (from 483 to 906, a region spanning 424 aa) was expressed in high quantities as a soluble protein. An earlier attempt to express the protein in a phosphate buffer at pH 7.2 produced largely insoluble fractions of the domain. After redefining the boundaries to include 4 amino acids downstream and expressing the protein in a Tris buffer at pH 8.0, we were able to produce AT_{CALE8} in its soluble form. The expression conditions, time of induction and incubation times were optimized to attain a higher yield of the soluble protein (~20 mg protein from 2 litres). The protein was purified by Ni-sepharose affinity chromatography followed by size exclusion chromatography using a superdex-75 gel-filtration column. The FPLC chromatogram (**Fig. 4.2**) showed that the AT_{CALE8} domain was expressed as a monomer. A peak

corresponding to the dimer fraction was also observed. The eluted fractions were run on an SDS-PAGE to confirm the presence of the protein.
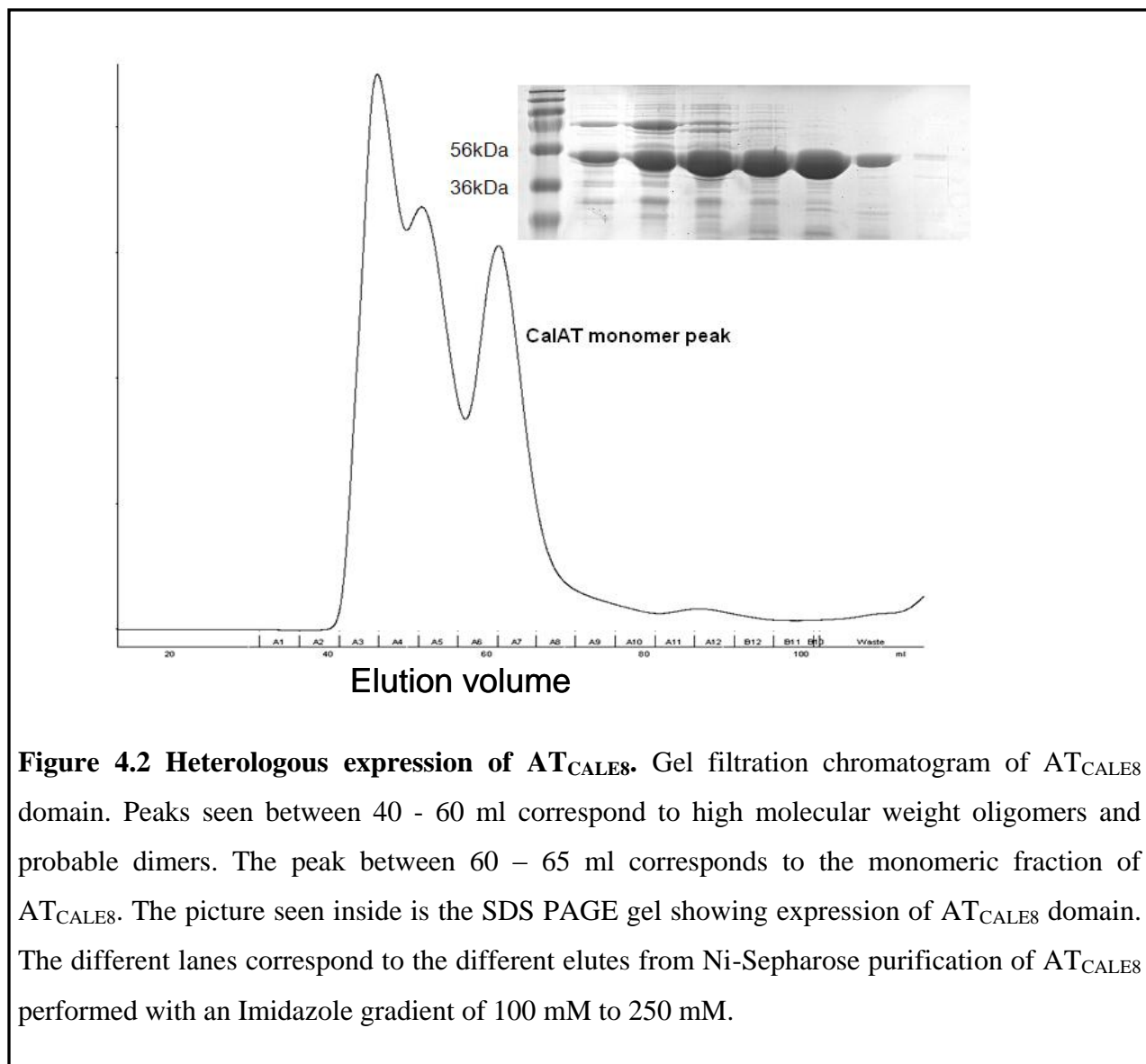
Figure 4.2 Heterologous expression of AT$_{CALE8}$. Gel filtration chromatogram of AT$_{CALE8}$ domain. Peaks seen between 40 - 60 ml correspond to high molecular weight oligomers and probable dimers. The peak between 60 – 65 ml corresponds to the monomeric fraction of AT$_{CALE8}$. The picture seen inside is the SDS PAGE gel showing expression of AT$_{CALE8}$ domain. The different lanes correspond to the different elutes from Ni-Sepharose purification of AT$_{CALE8}$ performed with an Imidazole gradient of 100 mM to 250 mM.

4.3.3 Validation of structural integrity – Circular Dichroism (CD) Spectroscopy

The integrity of the secondary structural elements and hence the folding of the AT$_{CALE8}$ domain were confirmed by CD spectroscopy (**Fig. 4.3**).
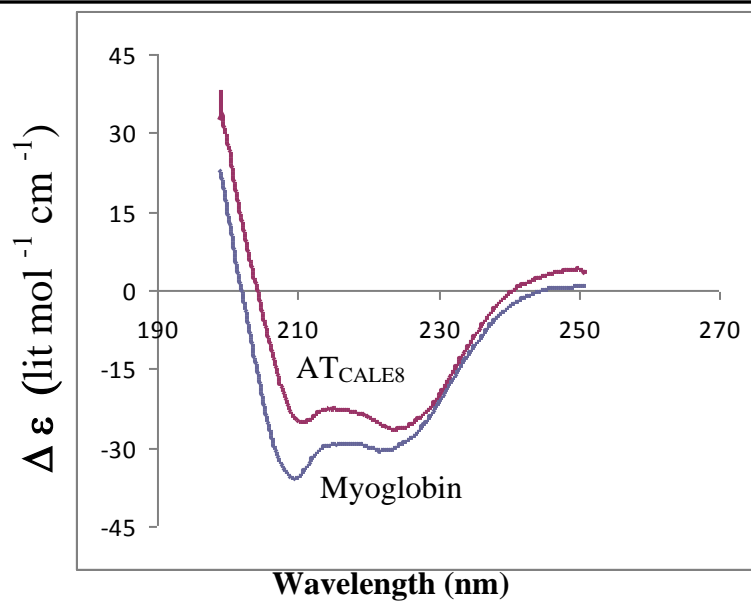
Figure 4.3 Circular dichroism (CD) spectra of AT_{CALE8} domain. The reference curve in blue is the spectra for myoglobin a predominantly helical protein. The spectra for AT_{CALE8} is shown in red. The CD spectrum of AT_{CALE8} conforms to the alpha-helix percentage (42%). Based on the reference curve, we can see the predominance of helixes in the structure.

The CD spectrum of AT_{CALE8} reveals a predominantly helical protein with an alpha-helix percentage of 42%. This conforms to the predicted secondary structure (Helix - 41.52%, β-sheet - 13.17%, random coil - 45.31%).

4.3.4 Preparation of *holo*-ACP by Sfp modification

Sfp, the type II PPTase from *B. subtilis* exhibits broad substrate specificity and modifies the carrier proteins from various families across FASs and PKSs. Sfp was used to test whether *me*ACP can undergo phosphopantetheinylation as explained in chapter II. The Sfp protein with a broad substrate spectrum was able to phosphopantetheinylate *me*ACP and the formation of *holo*-CalE8-ACP after the incubation with Sfp and CoA for 1 h was validated by HPLC. The overall turnover increased slightly with extended incubation time (data not shown).

4.3.5 Interaction between *holo*-CalE8-ACP and AT$_{CALE8}$ in the presence of acetyl CoA and malonyl CoA

The *holo*-CalE8-ACP, prepared using modification by Sfp was treated with AT$_{CALE8}$ protein in the presence of acetyl CoA and malonyl CoA. The samples were analyzed using HPLC. The results showed that the AT$_{CALE8}$ treatment of *holo*-CalE8-ACP in the presence of malonyl CoA showed a new shoulder peak alongside the *holo*-peak (**Fig. 4.4**), whereas there was no change observed in the *holo*-CalE8-ACP when treated with acetyl CoA. In order to substantiate this claim, the concentration of AT$_{CALE8}$ was increased expecting further changes in the new *holo* peak. However the AT$_{CALE8}$ elution peak occurs in the same region as the *holo*-CalE8-ACP peak and this hinders the examination of the *apo*-ACP, *holo*-ACP and AT$_{CALE8}$ peaks (**Table 4.1**). Various solvent systems and gradient systems were employed to resolve the CalE8-ACP and AT$_{CALE8}$ peaks. Though good resolution was observed when the domains were tested individually, when the peaks occurred almost at the same regions when the reaction mixtures were run and distinct resolution of the peaks became impossible.
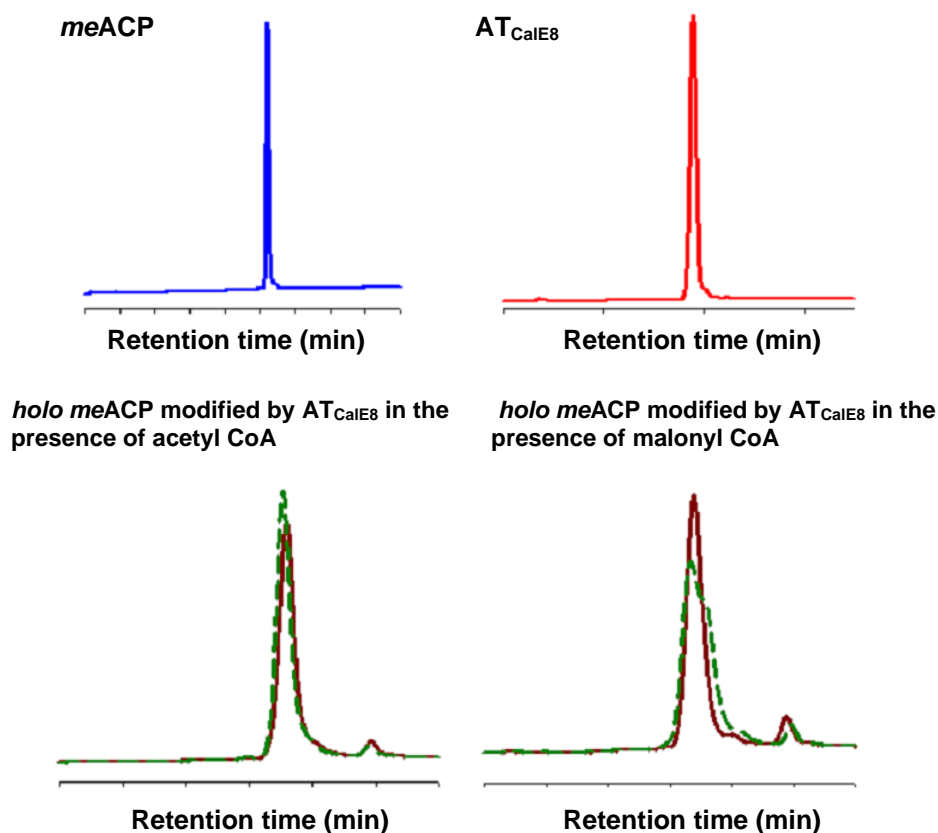
Figure 4.4 Activity of AT$_{CALE8}$ domain. HPLC analysis of Sfp modified *holo me*ACP (brown) treated with acetyl CoA and malonyl CoA (green) in the presence of AT$_{CALE8}$ domain. We can see that there is no shift in the *me*ACP peak when treated with AT$_{CALE8}$ and acetyl CoA. There is a peak widening and occurrence of a twin peak when *me*ACP is treated with AT$_{CALE8}$ and malonyl CoA

4.3.6 Examination of the domains using Bioinformatics

CalE8 provides an excellent platform for studying the structure of multidomain proteins and iterative biosynthesis. Examination of each of the domains and the sub domains of CalE8 *in silico* provides a better understanding of the domains and CalE8 as a whole. A series of bioinformatics techniques of sequence alignment and analysis, phylogenetic examination,

secondary structural analysis, homology modeling, three dimensional visualization and residue mapping were employed to examine and characterize the domains.

4.3.6.1 Structural modeling and characterization of KS$_{CALE8}$

The region at the N-terminus of CalE8 spanning 460 residues was expected to code for the KS domain. *In silico* examination of the proposed KS$_{CALE8}$ domain revealed substantial information on its structure and function. The KS domain is expected to catalyze the key decarboxylative Claisen reaction in the formation of the carbon-carbon bond between an acyl thioester and a malonyl thioester also determine the chain length of the growing polyketide. The active site triad of the KS domains is made up of the nucleophilic cysteine residue involved in the covalent attachment of substrates and the two catalytic histidine residues expected to promote the decarboxylation of the malonyl moieties. This active site triad, a characteristic feature of the KS family is well conserved in the primary structure of the domain. Multiple and pair-wise sequence alignments with KSs from the FAS (human mitochondrial FAS) modular PKS (DEBS) and iterative PKS (SgcE and DynE) reveal that these active site residues, Cys 211, His 344 and His 384, are conserved in KS$_{CALE8}$ domain (**Fig. 4.5. a**). The active site of the KS$_{CALE8}$ is made up of the Cys-His-His triad similar to the KS domains of modular DEBS PKS and human mitochondrial β-ketoacyl synthases.

Figure 4.5 Bioinformatics analysis of the KS domains. a) Partial multiple sequence alignment of the KS domains. Representative KS domains from FAS (human mitochondrial FAS) and human mitochondrial KS, modular PKS (DEBS) and iterative PKS (SgcE and DynE), R1128 (ZhuH) were aligned with KS$_{CALE8}$ domain. The conserved residues are boxed and highlighted in red. The active site Cys, His and His residues are marked by the green circle while the Gly and two Thr residues are marked by asterisk. The crystal structure of DEBS KS was used as the template for secondary structural alignment. b) Phylogenetic tree showing the evolutionary relationship between the KS domains from various families.

This Cys-His-His triad is in contrast to the active site triad, Cys-His-Asn, found in ZhuH, a priming ketosynthase involved in the initiation of the elongation of the polyketide chain in the biosynthetic pathway of a type II polyketide, R1128 (*147*) and the KS domains of *E. coli* and *Arachis hypogea*. A lysine residue (Lys 386 in KS$_{CALE8}$) of unclear function is conserved amongst the family of KSs including KS$_{CALE8}$. The conserved Gly 270 residue is expected to allow entrance into the substrate binding tunnel and the glycine rich moiety near the C-terminus contributes to the oxyanion hole. The phylogenetic examination (**Fig. 4.5 b**) of the KS$_{CALE8}$ also shows that it is similar to the KS domain of the modular DEBS PKS. Based on the sequence alignments and phylogenetic examination, the crystal structure of DEBS-KS domain (2hg4D), was chosen as the template and a model for KS$_{CALE8}$ was generated using Swiss Model server. The validated 3D model was examined for the conserved structural features. The KS$_{CALE8}$ model (**Fig. 4.6 a**) adopts the signature α-β-α-β-α thiolase fold as seen in the other representative KS domains across various families.

The residues that make up the catalytic triad, Cys 211, His 344 and His 384 are placed spatially close to each other and their spacing and alignment further confirm the integrity of the model (**Fig. 4.6 b**). The gatekeeper Gly 270 residue and the Thr 346 and Thr 348 residues are also spatially conserved. The two Thr residues are expected to form hydrogen bonds with the ACP domain.
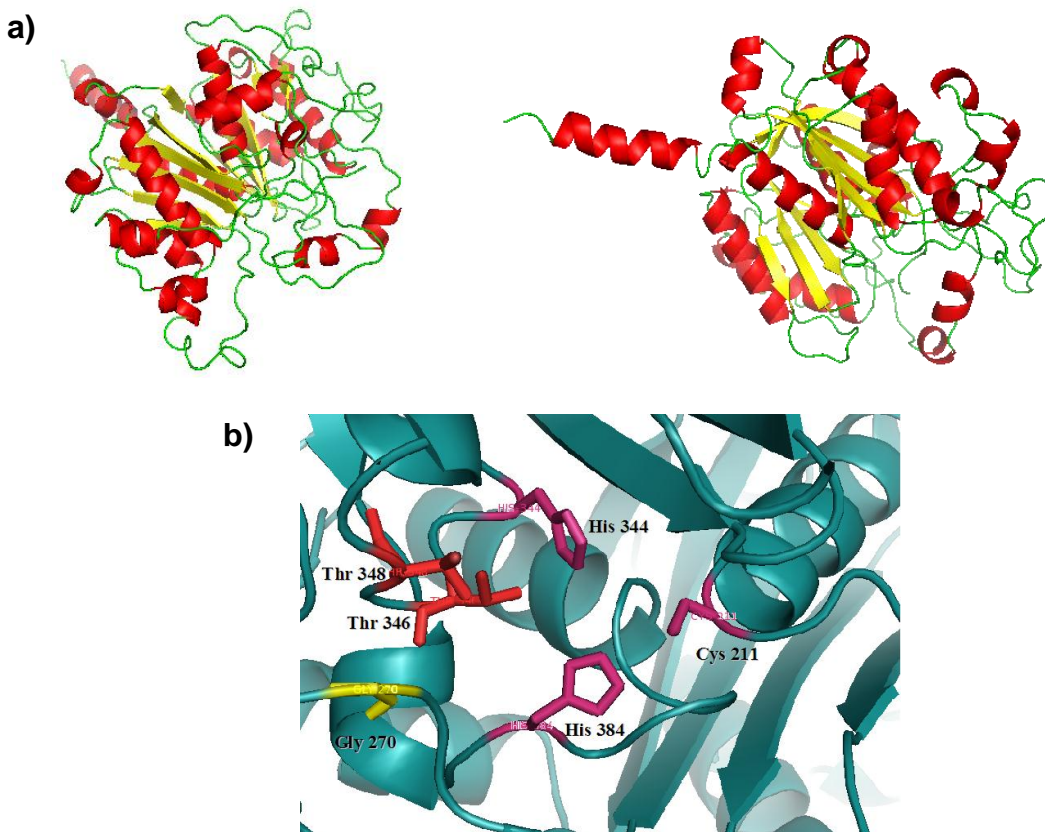
Figure 4.6 Structural modeling of KS$_{CALE8}$ domain. a) Different views of the model of KS$_{CALE8}$ built with DEBS KS (2hg4D) as a template. b) Spatial mapping of the active site residues in the model. The active site triad Cys 211, His 344 and His 384 (highlighted in pink) along with the conserved Thr 346, Thr 348 (highlighted in red) and the Gly 270 (highlighted in yellow).

The sequence and phylogenetic analysis coupled with modeling studies clearly indicate that the KS$_{CALE8}$ domain exhibits significant similarity to the KS domains of the modular PKSs like the DEBS KS. The KS$_{CALE8}$ domain exhibits the proposed characteristic thiolase fold. The active site triad is made of the Cys-His-His residues and is spatially conserved along with the other significant residues.

126

4.3.6.2 Structural modeling and characterization of AT$_{CALE8}$

The proposed AT$_{CALE8}$ domain spanning 424 residues (from 482 aa – 905 aa) has been successfully cloned and expressed. The secondary structural integrity of the domain has been verified by CD spectroscopy. The functional characterization of the AT$_{CALE8}$ domain has shown that it exhibits a substrate preference to the malonyl CoA over acetyl CoA. The AT domain was also examined by sequence comparison and structural modeling. Based on the sequence analysis (**Fig. 4.7 a**), the catalytic serine residue is conserved within the GHSxG (GHSLG) motif. This motif harboring the nucleophilic serine is conserved extensively over various families and subfamilies of the AT domain. The other catalytic residue, histidine is found within a highly conserved HAFH sequence motif, 100 residues downstream of the catalytic serine. Interestingly for some AT domains like DEBS AT, the catalytic histidine is conserved within a YASH sequence motif that is expected to prefer binding of methyl malonyl CoA while the HAFH motif has been discovered to influence preferential binding of malonyl CoA (*175*).

Figure 4.7 Bioinformatics analysis of the AT domains. a) Partial multiple sequence alignment of the AT domains. Representative AT domains from FAS (human mitochondrial FAS), modular PKS (DEBS) and iterative PKS (SgcE and DynE), *E. coli* FabD, malonyl transferases from *Burkholderia pseudomallei* and *Thermos thermopilus* were aligned with AT$_{CALE8}$ domain. The conserved residues are boxed and highlighted in red. The crystal structure of human mitochondrial MAT was used as the template for secondary structural alignment. b) Phylogenetic tree showing the evolutionary relationship between the AT domains from various families.

The phylogenetic examination (**Fig. 4.7 b**) also confirms the fact that AT_{CALE8} domain resembles malonyl transferases across various families compared to acetyl CoA binding AT domains. It resembles the AT domains from *E. coli* FabD (PDB: 2g2z), and malonyl transferases from *Burkholderia pseudomallei* (PDB: 3g87) and *Thermos thermophilus* (PDB: 2cuyA) and human mitochondrial malonyl transferases but is less similar to AT domains from human FAS malonyl acyl transferase (MAT) and also the modular PKS, DEBS AT. Though it exhibits high similarity and appears phylogenetically closer to the malonyl CoA binding AT domains, the highly conserved RVDVVQ sequence motif located 30 residues upstream of the catalytic serine residue, is absent in AT_{CALE8} domain. This makes it unique even within the families of malonyl transferases. The similarities and differences between the AT_{CALE8} domain to human mitochondrial malonyl CoA transferase (MT) and human cytosolic malonyl acetyl transferase, MAT were examined in detail. Pairwise sequence alignments reveal that AT_{CALE8} is similar to MT than MAT. MAT domain has three hydrophobic residues, Met499, Phe553 and Phe682 around a conserved Arg606 residue that favor acetyl CoA binding whereas the MT domain has two conserved non-hydrophobic residues (Gln34 and Gln85) around Arg142 which favor malonyl CoA binding over acetyl CoA. Sequence alignment of AT_{CALE8} with human MAT and MT reveal that the hydrophobic residues Met499 and Phe553 of MAT are not conserved, whereas the polar Gln34 and Gln85 residues of MT are conserved. All these findings from the investigation at the primary structure level are consistent with our functional studies where AT_{CALE8} exhibits clear preference for malonyl CoA.

Figure 4.8 Structural modeling of AT$_{CALE8}$ domain. a) Homology model of AT$_{CALE8}$ built with DEBS KS5 (PDB: 2hg4D) as a template. b) Spatial mapping of the active site residues in the model. The active site residues the Ser nucleophile 178 within the conserved GSLG motif (in pink) and His 279 within the conserved HAFH motif (in orange) are highlighted in the model. c) The spatial conservation of the Gln 109, Gln 150 and the Arg 203 residues (the residues are numbered according to the domain and not the whole CalE8 protein) mapped in the homology models of AT$_{CALE8}$ built on (i) PDB: 2c2n, (ii) PDB: 2g2z and (iii) PDB: 3g87 respectively.

For the homology modeling of AT$_{CALE8}$ domain, the crystal structure of malonyl CoA-acyl carrier protein transacylase from *Thermus thermophilus*, 2cuyA was chosen as the template. But when the conserved Gln residues were mapped on the 3D model, it was found that one of the Gln residues was placed far apart. Hence new models were built with

the crystal structures of human mitochondrial malonyl transferase 2c2n, *E. coli* FabD 2g2z and the malonyl CoA-acyl carrier protein transacylase from *Burkholderia pseudomallei*, 3g87 as templates (**Fig. 4.8 a, b**) . The structural motifs were examined and the residues were mapped in all these models and their spatial integrities were confirmed.

The active site residues, the Ser nucleophile 178 placed within the conserved GSLG motif and His 279 placed within the conserved HAFH motif were examined and found to be spatially compatible. The spatial conservation of the residues that influence malonyl CoA binding, Gln 109, Gln 150 and the Arg 203 were mapped in the homology models of AT_{CALE8} built on 2c2n, 2g2z and 3g87 respectively. All three models show that these residues are situated close to each other to promote binding of malonyl CoA.

The sequence and structural homology results together imply that AT_{CALE8} could possibly resemble the AT domains that prefer malonyl CoA as a starter and extender unit over acetyl CoA. The results from the functional studies also clearly indicate that malonyl CoA is the preferred substrate of AT_{CALE8} domain and the sequence and phylogenetic analyses and homology modeling have substantiated this claim. The AT domain of CalE8 must employ some form of a unique self-decarboxylation mechanism to convert the malonyl CoA substrate to acetyl CoA.

4.3.6.3 Structural modeling and characterization of KR$_{CALE8}$

The role of the KR domain in polyketide biosynthesis is to reduce the β-keto group to hydroxyl group. The KRs from both FAS and PKS families belong to the short-chain dehydrogenase/reductase (SDR) family. KR uses NADPH to stereo-specifically reduce the β-keto group to a hydroxyl group. In DEBS, two domains were contained in the KR domain - the structural and the catalytic subdomains. The catalytic subdomain adopted the Rossman fold. In the structural subdomain, a signature sequence motif (H/L/M/F/Y) XXXW was observed. This motif serves to identify the start of the KR domain since there is very little sequence conservation amongst KR domains.

Sequence alignment and analysis (**Fig. 4.9 a**) show that the KR$_{CALE8}$ domain harbors the typical SDR motif, TGxxxGxG conserved as TGggkGiT where the last glycine residue is replaced by threonine. The catalytic aspartic acid residue is conserved. The active site tetrad seen conserved in the KRs, Asn, Ser, Tyr and Lys is conserved in the KR$_{CALE8}$ as seen in actKR and other modular KR domains, except that in actKR the position of Asn and Lys are interchanged. Also, the characteristic NNAG motif in actKR1 is not conserved but modified as HGAG. This HxAG motif is conserved among the modular PKS KRs, DEBS KR and tylKRs. Despite these deviations, the spacing and conservation of the active site tetrad reveals that it is phylogenetically similar to the KR domains of modular PKSs like DEBS KR and tylKR (**Fig. 4.9 b**). One very striking observation is in regard to the determination of stereochemistry by the KR domains. The modular type I KR domains like DEBS-KR and tylKR have a conserved LDD motif whose presence or absence will determine the reduction stereochemistry. Along the same lines, type II KR domains like actKR have a PGG sequence motif that is important for determining stereospecificity. KR$_{CALE8}$ doesn't harbor any of these sequences. Sequence

alignment with these sequences shows a NTP motif at the same position. This is conserved in the SgcE8 and DynE8 KR domains as an NxP motif. There is a possibility that the proline residue in the NxP motif might be sufficient to carry out the determination of stereochemistry as it has been seen that the P residue in the PGG motif alone is sufficient to determine stereochemistry (*177*).

a)

DEBS_KR α1 ——β1—— TT ——β2—— ——β3—— α2

```
            α1                    β1      TT        β2                              β3              α2
DEBS_KR    ΩΩΩΩ                ———————►            ———————►                      ———————►      ΩΩΩΩΩΩΩΩΩΩ
              1         10            20              30              40            50          60
DEBS KR    ...AAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWK....PTGTVLVTGGTGGVGGQIARWLARE
tylKR      ...AAVLAGRDGEDQVAVRASGIYGRRVSRAAA.AGAASWQ....PSGTVLITGGMGAIGRRLARRLAAE
DynE-KR    LAAVEASRPWSGYQEVVYGADGVRTVPVTRLLDGQPRTDRRIPLGPGEVCLVTGGAKGIGAACAAALGEA
SgcE-KR    ...AAEVAATTHFSEVHLDRDGTRRVPVLRALPFAPDRTDQV.LGPDDVLLIVTGGGKGITAECALAVAER
CalE8-KR2  ...VADVAATDDFTEAVYDPGGRRRVPVLRPLPASDGEPGALPLGPADVLLIVTGGGKGITAESALMLARE
actKR      ...............................MATQDSEVALVTCATSGIGLEIARRLGKE
consensus>50  ...aa.va...gfdev...adgvr.vpv.ral..a.........lgp.ev.LITGg.kg!gae.Ar.la.e
```

```
            β4        η1              α3              β5          α4                    β6
DEBS_KR    ———————►  ΩΩΩ        ΩΩΩΩΩΩΩΩΩ           ———    ΩΩΩΩΩΩΩ              ———————►
              70         80              90            100          110          120        130
DEBS KR    GAPHLLLVSESGPDAD.GAGELVAELEALGARTTVAACDVTDRESVRELLGGI.GDDVPLSAVPHAAATL
tylKR      GAERLVLTSERGPRAP.GAAELARELRGHGCEVVHAACDVAERDALAALVTAY.....PPNAVPHTAGIL
DynE-KR    TGARLVLPGPS.PADDPEVVEAVRR.....TGAVYRSVDLADPVAVAAAVTAVRAEHGPVRGLLHAAGRN
SgcE-KR    TGAALAVLGPSDPGSDQDLAANLGRMRESGIRVAYAAADVTDPVRVAGAVAELTGALGSVTAVLHGAGRN
CalE8-KR2  SGARLAVLGPSDPTADEALADNLKRLADAASDLRYLRVDVTDAGAVAAAVATVTADWGPVTAVLHGAGQN
actKR      .GLRVPYCAP.....GEEGLRTTLKELREAGVEADGRTCDVRSVPEIEALVAAVVERYGPVDVLVNNAGRP
consensus>50  .garl.v.gRs.p.adeglae.l.elre.g.e...yaacDvtd..avaalvaav...d.gpv.avlh.Agrn
```

```
            η2        α5          α6                      β7      α7                  α8
DEBS_KR    ΩΩΩ     ΩΩΩΩΩΩΩ      ΩΩΩΩΩΩΩΩΩΩΩΩΩΩ            ———►  ΩΩΩΩ                ΩΩΩΩΩΩΩΩΩΩΩ
              140        150           160                170          180          190
DEBS KR    DDGTVDTITGERIERASRAKVLGARNLHELTRELD.L..TAFVLFSSFASAFCAPGLGGYAPGNAYLDGL
tylKR      DDAVIDTISPESFRTVRGAKVCGAELLHQLTADIKGL..DAFVLFSSVTGTWCNACQGAYAAANAALDAL
DynE-KR    EPAPVTDITAERLHRTLAPKAYGFEHVVAALDPDE.L..RFAVAFGSVIGRAGLPGEAAYATANEYLSRR
SgcE-KR    EPTALGGIDMAAVRSTLAPKVDGLRHVLDVVGEQN.L..RLLVTFGSIIGRACLRGEAHYATANEWLAGL
CalE8-KR2  TPAALADIDEAALRGVPAAKVDGLRAVLAAVDPAR.L..RLLVTPGSIIGRACLHGEAHYAAANEALAEL
actKR      GGGATARIADELWLDVVETNLTGVFPRVTKQVLKAGGMLERGTGRIVNIASTGCKQCVVHAAPYSASKHGV
consensus>50  dpaav.dL..e....v.aakv.G.r.v.e.v.e.d.$..r..v.fgsiigraGl.GeahyA.ane.l.gl
```

```
                            β8                      α9      β9      α10              β10  β11
DEBS_KR    ΩΩΩ..ΩΩΩ       ———►                    ΩΩΩΩΩΩ  ———►   ΩΩΩΩΩΩΩΩΩΩΩΩΩΩ        ——► ——►
              200        210           220              230          240          250        260
DEBS KR    AQQ..RRSDGLPATAVAWGTWAGSGMAEG.PVADRFRRHGVIEMPPETACRALQNALDRAEVCPIVIDVR
tylKR      AER..RRAAGLPATSVAWGLWGGGMAAG.AGBESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVD
DynE-KR    CAELAATVPQVRWLSIEWSVWSGTGMGVRLGVLDGLLRRGVSPIPPDEGTRLLLELL............
SgcE-KR    TEDVARRNPDCRALCMEWSVWSGVGMGVRLGVLDGLLRRGVSPVSPDQGIBILLRL............
CalE8-KR2  TREVAATRPECRAVCLEWSVWSGVGMGERLSVVESLSGSGVTPISPDDGLRVLREVV............
actKR      VGFTKALGLELARTGITVNAVCPGFVETPMAASVREHYSDIWEVSTEEAFDRITARVPIGRYVQPSEVAE
consensus>50  .ee.aar.pelrat.vewsvwsg.gmge.l.v.eslsr.gv.pisp#ea.d.lle.l...........
```

b)

```
                            ┌──────── tyl KR
                   ┌────────┤
          ┌────────┤        └──────── DEBSKR
          │        │        ┌──────── DynE8 KR
          │        └────────┤   ┌──── CalE8 KR
──────────┤                 └───┤
          │                     └──── SgcE8 KR
          └────────────────────────── act KR
```

Figure 4.9 Bioinformatics analysis of the KR domains. a) Partial multiple sequence alignment of the KR domains. Representative KR domains from modular PKS (tylKR, DEBS KR) and iterative PKS (SgcE and DynE), act KR were aligned with KR$_{CALE8}$ domain. The conserved residues are boxed and highlighted in red. The active site residues are marked with a brown asterisk, and the other important residues with a green circle. The crystal structure of DEBS KR was used as the template for secondary structural alignment. b) Phylogenetic tree showing the evolutionary homology between the representative KR domains.

Homology modeling of the KR_{CALE8} was initially attempted with the crystal structures of the KR domains of modular PKSs, DEBS KR, tylKR based on their phylogenetic similarities. However none of these models were satisfactory and did not conform to the spatial arrangement of the significant residues and the predicted structural arrangement. When actKR was employed as a template, the model obtained (**Fig. 4.10 a**) conformed to all the predicted spatial arrangement of the secondary structures and active site residues (**Fig. 4.10 b**). The 3D model of the structure clearly reveals that the KR domain follows the Rossman fold.
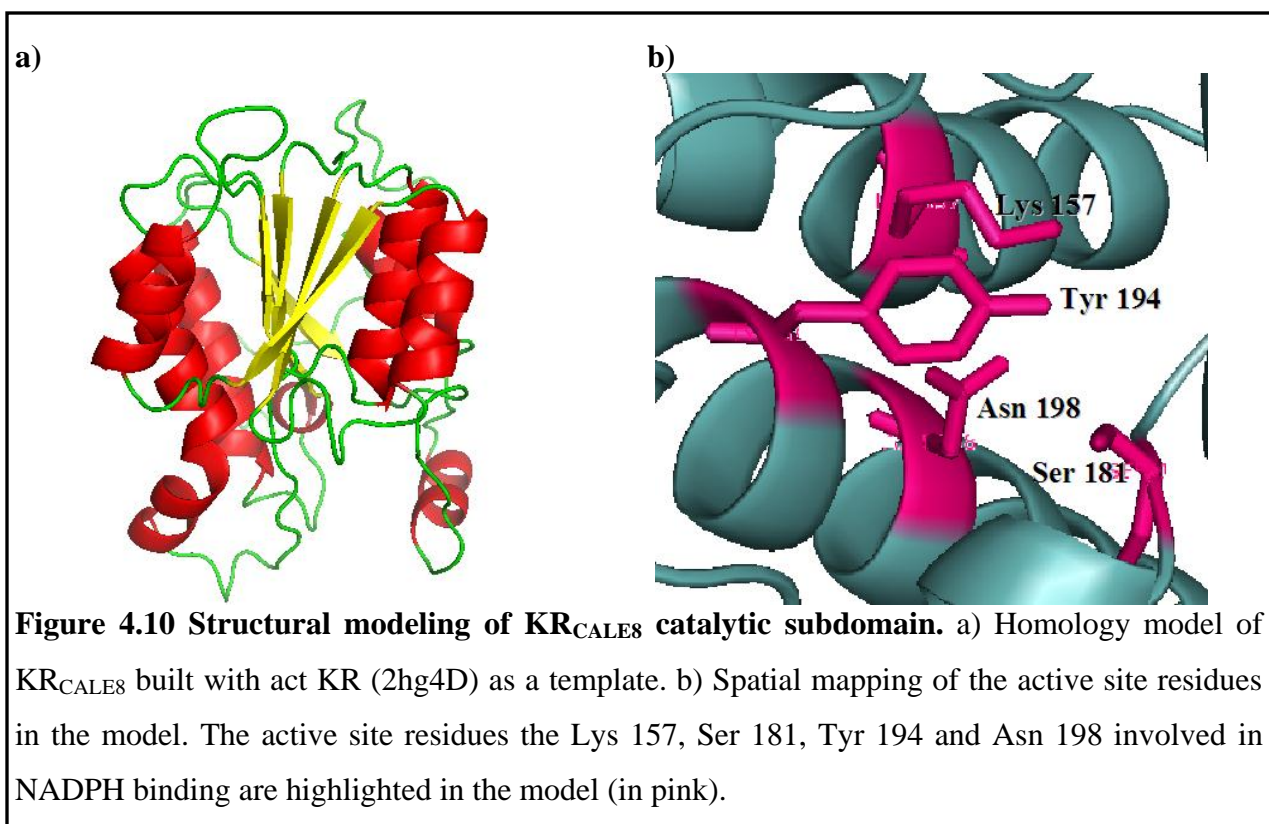
Figure 4.10 Structural modeling of KR$_{CALE8}$ catalytic subdomain. a) Homology model of KR$_{CALE8}$ built with act KR (2hg4D) as a template. b) Spatial mapping of the active site residues in the model. The active site residues the Lys 157, Ser 181, Tyr 194 and Asn 198 involved in NADPH binding are highlighted in the model (in pink).

The active site residues forming the tetrad (Lys 157, Ser 181, Asn 198 and Tyr 194) were mapped and their spatial integrity was confirmed. Examination of the sequence, active site residues and homology modeling allows us to predict the mechanism of activity of the KR$_{CALE8}$ domain. Similar to the DEBS KR domain (**Fig. 4.11 a**), in KR$_{CALE8}$ (**Fig. 4.11 b**), NADPH could

stack against the conserved Arg 77 residue and form hydrogen bonds with conserved Asp 105 and Val 106 residues. The β-carbonyl group is bound to Ser and Tyr and is attacked by the NADPH that is hydrogen bonded to Tyr 194 and Asn 198. The adenine ribose phosphate hydrogen bonds to Ser 181 and forms a salt bridge with Arg 77.

Figure 4.11 Catalytic residues of KR involved in NADPH binding. a) The active site tetrad made up of Lys, Ser, Tyr and Asn (highlighted in pink) along with the conserved Arg (highlighted in blue) and Asp and Val (highlighted in orange). b) The active site tetrad made up of Lys, Ser, Tyr and Asn (highlighted in pink) along with the conserved Arg (highlighted in blue) and Asp and Val (highlighted in orange)

When sequence alignment with both the KR domains was attempted, there was no similarity for the upstream first linker domain of unknown function. The upstream domain of KR after ACP had no homolog in the protein data bank. However when we tried *de novo* modelling on this ~100 amino acid fragment, it was found to adopt a methyl transferase-like structure. The important residues in functional methyl transferases necessary for the binding of the S-adenosine-L-methionine (SAM) cofactor are however absent in the sequence. The KR domains of the mammalian fatty acid synthase and modular PKS (DEBS) possess a degenerate methyl transferase (ψME) subdomain. It is highly possible that this methyl transferase-like domain is

part of the KR domain. Hence, it will be referred to as KR1$_{CALE8}$ domain. Interestingly, 3 of the 4 residues seen in KR domains involved in NADPH binding are found to be conserved in the KR1 domain. Though this is not highlighted in the sequence alignment, the spacing of these residues in the primary structure clearly indicates that they are conserved.

4.3.6.4 Structural modeling and characterization of DH$_{CALE8}$

The DH domain is involved in the dehydration of hydroxyacyl-ACP in FASs and 2R-methyl-3R-OH pentaketide in modular PKSs. The sequence alignment of DH$_{CALE8}$ domain with the representative DH domains (**Fig. 4.12. a**) did not give a satisfactory alignment. Despite low sequence homology, the characteristic HXXXGXXXXP motif was found conserved amongst all the DH domains. Though the DH$_{CALE8}$ domain harbors the significant HXXXGXXXXP and DXXXQ motifs, the other two significant motifs, GYXYGPXF and LPFXW are not maintained.

Despite its limited sequence similarities, based on the conserved residues and arrangement of secondary structures, the DH$_{CALE8}$ domains is relatively more similar to the dehydratase domain of the modular PKS, DEBS DH (**Fig. 4.12 b**). DEBS DH domain exhibits a clear double-hot-dog-fold structure with the alpha helix surrounded by the beta sheets.

a)

```
             η2    β4   β5            α1                  β6              TT       β7      TT
DEBS_DH    ℓ..ℓℓℓℓℓ →→ TT →   ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ       →                TT   →           TT
                70         80          90         100        110          120

DEBS_DH      EQP..WLAEHVVGGRTLVPGSVLVDLALAAGEDVGLPVLEELVLQRPLVLAGAGALLRMSVGAPDESGRR
rat_FAS_DH   ESSDHYLVDHCIDGRVLFPGTGYLYLVWKTLARSLSLSLEETPVVFENVTFHQATILPRTGTVPLEV..R
CalE8⊠DH     RTTDPYLEDHRLDGDLLFPAVLGLE...AMAQVATALARHPGVPVIEEVRFDRPVVVDPDTGTTVRVAAL
SgcE-DH      AGTDLYLTDHMLDGNLLPAVIGME...AMVQVGSAVTGRRDVPVIEDARFLRPIVVPPGGTTRIRIAAT
DynE-DH      LGADPYLDDHRIDGAAVLPAVLGLE...AMAQAAAALGAKQVPTVIRDVALTRPVTVGDREPRPVRVAAL
E.coli_Fabz  SVNEPFFQGH.FPGKPIFPGVLILE...AMAQATGILAFKSVGKLEPGELYYFAGIDEARFKRPVVPGDQ
consensus>50 ...dpyledH.idG..lfPavlgle...amaq...al...e.v.viedv.f.rpviv......pvrvaa.
```

```
                β8          η3          β9                   TT   β10      α2    β11    η4
DEBS_DH       →            ℓℓℓ       →                       →       ℓℓℓℓℓℓℓ  →      ℓℓℓ
           130        140        150        160        170        180        190

DEBS_DH      TIDVHAAEDVADLAD..AQWSQHATGTLAQGVAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQGYEYGPSF
rat_FAS_DH   LLEASHAFEVSDSGNLIVSGKVYQWEDPDSKLFDHPEVPIPAESESVSRLTQGEVYKELRLRGYDYGPHF
CalE8⊠DH     VRSEQVIDVVLRSAVTGFAADHFRARLRFAPDETYPAYTAAPAPAELPAVPLDPARDLYGDVLFQAG.RF
SgcE-DH      VTGTDRVDVAVHAQDTGFAAEHFRARLVY.....................................
DynE-DH      TGADGTVRAAVRSAETGFAVDHLTGTYGGAVTSGEP....VPPPSTGLLVPG..............
E.coli_Fabz  MIMEVTFEKTRRGLTRFKGVALVDGKVVCEATMMCARSREA..................
consensus>50 ...e..ve.v.rsadtgfa.dhf............p.......p.....vp............ye.g..f
```

b)

DEBS DH
E.coli FabZ
CalE8-DH
SgcE DH
DynE DH
Rat FAS DH

Figure 4.12 Bioinformatics analysis of the DH domains. a) Partial multiple sequence alignment of the DH domains. Representative DH domains from modular PKS (DEBS DH), FAS (rat FAS DH), iterative PKS (SgcE and DynE) and *E. coli* (FabZ) were aligned with DH$_{CALE8}$ domain. The conserved HXXXGXXXXP motif is boxed and highlighted in red. The crystal structure of DEBS DH was used as the template for secondary structural alignment. b) Phylogenetic tree showing the evolutionary homology between the representative KR domains.

The DH$_{CALE8}$ domain modeled (**Fig. 4.13 a**) with DEBS DH (PDB: 3el6) (**Fig. 4.13 b**) as a template yielded a clear hot-dog fold. In DEBS DH, the Gly residue within the highly conserved HXXXGXXXXP motif induces a turn in allowing van der Waals interactions between the His and Pro residues (**Fig. 4.13 d (i)).** When these residues were mapped in the modeled structure of

CalE8, they adopt a similar arrangement with Gly residue inducing the turn (**(Fig. 4.13 d (ii)**) and bringing the His and Pro residues closer.

Figure 4.13 Structural modeling of DH$_{CALE8}$ domain. a) Homology model of DH$_{CALE8}$1 built with DEBS DH, 3el6 as a template showing the hot dog fold, c) i & ii. The HXXXGXXXXP motif of DEBS DH and DH$_{CALE8}$ with the Gly residue, inducing a turn.
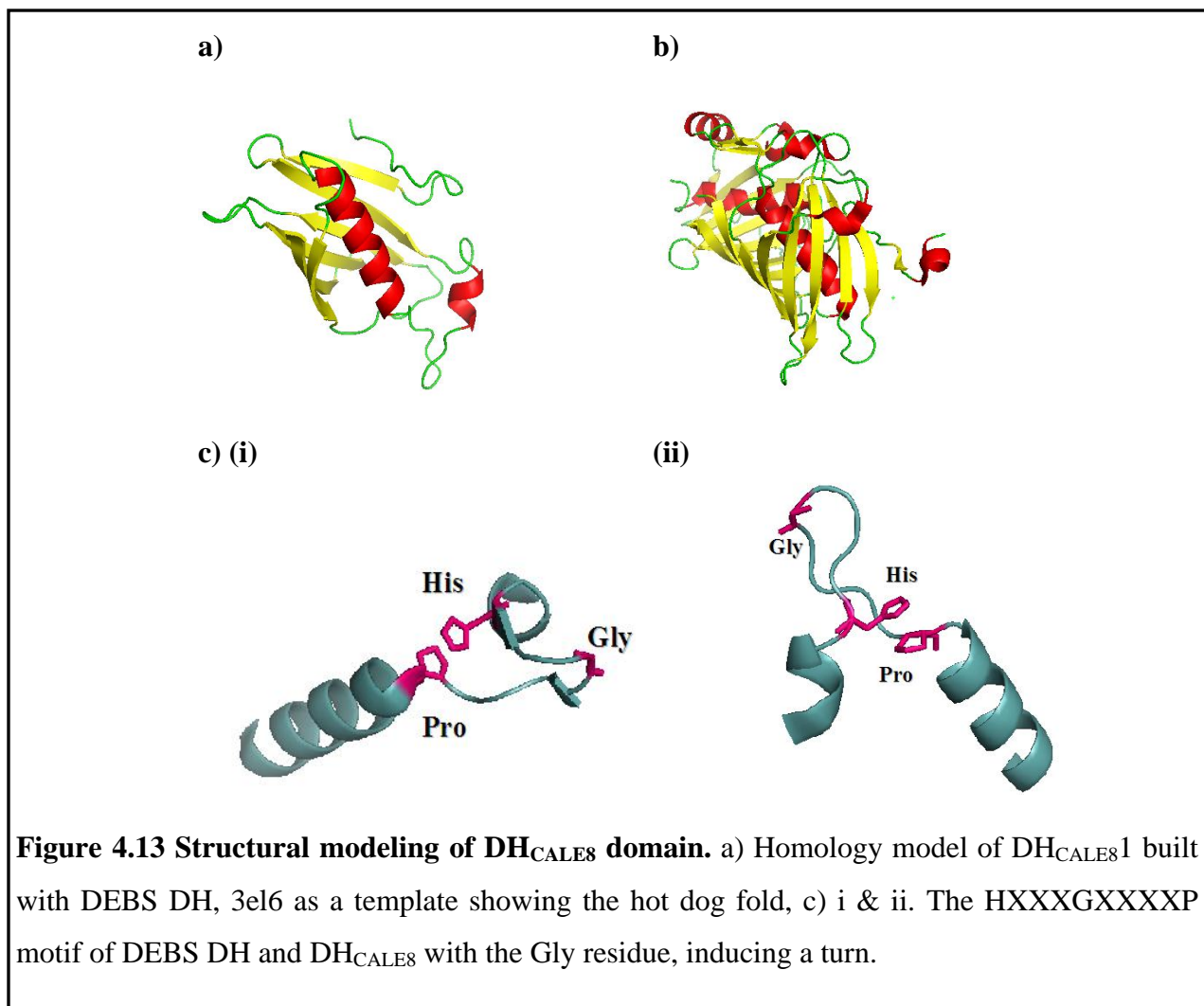
This structural similarity clearly validates the model and also provides us a platform to propose a mechanism of activity for DH$_{CALE8}$. DH$_{CALE8}$ can be assumed to adopt a catalytic mechanism similar to that proposed for DEBS DH. In this proposed mechanism, the conserved His residue acts as the active-site base in the deprotonation of the α-carbon, while the β-OH group maybe polarized by the helix-1 dipole thereby promoting dehydration.

Figure 4.14 Double hot dog fold in DH domains. a) Homology model of DH$_{CALE8}$2 showing the hot dog fold, b) Homology model of DH$_{CALE8}$1 showing the hot dog fold, c) Homology model of DH$_{CALE8}$1 and DH2 domains showing the double hot dog fold, d) crystal structure of DEBS DH (3EL6) exhibiting the double hot dog fold.
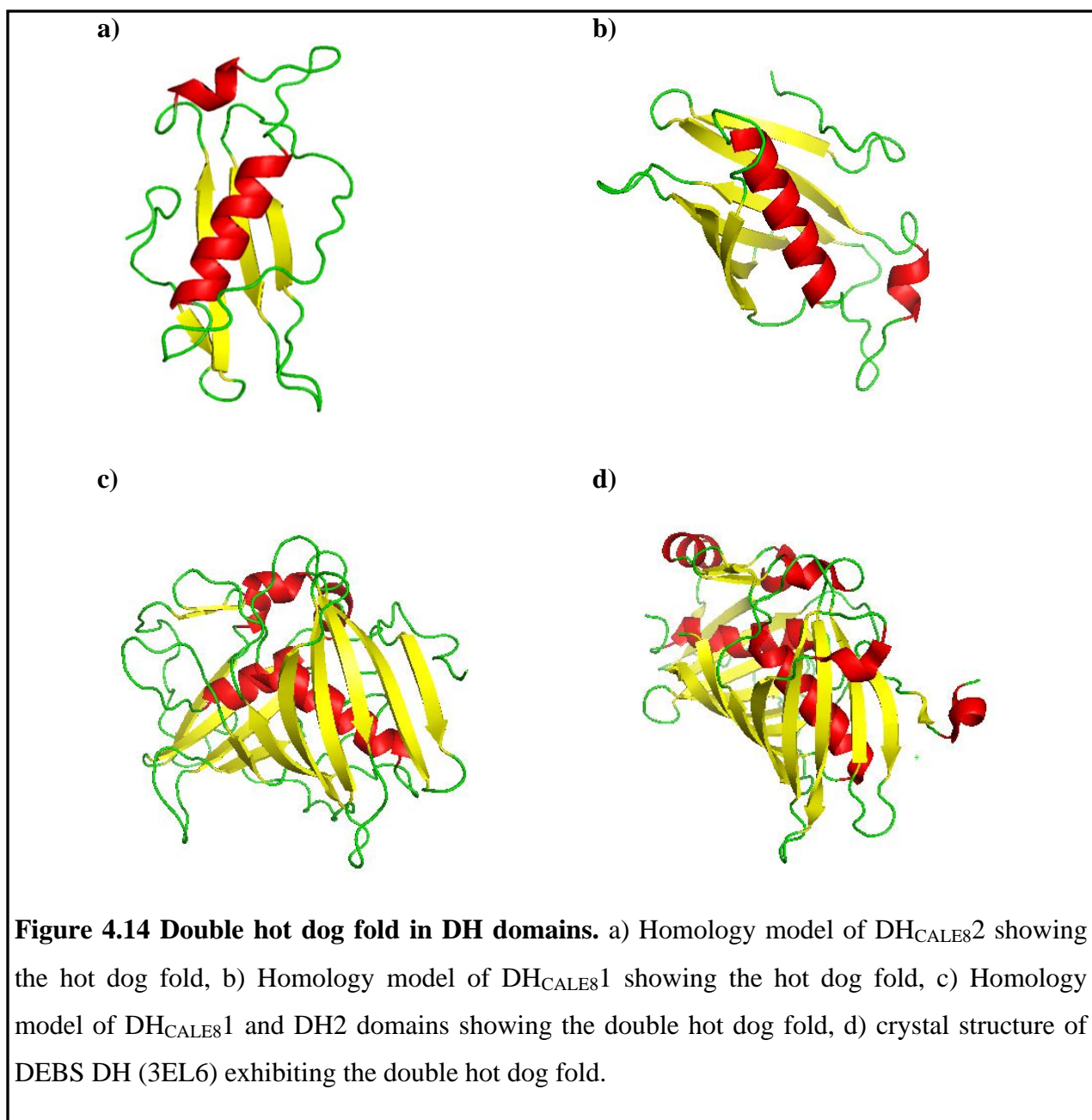
The ~150 aa sequence after the DH1 domain was previously speculated to be an evolutionary relic of an AcpS-like PPTase domain (*68*). We also found that the presence of this segment is imperative for the solubility and functioning of the PPTase. However, recent studies on the DH domains of the mammalian FAS and modular PKS reveal that they are actually comprised of two hotdog subdomains. Structural modeling analysis of the domain shows that the domain adopts a

hotdog fold (**Fig. 4.14 a**). The DH$_{CALE8}$1 domain was also found to exhibit the hotdog fold (**Fig. 4.14 b**). Homology modelling on the two domains together exhibited double hot dog fold (**Fig. 4.14 c**) similar to the double hot dog fold displayed by DH domains of modular PKS like DEBS (**Fig. 4.14 d**). The structural model of the DH1 and DH2 domains further confirms that the region between the DH1 and PPTase domains is probably the second half of the intact DH domain. Hence we assigned this domain as DH2 domain.

4.4 Conclusions

The ACP and PPTase domains in CalE8 were characterized and assigned boundaries in Chapter 2. The rest of the core domains and modifying domains were examined in this Chapter. Functional examination and *in silico* analyses were employed to assign the boundaries of the individual domains. The KS, AT, KR, DH domains and the two regions of unknown function were studied and examined. The AT$_{CalE8}$ domain was cloned and expressed as a functional protein, but the other domains could not be expressed as soluble proteins. All the domains were investigated by structural modelling based on homology. Functional validation and *in silico* analyses for the AT$_{CalE8}$ domain and the investigations conducted on the structural models for all the domains provided the following conclusions.

The KS$_{CALE8}$ domain has significant similarity to the KS domains of the modular PKSs like the DEBS KS. The KS$_{CALE8}$ domain exhibited the proposed characteristic thiolase fold and possesses the Cys-His-His active site triad spatially conserved along with the other significant residues.

The AT$_{CALE8}$ domain could possibly resemble the group of AT domains that prefer malonyl CoA as a starter and extender unit over acetyl CoA. The results from the functional studies also emphasize this aspect with clear indications that malonyl CoA is the preferred substrate of

AT$_{CALE8}$ domain. The sequence and phylogenetic analyses and homology modeling have also substantiated this claim. The AT$_{CALE8}$ domain must employ some form of a unique self-decarboxylation mechanism to convert the malonyl CoA substrate to acetyl CoA.

The 3D model of the KR$_{CALE8}$ domain clearly reveals the signature Rossman fold. The active site residues forming the tetrad (Lys 157, Ser 181, Asn 198 and Tyr 194) were mapped and their spatial integrity was confirmed. Based on the examination of the sequence, active site residues and homology modeling, the KR$_{CALE8}$ domain can be expected to adopt a mechanism of NADPH binding similar to the DEBS KR domain.

The region upstream of the KR domain was found to adopt a methyl transferase-like structure. Similar to mammalian FAS and modular PKSs that harbour a degenerate methyl transferase domain, this methyl transferase-like domain is likely a part of the KR domain and has been renamed as the KR1$_{CALE8}$ domain.

The structural model of the DH$_{CALE8}$ domain displays the well-preserved hot-dog fold a hallmark for dehydratase domains. The DH$_{CALE8}$ is similar to the DEBS DH domain and is expected to follow a similar dehydration mechanism. The region downstream of the DH domain, which was earlier expected to be a structural component of the PPTase domain (explained in Chapter 2), was examined and was included in developing the model. The model displayed the double hot dog fold as seen in the dehydratase domain of modular PKSs. The structural model of the DH1 and DH2 domains confirmed that the region between the DH1 and PPTase domains is probably the second half of the intact DH domain. Hence we assigned this domain as DH$_{CALE8}$2 domain.

Examination of the sequence and structure of the different domains has revealed that the domains and their arrangement were novel and displayed varied similarities and differences. The two

regions that were earlier designated as domains of unknown functions were found to be extensions of the KR and DH domains. Based on the observations we were able to assign the boundaries of the domains within CalE8 and also propose mechanisms for the functioning of the individual domains.

Chapter 5 Conclusions and Future directions

5.1 Conclusions

The primary objective of the thesis was to identify, characterize, examine and validate the discrete domains within the multidomain CalE8. The whole protein was examined through bioinformatics tools to demarcate the boundaries of the individual domains and the domain architecture of the multidomain CalE8 has been proposed. The domains boundaries were examined based on sequence and secondary structural analysis to preserve the integrity of the secondary structures. The whole protein, CalE8 was expressed and purified for analysis and examination. Expression and purification of the individual domains as stand-alone proteins were also attempted. Following the successful expression and purification of the individual domains like ACP, PPTase and AT domains, the predicted domain boundaries and assigned functionalities were tested and validated. The KS, KR and DH domains, which could not be expressed as soluble proteins were examined and characterized employing bioinformatics tools. With the structural and functional data, the regions of the whole length CalE8 were divided into domains with characteristic structure and function.
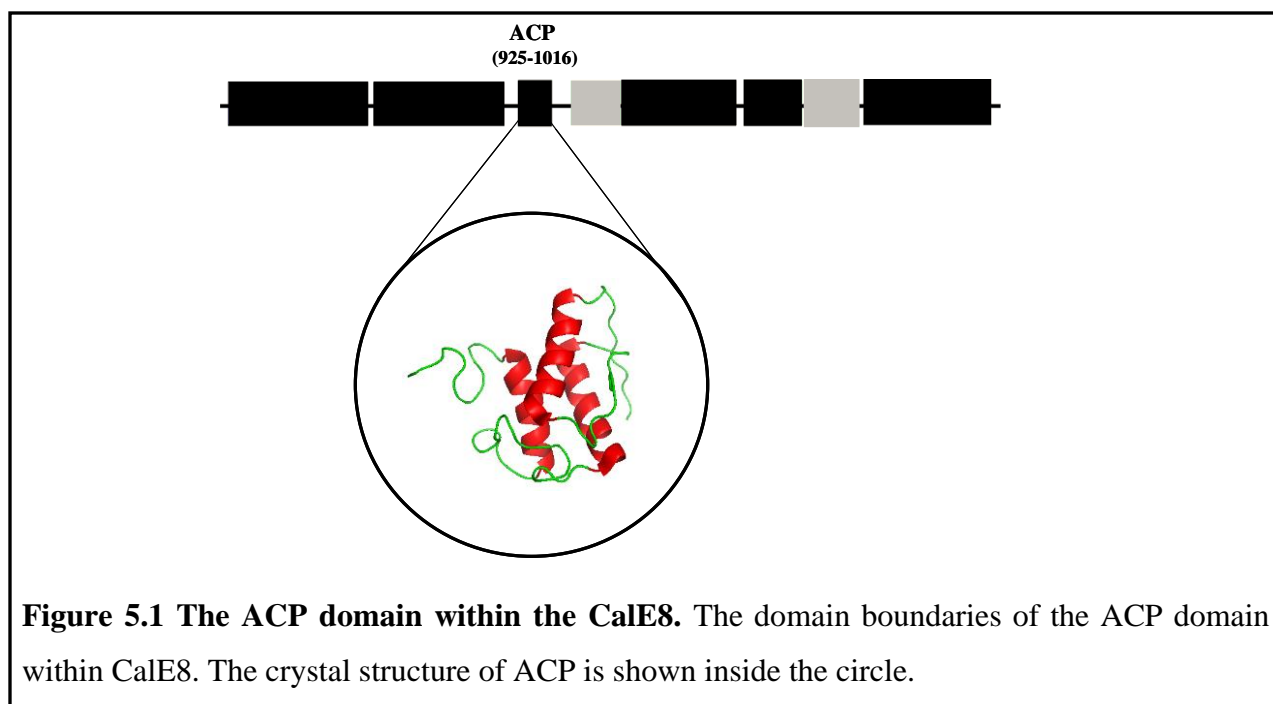
The thesis work constitutes the cloning, expression and purification of the whole protein (CalE8), 3 major domains (ACP, PPTase and AT domains), a novel phosphodiesterase from *P. aeruginosa* (*Pa*AcpH), a surfactin PPTase (Sfp) capable of modifying CPs that was used to study ACP and PPTase domains, 3 CPs from secondary metabolite pathways (EACP, GPCP and DACP) and 4 CPs from *P. aeruginosa* (PA1869, PA 2966, PA 3334 and PA4226) and purification of another PKS, SgcE for comparative analysis. Cloning, expression and purification of the proteins employed various techniques of molecular biology and protein engineering, like the long range PCR to generate 6 kb constructs, the pET-vector and BL21(DE3) expression

system, Ek/LIC cloning method, high performance Ni-sepharose and superdex-75 and 200 gel filtration systems in FPLC. Proteins were examined and characterized using SDS-PAGE, CD, MS and MALDI-TOF techniques. Functional assays were performed and the modifications and kinetics were studied employing techniques like HPLC and spectrophotometry. The KS, AT, KR, DH and two domains of unknown function were structurally modeled based on homology employing bioinformatics tools like ClustalW, Malign, Swiss-Modeller. The models were closely examined using PyMOL.

The dissertation emphasizes on the structural and functional characterization of the domains within the enediyne PKS CalE8. The domains exhibited significant novel characteristics − in terms of their similarities and homologies, to provide a very important topic of research. The major conclusions derived on the structural and functional aspects of the domains that have been examined and observed are summarized below.

5.1.1 Confirmation and Validation of ACP and PPTase domains of CalE8

The primary step in the polyketide synthesis is the addition of a 4'-phosphopantetheine moiety to the ACP domain within the PKS by a PPTase. Both these domains were expressed and characterized. Expression of the ACP domain of CalE8 was possible only after redefining the proposed domain boundaries to preserve the integrity of the secondary structural elements. The region (**Fig. 5.1**) spanning 92 aa (from 925 aa to 1016 aa) in CalE8 protein sequence was assigned as the functional ACP.

Figure 5.1 The ACP domain within the CalE8. The domain boundaries of the ACP domain within CalE8. The crystal structure of ACP is shown inside the circle.

The *me*ACP domain shared extremely low sequence homology with other known ACPs and carried an unusual HMS**S**I motif harboring the active site Ser971 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 similar to other ACPs with the exception of the absence of the short helix within loop-2. Owing to these structural diversities, the structure of *me*ACP was classified under a twisted three-helix bundle instead of the canonical four helix-bundle structure. Moreover the 'recognition helix' (α2) of *me*ACP was relatively deficient of the characteristic negative charges. Phylogenetic analysis showed that *me*ACP was similar to ACPs from the secondary metabolic pathways like DEBS ACP and PCP. The observation that *me*ACP was modified by Sfp and not its cognate partner, could be partially based on the unique structural features adopted by *me*ACP.

The C-terminal region (212 aa) of the iterative PKS in enediyne biosynthesis acted as a functional PPTase domain. This PPTase domain when expressed as a stand-alone protein

required the upstream linker region (one-half of the DH domain) to fold properly and to execute its activity. The C-terminal placement of the PPTase domain in enediyne PKSs bore a resemblance to the type III PPTase domains in yeast/fungal FASs, but the identification of catalytic residues conserved in type II PPTases suggested that the PPTase domain was most likely a type II PPTase. It consisted of two subdomains similar to Sfp PPTases. The identification of the conserved catalytic residues for type II PPTases from sequence alignment indicated that the PPTase domain might function similar to Sfp with only one active site for substrate binding and catalysis. The extremely low catalytic efficiency towards its own ACP partner and DACP led us to believe that the PPTase domain might not have been evolved to efficiently phosphopantetheinylate the ACP domains from secondary metabolic pathways.

5.1.2 Identification and characterization of a novel phosphodiesterase

A novel phosphodiesterase protein, *Pa*AcpH (**Fig. 5.2**), capable of cleaving the phosphopantetheinyl moiety from a modified CP was identified from *P. aeruginosa* through bioinformatics analysis.
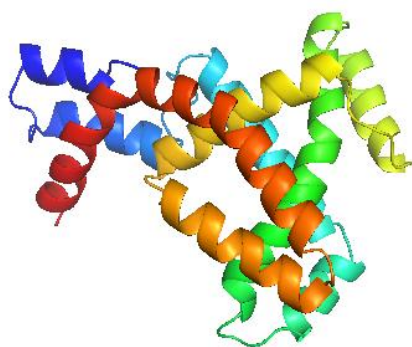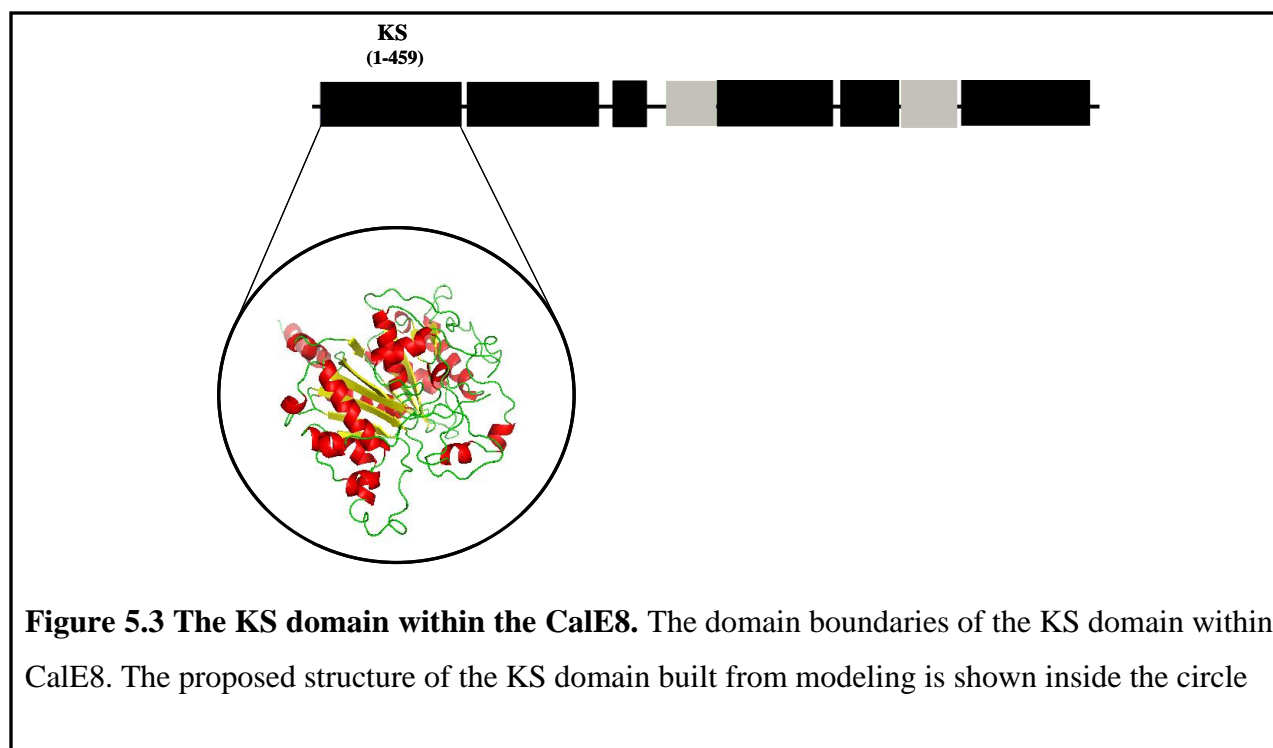

Figure 5.2 Model of *Pa*AcpH. The model structure of *Pa*AcpH based on *de novo* modeling. It is an all helical protein.

Only one protein from this rare non-essential phosphodiesterase family, the AcpH from *E. coli,* was characterized, but as an insoluble protein. In this project, the *Pa*AcpH protein was purified and expressed as a soluble protein that remained stable without any aggregation. The activity of the protein was examined against a family of homologous and heterologous carrier proteins. The ability of *Pa*AcpH to release the tethered intermediates from stand-alone acylated CPs has been confirmed. It has also been clearly demonstrated that the *Pa*AcpH protein was able to release the growing intermediates on acylated CPs within the PKSs.

5.1.3 Characterization of the core and the modifying domains of CalE8

The region spanning 460 residues at the N-terminus of CalE8 is the KS domain. The KS_{CALE8} domain exhibits the proposed characteristic α-β-α-β-α thiolase fold (**Fig. 5.3**). The active site of the KS_{CALE8} is made up of the Cys-His-His triad similar to the KS domains of modular DEBS PKS and human mitochondrial β-ketoacyl synthases. The residues that make up the catalytic triad, Cys 211, His 344 and His 384 were placed spatially close to each other and their spacing and alignment further confirmed the integrity of the model. The gatekeeper Gly 270 residue and the Thr 346 and Thr 348 residues expected to form hydrogen bonds with ACP were also spatially conserved. The sequence and phylogenetic analysis coupled with modeling studies indicated that the KS_{CALE8} domain exhibits significant similarity to the KS domains of the modular PKSs like the DEBS KS.

Figure 5.3 The KS domain within the CalE8. The domain boundaries of the KS domain within CalE8. The proposed structure of the KS domain built from modeling is shown inside the circle
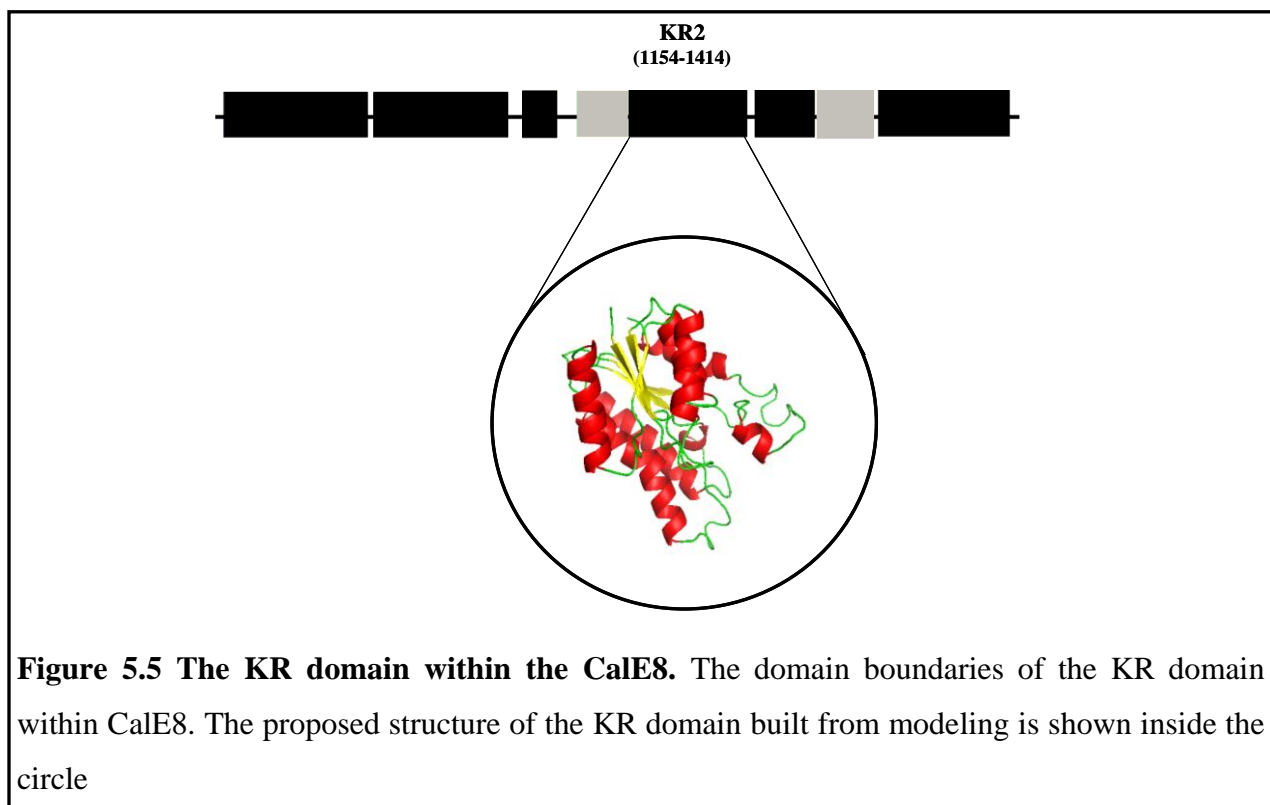
The region spanning 424 residues downstream of the KS domain (from 482 aa – 905 aa) is the AT domain (**Fig. 5.4**). It has been successfully cloned and expressed. The secondary structural integrity of the domain was verified by CD spectroscopy. The functional characterization of the AT_{CALE8} domain showed that it exhibits a substrate preference to the malonyl CoA over acetyl CoA. Inability to resolve the peaks (the *holo*-ACP peak and the AT_{CalE8} peak) also hindered further examination. This functional aspect was also verified by bioinformatics analyses. Based on the sequence analysis the catalytic serine residue was found conserved within the characteristic GHSxG (GHSLG) motif. The other catalytic residue, histidine is found within a highly conserved HAFH sequence motif, 100 residues downstream of the catalytic serine. Interestingly for some AT domains like DEBS AT, the catalytic histidine is conserved within a YASH sequence motif that is expected to prefer binding of methyl malonyl CoA while the HAFH motif has been discovered to influence preferential binding of malonyl CoA. The phylogenetic examination also confirms the fact that AT_{CALE8} domain resembles malonyl

transferases across various families compared to acetyl CoA binding AT domains. Though it exhibits high similarity and appears phylogenetically closer to the malonyl CoA binding AT domains, the highly conserved RVDVVQ sequence motif located 30 residues upstream of the catalytic serine residue, is absent in AT_{CALE8} domain. This makes it unique even within the families of malonyl transferases. Homology based structural modeling studies have shown that the active site residues, the Ser nucleophile 178 and His 279 along with the residues that influence malonyl CoA binding, Gln 109, Gln 150 and the Arg 203 were spatially compatible to promote binding of malonyl CoA.
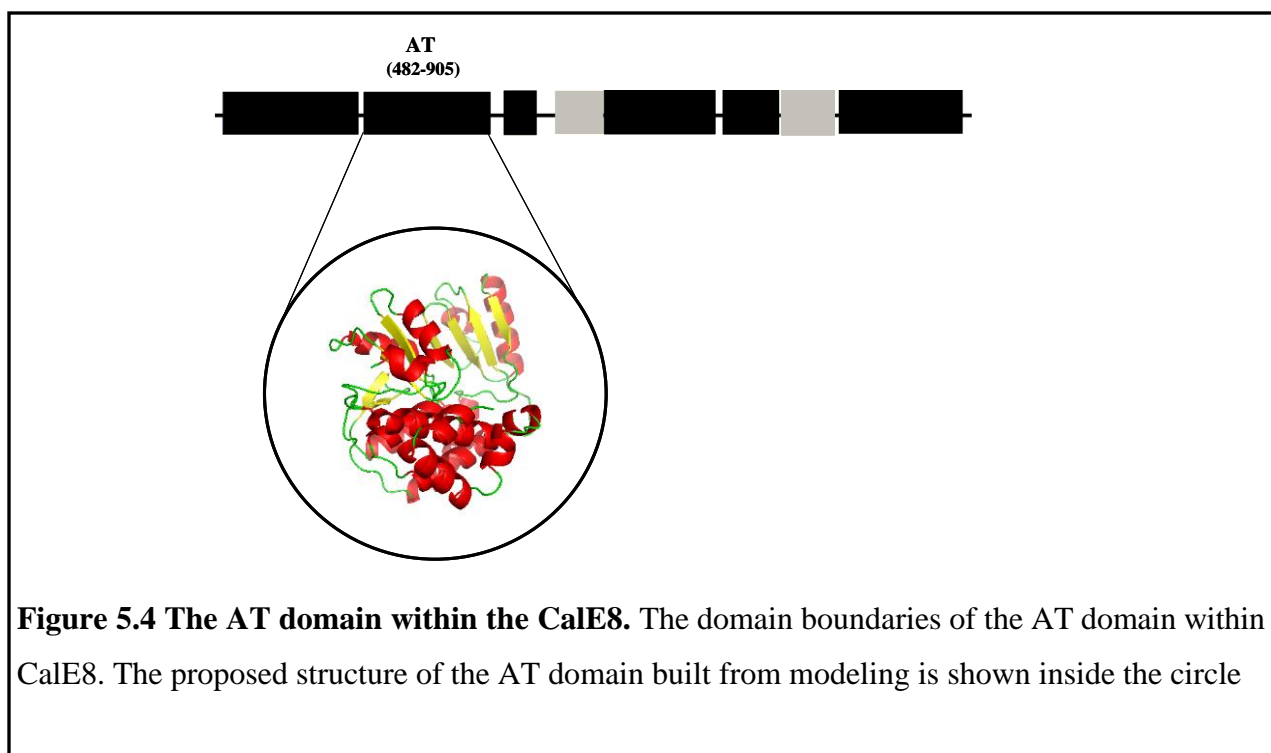
Figure 5.4 The AT domain within the CalE8. The domain boundaries of the AT domain within CalE8. The proposed structure of the AT domain built from modeling is shown inside the circle
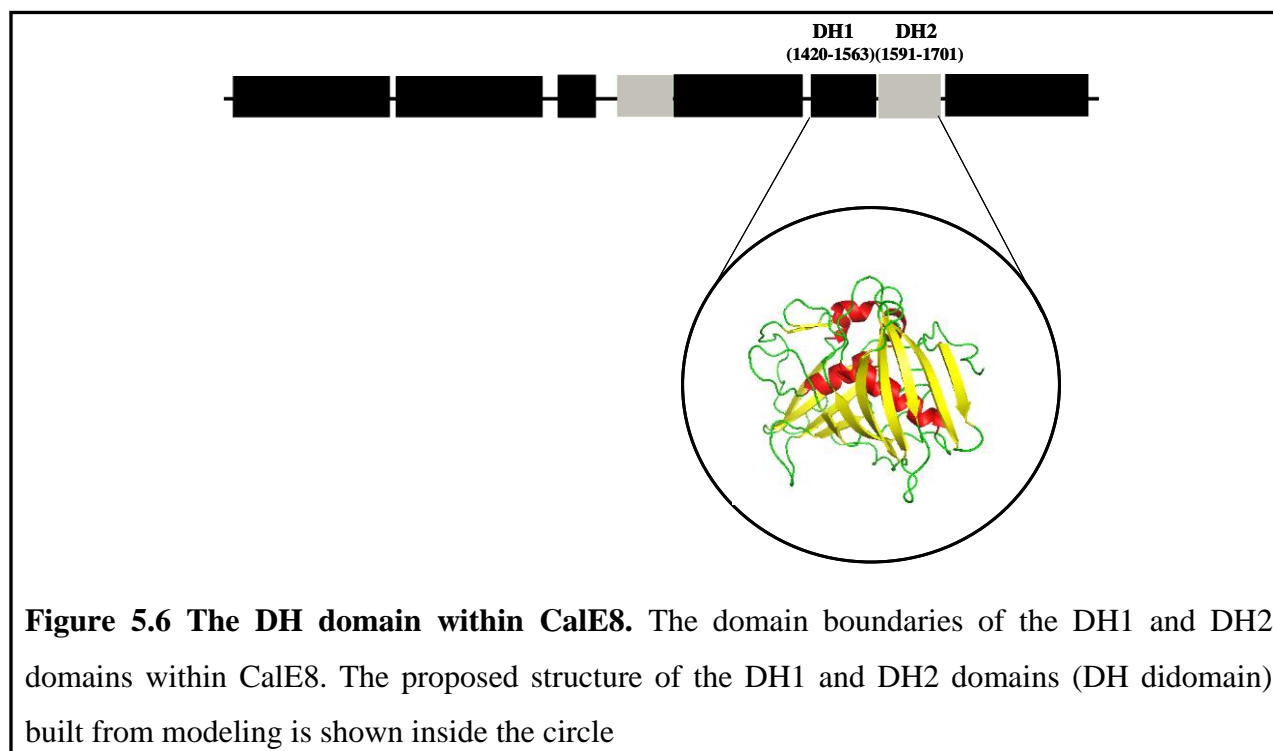
The results from the functional studies also clearly indicate that malonyl CoA is the preferred substrate of AT_{CALE8} domain and the sequence and phylogenetic analyses and homology modeling have substantiated this claim. The AT domain of CalE8, must therefore exhibit a unique self-decarboxylation mechanism to convert the malonyl CoA substrate to acetyl CoA.

The structural modeling studies have shown that the KR2$_{CalE8}$ domain adopts the Rossman fold (**Fig. 5.5**) and is phylogenetically closer to the modular PKSs like DEBS KR and tylKR compared to actKR.

KR2
(1154-1414)

Figure 5.5 The KR domain within the CalE8. The domain boundaries of the KR domain within CalE8. The proposed structure of the KR domain built from modeling is shown inside the circle

The role of the KR domain in polyketide biosynthesis is to reduce the β-keto group to hydroxyl group. Similar to the KRs from both FAS and PKS families CalE8 KR domain was also found to be made of two subdomains. The catalytic subdomain, assigned as KR2$_{CalE8}$ harbors the typical SDR motif, TGxxxGxG conserved as TGggkGiT where the last glycine residue is replaced by threonine. The catalytic aspartic acid residue is conserved. The characteristic active site tetrad Asn, Ser, Tyr and Lys seen in other representative KRs is conserved in the KR2$_{CALE8.}$ The HxAG motif conserved among the modular PKS KRs, DEBS KR and tylKRs is also found conserved in KR2$_{CALE8}$. However, KR2$_{CalE8}$ carries neither the LDD motif conserved in modular type I KR domains like DEBS-KR and tylKR nor the PGG motif conserved in type II KR domains like

151

actKR. These motifs are determinants of the reduction stereochemistry. Sequence alignment with these sequences shows a unique NTP motif at the same position. The Pro residue in the NxP motif is expected to be sufficient to determine the stereochemistry. Homology based structural modeling shows that the active site residues forming the tetrad (Lys 157, Ser 181, Asn 198 and Tyr 194) were spatially conserved. The unknown region found upstream KR was found to adopt a methyl transferase-like structure. Comparative analysis with other KR domains from modular PKSs and FASs has led us to assign this domain as the structural subdomain of the intact KR domain (referred to as $KR1_{CalE8}$) within CalE8. Hence, the intact KR domain of CalE8 has been confirmed to be composed of two units - the structural and the catalytic subdomains. The homology based structural modeling studies have shown that the DH_{CALE8} domain yielded a clear double hot-dog fold (**Fig. 5.6**).

Figure 5.6 The DH domain within CalE8. The domain boundaries of the DH1 and DH2 domains within CalE8. The proposed structure of the DH1 and DH2 domains (DH didomain) built from modeling is shown inside the circle

Despite its limited sequence similarities, the $DH1_{CalE8}$ domain carries the characteristic HXXXGXXXXP and DXXXQ motifs. The $DH1_{CALE8}$ domain was found to be phylogenetically closer to the DH domain of the modular PKS, DEBS DH. Homology based structural modeling revealed that the $DH1_{CALE8}$ domain modeled yielded a clear hot-dog fold. Similar to DEBS DH, the Gly residue within the highly conserved HXXXGXXXXP motif was found to induce a turn to allow van der Waals interactions between the His and Pro residues in $DH1_{CalE8}$. The unknown region downstream of the DH domain was initially assumed to be a part of the PPTase domain. However, the homology based structural modeling of this domain showed that the domain adopted a hotdog fold and was the other half of the intact DH domain. The structural modeling of the DH1 and DH2 domains exhibited a clear double hotdog fold and confirmed that the region between the DH1 and PPTase domains is the second half of the intact DH domain.

5.1.4 Domain architecture of CalE8

Structural and functional analyses the domains of the CalE8 using biochemical, biophysical and bioinformatics techniques, have given us a clear picture on the boundaries and the functions of the domains. The activities of domains like ACP, PPTase and AT domains were examined following heterologous expression of these domains as stand-alone proteins. The rest of the domains were examined using bioinformatics tools leading to the assignment of boundaries and functions (**Fig. 5.7**). The two unknown regions were also examined and assigned as the KR1 and DH2 domains respectively. The overall domain architecture has been validated and proposed.

Figure 5.7 The domain architecture of CalE8. The arrangement of the component functional domains within CalE8 and their domain boundaries.

5.2. Future directions

The aim of the project was to create the basic understanding of the domains of CalE8 at structural and functional levels. Simultaneously, examination of the activity of full length protein was carried out in our lab. With the experiments conducted in my thesis, we have gained a much deeper understanding into the organization and functions of the iterative type I PKS, CalE8. The multidomain CalE8 provides an excellent platform to conduct a multitude of investigations to understand iterative biosynthesis.

The assignment of the PPTase domain has given rise to a few important questions. With the inclusion of the DH2 subdomain (earlier thought of as the N-terminal extension of the PPTase itself), the protein is expressed as a soluble functional protein, whereas in the absence of this domain, it exists as non-functional soluble aggregates. Further examination of the fact that the DH2 domain plays a significant role in the folding of PPTase might also provide us new insights into inter-dependency between domains in the PKS to promote folding and solubility. The roles adopted by the DH2 domain are functional along with DH_{CalE8} and structural along with PPTase. Such an illustration is significant in understanding structural biology of proteins and protein engineering. This could be examined further from structural and functional perspectives. X-ray

crystallographic studies on the DH1, DH2 and PPTase domains individually or in combinations will provide us a clearer view of the structure and function of these domains.

Since the expression levels were too low for any reasonable analysis, further experimentation was not possible. Enhancement of solubility of the domains can be sought with more optimizations incorporating new vectors, expression systems, solubility tags and optimization of growth conditions. Combinatorial expression of the domains might also provide soluble proteins as observed in the case of $DH2_{CalE8}$-PPTase. Careful inspection of this "structural-symbiosis" model in the enhancement of solubility might generate newer understanding in protein structure, folding and solubility. The domains and didomains when expressed as soluble proteins could be subjected to intensive characterization.

Some of the domains can be cloned together and expressed as didomains and tridomains and their activities can be investigated. Attempts were made to express the KS and AT domains together as the KS-AT didomain and KR and DH domains as the KR-DH didomain as part of the thesis. Further optimizations can be performed with the combinatorial didomains of KS and AT domains. Both these domains are expected to work synergistically in determining the substrate specificity and the chain length of the polyketide in polyketide biosynthesis. Examination of the domains in combination with the KS-AT linker will elicit interesting details as the KS domain of CalE8 resembles modular PKSs, while the AT domain is unique with limited similarity to AT domains that employ malonyl CoA as starter molecules. The CalE8 has been shown to be one of the few PKSs that convert malonyl CoA to acetyl CoA. Studying the KS-AT didomain will shed light on this unique self decarboxylation reaction of CalE8.

For functional characterization, resolving the *holo-me*ACP peak and the AT_{CalE8} peak posed a problem since both the peaks appeared at the same region. Various solvent systems and columns should be tried to resolve the peaks. Resolving these two peaks will allow easier analysis and provide a better understanding of the AT_{CalE8} domain.

Structural modeling of the domains revealed that the two domains of unknown function are the N-terminal and C-terminal extensions of the KR and DH domains respectively. Hence the KR and DH domains can now be cloned along with these two subdomains. The reduction stereochemistry properties of KR and the dehydration properties of the DH domain can be studied in detail.

The combinatorial expression of the domains can also be exploited *in vivo* to generate a strain that has a complete or partial set of the selected domains being expressed to function in a co-operative manner. Such strains will be vital in studying the functionality of the domains and also in studying the polyketide precursors.

The long-term objective of the project will be to generate analogs of the Calicheamicin antitumor molecule that are potent and medicinally viable. With the knowledge obtained from the observations and results, the domains can be suitably employed to produce important intermediates.

The novel phosphodiesterase, *Pa*AcpH that was expressed and purified as a soluble protein, can be a potentially valuable tool in studying biosynthetic mechanisms, especially pertaining to PKSs. The phosphodiesterase activity of *Pa*AcpH can be employed to cleave the growing polyketides on CalE8 and other PKSs. The growing polyketide intermediates tethered on the CP within a PKS, form the templates of commercially and medicinally important antibiotics like

erythromycin and calicheamicin and also physiologically important non-ribosomal peptides. It could also be used to examine the modifications on CPs. Release of these intermediates from the polyketide synthase would be useful in examination of the biosynthetic intermediates and study of the mechanisms. The *Pa*AcpH phosphodiesterase has considerable potential to be patented as a commercially and medicinally important enzyme.

References

1   Collie, N. and Myers, W.S., *VII.-The formation of orcinol and other condensation products from dehydracetic acid.* J. Am. Chem. Soc, Transactions, 1893. 63: p. 122-128.

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

3   Thorson, J.S., et al., *Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research.* Curr. Pharm. Des. 2000. (6): p. 1841–1879.

4   Smith, A. L. and K. C. Nicolaou. *The enediyne antibiotics.* J. Med. Chem. 1996. (39): p. 2103-2117.

5   Edo, K., et al., *Naphthalenecarboxylic acid from neocarzinostatin (NCS).* J. Antibiot. 1980. 33(3): p. 347-351.

6   Edo, K., et al., *Unstability of neocarzinostatin-chromophore.* J. Antibiot. 1986. 39(4): p. 535-540.

7   Lee, M.D., et al., *Calicheamicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calichemicin .gamma.1I.* J. Am. Chem. Soc. 1987. 109(11): p. 3464-3466.

8   Lee, M.D., et al., *Calicheamicins, a novel family of antitumor antibiotics. 2. Chemistry and structure of calichemicin .gamma.1I.* J. Am. Chem. Soc. 1987. 109(11): p. 3466-3468.

9   Lee, M.D., et al.,*Calicheamicins, a Novel Family of Antitumor Antibiotics. Part 4. Structure Elucidation of Calicheamicins βBr 1, γBr 1, αI 2, αI 3, βI 1, γI 1, and δI 1* J. Am. Chem. Soc. 1992. 114, p. 985-997.

10   Golik, J., et al., *Esperamicins, a novel class of potent antitumor antibiotics. 2. Structure of esperamicin X.* J. Am. Chem. Soc. 1987. 109(11): p. 3461- 3462.

11   Golik, J., et al., *Esperamicins, a novel class of potent antitumor antibiotics. 3. Structures of esperamicins A1, A2, and A1b.* J. Am. Chem. Soc. 1987. 109(11): p. 3462-3464.

12   Konishi, M., et al., *Crystal and molecular structure of dynemicin A: a novel 1,5-diyn-3-ene antitumor antibiotic* J. Am. Chem. Soc. 1990. 112, p. 3715-3716.

13   Golik, J., et al., *AT2433-A1, AT2433-A2, AT2433-B1 and AT2433-B2 novel antitumor compounds produced by Actinomadura melliaura. II. Structure determination.* J. Antibiot (Tokyo) 1989. 42: p. 1784-1789.

14   Konishi, M., et al., *Esperamicins, a novel class of potent antitumor antibiotics. I. Physico-chemical data and partial structure.* J. Antibiot (Tokyo) 1985. 38: p. 1605-1609

15   Sugiura, Y., et al., *DNA intercalation and cleavage of an antitumor antibiotic dynemicin that contains anthracycline and enediyne cores.* Proc. Natl. Acad. Sci. U S A 1990. 87: p. 3831-3835.

16   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. 5548-53.

17   Miyagawa, N., et al., *DNA cleavage characteristics of non-protein enediyne antibiotic*

N1999A2. Biochem. Biophys. Res. Commun. 2003. 306(1): p. 87-92. .

18   Zein, N., et al., *Kedarcidin chromophore: an enediyne that cleaves DNA in a sequence-specific manner*. Proc. Natl. Acad. Sci. U S A. 1993. 90(7): p. 2822-6.

19   Zein, N., et al., *Selective proteolytic activity of the antitumor agent kedarcidin.* Proc. Natl. Acad. Sci. U S A 1993. 90: 8009-8012.

20   Okuno, Y*., et al., *Interaction of C-1027 Chromophore with d(GTATAC)$_2$: A Binding Model Based on NMR Experiments* J. Am. Chem. Soc. 1996. 118, p. 4729-4730

21   Xu J., et al., *DNA damage produced by enediynes in the human phosphoglycerate kinase gene in vivo: esperamicin A1 as a nucleosome footprinting agent.* Biochemistry. 1998. 37(7): p. 1890-7.

22   Biggins, J. B., et al., *A continuous assay for DNA cleavage: the application of "break lights" to enediynes, iron-dependent agents, and nucleases.* Proc. Natl. Acad. Sci. U S A 2000. 97: p. 13537-13542.

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* J. Am. Chem. Soc. 2002. 94, p. 660-661.

24   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.* J. Am. Chem. Soc. 1972. 94(2): p. 660-661.

25   Gredicak M, Jerić I. *Enediyne compounds - new promises in anticancer therapy.* Acta Pharm. 2007. 57(2): p. 133-50.

26   Grissom, J. W., et al., *The chemistry of enediynes, enyne, allens and related compounds,* Tetrahedron. 1996. 52, p6453–6518.

27   Dedon, P.C., et al., *Exclusive production of bistranded DNA damage by calicheamicin.* Biochemistry. 1993. 32(14): p. 3617-22. Erratum in: Biochemistry 1993. 32(27): p. 7064.

28   Ji, N., et al., *Enantioselective synthesis of N1999A2.* J. Am. Chem. Soc. 2006. 128: p. 14825-14827.

29   Kobayashi, S., et al., *Formal total synthesis of neocarzinostatin chromophore*. J. Org. Chem. 2006. 71(2): p. 636-44.

30   Groneberg, R. D, ., et al., *Total synthesis of calicheamicin .gamma.1I. 1. Synthesis of the oligosaccharide fragment.* J. Am. Chem. Soc, 1993. 115, p. 7593-7611.

31   Hitchcock, S.A., et al., *A convergent total synthesis of calicheamicin γ1 I.* Angew.Chem. Int. 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).* J. Am. Chem. Soc. 1995. 117(21): p. 5750- 5756.

33   Nicolaou, K. C., et al., *Design, synthesis and study of simple monocyclic conjugated – enediynes. The 10-membered ring enediyne moiety of the enediyne anticancer antibiotics,* J. Am. Chem. Soc. 1992. 114, p. 7360–7371.

34   Nicolaou, K. C., et al., *Total synthesis of calicheamicin .gamma.1I* J. Am. Chem. Soc. 1992.

114, p. 10082-10084.

35   Nicolaou, K. C., et al., *Chemistry and biology of natural and designed enediynes.* Proc. Natl. Acad. Sci. U S A 1993. 90(13): p. 5881-5888.

36   Nicolaou, K. C., et al., *Cell-specific regulation of apoptosis by designed enediynes.* Proc. Natl. Acad. Sci. U S A 1993. 90, p. 3142–3146.

37   Nicolaou, K. C., et al., *Total synthesis and stereochemistry of uncialamycin.* Ange. Chem. 2007. 46(25): p. 4704-4707.

38   Shair, M.D., et al., *Total synthesis of (±)-dynemicin A.* Ange. Chem. 1995. 34(16): p. 1721-1723.

39   Shair, M.D., et al., *The total synthesis of dynemicin A leading to development of a fully contained bioreductively activated enediyne prodrug.* J. Am. Chem. Soc. 1996. 118(40): p. 9509-9525.

40   Smith, A. L., et al., *Total synthesis of calicheamicin .gamma.1I. 2. Development of an enantioselective route to (-)-calicheamicinone* J. Am. Chem. Soc. 1993. 115, p. 7612-7624.

41   Kawata, S. S., et al., *Synthetic Study of Kedarcidin Chromophore: Revised Structure* J. Am. Chem. Soc. 1997. 119, p. 12012-12013.

42   Ren, F., et al., *Kedarcidin Chromophore: Synthesis of Its Proposed Structure and Evidence for a Stereochemical Revision* J. Am. Chem. Soc. 2007. 129, p. 5381-5383.

43   Jones, G. B., et al., *Synthesis and Photochemical Activity of Designed Enediynes* J. Am. Chem. Soc. 2000. 122, p. 9872-9873.

44   Kraka. E and Cremer. D., *Computer Design of Anticancer Drugs. A New Enediyne Warhead* J. Am. Chem. Soc. 2000. 122, p. 8245-8264.

45   Iida, K. I and Hirama. M., *Synthesis and Characterization of Nine-Membered Cyclic Enediynes, Models of the C-1027 and Kedarcidin Chromophores: Equilibration with a p-Benzyne Biradical and Kinetic Stabilization* J. Am. Chem. Soc. 2002. 117, p. 8875-8876.

46   Nuss, J.M., et al., *Transition-metal-catalyzed strategies for the synthesis of neocarzinostatin chromophore and analogs: intramolecular delivery of palladium controls construction of the biologically relevant dienediyne core* J. Am. Chem. Soc. 1993. 115, p. 6991-6992.

47   Toshima, K., *et al., Molecular Design, Chemical Synthesis, and Study of Novel Enediyne-Sulfide Systems Related to the Neocarzinostatin Chromophore* J. Am. Chem. Soc. 1995. 117, p. 4822-4831.

48   Fouad, F. S., J. M. Wright, et al., *Synthesis and protein degradation capacity of photoactivated enediynes.* J. Org. Chem. 2005. 70: p. 9789-9797.

49   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.* U S A. 2007. 104(45): p. 17632-17637.

50   Pandithavidana, D.R., et al., *Photochemical generation and reversible cycloaromatization of a nine-membered ring cyclic enediyn*e. J. Am. Chem. Soc. 2009. 131(1): p. 351-6.

51   Schreiber, S.L and Kiessling, L.L., *Synthesis of the bicyclic core of the*

esperamicin/calichemicin class of antitumor agents. J. Am. Chem. Soc. 1988. 110(2): p. 631-633.

52   Hensens, O.D., et al., *Biosynthesis of NCS Chrom A, the chromophore of the antitumor antibiotic neocarzinostatin.* J. Am. Chem. Soc., 1989. 111(9): p. 3295-3299.

53   Tokiwa, Y., et al., *Biosynthesis of dynemicin A, a 3-ene-1,5-diyne antitumor antibiotic* J. Am. Chem. Soc. 1992. 114(11): p. 4107-4110.

54   Davies, J., et al., *Uncialamycin, a new enediyne antibiotic.* Org. Lett 2005. 7, p. 5233-5236.

55   Lam, K.S., et al., *Biosynthesis of esperamicin A1, an enediyne antitumor antibiotic* J. Am. Chem. Soc. 1993. 115, p. 12340-12345.

56   Liu, W., et al., *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.* Antimicrob. Agents. Chem. other. 2000. 44(2): p. 382-92.

57   Whitwam, R.E., et al., *The Gene calC Encodes for a Non-Heme Iron Metalloprotein Responsible for Calicheamicin Self-Resistance in Micromonospora* J. Am. Chem.Soc. 2000. 122, p. 1556-1557

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

59   Ahlert, J. E. Shepard., et al., *The calicheamicin gene cluster and its iterative type I enediyne PKS,* Science 2002. 297 p. 1173–1176.

60   Zazopoulos, E., et al., *A genomics-guided approach for discovering and expressing cryptic metabolic pathways.* Nat. Biotechnol. 2003. 21: p. 187-190.

61   Gao, Q and Thorson, J.S. *The biosynthetic genes encoding for the production of the dynemicin enediyne core in Micromonospora chersina ATCC53710.* FEMS Microbiol. Lett. 2008. 282(1): p. 105-14.

62   Van Lanen, S., et al., *Characterization of the Maduropeptin Biosynthetic Gene Cluster fromActinomadura madurae ATCC 39144 Supporting a Unifying Paradigm for Enediyne Biosynthesis* J. Am. Chem.Soc. 2007. 129, p. 13082-13094.

63   Udwary, D. W., et al., *Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica.* Proc. Natl. Acad. Sci. U S A 2007. 104: p. 10376-10381.

64   Williams, P. G., et al., *New cytotoxic salinosporamides from the marine Actinomycete Salinispora tropica.* J. Org. Chem. 2005. 70: p. 6196-6203.

65   Reed, K.A., et al., *Salinosporamides D-J from the marine actinomycete Salinispora tropica, bromosalinosporamide, and thioester derivatives are potent inhibitors of the 20S proteasome.* J. Nat. Prod. 2007. 70(2): p. 269-76.

66   Liu, W., et al., *Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model.* Proc. Natl. Acad. Sci. U S A. 2003. 100(21): p. 11959-63.

67   Semmelhack, M. F. ., et al., *The effect on DNA cleavage potency of tethering a simple cyclic*

enediyne to a netropsin analog, J. Org. Chem. 1994. 59, p. 4357–4359

68   Sudhahar, C. G. and Chin. D. H., *Aponeocarzinostatin – A superior drug carrier exhibiting unusually high endurance against denaturants*, Bioorg. Med. Chem. 2006. 14, p. 3543–3552

69   Schor, N. F., et al., *Exploiting oxidative stress and signaling in chemotherapy of resistant neoplasms,* Biochemistry (Moscow) 2004. 69, p. 38–44.

70   Zein, N. M., et al., *Calicheamicin g1 I and DNA: Molecular recognition process responsible for site-specificity*; Science 1989. 244, p. 697–699.

71   Walker, S., et al., *Structural characterization of a calicheamicin-DNA complex by NMR*, J. Am. Chem. Soc. 1993. 115, p. 7954–7961.

72   Damle, N. K. *Tumour-targeted chemotherapy with immunoconjugates of calicheamicin*, Expert Opin. Biol. Ther. 2004. 4, p. 1445–1452.

73   Saito, I and Nakatani. K., *Design of DNA-cleaving agents*, Bull. Chem. Soc. Jpn. 1996. 69, p. 3007–3019.

74   Boghaert, E.R., et al., *Antibody-targeted chemotherapy with the calicheamicin conjugate hu3S193-N-acetyl gamma calicheamicin dimethyl hydrazide targets Lewisy and eliminates Lewisy-positive human carcinoma cells and xenografts.* Clin. Cancer. Res. 2004, 10(13): p. 4538-49.

75   DiJoseph, J.F., et al., *Potent and specific antitumor efficacy of CMC-544, a CD22-targeted immunoconjugate of calicheamicin, against systemically disseminated B-cell lymphoma.* Clin. Cancer. Res. 2004. 10(24): p. 8620-9.

76   Damle, N.K and Frost, P. *Antibody-targeted chemotherapy with immunoconjugates of calicheamicin.* Curr. Opin. Pharmacol. 2003. 3(4): p. 386-90.

77   Bhattacharyya, S. ., et al., *A unique approach to metal-induced Bergman cyclization: long range enediyne activation by ligand-to-metal charge transfer,* Angew.Chem. Int. Ed. 2005. 44, p. 592–595.

78   Boger, D. L. and Zhou, J., *CDPI3-enediyne and CDPI3-EDTA conjugates: a new class of DNA cleaving agents,* J. Org. Chem. 1993. 58, p. 3018–3024;

79   Ishida, N. K., et al.,*Neocarzinostatin, an antitumor antibiotic of high molecular weight,* J. Antibiot. 1965. 18, p. 68–76.

80   Watanabe, A and Ebizuka, Y. *Unprecedented mechanism of chain length determination in fungal aromatic polyketide synthases.* Chem Biol. 2004. 11(8): p. 1101-6.

81   Meier, J.L., et al., *The chemical biology of modular biosynthetic enzymes.* Chem. Soc. Rev. 2009. 38: p. 2012-2045.

82   Cox, R.J. *Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets.* Org. Biomol. Chem. 2007. 7;5(13): p. 2010-26.

83   Smith, S and Tsai, S.C. *The type I. fatty acid and polyketide synthases: a tale of two megasynthases.* Nat. Prod. Rep 2007. 24(5): p. 1041–1072.

84   Mercer, A.C and M.D. Burkart, *The ubiquitous carrier protein—a window to metabolite biosynthesis*, Nat. Prod. Rep. 2007. 24, p. 750–773.

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

86   Lai, J.R., et al., *Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis,* Biochemistry. 2006. 45, p. 14869–14879.

87   Prescott, D.J and Vagelos, P.R. *Acyl carrier protein.* Adv. Enzymol. Relat. Areas. Mol. Biol. 1972. 36: p. 269-311.

88   Chuman, L and Brody, S. *Acyl carrier protein is present in the mitochondria of plants and eucaryotic micro-organisms.* Eur. J. Biochem. 1989. 184(3): p. 643-9.

89   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,* J. Mol. Biol. 2007. 365 p. 135–145.

90   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.* J. Biol. Chem. 2009. 284(33): p. 22390-400.

91   Płoskoń, E., et al., *A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains.* J. Biol. Chem. 2008. 283(1): p. 518-28.

92   Reed, M.A., et al., *The type I rat fatty acid synthase ACP shows structural homology and analogous biochemical properties to type II ACPs.* Org. Biomol. Chem. 2003. 1(3): p. 463–471.

93   Leibundgut, M., et al., *Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase.* Science. 2007. 316(5822): p. 288-90.

94   Rawlings, M and Cronan, J.E Jr. *The gene encoding Escherichia coli acyl carrier protein lies within a cluster of fatty acid biosynthetic genes.* J. Biol. Chem. 1992. 267(9): p. 5751-4.

95   Ehmann, D.E., et al., *Lysine biosynthesis in Saccharomyces cerevisiae: mechanism of a-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5.* Biochemistry. 1999. 38, p. 6171–6177.

96   Cane, D.E and Walsh, C.T. *The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases.* Chem. Biol. 1999. 6(12): p. 319-25.

97   Alekseyev, V.Y., et al., *Solution structure and proposed domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase.* Prot. Sci. 2007. 16, p. 2093–2107.

98   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.

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

100   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.

101 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 Lett. 1996. 391(3): p. 302-6.

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

103 Mootz, H.D., et al., 4'- *Phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis.* J Biol Chem. 2001. 276, p. 37289–37298.

104 Joshi, A.K., et al., *Cloning, expression, and characterization of a human 40-phosphopantetheinyl transferase with broad substrate specificity.* J. Biol. Chem. 2003. 278, p. 33142–33149.

105 Fichtlscherer, F., et al., *A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation.* Eur J Biochem. 2000. 267(9): p. 2666-2671.

106 Lambalot, R.H., et al., *A new enzyme superfamily - the phosphopantetheinyl transferases.* Chem Biol. 1996. 3(11): p. 923-936.

107 Flugel, R.S., et al., *Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in Escherichia coli.* J. Biol. Chem 2000. 275, p. 959–968.

108 Austin, M.B and Noel, J.P. *The chalcone synthase superfamily of type III polyketide synthases.* Nat. Prod. Rep 2003. 20(1): p. 79–110.

109 Tang, Y., et al., *The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B. synthase.* Proc. Natl. Acad. Sci. USA 2006. 103(30): p. 11124–11129.

110 Witkowski, A., et al., *Characterization of the beta-carbon processing reactions of the mammalian cytosolic fatty acid synthase: role of the central core.* Biochemistry 2004. 43(32): p. 10458–10466.

111 Witkowski A., et al., *Mechanism of the β-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase.* Biochemistry 2002. 41(35): p. 10877–10887.

112 Zhang, Y.M., et al., *Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of Streptococcus pneumoniae.* J. Biol. Chem. 2006. 281(25): p. 17390–17399.

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

114 Olsen, J.G., et al., *The X-ray crystal structure of beta-ketoacyl [acyl carrier protein] synthase I.* FEBS Lett. 1999;460(1): p. 46– 52.

115 Price, A.C., et al., *The 1.3-Angstrom-resolution crystal structure of beta-ketoacyl-acyl carrier protein synthase II from Streptococcus pneumoniae*. J. Bacteriol. 2003. 185(14): p. 4136–4143.

116 Qiu, X., A. E. Choudhry, et al. . *Crystal structure and substrate specificity of the beta-ketoacyl-acyl carrier protein synthase III (FabH) from Staphylococcus aureus.* Prot. Sci. 2005. 14: p. 2087-2094.

117   Keatinge-Clay, A.T., et al., *An antibiotic factory caught in action.* Nat. Struct. Mol. Biol. 2004. 11: p. 888–893.

118   Pan, H., et al., *Crystal structure of the priming beta-ketosynthase from the R1128 polyketide biosynthetic pathway.* Structure 2002. 10(11): p. 1559–1568.

119   Tang, Y., et al., *Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B. synthase*. Chem. Biol. 2007. 14(8): p. 931–943.

120   Del Vecchio, F., et al., *Active-site residue, domain and module swaps in modular polyketide synthases.* J. Ind. Microbiol. Biotechnol. 2003. 30(8): p. 489–494.

121   Lau, J., et al., *Substrate specificity of the loading didomain of the erythromycin polyketide synthase.* Biochemistry 2000. 39(34): p. 10514–10520.

122   Lau, J., et al., *Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units*. Biochemistry 1999. 38(5): p. 1643–1651.

123   Liou, G.F, Khosla C. *Building-block selectivity of polyketide synthases*. Curr. Opin. Chem. Biol. 2003. 7 (2): p. 279–284.

124   Liou, G.F., et al., *Quantitative analysis of loading and extender acyltransferases of modular polyketide synthases.* Biochemistry 2003. 42(1): p. 200–207.

125   Marsden, A.F., et al., *Stereospecific acyl transfers on the erythromycin-producing polyketide synthase.* Science 1994. 263(5145): p. 378–380.

126   Khosla, C., et al., *Tolerance and specificity of polyketide synthases.* Annu. Rev. Biochem. 1999. 68: p. 219–253.

127   Haydock, S., et al., *Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases.* FEBS Lett. 1995. 374: p. 246–248.

128   Yadav, G., et al., *Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases*. J. Mol. Biol. 2003. 328(2): p. 335–363.

129   Reeves, C.D., et al., *Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations*. Biochemistry. 2001. 40(51): p. 15464–15470.

130   Keatinge-Clay, A.T and Stroud, R.M. *The structure of a ketoreductase determines the organization of the betacarbon processing enzymes of modular polyketide synthases.* Structure 2006. 14(4): p. 737–748.

131   Keatinge-Clay, A.T. *A tylosin ketoreductase reveals how chirality is determined in polyketides.* Chem. Biol. 2007. 14(8): p. 898–908.

132   Korman, T.P., et al., *Structural analysis of actinorhodin polyketide ketoreductase: Cofactor binding and substrate specificity*. Biochemistry 2004. 43(46): p. 14529–14538.

133   Korman, T.P., et al., *Inhibition kinetics and emodin cocrystal structure of a type II polyketide ketoreductase.* Biochemistry 2008. 47(7): p. 1837–1847.

134   O'Hagan, D. *Biosynthesis of fatty acid and polyketide metabolites.* Nat. Prod. Rep 1993.

10(6): p. 593–624.

135 Persson, B., et al., *Coenzyme-based functional assignments of shortchain dehydrogenases/reductases (SDRs).* Chem. Biol. Interact. 2003. p. 143–144:271–278.

136 Ostergaard, L.H., et al., *Stereochemistry of catalysis by the ketoreductase activity in the first extension module of the erythromycin polyketide synthase.* Biochemistry 2002. 41(8): p. 2719–2726.

137 Dutler, H., et al., *Fatty acid synthetase from pig liver. 2. Characterization of the enzyme complex with oxidoreductase activity for alicyclic ketones as a fatty acid synthetase.* Eur. J. Biochem. 1971. 22(2): p. 213–217.

138 Joshi, A.K and Smith, S. *Construction, expression, and characterization of a mutated animal fatty acid synthase deficient in the dehydrase function.* J. Biol. Chem. 1993. 268(30): p. 22508–22513.

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

140 Reid, R., et al., *A model of structure and catalysis for ketoreductase domains in modular polyketide synthases.* Biochemistry 2003. 42(1): p. 72–79.

141 Maier, T., et al., *Architecture of mammalian fatty acid synthase at 4.5 A resolution.* Science 2006. 311(5765): p. 1258–1262.

142 Keatinge-Clay, A.T., et al., *Crystal structure of the erythromycin polyketide synthase dehydratase.* J. Mol. Biol. 2008. 384(4): p. 941–953.

143 Dillon, S.C and Bateman, A. *The hotdog fold: wrapping up a superfamily of thioesterases and dehydratases.* BMC Bioinfo. 2004. 5: p. 109–123.

144 Kimber, M.S., et al., *The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from Pseudomonas aeruginosa.* J. Biol. Chem. 2004. 279(50): p. 52593–52602.

145 Leesong, M., et al., *Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site.* Structure 1996. 4(3): p. 253–264.

146 Rong, K., et al., *Characterization of a Carbonyl-Conjugated Polyene Precursor in 10-Membered Enediyne Biosynthesis*, J. Am. Chem. Soc. 2008. 130(26): p. 8142-8143.

147 Lim, J., et al., *Solution Structures of the Acyl Carrier Protein Domain from the Highly Reducing Type I Iterative Polyketide Synthase CalE8*. PLos ONE. 2011, 6 (6), e20549

148 Mofid, M. R., et al., *Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases.* Acta. Crystallogr. D Biol Crystallogr. 1999. 55(Pt 5): p. 1098-1100.

149 Mofid, M.R., et al., *Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from Bacillus subtilis: carrier protein recognition and reaction mechanism.* Biochem. 2004. 13;43(14) p4128-36.

150 Khosla, C., et al., *Structure and mechanism of the 6-deoxyerythronolide B synthase.* Ann. Rev. of Biochem. 2007. 76, p. 195–221.

151 Murugan, E and Liang, Z.X, *Evidence for a novel phosphopantetheinyl transferase domain in the polyketide synthase for enediyne biosynthesis*, FEBS Lett. 2008. 582, p. 1097–1103.

152 Barekzi, N., et al., *Genetic characterization of PcpS, encoding the multifunctional phosphopantetheinyl transferase of Pseudomonas aeruginosa.* Microbiology 2004. 150, p. 795–803.

153 Lambalot, R.H. and Walsh, C.T.*Cloning, overproduction, and characterization of the Escherichia coli holo-acyl carrier protein synthase.* J. Biol. Chem. 1995. 270, p. 24658–24661.

154 Kotada., et al., *Structure and catalytic mechanism of a bifunctional hotdog fold thioesterase in enediyne biosynthesis*, J. Biol. Chem. 2008. 284 p. 15739–15749.

155 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.

156 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.

157 Finking, R., et al., *Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in Pseudomonas aeruginosa.* J. Biol. Chem. 2002. 277, p. 50293–50302.

158 Mofid, M.R., et al., *Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp.* J. Biol. Chem. 2002. 277(19): p. 17023-31.

159 Yin, J., et al., *Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase.* Proc. Natl. Acad. Sci. USA 2005. 102, p. 15815–15820.

160 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, p. 883–895.

161 Lomakin, I.B., et al., *The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together*. Cell 2007. 129, p. 319–332.

162 Thomas, J., et al., *Acyl Carrier Protein Phosphodiesterase (AcpH) of Escherichia coli is a Non-Canonical Member of the HD Phosphatase/Phosphodiesterase Family*, Biochemistry 2007. 46, p 129-136.

163 Thomas, J, and Cronan, J.E, *The Enigmatic Acyl Carrier Protein Phosphodiesterase of Escherichia coli,* J Biol Chem. 2005. 280 (41), p. 34675-34683

164 Vagelos, P.R and Larrabee, A.R. *Acyl Carrier Protein. IX. Acyl Carrier Protein Hydrolase,* J Biol Chem. 1967. 242, p. 1776-1781.

165 Fischl, A.S and Eugene, P.K, *Isolation and Properties of Acyl Carrier Protein Phosphodiesterase of Escherichia coli,* J. Bactriol. 1990. p. 5445-5449,

166 Magnuson, K.S., et al., *Regulation of fatty acid biosynthesis in Escherichia coli,* Microbiol. Rev. 1993. 57, p. 522-542

167   Jackowski, S and Rock, C.O. *Turnover of the 4'-phosphopantetheine prosthetic group of acyl carrier protein.* J. Biol. Chem. 1984. 259(3): p. 1891-5.

168   Stover, C.K., et al., *Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen,* Nature 2000. 406, p. 959–964.

169   Raussens, V., et al., *"Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method."* Anal Biochem. 2003. 319, p. 114-21

170   Feng, R., et al., *The functional role of a conserved loop in EAL domain-based c-di-GMP specific phosphodiesterase*, J. Bacteriol. 2009. 191, p 4722–4731.

171   Feng, R., et al., *Catalytic mechanism of c-di-GMP specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa,* J. Bacteriol. 2008. 190, p. 3622–3631.

172   Quadri, L.E.N., et al., *Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases,* Biochemistry 1998. 37, p. 1585–1595.

173   Quadri, L.E.N., et al., *Assembly of the Pseudomonas aeruginosa nonribosomal peptide siderophore pyochelin: in vitro reconstitution of aryl-4,2-bisthiazoline synthetase activity from PchD, PchE, and PchF,* Biochemistry 1999. 38 p. 14941–14954

174   Walsh, C.T. *Polyketide and nonribosomal peptide antibiotics: modularity and versatility,* Science 2004. 303, p. 1805–1810.

175   Sun, H., et al., *Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis,* Chem. Comm. 2009. 47 p. 7399–7401.

176   Hrvoje Petković, et al., *Substrate specificity of the acyltransferase domains of EpoC from the epothilone polyketide synthase.* Org. Biomol. Chem. 2008. 6, p. 500-506

177   Javidpour, P., et al., *Structural and Biochemical Analyses of Regio- and Stereospecificities Observed in a Type II Polyketide Ketoreductase* Biochemistry, 2011. 50 (21), p. 4638–4649

PUBLICATIONS

1. **Elavazhagan Murugan,** Kong Rong, Sun Huihua, Rao Feng, Liang Zhao-Xun. Expression, purification and characterization of the acyl carrier protein phosphodiesterase from *Pseudomonas aeruginosa*. Protein Expression and Purification, 2010, 71, 132-138

2. **Elavazhagan Murugan,** Liang Zhao-Xun**.** Evidence for a novel phosphopantetheinyl transferase domain in the polyketide synthase for enediyne biosynthesis. FEBS Letters, 2008, April 2; 582 (7): 1097-103

3. Kong Rong, Goh Lan Pei, Liew Chong Wai, Ho Qin Shi, **Elavazhagan Murugan,** Li Ben, Tang Kai, Liang Zhao-Xun. Characterization of a carbonyl-conjugated polyene precursor in 10-memberedenediyne biosynthesis. Journal of American Chemical Society, 2008, 130, 8142-3

4. Feng Rao, Yaning Qi, **Elavazhagan Murugan,** Swati Pasunooti, Qiang Ji**.** 2',3'-cAMP hydrolysis bymetal-dependent phosphodiesterases containing DHH, EAL, and HD domains is non-specific:Implications for PDE screening. Biochemical and Biophysical Research Communications. 2010 July30; 398(3):500-5

5. Jackwee Lim, Kong Rong, **Elavazhagan Murugan,** Lawrence Chun Loong Ho, Zhao-Xun Liang, Daiwen Yang**.** Solution Structures of the Acyl Carrier Protein Domain from the Highly Reducing Type I Iterative Polyketide Synthase CalE8. PLos ONE. June 2011, Volume 6, Issue 6