# Structural and functional studies of enzymes from the enediyne polyketide biosynthetic pathway

Liew, Chong Wai

2013

Liew, C. W. (2013). Structural and functional studies of enzymes from the enediyne polyketide biosynthetic pathway. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/53510

https://doi.org/10.32657/10356/53510

# STRUCTURAL AND FUNCTIONAL STUDIES OF ENZYMES FROM THE ENEDIYNE POLYKETIDE BIOSYNTHETIC PATHWAY

LIEW CHONG WAI

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2013

## Acknowledgement

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

Apart from my supervisor, I would also like to thank for the spontaneous help and support given by my current colleagues (Lawrence Ho, Joe Yap, Qian Xin Lei, Yee Hwa, David, Ramya and Edmund) and former colleagues (Kotaka Masayo, Luo Dahai, Sujit Dutta, Mitchell, Rao feng, Qi Ya Ning, Ji Qiang). Thank you all for providing me with such an enjoyable experience in the laboratory. Not forgetting Kong Rong, Chen Ming Wei and Sun Hui Hua, the wonderful PKSE teammate who have been with me through out the many years of research for their extraordinary input and efforts. Moreover, I would like to thank our collaborators Assist. Prof. Tobias Convick and Martina, Protein Production Platform (KI-NTU), for their help in protein engineering.

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

**Table of contents**

**List of figures**

**List of tables**

**Abbreviations**

| | |
|---|---|
| 6-MSA | 6-Methylsalicyclic |
| Acetyl-CoA | Acetyl-Coenzyme A |
| ACP | Acyl Carrier Protein |
| ACY | Acetate |
| ARO | Aromatase |
| AT | Acyl Transferase |
| CalE8 | Calicheamicin Polyketide Synthase |
| CalE7 | Calicheamicin Thioesterase |
| CalU15 | Calicheamicin Acetylenase |
| CLC | Claisen-type Cyclase |
| CYC | Cyclases |
| CoA | Coenzyme A |
| DEBS | 6-Deoxyerythronolide B Synthase |
| DH | Dehydratase |
| DTT | Dithiothreitol |
| DynE8 | Dynemicin Polyketide Synthase |
| DynE7 | Dynemicin Thioesterase |
| DynU15 | Dynemicin Acetylenase |
| EDTA | Ethylenediaminetetra acetic acid |
| EspE | Esperamicin |
| EM | Electron Microscopy |
| ER | Enonyl Reductase |
| FAS | Fatty Acid Synthase |
| HCl | Hydrochloric Acid |
| HPLC | High Performance Liquid Chromatography |
| iPKS | Iterative Polyketide Synthase |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| kDa | Kilo Dalton |
| KR | Ketoreductase |
| KS | Keto-acyl Synthase |
| LB | Luria-Bertani Media |

| | |
|---|---|
| LC-MS | Liquid Chromatography Mass Spectroscopy |
| MadE | Maduropeptin |
| Malonyl-CoA | Malonyl-Coenzyme A |
| MPT | Malonyl/palmitoyl Transferase |
| Methylmalonyl-CoA | Methylmalonyl-Coenzyme A |
| mFAS | Mammalian Fatty Acid Synthase |
| MLA | Malonate |
| MW | Molecular Weight |
| *m/z* | Mass to Charge Ratio |
| NADPH | Nicotinamide Adenine Dinucleotide Phosphate |
| NcsE | Neocarzinostatin |
| $Ni^{2+}$-NTA | Nickel-nitrilotriacetic Acid |
| NR-PKS | Non-reducing Polyketide Syntase |
| PCP | Peptidyl Carrier Protein |
| PCR | Polymerase Chain Reaction |
| PKS | Polyketide Synthase |
| PKSE | Enediyne Polyketide Synthase |
| PDB | Protein Data Bank |
| PT | Product Template Domain |
| PPTase | Phosphopantetheinyl Transferase |
| *g* | Gravity Force |
| R.M.S.D | Root Mean Square Deviation |
| SAX | Small Angle X-ray Scattering |
| SDS-PAGE | Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis |
| SgcE | C-1027 Polyketide Synthase |
| SgcE3 | C-1027 Acetylenase |
| SgcE10 | C-1027 Thioesterase |
| TE/TEBC | Thioesterase |
| TFA | Trifluoroacetic Acid |
| UNBL/UNBU/UNBV | A family of unknown protein that appear specific to enediyne biosynthetic loci |
| UV-Vis | Ultraviolet-Visible |

**Abstract**

Enediyne polyketides are extremely potent antitumor antibiotics with a remarkable core structure consisting of two acetylenic group conjugated to a double bond within either 9-membered ring (bicycle-[7.3.0]-dodecadienediyne) or 10-membered ring (bicycle-[7.3.1]-tridecadiynene). Despite their structural differences, all enediyne antibiotics damage DNA via a rapid enediyne cycloaromatization to form highly reactive diradical species capable of inducing oxidative DNA strand scission. Due to their astonishing capacity in cleaving DNA, some of the enediynes have been conjugated to tumor specific monoclonal antibiotic. Besides, the application of polymer assisted delivery devices have also led to the clinical success of enediyne. With the sequencing of gene clusters responsible for enediyne production in 2002, studies on the biosynthesis of enediynes finally have ushered in a new era. The identification of the "minimal enediyne cassette" in the gene cluster also led the way for scientists to look into the biosynthetic pathway of the enediyne core. With the genomic information in hand, we could set out to explore the biosynthetic origin of the enediyne core of the enediyne polyketide synthase (PKSE). After successfully cloning, expressing and purifying the three proteins encoded by the genes from the minimal cassette, PKSE (CalE8, SgcE, and DynE8), TEBC (CalE7, SgcE10 and DynE7) and UNBL (CalU15, SgcE3 and DynU15), in *E. coli* expression system, a series of biochemical and structural studies have been carrier out to examine the structure and function of the proteins. The roles of these proteins in the early steps of enediyne core biosynthetic pathway were investigated. Various products of CalE8, SgcE and DynE8 have been synthesized and identified through *in vitro* biochemical assays. The crystal structure of the acytransferase domain of DynE8, as well as its binary complexes with glycerol, acetate and malonate, was determined to yield insight into the structure, and function of the substrate-transferring domain. Moreover, the crystal structure of thioesterases (CalE7, SgcE10 and DynE7) and ligand-enzyme complex were determined that together with mutagenesis studies reveal a remarkable induced-fit mechanism during substrate binding and a novel catalytic mechanism in releasing the polyketide product.

# CHAPTER 1 Introduction

## 1.1 Polyketide natural products

### 1.1.1 Polyketides

Natural products can be defined as a chemical compound or substances produced by living organism, found in nature, that usually has a pharmacological or biological activity for use in pharmaceutical drug discovery and drug design. They have been used in their crude forms by human beings for medicinal purposes for centuries. Natural products are biosynthesized as part of secondary metabolism in organisms. Primary metabolism encompasses all processes in living organism which is necessary for the maintenance of life; whereas, secondary metabolism is concerned with the production of molecules which are not absolutely necessary for life, such as antibiotics, pigments, pheromones and signaling molecules in quorum sensing.

Polyketides natural products are a remarkable class of secondary metabolites produced by bacteria, fungi, plants and sea urchin[1] with a vast variety of biological activities and pharmaceutical properties that involved in targeting both acute and degenerative diseases. Many polyketides are in clinical use as antibiotics (erythromycin and tetracycline), anticancer drugs (daunorubicin and epothilone), antiparasitics (avermectin), antifungals (amphotericin B), immunosuppressant (rapamycin), and antihypercholesterolemic (lovastatin)[2; 3]. The success of polyketides against a wide range of cellular targets is mainly due to their great structural diversity. The class ranges from 6-methylsalicyclic acid (6-MSA) and orsellinic acid, which contain only eight carbon, up to maitotoxin[4], a marine polyketide which is the largest and most toxic non-biopolymer natural product known with 168 carbons in its backbone. Despite this great array of structures, all polyketides are biosynthesized by a common mechanism from a shared pool of building blocks.

### 1.1.2 The acetate hypothesis

The study of polyketides was begun in 1890 by James Collie in an attempt to classify a type of aromatic molecules of synthetic origin during his studies on pyrones[5]. In Collie's theory, the biosynthesis of phenolic natural products with specific hydroxyl substitutions in living organisms could arise from a β-polyketone origin. Furthermore, his theory also raised the possibility that polyketones were being derived from polymerization of ketenes. Thus, the term polyketide came into existence with reference to polyketones and their phenolic derivatives. Sadly, Collie's notion did not receive serious attention from other influential organic chemist until half a century later. In 1950s, a leading natural product chemist, Arthur Birch, further developed Collie's theory and suggested the polyketones might originate from the repetitive condensation of acetate units. Birch tested his hypothesis by administering isotopically $^{14}C$ labeled acetates for the production of 6-methylsalicylic (6-MSA) in a polyketide-producing organism[6]. The serendipitous discovery of labeled 6-MSA along various phenolic natural products validated and proved his hypothesis. At last, the scientific world realized the significance of Collie's theory and the term polyketide was then universally adopted. Following the formulation of the polyketide theory, rigorous research in the field of polyketide natural products ensued and the number and diversity of polyketide structures discovered increased at a great pace[7; 8; 9; 10].

## 1.2 Fatty acid synthase and polyketide synthase

### 1.2.1 Fatty acid synthase

Much work of the early understanding in the field of polyketide biosynthesis came from insights gained in the related field of fatty acid biosynthesis. Fatty acids are key components of the cell, and their synthesis is essential for all organisms except archaea. They are major constituents of biological membranes, energy storage compounds, and messenger substances, and they act as posttranslational protein modifiers and modulate gene expression. The biosynthesis of fatty acid involves a conserved set of chemical reactions for the cyclic stepwise elongation of activated acetyl- and malonyl substrate from coenzyme A[11; 12]. Throughout this

process both acyl compounds are bound to the enzyme catalyzing the reaction, fatty acid synthase (FAS), as thioesters.

The fatty acid synthase (FAS) is divided into distinct functional domains, each of which catalyses a separate reaction in the biosynthetic pathway. In bacteria and plants, fatty acid biosynthesis is accomplished by a series of mono-functional proteins in a dissociated type II fatty acid synthase (FAS) system[13]. Meanwhile, the type I FASs of fungi and animals are huge multifunctional polypeptides that integrate all steps of fatty acid synthesis into large macromolecular assemblies. Fungal FAS is a 2.6-MD $\alpha_6\beta_6$-heterododecamer with the catalytic domains distributed along two polypeptides[11; 14; 15], whereas mammalian FAS (mFAS) consists of a 270-kD polypeptide chain, comprising all seven required domains, that assembles into homodimers for enzymatic activity[12; 16].

In spite of the variation in structural organization, all organisms employ a conserved set of chemical reactions for fatty acid biosynthesis[13; 16]. *De novo* synthesis involves the transfer of activated acetyl and malonyl substrate from coenzyme A (CoA) to the prosthetic phosphopantetheine group of acyl carrier protein (ACP) by acetyl transferase (AT) and malonyl/palmitoyl transferase (MPT). Ketoacyl synthase (KS) condenses them by malonyl decarboxylation to acetoacetyl-ACP, which is further modified at the β-carbon position by NADPH-dependent ketoacyl reductase (KR), dehydratase (DH) and NADPH-dependent enoyl reductase (ER) to yield a saturated acyl product elongated by two carbon units. This acyl group functions as a starter substrate for the next round of elongation, until the growing fatty acid chain reaches a length of 16 to 18 carbon atoms and is released from ACP. In fungi, the palmitoyl end product is transferred back from ACP to CoA by MPT, whereas in the mammalian system, the fatty acid is released from ACP by a thioesterase[17].

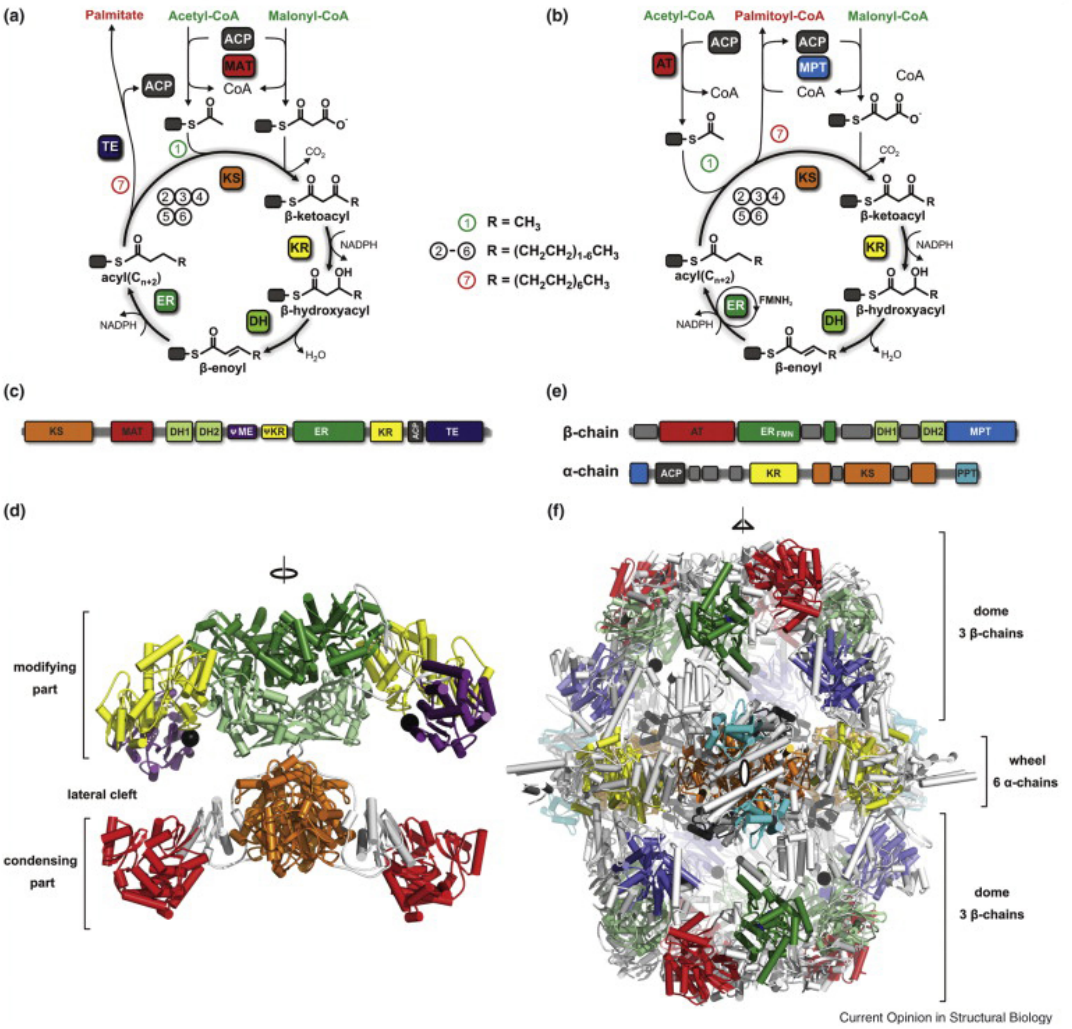



**Figure 1.1** Domain organization and crystal structure of the type I FAS system. (**a-b**) Catalytic reaction cycle of animal and fungal FAS. (**c** and **e**) Linear domain organization at approximate sequence scale showing catalytic (colored) and scaffolding (grey) domains in animal and fungal FAS. The non-catalytic pseudo-KR (ΨKR) and pseudo-methyltransferase (ΨME) domains of animals are colored in yellow and purple, respectively. In animals, all catalytic domains are present on one polypeptide, whereas in yeast, the domains are distributed among β- and α-chains. (**d** and **f**) Crystal structures of animal and fungal FAS. The domains are colored as in (**a**) and (**b**). ACP attachment points are shown as black spheres, two- and three-fold symmetry axes as ellipsoids and triangles, respectively. Animal FAS forms an X-shaped homodimer with a lower condensing and an upper modifying part, which confine two lateral reaction chambers. The fungal FAS folds into an $\alpha_6\beta_6$ heterododecamer consisting of a hexameric central wheel formed by six α-chains, which is capped on both sides by two domes predominantly formed by β-chain trimers. The central wheel divides the hollow interior into two reaction chambers, which are accessible for small molecules from the outside through several holes. The PPT is located at the periphery of the central wheel, outside of the reaction chamber. (This figure is adapted from *Current Opinion in Structural Biology*, Leibundgut, M.T., *et. al.*, 2008, vol. 18(6), 714-725)

While the mechanism of fatty acid biosynthesis has long been known, the architecture of the synthase has remained elusive until recently. High-resolution crystal structures of all components in type II FAS have been obtained, including a number of substrate- and inhibitor-enzyme complexes[13]. These high-resolution structures provide the basis for much of the work done on type I FAS system.

The crystal structure of the 2.6-megadalton $\alpha_6\beta_6$-heterododecamer fungal FAS[11; 14] from *Thermomyces lanuginosus* at 3.1 Å resolution has been solved by Ban's group. All catalytic domains necessary for the synthesis of $C_{16}$ fatty acids are distributed on two multidomain polypeptide chains, β chain (FAS1, 2080 amino acids, 230 kDa) and α chain (FAS2, 1878 amino acids, 210 kDA) (**Figure 1.1**). Two years later, Ban's group reported another crystal structure of the 270-kilodalton homodimeric mammalian FAS[12; 18; 19] from porcine mammary gland. mFAS assembles into an intertwined dimer approximating an "X" shape with an upper modifying portion (containing the DH, ER and KR domains) and a lower condensing portion (containing the KS and MAT domains) (**Figure 1.1**). Apart from that, there are two non enzymatic domains located at the periphery of the modifying part, namely the pseudo-KR (ψKR) and pseudo-methyltransferase (ψME) domains. The ACP and TE domains are not visible in the structure, probably due to their inherent flexibility.

### 1.2.2 Polyketide synthases

Polyketide synthases (PKSs) are structurally and functionally related to fatty acid synthases (FAS), as both enzyme classes catalyzes the condensation of activated primary metabolites (acetyl- and malonyl-CoA) to form β-ketoacetyl polymers linked to the enzyme by thioester bonds. In fatty acid synthesis, this condensation is followed by β-ketoreduction, dehydration and enoyl reduction to yield the final fully reduced (saturated) fatty acid. However, in polyketide synthesis, these reduction steps are partly or completely omitted in a controlled fashion, resulting in highly diverse polyketide chains with respect to the occurrence of β-ketone, β-hydroxyl and alkyl groups. Polyketide synthases have typically been categorized based on their number of subunits (single or multiple) and synthesis

(linear or iterative). Based on domains architecture, there are at least four distinct PKS types: modular type I, iterative type I, type II and type III.

### 1.2.2.1 Type I modular PKS

Type I modular PKS are multifunctional enzymes that are organized into modules, each of which harbors a set of distinct, non-iteratively acting activities responsible for the catalysis of one cycle of polyketide chain elongation. The overall series of reactions catalyzed by type I modular PKSs is analogous to FAS system. Modular PKSs are constructed of polypeptides that, like the FASs, may be 'unimodular' or may contain multiple modules, as many as nine in the gigantic mycolactone PKS[20]. The best-studied modular PKS is the 6-deoxyerythronolide synthase (DEBS) that function as a "molecular assembly line", passing the growing polyketide chain from one module to the next (**Figure 1.2**).

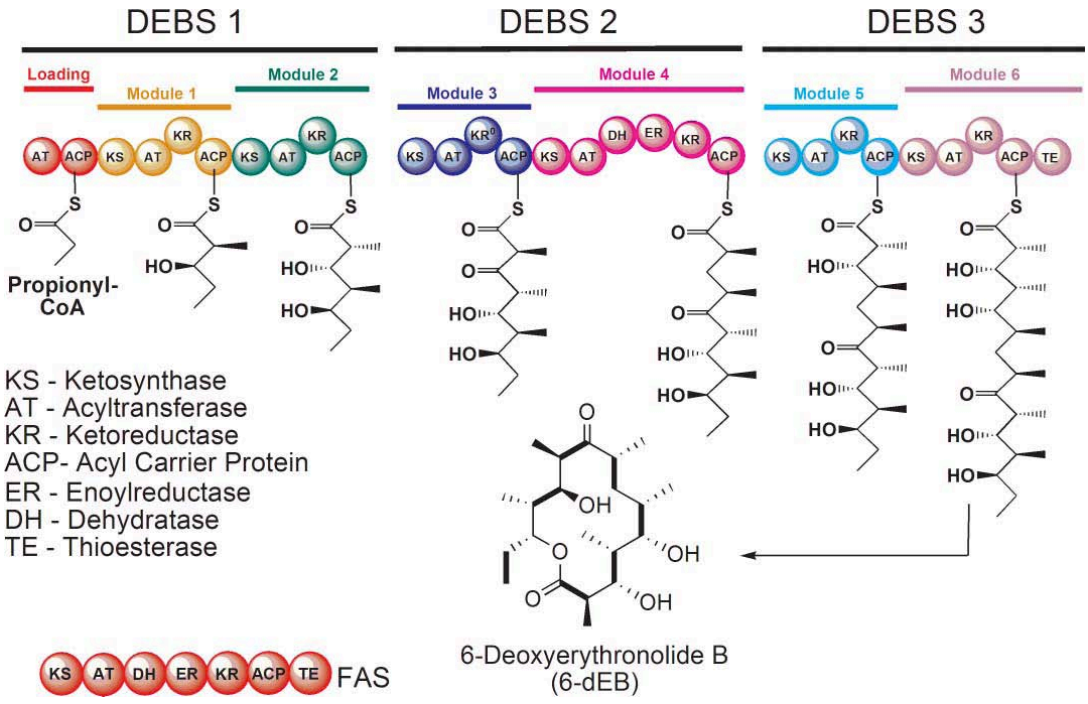



**Figure 1.2** Domain organization of the animal FAS and the DEBS modular PKS. In the FAS, the AT domain is responsible for loading both the starter and chain-elongating substrates, whereas in modular PKSs, these substrates are loaded by separate ATs in the loading and chain-extending modules. The intermediates formed by each of the DEBS modules are shown together with the final 14-membered ring product. The KR domain in module 3 of DEBS is inactive. (This figure adapted from *Natural Product Report*, Smith, S. & Tsai, S,-C., 2007, vol. 24, 1041 – 1072)

It has been proposed that modular type I PKSs adopt a similar structure as mFAS, a head-to-head, tail-to-tail dimer. This structure is considerably more complicated than the iterative mFAS since a modular PKS can contain more than one covalently linked set of modules and must be able to interact with modules on other polypeptide chains. Although the structure of the complete DEBS complex remains elusive, high-resolution x-ray crystal structures of several domains and multidomain components of the overall polyketide assembly line have been reported and are well characterized[18].

#### 1.2.2.2 Type I iterative PKS (iPKS)

Type I iPKSs are mostly found in fungi and consist of a single large polypeptide with multiple domains distributed along it. Basically, type I iPKS only possess a single copy of each catalytic domain that can be employed repeatedly during synthesis of a single polyketide molecule. Type I iPKSs are typically further subdivided into non-reducing, partially reducing and highly reducing, based on which modifications they can introduce into the growing polyketide chain during synthesis. However the action of the core domains (also known as the minimal PKS) remains the same in all subclasses. Type I iPKS in many aspects resembles the fatty acid synthesis, by utilizing the same active sites and reaction mechanisms. The synthesis can be divided into several steps to yield different final products (**Figure 1.3**).

The starter unit on the β- ketosynthase domain (KS) and the extender unit in the acyltransferase (AT) domain (or MAT in non-reducing PKS), loading and movement between the active sites within the enzyme are mediated by the phosphopantetheinylated acyl carrier protein (ACP) (**Figure 1.3**). The acetyl is delivered to the enzyme in the form of acetyl-CoA and bound to the enzyme via a thioester bond. ACP domains in non-reducing iPKSs have been shown to be able to auto-malonylate[21]. The KS domain catalyzes the Claisen condensation reaction between the starter and extender units, driven by decarboxylation of the extender unit[22]. At this point two different options exist: 1) add another ketide unit or 2)

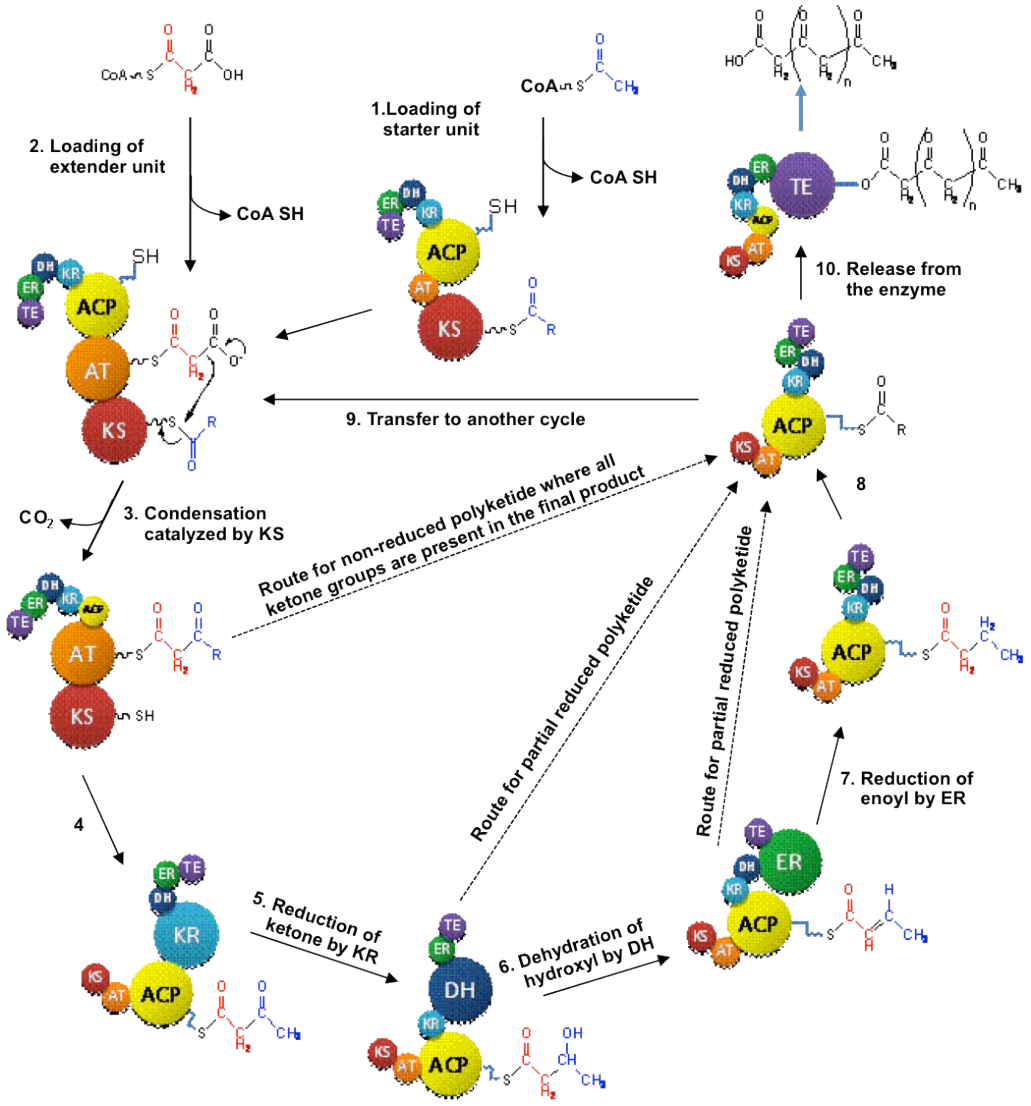



**Figure 1.3** Mechanism for polyketide biosynthesis in type I iterative PKS. Domains are shown as colored balls and active domains in the individual steps are enlarged.

release the polyketide chain from the enzyme. In option 1 the product is transferred back to the KS domain to prepare for a second iteration and another extender unit is loaded into the enzyme. For option 2 to occur, the polyketide must have reached its predetermined length which is unique for each iPKS. The polyketide is transferred to the thioesterase domain (TE) that catalyzes its release from the enzyme[23]. The products of non-reducing PKSs typically undergo intra-chain aldol or Claisen reactions catalyzed by a Claisen-type cyclase domain (CLC), which is related to the TE, resulting in the formation of aromatic structures[24]. Recent results have proven

the existence of a product template domain (PT) that is responsible for placing the polyketide chain correctly to ensure that only one product type is formed[25]. Besides, it also helps to define the molecular rules controlling NR-PKS cyclization specificity.

The ketide units that are added to the growing polyketide chain by the core set of PKS domains (KS, AT and ACP) can be subjected to modifications catalyzed by modifying domains, if present in the enzyme. In reducing iPKSs the ketone group of a ketide unit can be reduced to various degrees catalysed by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains[26]. The KR domain is responsible for reducing the ketone group to a hydroxyl group, the DH domain further reduces the hydroxyl group to an enoyl group, which in turn can be reduced to an alkyl group catalyzed by the ER domain. Reducing iPKSs can also contain domains that add methyl (CmeT) or acetyl (CacT) groups to the reduced polyketide chains, resulting in branching of the backbone chain[27].

#### 1.2.2.3 Type II PKS

Type II PKSs, like bacterial type II FASs, are iteratively associated complexes of discrete proteins, each of which has one or more specific functions in the pathway[28]. The "minimal PKS" consists of two KS-like condensing enzymes ($KS_\alpha$ and $KS_\beta$, only the first of which contributes the active site for the condensation) and an ACP domain shown to be involved in chain initiation (most often with acetate) and elongation with malonate (**Figure 1.4a**). A malonyl-CoA:ACP transferase (MCAT) in type II PKS is likely to assist in recruiting the extender units from primary metabolism. Additional enzymatic subunits, such as ketoreductase (KR), cyclases (CYC) and aromatases (ARO), cooperate with the minimal PKS to dictate the folding pattern of the nascent polyketide chain.

### 1.2.2.4 Type III PKS

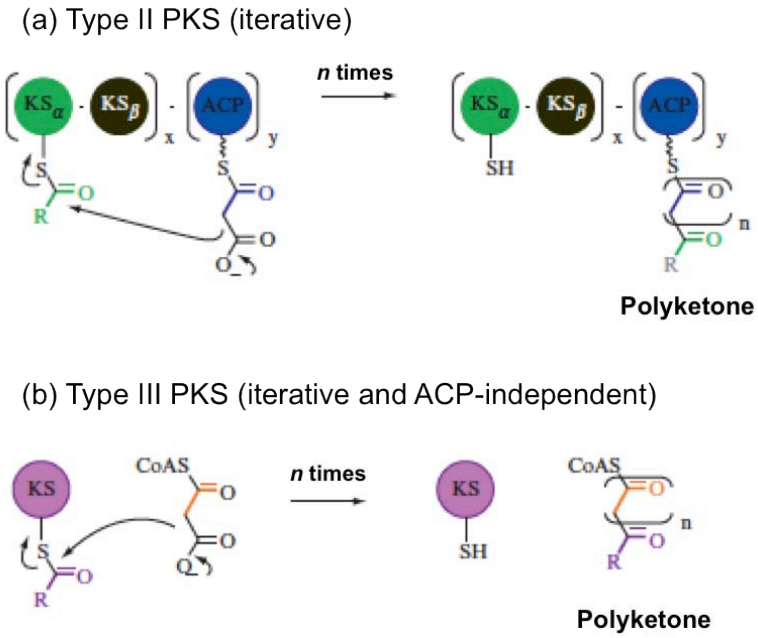



**Figure 1.4** A simplified diagram for the type II PKS and type III PKS. (**a**) Type II PKSs comprise discrete catalytic functions that associate into a productive complex. The "minimal PKS" includes the $KS_{\alpha}$, $KS_{\beta}$ and ACP domain. (**b**) Type II PKSs consist of a single multifunctional active site that, in cooperation with Co-A bound substrates, perform all the steps necessary to assemble a polyketone chain of defined length.

Unlike type I and II PKSs, enzymes from the type III PKS superfamily of plants, fungi and bacteria (also known as chalcone and stilbene) use single KS-like active sites to catalyze the repetitive condensation of acetate units to a CoA-derivatized starter molecule (**Figure 1.4b**). Typically, mono- and bi-cyclicaromatic products are obtained[29]. Chain extension is often followed by intramolecular condensation and aromatization of the linear intermediate and importantly, all in same PKS active site cavity. The "Über-adaptable" group of catalysts generates high diversity by varying the choice of acyl-CoA starter unit, the number of elongation steps and the mechanism of cyclization. Meanwhile, downstream enzymes cause additional pathway branching, variably transforming the initial scaffolds to produce a range of different compounds.

## 1.3 Enediyne polyketide

Among the polyketide products, naturally-occurring enediyne antibiotics are some of the most potent antineoplastic agents ever discovered. Since its discovery in the mid-1980s, the enediyne antibiotics have been a subject of great interest and intensive research among chemists and biologists. The enediyne antibiotics possess an exceptional biological profile due to their unique molecule structure, striking mode of action and high potency[30; 31]. To date, thirteen members of this unique family have been discovered and structurally confirmed, which includes two enediynes probably isolated as inactive degradation products.

The first member of enediyne antibiotic, neocarzinostatin chromophore (NCS), was reported in 1965[32] and further characterized by Edo and co-workers in 1985[33; 34]. The structure of neocarzinostatin features a bicyclo-[7.3.0]-dodecadienediyne core with an epoxide masked 1,5-diyne-3-ene unit encompassed in a 9-membered ring. Since the unveiling of the neocarzinostatin chromophore structure, the enediyne family has grown steadily. Two years later, the discovery of calicheamicin[35] and esperamicin[36], followed by the report on dynemicin[37] characterized in 1989, depicted a different ring structure from that of neocarzinostatin, with a bicyclo-[7.3.1]-tridecadiynene core containing a 1,5-diyne-3-ene unit. Since then, the enediyne are classified into 9-membered and 10-membered families, namely bicyclo-[7.3.0]-dodecadienediyne and bicyclo-[7.3.1]-tridecadiynene. The structure of the different members for the two families of enediyne is illustrated in **Figure 1.5** and **Figure 1.6**. The 9-membered enediyne family, known as chromoprotein enediyne, possesses a novel 9-membered ring chromophore core structure and requires a specific associated protein for chromophore stabilization and transportation within cells. In contrast to other chromoproteins, N1999A2 exists as an enediyne chromophore alone although its structure is very similar to those of the other chromoprotein chromophores[38]. In contrast, the 10-membered enediynes, known as non-chromoprotein enediyne, do not require any additional stabilization factors. Furthermore, they can be further sub-categorized into calicheamicin-like enediyne and dynemicin-like enediyne with

slight differences lying in both the bicyclic enediyne core and the anthraquinone moiety.

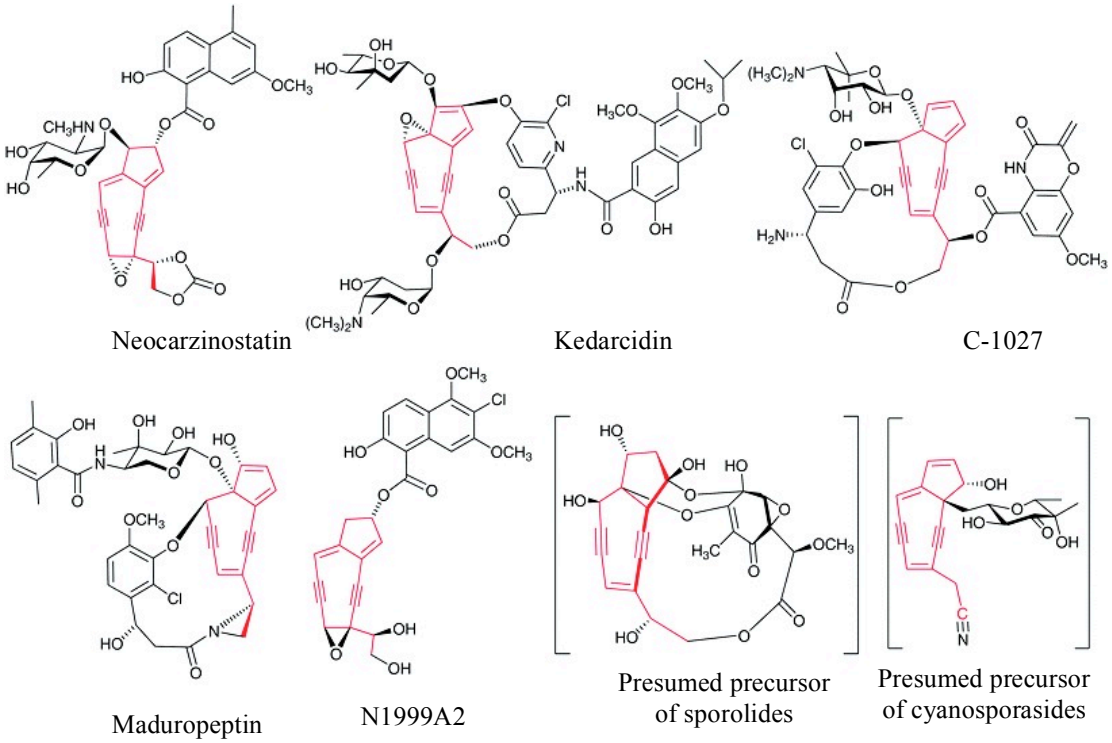



**Figure 1.5** 9-membered enediynes with enediyne moiety structure highlighted.

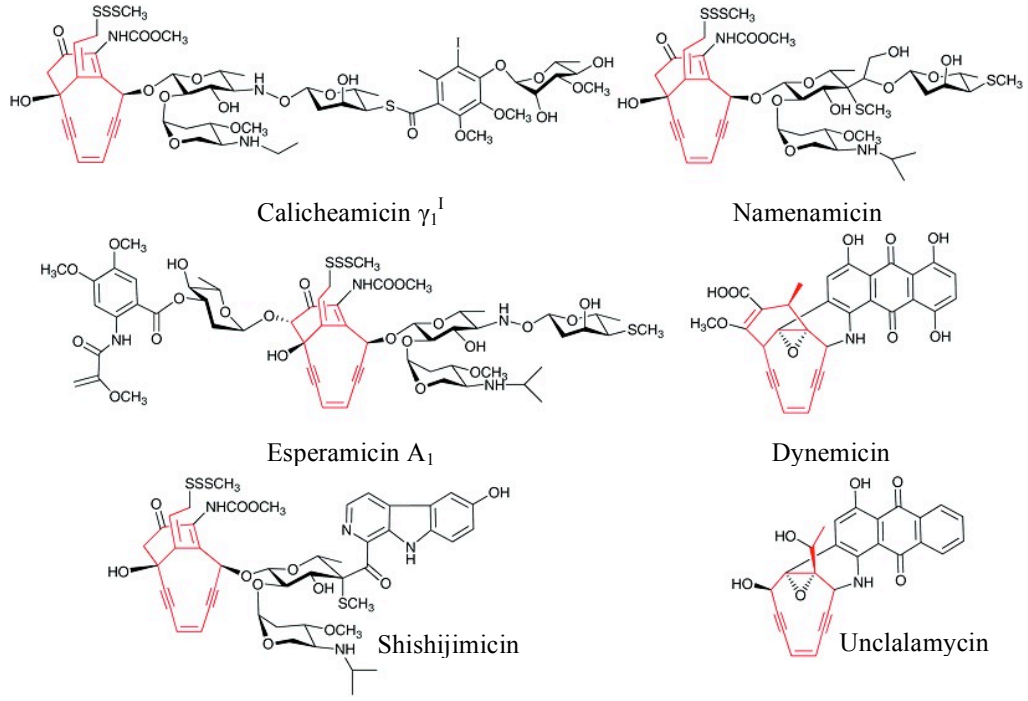


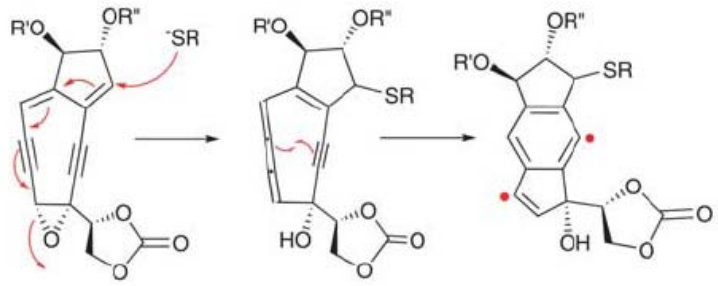

**Figure 1.6** 10-membered enediynes with enediyne moiety structure highlighted.

### 1.3.1 Structure and mechanism of action

The remarkable molecular architecture with its unusual reactivity has sparked great interest among chemists and biologist in the biosynthetic origin of the enediyne cores. Despite differences in the detailed structure, the common structural motif among enediyne antibiotics is the enediyne moiety (Z-hexa-1,5-diyn-3-ene) found embedded within a 9- or 10-membered ring. Basically, all enediyne antibiotic consist of three crucial functional domains responsible for their unprecedented DNA cleaving activity[31; 39]: 1) an aglycone warhead, consisting of two acetylenic group conjugated to a double bond or incipient double bond within either a 9- or 10-membered ring, attacks double helical DNA after cyclization; 2) a delivery system composed of oligosaccharide fragments that transports the molecule to the DNA target[40; 41] and 3) a triggering device that activates the molecule to generate highly reactive benzene diradicals that strips hydrogen atoms from the sugar phosphate backbone of DNA strands[30].

Two widely accepted cyclization mechanism of the enediyne core were proposed: Bergman cyclization (enediyne) and Myers-Saito (enyne-allene) reaction both of which yielding a benzenoid diradical[42; 43; 44] (**Figure 1.7**). In the case of C-1027 and calicheamicin-like enediynes, the formation of diradical intermediates are generated upon cycloaromatization via nucleophilic addition to the bicyclic enediyne warhead, whereas, the cycloaromatization of the free-standing chromophore neocarzinostatin undergoes spontaneously even at ambient temperature, due to the high reactivity of the chromophore. Once enediyne moieties are tightly bound to the minor groove of DNA, an abstraction of hydrogen atom occurs on the DNA molecule in the presence of the diradicals generated, which leads to the formation of a carbon centered radical on the ribose. Subsequently, the DNA molecule will undergo double- or single stranded scission via an oxidative radical mechanism with the presence of molecular oxygen.

**Myers-Saito mechanism**




**Bergman mechanism**

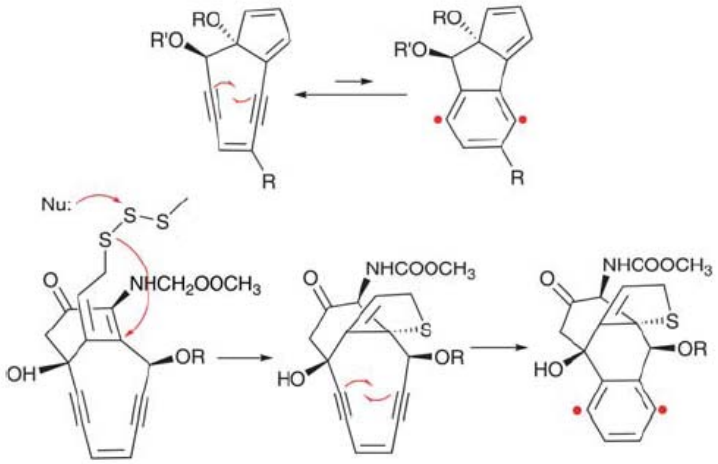



**Figure 1.7** Mechanisms of diradical formation for 9- and 10-membered enediynes.

Notwithstanding its remarkable antitumor activity, clinical usage of enediyne was initially disappointing because of inadequate selectivity and specificity. Therefore, to harness the rigorous DNA-cleavage activity of enediyne, polymer-based and antibody-conjugated derivatives of enediyne natural products has been developed[45; 46; 47]. The first polymer-conjugated form of neocarzinostatin was developed in 1994 and later known as zinostatin stimalamer (SMANCS). SMANCS showed remarkable specificity towards tumor cells and was being used for the treatment of hepatocellular carcinoma in Japan[48]. Meanwhile, the development of antibody-conjugated enediynes has also been explored, which has yielded a few successful drugs for targeting tumor cells through specific antibody-antigen interactions. Gemtuzumab ozogamicin (Mylotarg®), which consists of a

chemically modified calicheamicin $\gamma_1^I$ and a recombinant humanized IgG4κ antibody (anti-CD33 mAb), is one of the antibody-conjugated enediynes that has been approved by the US FDA for the treatment of refractory acute myeloid leukemia that against human tumor xenografts in preclinical model[47]. Three years later, another promising calicheamicin–antibody conjugate, inotuzumab ozogamicin (CMC-544), has been developed for the treatment of non-Hodgkin's lymphoma is currently in clinical trial[49].

### 1.3.2 Peripheral moieties of enediyne chromophores

The peripheral moieties, a part of the physical properties of the enediyne chromophores, play a critical role on the specific DNA sequence recognition. In general, the peripheral moieties or building blocks of most of the enediyne chromophores can be divided into aromatic and sugar moieties. The aromatic moieties, such as benzoxazolinate of C-1027 and orsellenic acid of calicheamicin, are essential for DNA-chromophore interaction through intercalation. Moreover, most of the enediyne chromophores contain mono- or polysaccharides, which are crucial for fine-tuning the interaction between the chromophore and DNA. The addition of these moieties at different positions of the enediyne core further expands the structural diversity of enediyne chromophores.

## 1.4 Study of enediyne biosynthesis from genomic approach

This highly unusual molecular architecture and reactivity has sparked great interest in the biosynthetic origin of the enediyne cores. However, little was known about the genomic aspect of the enediyne systems. Several efforts were devoted to locate the polyketide synthase gene clusters. The discovery of the polyketide synthase gene clusters was only successful in year 2000 when Shen's group accidentally stumbled upon a genetic fragment containing the C-1027 polyketide synthase gene cluster with probes designated for gene encoding chromophore-associated *apo*-protein CagA and two other that were hypothesized to be involved in the production of deoxysugar moiety[50]. The same year, *en route* to search for a gene conferring calicheamicin self-resistance in *Micromonospora echinospora*, Thorson's group discovered and identified the gene cluster responsible for

calicheamicin production by screening a genomic library for the cosmids that confer calicheamicin resistance to *M. echinospora*[51]. Furthermore, Thorson's group also succeeded in locating another DNA fragment that encompasses the gene cluster responsible for dynemicin[52]. Based upon the putative *dynE8* sequence[53; 54], 3 distinct probes targeting the gene encoding E8 (enediyne core), E7 (putative thioesterase) and T3 (unknown conserved protein) were used for screening *Micromonospora chersina* genomic cosmid library. With this shotgun approach, the three gene clusters were completely sequenced. The partial annotation of the C-1027, calicheamicin and dynemicin gene cluster was done for most of the genes based on homology[52; 55; 56]. Each of the gene clusters predicted to encode a variety of enzymes, transporters, transcriptional regulators and proteins with unknown functions.

Using shotgun approach, the calicheamicin locus was found to harbor 74 open reading frames with a size greater than 90-kb. Two separate iterative polyketide synthase genes, Cal05 and CalE8, were discovered among the ORFs residing in the calicheamicin gene cluster (**Figure 1.8**). CalO5 bears high homology to the polyketide synthase involved in the production of orsellinic acid and responsible for the construction of the aryltetrasaccharide moiety. Whereas, CalE8 is highly homologous to the polyketide synthases implicated in the biosynthesis of polyunsaturated fatty acids. Disruption of CalE8 totally abolished calicheamicin production, which implied that CalE8 was indeed required for warhead construction. Boundaries for most of the domains were assigned, including a keto-acyl synthase (KS), an acyl transferase (AT), a ketoreductase (KR) and a dehydratase (DH). Moreover, the sequence between AT and KR was speculated to encompass an acyl carrier protein domain (ACP). In another biochemical study, the C-terminal domain was later found out to be a phosphopantetheinyl transferase (PPTase)[57].

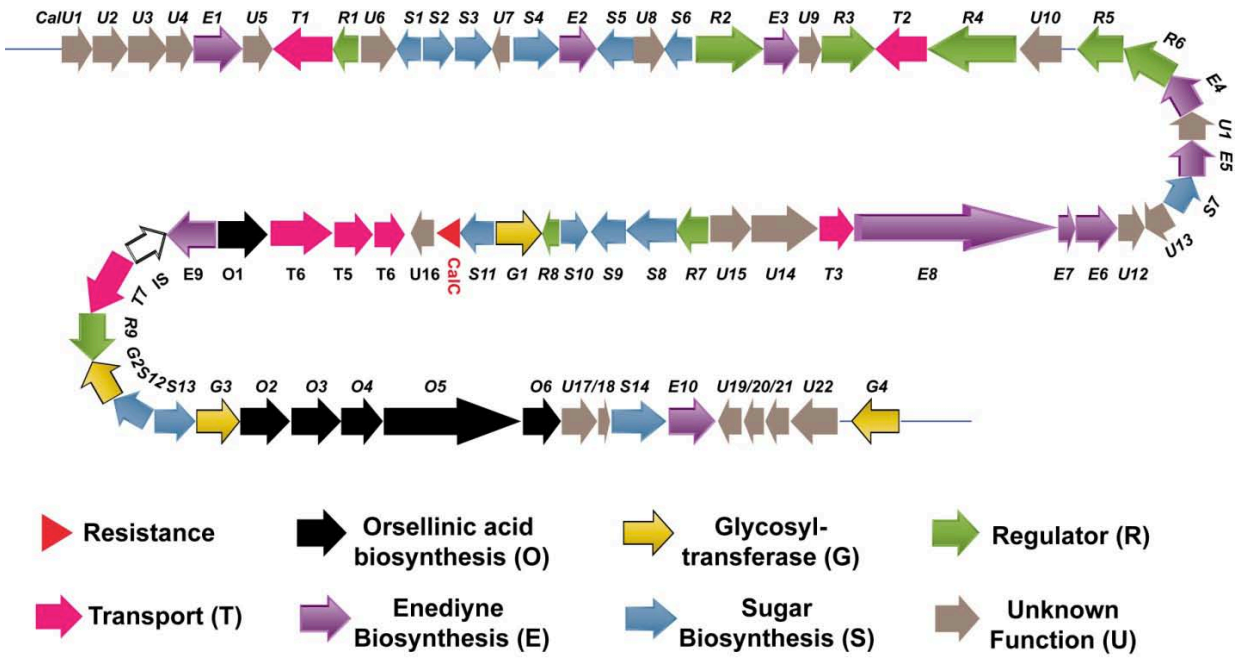



**Figure 1.8** The calicheamicin locus from *Micromonospora echinospora* spp. *Calichensis*.

A total of 60 open reading frames, embedded in the 76 kb of genomic cosmid library containing *dynE8*, were found in dynemicin locus and revealed a variety of genes conserved among known enediyne loci[52] (**Figure 1.9**). The development of a genetic system for the dynemicin producer *M. chersina* confirmed the requirement of the putative enediyne core biosynthetic genes (*dynE8*, *U14* and *U15*) and a tailoring oxidase gene (*orf 23*) for dynemicin production. Surprisingly, this DNA fragment excluded genes encoding for the biosynthesis of the anthraquinone suggesting that the location of genes encoding for the biosynthesis of the dynemicin enediyne core and the dynemicin anthraquinone are chromosomally distinct. Nevertheless, DynE8 demonstrated high sequence homology and identical domain organization in comparison to CalE8.

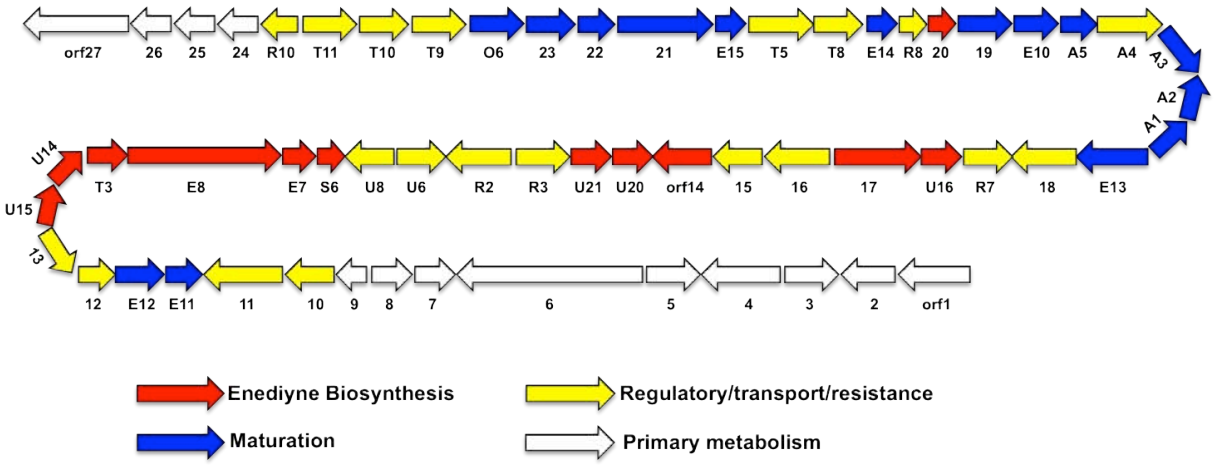



**Figure 1.9** The dynemicin locus from *Micromonospora chersina*.

Another shotgun sequencing experiment went on concurrently in Shen's group revealed the gene locus for C-1027 production. The gene locus, with size of 85 kb DNA, harbored 67 open reading frames. Following this, the boundaries of the cluster were determined by gene disruption techniques, which ruled out several irrelevant ORFs for the C-1027 production. The final layout of the ORFs led to the discovery of genes involved in the biosynthesis of enediyne, deoxy aminosugar, β-amino acid and benzoxazolinate. Among the genes discovered, SgcE was hypothesized to be a polyketide synthase responsible for the production of the enediyne core. A mutation experiment targeting SgcE KS domain totally abolished the biosynthesis of C-1027. Complementation by a plasmid harboring a functional copy of SgcE restored the production to a level comparable to that of the wildtype. Importantly, SgcE exhibits a high level of sequence homology and shares the same domain organization in comparison to CalE8 and DynE8. These results strongly suggest a convergent polyketide pathway for the production of both 9- and 10-membered enediyne cores.

## 1.5 Enediyne PKS gene cluster

### 1.5.1 Minimal enediyne PKS gene cassette

Four more genes in the vicinity of the PKS gene were identified in the calicheamicin, C-1027 and dynemicin gene clusters and found to be conserved between two gene clusters. Since calicheamicin, C-1027 and dynemicin do not share any common peripheral moieties, the five conserved genes are therefore regarded as the essential components for the assembly of the enediyne core. The four conserved genes together with the PKS gene were suggested to constitute the so-called 'minimal enediyne PKS gene cassette' or 'warhead gene cassette'[54; 56] (**Figure 1.10**). One of the four genes was proposed to encode a thioesterase based on sequence homology, but the rest of the genes did not share similarity with any of the genes with known functions in the database. With the high-throughput genomic scanning approach, research on the enediyne biosynthetic pathway gained forward rapidly. Just a couple of years later, the discovery and identification of the gene clusters for neocarzinostatin, maduropeptin and dynemicin further confirmed the conservation of the minimal enediyne gene cassette for all enediyne biosynthesis clusters[52; 61; 62].

In 2003, Thorson's group successfully uncovered several gene loci containing the gene cassette in soil-dwelling *actinomyces* through a high throughput genomic scanning method[54; 63]. The astonishing result indicates that the distribution of enediynes is likely more widespread than previously realized. Moreover, the ubiquitous presence of the minimal enediyne PKS gene cassette in a plethora of organisms was further substantiated when it was discovered in marine bacterium *Salinispora tropica*. The genomic sequencing of *Salinispora tropica* revealed the biosynthetic gene clusters of a large number of secondary metabolites, including two gene clusters that contain the minimal enediyne PKS gene cassettes[43;64; 65]. The first gene cluster was confirmed to play a vital role in the biosynthesis of enediyne precursor of sporolides. Based on the phylogenetic relationship with other PKS genes, the second discovered gene cluster was hypothesized to dictate the production of a 10-membered enediyne[53; 64].

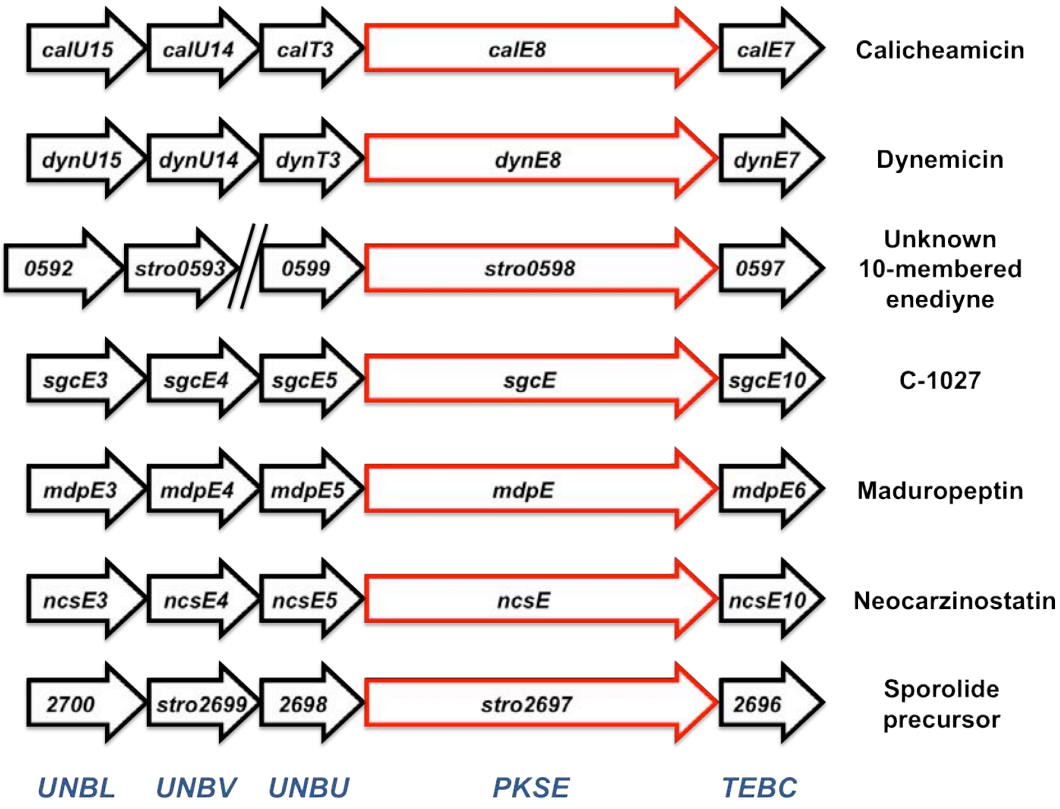



**Figure 1.10** Organization of the highly conserved warhead encoding gene cassette from different enediyne gene clusters. The top 3 clusters from the 10-membered enediyne family and the lower 4 from the 9-membered enediyne family.

The five conserved genes were assigned as *PKSE*, *TEBC*, *UNBL*, *UNBV* and *UNBU* (**Figure 1.10**). Basically, these genes were arranged in a hypothetical operon with unidirectional transcription in the minimal enediyne PKS gene cassette. An overlap in translational start and stop codons are common among the five genes. Thus, it implied that the expression of the five genes is highly coordinated and the protein products should be involved in performing related functions in the bacteria.

As mentioned, PKSE is the major enzyme involved in the production of the polyketide precursor for enediyne core. Located upstream of the PKSE, the TEBC protein is homologous to the 4-hydroxybenzoyl-CoA thioesterase[66] of *Pseudomonas* sp. Strain CBS-3 which belongs to a family of hotdog fold proteins characterized by a long central α-helix packed against a five-stranded antiparallel β-sheet. Such hotdog fold thioesterases are mainly known to utilize acyl-CoAs as

substrates. Hence, proteins from the TEBC family were speculated to catalyze the acyl chain release in polyketide production and also to possess a possible role in chain cyclization. The remaining three families of proteins, namely UNBL, UNBV and UNBU, bear no significant homology to any of the proteins with known functions in the protein database. However, these families of proteins are believed to carry-out specific functions in enediyne biosynthesis. From a structural perspective, proteins under UNBV family are most likely secreted since they all possess N-terminal signal sequences. A secondary structure analysis suggested that the protein is a large β-barrel protein with GAP**X**FG repeats and an ASPIC/UnbV motif. Usually, the ASPIC/UnbV motif can be found in the α-chain of integrin and a dozen proteins from *Rhodopirellula baltica* SH 1. The function of this family is predicted as carriers or surface-adhesive proteins. Meanwhile, proteins from UNBU family are speculated to be helical membrane protein that contains very short cytoplasmic segments membrane protein as it contains 7 – 8 putative membrane spanning α-helix. Although it has been annotated as a transporter[56], the precise role of UNBU family proteins remains to be established. Lastly, the proteins from UNBL family share a minimum sequence homology with a family of diiron cluster containing desaturases or acetylenases for fatty acid modification[67; 68; 69; 70]. Both enzymes catalyze the oxidation of single bonds to produce double bonds or triple bonds in the presence of $O_2$ and electron donors. Therefore, it was hypothesized that UNBL proteins might play a part in the formation of the acetylenic bonds of the 1,5-diyne-3-ene unit. However, these speculations are purely based on bioinformatics studies and require further substantiation from biochemical and structural experiments.

### 1.5.2 Enediyne core

Previously, feeding experiments with $^{13}C$-labeled precursors successfully established that both the 9- and 10-membered enediyne core were derived from eight head-to-tail acetate units[58; 59; 60]. Indeed, it remained controversial whether the enediyne cores are assembled by *de novo* polyketide biosynthesis or degradation from a fatty acid precursor. More recently, the elucidation of the biosynthetic loci for prototypical enediynes from each family led to the remarkable discovery of a

highly conserved enediyne polyketide synthase (PKS) gene involved in the biosynthesis of enediyne warheads via similar polyunsaturated polyketide intermediates and subsequently established PKS paradigm for enediyne biosynthesis[55; 56]. Consequently, it led to the discovery of functional enediyne PKS genes in diverse organisms previously not known as enediyne producers or where the presumed enediyne products were not elucidated structurally[54]. All enediyne core shared common overall enediyne PKS domain organization (KS, AT, ACP, KR, DH, PPTase) and exhibit significant head-to-tail sequence homology (**Figure 1.11**).

However, given the remarkable similarity in both sequence and organization of the enediyne PKS from both loci known to encode structurally characterized 9- or 10-membered enediynes, the biosynthetic divergence of the enediyne core among these architecturally distinct enediyne families remained to be defined. In 2003, Thorson's group invented a universal PCR method for the rapid amplification of minimal enediyne PKS genes, consisting of three enediyne from 9-membered producers (neocarzinostatin, C1027, and maduropeptin) and three from 10-membered producers (calicheamicin, dynemicin, and esperamicin), and successfully determined the presence of both family types as well as a genotypic distinction between two structural families through phylogenetic analysis[53]. This phylogenetic analysis revealed the separation of this essemble into two main branches and the partition coincided with the core enediyne ring size (**Figure 1.12**). The enediyne PKSs appear to here evolved from the same ancestor and underwent divergent evolution giving rise to the 9- and 10-member-specific enediyne enzymes. With this uniquely divergent relationship, the fundamental structure of the unknown enediyne core structure could be predicted directly based on a cloned enediyne PKS gene.

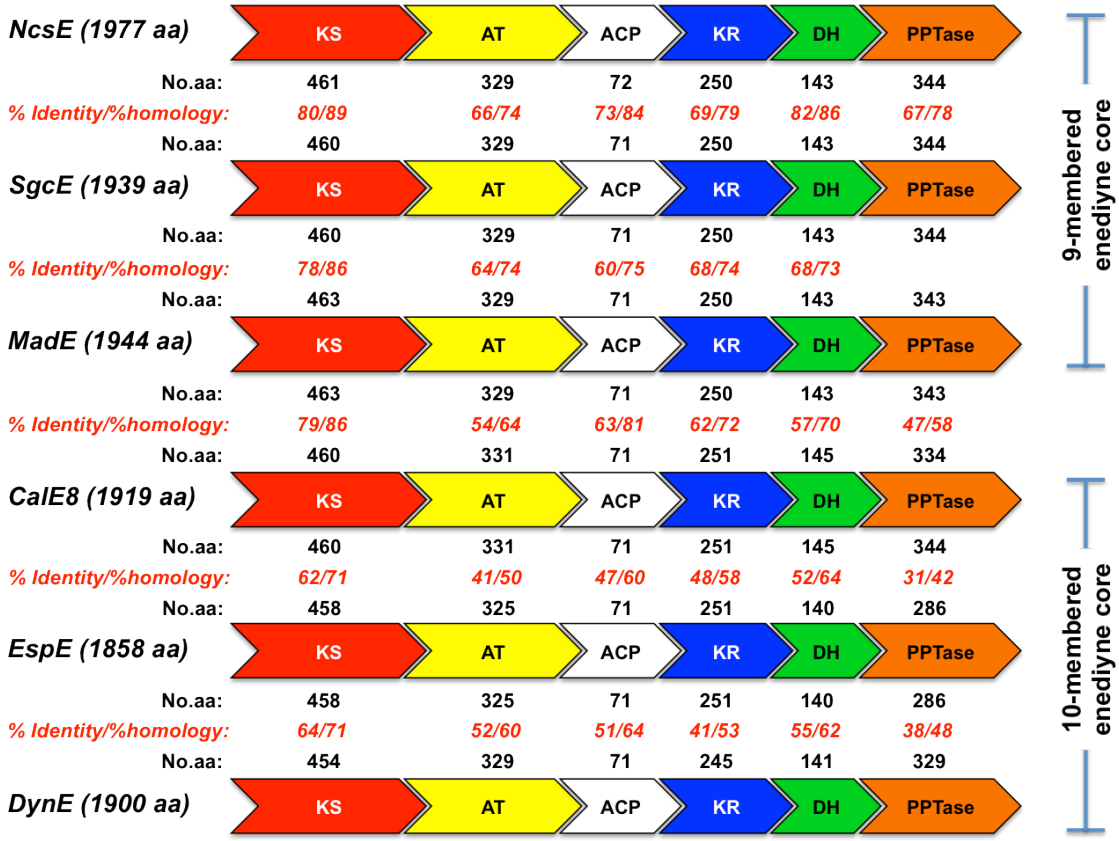



**Figure 1.11** Domain organization and sequence comparison between 9-membered (NcsE, neocarzinostatin; SgcE, C-107; MadE, maduropeptin) and 10-membered (CalE8, calicheamicin; EspE, esperamicin; DynE, dynemicin) enediyne core. Domain designations: KS, ketosynthase; AT, acyltransferase; ACP, acyl-carrier protein; KR, ketoreductase; DH, dehydratase; PPTase, phosphopantetheinyl transferase).

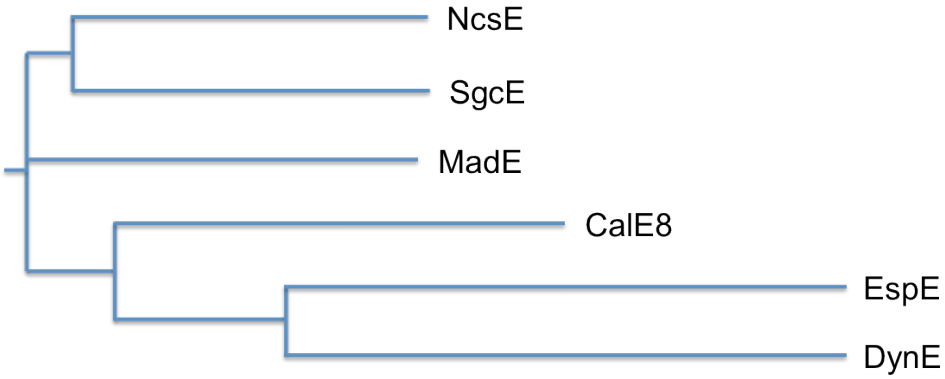



**Figure 1.12** A phylogenetic comparison of minimal enediyne PKSs.

In summary, research on the genes involved in enediyne biosynthesis not only confirmed the polyketide origin of the enediyne core, but also elucidated the biosynthetic pathways and mechanism. Engineering of the biosynthetic pathway for enediyne production also becomes possible, which provides us with opportunities for structural manipulation of enediyne products to cater for different biomedical applications

### 1.6 Iterative type I polyketide synthase for enediyne production

Many known bacterial PKSs belong to type I modular PKS or type II PKS[7; 71]. In fact, the enediyne PKSs are closely related to fatty acid synthases (FASs) that use a single set of catalytic domains for polyketides synthesis[19; 48; 72; 73], and responsible for producing a myriad of secondary metabolites. The enediyne PKS encompassed several domains that were readily identified as ketosynthase (KS), acyltransferase (AT), ketoreductase (KR) and dehydratase (DH) domains based on sequence homology[55; 56]. 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. (This part will be further discussed in Chapter 2)

### 1.7 Objective and organization of the thesis

The discovery of the minimal gene cassettes for enediyne biosynthesis has drawn attention to the biosynthetic pathway of the structurally complex enediyne core as well as the macromolecular protein structure. The first objective of this project is to purify the proteins and decipher the biosynthetic pathway of the enediyne core, as well as to characterize the structure and function of the biosynthetic enzymes. The precise role of the enediyne core enzymes (CalE8, SgcE8 and DynE8) in the production of the polyketide precursor for the enediyne core is still unknown. In addition, little was known about the polyketide precursors of the bicyclic rings and how the precursors are actually assembled from the acetate units. To explore the function of the enediyne core enzyme and its ancillary protein, a series of *in vitro* assays were set up. Three proteins encoded by the minimal cassette, namely PKSE (CalE8, SgcE, and DynE8), UNBL (CalU15, SgcE3 and

DynU15) and TEBC (CalE7, SgcE10, DynE7), were expressed using a heterologous *E. coli* expression system and purified. Their *in vitro* enzymatic activities were examined. Under appropriate experimental conditions, PKSE produces the polyketide precursor or the crude enediyne core itself with the presence of cognate substrates. The identification of the polyketide intermediate could bridge the gap between polyketide production and maturation of the polyketide into the enediyne core. At the same time, the comparative study of the homologous enediyne PKSs (CalE8, SgcE and DynE8) helped us understand the differences among the homologous PKSs as well as their closely-related evolutionary lineage.

In fact, the ultimate aim of this project was to obtain the crystal structures of the enediyne PKS as well as the ancillary proteins. Based on sequence alignments and domain organization, PKSEs appear to be similar to the mammalian fatty acid synthase that structurally consist of an upper modifying and lower condensing chain. A basic requirement for crystallization is high purity of PKSE and ancillary proteins. Protein degradation of PKSE was expected due to the enormous size of the PKSEs protein, approximately 410 kDa for the homodimeric protein. Various precaution steps during protein purification were taken to prevent unnecessary protein degradation. At the same time, the stability of the protein in various conditions, e.g. buffer pH, temperature, NaCl concentration, was examined prior to the crystallization trials. A total of 9600 crystallization conditions were used for screening protein crystallization. In parallel with the PKSE crystallization experiment, structural and functional studies of the individual domains (KS, AT, ACP KR, DH and PPTase) as isolated domains were successfully examined. By studying the structure and function of the individual domains, we were able to piece together the information to obtain a detailed mechanistic view of the iterative type I PKS.

In addition, crystallization of 2 family proteins encoded in the minimal gene cassette, namely UNBL and TEBC, was also my interest. The UNBL family proteins share little sequence homology with diiron desaturases or acetylenases with

the function of catalyzing the oxidation of single bonds to produce double or triple bonds in the presence of $O_2$ and electron donors. Thus, it was hypothesized that UNBL protein could be involved in the formation of the acetylenic bonds of the 1,5-diyne-3-ene unit. Meanwhile, the study on the TEBC, putative PKSE thioesterase (CalE7, SgcE10, DynE7), shed light on the structural and functional aspects of the novel enzyme as well. It is not known whether CalE7 functions as a type I or type II thioesterase. Due to the low homology of PKSE thioesterase to any known thioesterase in the database, it was almost impossible to deduce the function of PKSE thioesterase solely based on bioinformatics studies. Overall, structural studies provided us with a new perspective on the function of both UNBL and TEBC protein. With the structure of these proteins, it becomes possible to probe their catalytic mechanism.

# CHAPTER 2 Expression, purification, crystallization and functional studies of enediyne PKS and its ancillary proteins from the minimal gene cassette

## 2.1 Introduction

With the report published by Thorson and Shen in 2002, a milestone was reached with the successful sequencing of the calicheamicin and C-1027 gene clusters[55; 56]. Furthermore, the discovery of dynemicin[52] in 2008 further expanded the realm of iteractive type I PKS by manifesting the capability of the PKSEs in producing enediyne cores. The subsequent devise of a high throughput screening method unveiled a plethora of gene clusters containing similar enediyne PKSEs[53; 54]. Analysis of the enediyne PKS demonstrated the presence of six functional domains within the megasynthases (**Figure 2.1**)

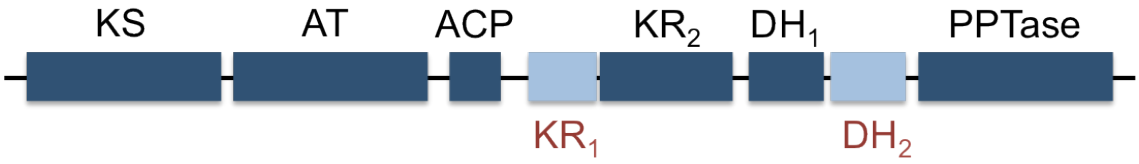



**Figure 2.1** Domain composition of enediyne PKS. From the N-terminal to C-terminal end (*left* to *right*): KS (ketoacyl synthase), AT (acyltransferase), ACP (acyl carrier protein), KR (ketoreductase, DH (dehydratase) and PPTase (phosphopantetheinyl transferase).

Many known bacterial PKSs belong to type I modular PKSs or type II PKSs[7; 71]. In contrast, the enediyne PKSs from the biosynthetic pathways of calicheamicin, C-1027 and dynemicin seem to resemble a family of fungal iterative type I PKSs that use a single set of catalytic domains for polyketide synthesis[72; 73; 74; 75]. Giving the recent surprising finding, several domains protein fold bears great similarity to the structural subdomain of modular PKS DEBS and the mammalian fatty acid synthase[19; 76; 77]. The enediyne PKSs encompass several domains that were readily identified as ketosynthase (KS), acyltransferase (AT), ketoreductase (KR) and dehydratase (DH) domains based on sequence homology (**Figure 2.1**)[55; 56]. The identities of the ACP (acyl carrier protein) and PPTase (phosphopantetheinyl transferase) domains were confirmed recently by Shen and

coworkers, showing that the ACP domain of SgcE is exogenous phosphopantetheinylated PPTase[78]. Meanwhile, Liang and coworkers demonstrated that the putative ACP domain of CalE8 is phosphopantetheinylated by a widely used PPTase with broad specificity, Sfp[57; 79], and the C-terminal region of CalE8 has a similar function capable of modifying heterologous ACPs[57].

The enediyne PKSs also contain two stretches of sequence located downstream of the ACP and DH (labeled as $DH_1$ in **Figure 2.1**) domains. Up to date, no homolog can be found for the ~100 and ~150 amino acid fragments in the protein data bank (PDB). Based on *de novo* structural modeling study[80], Liang speculated that ~100 amino acid segment before the KR domain (labeled as $KR_2$ in **Figure 2.1**) adopts an α/β fold with six β strands flanked by α-helices and bears great similarity to the structural subdomain of the KR domains of the modular PKS DEBS and the mammalian fatty acid synthase[19; 76]. Thus, it is likely that the unknown domain $KR_1$ is in fact part of the intact KR domain (~410 aa). Liang *et. al.* also speculated that the sequence between the $DH_1$ and PPTase domains might be the second half of the intact DH domain that adopts a hotdog fold[80]. This is supported by the recent finding that the DH domains of the mammalian FAS and modular PKS are actually comprised of two hotdog fold subdomains[19; 77]. The domain organization of the enediyne PKSs, with the two unknown regions tentatively assigned as the $KR_1$ and $DH_2$ subdomains, is depicted in **Figure 2.1**.

In the iterative enediyne polyketide synthesis, an intrinsic phosphopantetheinyl transferase (PPTase) domain, instead of a stand-alone PPTase protein in other bacteria PKS system, is found to modify the acyl carrier protein (ACP) domain by phosphopantetheinylation[54; 57; 78] (**Figure 2.2**). Spontaneously after modification, the phosphopantetheine moiety covalently linked to the ACP functions as a flexible arm that carries the extending chain and swings around to access the various enzymatic domains for chain processing. The acyltransferase domain (AT) load the ACP domain with acyl units from malonyl-CoA.

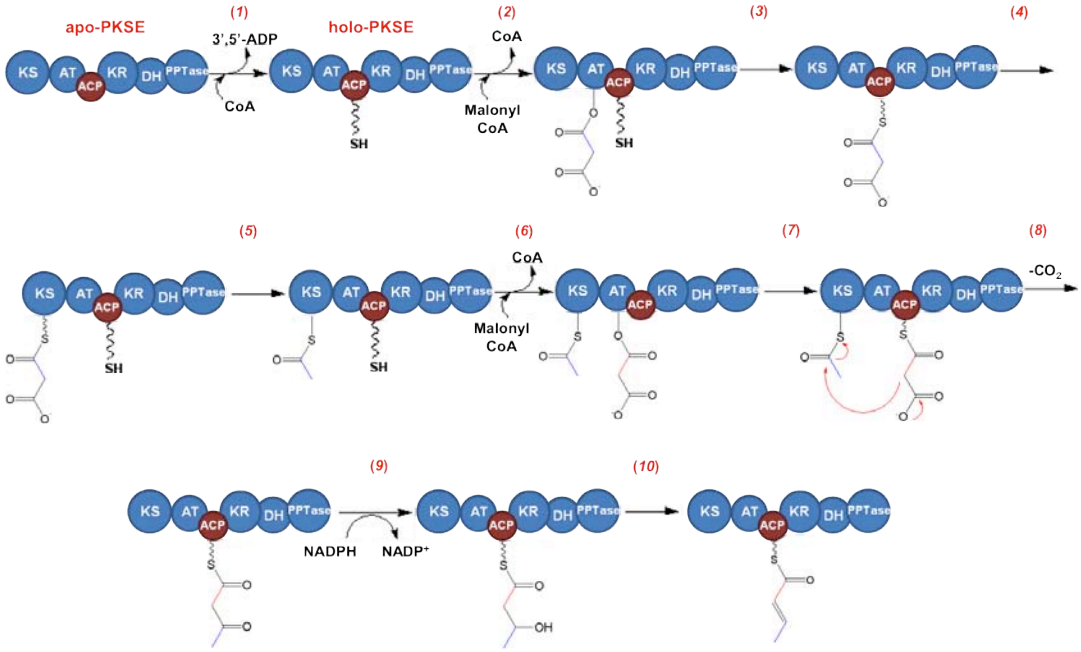



**Figure 2.2** Schematic diagram of individual steps of iterative enediyne polyketide synthesis. (1) The PPTase domain modifies the ACP domain by phosphopantetheinylation using CoA. (2 – 3) Malonyl-CoA is loaded onto the ACP domain by the AT domain via thioester linkage. (4 – 5) Acyl starter is transferred to the KS domain and undergoes an intrinsic decarboxylation to generate the acetate unit. (6) At the same time, the AT domain loads the ACP domain with malonyl-CoA as extender. (7 – 8) The KS domain mediates the condensation between the starter and extender molecules. As elongation goes on, (9) the KR domain reduces subsequent keto groups to corresponding hydroxyls group, which are then removed in (10) desaturation mediated by the DH domain to form conjugated double-bond system. The elongation process is stopped when a designated chain length is reached and the final product is cleavable by external thioesterase.

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

Once 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. The antitumor property of enediyne is largely due to the presence of the two chemically reactive triple bonds, therefore, the formation of the *–yne* group is of particular interest. Among the five genes in the minimal enediyne PKS gene cassette, UNBL is speculated to encode a non-heme di-iron protein[70; 81] that is potentially capable of catalyzing triple bond formation. At the same times, the control over the chain length of the polyketide intermediate and the release of product from PKS normally is catalyzed from either a type I or type II thioesterase. However, most of the type I thioesterases are normally *cis*-acting domains fused to the C-terminus of the most downstream module of PKS, whereas type II thioesterases are discrete proteins responsible for the *trans* hydrolytic release of the aberrant products[82; 83; 84]. Even though another gene in the minimal enediyne PKS gene cassette, TEBC, could be involved in product release from PKS, but the translated protein shares very low sequence homology to any known type I or type II thioesterases. Instead, the putative thioesterase encoded by TEBC shares moderate sequence homology with a family of hotdog fold proteins characterized by a long central α-helix packed against five-stranded anti-parallel β-sheet. The hotdog fold proteins include many dehydratases and thioesterases that use acyl-CoA as substrates[85].

Apart from *–yne* formation and cleavage of matured acyl chain, cyclization is also one of the crucial steps in the formation of enediyne core. To date there is no candidate for the cyclase protein, since the products encoded by the other two genes in the minimal cluster bear extremely low homology to any protein with known function in the database. With the completion of the whole series of complex modification process, the enediyne core is presumably ready to be assembled with orsellinic acid and sugar moieties by specific glycosyltransferases.

The main objective of this chapter is to crystallize ~205 kDa homodimeric PKSE proteins as well as the ancillary proteins (TEBC and UNBL). Due to the large size of homodimeric enediyne PKS and multiple flexible linker regions, crystallizing the protein proves to be a difficult task. Despite that, I have decided to take the risk as there were several successful crystallization models, such as mammalian fatty acid synthase and fungal fatty acid synthase. In this chapter, I will discuss the expression, purification and crystallization of PKSEs and the problems encountered as well as possible solutions. Also, the functional studies of PKSEs (CalE8, SgcE, and DynE8), TEBCs (CalE7, SgcE10, and DynE7) and UNBL (CalU15, SgcE3, and DynU15) will be investigated by using *in vitro* methods, as well as the production of premature or mature enediyne core.

## 2.2 Material and methods

### 2.2.1 Reagents and chemicals

CoA, acetyl-CoA, malonyl-CoA, NADPH and other chemicals were purchased from Sigma-Aldrich and stored at -20 °C. Expression vectors pET-28b(+) and pCDF-2 Ek/LIC were bought from Novagen, from where BL21(DE3) competent cell was also obtained. The genes encoding CalE8 (putative 10-membered enediyne PKSE), CalE7 (putative thioesterase), CalU15 (putative acetylenase), SgcE (putative 9-membered enediyne PKSE), SgcE10 (putative thioesterase), SgcE3 (putative acetylenase), DynE8 (putative 10-membered enediyne PKSE), DynE7 (putative thioesterase) and DynU15 (putative acetylenase) were purchased from GenScript Corporation with sequences optimized for protein expression in *E. coli*. The genes were synthesized based on the reported protein sequences and were inserted in pUC57-based plasmids.

### 2.2.2 Cloning, expression, purification and crystallization of CalE8, SgcE and DynE8

All highly homologous PKSE genes, *CalE8*, *SgcE* and *DynE8*, were sub-cloned into pET-28b(+) between *Nde*I and *Xho*I restriction sites (**Appendix Table 1**). Each recombinant plasmid, harboring the PKSE gene, was transformed into *E. coli* BL21 (DE3) expression host. The seed culture was prepared by inoculating the

freshly transformed colony into 50 mL LB medium supplemented with 50 μg/mL kanamycin and grown overnight at 37 °C. One percent of overnight seed culture was inoculated into 1 L of LB medium supplemented with 50 μg/mL kanamycin. After inoculation into LB medium, the culture was grown up to $OD_{600}$ = 0.6 before induction with 0.2 mM IPTG. The culture was shaken at 16 °C for 20 hours before harvesting the cells by centrifugation.

The cell pellet was suspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 25 mM imidazole, 20 % glycerol, and 10 mM β-mercaptoethanol) supplemented with 25 μL of Protease Inhibitor Cocktail Set III (Novagen) and lysed by sonication. After centrifugation at 53,000 x *g* for 30 minutes at 4 °C, the supernatant was filtered with 0.45 μm Minisart filters (Sartorius Stedim). The filtered supernatant was loaded onto 5 mL HisTrap HP column. A stepped gradient (25 mM and 50 mM) was employed for vigorous washing before elution with 400 mM imidazole containing elution buffer. The yellow protein eluent (due to the presence of polyene at 410 nm) that contains the N-terminal $His_6$-tagged PKSE was diluted 25 - 30 fold with IEX A buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 20 % glycerol, 2 mM DTT) and loaded onto 5 mL HiTrap Q anion exchanger. A linear gradient, from 100 mM NaCl to 1 M NaCl, was applied to elute the bound protein within 60 minutes. The yellow protein eluent was concentrated to 2 mL before loading onto Superdex S200 size exclusion chromatography column. Only the purest fractions of protein were collected and concentrated to ~15 mg/mL by using Vivaspin 20 – 100 kDa cut-off (Sartorius Stedim). The final protein concentration was determined by the Bradford method before the crystallization screening.

Prior to crystallization, freshly prepared protein solution (~15 mg/mL) was filtered using 0.1 μm Ultrafree centrifugal filter. An automated initial crystallization screen of 960 conditions was performed using the Phoenix (Art Robbins Instrument). Three different protein:crystallization buffer ratio (1:2, 1:1, and 2:1) were employed for the crystallization screening using the sitting drop, vapor-diffusion method. The remaining protein stock was flash frozen using liquid nitrogen for later use.

### 2.2.3 Co-expression and co-purification of PKSE, TEBC and UNBL proteins

The PCR primers for putative TEBC (CalE7, Sgc10 and DynE7) and putative UNBL (CalU15, SgcE3 and DynU15) were designed according to the instruction manuals for pCDF-2 Ek/LIC and LIC Duet Adaptor kit (Novagen). The amplified PCR products were treated with T4 DNA polymerase and ligated into pCDF-2 Ek/LIC vector (Novagen) to yield the recombinant plasmid DNA pCDF-2-*TEBC* (*CalE7* or *SgcE10* or *DynE7*) and pCDF-2-*UNBL* (*CalU15*, *SgcE3* or *DynU15*) (**Appendix Table 1**). With the use of LIC Duet Adaptor Kit (Novagen), both TEBC and UNBL were ligated into pCDF-2 EK/LIC vector to yield the recombinant plasmid DNA pCDF-2-*TEBC*-*UNBL* where only TEBC contains a $His_6$-tag (**Appendix Table 1**). The recombinant plasmid DNAs were co-transformed with pET-28b(+)-PKSE (CalE8, SgcE and DynE8) into *E.coli* BL21 (DE3) to yield 2 different expression systems: 1) pET-28b(+)-*PKSE*/pCDF-2-*TEBC*. 2) pET-28b(+)-*PKSE*/pCDF-2-*TEBC*-*UNBL*. The cells harboring the recombinant plasmids were selected against kanamycin and streptomycin. Successful co-transformation was confirmed by colony PCR and small-scale protein expression screening.

For co-expression, 1 liter of LB supplemented with 50 μg/mL kanamycin and 50 μg/mL streptomycin was inoculated with 5 mL of grown seed culture. The culture was grown at 37 °C at 200 rpm to $O.D_{600}$ = 0.6 and induced with 0.4 mM IPTG after the culture had cooled down to 16 °C for approximately 30 minutes. The culture was incubated at 16 °C for 20 hours at 200 rpm before harvesting the cells by centrifugation. The cells were lysed for protein purification using the procedures described above.

A three-step protein purification strategy, involved immobilized metal affinity chromatography (IMAC), ion exchange chromatography (IEX) and size exclusion chromatography, was applied to co-purify both PKSE and TEBC proteins. Standard purification buffer conditions, similar to the buffer condition mentioned above, were used for PKSE-TEBC co-purification. In the co-expression system, both PKSE and TEBC contain N-terminal $His_6$-tags, therefore, the

purification by $Ni^{2+}$-NTA yielded a mixture of yellow protein solution containing both PKSE and TEBC. The yellow color observed is due to the presence of polyene initially biosynthesized by PKSE and later transferred and bound to TEBC during co-expression. Through the use of ion exchange chromatography and size exclusion chromatography, the colorless PKSE and yellow-colored TEBC were separated and purified. Both PKSE and TEBC protein were concentrated to the ~15 mg/mL concentration using Vivaspin 20 and Vivaspin 15R, respectively.

The freshly prepared colorless PKSE and yellow TEBC protein were filtered with 0.1 μm Ultrafree centrifugal filter before crystallization screening in 96-well CrystalQuick plate using sitting-drop vapour diffusion method with the 10 commercially available screening kits from Hampton. Three different protein:crystallization buffer ratio (1:2, 1:1, and 2:1) were employed for crystallization screening using the sitting drop, vapor-diffusion method.

### 2.2.4 *In vitro* activity assay on CalE8 KR domain

Enzymatic reactions, involving only CalE8 KR domain, were conducted by monitoring the disappearance of the NADPH peak at 340 nm, using a spectrophotometer. The assays were carried out with 2.0 μM CalE8, 0.5 mM decalone and 0.25 mM NADPH in a total volume of 150 μL reaction buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl). A control reaction, containing same composition of reagents without the addition of decalone, was set up. Both reaction and control were subject to spectrophotometry scanning before and after 1.5 hours of incubation in 30 °C water bath.

### 2.2.5 *In vitro* activity assay by absorption spectroscopy

Enzymatic reactions involving PKSE (CalE8, SgcE and DynE8) and TEBC (CalE7, SgcE10 and DynE7) were conducted in a semi-micro quartz cuvette and monitored with a Shimazu UV-Vis 1700 spectrophotometer. The sample chamber was kept at 30 °C through an external temperature controller. Steady-state kinetic measurement was carried out for each PKSE/TEBC pair, with 1.0 μM of colorless PKSE, 10 μM of colorless TEBC, 0.1 mM malonyl-CoA, 0.5 mM NADPH in 200

μL of buffer (100 mM Tris pH 8.2, 300 mM NaCl and 1 mM DTT) at 30 °C. The reaction was incubated in the sample chamber for 15 minutes prior to the addition of 0.5 mM malonyl-CoA. The absorption spectrum of the reaction mixture was taken at various time intervals to monitor the progress of product formation. The reactions were also conducted with the addition of the putative acetylenase under similar conditions. Small molecule electron donors, such as ascorbate, sodium dithionite ($Na_2H_2SO_4$), as well as the ferredoxin/ferredoxin reductase system were tested in order to reduce the di-iron cofactor in the putative acetylenase. Under all the conditions tested, no oxidation of the products was observed.

### 2.2.6 *In vitro* activity assay and product analysis by HPLC

In order to test the activity of PKSE and TEBC proteins, a series of *in vitro* assay were conducted. A reaction mixture containing of 1.0 μM colorless PKSE and 5.0 μM colorless TEBC, 200 μM acetyl-CoA and 600 μM NADPH in buffer (100 mM Tris pH 8.2, 300 mM NaCl and 1 mM DTT) were incubated at 30 °C for 15 minutes prior to the addition of 1.2 mM of malonyl-CoA. At the same time, a similar enzymatic reaction mixture was set up without adding acetyl-CoA. The 200 μl of enzymatic reaction was carried out in a 30 °C water bath for 3 hours, together with two controls that lack either malonyl-CoA or PKSE. The reaction was then quenched with the addition of 5% TFA (later was replaced by HCl to bring pH to ~2). To extract the yellow pigment from TEBC, an equal volume of ethyl acetate was added into the reaction mixture and vigorous vortexing ensued. Subsequently the reaction mixture was centrifuged at 22,000 x *g* for 15 minutes. The organic solvent layer was then pipetted out and the reaction mixture was subject to extraction for the second time. The ethyl acetate extracts were pooled and the solvent was evaporated under a gentle flow of nitrogen gas. The yellow pigment was re-dissolved in a small volume of methanol for HPLC analysis using an eclipse XDB RP C18 column (4.6 x 150mm). The gradient employed was from 100% buffer A (HPLC grade water with 0.045% TFA) to 100% buffer B (90% acetonitrile, 10% HPLC grade water with 0.045% TFA) in 60 minutes.

### 2.2.7 LC-MS

The column used for LC-MS analysis was eclipse XDB C18 column with a dimension of 2.3 X 150mm. The gradient employed in the analysis was from 70% buffer A (HPLC grade water with 0.045% TFA) + 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B in 120 minutes. The ionization energy was set at 5.0 kV with an ESI ionization source for the Finnigan LTQ Orbitrap mass spectrometer (Thermo Electron). The result was analyzed with the software Xcalibur for the determination of plausible molecular compositions based on the observed molecular weight and fragmentation pattern. This experiment was carried out together with Dr. Kong Rong.

### 2.2.8 Concentration, pH and temperature dependence of enzymatic reactions involving PKSE

To obtain the optimum production of the enediyne precursor biosynthesis, a series of optimization involving protein concentration, pH and temperature were employed. A typical reaction mixture includes 1.0 μM CalE8 (or SgcE), 5.0 μM CalE7 (or SgcE10), 0.25 μM NADPH and 91.5 μL of buffer (100 mM Tris (pH 7.5 and 8.5), 300 mM NaCl and 1 mM DTT). After gentle mixing, the reaction was incubated in the sample chamber of UV-Vis spectrophotometer for 15 minutes prior to the addition of 2.5 μL malonyl-CoA (10 mM). The malonyl-CoA/NADPH concentration dependent reactions were done at constant NADPH/malonyl-CoA concentration, with varying malonyl-CoA/NADPH concentration. The reactions were carried out at pH 7.0 (or pH 8.5), 30 °C (25 °C or 37 °C) for 3 hours. At the end of 3 hours, the reaction was then quenched with HCl to bring the pH to ~2. An equal volume of ethyl acetate was added into the reaction mixture, vigorous vortexing, and subsequently centrifuged at 22,000 x *g* for 15 minutes. The organic solvent layer was pipetted out and the reaction mixture was subjected to extraction for the second time. The ethyl acetate extracts were pooled and the solvent was evaporated using a Speed-Vac. The yellow pigment was dissolved in methanol for HPLC analysis using an eclipse XDB RP C18 column (4.6 x 250mm). The gradient employed was from 70% buffer A (HPLC grade water with 0.045% TFA) + 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B in 60 minutes.

## 2.3 Results

### 2.3.1 Purification of PKSE proteins: CalE8, SgcE and DynE8

The protein expression level of CalE8, SgcE and DynE8 were analyzed by SDS-PAGE (**Appendix Figure 2.1**). Purification of PKSE proteins involved several purification methods. For all three o ver-expressed PKSE proteins, CalE8, SgcE and DynE8, immobilized metal ion affinity chromatography (IMAC HisTrap column) and size exclusion chromatography (Superdex 200 column) were initially employed for the purification process. Due to severe protein degradation, protein aggregation and impurities, anion exchange chromatography (HiTrap Q column) was used as an additional purification step before size exclusion chromatography. At the same time, the purification buffer conditions were redesigned to stabilize the protein as well as to enhance the purity of proteins (**Table 2.1**).

**Table 2.1**: Optimized buffer condition for the purification of PKSE proteins.

| Buffer | Initial buffer Condition | Optimized buffer condition |
|---|---|---|
| IMAC lysis buffer | 50 mM Tris pH 7.0, 0.3 M NaCl, 10 mM imidazole, 10 mM β-ME, 10 % glycerol | 50 mM HEPES pH 7.5, 0.5 M NaCl, 25 mM imidazole, 10 mM β-ME, 20 % glycerol |
| IMAC buffer A | 50 mM Tris pH 7.0, 0.3 M NaCl, 10 mM β-ME, 10 % glycerol | 50 mM HEPES pH 7.5, 0.5 M NaCl, 10 mM β-ME, 20 % glycerol |
| IMAC buffer B | 50 mM Tris pH 7.0, 0.3 M NaCl, 0.5 M imidazole, 10 mM β-ME, 10 % glycerol | 50 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 M imidazole, 10 mM β-ME, 20 % glycerol |
| IEX buffer A | 50 mM Tris pH 7.0, 0.1 M NaCl, 1 mM DTT, 10 % glycerol | 50 mM HEPES pH 7.5, 0.1 M NaCl, 2 mM DTT, 20 % glycerol |
| IEX buffer B | 50 mM Tris pH 7.0, 1.0 M NaCl, 1 mM DTT, 10 % glycerol | 50 mM HEPES pH 7.5, 1.0 M NaCl, 2 mM DTT, 20 % glycerol |
| Size exclusion buffer | 50 mM Tris pH 7.0, 0.3 M NaCl, 1 mM DTT, 10 % glycerol | 25 mM HEPES pH 7.5, 0.3 M NaCl, 2 mM DTT, 15 % glycerol |

In order to enhance the protein stability, high ionic strength purification buffer with 0.5 M NaCl was used in the IMAC purification step. Besides, increasing the concentration of glycerol to the range of 15 % - 20 % glycerol would help to reduce the chances of protein degradation and fragmentation. However, based on the basic requirement for protein crystallization, low salt concentration and glycerol concentration is more preferable. Therefore, both salt concentration and glycerol concentration were gradually reduced to 0.3 M and 15 %, respectively, in the last purification step.

IMAC was employed as a first purification step to isolate and concentrate the $His_6$-tagged target protein from other proteins and contaminants. A concentration of 25 mM imidazole in the IMAC lysis buffer was used to reduce non-specific binding of *E. coli* cellular protein to the $Ni^+$-NTA resin. The bound proteins was sequentially washed with 25 mM and 50 mM imidazole containing buffer and eluted with 400 mM imidazole. Although the target protein has been concentrated and enriched in the eluent, the desired purity level has yet to be achieved as many impurity and protein fragment resulting from protein degradation were seen from the SDS-PAGE analysis (**Figure 2.3**)

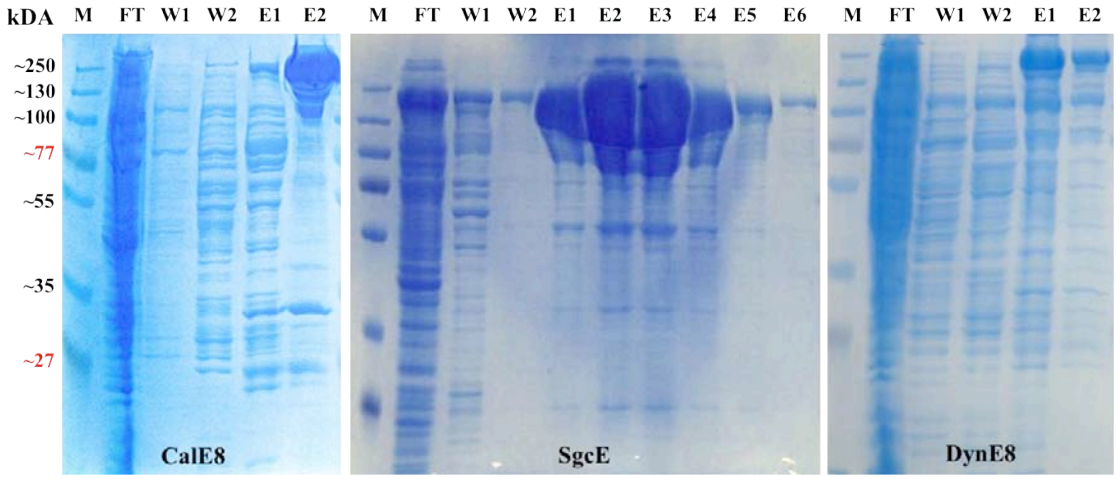



**Figure 2.3** SDS-PAGE analysis of cell-lysate and IMAC result. **M**: PageRuler Plus Prestained Protein Ladder (Fermentas). **FT**: flow-through, **W**: wash step with 25 mM (**W1**) and 50 mM (**W2**) imidazole. **E**: eluent in 1 mL fractions.

Since the IMAC purification step did not produce sufficiently pure protein, ion exchange chromatography step was introduced to enhance the protein purity. The theoretical pI of CalE8, SgcE8 and DynE8 are 5.44, 5.06 and 5.83, respectively, and therefore, an anion-exchanger (HiTrap Q Sepharose column) was used. HiTrap Q column allows the target protein to retain negative net charge at pH 7 and bind to the column. Other impurities and contaminant proteins that possess a positive net charge will flow through the column without any significant binding. The bound proteins were eluted with a linear gradient of increasing NaCl concentration, from 0.1 M to 1.0 M. Compared to the IMAC eluent, the IEX eluent showed improved purity with most of the contaminants removed (**Figure 2.4**). However, several additional protein bands were visible on SDS-PAGE analysis, suggesting both the presence of impurities and/or protein degradation. This could be due to drastic salt dilution of the protein sample using low salt IEX buffer A for protein binding in HiTrap Q column.

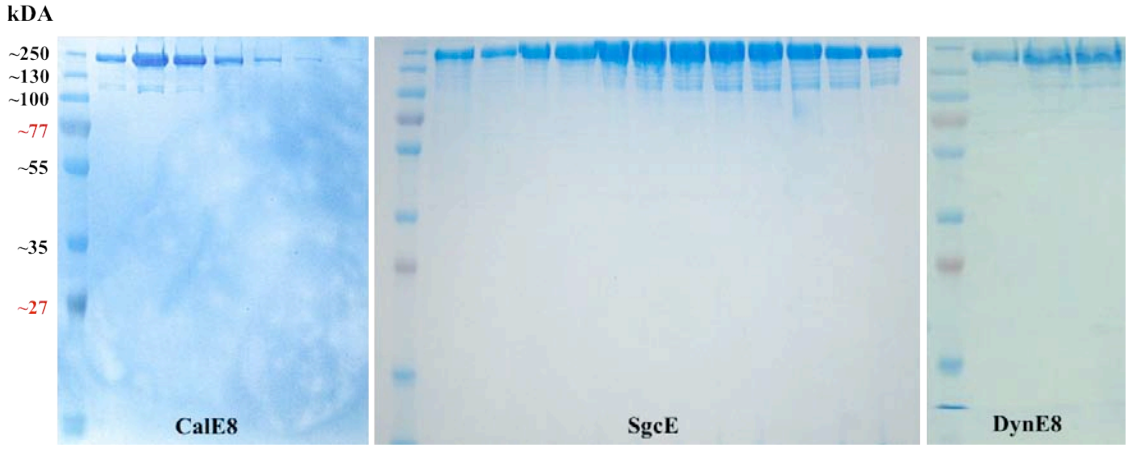



**Figure 2.4** SDS-PAGE analysis of IEX results for CalE8, SgcE and DynE8. Fragmented protein bands become more prominent in the later fractions as well as the contaminant bands.

In order to separate the native dimers from the non-native oligomeric aggregates as well as fragmented proteins and other impurities, gel filtration was introduced as a final purification step. Unlike ion exchange or affinity chromatography, gel filtration separates molecules according to differences in size as they pass through the medium packed in the column. This step is crucial as for crystallization mono-dispersed homogenous protein samples are strictly required.

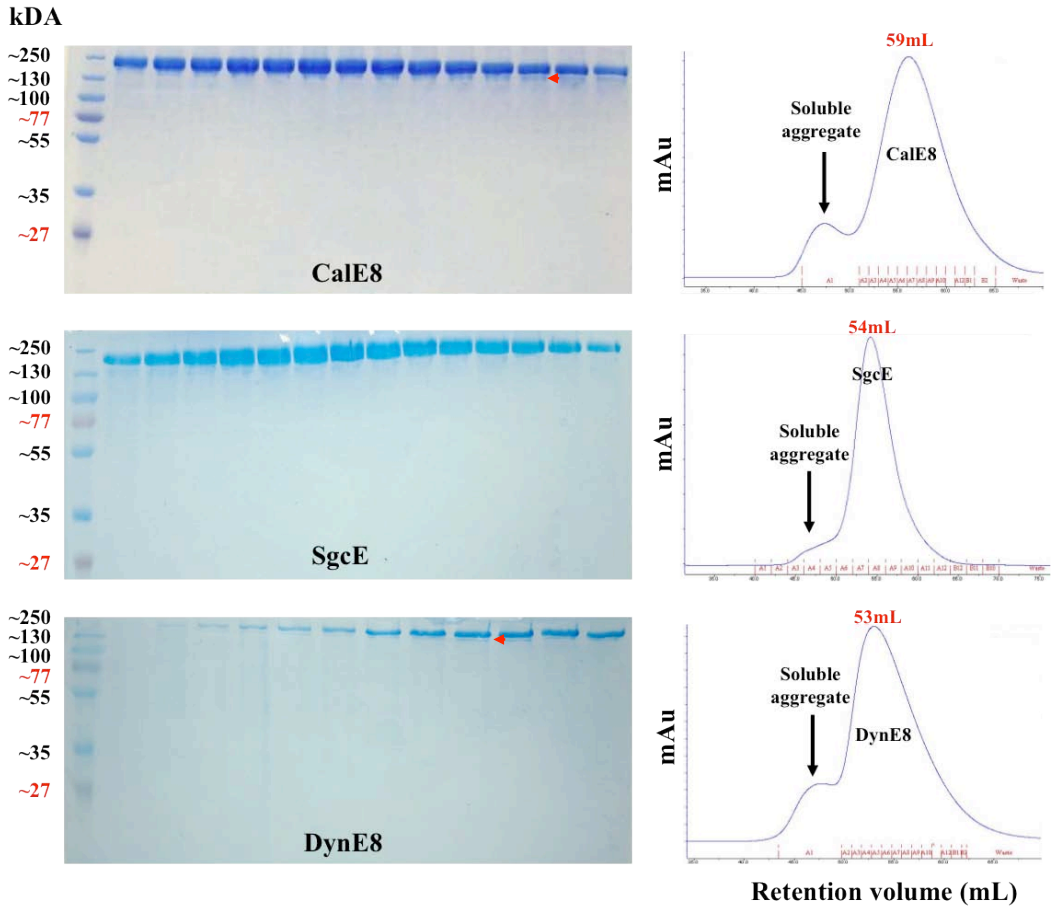



**Figure 2.5** A chromatogram of size exclusion chromatography and SDS-PAGE analysis. The *red* arrow indicated the fragmented PKSE proteins.

With a constant flow rate of 0.5 mL/min on Superdex 200, the homodimeric PKSE was separated from the soluble aggregate. Due to the limitation in size resolution of the Superdex 200, both soluble aggregate peak and PKSE peak are slightly overlapping (**Figure 2.5**). Therefore, the fine collection of 0.5 mL protein solution per fraction was performed. Only those fractions at the centre of the peak were collected and concentrated to final concentration of ~15 mg/mL. Obviously, the quality of the protein has improved. Most of the impurities are successfully removed from IEX-purified proteins. However, negligible amount of fragmented proteins, corresponding to the molecular weight of ~103 kDa, with consistent sizes were co-purified. The protein bands that correspond to the molecular weight of ~205 kDa and ~103 kDa, were verified by mass spectroscopy (Core Facility of the Biological Research Center, NTU). The MS results revealed that the protein (~205 kDa) mass obtained was similar to the molecular weight of PKSEs (**Appendix**

**Figure 2.2**). The protein sequences obtained from the MS of tryptic digested protein fragments was identified as PKSE protein, whilst the shorter fragment was identified as fragments of PKS protein resulting from degradation. The purified PKSE was found to be active and this would be further elaborated in **section 2.2.3** and **2.3.4**

Crystallization screening of PKSE proteins, CalE8, SgcE and DynE8, failed to produce crystals. In the single expression system, the enediyne polyketide intermediates are bound to the conserved serine residue, S966, in the ACP domain attached to a phosphopantetheine moiety[16]. Thus, we speculated that the product-bound PKSE might possess a number of conformational subpopulations and result in low homogeneity and difficulty to crystallize.

### 2.3.2 Co-purification of PKSEs and TEBCs

To improve the protein conformational homogeneity, the polyketide product attached to the ACP domain needs to be cleaved and off loaded. Therefore, co-expression of each PKSE (CalE8, SgcE8 and DynE8) with their cognate thioesterase (CalE7, SgcE10 and DynE7, respectively) was used to remove the polyketide product from PKSE.

A mixture of PKSE (CalE8 and SgcE) and thioesterase protein (CalE7 and SgcE10) was eluted at the same concentration of imidazole from IMAC (as described earlier) as both proteins carried a $His_6$-tag (**Figure 2.6 and 2.7**). In order to separate PKSE and thioesterase, an anion exchange HiTrap Q column (GE), was used for further purification. Two distinct peaks of PKSE (CalE8 and SgcE) and thioesterase (CalE7 and SgcE10) are observed on the chromatogram and eluted with ~ 350 mM and ~ 500 mM NaCl, respectively. PKSE (CalE8 and SgcE8) fractions and thioesterase (CalE7 and SgcE10) fractions were separately collected and subjected to a Superdex S200 column. The fractions of both PKSE and TEBC were concentrated to ~ 15 mg/mL and stored or used for biochemical analysis and subsequently crystallization trials.

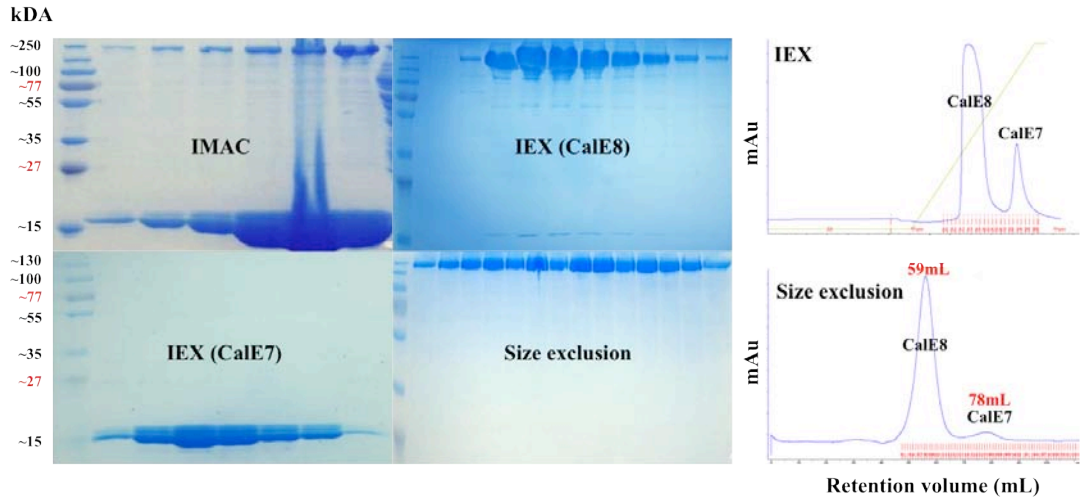



**Figure 2.6** Three steps sequential co-purification of CalE8 and CalE7. IMAC was first used to isolate the $His_6$-tagged protein from other impurities, followed by ion exchange chromatography and size exclusion chromatography. SDS-PAGE analysis of IMAC, ion exchange chromatography (IEX) and size exclusion chromatography results (*left panel*). (*Right panel*) Chromatogram of IEX (*top*) and size exclusion chromatography (*bottom*).

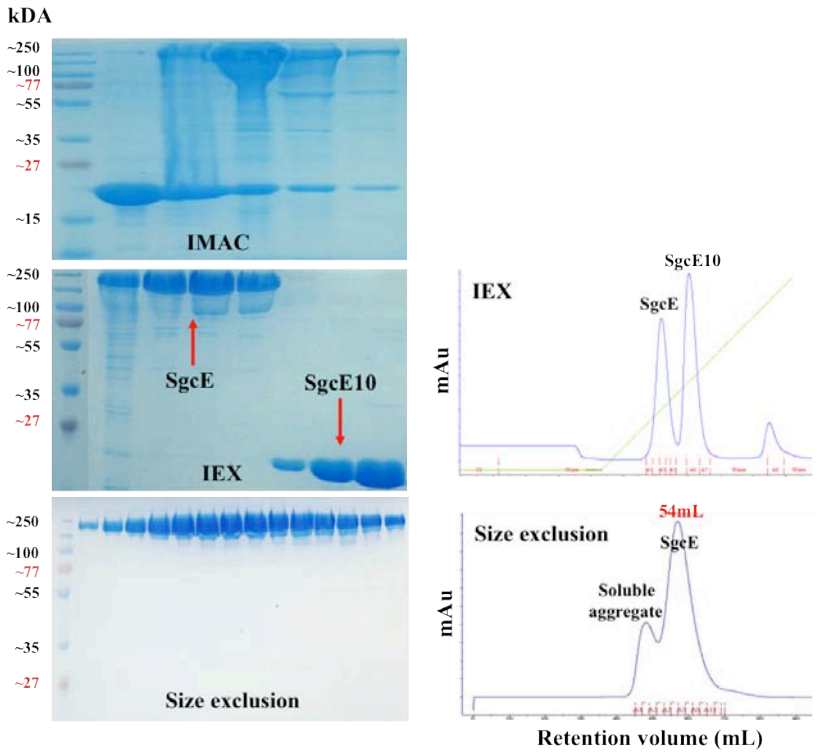



**Figure 2.7** Three steps sequential co-purification of SgcE and SgcE10. IMAC, ion exchange and size exclusion purification methods were used to co-purify SgcE and SgcE10 protein. (*Left panel*) SDS-PAGE analysis of purified protein. (*Right panel*) Chromatogram of ion exchange and size exclusion chromatography.

For the co-purification of DynE8 and DynE7, only IMAC and size exclusion were used to isolate PKSE and TEBC protein. A clear separation of DynE8 and DynE7 was observed (**Figure 2.8**). The IMAC eluent was further purified by size exclusion chromatography, using Superdex S200 column. Both DynE8 and DynE7 were collected separately and concentrated to ~15 mg/mL.

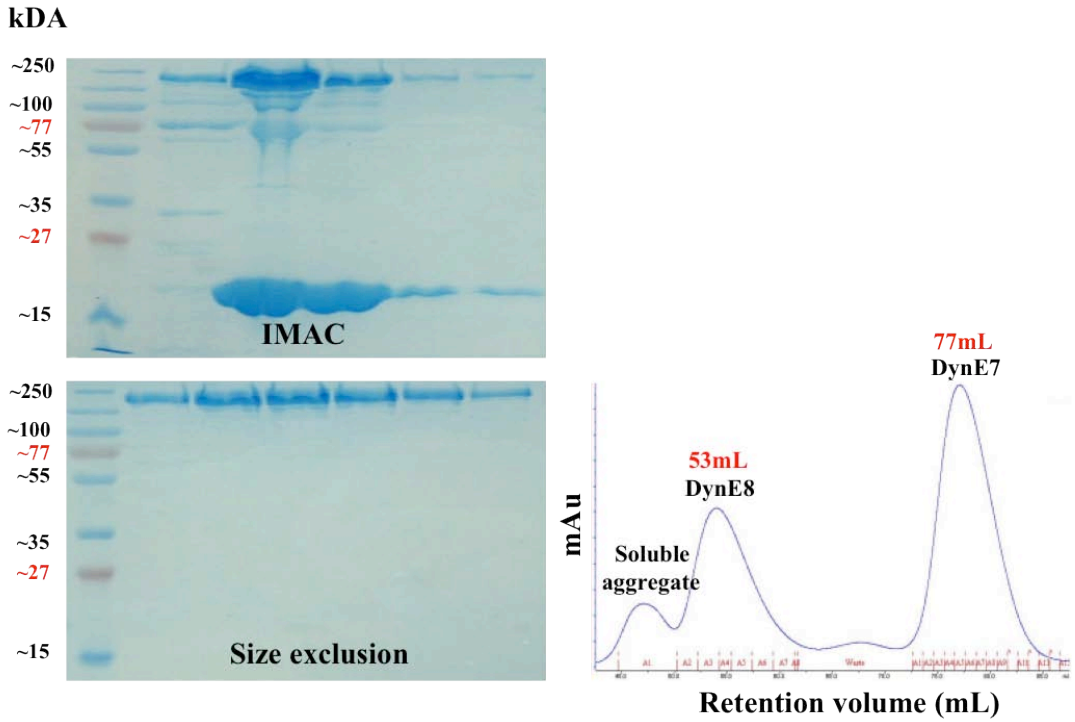



**Figure 2.8** Two steps sequential co-purification of DynE8 and DynE7. (*left panel*) SDS-PAGE analysis of IMAC and size exclusion chromatography. A mixture of DynE8 and DynE7 were isolated from the crude cell lysate by IMAC and further purified by size exclusion chromatography. (*right panel*) Chromatogram of size exclusion chromatography.

### 2.3.3 *In vitro* assay for the KR domain of PKSE

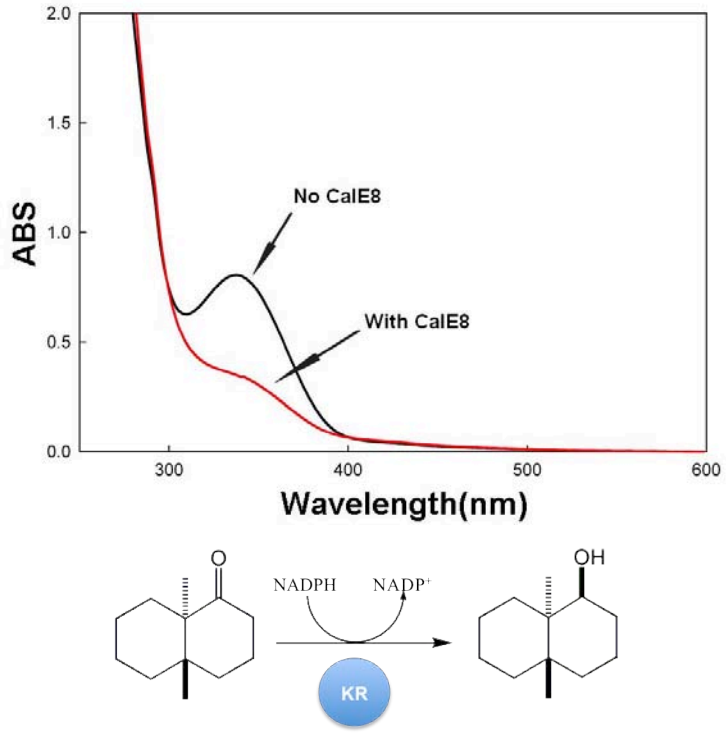



**Figure 2.9** Reaction of CalE8 with decalone and NADPH. The keto group is reduced to a hydroxyl group catalyzed by the KR domain of CalE8 in the presence of NADPH.

As mentioned in **section 1.2.2.2**, type I iterative PKSs can be further categorized into non-reducing, partially reducing and highly reducing type. To categorize and examine the PKSE activity, a simple reduction assay, using PKSE protein to reduce decalone in the presence of NADPH was conducted. Based on the result from the spectroscopic study (**Figure 2.9**), the KR domain of CalE8 appears fully functional as a reductase. The consumption of NADPH implied the conversion of the keto group of decalone to the hydroxyl group only in the presence of CalE8. Without CalE8, a decrease in the level of NADPH is not observed (**Figure 2.9**). As a control, decalone and NADPH were added in the presence of CalE7 instead of CalE8 and no reduction in the NADPH levels was observed. Consumption of NADPH in the presence of decalone is mainly due to the reductive function of the KR domain of CalE8. The reductive function of the KR domain is also expected in SgcE and DynE8 as the KR domains share great similarity.

### 2.3.4 *In vitro* enzymatic assays

Next, to further investigate the function of other proteins in the PKSE gene cassette, a series of *in vitro* enzymatic assays involving PKSE and its auxiliary proteins (TEBC and/or UNBL) were carried out. The yellow pigment is thought to become covalently attached to the phosphopantetheinyl group of the ACP domain, upon activation by self-phosphopantetheinylation with the integrated PPTase domain at the C-terminus[57], by a thioester linkage (**Figure 2.10**). I observed that the yellow pigment could be gradually removed from PKSE by incubation with purified TEBC.

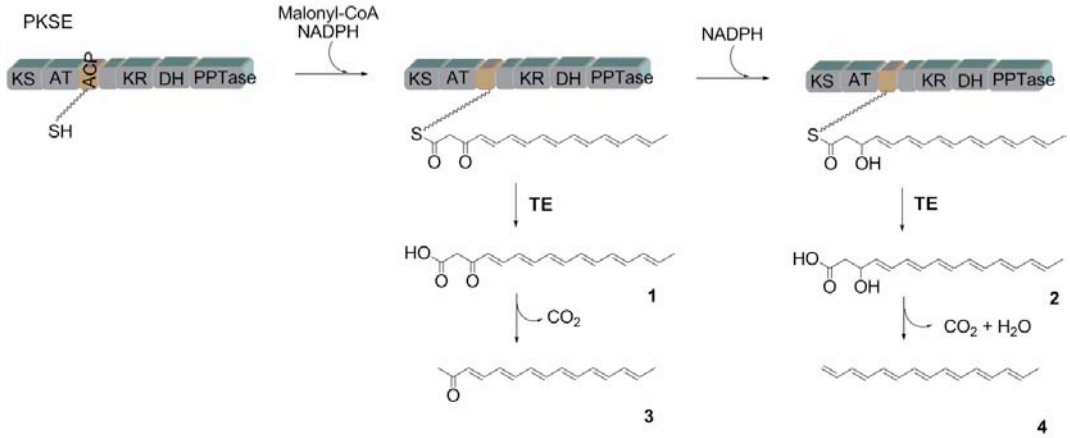



**Figure 2.10** Domain organization of the iterative PKSEs and its products. (KS, ketoacyl synthase; AT, acyl transferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; and PPTase, phosphopantetheinyl transferase).

Upon incubation of PKSE and TEBC, the reaction mixture readily generated products that absorbed light in the range of 350 – 450 nm. As a result, PKSE was shown to be the essential component of the reaction since no products were generated when PKSE was absent. TEBC was essential for the formation of the products in considerable quantities, indicating its plausible role in product release. The activities of the PKSE and TEBC were also investigated by incubating with the putative substrates, acetyl-CoA, malonyl-CoA and NADPH. As shown in **Figure 2.11**, the reaction can proceed without the addition of acetyl-CoA suggesting that PKSE is capable of generating the acetate unit from malonyl-CoA through an intrinsic decarboxylation mechanism. This will be further discussed in **Chapter 3**. Different quenching reagents produced different series of products (**Figure 2.11**). However, based on the cavity shape of the binding channel of TE, only product **3**

and **4** can be released from the PKSE and this has been proved by the TE structure. This will be further discussed in **Chapter 4**.

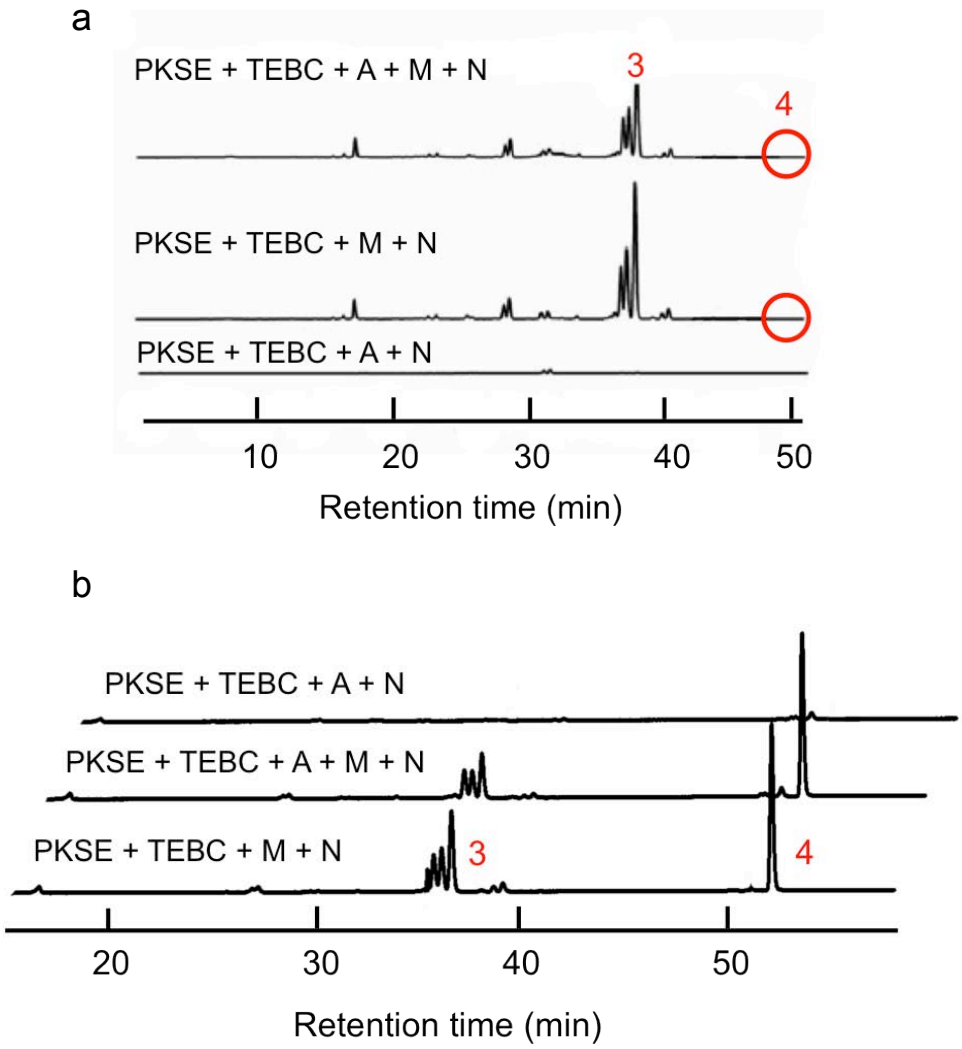



**Figure 2.11** HPLC analysis of *in vitro* products for CalE8/CalE7. Product distribution with enzymatic assays quenched by trifluoric acid, TFA (**a**) and hydrochloride acid, HCl (**b**). *In vitro* enzymatic assay involving PKSE, TEBC, malonyl-CoA, acetyl-CoA and NADPH revealed the formation of carbonyl conjugated polyene (**3**), conjugated polyene (**4**) and several compounds with differential product distribution. (M: malonyl-CoA; A: acetyl-CoA; N: NADPH)

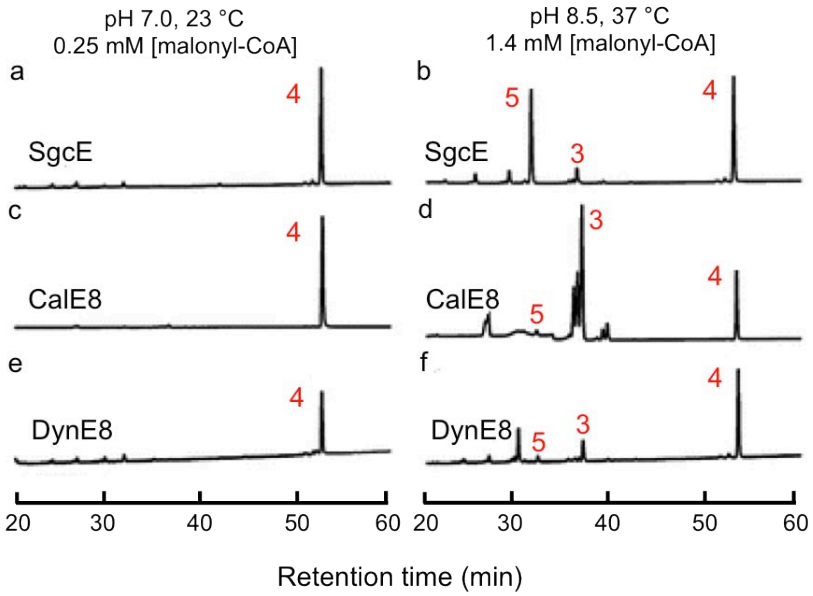



**Figure 2.12** Product analysis for the PKSEs. (a) – (f) HPLC analysis of the products of SgcE, CalE8 and DynE8 with the wavelength set at 400 nm. Two different enzymatic assay conditions were employed for the PKSEs product analysis.

The 9-membered enediyne natural product C-1027 was isolated from *Streptomyces globisporus* C-1027. The production of the insoluble conjugated polyene **4**, deposited in the *E. coli* cell debris after lysis, by iterative SgcE/SgcE10 has been reported by Shen and co-workers[78]. To examine the *in vitro* products of SgcE, an enzymatic assay involving SgcE/SgcE10, NADPH and malonyl-CoA was conducted. The initial results showed that the enzyme produces a series of products, including **3** ($MH^+$, $m/z$ = 215.1432), **4** ($MH^+$, $m/z$ = 199.1478) and several others by-products. Small amounts of **4** were initially observed when trifluoric acid (TFA) was used to quench the reaction prior to product extraction. A significant amount of **4** could be observed only when HCl was used for precipitating the protein (**Figure 2.12b**). The discrepancy in the yield of **4** is likely due to the chemical instability of the **4** in the presence of TFA. Meanwhile, exploration of the assay conditions revealed that **4** becomes the single dominant product under mild assay conditions (pH 7.0, 23 °C, 0.25 mM malonyl-CoA and 0.25 mM NADPH) that moderate catalytic rate was maintained (**Figure 2.12a**). In addition to **3** and **4**, several putative polyene of various lengths and a major by-product **5** ($MH^+$, $m/z$ = 241.1202) with unknown structure were observed where pH, temperature or

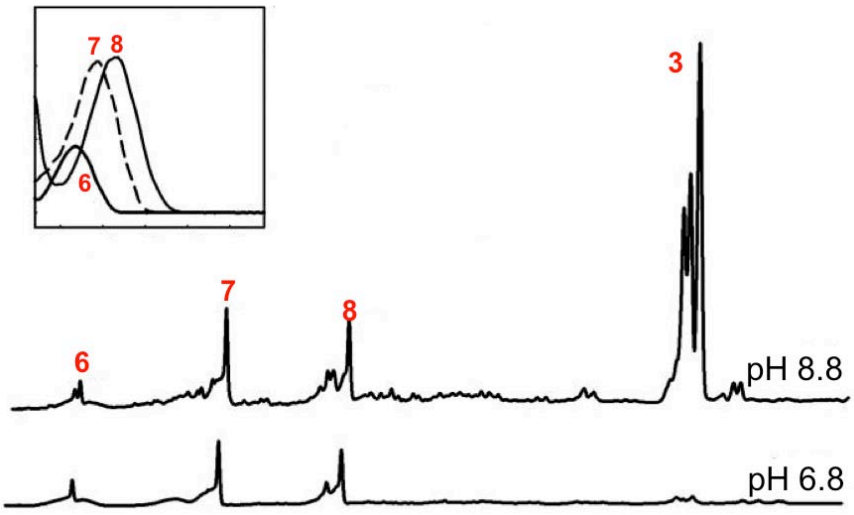



**Figure 2.13** HPLC analysis of the products generated by CalE8 and CalE7 under high and low pH conditions. HPLC chromatograms are shown along with the absorption spectra of **3**, **6**, **7** and **8**.

malonyl CoA concentration were increased (**Figure 2.12b**). Interestingly, only a small amount of **3** was produced under all assay conditions tested.

Similar to SgcE/SgcE10, CalE8/CalE7 produced **3** and negligible amount of **4** when TFA was used to quench the reaction[86; 87]. Given the instability of **4** in the presence of TFA as observed in the SgcE assay, the yield of **4** has been underestimated. With the used of HCl as quenching reagent, both **3** and **4** were observed. Indeed, three previously observed pyrone **6**, **7** and **8** (please refer to page 57) were obtained when HCl was used[87]. At relatively acidic pH, e.g. pH 6.6, the minor products, **6**, **7** and **8** with absorption maxima at 310 nm, 355 nm and 375 nm respectively, were formed as major peaks (**Figure 2.13**). Only at alkaline pH, e.g. pH 8.8, **3** appeared as the major product (**Figure 2.13**). The enzymatic assays further revealed that the product ratio varies considerably with substrate concentration and assay condition. Although **3** was produced as a major product at a high catalytic rate at elevated pH, temperature and malonyl-CoA concentration (**Figure 2.12d**), **4** became the single dominant product when the enzymatic reaction was performed at a moderate rate with sufficient NADPH (**Figure 2.12c**).

The enzymatic assay with purified DynE8 and DynE7 showed that DynE8 generated **4** as the dominant product at a moderate catalytic rate (**Figure 2.12e**). Likewise, DynE8 also produced a series of minor products that include **3** and **5** at high pH and temperature (**Figure 2.12f**). However, **3** was never observed as a major product in the enzymatic assay although both CalE8 and DynE8 have been classified under the same enediyne family. Probably, there are some subtle differences between CalE8 and DynE8. Meanwhile, the production of **4** indicates that CalE7 and DynE7 are fully capable of catalyzing the putative decarboxylation and dehydration of the intermediate **2** (**Figure 2.10**).

We could determine that PKSE has an indispensable role in polyketide product formation, whilst TE assists in product release. From the PKSE enzymatic assay, synthesis of product 4 favors basis pH condition and supplied with sufficient amount of NADPH in moderate rate. Under high rate conditions, it was found that product **3** was the predominant product formed. It was also determined that the product **4** formed is unstable in the presence of TFA.

Apart from PKSE and TEBC, the putative acetylenase UNBL (CalU15, SgcE3 and DynU15) was also tested for activities in the *in vitro*. Various reducing agents, such as ascorbic acid, sodium dithionite and ferrodoxin – ferrodoxin reductase pair, have been employed as electron donors in the oxidative reaction. Unfortunately, no detectable changes occurred to the products formed compared to reaction involving only PKSE and TEBC. The inability of UNBL to carry out the desired desaturation step is probably due to the lack of di-iron cluster or a proper electron donor that is thought to be crucial for the function of UNBL in the catalysis of triple bond formation.

### 2.3.5 LC-MS analysis

It has been established that the enediyne core is formed by at least seven acetate units connected in a head-to-tail fashion. Therefore, **6**, **7** and **8**, could only be a series of aberrant or immature products with the molecular formula of $C_8H_8O_3$ (**6**, [M+1$^+$], m/z 152.05), $C_{10}H_{10}O_3$ (**7**, [M+1$^+$], m/z 179.02) and $C_{12}H_{12}O_3$ (**8**,

[M+1$^+$], m/z 205.02), based on high-resolution mass spectrometry results (**Appendix Figure 2.3** and **2.4**). Although the characterization of the molecular structures of **6**, **7** and **8** by NMR remained difficult given the low yield of the three products, the high-resolution molecular formulae and MS/MS spectra implied that they are most likely the lactones formed in solution from the linear tetra-, penta- and hexaketide PKS products[88]. The observed m/z of 215.1432 ($MH^+$) for **3**, and the two minor components eluted immediately before **3** indicated that they are most likely geometrical isomers sharing the same molecular formula of $C_{15}H_{18}O$ (calculated m/z 215.1430 [$MH^+$]). The MS/MS spectrum exhibits a series of fragments, suggesting that the molecule is likely linear consisting of at least seven acetate units (**Appendix Figure 2.5**). This works were done together with one of our former PhD student, Mr. Kong Rong.

### 2.3.6 Product distribution with enzymatic assays quenched by HCl

Initially **3** was regarded as the only major product from the enzymatic reaction involving PKSE and TEBC after quenching by TFA. However, when hydrochloric acid (HCl) was used instead of TFA to quench the reaction, we discovered that addition of TFA into the reaction mixture actually had adverse effect on the composition of the products in the *in vitro* enzymatic assays. Due to the chemical lability of the **4** under variety of conditions including heat, acid and light[89], the likelihood that the majority of **4** produced was destroyed by TFA appeared as a possibility. Indeed, we could observe both **3** and **4** when HCl was used to quench the CalE8/E7 reaction. The three previously characterized pyrone derivatives (**6**, **7** and **8**) were still found to be present as minor products.

Based on the NADPH concentration-dependent experiment, the amount of NADPH present in the reaction mixture affected the product distribution. The formation of by-product was observed at low NADPH concentration. The peaks of product **3** became highly prominent at the lowest NADPH concentration, whilst the amount of product **4** increased with increasing NADPH concentrations in the reaction mixture. Taken together, the amount of by-product formed in the enzymatic reaction is inversely correlated with the concentration of NADPH at the

expense of the synthesis of **4**. As a result, much of acetate units are channeled into the production of **3** when NADPH is deficient (**Figure 2.14**).

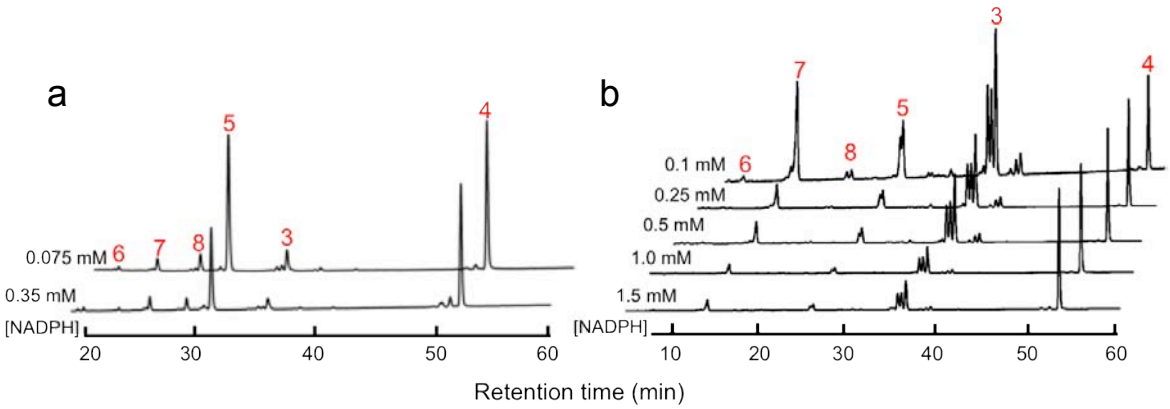



**Figure 2.14** HPLC analysis of the NADPH concentration-dependent production formation for SgcE (**a**) and CalE8 (**b**).

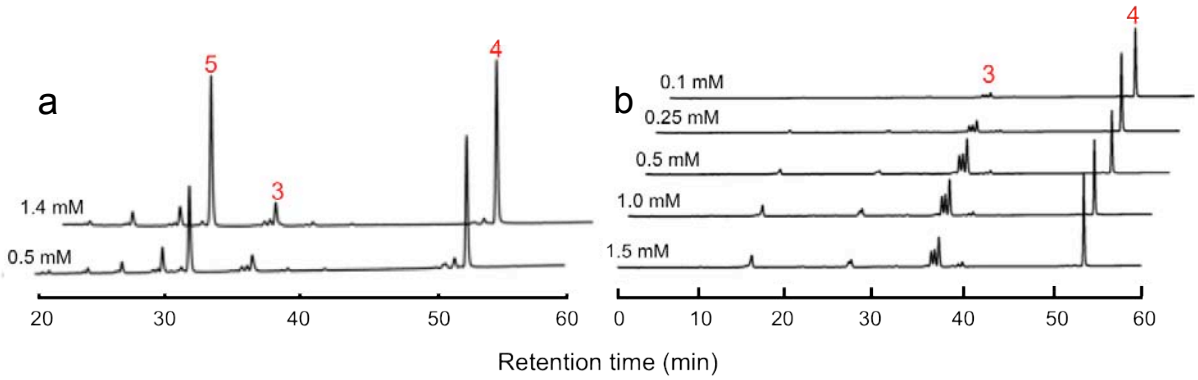



**Figure 2.15** HPLC analysis of the malonyl-CoA concentration-dependent production formation for SgcE (**a**) and CalE8 (**b**).

In the malonyl-CoA concentration-dependent experiment, an excess of malonyl-CoA enhances the production of **3** by a smaller margin during enzymatic assays in comparison to NADPH concentration-dependent enzymatic assays (**Figure 2.15**). Unlike the NADPH concentration-dependent experiments, the correlation between the amount of **4** produced and malonyl-CoA concentration was not obvious. The result raises the possibility that the rate of production of **4** is tightly controlled by SgcE (or CalE8) irrespective of the concentration of malonyl-CoA present.

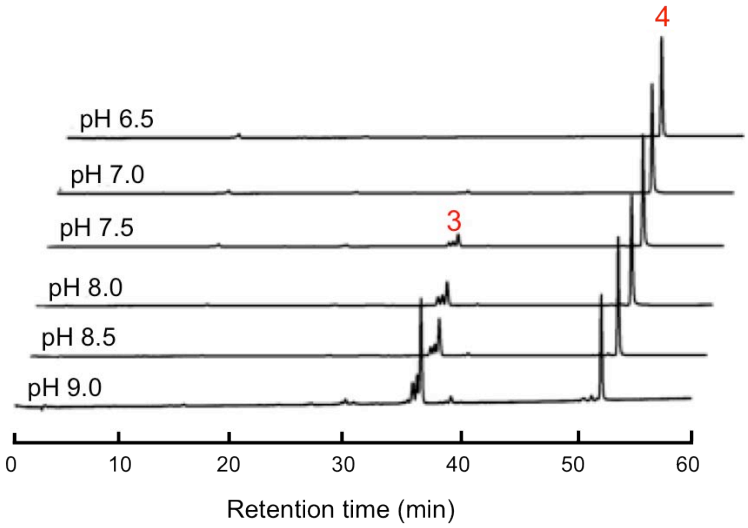



**Figure 2.16** HPLC analysis of the pH-dependent product formation for CalE8

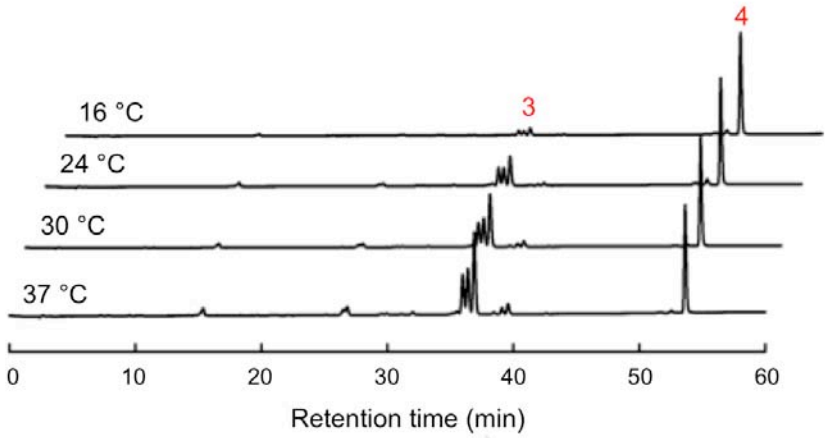



**Figure 2.17** HPLC analysis of the temperature-dependent product formation for CalE8.

Instead, the impact of experimental conditions on product distributions was investigated. As shown by these results, at slightly acidic or neutral pH, the formation of **3** was either abolished or highly suppressed, while the formation of **4** was compromised to a certain extent (**Figure 2.16**). The optimal pH for the production of **3** as well as **4** falls within the range of pH 7.5 to pH 8.5. Besides, the effect of temperature on the formation of 3 is highly prominent. The formation of **3** is totally suppressed at lower temperature and drastically increased at higher temperature. The presence of pyrones was almost negligible within the temperature range investigated (**Figure 2.17**). These results further substantiate the hypothesis that the formation of pyrones is correlated to the substrate concentrations rather than experimental conditions.

### 2.3.7 Crystallization of polyene-free PKSEs – CalE8, SgcE and DynE8

In order to elucidate the structure of PKSEs, crystals were set up. After 7 days of incubation at 13 °C, thin hexagonally shaped crystals (**Figure 2.18a**) were observed. Due to the size and fragility of these crystals, mounting them for X-ray diffraction was a major obstacle. Hence, the optimization of crystallization was performed to improve the quality of the CalE8 crystal by varying the CalE8 protein concentration, pH of the crystallization buffer, temperature and precipitant reagents. Changing incubation temperature from the initial crystallization screening temperature of 13 °C to 4 °C and 20 °C did not cause crystal growth. The replacement of PEG 3,350 with PEG monomethyl ether 2,000 in the original buffer condition enhanced the growth of crystal to a mountable size (**Figure 2.18b**). The size of optimized CalE8 crystal is slightly bigger compared to those crystals from the crystallization screening plate. However, no diffraction was obtained on R-AXIS IV$^{++}$ imaging plate detector (SBS, Nanyang Technological University, Singapore), neither at the ID-23-1 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France).

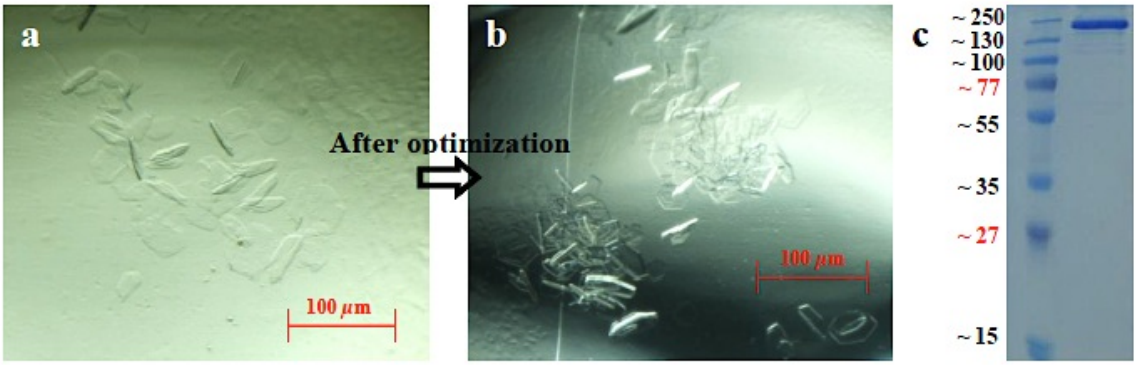



**Figure 2.18** The polyene-free CalE8 crystals. (***a***) A hexagonally shaped polyene-free CalE8 crystals was obtained after 1 week of incubation at 13 °C. (***b***) After several round of crystallization optimization and the replacement of PEG 3,350 with PEG monomethyl ether 2,000 in the original buffer, bigger and well-shaped crystals of polyene-free CalE8 were obtained. However, no diffraction could be recorded. (***c***) SDS-PAGE of the dissolved polyene-free CalE8 crystals. The protein band at ~205 kDa was observed.

In order to verify the content of these crystals, I dissolved polyene-free CalE8 crystals and subjected to SDS-PAGE analysis. From SDS-PAGE, a protein band corresponding to the molecular weight of ~ 205 kDa was observed (**Figure**

**2.18c**). This proved that the crystal contains CalE8 instead of other protein contaminants or salt. Meanwhile, polyene-free SgcE and polyene-free DynE8 failed to produce crystals. A variety of crystallization screening and methods were employed to promote the protein crystallization as well as to enhance the quality of the crystals. The crystallization optimization methods are summarized in **Table 2.2**.

**Table 2.2** Summary of the crystallization optimization

| Screening | Description | Result |
|---|---|---|
| Silver Bullets and Silver Bulltes Bio Screen (Hampton Research) | Screens a portfolio of small molecules for their ability to establish stabilizing, intermolecular, hydrogen bonding, hydrophobic and electrostatic interactions which could promote lattice formation and crystallization. | No crystals were observed after 4 weeks of incubation at 13°C |
| Microseeding | Influencing the nucleation event in standard crystallization screen, especially in cases where spontaneous nucleation is low or the crystal morphology is poor. | No crystals were observed after 4 weeks of incubation at 13°C |
| Additive Screen (Hampton Research) | A library of small molecule that can affect the solubility and crystallization of biological macromolecules as well as perturb and manipulate sample-sample and sample-solvent interactions. | No crystals were observed after 4 weeks of incubation at 13°C |
| Substrates/inhibitors screening | In cases where protein has high mobility, the addition of substrate or inhibitor would help to stabilize the overall structure of the protein | ~150 μm crystals were observed after 4 days of incubation at 13 °C, with $NADP^+$ **Figure 2.18** |
| **Method** | **Description** | **Result** |
| Crystal dehydration | Reduction of solvent content can produce more closely packed and better ordered crystals, extending the resolution of X-ray diffraction patterns. | No X-ray diffraction observed. |
| Microbatch crystallization under oil | Regulating the rate of water evaporation which permits fine-tuning of the crystallization conditions and subsequently increase the concentration of both protein and precipitant. | No X-ray diffraction observed. |

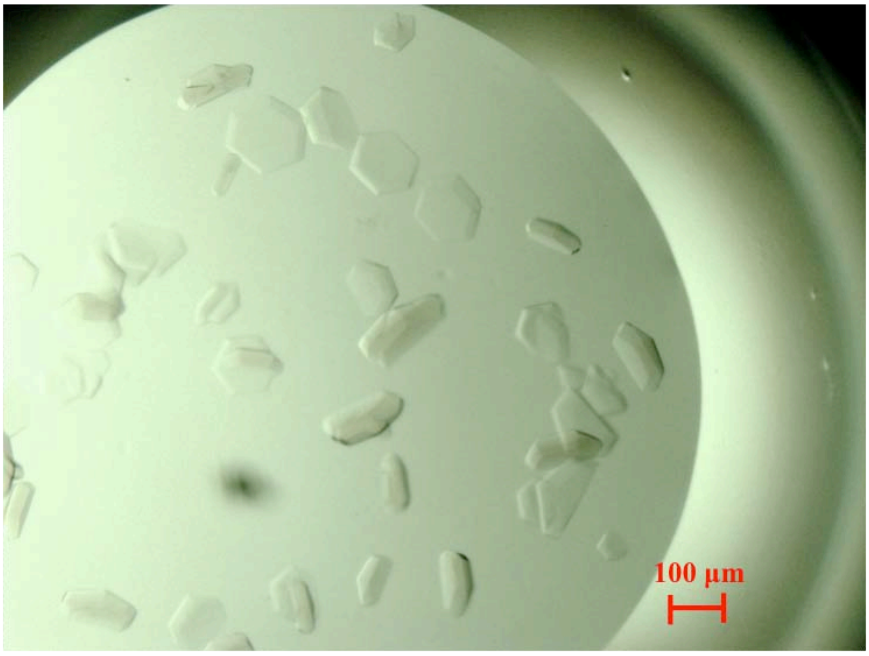



**Figure 2.19** CalE8 crystals. The crystals were obtained after 4 days of incubation at 13 °C, with $NADP^{+.}$ Each crystal is about 70 µm and more rigid compared to the previous batch of crystals obtained (see **Figure 2.17**).

Co-crystallization of CalE8 with $NADP^{+}$ yields a better shape and size of crystals (**Figure 2.19**). Despite the increase in the size of crystal, no diffraction was obtained. Various post-crystallization treatments were carried out to improve diffraction quality. However, no improvement was gained (**Table 2.2**). Currently, the crystallization optimization is still in progress. Meanwhile, no crystal was observed from the remaining crystallization screening of SgcE and DynE8.

## 2.4 Discussions

### 2.4.1 Protein degradation problem

Protein degradation has been a major problem encountered during purification of PKSE proteins. The recombinant PKSEs were $His_6$-tagged at both N- and C-termini. Hence, all incomplete fragments that include either terminus would be co-purified with the intact proteins during IMAC. The only way to reduce binding of fragments would be to cleave off the N-terminal $His_6$ tag by using thrombin. This could prevent fragments that lack C-terminal portions from being co-purified.

### 2.4.2 *In vitro* enzymatic assays

Based on sequence homology of the PKSEs, CalE8, SgcE and DynE8, the three PKSEs are highly homologous despite the observed differences in the enediyne cores produced. They share the same domain organization with a sequential arrangement of KS, AT, ACP, KR, DH and PPTase spanning from N- to C-terminus. Unlike most of the PKSs that require a starter molecule for the initiation of decarboxylation condensation, the KS domain of PKSE is presumably capable of generating starter molecule *de novo* from the malonyl-CoA substrate via intrinsic decarboxylation. Based on these experimental results, PKSEs does not require acetyl-CoA in the reaction mixture for the synthesis of full-length products **3** and **4**. In addition, the presence of the acetyl-CoA is potentially inhibiting the catalytic activity. Importantly, a negligible amount of **3** and **4** could be observed in the absence of TEBC under all conditions, indicating that TEBC is responsible for the release of the octaketides linear intermediates. Moreover, TEBC must undergo extensive conformational changes to produce an elongated substrate-binding channel in order to accommodate the lengthy thioester precursors of **3** and **4**. Along with the production of **3** and **4** (**Figure 2.10**), the immature products **6**, **7**, and **8** are also observed in the PKSE/TEBC reactions (**Figure 2.20**).

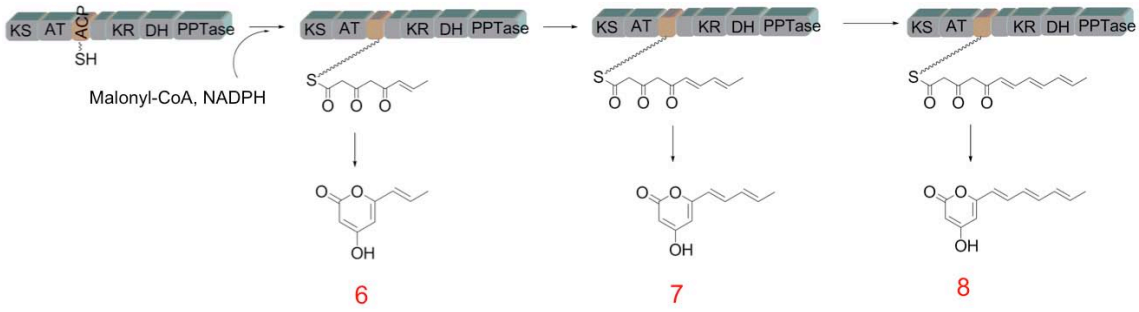



**Figure 2.20** The different pyrones generated by PKSE. A schematic representation of PKSE generating the truncated polyketides of **6**, **7** and **8**.

The formation of the by-product pyrones is highly dependent on the concentration of NADPH present in the reaction mixture[88]. The deficiency in NADPH impedes the function of the KR domain in the conversion of the keto group to hydroxyl group on the growing acyl chain. Due to the high acidity of α-hydrogen, 3,5-diketo thioester intermediates are highly susceptible to self-cyclization. The cyclization process is normally spontaneous and does not require the aid of a thioesterase for release. The formation of pyrones as by-products is inevitable in most of the FAS or PKS reactions. However, an ample supply of NADPH in the reaction mixture will reduce the formation of the pyrones. Most of the resources used for the formation of pyrones will be then channeled into the production of full-length products.

The synthesis of **4** as the major product under physiologically relevant conditions suggests that **4** is the polyketide intermediate *en route* to the construction of enediyne core. The formation of the **4** becomes the dominant product that suppresses the generation of other polyketides like **3**, and pyrones **6**, **7** and **8**. Thus, **4** could be the authentic biosynthetic intermediate for both 9- and 10-membered enediynes as suggested by the studies on DynE8 and SgcE as well[90]. Based on their propensity to generate different polyketide products, some intrinsic differences may exist among the PKSEs in 9 and 10-membered enediyne biosynthesis[90].

The capability of the PKSs to generate more than one full-length product is remains unclear at this moment. TE might play a role in the formation of different products. It is likely that the catalytic cavity of enediyne PKS is in fact easily accessible to TE. The event of thioester cleavage takes place once the acyl chain grows to a designated length. Provided the catalytic rate of TE is faster than that of the KR domain in PKS, the octaketide chain is readily cleaved even before the KR domain can reduce the β-keto group on the thioester intermediate. In accordance with the pH-dependent experiments, high pH is likely to facilitate the cleavage of the thioester through increased availability of hydroxide ions in the catalytic cavity which act as nucleophiles in the hydrolytic reaction. An increase in the conformational flexibility of PKS might grant easy access of the ACP domain to TE at elevated temperature, resulting in the increased production of 3 since less time would be allowed for the KR domain to reduce the β-keto group. Hence, a plausible conclusion can be drawn whereby the indiscrimination of TE towards the full-length acyl chains leads to observed product variations. Considering that iterative PKSs may or may not require ancillary proteins to generate the correct product, further evidence is required to fully establish **4**, or **3**, or even another PKS product as the biosynthetic intermediate.

In spite of the fact that both **4** and **3** are formed as the putative full-length linear products (**Figure 2.10**), we cannot rule out the possibility that both compounds are merely by-products of PKSE. Up to now, there is still no consensus on which PKS product is the real intermediate for 9- and 10-membered enediyne core. Therefore, the point at which the evolutionary divergence occurs between the 9- and 10-membered enediyne biosynthetic pathways remains to be established. One school of thought claims that the divergence between the two classes of enediyne takes place after the PKS-TE stage. With evidence from bioinformatics studies, as well as several *in vitro* experiments, a common precursor is envisioned for the early stage of enediyne biosynthesis, such as the **4** described above. The auxiliary enzymes of the enediyne system, potentially the cyclase or acetylenase, will dictate the differentiation of the common precursor into the respective core structures of the 9- and 10-membered enediynes. In contrast, another proposed

biosynthetic pathway depicted the occurrence of the evolutionary divergence at PKS level, whereby the subtle intrinsic differences in the PKSs resulted in the production of 9- and 10-membered enediynes via distinct polyketide intermediates. However, both models require further substantiation because the crucial link between the polyketide intermediate and the folded enediyne core is still veiled in mystery. Unraveling the subsequent post-PKS steps is the key to fully decipher the biosynthetic pathway of enediynes.

### 2.4.3 CalE8 from co-expression system more readily for crystallization

The conserved serine residue, S966, in the ACP domain attaches to a phosphopantetheine moiety, is essential for binding and shuttling substrates[16]. The prosthetic group is flexible and transfers intermediate molecules across domains in a "swinging arm" manner; it has also been proposed that entire ACP may mobilize to interact with other domains[91]. Hence, substrate-bound PKSE might possess a number of conformational subpopulations that covalently bind different reaction intermediates, resulting in lowered homogeneity and probability to crystallize. This could be the main reason that single expressed PKSE protein did not yield any crystals. Therefore, by co-purifying PKSE with TEBC, the octaketides linear intermediates covalently attached to the phosphopantetheinylated-ACP domain will be released and accommodated by TEBC. Without the polyketide intermediates bound to the ACP domain, the PKSE protein is expected to be more homogenous.

# CHAPTER 3 Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis

## 3.1 Introduction

As secondary metabolites produced by soil and marine microorganisms, fungi and plants, polyketides form a structurally diverse family of natural products endowed with a variety of biological and pharmaceutical properties. Many polyketides are synthesized by type I PKSs that can be further classified into modular or iterative PKSs[92]. While modular type I PKSs assemble the polyketides using multiple modules, each composed of several catalytic domains, iterative type I PKSs synthesize their polyketide products iteratively using a single module[72]. In contrast to the large body of knowledge accumulated for modular PKSs in the last two decades, much remains to be learned about the catalytic mechanism of iterative PKSs. Particularly, despite sharing the same repertoire of catalytic domains with modular type I PKSs, many iterative type I PKSs feature highly unusual programmed reduction, dehydration or methylation during the chain extension processes, possibly owing to some undisclosed properties of their catalytic domains.

Enediyne natural products are among the most potent antitumor and antibiotic agents ever discovered[31; 80; 93]. Their remarkable antitumor properties derive from the ability of the enediyne "warhead" to induce cell apoptosis through chromosomal DNA cleavage via an oxidative radical mechanism. Recent studies revealed that enediyne biosynthetic pathways involve an iterative type I PKS that initializes the biosynthesis of the 9- or 10-membered bicyclic enediyne warhead[54; 56]. The PKSs responsible for the biosynthesis of several naturally occurring enediynes have been characterized by *in vitro* and *in vivo* studies[50; 52; 55; 56; 61; 62]. In the presence of the hot-dog fold thioesterase DynE7, the enediyne PKS DynE8 produces several linear conjugated polyene products[90]. Enediyne PKSs catalyze the synthesis of the linear polyenes through iterative cycles of condensation, ketoreduction and dehydration (**Figure 3.1**). Based on amino-acids sequence alignment and experimental evidence, the enediyne PKSs, such as CalE8, SgcE8 or DynE8, possess a domain organization that is different from known PKSs[52; 54; 55; 56].

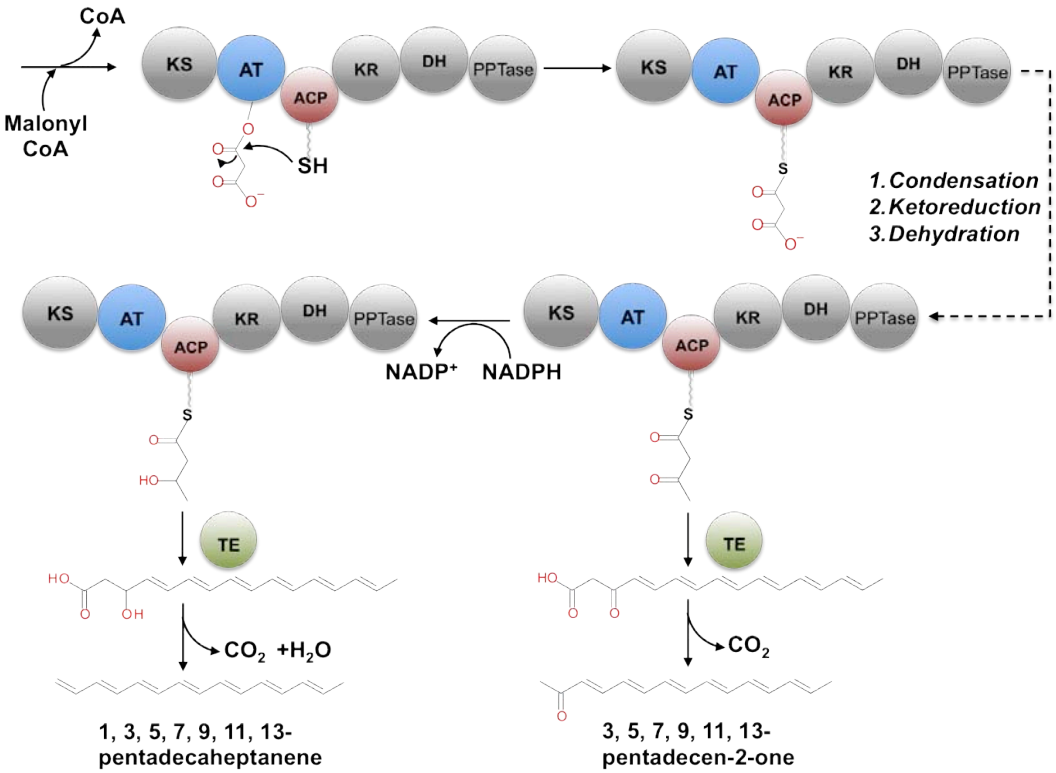



**Figure 3.1** A simplified diagram of the enediyne polyketide biosynthesis pathway and domain organization of DynE8. KS: ketoacyl synthase, AT: acyltransferas, ACP: acyl-carrier protein, KR: ketoreductase, DH: dehydratase, PPTase: phosphopantetheinyl transferase, TE: thioesterase.

In addition to the essential ketoacyl synthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domains required for polyketide chain elongation, the processing ketoreductase (KR) and dehydratase (DH) domains were also identified. A C-terminal 4′-phosphopantetheinyl transferase (PPTase) domain was also documented for the addition of the phosphopantetheinyl moiety onto the ACP domain of the enediyne PKS (**Figure 3.1**)[52; 54; 55; 57]. The NMR structure of the ACP domain of CalE8 has revealed some interesting features of the shuttling domain[94] and crystal structures of the thioesterases DynE7 and CalE7 responsible for off-loading the conjugated polyenes synthesized by CalE8 and DynE8 respectively were also reported[87; 95]. Thus, determination of the 3D structures of the other catalytic domains (KS, AT, KR and DH) of the enediyne PKSs should substantially enhance our understanding of the mechanism governing enediyne biosynthesis and perhaps open up possibilities for rational alterations of the warhead.

For chain elongation, the AT domains of enediyne PKSs catalyze the transfer of the malonyl group from malonyl-CoA to the ACP domain (**Figure 3.1**). Interestingly, unlike the AT domain of other iterative PKSs, the AT domains of the enediyne PKSs do not seem to use acetyl- CoA as a starter unit but solely malonyl-CoA. In this chapter, a high-resolution crystal structure of the AT catalytic domain and of the adjacent KS-AT linker domain of DynE8 from the dynemicin biosynthetic pathway are presented. This fragment of DynE8 is subsequently referred to as $AT_{DYN10}$. In addition to the unbound structure, the structure of the malonyl-enzyme covalent complex and the structure of the enzyme in complex with acetate and glycerol were also determined. To our knowledge, this is the first crystal structure for an AT domain from an iterative PKS. The crystal structures provide insight into the evolutionary relationship between the AT domain of an iterative PKS and the AT domains of FAS and modular PKSs. Structural determinants of the substrate preference for malonyl-CoA are also unveiled. Moreover, structural and biochemical data suggest that although $AT_{DYN10}$ does not use acetyl-CoA as starter unit, $AT_{DYN10}$ is able to upload the acetyl group of acetyl-CoA to form an acetyl-enzyme covalent complex that is susceptible to hydrolysis. Comparison of the structures of $AT_{DYN10}$ with the acetyl- and malonyl-specific AT domain of mammalian FAS reveals structural differences in the active sites of the two AT domains that may account for the different fate of the acetyl-enzyme intermediate.

## 3.2 Material and methods

### 3.2.1 Reagents and chemicals

Coenzyme A (CoA), malonyl-CoA, acetyl-CoA, NADPH, and all other chemicals were purchased from Sigma-Aldrich and stored at -20 °C. Selenomethionine base medium was purchased from Molecular Dimensions. Crystallization screening buffers were purchased from Hampton Research.

### 3.2.2 Construction, expression and purification of the $AT_{DYN10}$ fragment

Following gene annotations by Zazopoulos and co-workers[54], a construct encompassing residues 473 – 893 of DynE8 (hereafter named $AT_{DYN10}$ see **Figure 3.2** for its subdomain composition**)** from *Micromonospora chersina* was cloned into to pNIC28-Bsa4 (GenBank accession EF198106), resulting in a construct with an N-terminal hexahistidine tag and a TEV protease recognition site. The positive recombinant clone was re-transformed and expressed in T1-phage resistant BL21-DE3 Rosetta strain. For expression, cells were grown in a LEX system[96] using 0.75 L Terrific Broth medium supplemented with 8 g/L glycerol, 50 μg/mLkanamycin and 34 μg/mL chloramphenicol. Cells were incubated at 37 ºC and at $OD_{600}$ ~2, the temperature was reduced to 18 ºC. After 30 – 60 min, the expression of the target protein was induced by addition of 0.5 mM IPTG for 17 – 20 hours.

Cells were harvested by centrifugation and resuspended in lysis buffer (100 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10 % (v/v) glycerol, 0.5 mM TCEP, pH 8.0), supplemented with the Protease Inhibitor Cocktail Set III, EDTA free (Calbiochem) and 2000 U of benzonase (Merck) and stored at −80°C. Cells disruption was performed by sonication on ice (using Sonics Vibra-cell at 70% amplitude, 3s on/off for 3min). The lysate was clarified by centrifugation at 47,000 g for 25 minutes at 4 °C and the supernatant filtered through a 1.2 μm syringe filter. Filtered lysates were loaded onto 1 mL Ni-NTA Superflow (Qiagen) in IMAC Wash buffer 1 (20 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10 % (v/v) glycerol, 0.5 mM TCEP, pH 7.5) and subsequently washed with IMAC Wash buffer 2 (20 mM HEPES, 500 mM NaCl, 25 mM imidazole, 10 % (v/v) glycerol, 0.5 mM TCEP, pH 7.5). Bound proteins were eluted with 500 mM imidazole and loaded onto a HiLoad 16/60 Superdex-200 column (GE Healthcare) pre-equilibrated with 20 mM HEPES, 300 mM, NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, pH 7.5. Fractions containing $AT_{DYN10}$ were pooled. TCEP was added to a final concentration of 2 mM and the sample was concentrated in Vivaspin 20 filter concentrators (15 kDa M.W. cutoff) at 15 °C. The final protein concentration was 30 mg/mL in a volume of 0.7 ml. The protein batch was aliquot, frozen in liquid nitrogen and stored at -80 °C.

### 3.2.3 Preparation of selenomethyonylated $AT_{DYN10}$

To obtain crystallographic phase information, selenometylonylated $AT_{DYN10}$ protein was prepared. The pNIC28-Bsa4-*$AT_{DYN10}$* plasmid was transformed into *E. coli* B834 (DE3). A 100 mL seed culture was grown overnight at 37 °C in Luria Bertani medium supplemented with 50 µg/mL kanamycin and L-methionine. The cells were then pelleted, washed 3 times with Selenomethionine base medium (Molecular Dimensions) and inoculated into 2 L of pre-aerated, pre-warmed selemethionine expression medium supplemented with 40 mg/L of L-selenomethionine and 50 µg/mL kanamycin. At an $OD_{600}$ of 0.6, the culture was cooled to 18 ºC and protein expression was induced with 0.1 mM IPTG. Protein purification was carried out as described above for the free enzyme.

### 3.2.4 Crystallization and X-ray diffraction data collection of native, substrate-enzyme complexes and selenomethyonylated crystals

Crystals of free $AT_{DYN10}$ were obtained after screening 672 crystallization conditions (Hampton Research), using the CyBio®-Crystal Creator robot (Jena Biosciences). Native $AT_{DYN10}$ crystals were grown in sitting drops comprising equal volumes of protein solution (15 mg/mL) and precipitant solution (0.1 M Tris HCl and 30% PEG 6000, pH 8.5). Crystals of malonate- and acetate bound complexes were grown in an analogous manner with prior addition of 5.0 mM malonyl- CoA and acetyl-CoA respectively to the protein solution. Crystals of the substrate-enzyme complex were obtained at a temperature of 12 °C. After 48 hours of incubation of the co-crystallization mixture, these crystals were flash frozen in liquid nitrogen following a short transfer to the precipitant solution supplemented with 25% (v/v) glycerol. Crystals of the selenomethyonylated protein were obtained in the same conditions as the wild-type enzyme. Before data collection, crystals were transferred to a cryo protecting solution containing the precipitating solution supplemented with 25% (v/v) glycerol and cooled to 100 K in liquid $N_2$. Diffraction intensities were collected at the Swiss Light Source (Switzerland) beamline PXI and PXIII, using a Dectris Pilatus 6M detector, to a resolution of 1.40 Å for the free enzyme, to 1.65 Å for the covalent malonate enzyme complex, to 1.90 Å for the enzyme bound to acetate, to 1.5 Å for a complex with glycerol and to 2.5 Å for the

*Se* data set. Integration, scaling and merging of the intensities were carried out using programs MOSFLM and SCALA from the CCP4 suite[97]. Data collection statistics are summarized in **Table 3.1**.

### 3.2.5 Structure determination and refinement of the $AT_{DYN10}$ and complexes

The structure of the $AT_{DYN10}$ was determined using a combination of the molecular replacement and SAD techniques with data recorded at the Selenium (*Se*) absorption edge from the selenomethionine-substituted protein. A partial model of $AT_{DYN10}$ comprising 223 amino acids was obtained using program MOLREP and the *Spn*MCAT[98] structure from *Streptococcus pneumoniae* (PDB code 3IM8) as a search probe. Phases calculated from this partial model were used to locate four selenium sites (out of a total of five *Se* per asymmetric unit) using an anomalous difference Fourier map to 5 Å resolution at the *Se* edge. The position of the four peaks in the anomalous Fourier map matched the location of $Met^{680}$ (*Se*1), $Met^{690}$ (*Se*2), $Met^{757}$ (*Se*3) and $Met^{828}$ (*Se*4) of the $AT_{DYN10}$ partial model and were also in agreement with the *Se* positions generated by SHELXD[99], giving confidence in this partial solution from molecular replacement. The program SHARP[100] was then used to refine the four *Se* positions and calculate initial SAD phases and the residual map yielded the fifth *Se* position corresponding to $Met^{853}$. A new set of SAD phases was calculated with an overall figure of merit of 0.12 to 2.5 Å resolution. The program REFMAC[97] was used to refine the partial model using the maximum likelihood target and the SAD phase distribution calculated using the program SHARP[100], encoded as Hendrickson-Lattman coefficients. After several rounds of manual and automatic iterative model building using BUCANNEER[97] and the graphic display program COOT[101] followed by refinement with REFMAC[97], the updated partial model could be successfully fed into program Arp/Warp[102] which delivered a much more complete model with an *R*free of 26.3%, using native data to 1.4 Å resolution. The connectivity of this model was modified manually using COOT[101]. For each complex, a difference Fourier synthesis map revealed residual electron density in the active site corresponding to the bound ligand. Statistics for the binary complexes are summarized in **Table** 3.**1**. The quality of final models were assessed with PROCHECK[103]. All structural graphics were generated using PYMOL[104].

### 3.2.6 Modeling of the $AT_{DYN10}$-$ACP_{DYN}$ complex

A model for the ACP domain of DynE8 ($ACP_{DYN}$) was generated by homology using the iterative threading assembly refinement (ITASSER) server[105; 106]. Macromolecular docking calculations for $AT_{DYN10}$ and $ACP_{DYN}$ were performed with ClusPro[107; 108] (28,29). In the simulation, the orientation of $AT_{DYN10}$ was kept fixed, whereas $ACP_{DYN}$, as ligand, was allowed to rotate and translate. A total of 57 conformations with lower energies and larger cluster sizes were generated based on the multistage protocol that includes rigid body docking, energy-based filtering, clustering properties ranking and energy minimization refinement. The model for the $AT_{DYN10}$/$ACP_{DYN}$ complex was selected based on the lowest energy and largest cluster size criteria as well as comparisons with previously reported AT/ACP complexes.

### 3.2.7 Enzymatic assay and product analysis by HPLC

All enzymatic reactions were performed in a final volume of 15 µL. Each enzymatic reaction was measured at different time points of incubation. A reaction mixture includes 2.38 µL of $AT_{DYN10}$ (630 mM), 3.0 µL of acetyl-CoA (50 mM) and 9.62 µL of reaction buffer (100 mM HEPES pH 8.5 and 100 mM NaCl). The reaction mixtures were incubated at 30 °C for 1 hour, 2 hours, 9 hours and 16 hours. Each sample was boiled and spun down before application onto an analytical eclipse XDB C18 column (4.6 x 250 mm) using an Agilent 1200 HPLC. A 40-min linear gradient was used starting with 100% buffer A (HPLC grade water with 0.045 % trifluoroacetic acid) to 40 % buffer B (acetonitrile with 0.045 % trifluoroacetic acid). The standards, acetyl-CoA and CoA, were diluted to 10 mM with the reaction buffer and incubated at 30 °C for 16 hours prior to application onto a XDB C18 column. The diodearray detector was set at 254 nm with a reference wavelength of 600 nm.

### 3.2.8 Acetyl-CoA inhibition assay

The co-purified DynE8 and thioesterase DynE7 prepared from a co-expression system[95] were used for the inhibition assay. The assays were carried out with 1.12 μM DynE8, 10.5 μM DynE7, 0.25 mM NADPH, 2.5 mM malonyl-CoA and various concentrations of acetyl-CoA (1.0 mM, 2.0 mM, 2.5 mM and 5.0 mM) in a total volume of 200 μL of buffer (25 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM DTT) at 30 °C. The reaction mixture without acyl-CoA was equilibrated at 30 °C for 10 min in the temperature-controlled sample chamber of Shimadzu UV-visible 1700 spectrometer. Enzymatic reaction was initiated by the addition of malonyl- or acetyl-CoA. A full wavelength spectrum scan was performed at 2-min intervals throughout the course of the experiment. Positive control with 2.5 mM malonyl-CoA as a substrate was used for the comparison.

### 3.2.9 Accession numbers

The refined coordinates and structure factors amplitudes have been deposited in the PDB with accession code 4AMP for the malonate-enzyme complex, 4AMO for the acetate-enzyme complex, 4AMN for the glycerol-enzyme complex and 4AMM for the unliganded enzyme.

## 3.3 Results

### 3.3.1 Cloning and expression of the $AT_{DYN10}$ domain.

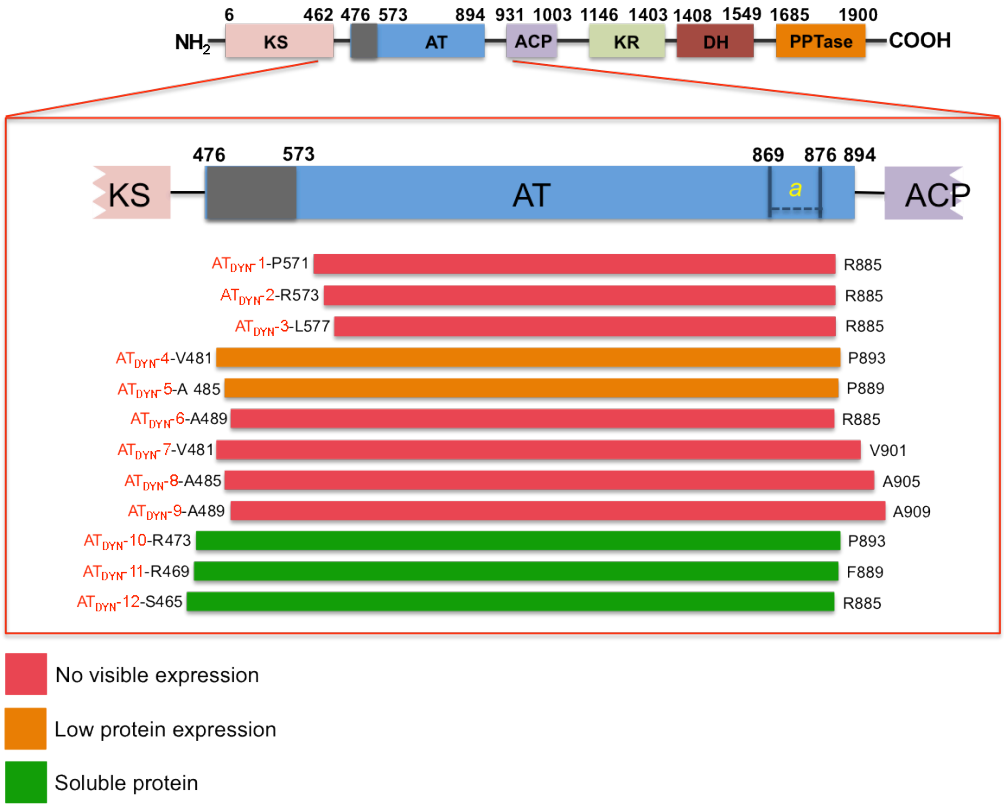



**Figure 3.2** Construction and expression of truncated AT fragments of DynE8. Three out of twelve recombinant protein could be purified (*green* bar), *Orange* and *red* bars indicate low or no soluble protein expression respectively. The boundaries of the KS-$AT_{DYN}$ linker domain are indicated by *grey* box. The construct that was crystallized is $AT_{DYN10}$

A soluble fragment ($AT_{DYN10}$ residues 473 – 893), comprising the predicted acyltransferase domain of the enediyne PKS DynE8 (1905 residues) from *Micromonospora chersina*, was successfully cloned and overexpressed in *E. coli* expression system (**Figure 3.2**). Construct boundaries were assigned based on annotations of the enediyne PKS biosynthesis genes reported by Thorson and co-workers[52; 54; 56] and sequence comparisons with other structurally characterized AT domains. The segment flanked by the KS and AT catalytic domain (named as KS-$AT_{DYN}$ linker domain) and the helix immediately downstream of the AT catalytic domain were retained because of their possible importance for the recognition of the $ACP_{DYN}$ domain of DynE8 and for acyl transfer[109; 110; 111; 112; 113]. The KS-$AT_{DYN}$ linker domain appears critical for solubility because constructs ($AT_{DYN1}$ – $AT_{DYN3}$)

that lack the KS-$AT_{DYN}$ linker domain yielded insoluble proteins (**Figure 3.2)**. Similar results were observed for constructs ($AT_{DYN6}$ – $AT_{DYN9}$) that include a partially truncated KS-$AT_{DYN}$ linker domain. Three fragments with an N terminus at residue 473, 469, or 465 respectively (named $AT_{DYN10}$ – $AT_{DYN12}$, **Figure 3.2**) yielded soluble proteins that could be purified and concentrated to approximately 30 mg/ml. These results demonstrate the crucial role played by residues 473 – 480 for the stability and solubility of the recombinant protein. The construct that spans residues 473-893 (named $AT_{DYN10}$) was used for crystallization and structure determination.

### 3.3.2 Crystal structure determination

The crystal structure of $AT_{DYN10}$ was solved using a combination of molecular replacement and single anomalous dispersion (SAD) at the Selenium absorption edge using the seleniated protein. The structure was refined to *R*work of 17.4% and *R*free of 20.0% using data to a resolution of 1.4 Å (**Table 3.1**). Statistics for the structures of the binary complexes of the protein fragment with glycerol, acetate and malonate are given in **Table 3.1**. The ligand of all binary complexes is well defined.

One $AT_{DYN10}$ protein is present in the asymmetric unit and the final model comprises residues 476 – 594 and 612 – 876 of the DynE8 protein. Due to poor electron density in this region, residues 595 – 611 are absent from the model. This region that corresponds to a helical flap in homologous AT proteins is presumably flexible in the absence of other DynE8 domains. The overall structure of $AT_{DYN10}$ is consistent with the important role played by the KS-$AT_{DYN}$ linker domain for protein solubility: close packing at the interface between the linker domain and the larger subdomain of $AT_{DYN10}$ shields non-polar surfaces and is important to overall protein stability.

**Table 3.1:** Data collection, phasing and refinement statistics

| | $AT_{DYN10}$ | Selenomethionine derivative | $AT_{DYN10}$ - glycerol complex | $AT_{DYN10}$ - acetate complex | $AT_{DYN10}$ - malonate complex |
|---|---|---|---|---|---|
| ***Data Collection*** | | | | | |
| X-ray source | PXI, SLS | PXIII, SLS | PXIII, SLS | PXI, SLS | PXI, SLS |
| Wavelength (Å) | 1.000 | 0.9791 | 1.072 | 1.072 | 1.072 |
| ***Crystallographic parameters*** | | | | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell dimensions (Å) | 65.5, 66.1, 85.4 | 65.9, 67.3, 85.2 | 65.8, 66.4, 85.5 | 66.5, 68.4, 85.7 | 65.7, 66.2, 86.1 |
| a,b,c (Å) ; α, β, γ (°) | 90°, 90°, 90° | 90°, 90°, 90° | 90°, 90°, 90° | 90°, 90°, 90° | 90°, 90°, 90° |
| ***Phasing*** | | | | | |
| $R_{cullis}$ (acentric)[a] | - | 0.960 | - | - | - |
| Phasing power (acentric)[b] | - | 0.436 | - | - | - |
| Acentric/centric | - | 0.115 / 0.034 | - | - | - |
| Number. of Se | | 5 | | | |
| ***Data Collection Statistics*** [c] | | | | | |
| Resolution limit (Å) | 23.3 – 1.4 | 48.1 – 2.5 | 18.8 – 1.5 | 29.7 – 1.9 | 26.3 – 1.7 |
| No. of observed reflections | 311,060 | 40,354 | 239,417 | 308,772 | 214,889 |
| No. of unique reflections | 72,439 | 8,016 | 54,690 | 30,505 | 45,596 |
| Completeness (%) | 98.6 (95.5) | 99.9 (100) | 91.0 (96.5) | 100 (100) | 99.4 (99.5) |
| Redundancy | 4.3 (3.3) | 5.0 (5.3) | 4.4 (4.4) | 10.2 (10.2) | 4.7 (4.8) |
| Average I/σ(I) | 9.7 (2.5) | 11.6 (4.9) | 11.0 (3.5) | 20.5 (5.1) | 11.1 (2.8) |
| $R_{merge}$ (%) | 0.081 (0.442) | 0.105 (0.361) | 0.069 (0.352) | 0.083 (0.430) | 0.071 (0.521) |
| ***Refinement*** | | | | | |
| *R* factor ($R_{work}$ / $R_{free}$) (%) | 17.4 / 20.0 | - | 20.1 / 24.0 | 17.7 / 22.0 | 17.4 / 21.0 |
| Model contents/average B (Å$^2$) | | | | | |
| Protein atoms | 2767 / 20.1 | - | 2760 / 16.6 | 2677 / 27.5 | 2711 / 23 |
| Water molecules | 331 / 31.5 | - | 361 / 30 | 190 / 36.2 | 302 / 37 |
| Ligand atoms | 1 / 21.3 (Chloride) | | 6 / 21.8 (Glycerol 1)<br>6 / 40 (Glycerol 2)<br>1 / 15 (Chloride) | 4 / 47 (Acetate 1)<br>4 / 60 (Acetate 2) | 6 / 38 (Malonate)<br>6 / 33 (Glycerol)<br>1 / 29 (Chloride) |
| r.m.s.d | | | | | |
| distances (Å) | 0.027 | - | 0.024 | 0.023 | 0.028 |
| bond angles (°) | 2.2 | - | 2.1 | 1.9 | 2.3 |
| Estimated coordinate error (Å) | 0.015 | - | 0.016 | 0.017 | 0.021 |
| ***Ramachandran plot*** [d] | | | | | |
| favored | 97.9 | - | 98.2 | 97.3 | 98.6 |
| allowed | 2.1 | - | 1.8 | 2.7 | 1.4 |
| outlier | 0 | - | 0 | 0 | 0 |

a *R*cullis calculated as $(\Sigma|E|\Sigma||FPH|-|FP||)$, where *F*PH is the amplitude of the protein plus the heavy atom and *F*P is the amplitude of the protein
b Phasing power calculated as r.m.s. $(|Fh|/E)$, where $|Fh|$ is the heavy atom structure-factor amplitude and *E* is the residual lack of closure error
c Statistic for the highest resolution shell are shown in parenthesis.
d As defined by PROCHECK

### 3.3.3 Overall structure of $AT_{DYN10}$

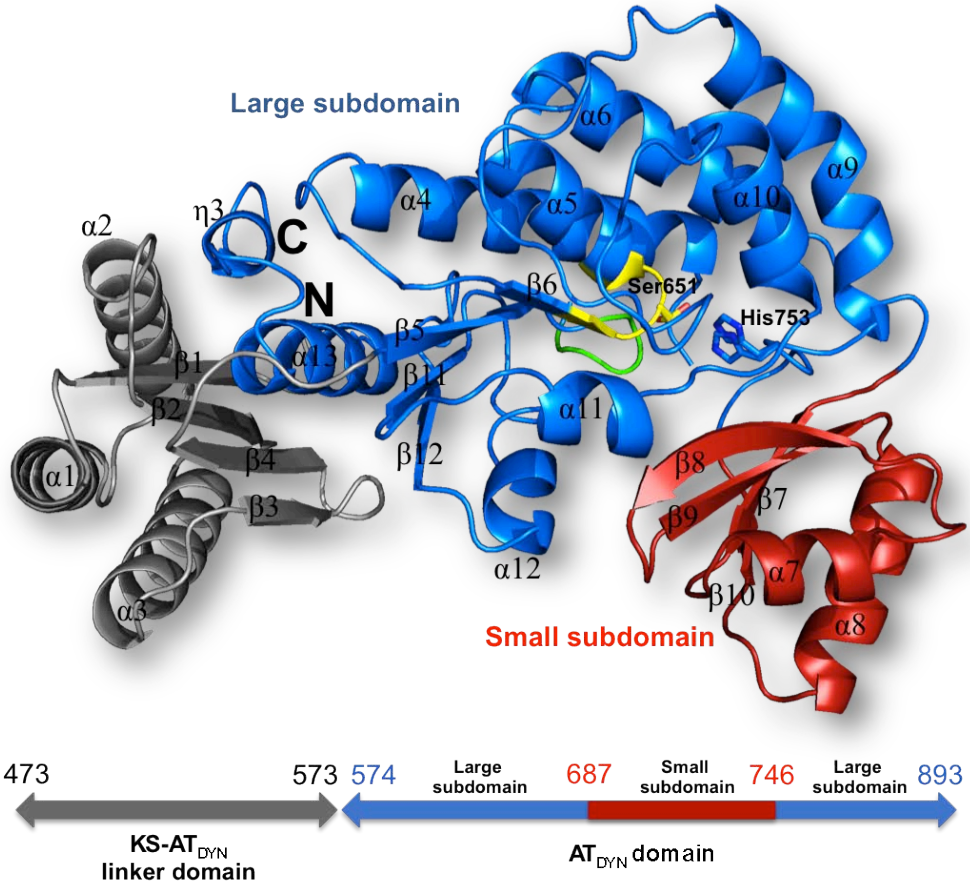



**Figure 3.3** Overall structure of $AT_{DYN10}$. $AT_{DYN10}$ is composed of a KS-$AT_{DYN}$ linker domain (*grey*) and the catalytic acyltransferase $AT_{DYN}$ domain. The catalytic $AT_{DYN}$ domain (in ribbon representation) consists of large hydrolase subdomain (*blue*) and a ferrodoxin-like smaller subdomain (*red*). The N- and C-termini and the secondary structure elements of $AT_{DYN10}$ are labeled. The $Ser^{651}$- $His^{753}$ catalytic residues are shown as *yellow* and *blue* sticks respectively. Residues from the GHS*X*G and PGQG*X*Q motifs are colored *yellow* and *green* respectively.

The $AT_{DYN10}$ protein fragment has approximate overall dimensions of 46 Å x 61 Å x 62 Å and is composed of three globular α/β domains (**Figure 3.3**). The AT catalytic domain ($AT_{DYN}$) folds into a large α/β hydrolase subdomain (residues 574 – 686 and 747 – 876) composed of a four-stranded parallel β-sheet, eight α-helices and the short left-handed C-terminal helix η3. The small subdomain of $AT_{DYN}$ (residues 687-746) adopts a ferredoxin-like fold comprising a four-stranded anti-parallel β-sheet and two α-helices (**Figure 3.3**). The helical flap, residues 595 – 611, is absent from the model due to poor electron density. However, the N-

terminal region of the helical flap adopts two markedly different orientations suggesting that the helical flap can take up different positions which may provide insights to the free enzyme and substrate enzyme complexes (**Figure 3.4**).

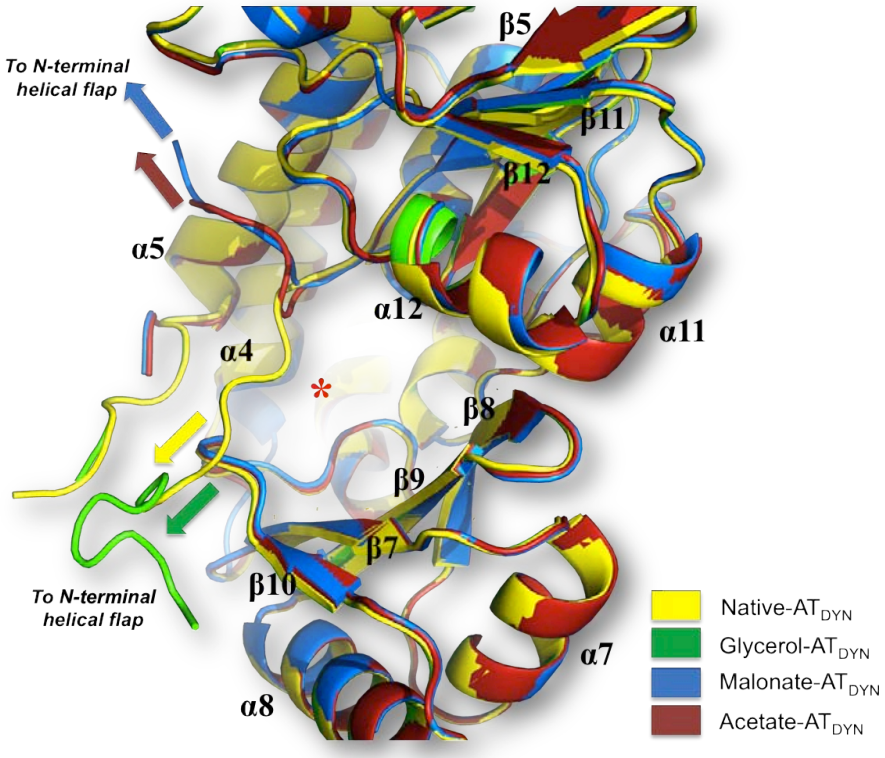



**Figure 3.4** Superposition of native $AT_{DYN}$ (*yellow*), glycerol-$AT_{DYN}$ complex (*green*), malonyl-$AT_{DYN}$ complex (*blue*) and acetyl-$AT_{DYN}$ complex (*red*). The arrows indicate the connection to the N-terminal end of the missing helical flap. Free and substrate-bound $AT_{DYN}$ exhibit different orientations of the connection. The entrance of the catalytic pocket is indicated by a *red* asterisk.

The fold of $AT_{DYN}$ is similar to several structurally characterized AT domains or individual AT proteins with an average amino acid sequence identity of ~ 30%. A structure based sequence alignment of $AT_{DYN}$ with various AT catalytic domains is shown in **Figure 3.5**. Structure comparison with AT catalytic domains from type I mammalian FAS (PDB code 2VZ9) and discrete type II AT domains such as the malonyl CoA:ACP transferases from *S. coelicolor* (PDB code 1NM2) and *E. coli* (PDB code 2G2Z) suggests that $AT_{DYN}$ bears closer similarity with type I AT domains. Noticeable differences are found in the length of the C-terminal α-helix α13 and η3 of $AT_{DYN}$ that are two turns longer.

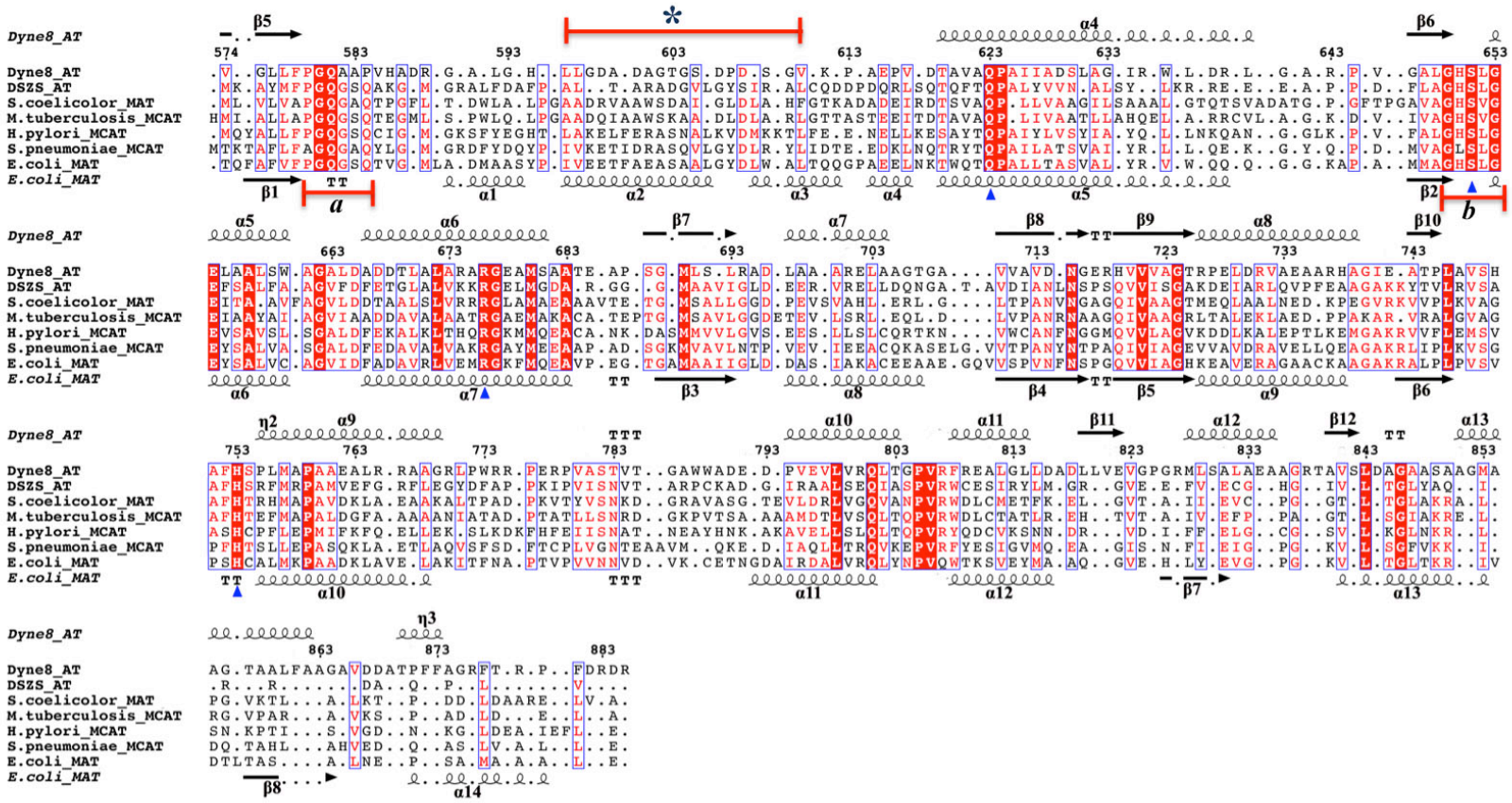



**Figure 3.5** Structure-based sequences alignment of structurally characterized ATs. $AT_{DYN}$ (*M. chersina*), DSZS AT (PDB code 3RGI), *Sc*MCAT (PDB code 1NM2), *Mt*MCAT (PDB code 2QC3), *Hp*MCAT (PDB code 2H1Y), *Sp*MCAT (*S.pneumoniae*), and *Ec*MCAT (PDB code 2G2Z) are malonyl-specific enzymes. The secondary structure for $AT_{DYN}$ is displayed on the *top* of the alignment and for *Ec*MCAT at the *bottom*. Strictly conserved residues are in *red* boxes and similar residues colored *red*. Catalytic residues are indicated by *blue* triangles. The PGQG*X*Q and GHS*X*G motif are labeled as "*a"* and "*b"*, respectively. The asterisk indicates the polypeptide stretch that is missing in the $AT_{DYN10}$ structure and that forms a helical flap in other AT structures. The alignment was made with ESPript[114].

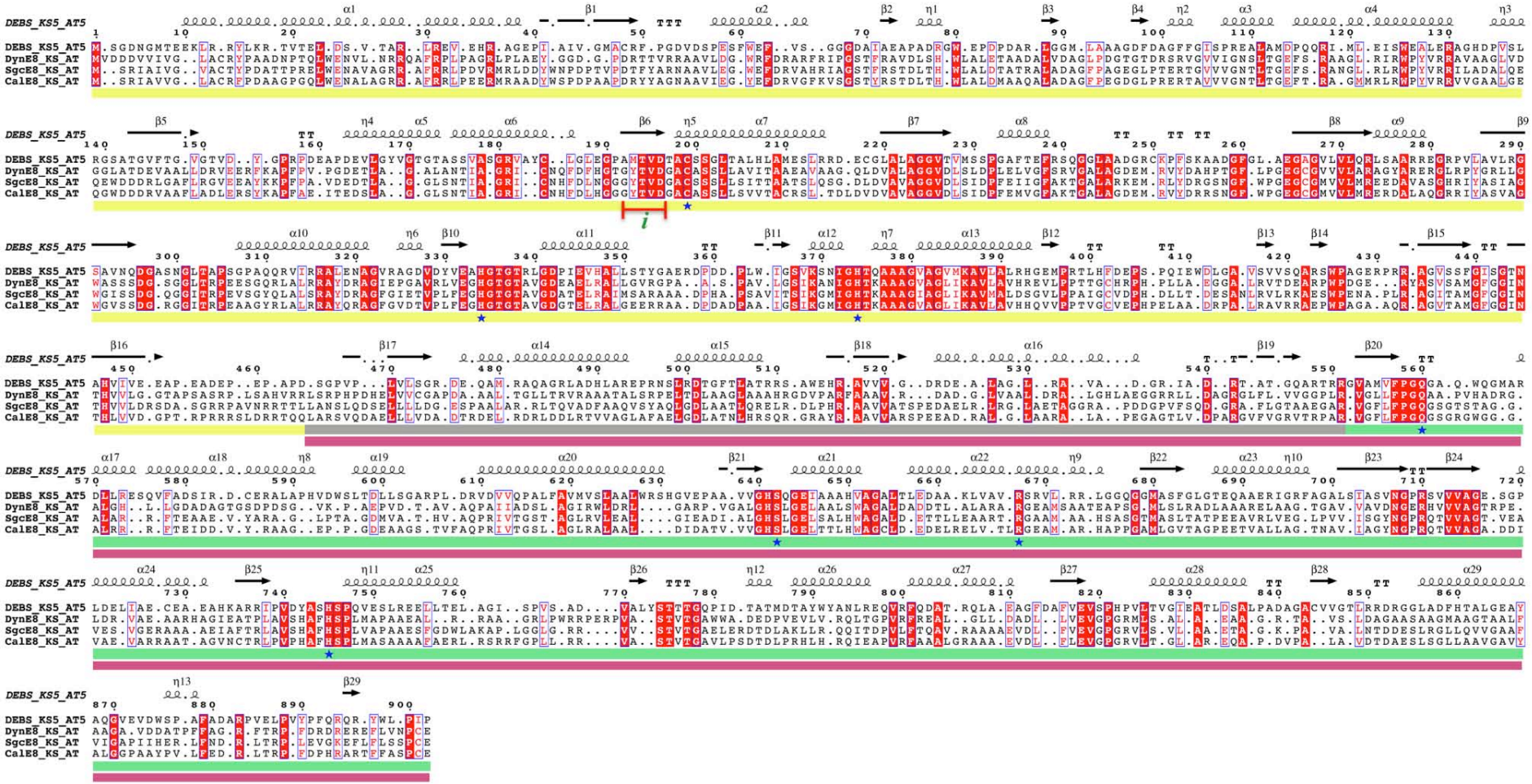



**Figure 3.6** Structure-based sequence alignment of [KS5][AT5] DEBS module 5 and the corresponding domains of enediyne PKSs DynE8, SgcE8 and CalE8. Strictly conserved residues are in *red* boxes and similar residues colored *red*. Catalytic residues are marked by *blue* asterisks. The evolutionary conserved dimerization region from the KS domain of [KS5][AT5] from DEBS is labeled *i*. The colored bars below sequences indicate the boundaries of the DEBS and $AT_{DYN10}$ proteins.

An automatic search for structurally homologous proteins using the Dali server[115] shows that $AT_{DYN10}$ shares closely related tertiary structures with type I modular PKS deoxyerythronolide B synthase (DEBS) module 5 (PDB code 2HG4, Z-score of 35.2), despite the fact that $AT_{DYN10}$ and the corresponding region of DEBS modules 3 or 5 only share ~25% sequence identity (**Figure 3.6**). The superimposition of $AT_{DYN10}$ with the corresponding domains of DEBS module 5 is displayed in **Figure 3.7**. This superimposition yields a r.m.s. deviation of 3.7 Å for 347 equivalent residues. In the complete DynE8 protein, the C6 terminal residue of $AT_{DYN10}$ (position 876) makes the connection to the N-terminal residue of the $ACP_{DYN}$ domain (position 925) via a linker region that was absent in the present protein fragment. The corresponding linker region adopts an extended conformation in the DEBS protein[110].

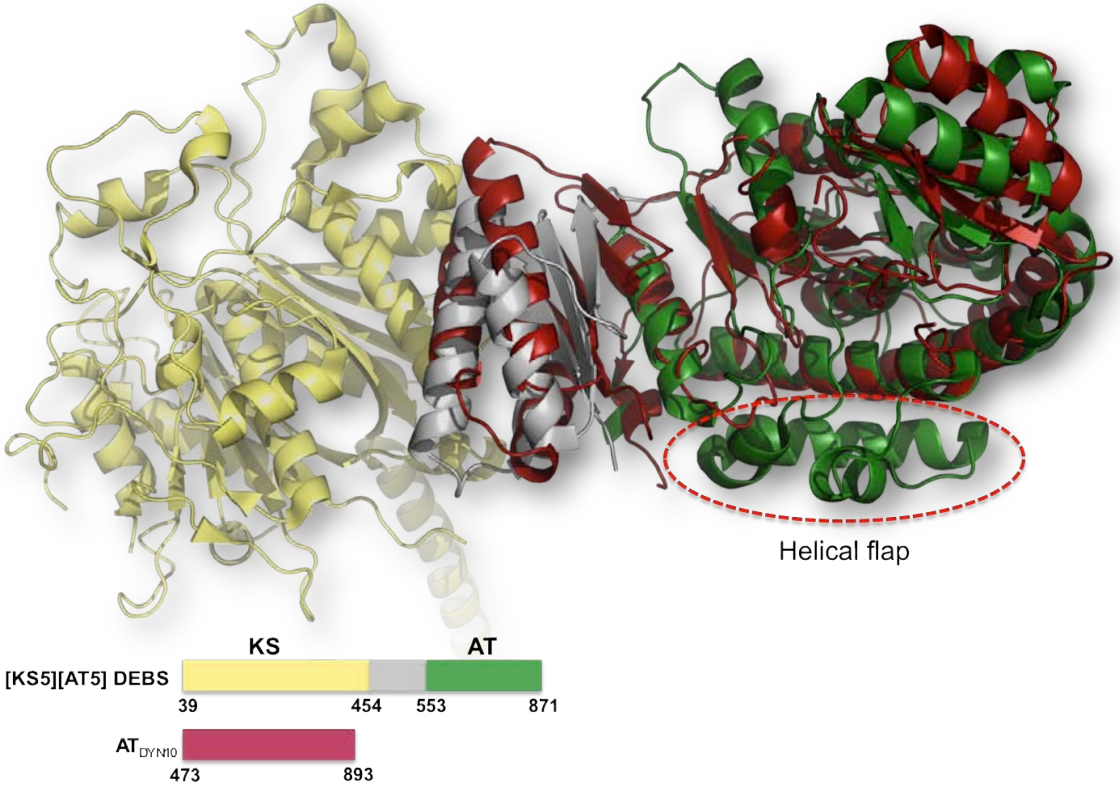



**Figure 3.7** Structure superposition of $AT_{DYN10}$ and [KS5][AT5] (PDB code 2HG4). The helical flap of the [AT5] domain, which is absent in $AT_{DYN10}$, is indicated with a dotted *red* circle.

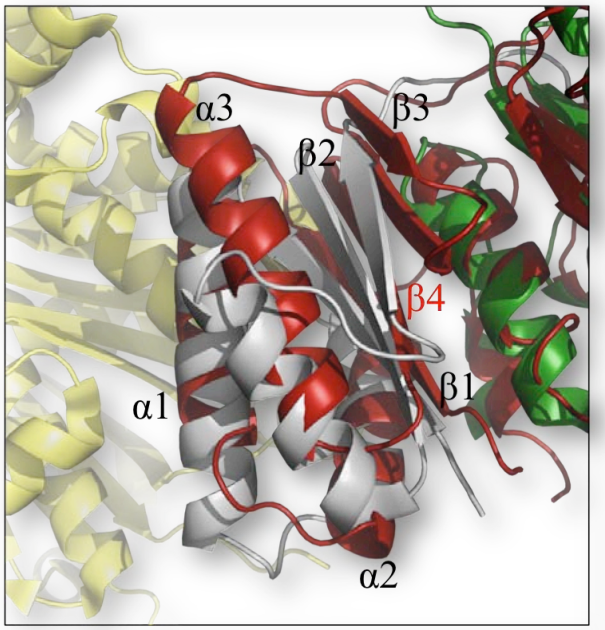



**Figure 3.8** Superposition of the linker domains from [KS5][AT5] and $AT_{DYN10}$. The secondary structure elements of the $AT_{DYN10}$ linker domain are labeled. The extra strand β4 (compared to [KS5][AT5], see text for details) is labeled in *red*.

In addition to the AT catalytic domain, the fragment of DynE8 that was crystallized contains an extra 10 kDa domain (residues 473 – 573) that connects the adjacent $KS_{DYN}$ domain to the N-terminal end of the AT catalytic domain and that is crucial for solubility (**Figure 3.2 and 3.3**). The linker domain also adopts a α/β fold that abuts on the large $AT_{DYN}$ subdomain with a buried surface area of 2436 $Å^2$ at the interface. The linker domain comprises four anti-parallel β strands and three α-helices running roughly parallel to each other. A search for similar protein structures also reveals closest homology with the corresponding [KS5][AT5] linker domain in the DEBS protein[91] and a magnified view of their superimposition is displayed in **Figure 3.8**. Hence, the overall structure of the KS-AT linker domain of modular PKSs is conserved in iterative PKSs. Nevertheless, compared to the [KS5][AT5] linker domain, KS-$AT_{DYN}$ possesses the extra β-strand β4 at its C-terminus that makes the connection to the AT domain (**Figure 3.3** and **3.8**). Compared to the [KS5][AT5] linker domain of the DEBS protein, KS-$AT_{DYN}$ has longer helices (spanning 14 – 17 residues) and shorter strands (4 – 5 residues). Interestingly, in the case of the DEBS protein, the linker regions were proposed to confer rigidity to the protein and to serve as anchoring sites to ACP[110]. Given the

overall structural homology, helices α1 – α3 helix of KS-$AT_{DYN}$ may also interact with the neighboring KS domain in DynE8.

### 3.3.4 Active site of $AT_{DYN}$ and amino acid sequence motifs

In agreement with previously characterized AT structures including the malonyl CoA:ACP acyltransferases from *S. coelicolor* and *E. coli*, and the AT domains of mammalian FAS and 6-DEBS, the active site of $AT_{DYN}$ is located in a deep cleft between its large and small subdomains (**Figure 3.3)**. Inspection of this cleft reveals the presence of several evolutionary conserved amino-acid sequence motifs required for catalysis and substrate specificity. The essential $Ser^{651}$ residue from the highly conserved GHS*X*G motif (GHSLG in $AT_{DYN}$ "motif *b*" in **Figure 3.5)** is located in a turn between strand β6 and helix α5 (**Figure 3.3)**. Two histidine residues: $His^{650}$ and $His^{753}$ can be seen in the $AT_{DYN}$ active site (**Figure 3.9)**.

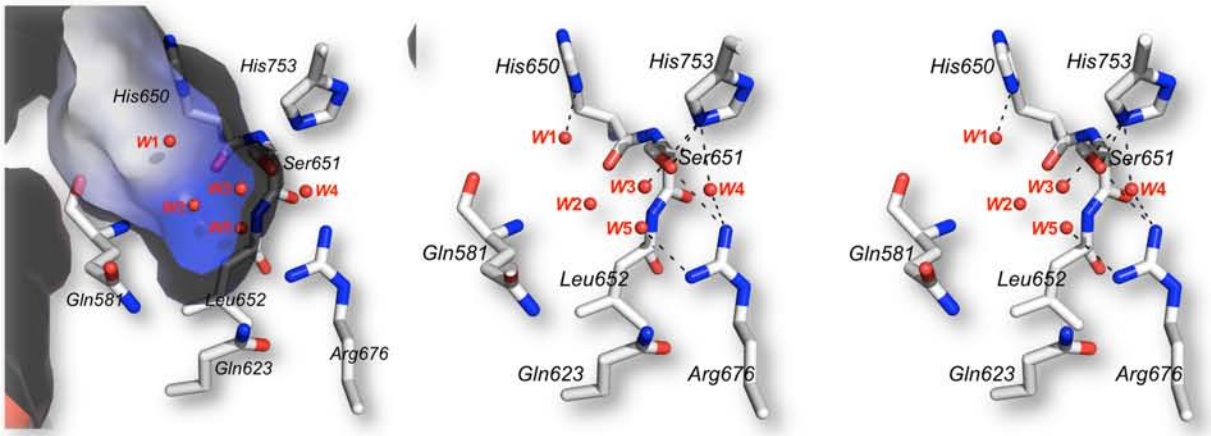



**Figure 3.9** The catalytic pocket and active site of the unbound $AT_{DYN}$ domain. The *left* panel shows cross-sections through the electrostatic potential surface of the catalytic pocket (Positive potential, *blue*; negative potential, *red* at the 10 $kTe^{-1}$ level). The *right* panel (stereo view) shows the active site of the unbound $AT_{DYN}$ domain. Water molecules (red spheres) are labeled "*W"*. The views are in the same orientation. The stereo view was generated using program PYMOL[104]

In other characterized AT domains from FAS or PKS, the residue corresponding to $His^{753}$ was suggested to act as a general base/acid catalyst in the acyl transfer reaction. In $AT_{DYN10}$, the catalytic role played by $His^{753}$ is supported by the short distance of 2.7 Å between its $N^{\varepsilon 2}$ atom and the $O^{\gamma}$ atom of $Ser^{651}$. In

contrast, the $N^{\varepsilon 2}$ atom of $His^{650}$ is located at a distance of 5 Å from the $O^{\gamma}$ atom of $Ser^{651}$, excluding a direct role in catalysis for $His^{650}$. Spatial conservation of the catalytic Ser-His dyad suggests that $AT_{DYN}$ uses a catalytic mechanism similar to other FAS and PKS AT domains, where transfer of the malonyl group from malonyl-CoA onto $Ser^{651}$ occurs via a *pingpong* kinetic *bi bi* mechanism, followed by the transfer of the malonate group to the shuttling ACP domain[116; 117](**Figure 3.1)**. In the malonyl acetyl transferase (MAT) FabD, the backbone amide groups of $Gln^{11}$ and $Leu^{93}$ form the oxyanion hole stabilizing the transition state. Previous work showed that the conformation of the PG**Q**$^{11}$GXQ motif has a direct influence on the integrity of the oxyanion hole[117; 118; 119]. In $AT_{DYN}$, motif PG**Q**GXQ is replaced by PG**Q**$^{581}$AAP (labeled "motif *a*" in **Figure 3.5** and **Figure 3.3**) that contains the conserved Gln581, and $Leu^{652}$, the corresponding residue $Leu^{652}$ from "motif *b*" is also conserved. Moreover, it was proposed that the **Z**T**XX**[AT][QE] motif (where **Z** is a hydrophilic residue and **X** is hydrophobic) located 30 residues upstream of the essential Ser determines the enzyme specificity for malonyl-CoA[120]. In $AT_{DYN}$, the corresponding motif (DTAVAQ$^{623}$) is present with $Gln^{623}$ hydrogen-bonded to the highly conserved $Arg^{676}$ residue. Finally, a 'HAFH$^{753}$, motif is also found in $AT_{DYN}$ (**Figure 3.5**), which further supports the correlation between malonyl-CoA specificity and the presence of the [HTVY]AFH motif[120].

### 3.3.5 Enzyme-malonate covalent complex

$AT_{DYN}$ catalyzes the transfer of the malonate group from malonyl-CoA onto $ACP_{DYN}$. According to studies on the AT domains of FAS and modular PKSs, catalysis is likely to proceed through a covalent intermediate with the malonate group attached to $Ser^{651}$. By co-crystallizing $AT_{DYN10}$ with malonyl-CoA, the malonyl group was transferred onto the AT domain yielding the structure of the enzyme malonate covalent complex (**Figure 3.10**). The crystal structure of this complex was refined to *R*work and *R*free values of 17.4% and 21.0%, respectively (**Table 1**). After refinement, the distance between the Cβ atom of the $Ser^{651}$ side chain and the oxygen of the malonate moiety is 1.4 Å, suggesting formation of a covalent complex that represents the acyl-enzyme intermediate of the transacylation reaction. A bidendate salt bridge is formed between the carboxylate of the malonate

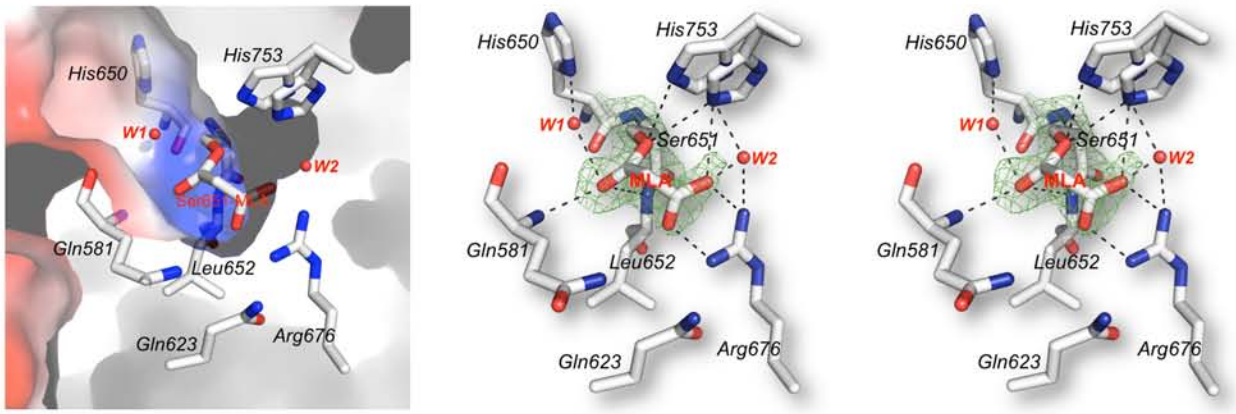



**Figure 3.10** The catalytic pocket and active site of the malonate-enzyme complex. The left panel shows cross-section of the catalytic pocket of the malonate-enzyme complex. Positive potential, *blue*; negative potential, *red* at the 10 $kTe^{-1}$ level. The *F*o – *F*c difference electron density maps of the malonate-enzyme complex are countered at level of 3.0 σ, where the ligand has been omitted from phase calculation (*right panel*). The hydrogen-bond network stabilizing the ligand is indicated as dotted lines. Water molecules (*red* spheres) are labeled "*W"*. The views are in the same orientation. The stereo view was generated using program PYMOL[104]

and the guanidinium group of $Arg^{676}$ (**Figure 3.10**). The oxygen atoms of the malonate moiety are anchored through a water-mediated hydrogen bond with the $N^{\delta 1}$ atoms of the surrounding $His^{650}$ and $His^{753}$ and with the amide nitrogen of $Gln^{581}$ (**Figure 3.10**). The structure of the enzyme-malonate covalent complex also reveals the presence of a Phe residue in the active site that is likely to play a key role in discriminating against methylmalonyl-CoA as a substrate. The bulky side chain of $Phe^{752}$ is in direct contact with the methylene carbon of the covalently tethered malonate (**Figure 3.11**). The close distance excludes the possibility to accommodate methylmalonate in a similar position in the active site.

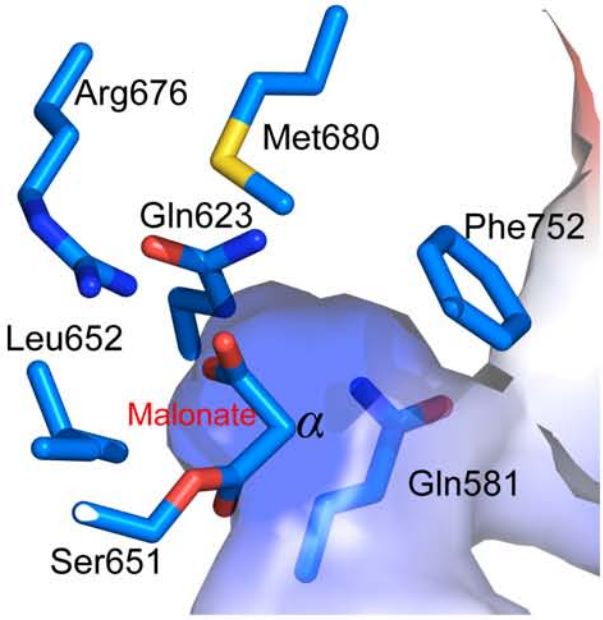



**Figure 3.11** Cross-section of the $AT_{DYN}$ catalytic pocket. The catalytic pocket of $AT_{DYN}$ is occupied by malonate, covalently linked to $Ser^{651}$. The electrostatic potential surface represents positive potential in *blue*, negative potential in *red* at the 10 kT $e^{-1}$ level.

Comparison of the free enzyme and malonate-enzyme structures (**Figure 3.9-3.10** and **Figure 3.12**) reveals several subtle conformational changes in the active site upon formation of the $Ser^{652}$-malonate linkage. Following malonate attachment, the rotamer conformation of $Ser^{652}$ changes (**Figure 3.12**). In the free enzyme structure, the $N^{\delta 1}$ atom of $His^{753}$ and the $O^{\gamma}$ atom of $Ser^{651}$ are hydrogen bonded. This is in agreement with the role of $His^{753}$ as a general base catalyst to activate $Ser^{651}$ for nucleophilic attack of the thioester carbonyl of malonyl-CoA. In the enzyme-malonate complex structure, the imidazole ring of $His^{753}$ is reoriented and the surrounding hydrogen bond network is altered. The active site pocket also expands slightly following the reorientation of the $His^{753}$ and $Leu^{802}$. The transfer of the malonate moiety from malonyl-CoA to the enzyme and its subsequent transfer from the enzyme to the ACP domain are believed to proceed through tetrahedral transition states and intermediates. An oxyanion hole formed by two main chain amide groups was proposed to play an essential role in stabilizing these transition states and intermediates[118]. The hydrogen bond network formed between malonate

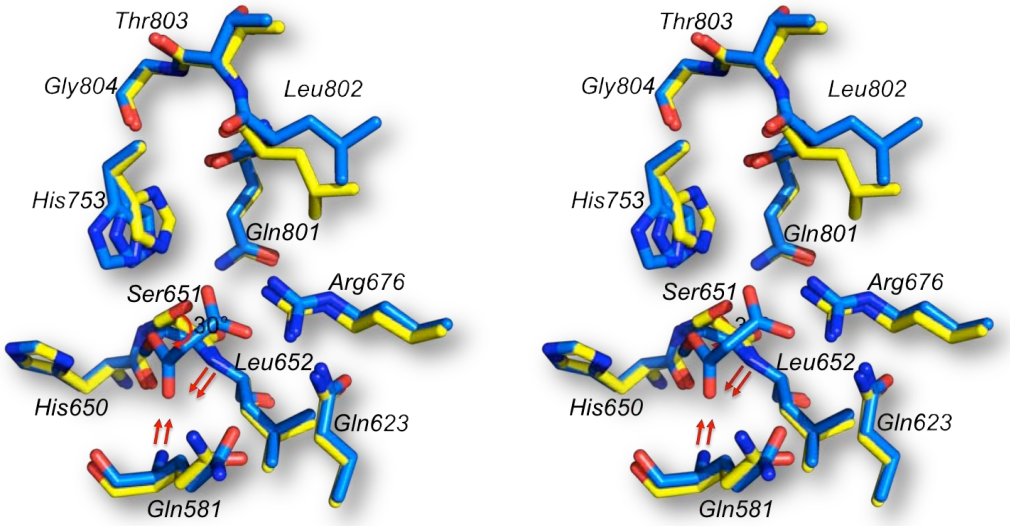



**Figure 3.12** Stereo-view of a superposition of the active site for the free enzyme (*yellow* sticks) and malonate-bound $AT_{DYN}$ (*blue* sticks). Residues $His^{753}$ and $Ser^{651}$ form the catalytic dyad. Compared to their positions in the unliganded structure, the backbone amides of residues $Gln^{581}$ and $Leu^{652}$ are shifted by about 3 Å to shape the oxyanion hole (the movement is indicated by *red* arrows). Note that the malonate-bound $Ser^{651}$ rotates downwards by about 30° (indicated by *red* arrows).

and nearby residues and structural comparison with FabD (PDB: 1MLA) points to the amide groups of $Leu^{652}$ and $Gln^{581}$ to constitute the oxyanion hole in $AT_{DYN}$. In the malonyl-enzyme structure, the $Gln^{581}$ shifts towards the $Leu^{652}$ amide and the distance between the two amides is decreased from a value of 5.5 Å (unliganded structure) to 5.2 Å (malonyl complex) (**Figure 3.12**).

### 3.3.6 Enzyme-acetate complex

We previously showed that DynE8 and other enediyne PKSs do not use acetyl-CoA to generate the acetate starter unit[86; 90]. Instead, the starter acetate unit of the enediyne PKS products is derived from malonyl-CoA through an intrinsic decarboxylation mechanism. This is in contrast to the AT domain of the mammalian FAS that exhibits dual specificity for malonyl- and acetyl- CoA. It was hypothesized that the inability for enediyne PKSs to utilize acetyl-CoA derives from their AT domains that bind acetyl-CoA in an unproductive configuration that

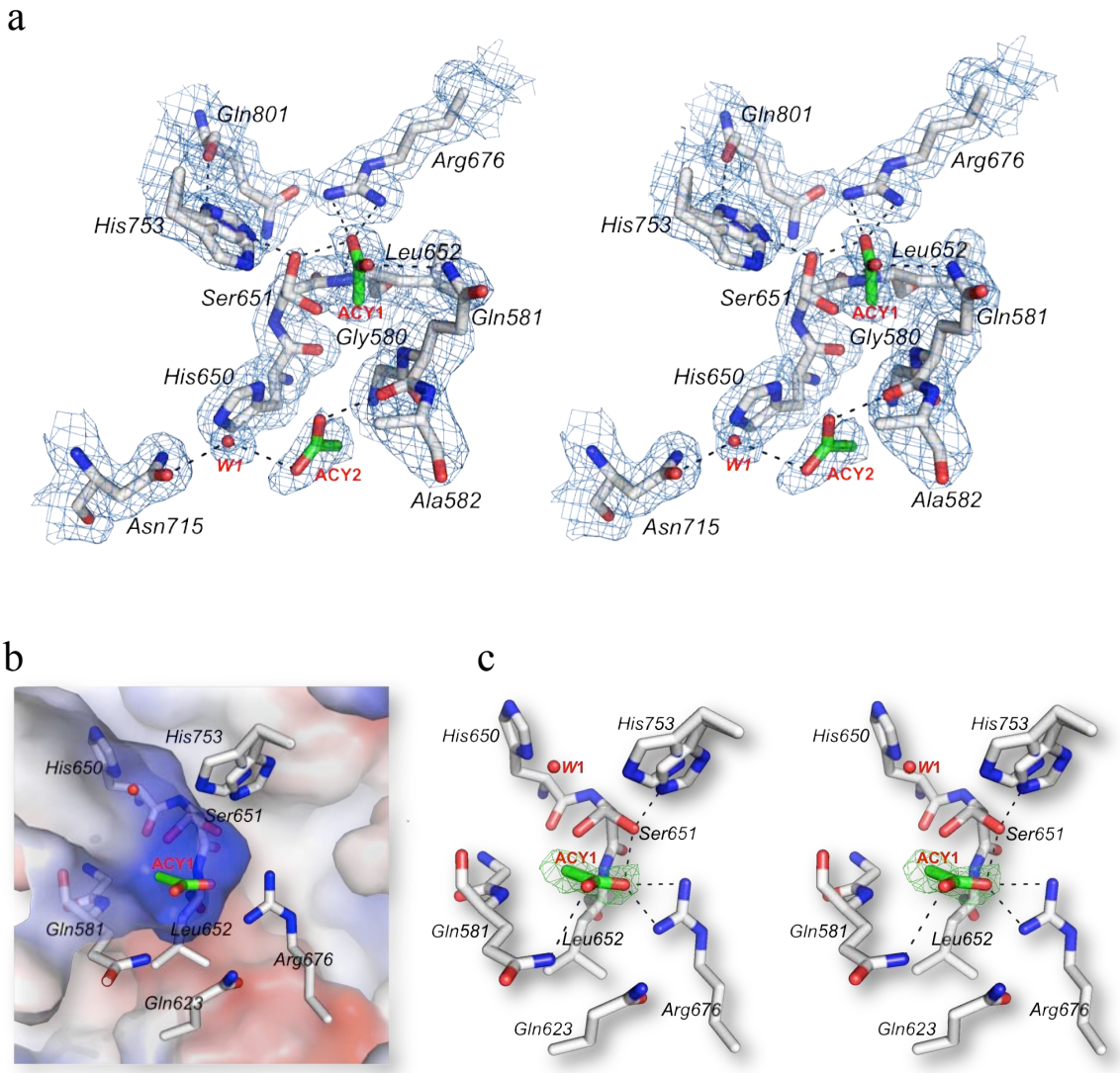



**Figure 3.13** The catalytic pocket and active site of acetate-enzyme complex. (**a**) Stereogram of the active site of the binary complex of $AT_{DYN}$ with acetate. The ACY1 and ACY2 with temperature factors of 47 $Å^2$ and 60 $Å^2$ respectively, were found uncovalently bound to the $AT_{DYN}$. (**b**) The cross-section of the catalytic pocket of the acetate-enzyme complex. Positive potential, *blue*; negative potential, *red* at the 10 $kTe^{-1}$ level. (**c**) The *F*o – *F*c difference electron density maps of the acetate-enzyme complex are countered at level of 3.0 σ, where the ligand has been omitted from phase calculation (*right panel*). The hydrogen-bond network stabilizing the ligand is indicated as dotted lines. Water molecules (*red* spheres) are labeled "*W*". The views are in the same orientation. The stereo views were generated using program PYMOL[104].

prohibits acyl transfer. To understand the molecular basis for substrate discrimination, we co-crystallized $AT_{DYN10}$ with acetyl-CoA (**Figure 3.13**). Intriguingly, co-crystallization of $AT_{DYN10}$ with acetyl-CoA yielded a non-covalent enzyme-acetate complex (1.9 Å resolution). The electron density observed near $Ser^{651}$ rules out the formation of an acetyl-enzyme complex.

Instead, two non-covalently bound acetates ACY1 and ACY2 are observed with average temperature factors of 47 $Å^2$ and 60 $Å^2$ respectively, whilst the average temperature factor for the protein is 28 $Å^2$ (**Figure 3.13a**). ACY1 is bound in the active site, with its carboxylic end anchored by two hydrogen bonds with $Arg^{676}$ and the $N^{\varepsilon 2}$ atom of the side chain of $Gln^{581}$ (**Figure 3.13b-c**). A weak hydrogen bond (bond angle 97.4 ° and bond length = 3.7 Å) is also formed with the backbone amide group of $Leu^{652}$. This mode of interaction, where the acetate moiety mimics interactions formed by the carboxylic end of the malonyl group is similar to the *Sc*MAT-acetate complex[117; 121]. However, the orientation of ACY1 in $AT_{DYN}$ differs slightly from the one adopted in *Sc*MAT since ACY1 is positioned parallel to $Ser^{651}$ and almost perpendicular to the $Arg^{676}$ guanidino plane. Upon ACY1 binding, an alternate conformation of $Ser^{651}$ side chain is observed with a change in *chi* angle of 104°. Based on the electron density map, however, we can not rule out an alternative binding mode for ACY1 with its carboxylate group hydrogen bonded to the $O^{\gamma}$ atom of $Ser^{651}$ and backbone amide group of $Gln^{581}$. The second acetate ACY2 is located in the immediate vicinity of the AT active site in a pocket formed by $His^{650}$, $Asn^{715}$, $Gly^{580}$, $Gln^{581}$ and $Ala^{582}$ (**Figure 3.13a**). ACY2 is stacked against the imidazole ring of $His^{650}$ and is stabilized through a weak hydrogen bond (bond angle 124.6 ° and bond length = 3.2 Å) with the carbonyl oxygen of $Gln^{581}$ and a water-mediated hydrogen bond with $Asn^{715}$. ACY2 is found in a position equivalent to the second acetate bound to *Sc*MAT, demonstrating the similarity between the two active sites.

### 3.3.7 Enzyme-glycerol complex

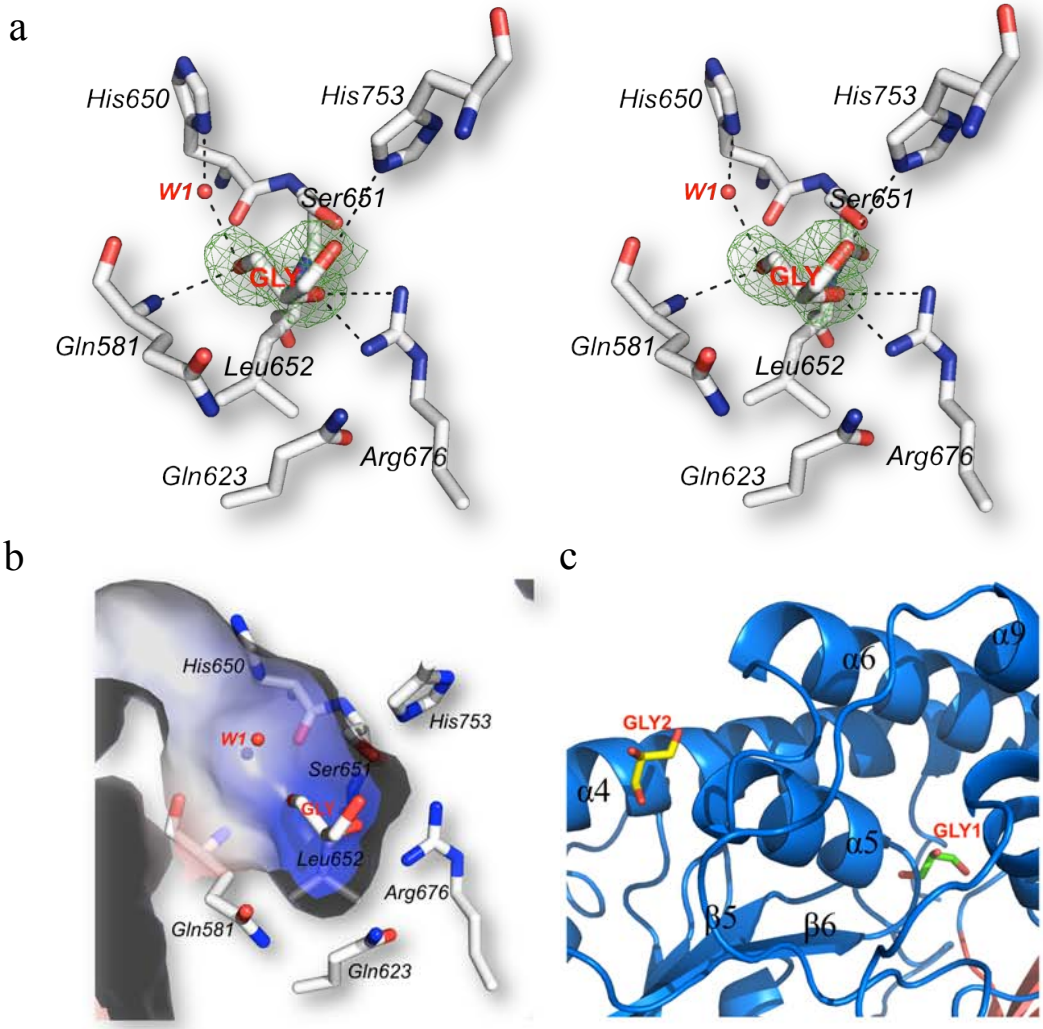



**Figure 3.14** The catalytic pocket and active site of glycerol-enzyme complex. (**a**) The *F*o – *F*c difference electron density maps of the glycerol-enzyme complex are countered at level of 3.0 σ, where the ligand has been omitted from phase calculation. (**b**) The cross-section of the catalytic pocket of the glycerol-enzyme complex. (**c**) The binding site of the second glycerol molecule (GLY2) in glycerol-$AT_{DYN}$ complex. Glycerol 1 (GLY1) and glycerol 2 (GLY2) are represented by *green* and *yellow* stick

A native $AT_{DYN10}$ crystal, soaked with 25% glycerol as cryoprotectant, clearly showed two bound glycerol molecules. One glycerol molecule, with a temperature factor of 21.8 $Å^2$, is located in the active site (**Figure 3.14a-b**). Its oxygen atoms make hydrogen bonds with the side-chains of $Ser^{651}$, $Arg^{676}$ and with $His^{650}$ via a bridging water molecule. Comparison with the malonyl-enzyme

covalent complex reveals that one hydroxyl group overlaps with the oxygen atom of the thioester carbonyl group. The hydroxyl group also forms hydrogen bonds with the main chain amide groups of $Leu^{652}$ and $Gln^{581}$ that form the oxyanion hole. A second glycerol molecule, with a temperature factor of 40.0 $Å^2$, is found at the surface of the large $AT_{DYN}$ subdomain next to loop α4 – α6 (**Figure 3.14c**).

### 3.3.8 $AT_{DYN10}$ is able to hydrolyze the thioester bond of acetyl-CoA

The surprising observation of acetate ligands in the $AT_{DYN10}$ crystal structure raised the question of how they were generated from acetyl-CoA during the crystallization process. To distinguish between an enzyme or a buffer catalyzed hydrolysis mechanism, we performed enzymatic assays with acetyl-CoA. This experiment showed that $AT_{DYN}$ is able to hydrolyze the thioester bond of acetyl-CoA to produce CoA and acetate (**Figure 3.15a**), whereas, under the same conditions, acetyl-CoA remains intact in the absence of enzyme. This shows that the acetate moieties seen in the crystal structure are generated by enzyme-catalyzed hydrolysis of acetyl-CoA.

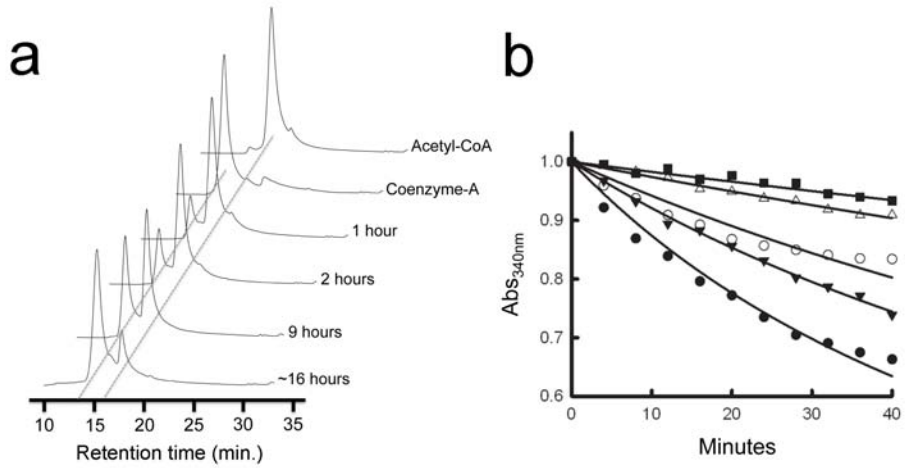



**Figure 3.15** The catalytic activity of $AT_{DYN10}$. (**a**) HPLC analysis of $AT_{DYN}$ catalyzed hydrolysis of acetyl-CoA. The HPLC chromatograms show the depletion of the acetyl-CoA substrate along with the increment of the CoA by-product. (**b**) Substrate-competition inhibition assay. The absorption spectra of NADPH are monitored using UV-Visible spectrophotometry at an absorbance of 340 nm. The positive control (*black* circles), with only 2.5 mM malonyl-CoA as substrate, shows depletion of NADPH, while the reaction with the addition of 1.0 mM (*black* triangles), 2.0 mM (*white* circles), 2.5 mM (*white* triangles), and 5.0 mM (*black* square) acetyl-CoA shows slower depletion of NADPH with increasing concentrations of acetyl-CoA.

Further support for the idea that $AT_{DYN}$ can accept acetyl-CoA as substrate comes from competitive inhibition experiments using the full-length DynE8 protein. We found that addition of acetyl-CoA to the reaction mixture can significantly slow down the PKS reaction in a concentration-dependent fashion (**Figure 3.15b**), indicating that acetyl-CoA is competing with malonyl-CoA for the substrate binding pocket. In contrast to the previous conclusions, this mechanism indicates that $AT_{DYN}$ can also use acetyl-CoA as substrate by transferring the acetyl moiety onto $Ser^{651}$. However, unlike the malonyl moiety that stays attached to $Ser^{651}$ and is later transferred to the ACP domain, the acetyl moiety is likely to be removed from the protein by hydrolysis.

### 3.3.9 ACP docking site of $AT_{DYN10}$

In fatty acid and polyketide synthesis, specific protein interaction between the ACP and the AT domains are critical for the efficient transfer of acyl groups[117; 122]. Considerable efforts have been devoted to identify the structural factors influencing the interaction between ACP and AT in FAS and modular PKSs[117; 122; 123], while little information is available for iterative PKSs. We found that the weak and transient interaction between $AT_{DYN10}$ and $ACP_{DYN}$ precluded crystallization of the complex. As an alternative strategy, we performed protein-protein docking simulations using the crystal structure of $AT_{DYN10}$ and a model for $ACP_{DYN}$ generated using the highly similar *me*ACP (15) domain structure (PDB code 2L9F) as template and adding the phosphopantetheine group to $Ser^{965}$. The resulting model resembles the [KS3][AT3]-ACP3 complex from DEBS (30). The ACP tethered phosphopantetheine extends into the AT active site (**Figure 3.16a)**. The distance between the active site $Ser^{651}$ of the $AT_{DYN}$ domain and the phosphopantetheine attachment site on $ACP_{DYN}$ is approximately 18 Å, consistent with the length of a fully extended phosphopantetheine arm (**Figure 3.16a**). The $ACP_{DYN}$ protein makes contacts with the KS-$AT_{DYN}$ linker domain and the larger $AT_{DYN}$ subdomain. Interestingly, this mode of interaction differs from the predicted *Hp*MCAT-*Hp*ACP[122] and FabH-ACP[124] complexes, where the AT-ACP interactions mainly involve the larger and smaller AT subdomains. The predicted interface includes an area ("area A", see

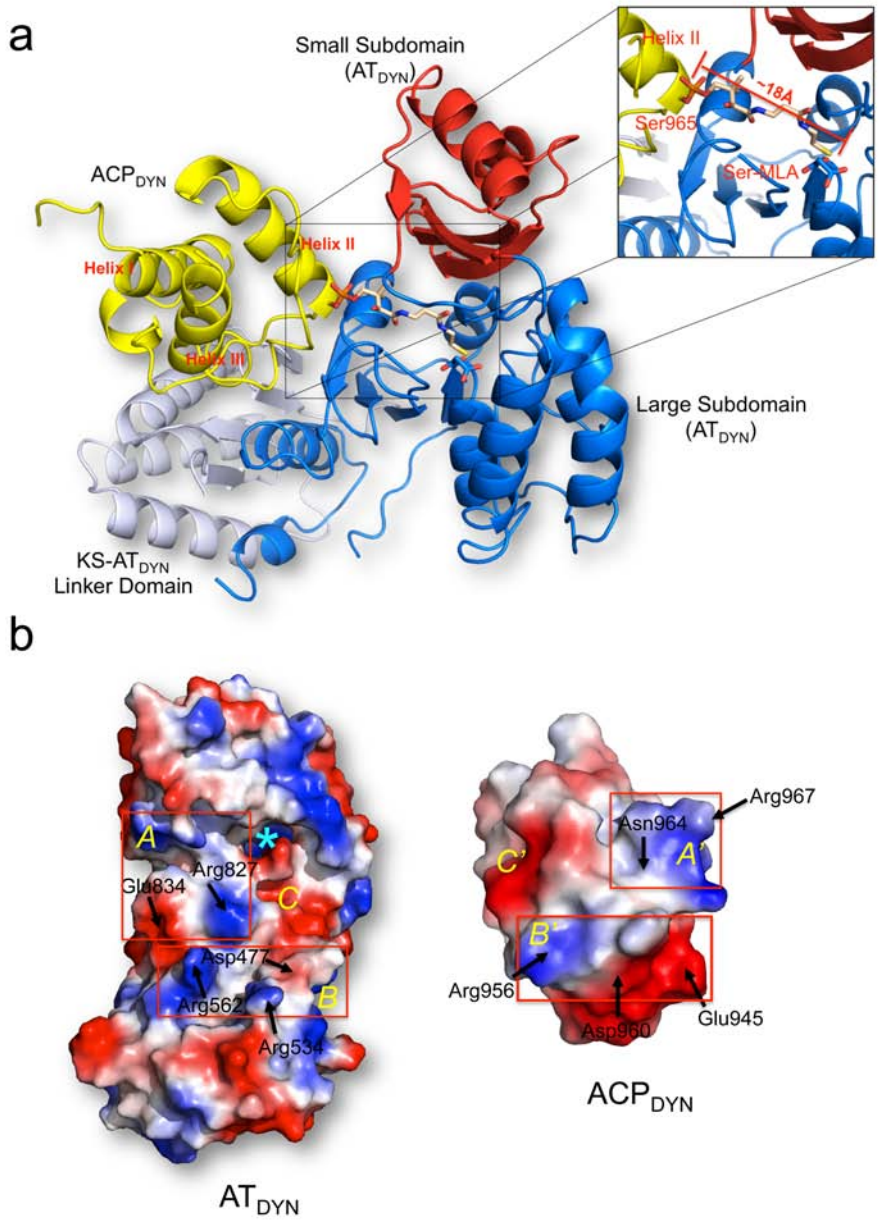



**Figure 3.16** Docking model of $AT_{DYN10}$ and $ACP_{DYN}$. (**a**) The $ACP_{DYN}$ homology model (colored in *yellow*), generated by I-TASSER[105; 106], was used as a ligand protein to dock with the unliganded $AT_{DYN10}$ structure. The phosphopantetheine arm and malonate were modeled into the catalytic pocket of $AT_{DYN}$. Distance between the $Ser^{651}$ of the $AT_{DYN}$ and the $Ser^{965}$ phosphopantetheine attachement site on $ACP_{DYN}$ is approximately 18 Å. (**b**) Electrostatic surface maps of $ACP_{DYN}$ and $AT_{DYN10}$ docking interfaces are generated with Pymol[104]. Colors range from *blue* (positive) to *white* to *red* (negative). Three main electrostatic surface contact areas with the interacting residues are labeled. A blue asterisk indicates the entrance of the catalytic pocket. Docked $ACP_{DYN}$ is rotated 180° from $AT_{DYN10}$–$ACP_{DYN}$ interface such that annotations of A – B on $AT_{DYN10}$ and $ACP_{DYN}$ interfaces should match up.

**Figure 3.16b**) from the larger subdomain located at the entrance of the substrate tunnel leading to the active site serine. Electrostatic interactions between the KS-AT linker domain ("area B") and loops I and II of $ACP_{DYN}$ are probably also important for the interaction. Residues $Asp^{477}$, $Arg^{534}$ and $Arg^{562}$ ("area B") of KS-$AT_{DYN}$ are predicted to form salt bridges with $Arg^{956}$, $Asp^{960}$ and $Glu^{945}$ on the surface of $ACP_{DYN}$ (**Figure 3.16b**, **Figure 3.17; Table 3.2**). Finally, interactions between the helical flap of the larger subdomain of $AT_{DYN}$ ("area C") with $ACP_{DYN}$ could also play a role for orienting the ACP domain.

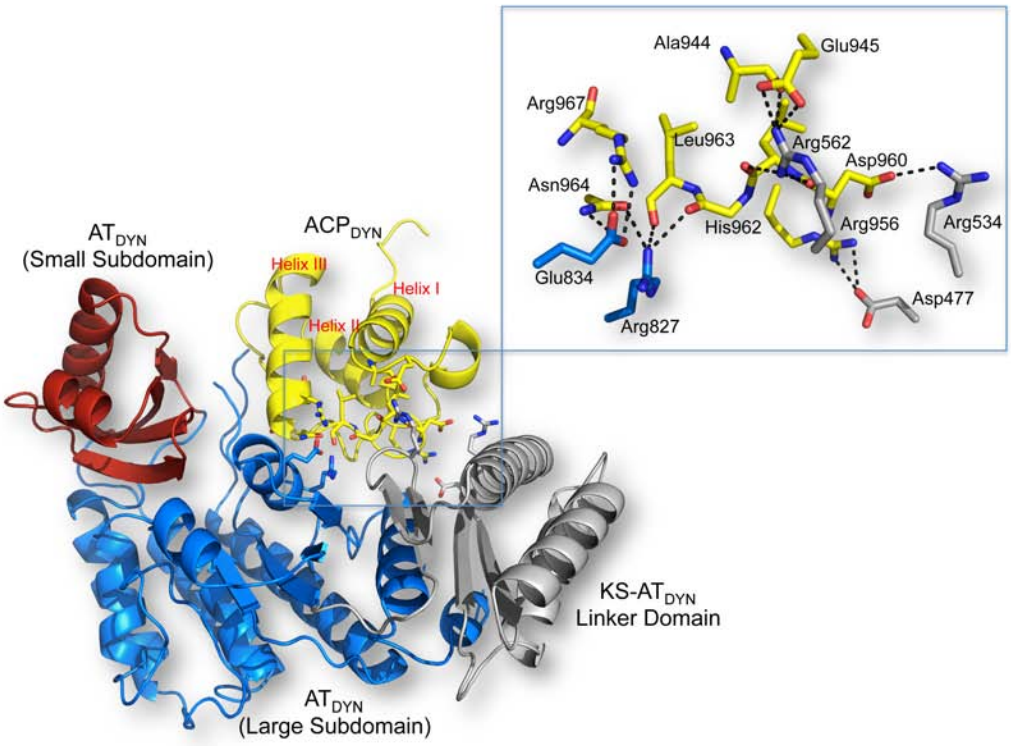



**Figure 3.17** Docking model of the complex formed between $AT_{DYN}$ and $ACP_{DYN}$. The interaction between $AT_{DYN}$ and $ACP_{DYN}$ primarily involves $ACP_{DYN}$ with the large subdomain of $AT_{DYN}$ and the KS-$AT_{DYN}$ linker domain. Residues predicted to contribute to hydrogen bond interactions for $AT_{DYN}$/$ACP_{DYN}$ complex formation are labeled (inset).

**Table 3.2** Hydrogen-bond interactions in the $AT_{DYN}$/$ACP_{DYN}$ docking model

| Interaction amino acid residues | | | | |
|---|---|---|---|---|
| $AT_{DYN10}$ | | $ACP_{DYN}$ | | Distance (Å) |
| Asp477 | $O^{\delta 2}$ | Arg956 | $N^{\eta 1}$ | 3.1 |
| Asp477 | $O^{\delta 2}$ | Arg956 | $N^{\eta 2}$ | 2.6 |
| Arg534 | $N^{\eta 2}$ | Asp960 | $O^{\delta 1}$ | 2.8 |
| Arg562 | $N^{\eta 1}$ | Leu961 | O | 2.8 |
| Arg562 | $N^{\eta 2}$ | Glu945 | $O^{\varepsilon 1}$ | 2.8 |
| Arg562 | $N^{\eta 2}$ | Glu945 | $O^{\varepsilon 2}$ | 2.7 |
| Arg562 | $N^{\eta 2}$ | Ala944 | O | 2.8 |
| Arg827 | $N^{\eta 1}$ | His962 | O | 3.1 |
| Arg827 | $N^{\eta 1}$ | Leu963 | O | 2.8 |
| Arg827 | $N^{\eta 1}$ | Asn964 | $O^{\delta 1}$ | 2.7 |
| Glu834 | $O^{\delta 1}$ | Asn964 | $N^{\eta 2}$ | 3.0 |
| Glu834 | $O^{\delta 1}$ | Arg967 | $N^{\eta 1}$ | 3.0 |
| Glu834 | $O^{\delta 2}$ | Arg967 | $N^{\eta 2}$ | 2.7 |

### 3.4 Discussions

AT domains or discrete AT proteins are responsible for the loading of starter or extender units for FASs, modular and iterative PKSs. Extensive structural and biochemical studies have been conducted to understand the catalytic mechanism and substrate specificity of the AT domains of FASs and modular PKSs. The crystal structure of $AT_{DYN10}$ presented here, which represents the first AT domain structure for an iterative PKSs, provides valuable information on the relationship between AT domains from the iterative PKS, FASs and modular PKSs. The conserved KS-AT linker domain, overall sequence homology and similarities in active site, suggests that $AT_{DYN}$ is most closely related to the malonyl-CoA specific AT domains of modular PKSs.

#### 3.4.1 Substrate specificity

Our previous studies have shown that the iterative PKSEs (CalE8, SgcE8 and DynE8) assemble their polyketide intermediates using the acetate units of malonyl-CoA alone[86; 87; 90]. The crystal structure of $AT_{DYN10}$ reveals the molecular basis for substrate preference, with several conserved key residues/motifs identified for malonate binding. Similar to the malonyl-specific AT domains from FAS and modular PKSs, the specificity toward malonyl-CoA is determined by several key

residues as well as the properties of the active site. In addition to the catalytic residues Ser[651] and His[753], the other key residues contributing to substrate specificity include Arg[676], Phe[752], Gln[581], Leu[652], Met[680] and Gln[623]. The key determinant Arg[676] not only defines the depth of the binding pocket but also forms a salt bridge with the carboxylate of the malonyl group. The specificity for malonyl-CoA, but not α -substituted malonyl-CoA (e.g. methylmalonyl CoA), appears primarily achieved by the bulky residues Phe[752] (from the HAFH motif) and Met[680], as the latter would restrict the ability of Phe[752] to move out the way to accommodate the additional methyl group (**Figure 3.18**).

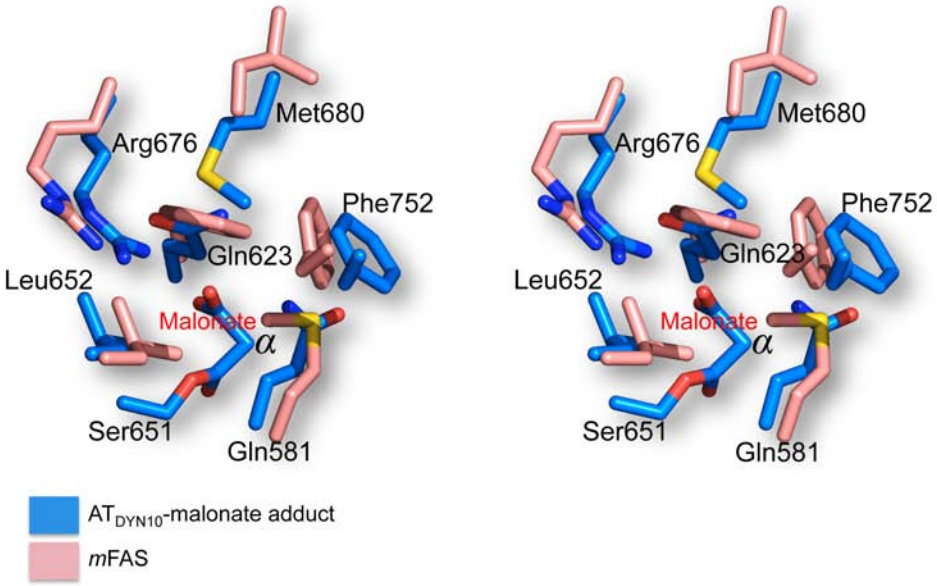



**Figure 3.18** Overlay of $AT_{DYN10}$-malonate adduct with AT mammalian FAS. Only residues of $AT_{DYN}$ are labeled. The malonate moiety is covalently attached to residue Ser[651] of $AT_{DYN}$. Residues Met[680] and Phe[752] of $AT_{DYN}$ would hinder binding of a bulkier substrate (e.g. methylmalonyl-CoA) with substituents projecting from the α -carbon. The active site of mammalian FAS is more hydrophobic compared to $AT_{DYN}$ and other AT because of the substitution of Gln[581] and Gln[623] by Met[499] and Phe[553] respectively. The stereo view was generated using PYMOL[104].

Both DynE8 and mammalian FAS assemble their products using a single set of catalytic domains iteratively and they both contain an acetate starter unit in their final products. However, $AT_{mFAS}$ is known to be an acetyl- and malonyl- dual specific domain[19]; while the $AT_{DYN}$ domain does not use acetyl-CoA as a substrate.

### 3.4.2 The $AT_{DYN10}$ catalytic mechanism

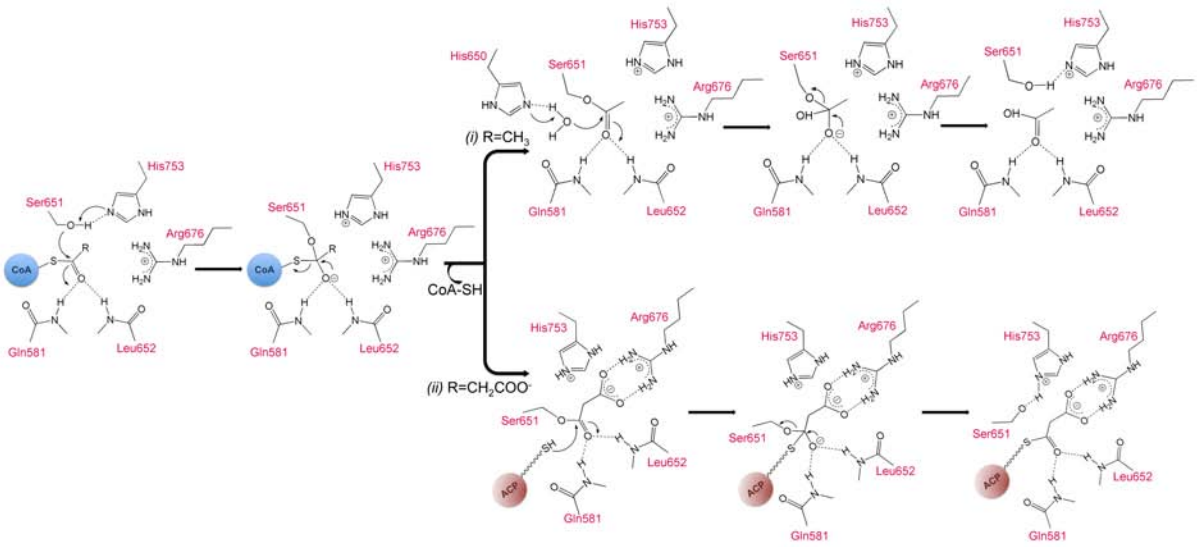



**Figure 3.19** Proposed pathway following post-catalytic mechanism for acetyl- and malonyl-enzyme intermediate. (i) Acetyl-$AT_{DYN}$ intermediate: $His^{650}$ deprotonate the adjacent water molecule and subsequently hydrolyzes the covalent bond of acetyl-$Ser^{651}$. (ii) Malonate-$AT_{DYN}$ intermediate: A phosphopantetheined-ACP domain extracts malonate by breaking the malonyl-Ser651 covalent bond.

Our data indicate that the terminal acetate unit of the DynE8 product is derived from malonyl-CoA through an intrinsic decarboxylation mechanism. Previously, it was not known whether the decarboxylation is catalyzed by the ketosynthase (KS) domain or the AT domain. Observation of the malonyl-enzyme covalent complex confirms that the AT domain does not convert the malonyl group to acetyl group, and that the decarboxylation is thus most likely catalyzed by the KS domain. Our structural and biochemical data also suggest that $AT_{DYN}$ can transfer the acetyl moiety of acetyl-CoA onto the enzyme. The possible mechanisms for the trans-acylation and hydrolysis reactions are depicted in **Figure 3.19**. In contrast to $AT_{mFAS}$ which subsequently transfers the acetyl moiety to the ACP domain, the acetyl group attached to $AT_{DYN}$ is likely to be off-loaded from the protein by a hydrolytic process. Hence, one of the key differences between these two AT domains is that while $AT_{mFAS}$ is able to protect both the acetyl- and malonyl-enzyme intermediate from hydrolysis, $AT_{DYN}$ only protects the malonyl enzyme intermediate, but not the acetyl-enzyme intermediate. Comparison of the two AT

domains points to some subtle structural differences that may account for the different mode in acetyl transfer. Compared to $AT_{mFAS}$, the active site of $AT_{DYN}$ is more spacious and hydrophilic with $Gln^{581}$ and $Gln^{623}$ replacing residues $Met^{499}$ and $Phe^{553}$ in $AT_{mFAS}$ (**Figure 3.18**). As a result, the active site of $AT_{DYN}$ is more likely to contain water molecules to hydrolyze the acetyl-enzyme intermediate. In support of this model, a few ordered water molecules are seen in the substrate binding pocket of the free enzyme occupying the position of the malonate carboxylate in the malonyl-enzyme complex (**Figure 3.10**).

In summary, the crystal structures reported here yield insight into the close evolutionary relationship with the malonyl specific AT domains of modular PKSs and the acetyl/malonyl specific AT domain of mammalian FAS. Observation of a malonyl enzyme covalent complex settles some uncertainties in the enediyne biosynthetic mechanism by suggesting that the terminal acetate unit of the PKS products is likely to be generated by a KS domain-catalyzed decarboxylation. In contrast to the previous belief that the malonyl-specific AT domains of enediyne PKSs do not recognize acetyl-CoA, the current structural and biochemical data reveal the surprising uploading of both acetyl and malonyl groups to the malonyl-specific $AT_{DYN}$ domain. However, $AT_{DYN}$ only transfers malonyl group, but not the acetyl group, to the ACP domain by protecting the malonyl-enzyme intermediate from hydrolysis and facilitating the second malonyl transfer, presumably by activating the oxyanion hole.

# CHAPTER 4 Crystal Structure and Catalytic Mechanism of the Hot-dog Fold Thioesterase CalE7, SgcE10 and DynE7 in Enediyne Biosynthesis

## 4.1 Introduction

Naturally-occurring enediynes constitute a small class of potent antitumor secondary metabolites that have been exploited as anticancer drugs[10; 80; 125].A thorough understanding of their biosynthetic pathway is a prerequisite for future rational modifications, combinatorial biosynthesis and metabolic engineering efforts. The enediynes contain a 9- or 10- membered bicyclic structural unit consisting of two acetylenic groups conjugated to a double bond. Sequencing of the biosynthetic gene clusters for several enediyne natural products has uncovered de novo biosynthetic pathways centered around a family of type I iterative polyketide synthases (iPKS)[52; 54; 61; 62; 126]. CalE8, SgcE and DynE8 are examples of multifunctional and modular PKSE that catalyze stepwise decarboxylative condensation of acyl-coenzyme A (CoA) precursors with the elongation substrate malonyl-CoA (**Figure 4.1**) and are responsible for the biosynthesis of the enediyne natural products calicheamicin, C-1027 and dynemicin A respectively[78; 86; 90; 127]. CalE8, SgcE and DynE8 cooperate with their cognate thioesterases (TE), CalE7, SgcE10 and DynE7, that are discrete proteins responsible for the *trans* hydrolytic release from the terminal thiol of the phosphopantheteine cofactor of the final acyl carrier protein (ACP)-tethered acyl-product (**Figure 4.1**). Recent biochemical studies on the PKSE proteins from several enediyne pathways led to the characterization of two conjugated polyene products: carbonyl conjugated polyene 3,5,7,9,11,13-pentadecen-2-one (**3**) and conjugated polyene 1,3,5,7,9,11,13-pentadecaheptanene (**4**) (**Figure 4.1**). Although it remains to be seen whether these conjugated polyenes are authentic biosynthetic intermediates, it is reasonable to believe that the enediyne moieties derive from structurally related polyene precursors[76; 85; 89; 126].

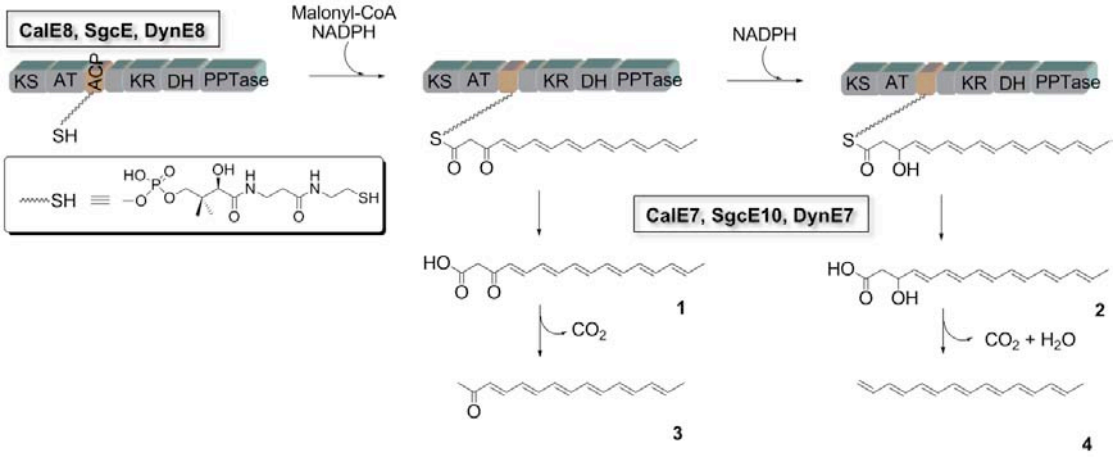



**Figure 4.1** Domain organization of the iterative PKSEs, CalE8, SgcE8 and DynE8, and their products. The structure of phosphopantetheine moiety is shown in the box. (KS, ketoacyl synthase; AT, acyl transferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; and PPTase, phosphopantetheinyl transferase).

Genetic, biochemical and structural studies revealed that TEs involved in the early steps of enediyne biosynthesis belong to a subfamily of hotdog fold thioesterases (TE). *In vitro* biochemical assays showed that hot-dog fold TEs release products **3** and **4** from CalE8, SgcE or DynE8[76; 85; 89; 126]. Polyketide and nonribosomal peptide (NRPS) synthesis generally involves the type I and type II thioesterases for the release of final product or removal of aberrant products. Type I thioesterase (TE I) are *cis*-acting domains fused to the C-terminal of the most downstream module of PKS or NRPS for the release and cyclization of the final product[128; 129]. Type II thioesterases (TE II) are discrete proteins responsible for the *trans* hydrolytic release of aberrant products[82; 84; 130]. TE II proteins are structurally and evolutionarily related to a family of well-known α/β hydrolase that contain 240 – 260 residues[131]. A common serine esterase motif G*X*S*X*G and another downstream motif GXH are conserved in TE II proteins[132; 133]. Based on the sequence information, TEs from the enediyne pathways including CalE7, SgcE10 and DynE7, do not share any significant amino acid sequence homology with the typical type I and II TEs. Instead, CalE7, SgcE10 and DynE7 share moderate sequence homology with a family of hotdog fold proteins characterized by a long central α-helix packed against a five-stranded anti parallel β -sheet. The three-dimensional structure and substrate specificity of several hotdog fold thioesterases

have been determined, including YbgC from *Helicobacter pylori*[134], homotetrameric medium chain acyl-CoA thioesterase II from *E. coli*[135], Paal from *Thermos thermophilis* HB8[136], FcoT from *Mycobacterium tuberculosis*[137], YciA from *Haemophilus influenza*[138], human THEM2[139] and 4-hydroxylbenzoyl-CoA thioesterases (4-HBT) from *Pseudomonas sp.* Strain CBS and *Arthrobacter sp.* strain SU[66; 140; 141]. Despite their diverse specificity towards acyl substrates, almost all known hotdog fold thioesterases catalyze the hydrolysis of the thioester bond by using a Glu/Asp residue as nucleophile or general-base catalyst[85; 139].

In this chapter, I present the crystal structure and biochemical data that show CalE7, SgcE10 and DynE7 from the enediyne biosynthesis pathway do not contain catalytically critical acidic residues in the active site. Thus, the three hotdog fold proteins, in contrast to the canonical type I and II thioesterases, are likely to employ a novel catalytic mechanism for thioester hydrolysis. The crystal structure of CalE7, SgcE10 and DynE7 revealed a homo-tetrameric protein with each subunit adopting a hot-dog fold. The structural data suggest that CalE7, SgcE10, are DynE7 are likely to use an arginine residue for catalyzing the hydrolysis of the thioester bond. From the data collected, the crystal structure of CalE7 and SgcE10 in its closed conformation displays a kinked substrate-binding channel which is too short to accommodate a full length product. On the other hand, the DynE7 crystal reveals an extended substrate-binding channel with the carbonyl-conjugated polyene (**3**) trapped inside. This structure provides insight into substrate binding and the enzymatic mechanism for thioester bond hydrolysis by DynE7. Furthermore, a comparison of the structure of ligand-bound DynE7 with the unliganded enzyme at 2.7 Å resolution reveals extensive conformational changes that are crucial for substrate binding. Together with mutagenesis analysis, the detailed information gleaned from the present structural studies further our understanding of the structure and function of this highly versatile hotdog fold superfamily of thioesterases. Thus, the comparison between the three closely-related homologs revealed a remarkable induced-fit mechanism for the substrate binding of the TEBC family proteins and suggested a novel catalytic mechanism for this subfamily of thioesterases.

## 4.2 Material and methods

### 4.2.1 Reagents and chemicals

Coenzyme A (CoA), acetyl-CoA, malonyl-CoA, NADPH and other chemicals are purchased from Sigma-Aldrich and stored at -20 °C. Expression vectors pET-28b(+), pCDF-2 Ek/LIC and *E. coli* BL21(DE3) were obtained from Novagen.

### 4.2.2 Cloning, co-expression and co-purification of PKSEs and TEs

Cloning, co-expression and co-purification of the PKSEs (CalE8, SgcE and DynE8) and TEs (CalE7, SgcE10 and DynE7) have been described in details in Chapter 2.

### 4.2.3 Site-directed mutagenesis on CalE7 and DynE7

Site-directed mutagenesis of CalE7 and DynE7 was performed using the QuickChange II kit (Strategene) following the manufacturer instructions. Eleven single CalE7 mutants and two single DynE7 mutants, harbored by pCDF-2-Ek/LIC were mutated and subsequently confirmed by DNA sequencing. All the thioesterase mutants are listed in **Appendix Table 4.1**. The DynE7 mutant proteins, DynE7-R35A and DynE7-E36A, were separately co-expressed and co-purified with pET-28b(+)-*DynE8* following the same procedure as for the wild-type protein. The absorption spectra of each co-purified DynE7-R35A and DynE7-E36A proteins were measured using a semi-micro cuvette on a Shimazu 1700 UV-Vis spectrophotometer.

### 4.2.4 Activity assay and product analysis by absorption spectroscopy and HPLC for CalE7 and its respective mutants

Enzymatic assay was carried out with 7.2 μM CalE8, 27.8 μM CalE7 (or CalE7 mutant), 1.05 mM NADPH and 1.05 mM malonyl-CoA in 900 μL of pH 8.2 buffer (100 mM Tris, 300mM NaCl, 1mM DTT). The reaction mixture in the absence of malonyl-CoA was equilibrated at 30°C for 15 min in the temperature-controlled sample chamber of Shimazu UV-Vis 1700 spectrometer. Enzymatic reaction was initiated with the addition of malonyl-CoA, and was carried out for 2

hours at 30°C with values for $OD_{384}$ and $OD_{409}$ taken at 5 minutes intervals. For the analysis of products by HPLC, an equal volume of the ethyl acetate was added into the yellowish CalE7 protein solution and vigorous vortexing ensured. The mixture was centrifuged at 22,000 x *g* for 15 minutes. The organic solvent layer was then pipetted-out and subjected to extraction for a second time. The ethyl acetate extracts were pooled and the solvent was evaporated using a Speedvac centrifuge. The dried sample was re-dissolved in methanol before application onto an analytical eclipse XDB C18 column (4.6 x 250mm) using an Agilent 1200 HPLC. A full gradient was employed from 100% Buffer A (HPLC grade water with 0.045% TFA) to 100% Buffer B (acetonitrile with 0.045% TFA) in 60 minutes. HPLC UV detector was set at multiple wavelengths, such as 280nm, 375nm, 400nm, 420nm with a reference wavelength of 600nm.

### 4.2.5 Kinetic measurement of reaction involving DynE8, DynE7 and its respective mutants.

Steady-state kinetic measurements were carried out. The reaction mixture consists of 3.2 μM DynE8, 66.8 μg DynE7-R35A or DynE7-E36A, 140 μM malonyl-CoA and 150 μM NADPH, topped up with Tris HCl buffer (100mMTris HCl, 300mMNaCl and 1 mM DTT, pH 8.2) to 200 μL in cuvette. The reaction mixture was pre-incubated at 30 °C prior to malonyl-CoA addition. Full wavelength scan was performed at 2 minutes interval for duration of 2 hours. The reaction kinetics was monitored at 340 nm which corresponds to the decay of NADPH in the reaction mixture.

### 4.2.6 Crystallization and X-ray diffraction data collection

The co-purified CalE7, SgcE10 and DynE7 proteins, ~15 mg/mL, from PKSE-TE co-expression system were used for the crystallization screen. An automated initial crystallization screen of 672 conditions was performed using the Cybio®-Crystal Creator (Jena Biosciences). A volume of 200 nL of purified protein (10 mg/mL) was added to an equal volume of crystallization solution using the sitting drop, vapor-diffusion method. Crystals of free CalE7 and SgcE10 were obtained at 291 K with Index Screen condition 38 (0.1M HEPES pH 7.0, 30%

Jeffamine-M600, Hampton Research) and Index Screen condition 82 (0.2 M Magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350), respectively. Meanwhile, crystals of unliganded DynE7 and ligand-bound DynE7 were obtained a 291 K with PEG Ion Screen II condition 18 (8% tacsimate, pH 8.0, 20% w/v PEG 3350, Hampton Research) and PEG Ion Screen I condition 36 (0.2 M sodium tartrate dibasic dehydrate, 20% w/v PEG3350, Hampton Research). Before data collection, each crystal was transferred to a cryo-protecting solution containing 25% glycerol and cooled to 100 K in a gaseous $N_2$ stream using an Oxford cryosystem. Diffraction intensities were collected at the NSRRC synchrotron (Hsinchu, Taiwan) to a resolution of 1.75 Å for unliganded CalE7, to 2.8 Å for unliganded SgcE10, to 2.7 Å for unliganded DynE7 and to 2.1 Å for ligand-bound DynE7. Integration, scaling and merging of the intensities were carried out using programs MOSFLM and SCALA from CCP4 suite[97]. Data collection statistics are summarized in **Table 4.1**.

### 4.2.7 Structure determination and refinement

The structure of CalE7 was solved using the molecular replacement software MrBUMP[142] with the crystal structure of a hypothetical protein (AQ1494) from *Aquifex aeolicus* (PDB code 2EGI) having an overall homology of only 28% of 126 aligned residues with CalE7, as a search probe. Automated model building was carried out with Arp/warp[143]. The structure was refined using molecular dynamics and simulated annealing as implemented in the program CNS[144], with positional and individual temperature factor refinement without NCS restrains. The computer graphics software O was used for model rebuilding between refinement cycles[145]. Analysis of the atomic model was carried out with the CCP4 suite of programs[97].

Structures for the unliganded enzyme SgcE10, the unliganded enzyme DynE7 and the ligand-bound were solved using the molecular replacement software Molrep from the CCP4 suite[97] with the CalE7 protein from *Micromonospora echinospora* spp. *calichensis* (PDB: 2W3X) as a search probe. These proteins have an overall identity of 45% for 150 aligned amino-acid residues. Automated model

building was carried out with the program Chainsaw from the CCP4 suite[97]. Refinement cycles carried-out using REFMAC5[97] without NCS restraints between six subunits of the SgcE10, the eight subunits of the two unliganded-tetramers of the DynE7, and the four subunits of the ligand-bound tetramer of DynE7 were interspersed with model rebuilding using Coot[101]. The program BUSTER was introduced for the last refinement steps[100].

Refinement statistics are shown in Table 4.1. The quality of the structures was analyzed using PROCHECK[103]. The stereochemistry for the 15-carbon polyene ligand was built using the PRODRG server[146]. Figures were prepared using the program Pymol[104]. The refined coordinates and structure factor amplitudes have been deposited in the PDB (PDB code: 2W3X).

### 4.2.8 Analysis of the product bound by DynE7 in solution and crystal by HPLC

In order to analyze the chemical composition of the yellow compound bound to the DynE7, an equal volume of ethyl acetate was added into the yellowish DynE7 protein solution and vigorous vortexing ensued. Subsequently, the mixture was centrifuged at 22,000 x *g* for 15 minutes. The organic solvent layer was then pipetted-out and subjected to extraction for a second time. Ethyl acetate extracts were pooled and the solvent was evaporated using a Speedvac centrifuge. The yellow pigment was re-dissolved in a small volume of methanol before application onto an analytical eclipse XDB RP C18 column (4.6 x 250 mm) using an Agilent 1200 HPLC. The gradient employed was from 70% buffer A (HPLC grade water with 0.045% TFA) and 30% buffer B (100% acetonitrile with 0.045% TFA) to 100% buffer B in 60 minutes. The diode-array detector was set at 410 nm with a reference wavelength of 600 nm. For liganded DynE7 crystals, the bound product was extracted from dissolved crystals following the procedure described above. The standards, 1,3,5,7,9,11,13-pentadecaheptanane (4) and 3,5,7,9,11,13- pentadecen-2-one (3), were produced by the *in vitro* enzymatic reaction with CalE8 and CalE7.

### 4.2.9 Accession numbers

The refined coordinates and structure factors amplitudes have been deposited in the PDB with accession code 2W3X for the unliganded CalE7, 2XFL for the unliganded DynE7 and 2XEM for the ligand-bound DynE7. SgcE10 structure has not been deposited in the PDB.

## 4.3 Results

### 4.3.1 Overall structure of CalE7

The production of yellow colored product-bound CalE7 were previously observed when the protein was co-expressed with CalE8[86]. The product-bound CalE7 was used for crystallization in an attempt to obtain the 3D structure of an enzyme-product complex. However, despite the intense yellow appearance of the original protein solution due to the binding of the product, crystallization in the presence of Jeffamine M-600, a polyether amine, appears to select the subpopulation of unliganded CalE7. As a result, only the structure of the unliganded enzyme was obtained and refined to a final $R_{work}$ of 19.6% and $R_{free}$ of 23.4% (**Table 4.1**). The asymmetric unit contains six molecules that can be overlaid with an average r.m.s.d. of 0.56 Å. The CalE7 monomer (**Figure 4.2a**) exhibits a typical α/β hotdog fold with the five-stranded antiparallel β-sheet wrapping around the long α-helix α1 and two shorter α-helices α2 and α3 at the N-terminus of the central helix[85]. The main conformational differences between the six independent monomers are observed in the β5–α3 loop.

An automatic search for structurally homologous proteins using the DALI server[115] revealed that CalE7 is closely related to family of hotdog fold proteins, including YbgC[134], Paal[136], FcoT[137], YciA[138], hTHEM2[139] and 4-HBT[66; 140; 141] (**Figure 4.3**). In solution, CalE7 is a tetrameric protein. Six subunits were seen in the asymmetric unit: four form a 222 (or *D2*) tetramer and two form a 2-fold symmetric dimer. A second complete 222 tetramer can be generated from the latter dimer via the crystallographic dyad. A crystal packing system of CalE7 is shown in **Appendix Figure 4.1**. As seen for other hotdog fold thioesterases, dimerization proceeds through the formation of a continuous 10-stranded antiparallel β-sheet

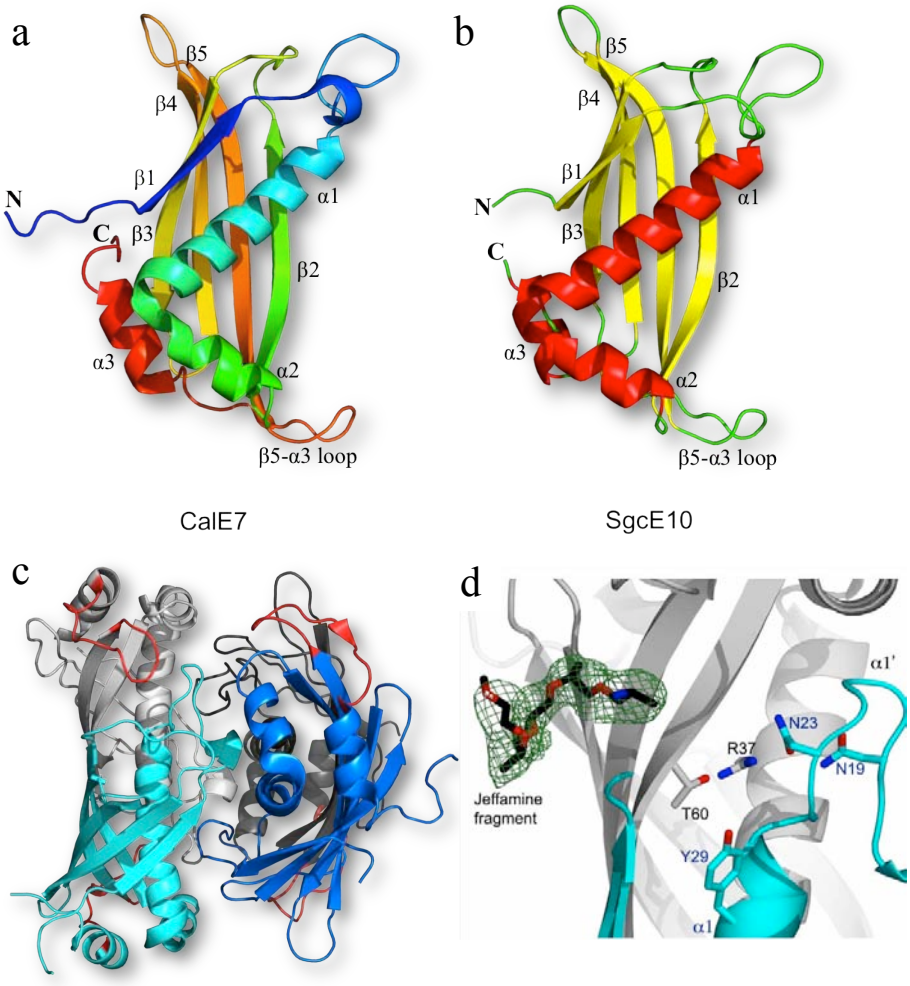



**Figure 4.2** Crystal structures of CalE7 and SgcE10. Representation of the CalE7 (**a**) and the SgcE10 (**b**) monomer. The protein secondary structures elements are labeled and colored. Both N- and C-terminus are labeled. (**c**) Cartoon representation of the CalE7 tetramer. Subunits are colored grey and cyan for one dimer and blue and dark grey for the other. The β5- α3 loops at the active sites are colored red. (**d**) Simulated annealing difference Fourier map with coefficients $|F_{obs}|$ - $|F_{calc}|$ and phases calculated from the protein model with atoms from the Jeffamine fragment omitted and contoured at level of 3 σ.

with the central helices on the inner side of the β-sheet (**Figure 4.2c**). This buried dimer interface (4540 $Å^2$) is stabilized via main-chain interactions between strands β2 and by residues 25 – 34 of the central helix α1 with their 2-fold counterparts, as well as by hydrophobic interactions between α2 of one subunit and the β1–α1 loop of the other subunit. Tetramer formation leads to the burying of the four central helices while the β-sheets are exposed to the solvent, as observed in other hotdog fold thioesterases with the εγ oligomeric arrangement[66; 134; 140]. A hydrophilic channel ~8 Å wide lies at the center of the tetramer along a 2-fold symmetry axis. Pyrophosphate molecules, possibly carried over from bacterial expression, are observed at both ends of this solvent-accessible channel.

### 4.3.2 Overall structure of SgcE10

The overall structure of SgcE10 reflects the structural fold of unliganded CalE7[87]. The secondary structure alignment shows only minor differences between unliganded CalE7, with an overall sequence identity of 60%. The structure of the unliganded SgcE10 was refined to a final $R_{work}$ of 23.2% and $R_{free}$ of 24.1% (**Table 4.1**). The monomer of SgcE10 shares a similar structural fold with CalE7 (**Figure 4.2b**). Similar to CalE7, dimerization creates a continuous 10-stranded antiparallel β-sheet with the central helices on the inner side of the β-sheet. The stabilization of the buried dimer interface is due to main-chain interactions between strands β2 and central helix α1 with their 2 fold-counterparts, and the hydrophobic interactions between α2 of one subunit and β1 – α1 loop of the other subunit. In solution, SgcE10 is a tetramer. Six subunits were seen in the asymmetric unit: four form a 222 (or *D2*) tetramer and two form a 2-fold symmetric dimer. A ~ 7Å-wide hydrophilic channel is observed at the center of the tetramer along 2-fold symmetry axis.

**Table 4.1** Data collection, phasing and refinement statistics

| | Unliganded-CalE7 | Unliganded-SgcE10 | Unliganded-DynE7 | Ligand-bound DynE7 |
|---|---|---|---|---|
| ***Data collection*** | | | | |
| X-ray source | BM13B1, NSRRC | BM13B1, NSRRC | BM13B1, NSRRC | BM13B1, NSRRC |
| Wavelength (Å) | 1.027 | 1.027 | 1.027 | 1.027 |
| Spacegroup | *C*2 | $P3_221$ | $P2_1$ | $P2_1$ |
| Unit cell dimensions | | | | |
| a, b, c (Å) | 102.0, 73.3, 124.4 | 75.9, 75.9, 347.43 | 60.1, 178.5, 64.0 | 58.5, 65.4, 94.4 |
| α, β, γ (°) | 90.0, 98.7, 90.0 | 90, 90, 90 | 90.0, 110.8, 90.0 | 90.0, 100.0, 90.0 |
| Molecules per asymmetric unit | 6 | 6 | 8 | 4 |
| Resolution[a] | 30.0-1.75(1.81-1.75) | 24-2.8 (2.70-2.80) | 25.0-2.7 (2.65-2.70) | 31.0-2.10 (2.25-2.10) |
| Completeness | 99.3 (99.6) | 98.6 (99.6) | 98.3 (89.1) | 97.3 (96.7) |
| Measured reflections | 329,627 | 310,245 | 136,152 | 147,259 |
| Unique reflections | 90,879 | 89,134 | 34,008 | 39,917 |
| Average I/σ(I) | 24.2 (1.2) | 18.4 (2.9) | 9.8 (2.1) | 6.5 (2.4) |
| $R_{merge}$ (%) | 0.084 (0.66) | 0.056 (0.60) | 0.087 (0.35) | 0.122 (0.43) |
| ***Refinement*** | | | | |
| R factor ($R_{work}$ / $R_{free}$) (%) | 0.196/0.235 | 0.232/0.241 | 0.224/0.247 | 0.162/0.194 |
| No. of atoms (protein/water/ligand) | 7,011/530/62 | 6994/380 | 8,637/280 | 4955/467/16 |
| Mean B factor (Å$^2$) (protein/water/ligand) | 39.0/49.1/61.7 | 35/45.2 | 35/45.6/- | 30/37.4/60 |
| r.m.s.d. distance (Å$^2$) | 0.006 | 0.008 | 0.010 | 0.010 |
| r.m.s.d. bond angles (°) | 1.37 | 1.38 | 0.98 | 1.07 |
| ***Ramachandran plot***[b] | | | | |
| Residues in the favoured regions (%) | 91.5 | 91.8 | 96.0 | 98.3 |
| Residues in the allowed regions (%) | 8.5 | 8.1 | 3.9 | 1.7 |
| Residues in the disallowed regions (%) | 0 | 0.1 | 0.1 | 0 |
| PDB code | 2W3X | - | 2XFL | 2XEM |

[a] Statistic for the highest resolution shell are shown in parenthesis
[b] As defined by PROCHECK

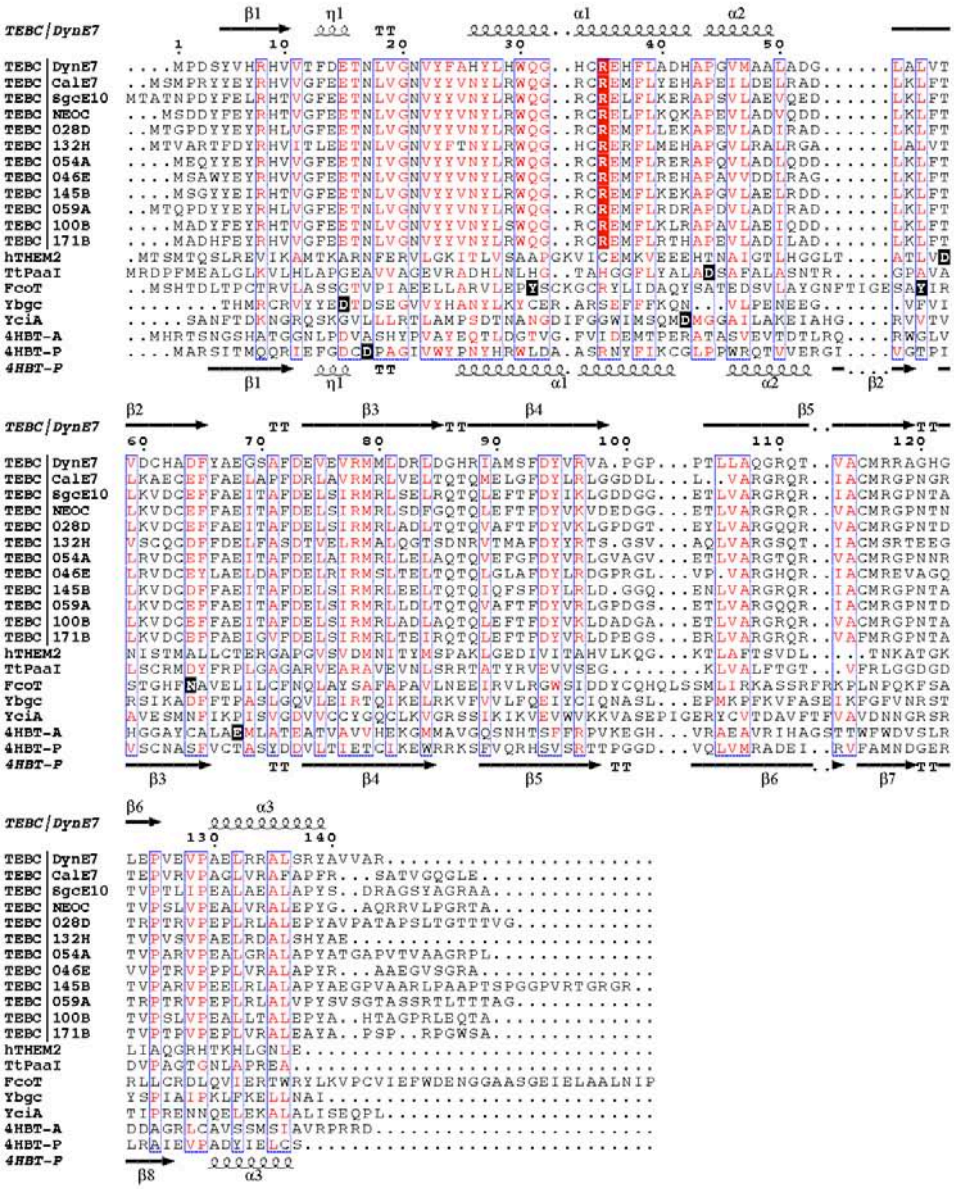


**Figure 4.3** Structure-based sequence alignment of CalE7, SgcE10, DynE7 with TEBC family proteins and other hotdog fold thioesterases. The catalytically essential residue arginine that is strictly conserved among TEBC family proteins but not in other hot-dog fold proteins is highlighted in red. The catalytic residues for other hotdog fold thioesterase are highlighted in black. TEBC family proteins: TEBC|CalE7 (*Micromonospora echinospora*), TEBC|DynE7 (*Micromonospora chersina*) and TEBC|SgcE10 (*Streptomyces globisporus*), TEBC|NEOC (*Streptomyces carzinostaticus* subsp. *neocarzinostaticus*), TEBC|028D (*Kitasatospora* sp. CECT 4991), TEBC|132H (*Lechevalieria aerocolonigenes*), TEBC|054A (*Micromonospora megalomicea* subsp. *nigra*), TEBC|046E (*Micromonospora* sp. 046/Eco11), TEBC|145B (*Streptomyces citricolor*), TEBC|059A (*Streptomyces griseus*), TEBC|100B (*Streptomyces* sp. 100/Eco52), TEBC|171B (*Streptomyces* sp. 171/Eco105). Hotdog fold thioesterase family proteins: hTHEM2 (*Homo sapiens*), TtPaaI (*Thermos thermophilis* HB8), FcoT (*Mycobacterium tuberculosis*), YbgC (*Helicobacter pylori*), YciA (*Haemophilus influenza*), 4HBT-A (*Arthrobacte*r sp. Strain SU) and 4HBT-P (*Pseudomonas* sp. Strain CBS-3).

### 4.3.3 Substrate-binding pocket CalE7 and SgcE10

An L-shaped putative substrate-binding pocket of CalE7, as well as SgcE10, is visible at each dimer interface, over the β-sheet edge (**Figure 4.4**). The location of the substrate-binding pocket agrees well with other characterized hotdog fold thioesterases. The inner channel extends inside the other subunit, surrounded by residues from helices α1, α2, and sealed by the short C-terminal helix α3 (**Figure 4a** and **4c**). The putative substrate-binding channel of CalE7 is lined with hydrophobic residues that project from the two neighboring subunits, as well as five polar residues that include $Arg^{37}$ and $Thr^{60}$ from one subunit and $Asn^{19}$, $Asn^{23}$, and $Tyr^{29}$ from the adjacent subunit (**Figure 4a-b**). An additional ~7-Å deep side pocket capped by the phenolic ring of $Tyr^{29}$ faces the inner segment of the substrate-binding channel with several water molecules making hydrogen bonds to residues $Tyr^{29}$ (2.9 Å), $Asn^{19}$ (2.84 Å), $Arg^{37}$ (2.85 Å), and $Thr^{60}$ (2.7 Å).

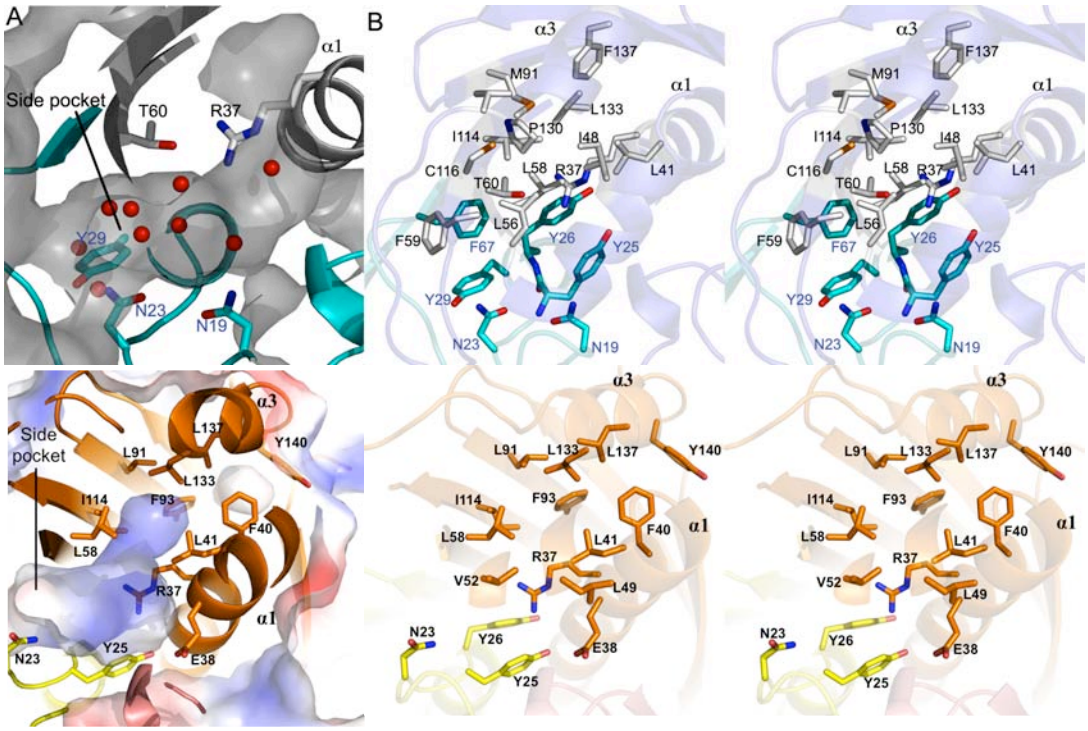



**Figure 4.4** Substrate-binding pocket of CalE7 and SgcE10. Cross-section through the surface of the (**a**) CalE7 and (**c**) SgcE10 substrate-binding pocket. The side pockets are shown. (**b-d**) Stereoview of the residues lining the substrate-binding channel. The polar and hydrophobic residues are labeled.

Meanwhile, SgcE10 possess a shorter substrate-binding channel in comparison to CalE7 (**Figure 4.4c**). The substrate-binding channel is likely gated by $Phe^{40}$, $Leu^{133}$, $Phe^{93}$ and $Leu^{137}$ that shorten the length of the channel (**Figure 4.4d**). No water molecules are observed in the binding channel, suggesting that SgcE10 exhibits a closed conformation that seals the entire entrance from any entries.

As inferred from a superposition with substrate-bound structures of 4-HBT from *Pseudomonas*[140], the segment forming the entrance of the channel at the dimer interface is likely to accommodate the phosphopantetheinyl moiety of the substrate. Interestingly, in the case of CalE7, only one (subunit E) out of the six substrate-binding pockets adopts an open conformation, with the substrate-binding channels of the other five subunits (A–D and F) shielded from bulk solvent by loop β5– α3. Structural comparison reveals that a movement of 2.8 Å of the β5– α3 loop results in the opening of the substrate-binding pocket to bulk solvent. This observation indicates that substrate binding is likely to be accompanied by movement of the flexible β5– α3 loop that acts as a gate (**Figure 4.5**). For SgcE10, all substrate-binding channel remained in the closed conformation.

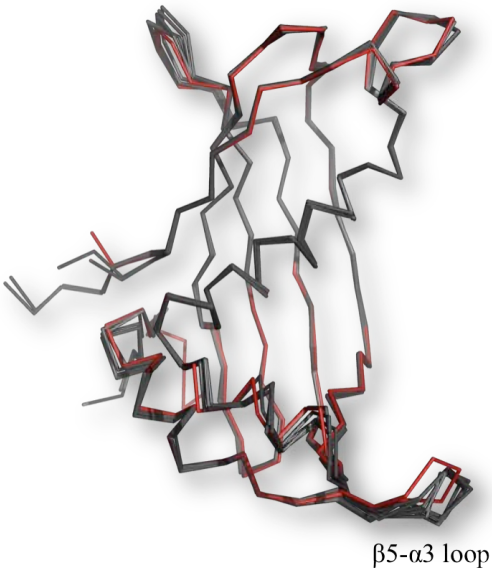



**Figure 4.5** Superposition of six subunits of CalE7 in an asymmetric unit. Subunit E which is most diverse is shown in *red* while the other molecules are shown in shades of gray. Not that the displacement of the β5– α3 loop in subunit E blocks the entrance of the substrate-binding channel.

No bound substrate was observed in the substrate binding pocket of both CalE7 and SgcE10. However, in the case of CalE7, a Jeffamine molecule from the crystallization solution is clearly visible in the electron density map (**Figure 4.2d**). The Jeffamine molecule sits at the edge of the dimer interface close to the entrance of the substrate-binding channel, and mostly stabilized by hydrophobic interactions. The location of the Jeffamine fragment may mimic interactions formed by the exposed region of the phosphopantetheinyl arm tethered to the ACP domain with the CalE7 protein.

### 4.3.4 Overall structure of unliganded-DynE7 and product-bound DynE7

Shortly after the elucidation of the crystal structure of CalE7 and SgcE10, I also determined the structure of the highly homologous DynE7. The statistics for the free and ligand bound DynE7 proteins crystals are presented in **Table 4.1**. Crystal of the unliganded DynE7 enzyme contains two homotetramers per asymmetric unit, with point group symmetry 222 and the ligand-bound DynE7 crystal form contains one homotetramer. The structural conformation of DynE7 protein resembles that of CalE7 and other hotdog fold TEs with five stranded antiparallel β-sheet wrapping around the long α-helix α1 (**Figure 4.6**).

The overall structure of the DynE7 monomer is highly similar to that of CalE7 with superposition of 134 equivalent Cα atoms from the two proteins yielding an r.m.s. deviation of 1.46 Å. The dimer interface is primarily stabilized through main-chain interactions between strand β2 and its two-fold counterpart that creates a continuous intermolecular β -sheet. Contacts within the primary dimer create the substrate-binding channel with an entrance in one subunit and exit in the other subunit (subunits D and B respectively in **Figure 4.6a-c**). Each substrate-binding channel comprises a short hydrophilic segment (outer binding channel) at its entrance, which constitutes the phosphopantetheinyl group binding site, and a hydrophobic segment (inner binding channel) shielded from the solvent. The total length of the prolonged channel is estimated to be around 30 Å, which could accommodate either of the two ACP-tethered β-keto and β-hydroxyl thioesters.

Similar to CalE7, the movement of β5-α3 loop leads to the opening of the hydrophilic segment at the beginning of the channel for incoming substrate. The finding confirmed the role of β5-α3 loop as part of the gating mechanism that controls the accessibility of the channel for the ligands. In addition, a minor translation of helix α2 and β2 strand further expands the entrance of the inner channel. At the same time, the differential conformation of $Leu^{136}$ and other residues that project from the C-terminal helix α3 in DynE7 lead to the extension of the rear portion thus increasing the length of the substrate-binding channel.

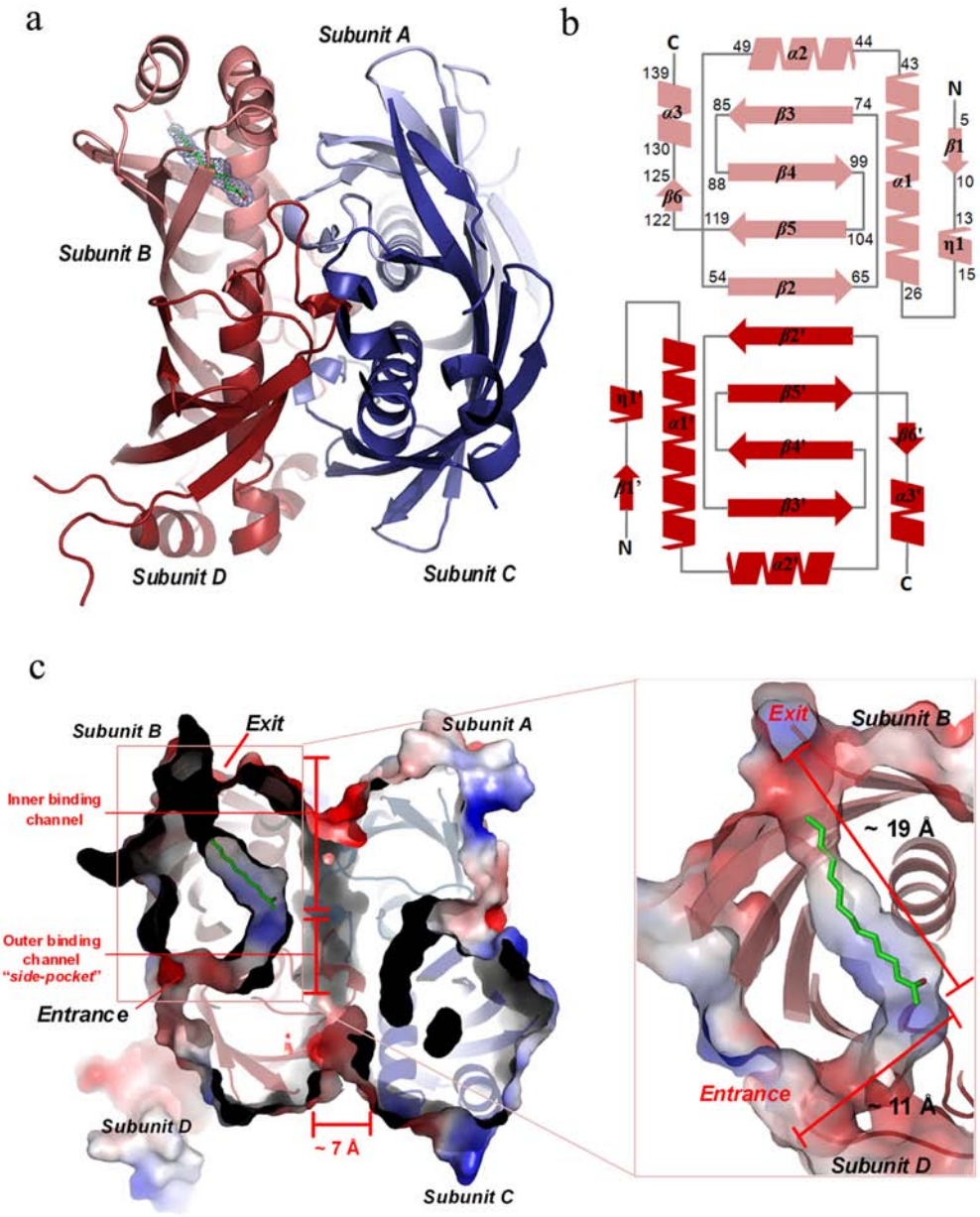



**Figure 4.6** Overall structure and topology of DynE7. (a) Schematic representation of the DynE7 tetramer showing the ligand as sticks colored in green at the interface of two subunits. Subunits are colored pink and red for one dimer and dark blue and light blue for the other. (b) Topology diagram of the dimer of dimers. The secondary structure elements are numbered according to CalE7[87]. (c) The L shaped ligand binding channel of DynE7. An open conformation is observed at the entrance and exit of this channel when it contains a ligand. The carbonyl-conjugated polyene colored in green is observed in a ~19 Å inner binding channel of subunit B of DynE7.

### 4.3.5 Co-expressed DynE7 crystal structure and product bound analysis

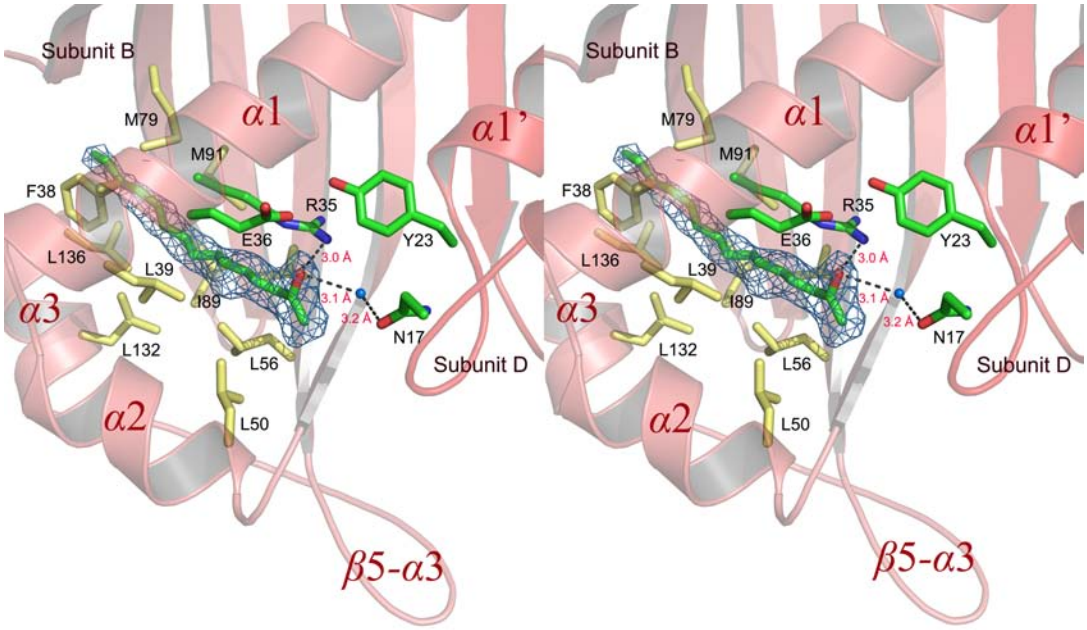



**Figure 4.7** A stereo ribbon diagram of the ligand-protein interaction in the binding channel. The carbonyl group of polyene is hydrogen-bonded with $Arg^{35}$ from helix α1 and to $Asn^{17}$ from the adjacent subunit D via a bridging water molecule (*blue* sphere). The hydrophobic poly-carbon tail of the carbonyl-conjugated polyene is stabilized by hydrophobic residues along the channel. Hydrophobic residues are colored in *yellow* and conserved polar residues in *green*. A simulated annealing Fourier map with coefficient $|F_{obs}|$-$|F_{calc}|$ and phases calculated from the protein model with atoms from the carbonyl-conjugated polyene fragment omitted is contoured at a level of 3.0 σ.

Using the co-expressed DynE7 protein, crystal structure of a product-protein binary complex was solved at 2.1 Å. Electron density maps unambiguously revealed an elongated ligand bound within the inner segment of this putative substrate-binding channel (**Figure 4.6 and 4.7)**. Using an HPLC analysis from dissolved protein crystals and comparison with the standards (**Figure 4.8)**, the ligand was identified as the carbonyl-conjugated polyene **3**, one of the major products synthesized from the malonyl-CoA precursor by the DynE8 protein. To confirm the identity of the ligand observed in only one out of the four subunits (monomer B) for the DynE7-product complex, the carbonyl-conjugated polyene **3** was manually placed in the same orientation in each of the three other monomers

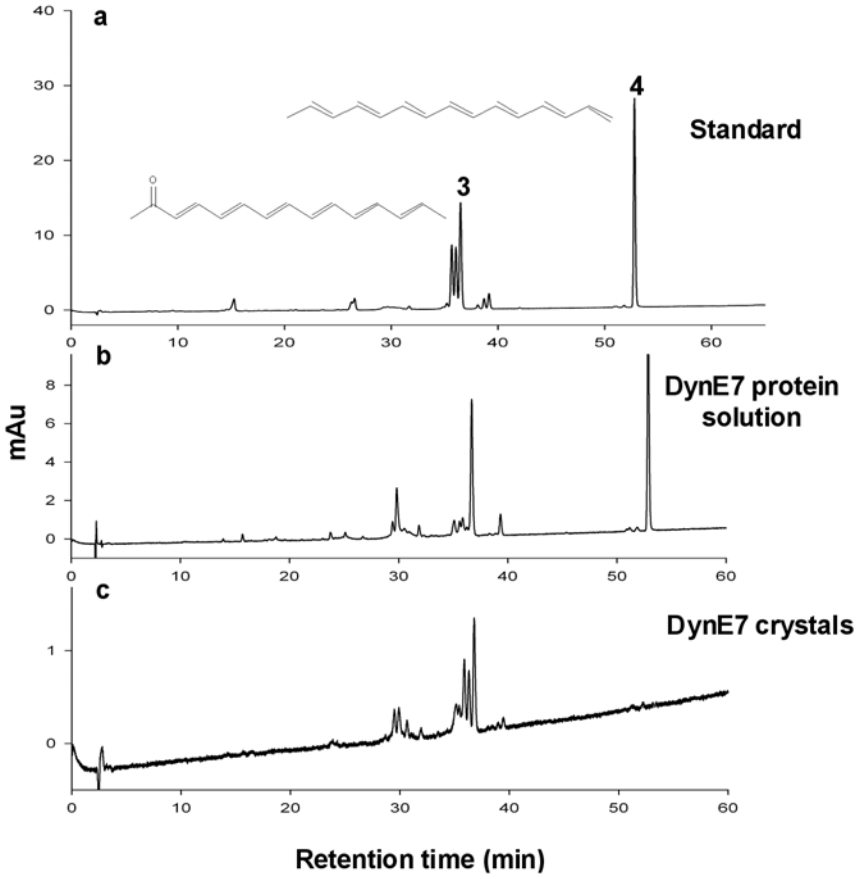



**Figure 4.8** HPLC analysis of products generated from DynE8 and DynE7 co-expression, recorded at 410 nm. (a) The standard composed of products 3 and 4. (b) Extract from ligand-bound DynE7 protein solution. Both product 3 and 4 are observed. However, for substrates extracted from ligand-bound DynE7 crystal (c) a majority of product 3 is observed.

and crystallographic refinement was performed in each case. Electron density maps with Fourier coefficient $|F_{obs}|$ - $|F_{calc}|$ and atoms from the ligand excluded from the phase calculation were compared, as well as temperature-factors for the putative ligand, after refinement. As a result, electron density maps convincingly indicate the presence of a bound ligand in subunit B only and not in the other three subunits A, C and D of the DynE7 tetramer (**Figure 4.9)**. In addition, a comparison of refined temperature factors for the polyene molecule placed in subunits A, C, D returns values of ~100 $Å^2$ whereas the ligand placed in subunit B has an average temperature factor of ~30 $Å^2$, a value comparable with those of neighboring coordinating atoms (~25 $Å^2$) (**Table 4.2**). This indicates that the ligand is bound in

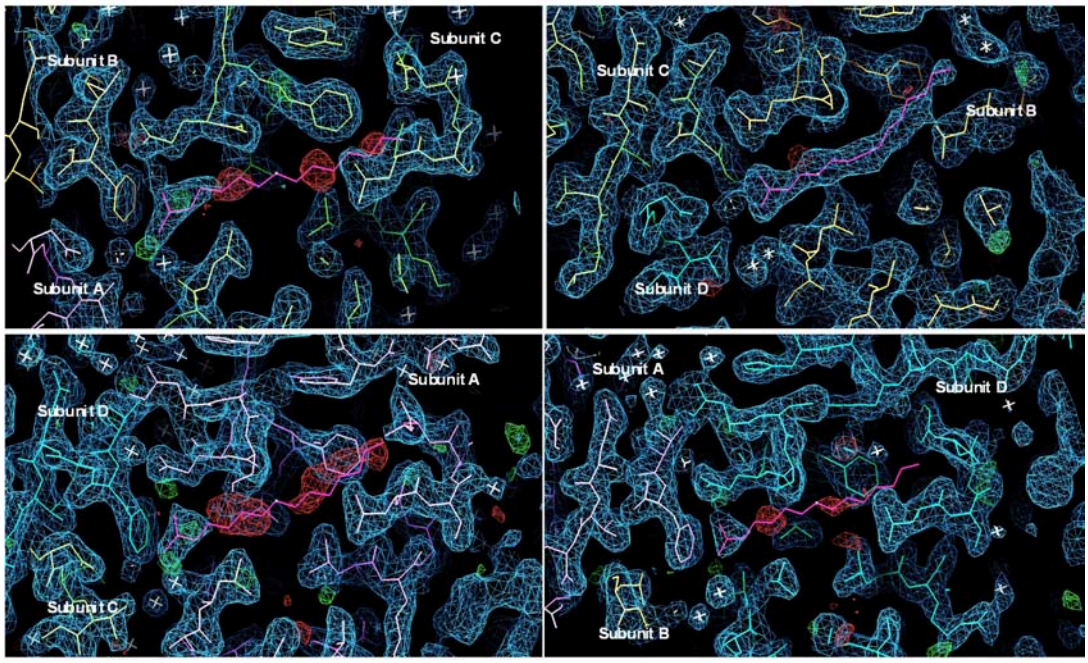



**Figure 4.10** Carbonyl-conjugated polyene binding. The final $|F_{obs}|$- $|F_{calc}|$ electron density map (positive map in *green* and negative map in *red*) contoured at 3.0 $\sigma$ indicates the absence of the ligand in the binding channels of monomers A, C and D colored in *purple*, *yellow*, *green* and *cyan*, respectively. The ligand, carbonyl-conjugated polyene, is colored in *red*. $2|F_{obs}|$- $|F_{calc}|$ map is colored in *blue*.

**Table 4.2** Carbonyl-conjugated polyene induced fit test

| | **Rfree** | **Rfactor** | ***B*-Factor (Å$^2$)** |
|---|---|---|---|
| **Monomer A** | 0.1998 | 0.1642 | 100.7 |
| **Monomer B** | 0.1940 | 0.1621 | 30.4 |
| **Monomer C** | 0.1987 | 0.1644 | 101.7 |
| **Monomer D** | 0.1954 | 0.1652 | 100.5 |

subunit B only and is also consistent with the longer and wider substrate-binding channel adopted in subunit B that has both its ends open. A list of interactions established between the protein and the polyene ligand is given in **Table 4.3**. Overall, the ligand seems to be sequestered into the channel through extensive hydrophobic packing, given that the hydrocarbon moiety **3** makes multiple contacts with the hydrophobic residues lining the inner segment of the channel (**Table 4.3**

**Table 4.3** Polyene-protein interactions in the substrate-binding channel

| ***Hydrogen bonds interactions*** [a] | | | |
|---|---|---|---|
| Residues (atom) | | Polyene (atom) | Distance (Å) |
| $Arg^{35}$ | Nη2 | O | 3.0 |
| $Asn^{17}$ | Oδ1 | O | 3.2 / 3.1 (water-mediated) |
| ***Hydrophobic interactions*** | | | |
| Residues (atom) | | Polyene (atom) | Distance (Å) |
| $Phe^{38}$ | Cζ | C3 | 3.6 |
| $Leu^{39}$ | Cδ1 | C10 | 3.4 |
| $Leu^{50}$ | Cδ1 | C14 | 4.8 |
| $Leu^{56}$ | Cδ1 | C11 | 3.2 |
| $Met^{79}$ | Cξ | C4 | 2.7 |
| $Ile^{89}$ | Cδ1 | C7 | 4.8 |
| $Met^{91}$ | Sδ | C5 | 3.3 |
| $Val^{113}$ | Cγ1 | C9 | 4.1 |
| $Leu^{132}$ | Cδ1 | C8 | 4.6 |
| $Leu^{136}$ | Cδ2 | C3 | 2.4 |

[a] In parentheses, potential hydrogen bonds but outside the limits given by Baker and Hubbard[147]

and **Figure 4.7**). In addition, several conserved polar residues can be identified within the ligand-binding channel, including $Arg^{35}$ and $Glu^{36}$ emanating from subunit B and $Asn^{17}$ and $Tyr^{23}$ that project from the adjacent subunit D (**Figure 4.**7). The carbonyl group of polyene **3** interacts directly with the side chains of residues $Arg^{35}$ through hydrogen bonds, and indirectly with residue $Asn^{17}$ through a water-mediated hydrogen bond.

### 4.3.6 Probing the catalytic mechanism by site-directed mutagenesis on CalE7

With the absence of an acidic residue Glu or Asp in the immediate lining of the binding pocket, CalE7 must catalyze the hydrolysis of the thioester bond using a novel mechanism from other hotdog fold thioesterases (refer to **Figure 4.18**). The binding pocket mainly consists of hydrophobic residues with the exception of five conserved polar residues, namely $Asn^{19}$, $Asn^{23}$, $Tyr^{29}$, $Arg^{37}$ and $Thr^{60}$. Five additional polar residues ($Glu^{17}$, $Cys^{36}$, $Glu^{38}$, $Glu^{70}$ and $Asp^{75}$) in the vicinity of the substrate-binding channel are also well conserved among CalE7 homologs (**Figure 4.11**).

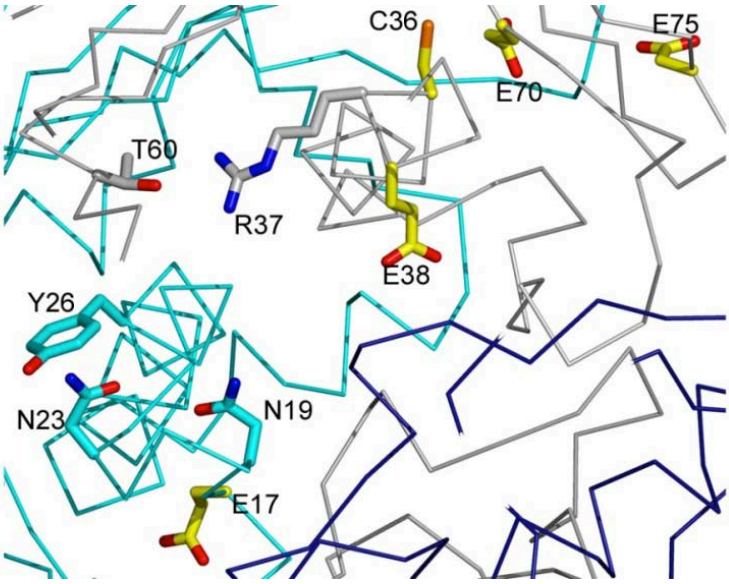



**Figure 4.11** Location of the five conserved residues (shown in *yellow*) outside of the substrate-binding pocket that were examined by site-directed mutagenesis and kinetic study.

Site-directed mutagenesis and enzymatic assay were carried out to examine the roles of the ten residues in catalysis. Incubation of CalE8 and CalE7 mutant with substrates at pH 8.2 produced carbonyl-conjugated polyene **3** and its geometrical isomers as major products, whereas the mutants generated various amount of product relative to the wild-type CalE7 (**Figure 4.12**). For the five residues located in the pocket, the mutation N23A had a negligible effect on product yield. The greatest effect was observed for the mutation R37Q or R37K, which completely abolished the enzymatic activity with negligible product formation. The mutations T60A and Y29F have significant impact by reducing the yield to 17 and 28%, respectively. Meanwhile, the mutation N19A seems to facilitate the formation of carbonyl-conjugated polyene **3** with an increase in the relative yield of more than 300% compared to CalE7 wild type. The enhancement on product formation by N19A was extremely prominent for carbonyl-conjugated polyene **3** at pH 7.0. For the five conserved residues located outside the binding pocket, the mutation E17Q caused a notable decrease of product formation (80%), whereas mutations C36A, E38Q, E70Q, and D75N only caused moderate decreases in product yield. The effect of the E17Q mutation was surprising given that the side chain of the distal $Glu^{17}$ is distant from the substrate-binding pocket (**Figure 4.11**).

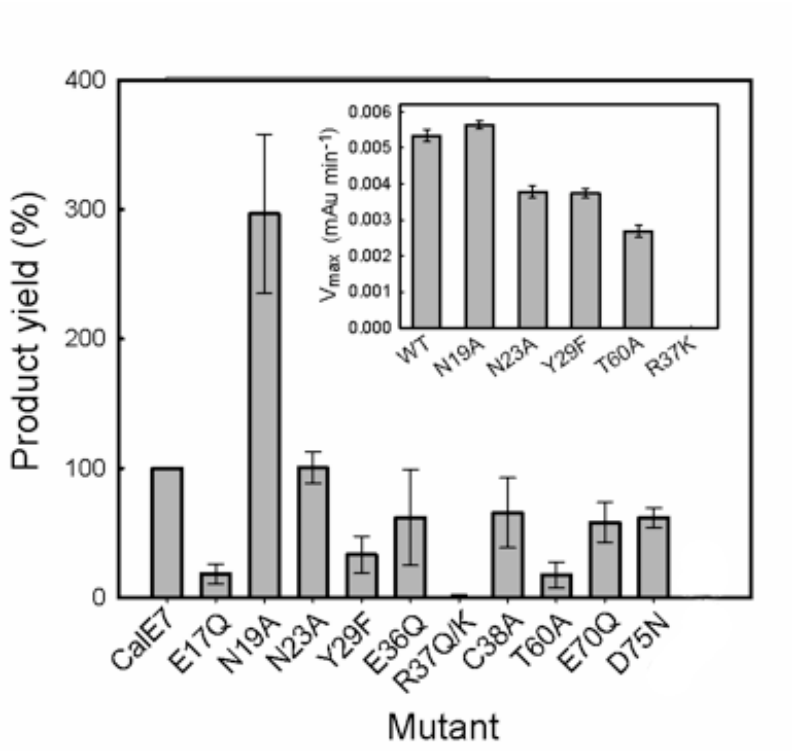



**Figure 4.12** Enzymatic activity of the wild type and mutant CalE7. Reaction conditions are described in the experimental section and the relative yield was calculated based on the total peak area of the product (**3**) and its isomers obtained from HPLC chromatogram. (*Inset*) Comparison of $V_{max}$ for CalE7 and its mutants. The reactions were followed by monitoring the formation of the products at 420 nm using a UV-Vis spectrophotometer. The catalytic activity for the R37Q and R37K mutants was too low to be measured.

The E17Q mutant was found to be a dimer in solution by size-exclusion chromatography, whereas all the other mutants remain as tetramer. From the structure, it can be seen that $Glu^{17}$ is located at the tetrameric interface of CalE7 and surrounded by positively charged residues (**Figure 4.13**). Hence, the effect of the E17Q mutation is most likely due to the structural perturbation at the tetramer interface.

The activities of wild type and the five mutants lining the catalytic cavity were further characterized by steady-state kinetic measurement. The effect of the mutation on enzymatic rate for the five residues in the substrate pocket was examined by following product formation using absorption spectroscopy. The effect

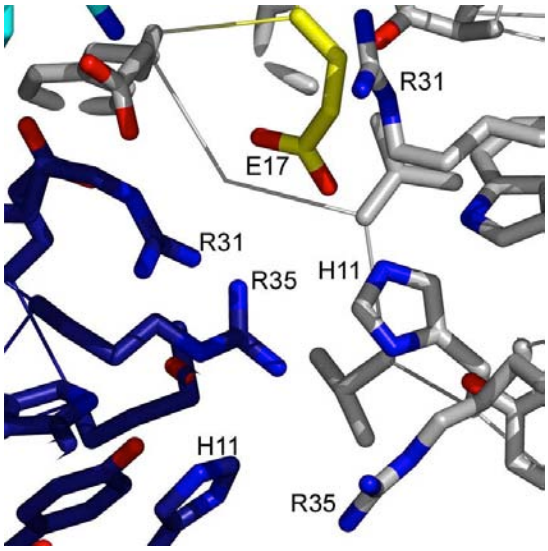



**Figure 4.13** Surrounding residues of $Glu^{17}$ at the tetramer interface. $Glu^{17}$ is shown in *yellow*. Residues from the same subunit as $Glu^{17}$ are shown in grey while the residues from the adjacent dimer are shown in dark *blue*

of the mutation on Vmax for the mutants generally correlates well with the effect on product yield (**Figure 4.12** ***inset***). Not surprisingly, the greatest effect was observed for mutation R37Q or R37K, which displays negligible activity even at high enzyme concentration. The T60A mutant exhibited a ~2- fold reduction in Vmax relative to the wild type enzyme which is probably due to the suppressed formation of carbonyl-conjugated polyene **3**; whereas the N19A mutant showed a 1.2-fold greater Vmax than that of the wildtype CalE7. The mutations Y29F and N23A caused similar decreases (~1.4-fold) in catalytic rate. Together, the data suggested that Arg37 is the only essential residues for catalysis.

### 4.3.7 Mutagenesis studies on $Arg^{35}$ and $Glu^{36}$ in DynE7

To assess the importance of arginine and glutamine in catalysis, we examine the catalytic efficiency of the wild type, R35A and E36A single mutants of DynE7. An enzymatic assay was carried out to monitor the rate of consumption of NADPH in reactions involving DynE8 and the two DynE7 mutants. The results obtained were congruent with that of CalE7 mutations. The mutant R35A abolished the

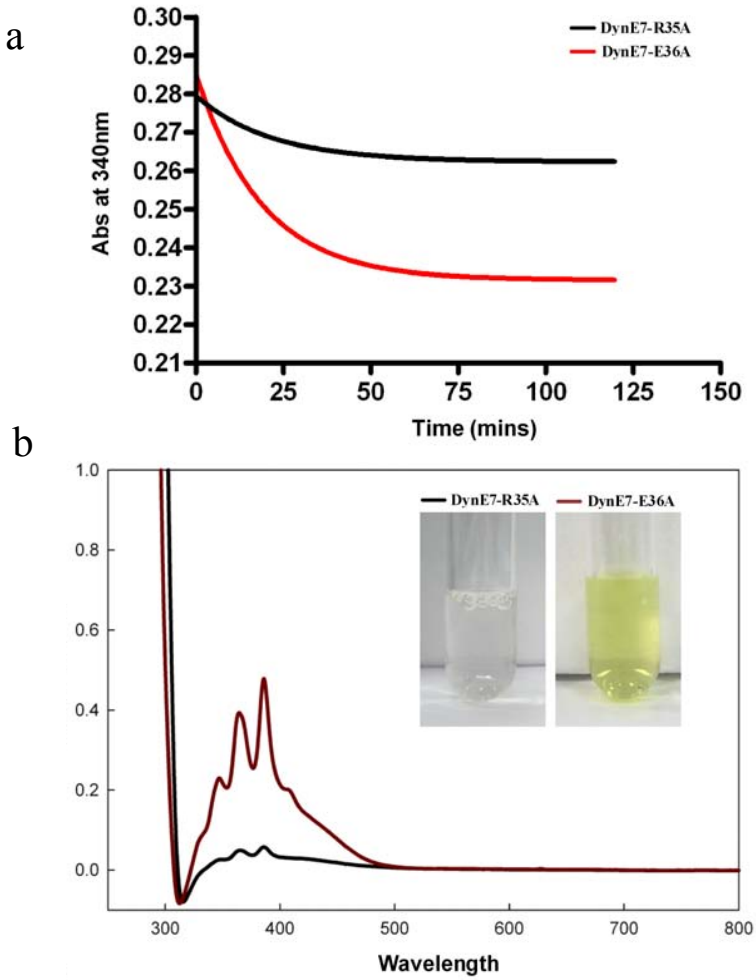



**Figure 4.14** Heterologous co-expression in *E. coli* of the DynE8 protein with either DynE7-R35A or the DynE7-E36A mutant. (**a**) Enzymatic assay of DynE8 with DynE7 mutants. R35A mutant shows negligible activity in terms of NADPH consumption, leading further support to its critical role in thioester cleavage. The E36A mutant displays a much higher activity since the rate of NADPH consumption is much faster. (**b**) The bright yellow color of the protein solutions indicates the present of linear conjugated polyene. UV-Vis absorption spectra of R35A and E36A mutant. The protein concentration is ~ 15 mg/mL.

catalytic activity completely, leaving NADPH levels largely unchanged at the end of the reaction. On the other hand, the E36A mutant displayed a considerable level of NADPH consumption, which implied that the catalytic capability of the mutant was largely retained (**Figure 4.14a**). In addition, by the co-expression of DynE8 and the DynE7 mutants in *E. coli*, we examined *in vivo* product formation to further assess the importance of these two residues. As shown in **Figure 4.14b**, the comparison of the absorption spectra of the purified DynE7 mutants indicates that

the E36A mutant is capable to release the polyenes with a slightly lower efficiency compares to wild-type, whereas the protein solution of the R35A mutant appears colorless with a negligible absorbance around 400 nm, indicating that this mutation essentially abolishes the enzymatic activity of DynE7 for hydrolytic release of the products from DynE8. These observations are fully consistent with in vitro enzymatic assays carried out for CalE7[87].

## 4.4 Discussions

Thioesterases associated with fatty acid synthase (FAS), PKS and non-ribosomal peptide synthase (NRPS) generally belong to the so-called type I and II thioesterases. On the other hand, hotdog fold thioesterases are mainly known to utilize acyl-CoA as substrates, with only a few known examples that act on peptidyl carrier protein (PCP) or ACP-tethered acyl substrates[85; 148; 149], including the EntH protein (or YbdB) involved in the biosynthesis of the nonribosomal peptide-derived enterobactin[148; 150; 151]. Hence, CalE7, SgcE10 and DynE7 appear to represent unconventional examples where hotdog fold thioesterases are recruited for polyketide synthesis.

### 4.4.1 Induced-fit in DynE7 upon ligand binding

Generally, the overall structure of CalE7, SgcE10 and DynE7 resemble other hotdog fold thioesterases with the εγ oligomeric arrangement. However, with a sharply kinked substrate-binding channel and an unusually large side pocket (**Figure 4.4**), the substrate-binding pocket of CalE7 and SgcE10 in the closed conformation differs significantly from other thioesterases despite the common location of the pocket at the subunit surface. According to the crystal structure of CalE7 and SgcE10, the substrate-binding channel has been shortening by several conformational changes, involving residues $Leu^{133}$ and $Phe^{137}$ from helix α3 in CalE7 ($Phe^{40}$, $Phe^{93}$, $Leu^{133}$ and $Leu^{137}$ in SgcE10). With the unliganded structure of CalE7 (or SgcE10) and the structure of ligand-bound DynE7, the binding of the ACP-tethered substrate requires several conformational changes in the secondary structural elements.

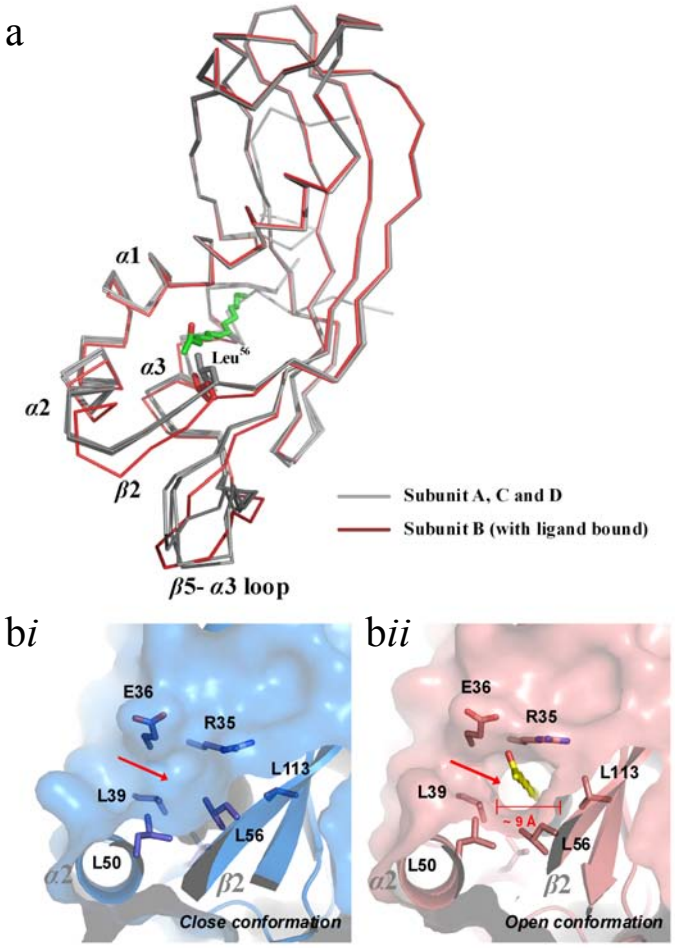



**Figure 4.15** Superposition of the four monomers for the ligand-bound DynE7 tetramer. (**a**) Conformational changes are visible for the β2 strand, α2 helix and the β5-α3 "gating loop. A shift of the $Leu^{56}$ side chain that projects from the β2 strand of subunit B leading to the opening of the substrate binding channel is visible. Closed (monomers A, C, D) (**bi**) vs open conformation (monomer B) (**bii**) of the ligand binding channel. The red arrows indicate the entrance of the binding channel.

In the crystal structure of the ligand-bound DynE7, the ligand-binding channel of monomer B differs significantly from the other non-bound monomers (labeled A, C and D) (**Figure 4.15a**). In contrast to monomer B, the other three monomers adopt a closed conformation with the hydrophobic segment of the channel completely shielded from the solvent (**Figure 4.15b**, ***left panel***), and with the hydrophilic segment partially accessible to solvent. When the ligand-bound monomer is superimposed onto the other three, r.m.s.d values for Cα atoms of ~0.77 Å are returned, about twice the corresponding values for pairwise

**Table 4.4** R.M.S.D for α-carbon atom between monomers after superposition

| Superposition | Primary Superposition | | | Residual Shift r.m.s.d (Å) | | |
|---|---|---|---|---|---|---|
| | Reference | Target | r.m.s.d (Å) | *α*2 (45-50) | *β*2 (55-66) | *β*5-*α*3 loop (120-132) |
| **Ligand-bound DynE7**[a] | B | A | 0.77[+] | 0.18 | 2.19 | 1.24 |
| | | C | 0.78[+] | 0.33 | 2.19 | 1.10 |
| | | D | 0.76[+] | 0.27 | 2.20 | 0.77 |
| | A | C | 0.39[+] | 0.06 | 0.06 | 0.47 |
| | | D | 0.37[+] | 0.07 | 0.04 | 0.42 |
| | C | D | 0.48[+] | 0.09 | 0.05 | 0.96 |
| **Ligand-bound and unliganded DynE7**[b] | **Unliganded-DynE7 (Tetramer 1)** | **Ligand-bound DynE7** | 0.88[#] | | | |
| | A | A | 0.61[&] | 0.64 | 0.31 | 0.45 |
| | B | B | 0.56[&] | 0.72 | 2.17 | 0.97 |
| | C | C | 0.68[&] | 0.61 | 0.19 | 0.59 |
| | D | D | 0.76[&] | 0.69 | 0.28 | 0.50 |
| | **Unliganded-DynE7 (Tetramer 2)** | **Ligand-bound DynE7** | 0.86[#] | | | |
| | E | A | 0.55[&] | 0.65 | 0.21 | 0.45 |
| | F | B | 0.56[&] | 0.74 | 2.13 | 1.03 |
| | G | C | 0.69[&] | 0.60 | 0.25 | 0.57 |
| | H | D | 0.76[&] | 0.62 | 0.23 | 0.60 |

[a] Conformational differences within the ligand-bound DynE7 tetramer were first evaluated by superposing all α-carbon atoms of individual monomers[+], followed by a second superposition of each of the three secondary structure elements that undergo significant conformational changes (labeled "residual shifts"). The residue range used for the latter superposition is listed.

[b] Conformational changes between unliganded- and ligand-bound DynE7 were first evaluated by superposing the α-carbon atoms of the ligand bound DynE7 tetramer onto the corresponding atoms of each of the two unliganded DynE7 tetramers (Tetramer 1 and Tetramer 2). These overall superimpositions are referred to as "primary superposition". After this rigid body superposition of the entire tetramer[#], pairwise comparisons of each monomer was performed in turn[&].

comparisons between non-bound monomers (**Table 4.4**). These observations strongly suggest structural remodeling within the DynE7 protein upon binding of the ACP domain-tethered substrate. In its ligand-bound state, several secondary structure elements of DynE7, such as helix α2, β2 strand and the β5- α3 loop, are displaced (**Figure 4.15a** and **Table 4.4**). First, movement of the β5- α3 loop, with the r.m.s.d values of ~0.77 – 1.24 Å, results in the opening of the hydrophilic segment to incoming substrate. The β5-α3 loop may therefore act as a "gate" granting access to the substrate-binding channel. Second, the β2 strand of subunit B undergoes a 2.0 Å shift coupled with a minor translation of the neighboring helix α2. A ~0.3 Å displacement of the side chains of $Leu^{56}$ on strand β2 and $Leu^{39}$ on

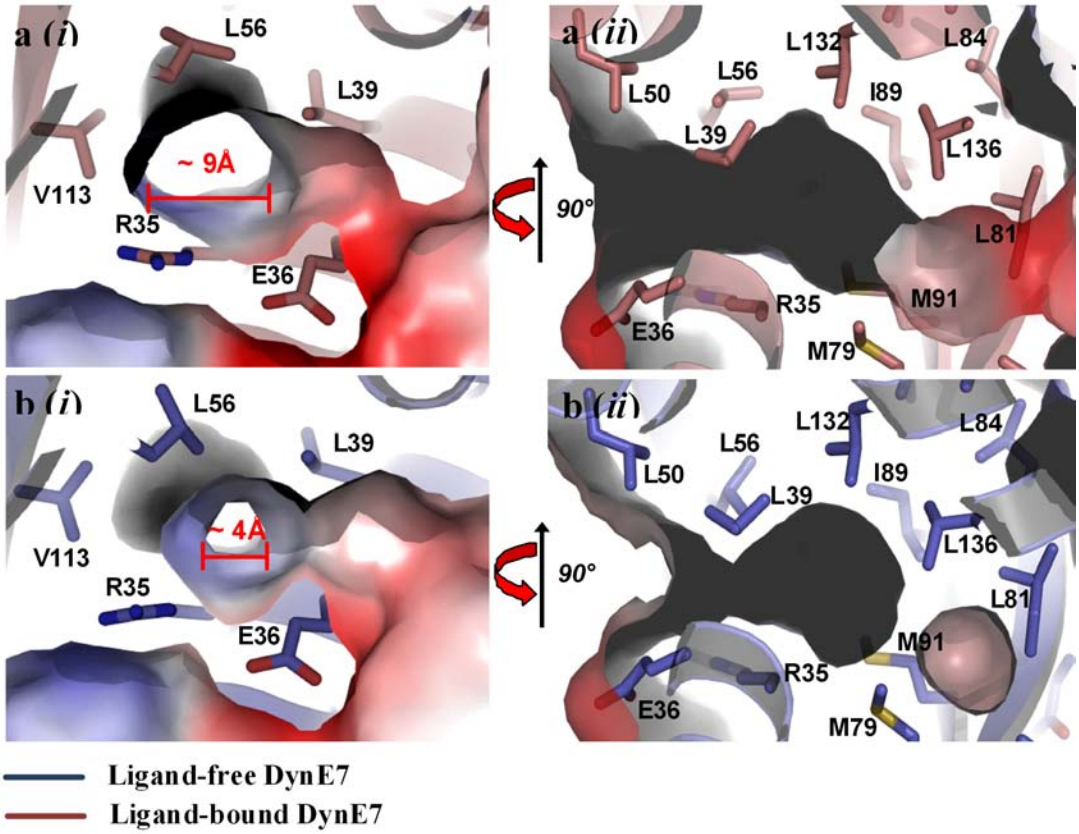



**Figure 4.16** Various conformations adopted by the ligand binding channel of DynE7 tetramers. (**a**) The ligand bound DynE7 tetramer. (**i**) A translation of the side chain of Leu[56] enlarges the entrance of the tunnel. (**ii**) The length of the binding channel is determined by positioning of the side chain of Leu[136]. (**b**) The unliganded DynE7 has a narrower entrance (**i**) and a shorter binding channel (**ii**) compared to ligand-bound DynE7.

helix α1 further expands the entrance of the inner channel to grant access to the hydrophobic section of the channel. In addition, a breathing motion of the hydrophobic residues that line the channel for binding the hydrocarbon chain through Van der Waals interactions is observed (**Figure 4.15**a). Third, shifts of Leu[136] and of other residues that project from the C-terminal helix α3 lead to an increase in the length of the channel and opening at the end of the tunnel. Together, these remarkable movements throughout the channel lead to the formation of a wider and longer channel with an approximate diameter of 9 Å and a total length of ~30 Å. This expanded hydrophobic tunnel features a greater diameter and length able to accommodate the carbonyl conjugated polyene that spans about 19 Å

(**Figure 4.16a**). These observed conformational changes suggest that one DynE7 tetramer can only bind one ACP domain-tethered substrate at a time with a 4:1 stoichiometry. This is because once a channel is bound, the other channels appear to close, hence preventing further ligand binding (**Figure 4.15**). These such allosteric effects caused by substrate binding were observed in the hot-dog fold TE Paal[136]. In the case of Paal, binding of two acyl- CoA molecules induces a cooperative rigid-body rearrangement to prevent further binding at the two other pockets.

### 4.4.2 Interaction between DynE7 and product-tethered PKS

The structural analysis shows that the precursor of the carbonyl-conjugated polyene **3** is covalently attached to the phosphopantetheinyl group of the ACP domain of DynE8 (**Figure 4.1**). Efficient cleavage of the product presumably requires a direct interaction between DynE7 and the ACP domain of DynE8. Initial pull-down experiments to test whether the wild type CalE8 and CalE7 proteins form a stable protein complex suggest their interaction is at most weak or transient (not shown). Despite the lack of specific protein-protein interaction, the ACP domain of DynE8 is expected to be located at the entrance of the substrate-binding channel. Modeling the phosphopantetheinyl group in the channel shows that it can fit in the hydrophilic segment of the "L" shaped substrate-binding channel (**Figure 4.17**). Considering the conformational change induced by ligand binding, the current model constitutes an improvement compared to models based on the structure of the ligand-free CalE7[87] and SgcE10. Importantly, the length of the phosphopantetheinyl group closely matches that of the channel in such as way that the phosphate group of the phosphopantetheinyl group is positioned at the edge of channel entrance (**Figure 4.17**). This modeling also positions the $Arg^{35}$, but not $Glu^{36}$, near the thioester carbonyl group, consistent with their respective roles in the proposed catalytic mechanism. Nonetheless, we can not completely exclude the possibility that the binding mode for the PKS-tethered substrate might differ slightly from the binding mode of carbonyl conjugated polyene **3** given the formation of additional interactions between the phosphopantetheinyl moiety and the TE. For this reason, we are cautious of proposing a detailed catalytic mechanism by defining the precise roles of the putative catalytic residues. Examination of the electrostatic surface of

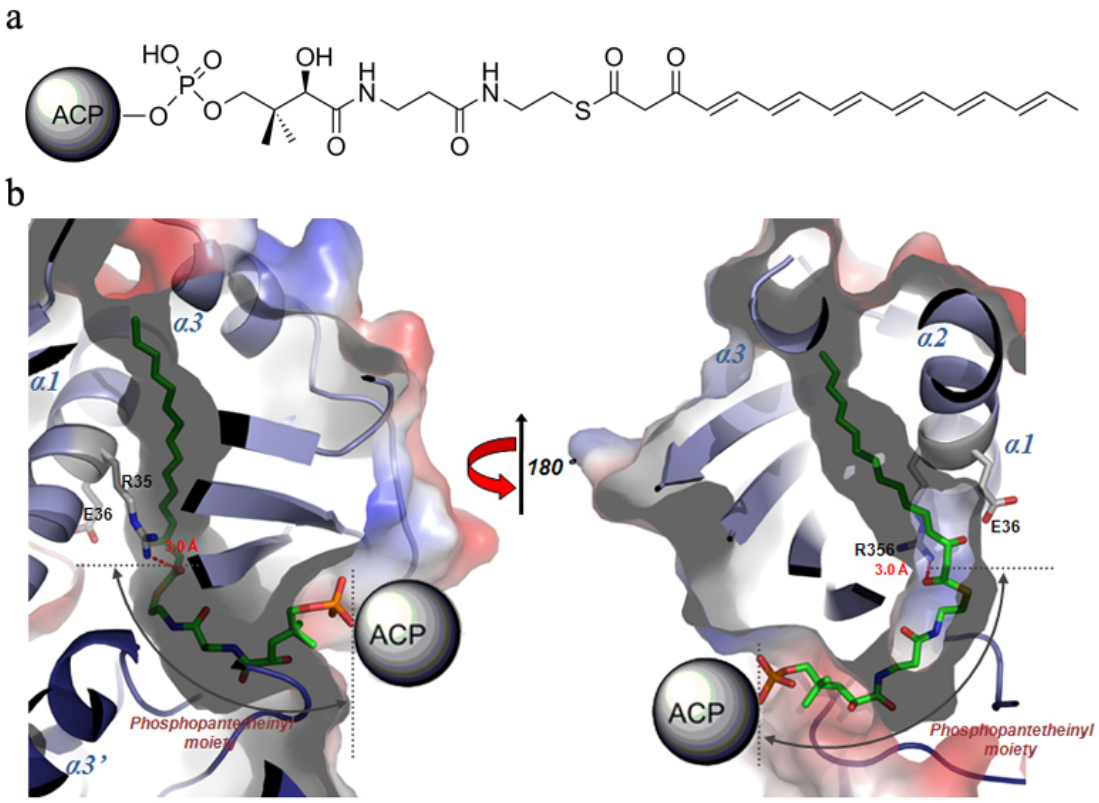



**Figure 4.17** Modeling of the phosphopantetheinyl-linked carbonyl-conjugated polyene (a) into the L-shaped binding channel of ligand-bound DynE7. (b) The phosphopantetheinyl-moeity fits into the outer binding channel, keeping the carbonyl-conjugated at its experimental position. The DynE7-ACP domain is proposed to be located adjacent to the dimer interface.

CalE7 and DynE7 also reveals rather different charge distribution near the entrance of the channel. Given the observation that there is no significant formation of protein- protein complex between CalE7 and CalE8, the lack of conserved electrostatic surface at the channel entrance seems to suggest that the TEs most likely bind to the ACP domain tethered product primarily by recognizing the hydrophobic polyene. This hypothesis is in agreement with the observation that helix II, the universal "recognition helix" from the ACP domains of CalE8 and DynE8 have very different charge character.

### 4.4.3 Catalytic mechanism of DynE7 and its homolog

Most known hot-dog fold TEs utilize acyl-CoA substrates, rather than acyl or peptidyl carrier protein substrates. Hence, the TEBC subfamily seems to represent an uncommon family of hotdog fold TEs that have been recruited for the release of polyketide products from the PKS proteins. Furthermore, most hot-dog fold TEs that have been functionally characterized such as 4-HBT from *Pseudomonas* sp. Strain CBS-329[66] and *Arthrobacter* sp. strain SU[141] use an acidic residue (Glu or Asp) as a nucleophile or general-base catalyst. Although the central helix α1 contains a conserved Glu residue ($Glu^{38}$ in CalE7, $Glu^{40}$ in SgcE10 and $Glu^{36}$ in DynE7), it was found to be dispensable for thioester hydrolysis, and thus, an identical mechanism is unlikely to be at work. The unnecessary of an acidic residues as nucleophile or general base was recently reported for the thioesterase FcoT from *Mycobacterium tuberculosis*[137], an acyl-CoA thioesterase that activates the thioester bond and stabilizes the oxyanion intermediate via the side chains of $Tyr^{33}$, $Tyr^{66}$ and $Asn^{74}$. No such residues can be found in CalE7/SgcE10/DynE7 for either bond activation or intermediate stabilization. Therefore, DynE7 is likely to utilize a catalytic mechanism different from FcoT.

With information on the catalytic mechanism of conventional thioesterase, the search for the catalytically critical residues in DynE7 began with acidic amino acids located in the vicinity of the substrate-binding channel. As previously noted for CalE7, every member of the TEBC family possesses a strictly conserved arginine residue located on the central helix α1. Our comprehensive kinetic study conducted on a series of ten single CalE7 mutants revealed that four conserved residues ($Asn^{21}$, $Arg^{39}$, $Glu^{40}$ and $Thr^{62}$ in SgcE8; $Asn^{17}$, $Arg^{35}$, $Glu^{36}$ and $Thr^{58}$ in DynE7) may play various roles in catalysis, based on the impact of the mutations on product formation[87]. This mutagenesis study on CalE7, as well as DynE7, pointed to $Arg^{37}$ ($Arg^{35}$ in DynE7) as a crucial catalytic residue for hydrolysis of the thioester bond. An essential role of the Arg residue for catalysis is supported by the abrogation of the enzymatic activity by the Arg to either Lys or Gln mutation in CalE7[87].

Based on our structural and biochemical data on both CalE7 and DynE7, we propose a catalytic mechanism for the thioester hydrolysis. The transient interaction between hotdog fold TE and product-tethered PKSE induces major conformational changes in the overall structure of the thioesterase. Eventually, the conformational changes in the β5-α3 loop and helix α3 generates wider and longer substrate-binding channel. Immediately after insertion into the catalytic chamber, the substrate is anchored within the channel by extensive hydrophobic interactions between the polyene segment and the side chains of the hydrophobic residues lining the cavity. In accordance with the ligand-bound structure of DynE7, this binding configuration places the carbonyl oxygen of the thioester group within hydrogen-bond distance of $Arg^{35}$ ($Arg^{37}$ in CalE7). As for $Asn^{17}$ ($Asn^{19}$ in CalE7), a water molecule is utilized to mediate the interaction between the carbonyl oxygen of the thioester and the amide group of Asn. The keto group of the bound ligand **3** in the crystal indeed interacts with $Arg^{35}$ and $Asn^{17}$ ($Arg^{37}$ and $Asn^{19}$ in CalE7) through hydrogen bonds and water mediated hydrogen bonds respectively in place of the carbonyl group of the thioester (**Figure 4.7**).

In the proposed catalytic mechanism, the nucleophilic attack of the thioester carbonyl by a water or hydroxide anion is facilitated with $Arg^{35}$ ($Arg^{37}$ in CalE7 and $Arg^{39}$ in SgcE10). $Arg^{35}$ is hypothesized to stabilize the oxyanion of the tetrahedral intermediate formed by the attack of a water molecule or nucleophilic hydroxide ion (**Figure 4.18**). This mechanism is consistent with the presence of several water molecules in the hydrophilic section of the substrate-binding channel of DynE7 that holds the phosphopantetheinyl group and thioester bond. The hydroxide ion may come directly from the bulk solvent. This hypothesis is consistent with the observation that alkaline pH seems to favor product formation[87]. In the typical hot-dog fold TEs, a general-base catalyst is required for the deportation of water to generate the hydroxide ion. A general base catalyst may not be necessary for DynE7 because of the polarization and activation of the C=O group by $Arg^{35}$. Although a potential role of the $Glu^{36}$ or $Tyr^{27}$ as general-base catalyst in the generation of the hydroxide ion can not be completely ruled out, current structural and biochemical data are more supportive of a mechanism in which $Glu^{36}$ plays a

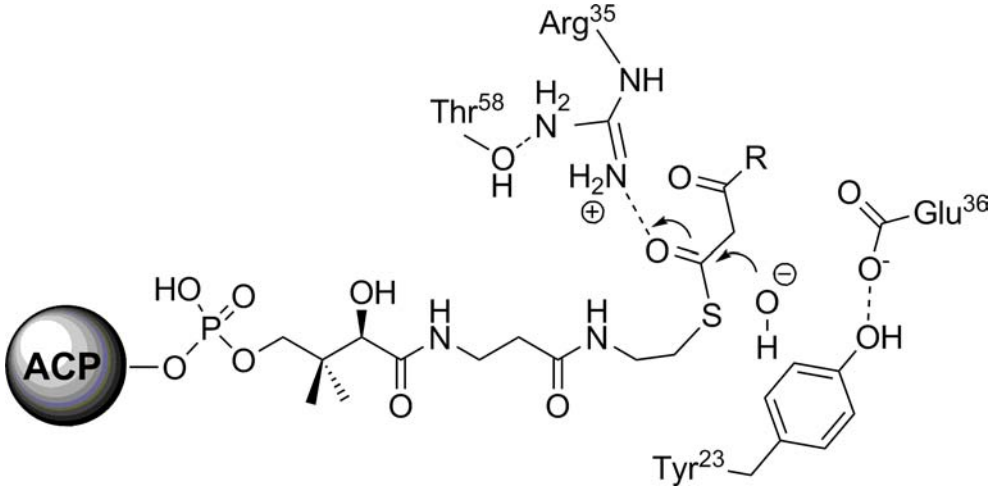



**Figure 4.18** Proposed hydrolytic mechanism for DynE7 with the four polar residues affecting catalytic efficiency shown. The essential residue Arg35 stabilizes the oxyanion of the tetrahedral intermediate formed by the attack of hydroxide ion.

secondary role for catalysis, considering the position of the residue and the relatively minor effect of the mutation on catalytic efficiency. Despite a similar position in the central helix, $Glu^{36}$ is also unlikely to play an equivalent role as the essential acidic catalytic residues in other hot-dog fold TEs such as PaaI. Based on the structure of the DynE7-product complex, the estimated distance between the scissile thioester bond and $Glu^{36}$ seems to be too long for a direct involvement of $Glu^{36}$ as a nucleophile or general-base catalyst. However, such deduction remains speculative because the distance is inferred from the structure of an enzyme-product complex and not from an enzyme-substrate complex.

In the case of carbonyl conjugated polyene **3** formation, a facile decarboxylation of the β-keto carboxylic acid could occur with an intramolecular proton transfer between the carbonyl and carboxylic groups within a six-centered transition state. The role of CalE7/DynE7 in the decarboxylation step of the β-keto carboxylic acid remains unclear. A potential proton acceptor might be present in the catalytic channel to assist the decarboxylation of the β-keto carboxylic acid. Eventually, the carbonyl-conjugated polyene **3** is generated after the immediate tautomerization of the enol intermediate with a water molecule in the side pocket acting as proton donor. The formation of polyene **4** is considerably more

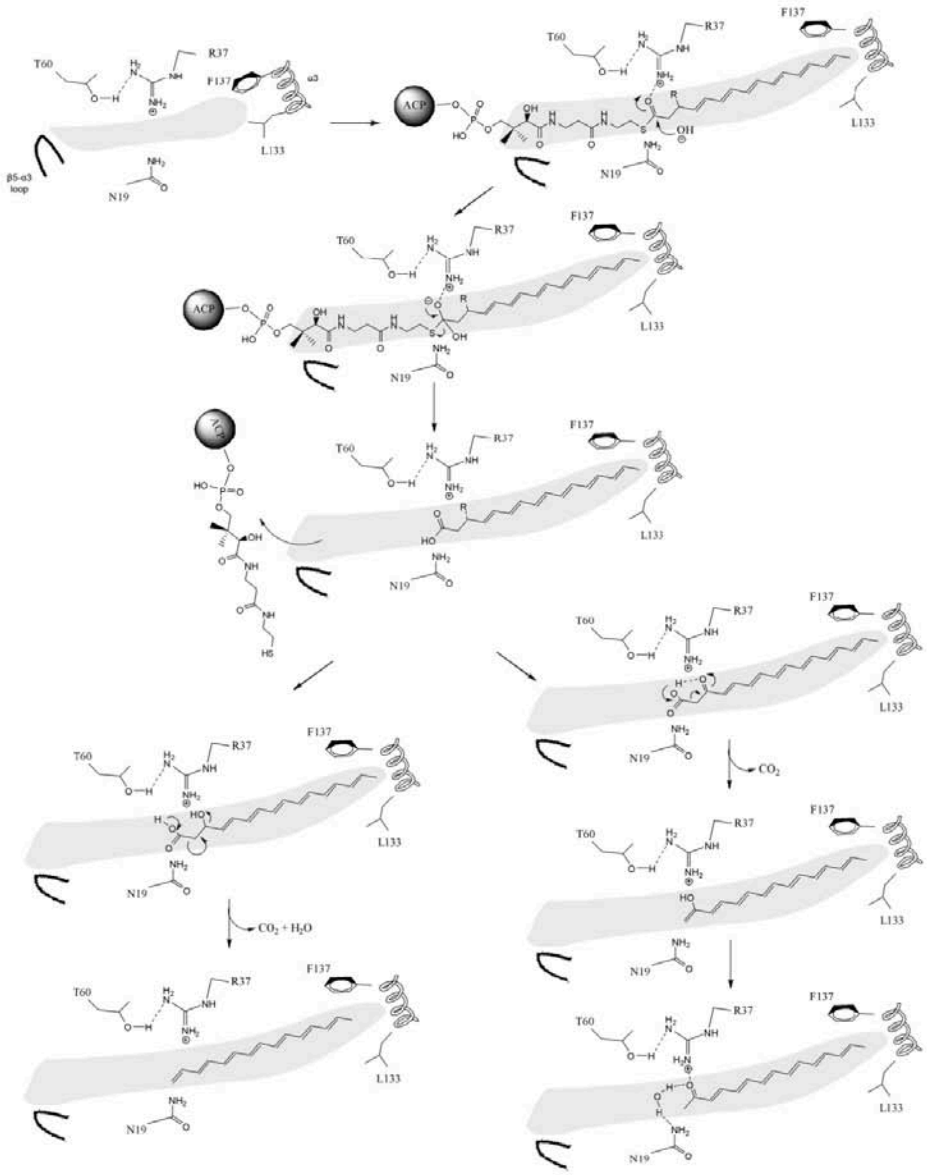



**Figure 4.19** Proposed catalytic mechanism based on CalE7/DynE7 structure. The substrate-binding channel is represented by the grey shade. In the hydrolysis step, the essential residue $Arg^{37}$ ($Arg^{35}$ in DynE7) functions as an oxyanion hole. The step following hydrolysis represents the breakage of the thioester bond. The decarboxylation occurs with an intramolecular proton transfer within the six-centered transition state, and with the tautomerization to form product **3**. As for **4**, an additional step of dehydration occurs to yield the conjugated polyene **4**. Whether the decarboxylation and dehydration steps are assisted by CalE7 is not clear. The last step shows the repositioning of the final product in the channel in accordance with the data obtained from the DynE7-ligand (**3**) complex crystal structure.

complicated compared to that of carbonyl conjugated polyene **3**. The mechanism for decarboxylation *en route* to the formation of polyene **4** might be similar to that of carbonyl conjugated polyene **3**. In addition to the decarboxylation, a dehydration step ensues the elimination of the β-hydroxyl group together with a hydrogen atom from the α-carbon of the thioester to form the final double bond in the polyene **4** (**Figure 4.19**). The presence of a proton acceptor might be imperative for either the decarboxylation or the dehydration step. Nonetheless, the mechanism for dehydration is still beyond our comprehension and whether the decarboxylation and dehydration steps are taking place simultaneously or sequentially is not well understood either.

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

With respect to the mutagenesis data, the role of $Thr^{60}$ in CalE7 catalysis presents yet another mystery to be solved. The T60A mutation resulted in the preferential production of conjugated polyene **4**, indirectly ruling out the possibility that the hydroxyl side chain functions as the nucleophilic group attacking the thioester. Instead, $Thr^{60}$ is most likely to play a secondary or structural role in catalysis, by positioning the side chain of $Arg^{37}$ through hydrogen-bonding. The mutation of $Thr^{60}$ to alanine probably alters the conformation of the side chain of

Arg[37], leading to the positioning away of the critical amine groups from the catalytic channel. However, due to the inherent flexibility of the β -hydroxyl intermediate, bond rotation at C2 and C3 positions is still able to place the carbonyl group of the thioester within the vicinity of Arg[37]. On the other hand, the abstraction of the highly acidic hydrogen on α -carbon may result in the tautomerization of the β -keto intermediate into the corresponding enol form. Therefore, the rotational flexibility around C2 and C3 position is entirely obliterated. The increased distance between Arg[37] and the carbonyl group of the thioester intermediate due to keto-enol tautomerism would lead to the drastic reduction in the formation of carbonyl conjugated polyene **3** (**Figure 4.20**).

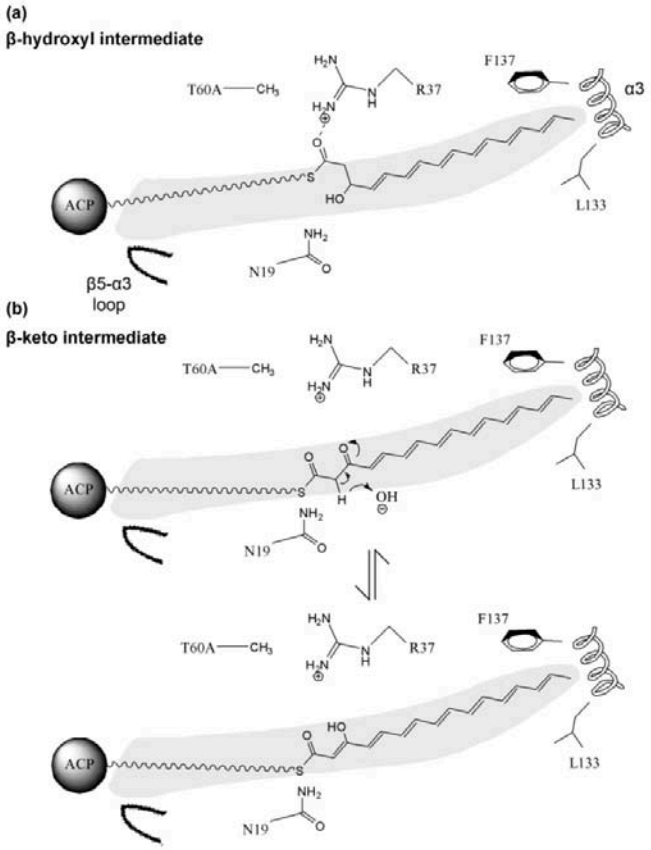



**Figure 4.20** Proposed mechanism of mutation T60A on thioester hydrolysis. (**a**) The conformational changes in β-hydroxyl intermediate still allow the hydrolysis of thioester to occur. (**b**) The keto-enol tautomerism makes the carbonyl oxygen of the thioester in β-keto intermediate difficult to reach by Arg[37].

The exact function of the essential residue $Arg^{37}$ ($Arg^{35}$ in DynE7) remains to be fully established in the future. Current structural and biochemical data do not distinguish between the Arg acting as transition state stabilizer or general base, even though the latter is less likely. We propose that hydrolysis may take place through a direct attack of a hydroxide anion, similar to the mechanism proposed for $FcoT^{137}$. Since the rate of such nucleophilic reaction depends on the concentration of the hydroxide anion, this is to some extent in line with the observation that high pH favored product formation. Although it is rare for Arg to act as general base in enzymes due to its high p*Ka*, we could not totally rule out the possibility that the water molecule could be activated by the neutral form of $Arg^{37}$ ($Arg^{35}$ in DynE7) if the p*Ka* of the Arg is drastically lowered by several pH units in CalE7/DynE7.

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

# CHAPTER 5 Conclusions and future directions

## 5.1 Conclusions

The biosynthetic origin of the enediyne warheads has been at the center of speculation and discussion since the unveiling of the first enediyne structure. The unique structural and chemical properties of the enediyne warheads, together with a myriad of diverse peripheral moieties attached to them, entail an entire array of novel mechanisms involved in the biosynthesis pathway. Among all the advancement, the identification and the complete sequencing of the enediyne gene clusters revitalized the field and offered numerous research opportunities.

Identification of the genes encoding PKSs within the clusters from various enediyne-producing bacteria not only confirms the postulated polyketide origin of the enediyne warheads, but also indicates that the bicylo-[7.3.1.]-tridecadiynene and bicylo-[7.3.0]-dodecadienediyne cores share similar biosynthetic pathways. From a series of *in vitro* experiments conducted by me and former PhD student, recombinant enediyne PKSs produced a wide range of *in vitro* products of different chain lengths, with the conjugated polyene **4** as the exclusive product under certain experimental conditions. Due to the high structural similarity between the carbonyl conjugated polyene **3** and conjugated polyene **4**, the intermediate *en route* to the biosynthesis of enediyne core remains controversial, in spite of the observation from our DynE7 crystal structure[95]. The possibility that both carbonyl conjugated polyene **3** and conjugated polyene **4** are aberrant products could not be completely ruled out either. Since the likelihood of CalE7/SgcE10/DynE7 functioning as a type II thioesterase still exits, a conceivable scenario is that the function of the thioesterase is to remove the derailed products and that the authentic intermediate remains anchored to the PKS for downstream oxidation and cyclization to furnish the enediyne core.

By intervening the critical steps of the polyketide synthesis at the PKS stage, the regulatory factors may function to direct the PKSs to produce the authentic biosynthetic intermediate. However, it is still no clear which gene may play the role as the regulator for the enediyne biosynthetic pathway. From the bioinformatics

studies on the gene clusters, CalU15/SgcE3/DynU15 from UNBL family is speculated to be an acetylenase responsible for the formation of the –yne groups. Nonetheless, no positive results were obtained from all our experiments designed to probe the function of CalU15. We suspect that CalU15 lacks the activity of acetylenase could be due to the absence of the cognate electron donor in the assay system. Even after various optimization and reconstitution of the experimental condition, CalU15 remained inert for the enediyne biosynthesis. Meanwhile, the crucial cyclase that regulates the different folding pattern of enediyne precursors, presents another missing link in the biosynthesis pathway. In this study, only two of the five conserved genes *PKSE* (encoding PKS) and *TEBC* (encoding TE) are well characterized, while the elusive *UNBL* is still under intensive study. The other two genes, *UNBU* and *UNBV*, stand as insuperable even at the stage of protein expression, not to mention any structural and biochemical characterizations. In the protein structural perspective, we successfully crystalized one of the PKSE full-length protein, CalE8, from the co-expression system. Sadly, the CalE8 crystals did not show any X-ray diffraction spot. In spite of our continued attempts in protein crystallization optimization, the CalE8 crystals remain reluctant to unveil itself. An alternate way to solve the PKSE structure is to dissect each functional domain into individual expressed protein and it worked out for the AT domain of DynE8.

The novelty in the structure of proteins encoded by the minimal gene cassette is also manifested in the individual domains of the enediyne PKSs. The crystal structure of the AT domain from DynE8, encompassing its entire catalytic AT domain and the adjacent linker domain, unveils a α/β hydrolase and ferrodoxin-like subdomain with the Ser-His dyad located in the cleft between the two subdomains. The linker domain also adopts α/β fold abutting the AT catalytic domain. To our knowledge, this is the first crystal structure for an AT domain from an iterative PKS. The crystal structure provide insight into the evolutionary relationship between AT domain of an iterative PKS and the AT domains of fatty acid synthase (FAS) and modular PKSs. Functioning as a chain elongation, the AT domain catalyze the transfer of the malonyl group from malonyl-CoA to the ACP domain. Interestingly, unlike the AT domain of other iterative PKSs, the AT

domains of the enediyne PKSs do not seem to use acetyl-CoA as a starter unit but solely malonyl-CoA. This differs from the dual specificity exhibited by acyltransferases of mammalian FAS as well as other iterative PKSs. Our structure explains the preference for malonyl-CoA with a conserved arginine orienting the carboxylate group of malonate and several nonpolar residues that preclude α-alkyl malonyl-CoA binding. The observation of malonyl-enzyme complex settles some uncertainties in the enediyne biosynthetic mechanism by suggesting that the terminal acetate unit of the PKS products is likely to be generated by a KS domain catalyzed decarboxylation. Meanwhile, co-crystallization with acetyl- CoA revealed two non-covalently bound acetates generated by the enzymatic hydrolysis of acetyl-CoA that acts as an inhibitor for DynE8. In contrast to the previous belief that the malonyl specific AT domains of enediyne PKSs do not recognize acetyl-CoA, our structural and biochemical data suggest that although $AT_{DYN10}$ does not use acetyl-CoA as starter unit, $AT_{DYN10}$ is able to upload the acetyl group of acetyl-CoA to form an acetyl-enzyme covalent complex that is susceptible to hydrolysis.

With the determination of the structures of CalE7 (or SgcE10) and ligand-bound DynE7 showing an expanded entrance of the inner channel to grant access to the hydrophobic section, the function and catalytic mechanism of the novel type of thioesterase came to light. Unlike other hotdog fold thioesterases, the absence of canonical Glu/Asp for hydrolysis of thioester bond makes CalE7 and DynE7 unique in thioester hydrolytic catalysis. The structural change upon the binding of the PKS product revealed a remarkable induced-fit mechanism for substrate binding. Despite the production of **3** implies that the proteins encoded by *TEBC* family may function as type I thioesterase, the precise role of the TEs as type I or type II TEs in enediyne biosynthesis remains to be established fully.

### 5.2 Future directions

The mystery shrouding the structure and function of the individual domains of PKS is but only a small part of the entire puzzle. Unraveling the secrets of PKS presumably entails the complete structural and functional characterization of the megasynthase itself. Up to date, there are no crystal structure of any full length PKS

that has been determined and deposited. The total elucidation of the structure for iterative enediyne PKS would represent a major breakthrough in the field of PKS research. The crystallization of full length CalE8, DynE8 and SgcE is already under way in our laboratory. The expression and purification processes have been optimized and preliminary crystallization screening on the colorless co-expressed CalE8 has yielded planar-hexagonally shaped crystals. In the future, it is planned to generate a $Ser^{651}$ mutant, which will inevitably inactivate the AT domain of PKS thus maximizes the homogeneity of the PKS protein. On the other hand, co-crystallization of PKS can also be tried out with cognate TE. Given that the R37Q/K mutants of CalE7 have been proven to be completely inactive in thioester cleavage, co-crystallization of R37Q/K with product-tethered CalE8 can be carried out in the screening process since the acyl chain will be stuck inside the cavity of the CalE7 mutants without being cleaved off. As a result, it should be feasible to trap both CalE8 and CalE7-R37Q/K in the crystals because of the constraint imposed by protein-ligand interactions between the two. Moreover, SAX (small angle X-ray diffraction) and EM (electron microscopy) has been carried out by another PhD student, as well to study the overall topology of CalE8. With the existing homologous structures of some of the PKS domains, the structural information of individual domain can be pieced together to generate a structural model for the entire PKS.

Biochemical and structural studies on *UNBL* family protein, CalU15, SgcE3 and DynU15, will continue to be the focus of our research. High throughput screening of reducing agents that may act as the electron donor will be carried out. At the same time, crystallization screening on CalU15 is currently being performed in an effort to deduce the function of the protein from the structural perspective. Although the enediyne biosynthetic mechanism still waits to be unraveled fully, the on-going research in our laboratory as well as in other groups will eventually reveal the secrets locked in the biosynthetic pathway of enediynes.

**References**


1. Calestani, C., Rast, J. P. & Davidson, E. H. (2003). Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* **130**, 4587-4596.

2. Schmalz, H.-G. (2001). Book Review: Comprehensive Natural Products Chemistry. Vol. 1-9. Edited by Sir Derek Barton, Koji Nakanishi, and Otto Meth-Cohn. *Angewandte Chemie International Edition* **40**, 1135-1136.

3. O'Hagan, D. (1991). *The polyketide metabolites*, E. Horwood.

4. Yokoyama, A., Murata, M., Oshima, Y., Iwashita, T. & Yasumoto, T. (1988). Some chemical properties of maitotoxin, a putative calcium channel agonist isolated from a marine dinoflagellate. *The Journal of Biochemistry* **104**, 184-187.

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

6. Birch, A. J., Massy-Westropp, R. A. & Moye, C. J. (1955). Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Australian Journal of Chemistry* **8**, 539-544.

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

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

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

10. Van Lanen, S. G. & Shen, B. (2008). Biosynthesis of enediyne antitumor antibiotics. *Current Topics in Medicinal Chemistry* **8**, 448-459.

11. Jenni, S., Leibundgut, M., Maier, T. & Ban, N. (2006). Architecture of a fungal fatty acid synthase at 5Å resolution. *Science* **311**, 1263-1267.

12. Maier, T., Jenni, S. & Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**, 1258-1262.

13. White, S. W., Zheng, J., Zhang, Y.-M. & Rock, C. O. (2005). The structural biology of type II fatty acid biosynthesis. *Annual Review of Biochemistry* **74**, 791-831.

14. Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B. & Ban, N. (2007). Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-261.

15. Lomakin, I. B., Xiong, Y. & Steitz, T. A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**, 319-332.

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

17. Smith, S., Witkowski, A. & Joshi, A. K. (2003). Structural and functional organization of the animal fatty acid synthase. *Progress in Lipid Research* **42**, 289-317.

18. Katz, L. (2009). Chapter 6 The DEBS paradigm for type I modular polyketide synthases and beyond. *Methods in Enzymology* (David, A. H., ed.), Vol. Volume 459, pp. 113-142. Academic Press.

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

20. Stinear, T. P., Mve-Obiang, A., Small, P. L. C., Frigui, W., Pryor, M. J., Brosch, R., Jenkin, G. A., Johnson, P. D. R., Davies, J. K., Lee, R. E., Adusumilli, S., Garnier, T., Haydock, S. F., Leadlay, P. F. & Cole, S. T. (2004). Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1345-1349.

21. Hitchman, T. S., Crosby, J., Byrom, K. J., Cox, R. J. & Simpson, T. J. (1998). Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. *Chemistry & Biology* **5**, 35-47.

22. Proctor, R. H., Desjardins, A. E., Plattner, R. D. & Hohn, T. M. (1999). A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. *Fungal Genetics and Biology* **27**, 100-112.

23. Hendrickson, L., Ray Davis, C., Roach, C., Kim Nguyen, D., Aldrich, T., McAda, P. C. & Reeves, C. D. (1999). Lovastatin biosynthesis in *Aspergillus terreus*: Characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. *Chemistry & Biology* **6**, 429-439.

24. Fujii. I., Watanabe, A., Sankawa, U., Ebizuka, Y. (2001) Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chemistry & Biology* **8**, 189-97

25. Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpeh, O., Tsai, S.-C. & Townsend, C. A. (2009). Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139-1143.

26. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences* **100**, 15670-15675.

27. Song, Z., Cox, R. J., Lazarus, C. M. & Simpson, T. J. (2004). Fusarin C biosynthesis in *Fusarium moniliforme* and *Fusarium venenatum*. *ChemBioChem* **5**, 1196-1203.

28. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports* **24**, 162-190.

29. Austin, M. B. & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* **20**, 79-110.

30. Nicolaou, K. C., Smith, A. L. & Yue, E. W. (1993). Chemistry and biology of natural and designed enediynes. *Proceedings of the National Academy of Sciences* **90**, 5881-5888.

31. Smith, A. L. & Nicolaou, K. C. (1996). The enediyne antibiotics. *Journal of Medicinal Chemistry* **39**, 2103-2117.

32. Semmelhack, M. F., Gallagher, J. J., Ding, W. d., Krishnamurthy, G., Babine, R. & Ellestad, G. A. (1994). The effect on DNA cleavage potency of tethering a simple cyclic enediyne to a netropsin analog. *The Journal of Organic Chemistry* **59**, 4357-4359.

33. Edo, K., Katamine, S. & Kitame, F. (1980). Naphthalenecarboxylic acid from neocarzinostatin (NCS). *Journal of Antibiotics* **33**, 347-351.

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

35. Lee, M. D., Dunne, T. S., Siegel, M. M., Chang, C. C., Morton, G. O. & Borders, D. B. (1987). Calichemicins, a novel family of antitumor antibiotics. 1. Chemistry and partial structure of calichemicin $\gamma 1^{I}$. *Journal of the American Chemical Society* **109**, 3464-3466.

36. Golik, J., Clardy, J., Dubay, G., Groenewold, G., Kawaguchi, H., Konishi, M., Krishnan, B., Ohkuma, H., Saitoh, K. & Doyle, T. W. (1987). Esperamicins, a novel class of potent antitumor antibiotics. 2. Structure of esperamicin X. *Journal of the American Chemical Society* **109**, 3461-3462.

37. Konishi, M., Ohkuma, H., Tsuno, T., Oki, T., VanDuyne, G. D. & Clardy, J. (1990). Crystal and molecular structure of dynemicin A: a novel 1,5-diyn-3-ene antitumor antibiotic. *Journal of the American Chemical Society* **112**, 3715-3716.

38. Thorson, J. S., Shen, B., Whitwam, R. E., Liu, W., Li, Y. & Ahlert, J. (1999). Enediyne Biosynthesis and Self-Resistance: A Progress Report. *Bioorganic Chemistry* **27**, 172-188.

39. Thorson, J. S., Sievers, E. L., Ahlert, J., Shepard, E., Whitwam, R. E., Onwueme, K. C. & Ruppen, M. (2000). Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research. *Current pharmaceutical design* **6**, 1841-79.

40. Walker, S., Valentine, K. G. & Kahne, D. (1990). Sugars as DNA binders: a comment on the calicheamicin oligosaccharide. *Journal of the American Chemical Society* **112**, 6428-6429.

41. Drak, J., Iwasawa, N., Danishefsky, S. & Crothers, D. M. (1991). The carbohydrate domain of calicheamicin $\gamma 1^I$ determines its sequence specificity for DNA cleavage. *Proceedings of the National Academy of Sciences* **88**, 7464-7468.

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

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

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

45. Boghaert, E. R., Sridharan, L., Armellino, D. C., Khandke, K. M., DiJoseph, J. F., Kunz, A., Dougher, M. M., Jiang, F., Kalyandrug, L. B., Hamann, P. R., Frost, P. & Damle, N. K. (2004). Antibody-targeted chemotherapy with the calicheamicin conjugate hu3S193-*N*-Acetyl γ calicheamicin dimethyl hydrazide targets Lewis$^y$ and eliminates Lewis$^y$-positive human carcinoma cells and xenografts. *Clinical Cancer Research* **10**, 4538-4549.

46. DiJoseph, J. F., Armellino, D. C., Boghaert, E. R., Khandke, K., Dougher, M. M., Sridharan, L., Kunz, A., Hamann, P. R., Gorovits, B., Udata, C., Moran, J. K., Popplewell, A. G., Stephens, S., Frost, P. & Damle, N. K. (2004). Antibody-targeted chemotherapy with CMC-544: a CD22-targeted immunoconjugate of calicheamicin for the treatment of B-lymphoid malignancies. *Blood* **103**, 1807-1814.

47. Damle, N. K. & Frost, P. (2003). Antibody-targeted chemotherapy with immunoconjugates of calicheamicin. *Current Opinion in Pharmacology* **3**, 386-390.

48. Maeda, H. (2001). SMANCS and polymer-conjugated macromolecular drugs: advantages in cancer chemotherapy. *Advanced Drug Delivery Reviews* **46**, 169-185.

49. Takeshita, A., Yamakage, N., Shinjo, K., Ono, T., Hirano, I., Nakamura, S., Shigeno, K., Tobita, T., Maekawa, M., Kiyoi, H., Naoe, T., Ohnishi, K., Sugimoto, Y. & Ohno, R. (2009). CMC-544 (inotuzumab ozogamicin), an anti-CD22 immuno-conjugate of calicheamicin, alters the levels of target molecules of malignant B-cells. *Leukemia* **23**, 1329-1336.

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

51. Whitwam, R. E., Ahlert, J., Holman, T. R., Ruppen, M. & Thorson, J. S. (2000). The gene calC encodes for a non-heme iron metalloprotein responsible for calicheamicin self-resistance in *Micromonospora*. *Journal of the American Chemical Society* **122**, 1556-1557.

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

53. Liu, W., Ahlert, J., Gao, Q., Wendt-Pienkowski, E., Shen, B. & Thorson, J. S. (2003). Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. *Proceedings of the National Academy of Sciences* **100**, 11959-11963.

54. Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B. & Farnet, C. M. (2003). A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nature Biotechnology* **21**, 187-190.

55. Liu, W., Christenson, S. D., Standage, S. & Shen, B. (2002). Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170-1173.

56. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M. & Thorson, J. S. (2002). The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173-1176.

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

58. Hensens, O. D., Giner, J. L. & Goldberg, I. H. (1989). Biosynthesis of NCS Chrom A, the chromophore of the antitumor antibiotic neocarzinostatin. *Journal of the American Chemical Society* **111**, 3295-3299.

59. Tokiwa, Y., Miyoshi-Saitoh, M., Kobayashi, H., Sunaga, R., Konishi, M., Oki, T. & Iwasaki, S. (1992). Biosynthesis of dynemicin A, a 3-ene-1,5-diyne antitumor antibiotic. *Journal of the American Chemical Society* **114**, 4107-4110.

60. Lam, K. S., Veitch, J. A., Golik, J., Krishnan, B., Klohr, S. E., Volk, K. J., Forenza, S. & Doyle, T. W. (1993). Biosynthesis of esperamicin A1, an enediyne antitumor antibiotic. *Journal of the American Chemical Society* **115**, 12340-12345.

61. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F. & Shen, B. (2005). The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. *Chemistry & Biology* **12**, 293-302.

62. Van Lanen, S. G., Oh, T.-j., Liu, W., Wendt-Pienkowski, E. & Shen, B. (2007). Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. *Journal of the American Chemical Society* **129**, 13082-13094.

63. Williams, P. G., Buchanan, G. O., Feling, R. H., Kauffman, C. A., Jensen, P. R. & Fenical, W. (2005). New cytotoxic salinosporamides from the marine actinomycete *Salinispora tropica*. *The Journal of Organic Chemistry* **70**, 6196-6203.

64. Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S. (2007). Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proceedings of the National Academy of Sciences* **104**, 10376-10381.

65. Reed, K. A., Manam, R. R., Mitchell, S. S., Xu, J., Teisan, S., Chao, T.-H., Deyanat-Yazdi, G., Neuteboom, S. T. C., Lam, K. S. & Potts, B. C. M. (2007). Salinosporamides D-J from the marine actinomycete *Salinispora tropica*, bromosalinosporamide, and thioester derivatives are potent inhibitors of the 20S proteasome. *Journal of Natural Products* **70**, 269-276.

66. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D. & Holden, H. M. (1998). The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. Strain CBS-3. *Journal of Biological Chemistry* **273**, 33572-33579.

67. Reed, D. W., Polichuk, D. R., Buist, P. H., Ambrose, S. J., Sasata, R. J., Savile, C. K., Ross, A. R. S. & Covello, P. S. (2003). Mechanistic study of an improbable reaction: Alkene dehydrogenation by the Δ2 acetylenase of *Crepis alpina*. *Journal of the American Chemical Society* **125**, 10635-10640.

68. Carlsson, A. S., Thomaeus, S., Hamberg, M. & Stymne, S. (2004). Properties of two multifunctional plant fatty acid acetylenase/desaturase enzymes. *European Journal of Biochemistry* **271**, 2991-2997.

69. Fox, B. G., Lyle, K. S. & Rogge, C. E. (2004). Reactions of the diiron enzyme stearoyl-acyl carrier protein desaturase. *Accounts of Chemical Research* **37**, 421-429.

70. Fox, B. G., Shanklin, J., Somerville, C. & Münck, E. (1993). Stearoyl-acyl carrier protein delta 9 desaturase from Ricinus communis is a diiron-oxo protein. *Proceedings of the National Academy of Sciences* **90**, 2486-2490.

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

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

73. Shao, L., Qu, X.-D., Jia, X.-Y., Zhao, Q.-F., Tian, Z.-H., Wang, M., Tang, G.-L. & Liu, W. (2006). Cloning and characterization of a bacterial iterative type I polyketide synthase gene encoding the 6-methylsalicyclic acid synthase. *Biochemical and Biophysical Research Communications* **345**, 133-139.

74. Ma, S. M., Li, J. W.-H., Choi, J. W., Zhou, H., Lee, K. K. M., Moorthie, V. A., Xie, X., Kealey, J. T., Da Silva, N. A., Vederas, J. C. & Tang, Y. (2009). Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589-592.

75. Kennedy, J., Auclair, K., Kendrew, S. G., Park, C., Vederas, J. C. & Richard Hutchinson, C. (1999). Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368-1372.

76. Keatinge-Clay, A. T. & Stroud, R. M. (2006). The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748.

77. Keatinge-Clay, A. (2008). Crystal structure of the erythromycin polyketide synthase dehydratase. *Journal of Molecular Biology* **384**, 941-953.

78. Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L. & Shen, B. (2008). A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proceedings of the National Academy of Sciences* **105**, 1460-1465.

79. Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595.

80. Liang, Z.-X. Complexity and simplicity in the biosynthesis of enediyne natural products. *Natural Product Reports* **27**, 499-528.

81. Lee, M., Lenman, M., Banaś, A., Bafor, M., Singh, S., Schweizer, M., Nilsson, R., Liljenberg, C., Dahlqvist, A., Gummeson, P.-O., Sjödahl, S., Green, A. & Stymne, S. (1998). Identification of non-heme diiron proteins that catalyze triple bond and epoxy group formation. *Science* **280**, 915-918.

82. Hu, Z., Pfeifer, B. A., Chao, E., Murli, S., Kealey, J., Carney, J. R., Ashley, G., Khosla, C. & Hutchinson, C. R. (2003). A specific role of the Saccharopolyspora erythraea thioesterase II gene in the function of modular polyketide synthases. *Microbiology* **149**, 2213-2225.

83. Liu, T., Lin, X., Zhou, X., Deng, Z. & Cane, D. E. (2008). Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin. *Chemistry & Biology* **15**, 449-458.

84. Heathcote, M. L., Staunton, J. & Leadlay, P. F. (2001). Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry and Biology* **8**, 207-220.

85. Dillon, S. & Bateman, A. (2004). The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinformatics* **5**, 109.

86. Kong, R., Goh, L. P., Liew, C. W., Ho, Q. S., Murugan, E., Li, B., Tang, K. & Liang, Z. X. (2008). Characterization of a carbonyl-conjugated polyene precursor in 10-membered enediyne biosynthesis. *Journal of the American Chemical Society* **130**, 8142-8143.

87. Kotaka, M., Kong, R., Qureshi, I., Ho, Q. S., Sun, H., Liew, C. W., Goh, L. P., Cheung, P., Mu, Y., Lescar, J. & Liang, Z. X. (2009). Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis. *Journal of Biological Chemistry* **284**, 15739-15749.

88. Nixon, J. E., Putz, G. R. & Porter, J. W. (1968). Synthesis of triacetic acid lactone by the pigeon liver fatty acid synthetase complex. *Journal of Biological Chemistry* **243**, 5471-5478.

89. Tee, E. S. & Lim, C.-L. (1991). The analysis of carotenoids and retinoids: A review. *Food Chemistry* **41**, 147-193.

90. Sun, H., Kong, R., Zhu, D., Lu, M., Ji, Q., Liew, C. W., Lescar, J., Zhong, G. & Liang, Z.-X. (2009). Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis. *Chemical Communications*, 7399-7401.

91. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. (2006). The 2.7Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences* **103**, 11124-11129.

92. Walsh, C. T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805-1810.

93. Shen, B., Liu, W. & Nonaka, K. (2005). Enediyne natural products: biosynthesis and prospect towards engineering novel antitumor agents. *Frontiers in Medicinal Chemistry - Online* **2**, 357-369.

94. Lim, J., Kong, R., Murugan, E., Ho, C. L., Liang, Z.-X. & Yang, D. (2011). Solution structures of the acyl carrier protein domain from the highly reducing type I iterative polyketide synthase CalE8. *PLoS ONE* **6**, e20549.

95. Liew, C. W., Sharff, A., Kotaka, M., Kong, R., Sun, H., Qureshi, I., Bricogne, G. r., Liang, Z.-X. & Lescar, J. (2010). Induced-fit upon ligand binding revealed by crystal structures of the Hot-dog fold thioesterase in dynemicin biosynthesis. *Journal of Molecular Biology* **404**, 291-306.

96. Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G. A., Gao, M., Hills, T., Brokx, S., Qiu, W., Sharma, S., Diassiti, A., Alam, Z., Melone, M., Mulichak, A., Wernimont, A., Bray, J., Loppnau, P., Plotnikova, O., Newberry, K., Sundararajan, E., Houston, S., Walker, J., Tempel, W., Bochkarev, A., Kozieradzki, I., Edwards, A., Arrowsmith, C., Roos, D., Kain, K. & Hui, R. (2007). Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Molecular and Biochemical Parasitology* **151**, 100-110.

97. Collaborative. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **50**, 760-763.

98. Hong, S. K., Kim, K. H., Park, J. K., Jeong, K.-W., Kim, Y. & Kim, E. E. (2010). New design platform for malonyl-CoA-acyl carrier protein transacylase. *FEBS letters* **584**, 1240-1244.

99. Sheldrick, G. (2008). A short history of SHELX. *Acta Crystallographica Section A* **64**, 112-122.

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

101. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132.

102. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3**, 1171-1179.

103. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291.

104. Delano, W. L. (2002). The PyMol Molecular Graphics System, Delano Scientific, San Carlos, CA, USA.

105. Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725-738.

106. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.

107. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. (2004). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45-50.

108. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. (2004). ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Research* **32**, W96-W99.

109. Chen, A. Y., Cane, D. E. & Khosla, C. (2007). Structure-based dissociation of a type I polyketide synthase module. *Chemistry & Biology* **14**, 784-792.

110. Wong, F. T., Chen, A. Y., Cane, D. E. & Khosla, C. (2009). Protein-protein recognition between acyltransferases and acyl carrier proteins in multimodular polyketide synthases. *Biochemistry* **49**, 95-102.

111. Chen, A. Y., Schnarr, N. A., Kim, C.-Y., Cane, D. E. & Khosla, C. (2006). Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society* **128**, 3067-3074.

112. Kim, C.-Y., Alekseyev, V. Y., Chen, A. Y., Tang, Y., Cane, D. E. & Khosla, C. (2004). Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. *Biochemistry* **43**, 13892-13898.

113. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. (2007). Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Science* **16**, 2093-2107.

114. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

115. Holm, L. & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**, W545-W549.

116. Joshi, V. C. & Wakil, S. J. (1971). Studies on the mechanism of fatty acid synthesis: XXVI. Purification and properties of malonyl-coenzyme A-acyl carrier protein transacylase of *Escherichia coli*. *Archives of Biochemistry and Biophysics* **143**, 493-505.

117. Keatinge-Clay, A. T., Shelat, A. A., Savage, D. F., Tsai, S.-C., Miercke, L. J. W., O'Connell, J. D., Khosla, C. & Stroud, R. M. (2003). Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA:ACP transacylase. *Structure (London, England : 1993)* **11**, 147-154.

118. Liu, W., Han, C., Hu, L., Chen, K., Shen, X. & Jiang, H. (2006). Characterization and inhibitor discovery of one novel malonyl-CoA: Acyl carrier protein transacylase (MCAT) from Helicobacter pylori. *FEBS Letters* **580**, 697-702.

119. Serre, L., Swenson, L., Green, R., Wei, Y., Verwoert, I. I. G. S., Verbree, E. C., Stuitje, A. R. & Derewenda, Z. S. (1994). Crystallization of the malonyl coenzyme A-acyl carrier protein transacylase from *Escherichia coli*. *Journal of Molecular Biology* **242**, 99-102.

120. Yadav, G., Gokhale, R. S. & Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *Journal of Molecular Biology* **328**, 335-363.

121. Dreier, J., Li, Q. & Khosla, C. (2001). Malonyl-CoA:ACP transacylase from *Streptomyces coelicolor* has two alternative catalytically active nucleophiles. *Biochemistry* **40**, 12407-12411.

122. Zhang, L., Liu, W., Xiao, J., Hu, T., Chen, J., Chen, K., Jiang, H. & Shen, X. (2007). Malonyl-CoA: acyl carrier protein transacylase from *Helicobacter pylori*: Crystal structure and its interaction with acyl carrier protein. *Protein Science* **16**, 1184-1192.

123. Wong, F. T., Jin, X., Mathews, I. I., Cane, D. E. & Khosla, C. (2011). Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* **50**, 6539-6548.

124. Zhang, Y.-M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O. & White, S. W. (2001). Identification and analysis of the acyl carrier protein (ACP) docking site on β-Ketoacyl-ACP synthase III. *Journal of Biological Chemistry* **276**, 8231-8238.

125. Grdičak, M. & Jerić, I. (2007). Enediyne compounds - New promises in anticancer therapy. *Acta Pharmaceutica* **57**, 133-150.

126. Qunjie, G. & Jon, S. T. (2008). The biosynthetic genes encoding for the production of the dynemicin enediyne core in *Micromonospora chersina* ATCC53710. *FEMS Microbiology Letters* **282**, 105-114.

127. Belecki, K., Crawford, J. M. & Townsend, C. A. (2009). Production of octaketide polyenes by the calicheamicin polyketide synthase CalE8: Implications for the biosynthesis of enediyne core structures. *Journal of the American Chemical Society* **131**, 12564-12566.

128. Akey, D. L., Kittendorf, J. D., Giraldes, J. W., Fecik, R. A., Sherman, D. H. & Smith, J. L. (2006). Structural basis for macrolactonization by the pikromycin thioesterase. *Nature Chemical Biology* **2**, 537-542.

129. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215-218.

130. Kim, B. S., Cropp, T. A., Beck, B. J., Sherman, D. H. & Reynolds, K. A. (2002). Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *Journal of Biological Chemistry* **277**, 48028-48034.

131. Koglin, A., Lohr, F., Bernhard, F., Rogov, V. V., Frueh, D. P., Strieter, E. R., Mofid, M. R., Guntert, P., Wagner, G., Walsh, C. T., Marahiel, M. A. & Dotsch, V. (2008). Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* **454**, 907-911.

132. Kotowska, M., Pawlik, K., Butler, A. R., Cundliffe, E., Takano, E. & Kuczek, K. (2002). Type II thioesterase from *Streptomyces coelicolor* A3(2). *Microbiology* **148**, 1777-1783.

133. Linne, U., Schwarzer, D., Schroeder, G. N. & Marahiel, M. A. (2004). Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *European Journal of Biochemistry* **271**, 1536-1545.

134. Alessandro, A., Laura, C., Susana, G., Giuseppe, Z. & Laurent, T. (2008). Structural and enzymatic characterization of HP0496, a YbgC thioesterase from *Helicobacter pylori*. *Proteins: Structure, Function, and Bioinformatics* **72**, 1212-1221.

135. Li, J., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. *Nature Structural & Molecular Biology* **7**, 555-559.

136. Kunishima, N., Asada, Y., Sugahara, M., Ishijima, J., Nodake, Y., Sugahara, M., Miyano, M., Kuramitsu, S., Yokoyama, S. & Sugahara, M. (2005). A novel induced-fit reaction mechanism of asymmetric hot-dog thioesterase PaaI. *Journal of Molecular Biology* **352**, 212-228.

137. Wang, F., Langley, R., Gulten, G., Wang, L. & Sacchettini, J. C. (2007). Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence. *Chemistry and Biology* **14**, 543-551.

138. Willis, M. A., Zhuang, Z., Song, F., Howard, A., Dunaway-Mariano, D. & Herzberg, O. (2008). Structure of YciA from *Haemophilus influenzae* (HI0827), a hexameric broad specificity acyl-coenzyme A thioesterase. *Biochemistry* **47**, 2797-2805.

139. Cheng, Z., Song, F., Shan, X., Wei, Z., Wang, Y., Dunaway-Mariano, D. & Gong, W. (2006). Crystal structure of human thioesterase superfamily member 2. *Biochemical and Biophysical Research Communications* **349**, 172-177.

140. Thoden, J. B., Holden, H. M., Zhuang, Z. & Dunaway-Mariano, D. (2002). X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. *Journal of Biological Chemistry* **277**, 27468-27476.

141. Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. & Holden, H. M. (2003). The structure of 4-hydroxybenzoyl-CoA thioesterase from *Arthrobacter* sp. strain SU. *Journal of Biological Chemistry* **278**, 43709-43716.

142. Keegan, R. M. & Winn, M. D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallographica Section D* **63**, 447-457.

143. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. In *Methods in Enzymology* (Charles W. Carter, Jr. & Robert, M. S., eds.), Vol. Volume 374, pp. 229-244. Academic Press.

144. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D* **54**, 905-921.

145. Jones, T. A. & Kjeldgaard, M. (1997). Electron-density map interpretation. In *Methods in Enzymology* (Charles W. Carter Jr, R. M. S., ed.), Vol. Volume 277, pp. 173-208. Academic Press.

146. Schuttelkopf, A. W. & van Aalten, D. M. F. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D* **60**, 1355-1363.

147. Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* **44**, 97-179.

148. Leduc, D., Battesti, A. l. & Bouveret, E. (2007). The hotdog thioesterase EntH (YbdB) plays a role *in vivo* in optimal enterobactin biosynthesis by interacting with the ArCP domain of EntB. *Journal of Bacteriology* **189**, 7112-7126.

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

150. Chen, D., Wu, R., Bryan, T. L. & Dunaway-Mariano, D. (2009). *In vitro* kinetic analysis of substrate specificity in enterobactin biosynthetic lower pathway enzymes provides insight into the biochemical function of the hot dog-fold thioesterase EntH. *Biochemistry* **48**, 511-513.

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

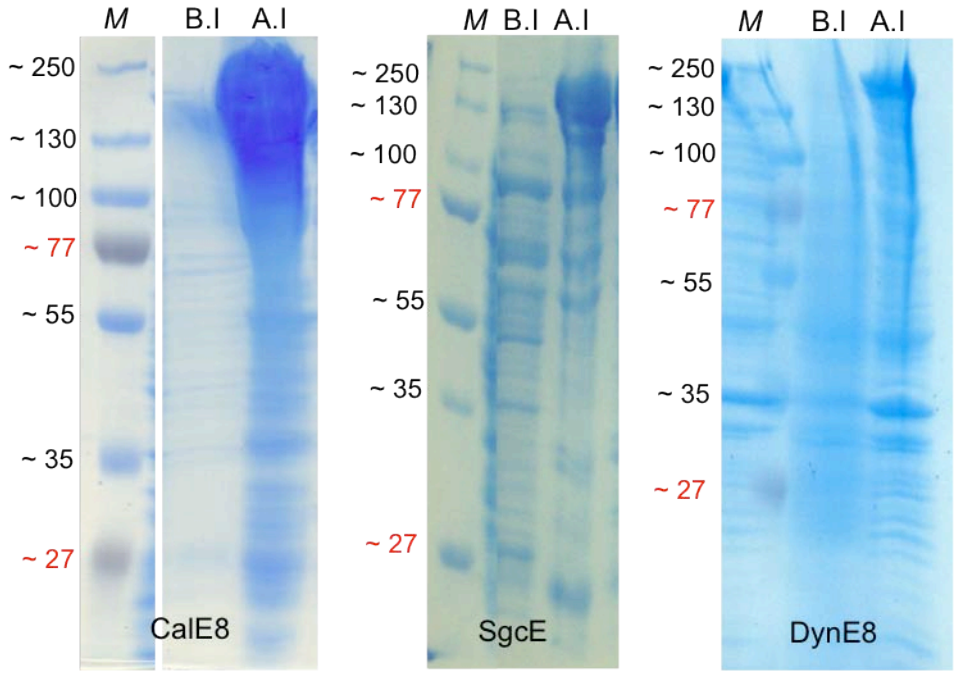
## Appendix

**Appendix Table 2.1** The design of PKSE, TEBC and UNBL plasmid DNA construct used for single, dual and triple protein expression system.

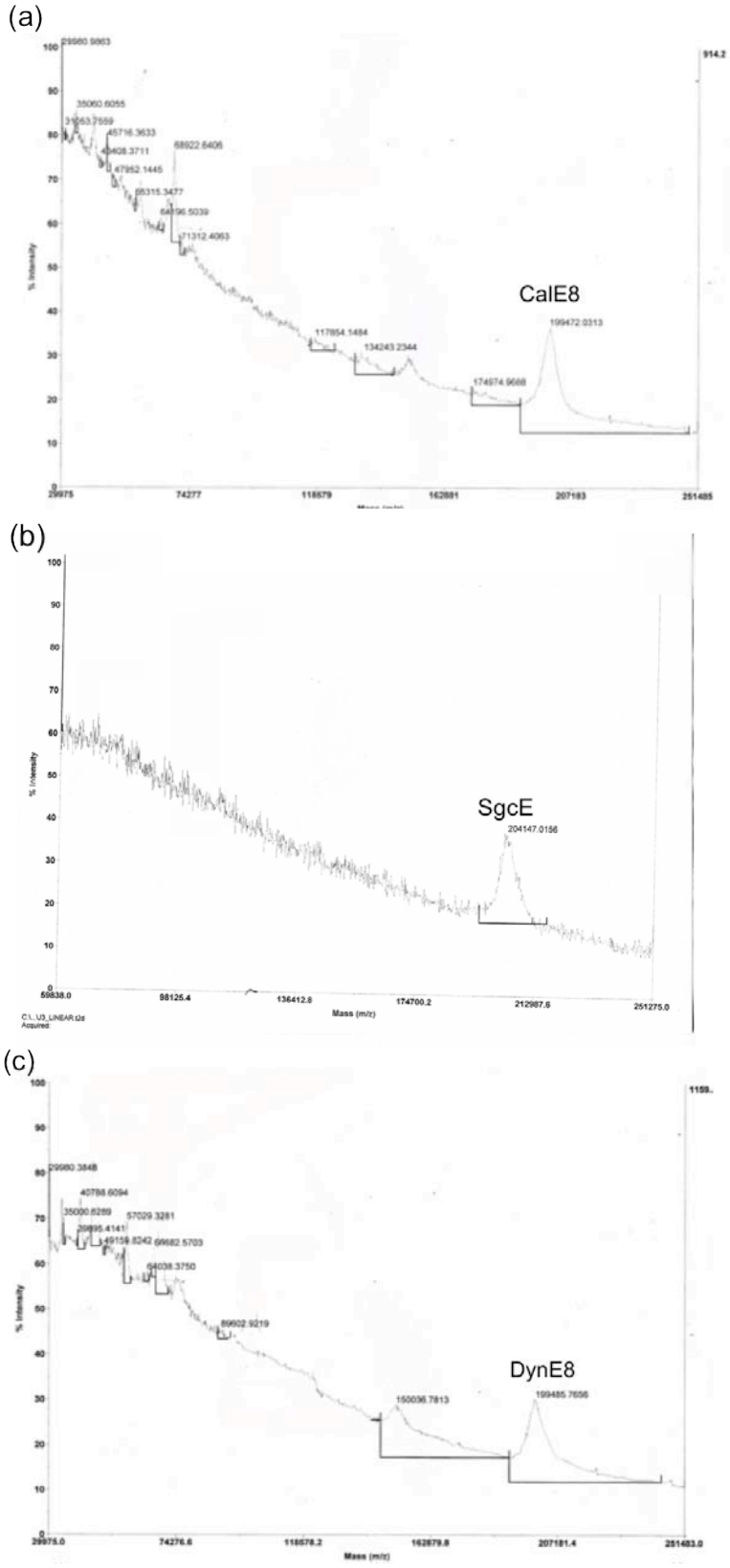
| **Single Protein Expression System** |
|---|
| PET28b(+) – *CalE8*<br>pET28b(+) – *SgcE*<br>pET28b(+) – *DynE8*<br>pCDF-2 – *CalE7*<br>pCDF-2 – *SgcE10*<br>pCDF-2 – *DynE7*<br>pCDF-2 – *CalU15*<br>pCDF-2 – *SgcE3*<br>pCDF-2 – *DynU15* |
| **Dual Protein Expression System** |
| pET28b(+) – *CalE8* + pCDF-2 – *CalE7*<br>pET28b(+) – *SgcE* + pCDF-2 – *SgcE10*<br>pET28b(+) – *DynE8* + pCDF-2 – *DynE7* |
| **Triple Protein Expression System** |
| pET28b(+) – *CalE8* + pCDF-2-*CalE7*-Adaptor-*CalU15*<br>pET28b(+) – *SgcE* + pCDF-2-*SgcE10*-Adaptor-*SgcE3*<br>pET28b(+) – *DynE8* + pCDF-2-*DynE7*-Adaptor-*DynU15* |

**Appendix Table 4.1** Site-directed mutagenesis on CalE7 and DynE7

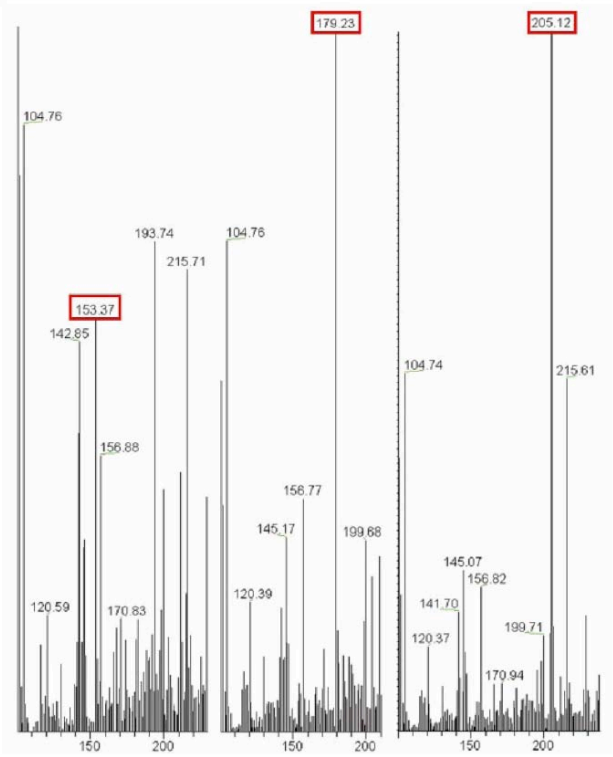
| CalE7 mutants | DynE7 mutants |
|---|---|
| E17Q | R35A |
| N19A | E36A |
| N23A | |
| Y29F | |
| C36A | |
| R37Q | |
| R37K | |
| E38Q | |
| T60A | |
| E70Q | |
| D75N | |



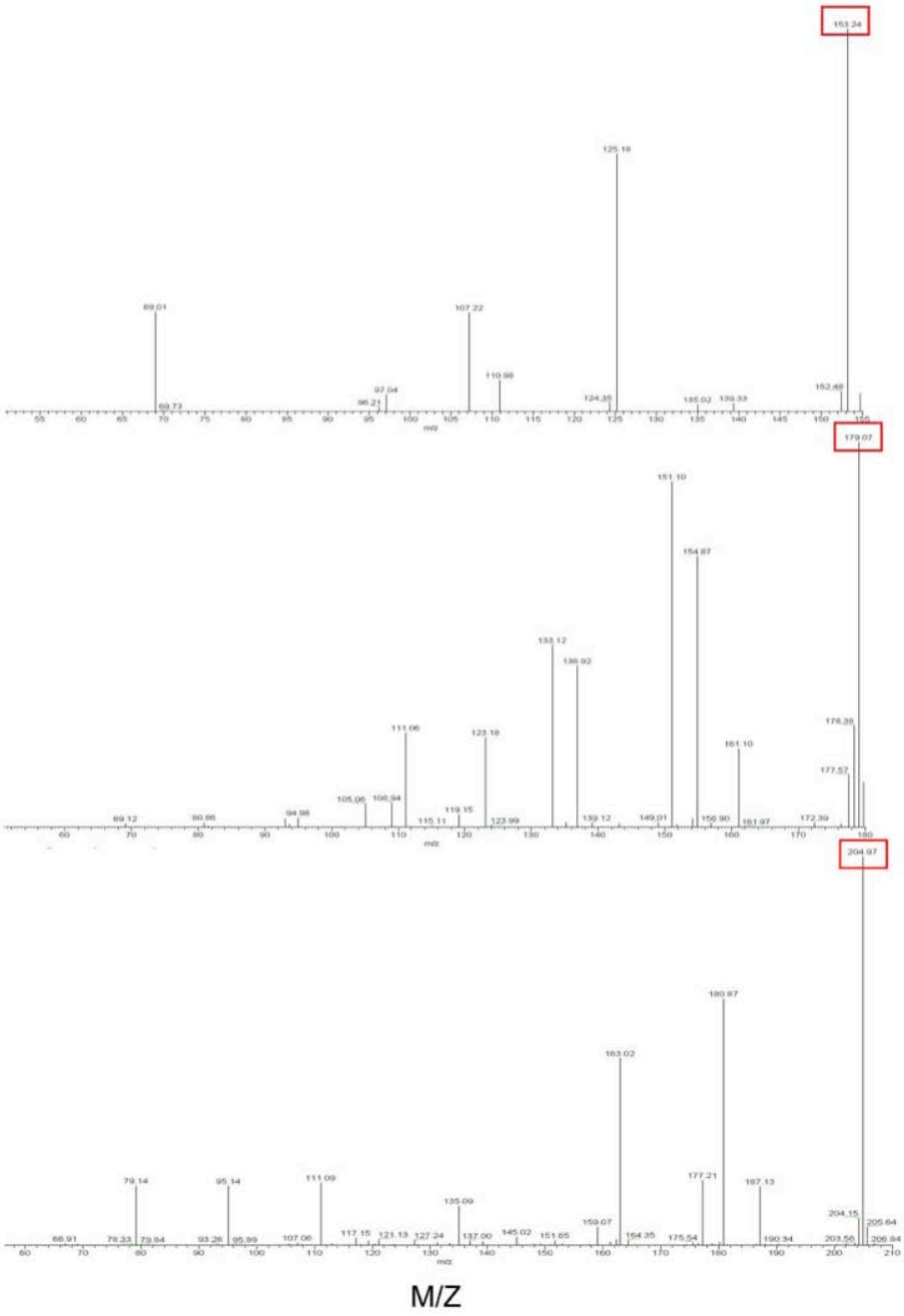

**Appendix Figure 2.1** Expression of CalE8, SgcE and DynE8 protein. ***M***: PageRuler Plus Prestained Protein Ladder (Fermentas). **B.I**: Before induction. **A.I**: After induction with 0.2 mM IPTG.



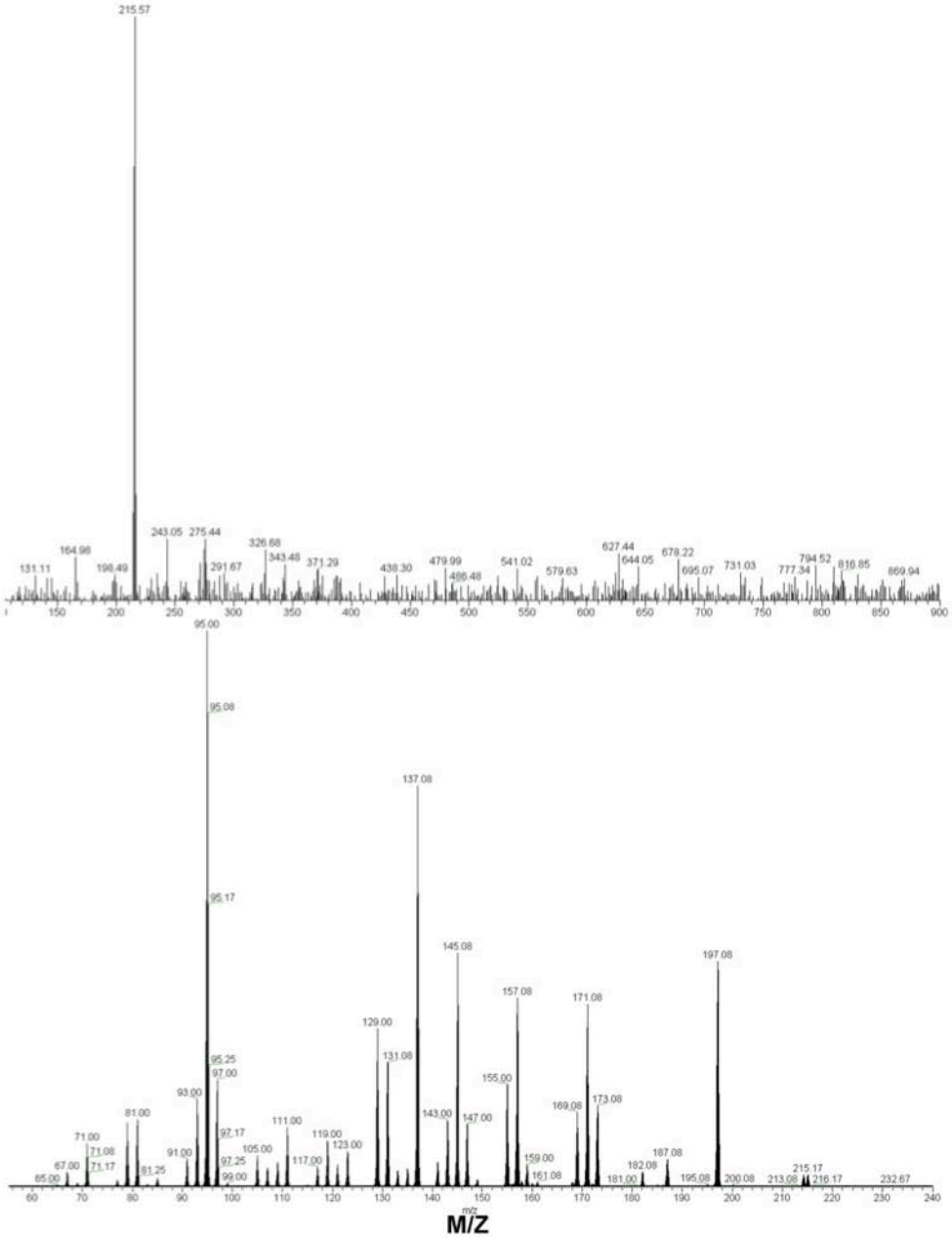

**Appendix Figure 2.2** mass spectroscopy analyses of CalE8 (**a**), SgcE (**b**) and DynE8 (**c**) purified sample.



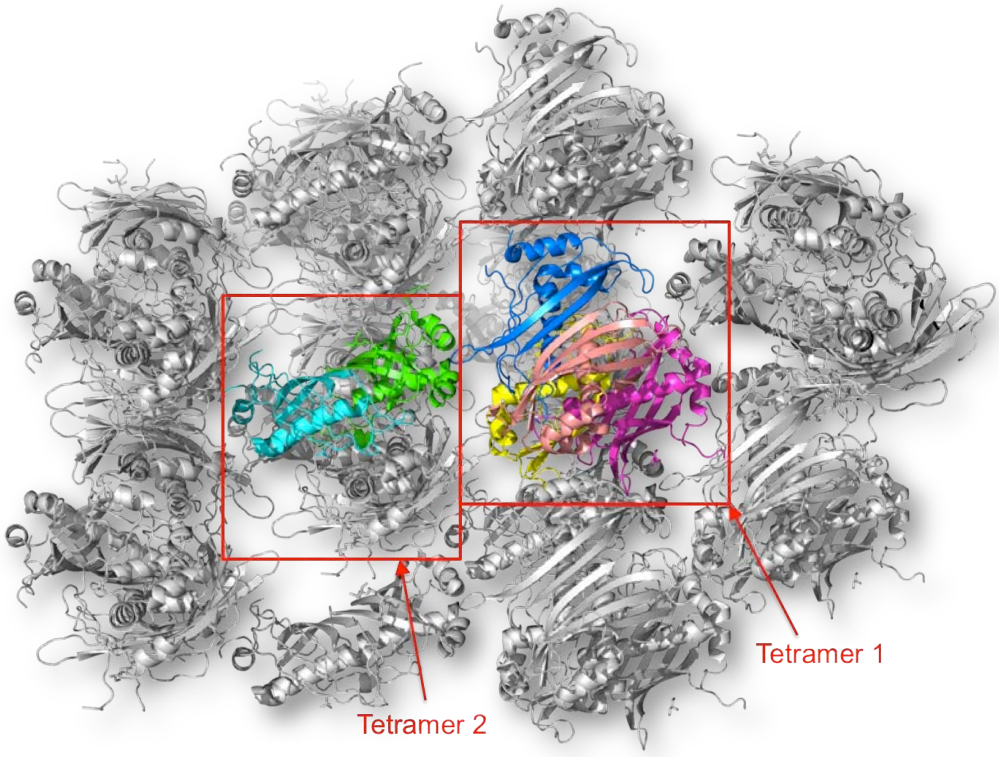

**Appendix Figure 2.3** Full MS for product **6**, **7** and **8**. Since PKSE products are comprised of at least 8 acetate units connected head-to-tail, the masses (circled in red) shown here can only be aberrant products with predicted molecular formula of $C_8H_8O_3$(**4**), $C_{10}H_{10}O_3$ (**5**) and $C_{12}H_{12}O_3$(**6**).




**Appendix Figure 2.4** Fragmentation pattern for product **6**, **7** and **8** in positive mode MSMS. The intensity for fragment ions are extremely low compared to that of parental ions for all 3 species (figures are not shown in proportion). Therefore, it is likely that all 3 aberrant products are cyclic in nature. The parental species are highlighted in red box.




**Appendix Figure 2.5** Full MS and MSMS for product **3**. The high resolution mass obtained for protonated **3** (predicted molecular formula of C15H18O) is 215.1430. Fragmentation for **3** generates a series of broken-down masses of considerable intensity with respect to parental ion. Given the fact that the long chain acyl product only contains a single oxygen, it is highly plausible that **3** is a linear polyunsaturated polyketide product.




**Appendix Figure 4.1** A crystal packing of CaLE7. A total of six molecules in the asymmetric unit: four form a 222 (or *D2*) tetramer (Tetramer 1) and two form a 2-fold symmetric dimer. A second complete 222 tetramer can be generated from the latter dimer via the crystallographic dyad (Tetramer 2).

**Publications**

Kong R, Goh LP, **Liew CW**, Ho QS, Murugan E, Li B, Tang K, Liang ZX., *Characterization of a carbonyl-conjugated polyene precursor in 10-membered enediyne biosynthesis.* Journal of the American Chemical Society, 2008. 130(26): p. 8142-8143.

Kotaka M, Kong R, Qureshi I, Ho QS, Sun H, **Liew CW**, Goh LP, Cheung P, Mu Y, Lescar J, Liang ZX., *Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis.* Journal of Biological Chemistry, 2009. 284(23): p. 15739-15749.

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

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

**Liew CW**, Nillson M, Chen MW, Sun HH, Cornvik T, Liang ZX and Lescar J., *Crystal structure of acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis*. Journal of Biological Chemistry. 2012. 287(27): p. 23203-23215.