THE UNIVERSITY
of EDINBURGH

Studies on the structure, mechanism and protein engineering of *Bacillus subtilis* pimeloyl-CoA synthetase (PCAS)

Menglu Wang

A Thesis Submitted for the Degree of Doctor of Philosophy

The University of Edinburgh

2016

Abstract

Biotin is an essential vitamin in plants and mammals functioning as the carbon dioxide carrier within central lipid metabolism. Biotin is composed of a fused bicylic ring system and a five carbon, carboxylic acid chain. Biotin biosynthesis in bacteria is catalysed by a series of enzymes that use fatty acid, amino acid and sulfur-containing substrates. In *Bacillus subtilis*, pimeloyl-CoA synthetase (PCAS, EC 6.2.1.14, UNIPROT code: P53559, 29.6 kDa) is the first enzyme in the biotin biosynthetic pathway and acts as a highly specific substrate selection gate ensuring the integrity of the carbon chain in biotin synthesis. PCAS catalyses the synthesis of the key acyl-thioester, pimeloyl-CoA in two steps; the first involves activation of pimelic acid (C7 dicarboxylic acid) using ATP to give an acyl-adenylate, enzyme-bound intermediate and pyrophosphate (PPi), and in the second step, this pimeloyl-adenylate reacts with coenzyme A (CoASH) to form the pimeloyl-CoA thioester. This thesis describes the results of biochemical, structural and mechanistic studies of *B. subtilis* PCAS. Recombinant PCAS was prepared by expressing the *B. subtilis BioW* gene in *E. coli* in various hexa-histidine affinity-tagged forms and the enzyme purified in high purity and yield. Enzyme activity and kinetic constants were measured using reverse-phase HPLC and enzyme coupled spectroscopic assays. These revealed the enzyme to have a strict carboxylic acid specificity. In collaboration with colleagues at the University of St. Andrews various commercial and in-house screens were used to obtain diffraction-quality crystals suitable for X-ray crystallography. This also

included the generation of seleno-methionine (SeMet) labelled PCAS, as well as heavy-metal derivatives. Structures of *B. subtilis* PCAS in complex with the substrate pimelic acid and the pimeloyl-adenylate intermediate and product PP_i were determined at 2.04 Å and 2.34 Å resolution respectively. The *B. subtilis* PCAS displays a novel 3D fold and defines a new class (Class IV) in the ANL superfamily of adenylate forming enzymes. The enzyme is a homodimer composed of two domains, a short N-terminus and a large C-terminal domain and the ligand-bound structures revealed the residues potentially involved in substrate specificity and enzyme catalysis. The enzyme uses an internal ruler composed of a number of conserved arginine residues (Arg213, Arg227 and Arg170) to select the correct dicarboxylic acid substrate. The X-ray structures guided the production of a number of site directed mutants to identify residues involved in the catalytic mechanism and stabilising the acyl-adenylate intermediate. This also allowed rational engineering of the PCAS active site to generate mutants with altered substrate specificity. Mutant PCAS Y211F was shown to synthesise both heptanoyl (C7) and octanoyl (C8) mono carboxylic acid-CoA and C8 dicarboxylic-CoA thioester products, highlighting the synthetic potential of PCAS.

The PCAS pimeloyl-CoA product is the substrate for the next enzyme in the biotin pathway, a pyridoxal 5'phosphate (PLP)-dependent 8-amino 7-oxononanoate synthase (AONS). AONS catalyses the condensation of pimeloyl-CoA with L-alanine to give AON which is converted to biotin by the action of three other enzymes. We used

genome mining to identify a putative ~66 kDa, bi-functional PCAS/AONS enzyme with an N-terminal PCAS domain fused to C-terminal AONS domain in the organism *Corynebacterium amycolatum*. A recombinant *C. amycolatum* PCAS/AONS fusion protein was expressed and purified from *E. coli* and initial studies suggest that it forms a functional, fused, dimeric enzyme.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Dominic Campopiano for giving me the opportunity to work in his group. His academic support, valued input, encouragement and contagious passion for science are very much appreciated.

I am indebted to the University of Edinburgh and *Ingenza Ltd* for funding, without which this project would not have been possible.

I would like to express my gratitude to Prof. Jim Naismith and Dr. Lucile Moyni é. They have given me the opportunity to work in their lab in St. Andrews and have provided useful scientific expertise and suggestions in the X-ray crystallography study of PCAS project.

Thanks to Dr. Peter Harrison for his kind help with more lab results during the paper revision process after I went back to China.

I am indebted to Dr. Van Kelly and Dr. Dave Clarke for their kindly help with the mass spectrometry analysis and also for their astute academic advice and suggestions during my studies.

I would especially like to thank Dr Scott Baxter for his tremendous guidance and help in my first year PhD. I am grateful for all I have learnt under his supervision in *Ingenza*. Thanks to all the people in *Ingenza* for their enthusiastic help.

I would also like to extend my gratitude to the rest of the Campopiano research group (Lab229). Daynea, Chris, Guiomar, Annabel, Piera, Bohdan and Alex thanks a lot for the fruitful discussions we have had, that have aided in my scientific growth. Thanks to project student Andrew, Lauren, Cat.

I am also thankful for the encouragement and support of my friends both in Edinburgh and from China. Love you all.

I would like to say a big thank you to my mom, dad and grandparents for your love. You have truly been my strength throughout my studies and also my life.

Declaration

I, Menglu Wang, hereby certify that this thesis has been completed by myself, and that it is a record of my work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Menglu Wang

The University of Edinburgh

August, 2016

Table of Contents

Abbreviations

ACP	Acyl Carrier Protein
5'-DOA	5'-deoxyadenosine
Abs	Absorbance
ADP	Adenosine diphosphate
Ala	Alanine
AMP	Adenosine-5'-monophosphate
AMP_{100}	Ampicillin (100µg/mL)
AMTOD	S-adenosyl-2-oxo-4-thiomethylbutyrate
AON	8-Amino-7-oxononanoate
AONS	8-Amino-7-oxononanoate synthase
APS	Ammonium persulfate
Arg	Arginine
ATP	Adenosine triphosphate
B. subtilis	*Bacillus subtilis*
BME	β-mercaptoethanol
C. amycolatum	*Corynebacterium amycolatum*
CFE	Cell free extract
CoA/CoASH	Coenzyme A
Da	Dalton
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic aci
ESI	Electrospray ionisation
FAS	Fatty acid synthase
His	Histidine
IPTG	Isopropyl-β-D-galactoside
KAN_{50}	Kanamycine (50µg/mL)
k_{cat}	Catalytic constant
K_D	Dissociation constant
K_M	Michaelis-Menten constant
Kphos	Potassium phosphate
LB	Luria-Bertani Medium
LC-MS	Liquid chromatography-mass spectrometry
LMW	Low molecular we ght marker
MESG	7-methyl-6-thioguanosine
Met	Methionine
Mr	Protein relative molecular mass
MS	Mass spectrometry

NIS	NRPS-independent siderophores
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein data bank
Pi	Phosphate
PLP	Pyridoxal 5'-phosphate
PMSF	Phenylmethanesulfonyl fluoride
PNP	Purine nucleoside phosphorylase pyrophosphorylase
PPase	Pyrophosphatase
PPi	Pyrophosphate
RP-HPLC	Reverse phase high performance liquid chromatography
SAD	Single-wavelength anomalous diffraction
SAH	S-adenosylhomocysteine
SAM	S-adenosyl-L-methionine
SDM	Site directed mutagenesis
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SeMet	Seleno-methionine
SOC	Super Optimal Culture
TAE	Tris base, acetic acid and EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TGS	Tris-Glycine-SDS
Tris	Tris(hydroxymethyl)aminomethane
Tris-HCl	2-Amino-2-hydroxymethyl-1,3-propanediol-Hydrochloric acid
UV-vis	Ultraviolet-visible
Ve	Elution volume
Vo	Void volume
WT	Wild type

Chapter I. Introduction

1.1 Discovery of biotin

Biotin (5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno [3,4-*d*] imidazol-4-yl] pentanoic acid, $C_{10}H_{16}N_2O_3S$, **Fig. 1.1**), also known as coenzyme R, vitamin H or vitamin B7, plays an essential role in the anabolism of fatty acids, where it acts as an important carbon dioxide-carrying cofactor within large enzyme complexes such as acetyl-CoA carboxylase (ACC). ACC catalyses the carboxylation of acetyl-CoA to form malonyl-CoA in the first step of fatty acid biosynthesis [1].

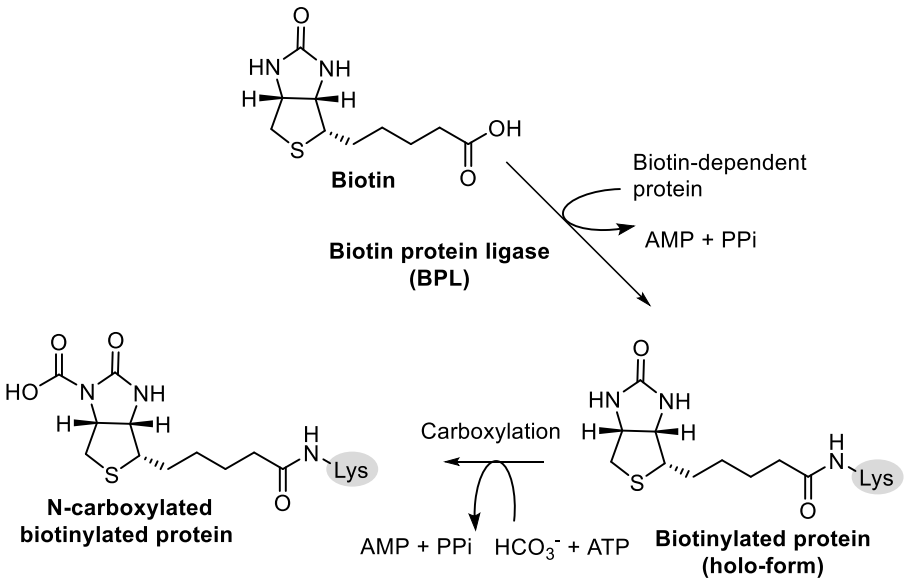


Fig. 1.1 This shows the structure of biotin, how it is attached to a specific lysine residue of its cognate biotinylated (holo-form) of a biotin-dependent enzyme (e.g. ACC) and carries the CO_2 moiety on the N8 atom of the ureido ring.

In recent years, new roles for biotin in cell signalling, chromatin structure and gene expression have been discovered [2]. While biotin is required by all three domains of

life, it can only synthesised *de novo* by bacteria, yeast and certain higher plants, mammals intake biotin from the diet. This provides a feasible way to develop novel antibiotics and herbicides by targeting enzymes involved in the biotin biosynthesis. For example, plumbagin (5-hydroxy- 2-methyl-1,4-naphthoquinone), which is an inhibitor of *Arabidopsis thaliana* 8-amino-7-oxononanoate synthase (AONS), has been identified as a herbicide in certain plants [3]. In humans, biotin deficiency causes illnesses including periorificial dermatitis, alopecia, seizures and developmental delay in children etc. and also has adverse effects on immune functions and cell stress [4], but it usually remedied by biotin supplementation. Biotechnological applications of biotin are based around its avidin complex, one of the strongest non-covalent interactions known. Biotin conjugated to biomolecules *in vitro* and *in vivo* can be used to isolate proteins, DNA and cell surfaces [5, 6].

Biotin was first discovered by Wilders in 1901 [7] and was associated with different names in the following years [8] including protective factor X [9] and vitamin H [10] which was isolated from egg-white as a curative factor, coenzyme R from rhizobia as a growth factor [11] and biotin from yeast [12]. Until 1942, Vigneaud was first determined its chemical structure and by this the identity of biotin becomes clear [13], later the synthesis of biotin was accomplished in 1943 [14]. Biotin is comprised of a valeric acid attached to a bicyclic ring as shown in **Fig. 1.1** [15]. The bicyclic ring consists of a planar ureido ring fused to an envelope conformation tetrathiophene ring. Biotin is covalently attached to its cognate biotin-dependent protein in a reaction catalyzed by the enzyme biotin protein ligase (BPL) [16] via an amide bond between its carboxyl

group and the ε-amino group of a specific lysine in the conserved 70-80 residues domain of the biotin-dependent protein [5, 17] . The N8 nitrogen atom on the ureido ring carries carbon dioxide and forms N-carboxybiotin [18, 19] as shown in **Fig. 1.1**.

1.2 Pathways of microbial biotin biosynthesis

Biotin possesses two fused rings (ureido and thiophane) whose carbon skeleton derives from pimelic acid which is an α, ω-dicarboxylic acid composed of seven carbons. Pimelic acid is quite unusual because natural fatty acid biosynthesis pathways mainly generates even numbered carbon chains of fatty acids such as palmitic acid (C16) [20]. In fact, the derivatives of pimelic acid are best known in the lysine biosynthetic pathway [5]. The origin of the pimelic acid varies across bacterial species and it has been found in various forms: as a free di-acid, a mono-methyl ester, as well as being transformed into a pimeloyl thioester derivative attached to either coenzyme A (CoASH) or acyl carrier protein (ACP) [6]. It is this pimeloyl thioester that is condensed with L-alanine by the 8-amino-7-oxononanoate synthase (AONS, *BioF* gene product) to begin the late stages of biotin biosynthesis [21, 22]. The genes encoding the enzymes of biotin biosynthesis are often found co-expressed as operons [21, 22].

As summarised in **Fig. 1.2**, biotin biosynthesis pathway can be easily divided into two stages: (1) synthesis of the pimelate thioester through the BioC-BioH pathway of *Escherichia coli* (details shown in **Section 1.2.1**) or BioI-BioW pathway of *Bacillus subtilis* (**Section 1.2.2**), and (2) assembly of the bicyclic rings (**Section 1.2.3**).

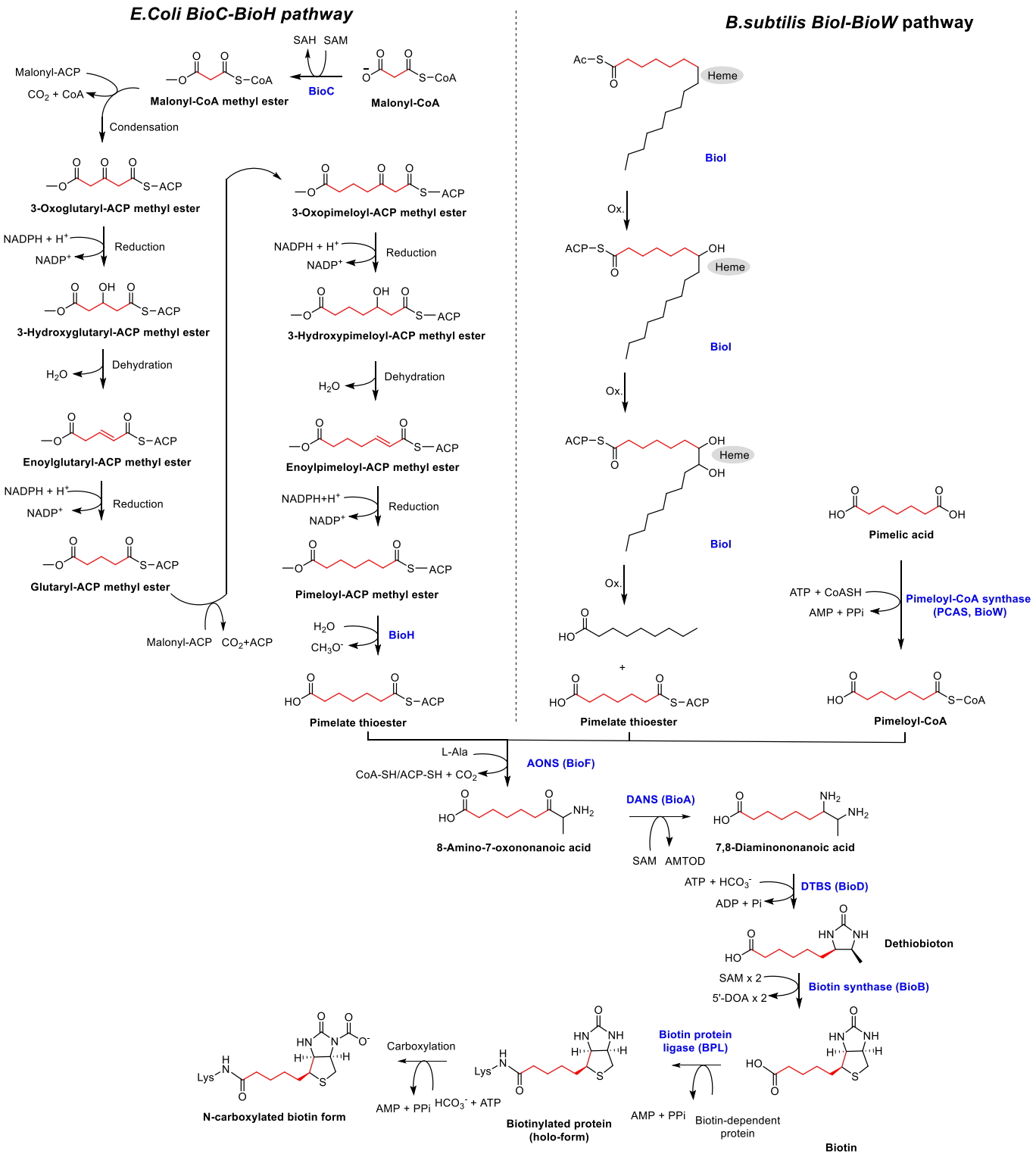


Fig. 1.2 This shows the proposed biotin biosynthesis pathway in various microbes. The pimeloyl thioester is first synthesised through either the *E. coli* BioC-BioH pathway or the *B. subtilis* BioI-BioW pathway, and then used as an

important building block for the bicyclic ring assembly step of biotin. Then biotin is attached to biotin-dependent protein and carries CO_2. (SAM=S-adenosyl-L-methionine, SAH=S-adenosylhomocysteine, AMTOD= S-adenosyl-2-oxo-4-thiomethylbutyrate, 5'-DOA=5'-deoxyadenosine).

1.2.1 The BioC-BioH pathway of *E. coli*

The organism *E. coli* was the model microorganism in which most of the early biosynthetic work was done to characterise any chemical intermediates involved, isolate the enzymes that catalyse their conversion and clone the genes that encoded each enzyme. In 1968, Eisenberg found that *E. coli* ΔbioC and ΔbioH mutation strains require biotin for cell growth, moreover, detectable cell growth could also be observed by adding 8-amino-7-oxononanoate (AON, also known as 7-keto-8-aminopelargonic acid, KAPA) or any later intermediates of the biotin biosynthesis pathway to the growth media [23]. This indicated that both the BioC and BioH gene products could be involved in the synthesis of the pimelate thioester which is an essential substrate for AONS reaction. Also, the lack of isolatable pathway intermediate suggested that the possibility of these pathway intermediates being protein bound [24].

In the following years, Lemoine proposed that BioC involves in the synthesis of pimeloyl-CoA or the condensation of malonyl-CoA [25], Akatsuk thought BioH catalyses the reaction which transferred pimeloyl from BioC to CoASH [26], whereas Sanishvili hypothesized BioH is a pimeloyl-CoA synthetase and together with BioC,

it can catalyse the reaction which attaches the C7 pimelic acid onto CoASH [27]. However, all these models failed to explain the details about how the α, ω-dicarboxylic acyl chain is assembled by BioC and BioH. In recent years, this biosynthetic "enigma" was resolved; a more acceptable model was put forward by Cronan, he proposed that the pimeloyl-ACP thioester is produced from acetyl-CoA and malonyl-CoA by an ingenious modification of the classic fatty acid pathway in *E. coli* and in effect, two enzymes from the biotin pathway, BioC (a methyl transferase) and BioH (an esterase) "hijack" fatty acid biosynthesis by methylating malonyl-ACP then extending it through two rounds of the fatty acid synthase cycle [5, 28]. As shown in **Fig. 1.3**, the initial step of the BioC-BioH branch of the pathway involves the BioC-catalysed transfer of the methyl group from SAM onto the ω-carboxyl group of a malonyl-CoA to form a malonyl-CoA methyl ester; this methyl ester is then transferred to the ACP of the fatty acid synthase (FAS) complex and the pimeloyl-CoA methyl ester is produced through two fatty acid chain elongation steps. Finally the methyl group is hydrolysed by BioH to produce pimeloyl-ACP which can then act as a substrate for AONS directly.

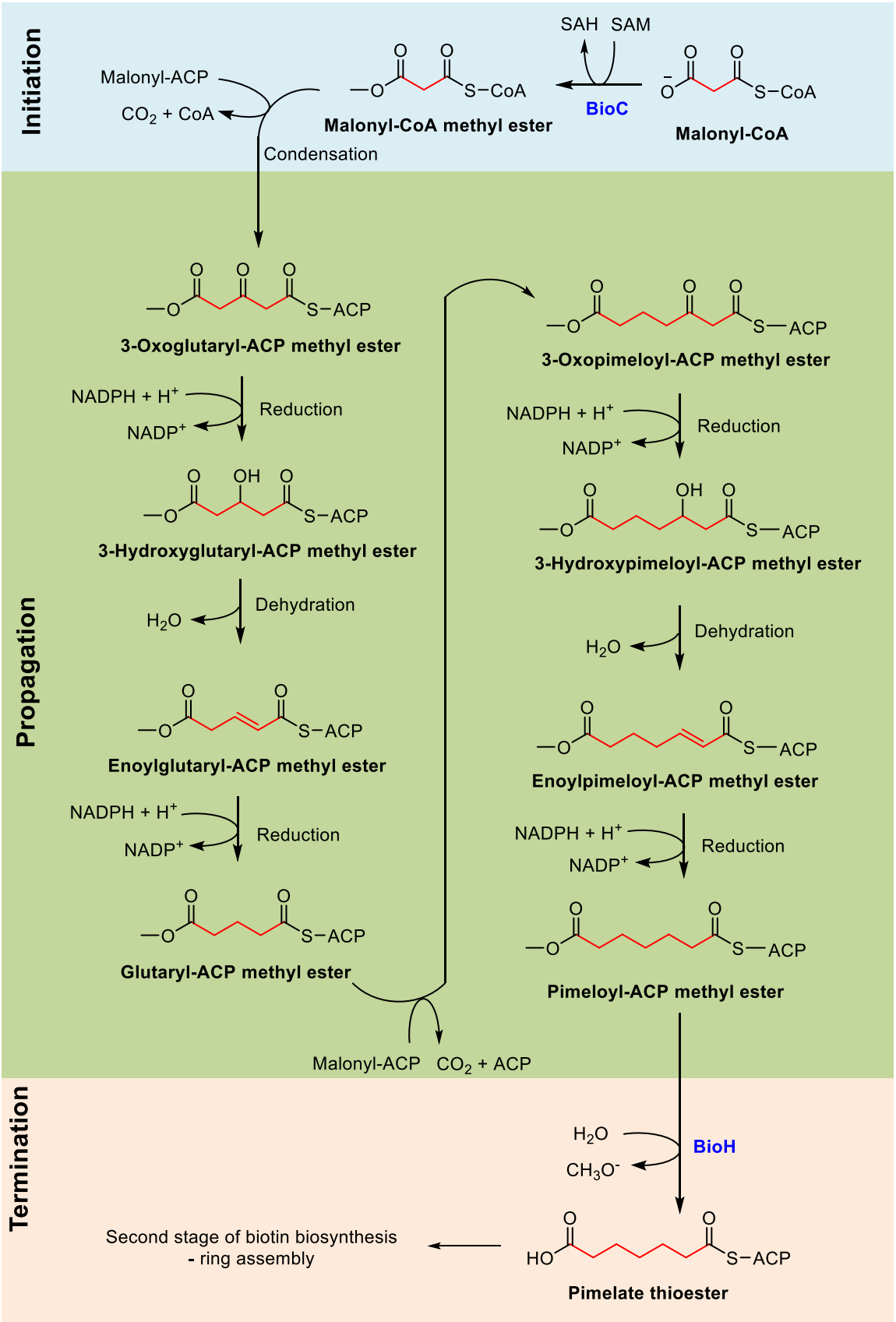


Fig. 1.3 The proposed BioC-BioH branch of the biotin biosynthetic pathway in *E. coli*

for pimeloyl-ACP synthesis contains three main steps: initiation, chain elongation by

fatty acid synthesis and termination. (SAM = S-adenosyl-L-methionine, SAH = S-adenosylhomocysteine).

1.2.2 The BioI-BioW pathway of *B. subtilis*

In contrast to the pathway in *E. coli* where fatty acid biosynthesis is hijacked, in *Bacillus subtilis* pimelic acid is generated by the oxidative cleavage of longer fatty acids catalysed by BioI [29] as shown in **Fig. 1.4**. BioI, an cytochrome P450 (CYP) enzyme, catalyses the oxidative carbon-carbon bond cleavage of acyl thioesters with chain lengths of 10-20 carbons between carbons 7 and 8 (**Fig. 1.4**) [30]. BioI uses the cysteine-ligated heme to activate O_2 to form one equivalent of the highly reactive iron (V)-oxo species and a water molecule. The two electrons required for the oxygen activation are derived from NAD(P)H [31]. The general reaction can be written as follows:

$$RH + O_2 + NAD(P)H + H^+ \longrightarrow ROH + H_2O + NAD(P)^+$$

There are three main steps involved in the BioI pathway: (1) the initial monohydroxylation step at carbon 7 position [32], (2) the monohydroxylation step at carbon 8 resulting in a dihydroxylated, and (3) the final oxidation step involving in the cleavage of the C7-C8 bond which produces two carboxylic acids (**Fig. 1.4**). One acid is the C7 pimeloyl-ACP which is then used for biotin assembly by AONS; while another will vary in carbon chain length depending on the substrate but will always contains odd number of carbons.

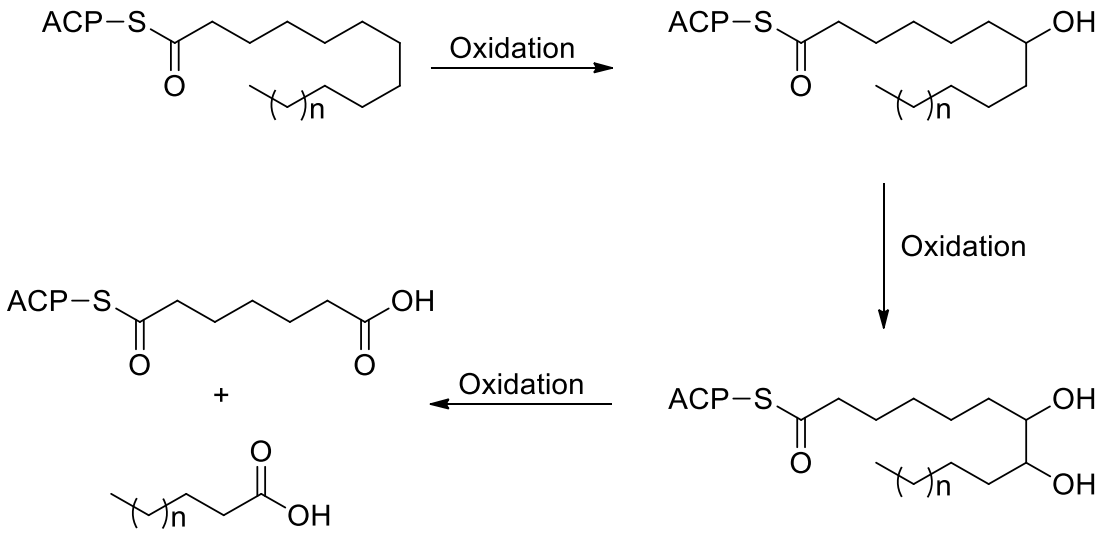


Fig. 1.4 The three oxidation steps of the P450 enzyme BioI generates pimeloy-ACP as suggested by Cryle et al [33].

The regiospecificity of BioI oxidative cleavage of fatty acyl-ACPs is attributed to (1) the geometry of BioI active site and (2) enzyme/ACP interaction. The crystal structure of BioI/acyl-ACP complexes (as **Fig. 1.5**) reveals that the acyl chain together with the 4'-PPant group is inserted into the U-shaped substrate pocket with C7-C8 positioned at the bend of the "U" above the active site heme, and ACP is bound to the BioI surface [34] . The U-shaped cavity extends beyond the heme to accommodate different length of the fatty acyl chains.

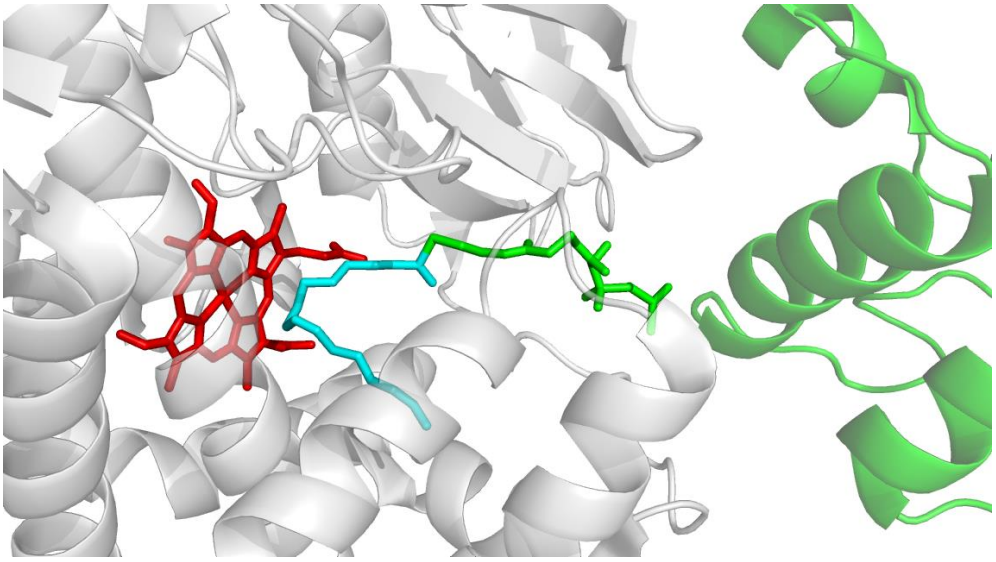


Fig. 1.5 The crystal structure of *B. Subtilis* BioI in complex with *E. coli* hexadec-9Z-enoyl-ACP (PDB code: 3EJD) [34], heme group is in red, fatty acyl chain in cyan, 4'-PPant group as green stick and ACP as green carton, BioI enzyme is in grey.

In *B. subtilis*, as well as formation of pimeloyl-ACP, the synthesis of pimeloyl-CoA, the alternative acyl-thioester substrate for AONS, is catalysed by the enzyme pimeloyl-CoA synthetase (PCAS, *BioW* gene product). PCAS uses the free C7 di-acid as a substrate and MgATP to generate a pimeloyl-adenylate intermediate which subsequently reacts with CoASH to give the acyl-CoA thioester product. The details of PCAS will be discussed in **Section 1.4**.

1.2.3 Ring assembly

As shown in **Fig. 1.6** [5], a key early step in ring assembly step requires the decarboxylative condensation of two components: L-alanine and a thioester activated form of pimelic acid (either a pimeloyl-CoA or pimeloyl-ACP thioester). This 8-amino-7-oxononanoate synthase (AONS, the *bioF* gene product) catalysed reaction generates the nine carbon AON intermediate and is highly conserved across most bacterial species, as are the latter stages of the construction of the two rings of biotin from AON. Then the fused bicyclic structure is formed in the latter conserved three steps of the biotin biosynthetic pathway involved in (1) the initial reaction that converts ketone to amine using S-adenosyl-L-methionine (SAM) as the amine source which catalysed by 7, 8-diaminononanoate synthase (DANS, *bioA* gene product), (2) an ATP dependant reaction inserts a carbonyl group between the amines of 7,8-diaminononanoic acid to produce a five-membered ring which catalysed by dethiobiotin synthase (DTBS, *bioD* gene product) and (3) the final biotin synthase (BS, *bioB* gene product)-catalysed reaction that converts dethiobiotin to biotin using two S-adenosyl methionine (SAM) is a member of the radical SAM-dependent enzyme superfamily.

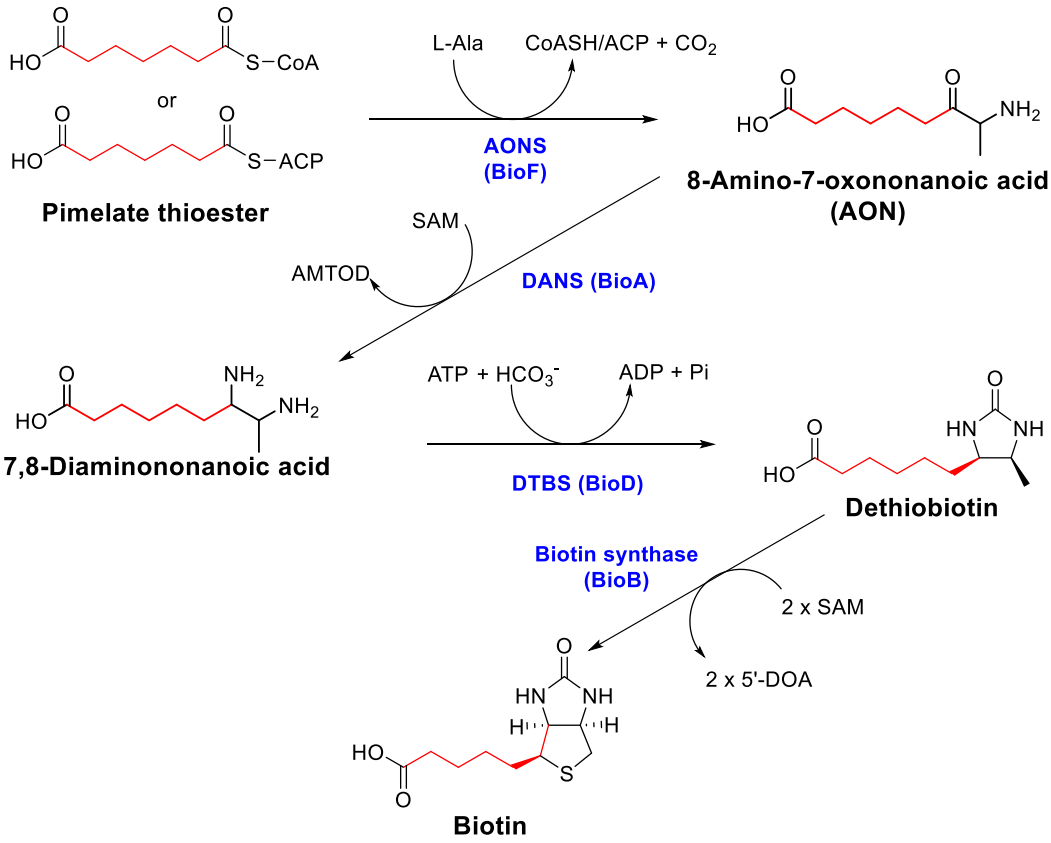


Fig. 1.6 The second step of biotin biosynthesis that is conserved in all biotin producing organisms. It is composed of four reactions involved in 8-amino-7-oxononanoate synthase (AONS, BioF), 7,8-diaminononanoate synthase (DANS, BioA), dethiobiotin synthase (DTBS, BioD) and biotin synthase (BS, BioB) to form the bicyclic rings of biotin. A pimelate thioester acts as a precursor to provide the biotin carbon backbone. (AMTOD = S-adenosyl-2-oxo-4-thiomethylbutyrate, 5'-DOA = 5'-deoxyadenosine, SAM = S-adenosyl-L-methionine,)

1.2.4 Structural information of enzymes involved in biotin biosynthesis

Until now, the structural information of nearly all of the enzymes in the biotin biosynthesis pathway is available in protein Data Bank (PDB) except the methyl-transferase BioC and pimeloyl-CoA synthetase (PCAS, BioW) (**Table 1.1** and **Fig. 1.7**).

Gene	Enzyme	EC	Classification	PDB	Fig.
BioH	Pimeloyl-ACP methyl esterase	3.1.1.85	Hydrolase	4ETW [35]	Fig.1.6A
BioI	Pimeloyl-ACP synthase	1.14.15.12	P450 oxidoreductase	3EJE [34]	Fig. 1.6B
BioF	8-amino-7-oxononanoate synthase (AONS)	2.1.3.47	PLP-dependent synthase	1DJ9 [22]	Fig. 1.7A
BioA	7, 8-diaminononanoate synthase (DANS)	2.6.1.62	PLP-dependent aminotransferase	1QJ3 [36]	Fig. 1.6C
BioD	Dethiobiotin synthase (DTBS)	6.3.3.3	ATP-dependent ligase	3MLE [37]	Fig. 1.6D
BioB	Biotin synthase (BS)	2.8.1.6	SAM-dependent transferase	1R30 [38]	Fig. 1.6E

Table 1.1 Structural information of enzymes involved in biotin biosynthesis

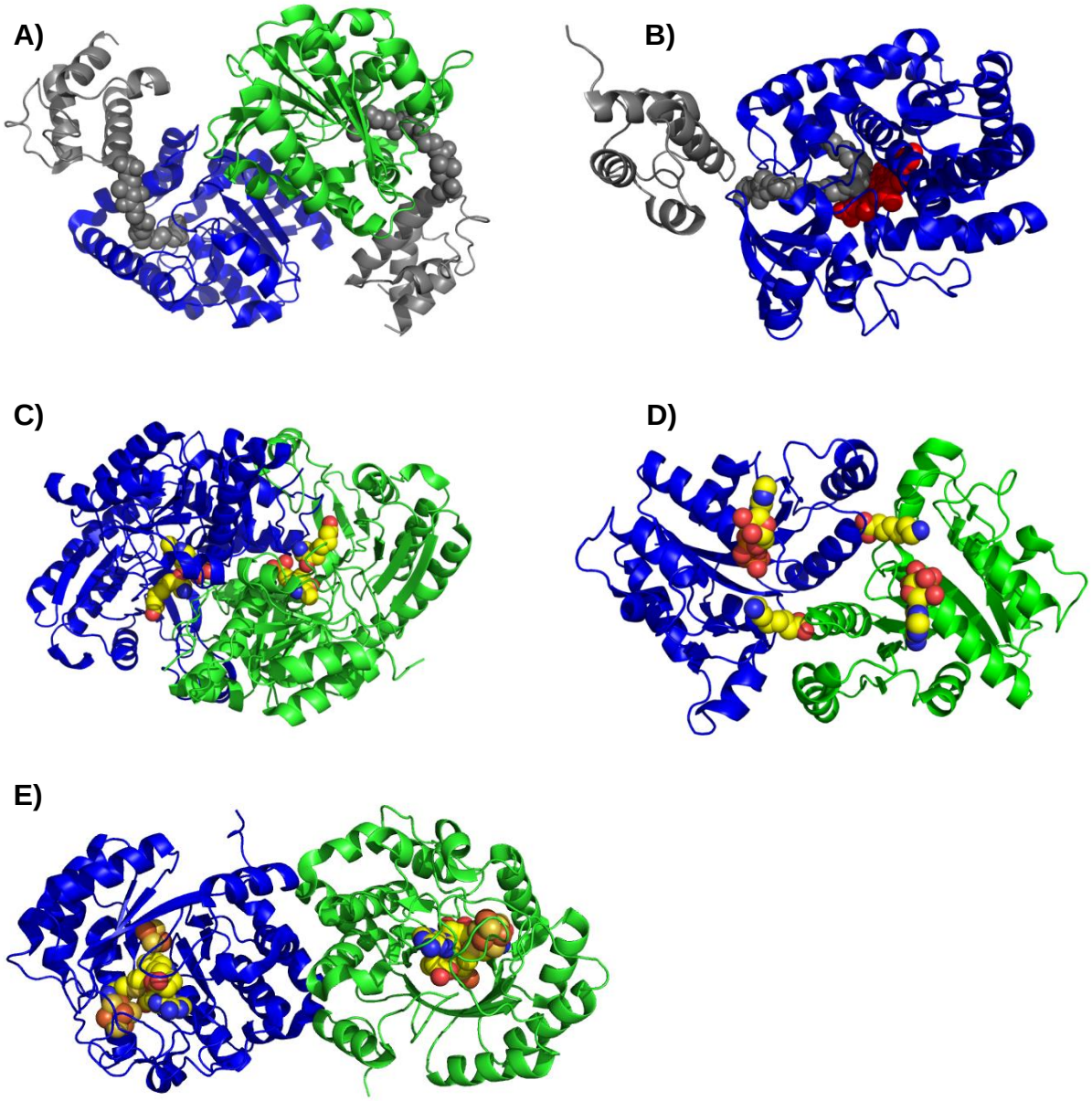


Fig. 1.7 X-ray structures of enzymes from the biotin biosynthetic enzymes **A)** *Shigella flexneri* BioH in complex with ACP (grey cartoon) and the phosphopantetheine-linked pimeloyl-methyl ester (grey spheres, PDB code: 4ETW [35]). **B)** *B. subtilis* BioI in complex with ACP, phosphopantetheine-linked fatty acid (shown in grey) and heme (in red, PDB code: 3EJE [34]). **C)** *E. coli* DANS in complex with cofactor PLP and substrate AON (AON, spheres with carbon atoms colored yellow, PDB code: 1QJ3 [36]). **D)** *Helicobacter pylori*

DTBS in complex with ATP (PDB code: 3MLE [37]). **E)** *E. coli* biotin synthase in complex with substrate dethiobiotin, S-adenosyl-L-methionine (SAM), clusters Fe_4S_4 and Fe_2S_2 (shown as blue/red/yellow spheres, PDB code: 3MLE [38]).

1.3 8-amino-7-oxononanoate synthase (AONS, BioF)

1.3.1 Introduction of AONS

8-amino-7-oxononanoate synthase (also known as 7-keto-8-aminopelargonic acid, AONS, EC 2.1.3.47, *BioF* gene product) catalyses the condensation of pimeloyl-CoA with L-alanine to produce 8-amino-7-oxononanoate (AON) in the first committed step of biotin biosynthesis. In 1998, Alexeev and colleagues in Edinburgh resolved the X-ray crystal structures of the pyridoxal 5'-phosphate (PLP) bound form of *E. coli* AONS (PDB code: 1BS0), which showed great homology with other PLP-dependent enzymes [39]. As shown in **Fig. 1.8 A**, AONS presents as a symmetric homodimer, each monomer composes of three domains: N-terminal, central and C-terminal. Residue Lys 236 is the conserved, active site lysine that forms the internal aldimine with PLP cofactor. The amino acid sequence alignment between AONS and other members of the α-oxoamine synthase (AOS) family such as the catalytic domain of human erythroid5-aminolevulinatesynthase (ALAS), serine palmitoyl transferase (SPT) and the 2-amino-3-oxobutyrateCoA ligase (AKB or KBL) shows the active site residues to be highly conserved [39]. In 2000, Webster reported the structure of *E. coli* AONS in complex with the product AON in the form of a PLP:AON external aldimine (PDB

code: 1DJ9) [22]. As shown in **Fig. 1.8 B**, the terminal carboxylate of AON interacts with Arg21 which locates in the N-terminal domain of AONS.

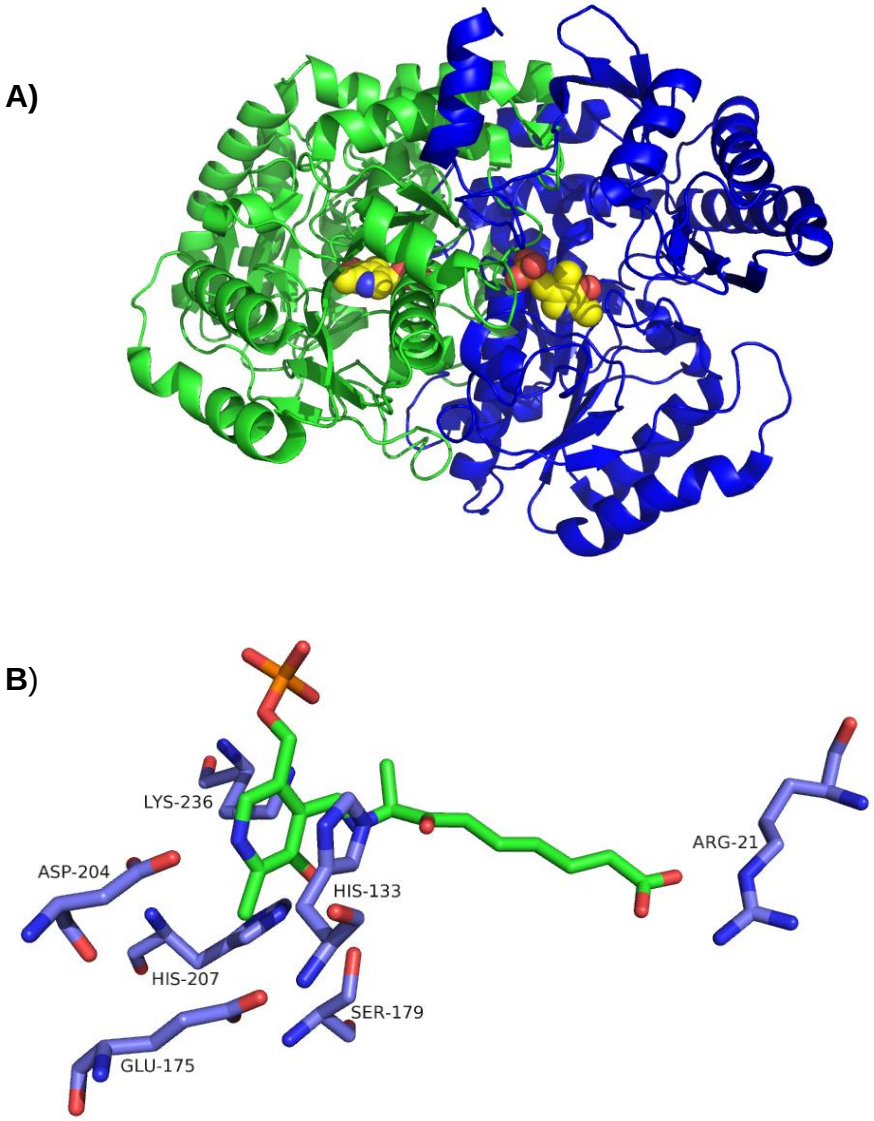


Fig. 1.8 X-ray crystal structures of **A)** the dimer of *E. coli* AONS, one monomer in green, one in blue, PLP cofactor is shown as spheres with carbon atoms colored yellow (PDB code: 1DTY); **B)** Conserved active site residues of AONS (in blue) bind to the PLP:AON product (in green) (PDB code: 1DJ9) [22].

Along with the structural study, the K_M values for substrates pimeloyl-CoA and L-alanine of *E. coli* AONS were also reported by Webster [22] using the DTNB (5,5'-dithiobis(2-nitrobenzoic acid)) assay that monitors the release of the free thiol of CoASH. The K_M for pimeloyl-CoA and L-alanine was found to be 25 ± 2 μM and 0.50 ± 0.04 mM respectively.

1.3.2 Proposed mechanism of AONS

The mechanism of AONS is consistent with other PLP-dependent enzymes. As shown in **Fig. 1.9**, a conserved lysine residue (Lys 236) of AONS firstly reacts with the PLP cofactor to form an internal aldimine **(I)**. The binding of L-alanine displaces Lys 236 to form an external aldimine complex **(II)** which is relatively stable and deprotonation only happens in the presence of the second substrate, pimeloyl-CoA. The formation of this external aldimine complex causes the increase in the absorbance at 425nm, thus it allows the K_d of L-alanine with AONS to be measured. The binding of pimeloyl-CoA then leads to rotation of the external aldimine complex about the Cα, thus forming the Dunathan intermediate **(III)**. The conserved Lys 236 acts as a base to deprotonate the external aldimine complex, causing the formation of a quinonoid intermediate **(IV)**. Next a Claisen condensation happens between the quinonoid and pimeloyl-CoA to form the β-ketoacid intermediate **(V)** and release of CoASH. The β-ketoacid intermediate then undergoes a decarboxylation reaction to form the quinonoid **(VI)**, which is protonated after then to produce the external aldimine **(VII**). Finally, the

product AON is released from the enzyme and along with that, the internal aldimine **(I)** is regenerated.

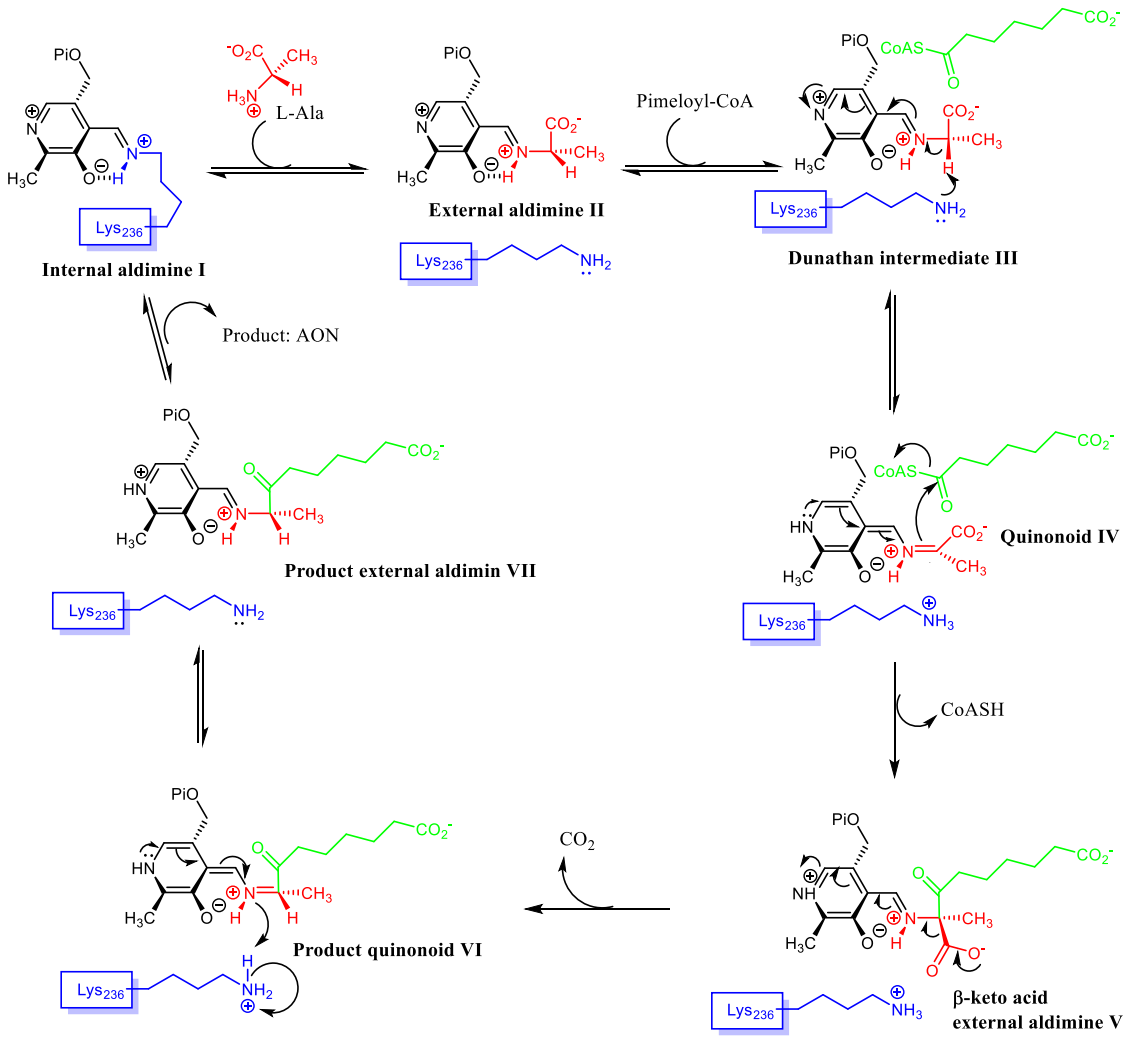


Fig. 1.9 Proposed catalytic mechanism of enzyme AONS from Webster et al [22]. L-alanine is in red, pimeloyl-CoA is in green, enzyme amino acids in blue, PLP cofactor is in black.

1.4 Pimeloyl-CoA synthetase (PCAS，BioW)

1.4.1 Introduction of PCAS

Pimeloyl-CoA synthetase (also known as 6-carboxyhexanoate-CoA ligase and pimeloyl-coenzyme A (CoA) ligase, PCAS, EC 6.2.1.14, *BioW* gene product) is an important enzyme involved in the first step of biotin biosynthesis in some specific organisms (such as *Bacillus megaterium*, *Bacillus sphaericus* and *Pseudomonas mendocina*) [5]. It catalyses the ATP-dependent transformation of exogenous pimelic acid into pimeloyl-CoA, pyrophosphate (PP_i) and AMP [40, 41] (as shown in **Fig. 1.10**).

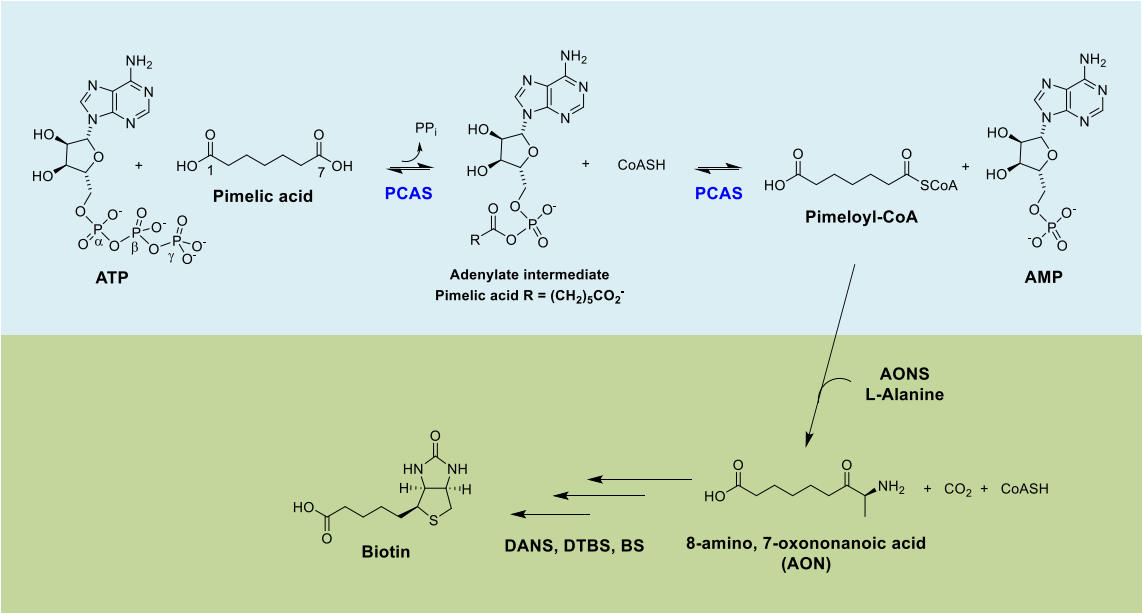


Fig. 1.10 The pimeloyl-CoA synthetase (PCAS) reaction and its role in biotin biosynthesis. The blue part shows the reaction catalysed by PCAS undergoes an adenylate forming mechanism which comprises two partial steps. Firstly, PCAS activates the unreactive carboxylic acid of pimelic acid by transforming the hydroxyl group into an adenosine monophosphate (adenylate) intermediate which is a much better leaving group. Then the pimeloyl-adenylate is attacked by CoASH to produce pimeloyl-CoA and liberate AMP. After this, pimeloyl-CoA acts as the substrate for AONS which is then further converted to biotin as shown in the green part.

At the beginning of this project, the function of this *BioW*-coded protein from *Bacillus subtilis* (Uniprot cod P53559) has been inferred from its sequence homology to other enzymes, in particular the *Bacillus sphaericus* PCAS studied by Ploux in 1992 with which it shares 45.0 % sequence homology [42], and it was thought to be involved in the ligation of pimelic acid and CoASH. In the second year of my PhD, Cronan and Manandhar at the University of Illinois, published a paper of the same *Bacillus subtilis* PCAS reporting on its purification and initial biochemical analysis, and this confirmed previous speculations [43]. The *B. subtilis* PCAS is composed of 259 aa, and its predicted molecular weight is 29,630 Da. The proposed mechanism is also shown in **Fig. 1.10**. The enzyme firstly activates the carboxylic acid by transforming the hydroxyl group into a pimeloyl-adenylate intermediate which can be attacked by a CoASH nucleophile to form the final product.

1.4.2 Adenylate forming enzymes

1.4.2.1 Introduction

The activation of carboxylic acids by transformation to adenylates is a common strategy in nature and exemplified in the ANL superfamily of adenylating enzymes that includes the adenylation domains of Non-Ribosomal Peptide Synthetases (NRPSs), the acyl- and aryl-CoA synthetases and oxidoreductases (firefly luciferases) [44, 45]. Indeed, the carboxylate of biotin is also activated by adenylation by a biotin protein ligase (BPL) for final transfer to the lysine side chain of carboxylase enzymes to form an amide bond. This strategy is quite similar to the use of triflates or acid

chlorides in organic synthesis which creates good leaving groups for the next step which is nucleophilic attack.

The core of the reaction is the attack of the carboxylate at the α-phosphate of ATP to form the adenylate with PP_i elimination driving the reaction as shown in **Fig. 1.11**. The reaction transition state is proposed to be the negatively charged penta-coordinate trigonal bipyramidal phosphorous atom, which is common to other triphosphate nucleotide utilising enzymes including kinases and DNA polymerases [45]. Interestingly a structural link between kinases and one type of adenylase has been identified [46, 47].

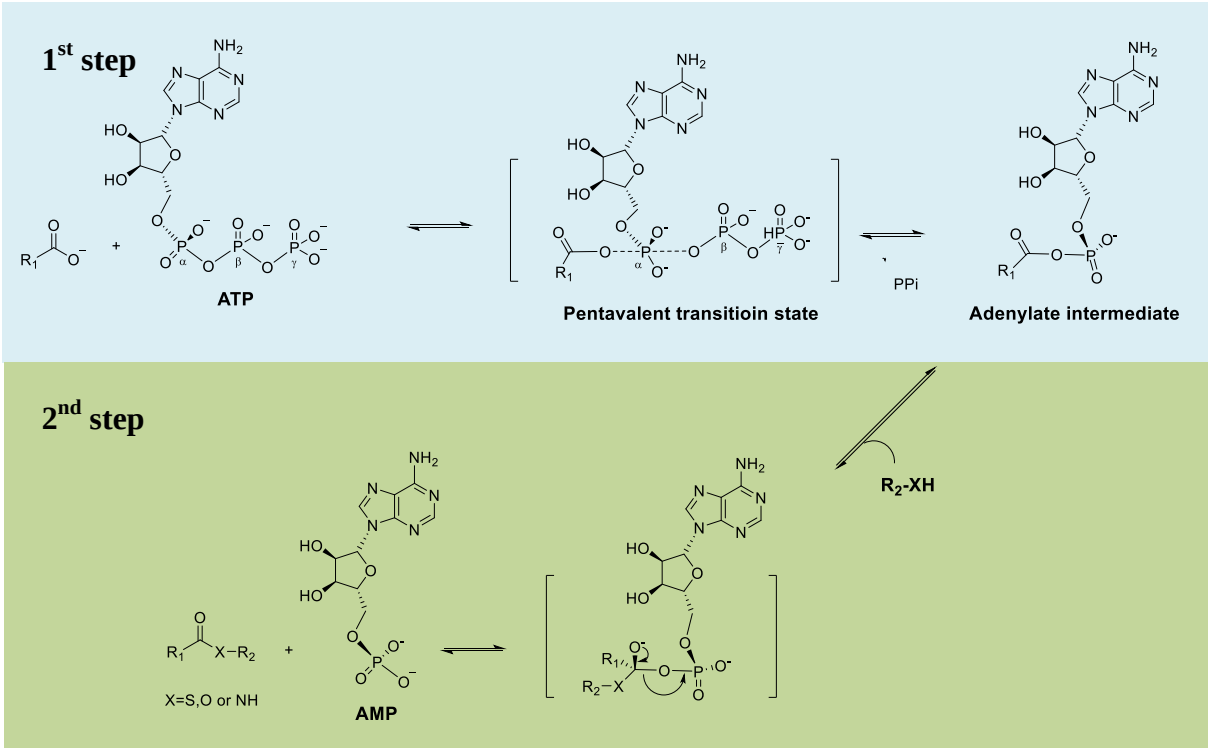


Fig. 1.11 This shows the proposed general two-step reaction catalyzed by the adenylate-forming enzymes of the ANL superfamily.

Since the initial discovery of acyl-CoA synthetases in the 1950s [44], the family of adenylate forming enzymes began to grow as enzymes that generated an adenylate

intermediate were isolated. In the following years, studies of other members of adenylating enzyme family including firefly luciferase identified by McElroy *et al*. in 1967 [48] and medium-chain fatty acyl-CoA synthetase identified by Bar-Tana in 1968 [49] revealed that the changes of the conformational states might catalyze the two partial reactions. In recent years, these statements from 40 years ago were confirmed by the investigation of the structural biochemistry of these adenylating enzymes.

1.4.2.2 Classification scheme of adenylate forming enzymes

The classification scheme of adenylate forming enzymes was previously put forward by Fulda based on amino acid sequence analysis [50], which were divided into three classes - the adenylation domains of Non-Ribosomal Peptide Synthetases (NRPSs), the acyl- and aryl-CoA synthetases and oxidoreductases (e.g. firefly luciferases). In a review in 2009, Schmelz and Naismith proposed a broader classification scheme that also includes structurally distinct adenylating enzymes [45]. As shown in **Table 1.2**, the first class of adenylating enzymes comprises the former ANL superfamily with three subclasses Ia,Ib and Ic; the second class comprises constitutes aminoacyl-tRNA synthetases with two subclasses IIa and IIb; and the third class contains NRPS-independent siderophores (NIS) enzymes with four subclasses IIIa, IIIb, IIIc and IIId.

Class I – ANL family	
Subclass Ia	The adenylation domains of NRPSs (1AMU)
Subclass Ib	Acyl- and aryl-CoA synthetases (3DMY, 2P2F, 3CW8)
Subclass Ic	Oxidoreductases (firefly luciferases) (2D1Q)
Class II	
Subclass IIa	Aminoacyl-tRNA synthetases class I (1M83,1OBC, 2QUI)
Subclass IIb	Aminoacyl-tRNA synthetases class II (2PME, 1E24)
Class III	
Subclass IIIa	NIS type A (2W02)
Subclass IIIb	NIS type B
Subclass IIIc	NIS type C (AlcC, 2X0Q)
Subclass IIId	NIS type A'

Table 1.2 This shows the new classification scheme proposed by Schmeltz and Naismith of adenylate-forming enzymes which is composed of three main subfamilies. The PDB codes of representative crystal structures of the enzymes in different classes are shown in red colour.

Class I enzymes are all composed of a small C-terminal domain and a larger N-terminal domain, they are linked via a flexible loop and the interface between these two domains forms the ATP binding site (**Fig. 1.12a**). Class I enzymes undergo large conformational changes during their catalytic cycle, the C-terminal domain rotates in relation to the N-terminus in the presence of CoASH [44, 48]. Class II enzymes shares no conserved domain arrangement and here structure of human glycyl-tRNA synthetase (PDB code: 2pme, **Fig. 1.12b**) [51] is shown as an example. Class III enzymes contains three domains: N-terminal, central and C-terminal (**Fig. 1.12c**).

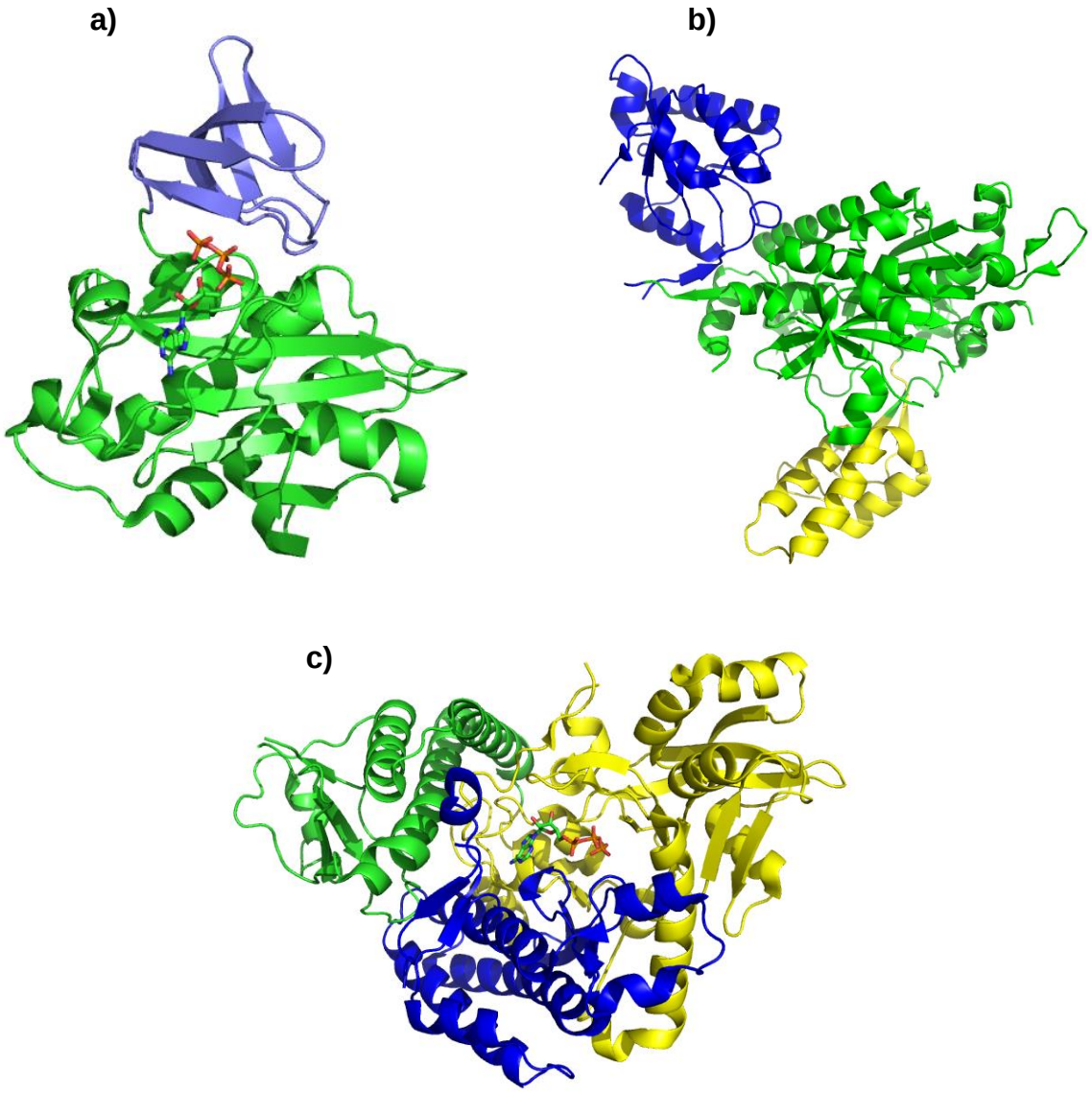


Fig. 1.12 Structures of different classes of adenylating enzymes. **(a)** Biotin protein ligase (BPL) from *Aquifex aeolicus* in complex with ATP (sticks) (PDB code: 3efs) [16] , the N-terminal domain is in green and C-terminal domain in blue. **(b)** Human glycyl-tRNA synthetase (PDB code: 2pme) [51], the central catalytic domain is in green and the anticodon-binding domain in blue, insertion I domain in blue. **(c)** Achromobactin synthetase protein D (AcsD) from *Pectobacterium chrysanthemi* in complex with ATP (sticks) (PDB code: 2W02)

[47], the N-terminal domain is in green, central domain in yellow and C-terminal domain in blue.

All adenylating enzymes require Mg^{2+}, a divalent ion that coordinates the phosphate by neutralizing the charge of ATP and stabilizing the pyrophosphate (PPi) leaving group. The α phosphate of ATP always contacts with positively charged residues including Arg, Lys and His. This is because these positively charged residues enhance the electrophilicity of the α phosphate of ATP and at the same time, polarize the negatively charged carboxylate group of the incoming substrate thus enhancing its nucleophilicity [45].

1.4.2.3 The role of PCAS in the family of adenylate forming enzymes

Given the diversity of carboxylic acid substrates used in numerous biosynthetic pathways, it is perhaps not surprising that the adenylating enzymes seem to be a very diverse enzyme family [45]. In my first year PhD, based on its function, PCAS was thought to be a class Ib adenylating enzyme (acyl- or aryl CoA synthetases) in the classification scheme put forward by Schmeltz and Naismith. However, sequence analysis of the *B. subtilis* PCAS showed no homology with other known adenylases or ATP-dependent synthetases, suggesting that PCAS may represent another member of the adenylating enzyme superfamily. The lack of sequence alignment to known enzymes meant there was no guide as to the molecular basis of recognition of the substrates, as well as leaving open the question of the constellation of active site

residues that catalyse the two steps in the reaction; pimeloyl-adenylate and acyl-CoA thioester formation. The proposed mechanism is shown in **Fig. 1.10** – the enzyme firstly activates the carboxylic acid by transforming the hydroxyl group into a pimeloy-adenylate intermediate which can be attacked by a nucleophile to form the final product. In PCAS, the thiol group of CoASH is the nucleophile; but an alcohol or amine groups are also used by other adenylate-forming enzymes.

1.5 Directed evolution

1.5.1 Introduction of biocatalysis

Biocatalysis is the application of natural catalysts, such as enzymes and microbes in synthetic chemistry [52]. It is increasingly being used as an addition to traditional chemical catalysis both in academic and industrial laboratory [53]. Its products range from bulk and fine chemicals such as pharmaceuticals, to agrochemicals, food, animal feed and biofuels [54-56]. Compared to metallocatalysis and organocatalysis, biocatalysis shows high selectivity (chemo-, regio-, an enantio-), high efficiency with environmentally friendly reaction conditions [52, 57, 58]. Unfortunately, in most cases naturally occurring enzymes are not optimal for the practical applications of interest [59], and in this respect the directed evolution of enzymes has emerged as a powerful enabling technology [54].

In 1972, directed evolution was firstly used by Francis and Hansche for improving certain functional properties of acid phosphatase in *Saccharomyces cerevisiae* [60], but due to the limitation of molecular biology technology at that time, the library was

simply composed of several naturally occurring mutated stains and was screened only *in vivo* without any *in vitro* protein analysis. From the late 1990s, attributed to the emerging progress in bioinformatics, DNA-recombination and high-throughput screening methods, directed evolution entered a new era with the work of Frances Arnold, Willem Stemmer and Manfred Reetz.

1.5.2 Procedures of directed evolution

Directed evolution, that mimics Darwinian 'survival of the fittest' in the test tube, involves repeated cycles of mutagenesis and selection/screening of resulting variants for mutations with desired properties [52, 53]. It deals with various enzyme engineering problems, including enhancing stability [61-63], tolerance towards pH and organic solvents [64, 65], improving activity, changing or extending specificity and altering the enantioselectivity [66].

In general, directed evolution is composed of four steps (**Fig. 1.13**) [53].The initial step is to select an appropriate parent protein. Usually, a protein containing a similar target function but being not optimal for practical application will be considered. Secondly, a mutant variant library will be generated based on the parent protein. This step involves 1) creating a gene library by mutation or recombination; 2) inserting the gene into an expression plasmid; 3) expressing the gene library an appropriate bacterial host. The third step is to select/screen variants with improved desired functions. And finally, the entire process needs to be repeated until the target with the desired properties has been achieved.

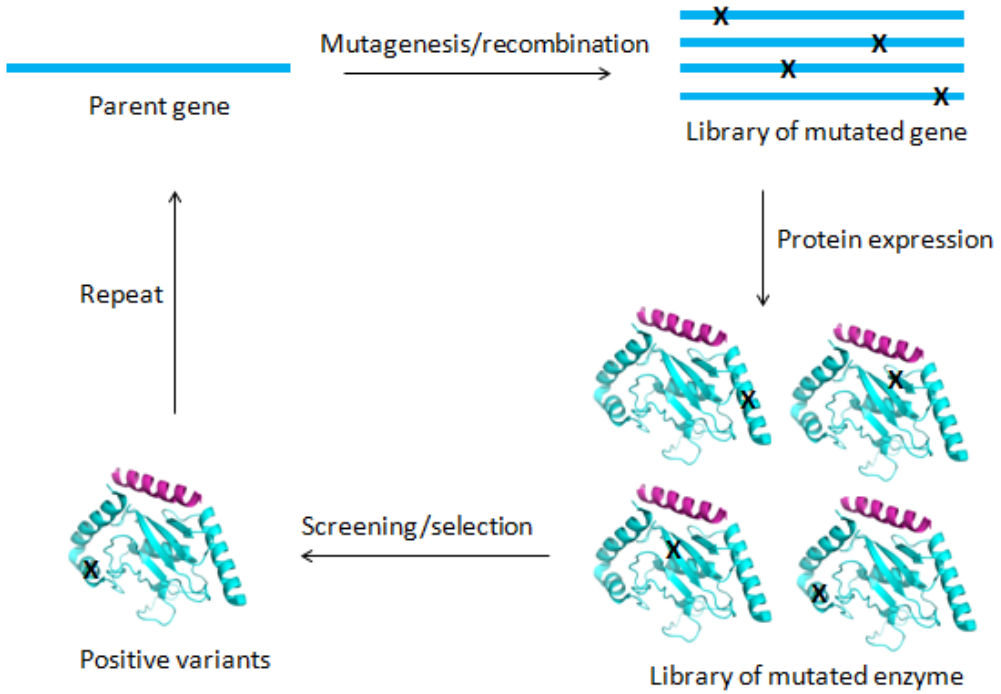


Fig. 1.13 General scheme of directed evolution involved in: selection of parent protein, generation of mutant variant library, selection/screening of the positive variants and finally repeating the whole process until the target with the desired function has been obtained.

1.5.3 Generating libraries

There are two types of directed evolution libraries – a "random mutant" library and a "smart" library. When only limited prior knowledge of the parent enzyme is available (for example, only an amino acid sequence with no crystal structure or high homology model), the library can be vast with every amino acid (except the ATG start codon) open to mutagenesis. A wide variety of methods have been developed to create a mutant library, including using mutator strains, error-prone PCR, *in vivo* and *in vitro* chemical mutagenesis, DNA shuffling, etc.

On the other hand, unlike the random mutagenesis method, much smaller libraries may also be used where mutations are limited to a small number of amino acid positions (catalytic or substrate binding residues) in the target protein. [67] The successful design of a smart library relies largely on three main factors: 1) understanding of structure – function relationships of the target protein; 2) the consequences of mutations on protein folding and activity; 3) methods for identifying positive mutations [53]. Although more bio-informatic analysis of the parent protein is required, a smart library can largely minimise the library size and enhance the success rate of directed evolution.

Currently, whole gene randomization, specific position randomization and recombination remain the cornerstones of the diversity generating methods [68]. All these methods aim to create comprehensive but less biased libraries.

Aims of this thesis

The aims of Chapter III were:

1) To express of *Bacillus subtilis BioW* gene in *E. coli* strain, and then isolate and characterise of the PCAS protein. *Bacillus subtilis* PCAS was chosen in this project because the X-ray crystal structure of BioI was obtained from the same species [30].
2) Determine the X-ray crystal structure of *Bacillus subtilis* PCAS to understand its catalytic mechanism and to provide information for protein engineering.
3) To generate mutants with altered substrate specificity by re-designing the active site of PCAS.

The aims of Chapter IV were:

1) To express and isolate *Escherichia coli* 8-amino-7-oxononanoate synthase (AONS) from an *E. coli* host.
2) To express and purify C*orynebacterium amycolatum SK46* BioWF fusion from an *E. coli* host.

Chapter II. Materials and Methods

2.1 Materials and Reagents

All standard reagents, chemicals and media were from Sigma-Aldrich, New England Biolabs or Fisher unless otherwise stated. CoASH was purchased from Chem-Impex International. 7-methyl-6-thioguanosine (MESG) was from Berry & Associates. All chromatography columns were from GE Healthcare. BioTek Synergy HT plate reader was used in this project and the 96 well UV-transparent plates were purchased from Costar. The synthetic DNA construct for *Bacillus subtilis* pimeloyl-CoA synthetase (PCAS, *BioW*) and *Corynebacterium amycolatum SK46* BioWF fusion were designed and purchased from GenScript. The plasmid contains *E*sch*e*richia *coli* 8-amino-7-oxononanoate synthase (AONS, *BioF*) was available from our lab prior to the AONS project. The expression plasmid for TEV protease was a generous gift from Prof. Jim Naismith's group.

2.1.2 Cell stocks

The *E. coli* host strain was plated onto LB agar and incubated at 37^{o}C overnight. A single colony was picked into LB (5-10 ml) supplemented with appropriate antibiotics under sterile conditions and grown overnight at 37 ^{o}C with agitation. Next day, autoclaved glycerol was added to a final conc. of 20%. The cell stock was then stored at -80^{o}C freezer.

2.1.1 Competent *E. coli* host strains

Cell line	Company	Genotype	Application
BL21 (DE3)	Novagen	F$^-$ *omp*T *hsd*S$_B$ (r_B^- m_B^-) *gal dcm* (DE3)	Transformation and expression host.
Rosetta (DE3)	Novagen	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)	Transformation and expression host.
C2987	New England Biolabs	fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Transformation of PCR and ligations, host for plasmid storage.
DH5α™	Invitrogen	F$^-$ φ80d*lac*ZΔM15 Δ(*lac*2YA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (r_K^-, m_K^-) *pho*A *sup*E44λ$^-$ *thi*-1 *gyr*A96 *rel*A1	Transformation of plasmid, host for plasmid storage.
E. coli ΔBioH	Coli Genetic Stock Centre (CGSC# 12019)	F$^-$ Δ*(araD-araB)567* Δ*lacZ4787*(::rrnB-3) λ$^-$ Δ*bioH756::kan* *rph-1*Δ*(rhaD-rhaB)568 hsdR514*	*E. coli* ΔBioH mutation strain for pimeloyl-CoA synthetase (*BioW*) complementation assay.
E. coli ΔBioH (DE3)	This work	F$^-$ Δ*(araD-araB)567* Δ*lacZ4787*(::rrnB-3) λ$^-$ Δ*bioH756::kan* *rph-1*Δ*(rhaD-rhaB)568 hsdR514* (DE3)	T7 promoter compatible *BioW* selection host.
B834 (DE3)	A gift from Dr Jon Marles-Wright's group	*E. coli* methionine auxotroph strain	Expression host for seleno-methionine labelled protein.

Table 2.1 Competent *E. coli* host strains and applications

2.1.3 Sterile conditions

All flasks, eppendorfs, tips, liquid and solid media were sterilized using the LTE touchclave lab autoclave (at 120 ^oC, 15 psi for 20 mins) before use.

2.1.4 Growth media

All media were prepared by dissolving the components (**Table 2.2**) in distilled water, adjusting pH to 7.5 and sterilizing. Appropriate antibiotic of the required concentration was added on cooling.

Growth Media	Components
Luria Bertani (LB)	Tryptone (10 g/L), yeast extract (5 g/L), sodium chloride (10 g/L)
Super optimal broth (SOC)	Tryptone (20 g/L), yeast extract (5 g/L), sodium chloride (10 mM), potassium chloride (2.5 mM), magnesium chloride (10 mM), magnesium sulphate (10 mM), glucose (2 % w/v) – filter sterilized.
LB Agar	Tryptone (10 g/L), yeast extract (5 g/L), sodium chloride(5 g/L), agar (15 g/L)
S-Gal Agar	Tryptone (10 g/L), yeast extract (5 g/L), sodium chloride (10 g/L), Agar (12 g/L), S-Gal (0.300 g/L), Ferric ammonium citrate (0.500 g/L), IPTG (0.030 g/L)

Table 2.2 Components for preparing growth media

2.1.5 Antibiotic solutions

All antibiotic stock solutions were prepared according to **Table 2.3** and then 0.22 μm filter sterilized.

Antibiotic	Protocol
Ampicillin	A 100 mg/mL stock solution was made using distilled water and stored at 4 ℃. The required volsume was added to media to give a final concentration of 100 μg/mL.
Kanamycin	A 50 mg/mL stock solution was made using distilled water and stored at 4 ℃. The required volume was added to media to give a final concentration of 50 μg/mL.
Chloramphenicol	A 25 mg/mL stock solution was made using ethanol and stored at -20 ℃. The required volume was added to media to give a final concentration of 25 μg/mL.

Table 2.3 Components for preparing antibiotic solutions

2.1.6 Buffers

All buffers were prepared by dissolving the components in distilled water based on **Table 2.4**. All buffers were 0.22 μm filtered and degassed before using for chromatography.

Table 2.4 Components for preparing buffers

Buffer	Components
1A	20 mM sodium phosphate pH 7.5, 500 mM NaCl, 10 mM imidazole, 2 mM TCEP, 10 % glycerol with the addition of complete EDTA-free protease inhibitor tablets (Roche), DNase (0.2 mg per 10 mL buffer, Sigma) and lysozyme (1 mg per 10 mL buffer, Sigma)
1B	20 mM sodium phosphate pH 7.5, 500 mM NaCl, 25 mM imidazole, 2 mM TCEP, 10 % glycerol
1C	20 mM sodium phosphate pH 7.5, 500 mM NaCl, 300 mM imidazole, 2 mM TCEP, 10 % glycerol
1D	50 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM β-mercaptoethanol (BME), 10% glycerol, 0.25 mM Phenylmethanesulfonyl fluoride (PMSF)*
1E	20 mM sodium phosphate pH 7.5, 500 mM NaCl, 20 mM imidazole, 2 mM TCEP, 10 % glycerol
1F	10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.25 mM Phenylmethanesulfonyl fluoride (PMSF)*, 1 mM Dithiothreitol (DTT)
2A	PBS, 300 mM NaCl, 10 mM imidazole, 10 % glycerol, 1 mM Phenylmethanesulfonyl fluoride (PMSF)*, 1 mM benzamidine**
2B	PBS, 300 mM NaCl, 30 mM imidazole, 10 % glycerol, 1 mM Phenylmethanesulfonyl fluoride (PMSF)*, 1 mM benzamidine
2C	PBS, 300 mM NaCl, 300 mM imidazole, 10 % glycerol, 1 mM Phenylmethanesulfonyl fluoride (PMSF)*, 1 mM benzamidine**
2D	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 % glycerol, 1 mM Dithiothreitol (DTT), 0.5 mM Ethylenediaminetetraacetic acid (EDTA)
2E	50mM Tris-HCl pH 8.0, 300 mM NaCl, 50 % glycerol, 1 mM Dithiothreitol (DTT), 0.5 mM Ethylenediaminetetraacetic acid (EDTA)
3A	20 mM potassium phosphate pH 7.5, 150 mM NaCl,10 mM imidazole
3B	20 mM potassium phosphate pH 7.5, 150 mM NaCl, 1 M imidazole
3C	50 mM Tris-HCl pH 8.0, 100 mM NaCl

4A	50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol, 5 mM TCEP with the addition of complete EDTA-free protease inhibitor tablets (Roche)
4B	50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol,1mM Dithiothreitol (DTT)
4C	50 mM Tris-HCl pH 7.5, 1M NaCl, 10% glycerol,1mM Dithiothreitol (DTT)
4D	50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10% glycerol,1mM Dithiothreitol (DTT), 1 M Ammonium sulfate
4E	10 mM Tris-HCl pH 8.0, 150 mM NaCl
5A	20 mM potassium phosphate pH 7.5, 100 mM NaCl, 100 mM PLP with the addition of complete EDTA-free protease inhibitor tablets (Roche)
5B	20 mM potassium phosphate pH 7.5, 100 mM NaCl, 100 mM PLP, 10mM midazole
5C	20 mM potassium phosphate pH 7.5, 100 mM NaCl, 100 mM PLP, 300mM midazole
5D	20 mM potassium phosphate pH 7.5,100 μM PLP, 2 mM EDTA
6A	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol with the addition of complete EDTA-free protease inhibitor tablets (Roche)
6B	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6C	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 50 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6D	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 100 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6E	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 150 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6F	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 200 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6G	20 mM potassium phosphate pH 7.5, 500 mM NaCl, 300 mM imidazole, 2 mM β-mercaptoethanol (BME), 10% glycerol
6H	20 mM potassium phosphate pH 7.5, 100 mM NaCl, 25 mM PLP, , 10% glycerol

* Prepared from a 100 mM PMSF stock (toxic, 0.174 g PMSF dissolved in 10 mL of isopropanol or ethanol, used fresh stock every time)

** Prepared from a 100mM benzamidine stock (in methanol)

2.2 Cloning

2.2.1 Plasmids

The plasmids in **Table 2.5** were used for cloning and protein expression in this thesis.

Plasmid	Antibiotic resistance	Application	
pGEM-T Easy (Promega)	Ampicillin	Cloning in *E. coli*	Used for the cloning of PCR products, is compatible with blue/white screening on S-gal plates
pET28a	Kanamycin	Expression in *E. coli*	Protein can be expressed with an N-terminal his_6 tag.
pET-22b	Ampicillin	Expression in *E. coli*	Protein can be expressed with a C-terminal his6 tag.
pEHISTEV [69]	Kanamycin	Expression in *E. coli*	Protein can be expressed with an N-terminal TEV protease cleavable his_6 tag.

Table 2.5 Plasmids and their application

2.2.2 Primers

Primer oligonucleotides used in polymerase chain reaction (PCR) for gene amplification were designed to be around 20 to 30 base pairs in length as shown in **Table 2.6**. All of them were purchased from Sigma-Aldrich and stored at -20 ℃ as 10 μM stocks.

Primers	Sequence (5'-3')
PCAS/NTH forward	CCATGGAAGAAGAAACTTT (NcoI)
PCAS/NTH reverse	CTAGTTATTGCTCAGCGGT
AONS/CHis forward	GCCCATATGAGCTGGCAGGA (NdeI)
AONS/CHis reverse	CTCGAGACCGTTGCCATGCAGCACCT (BamHI)

Table 2.6 Primers used for PCR. The bases underlined are the restriction sites.

2.2.3 Polymerase chain reaction (PCR)

The reaction mixture contains DNA template (100 ng), forward primer (5 μL of 10 μM stock), reverse primer (5 μL of 10 μM stock), dNTP (1 μL of 10mM stock), reaction buffer (10 μL, x5), Phusion High-Fidelity DNA polymerase (1 μL, 2 U, New England Biolabs) and filter sterilized water to obtain a final volume of 50 μL was prepared on ice. The thermal cycling was then carried out in a PCR machine according to the protocol shown in **Table 2.7**. For A-tailing reaction, one Taq Ready-To-Go PCR beads was added to the 50 μl PCR reaction mixture from above and incubated for another 20 min at 72 ^{o}C in the PCR machine. The reaction is then terminated by adding DNA blue loading dye (10 μl, x6, New England Biolabs) and purified by DNA gel electrophoresis, which is cloned into a pGEM T-Easy empty vector.

Steps	Cycle	Temperature (^{o}C)	Time
Initial Denaturing	1	98	30 s
Denaturing	30	98	10 s
Annealing		55	30 s
Extension		72	30s/kbp
Final Extension	1	72	10 min
Hold at 4 ^{o}C			

Table 2.7 Polymerase Chain Reaction (PCR) protocol

2.2.4 DNA gel electrophoresis

1 g of agarose was added to 100 mL of TAE buffer (x1) and heated in a microwave until the agarose powder all dissolved. Upon cooling down to around 50 ^{o}C, gel red

(10 μL, Cambridge Bioscience) was then added and they were mixed well by swirling. After this, the agarose solution was poured into a casting mould and allowed to set at room temperature. The settled 1 % agarose gel was then transferred to an electrophoresis tank containing TAE buffer (x1). DNA samples containing DNA blue loading dye were then loaded onto the gel together with the DNA marker in a separate lane and run at 100 volts for 50 min.

2.2.5 Gel extraction

The DNA was cut from the agarose gel and then purified using the gel extraction QIAquick kit (QIAGEN).

2.2.6 Plasmid ligation

The amplified gene was incubated with vector (the ratio of vector/insert is 1:5 w/w, and the total DNA weight is around 100 ng), T4 DNA ligase (1 μL, 1 x 10^6 ligations/unit), reaction buffer (2 μL, x10) and filter sterilized water to obtain a final volume of 20 μL for 16 hours at 16 ^oC. After this, 4 μL of the ligation product was used to transform 50 μL of C2987 competent cells. The cells were then spread on LB agar plate supplemented with the appropriate antibiotic and incubated at 37^oC overnight. DNA cloned to pGEM T-Easy vector was grown on S-Gal agar plates supplemented with the appropriate antibiotic for blue/white screening.

2.2.7 Restriction digest

Reaction mixture contains template DNA (3 μL), restriction buffer (1 μL, 10x, New England Biolabs), restriction endonuclease (1 μL, New England Biolabs) and filter sterilized water to obtain a final volume of 10 μL was prepared and incubated for 3 h at 37 ℃. After addition of 2 μL of DNA blue loading dye (6x, New England Biolabs), this reaction mixture was loaded onto a 1% agarose gel, followed by DNA gel electrophoresis separation.

2.2.8 Plasmid DNA isolation

Plasmid DNA was isolated from 5 mL of LB overnight growth culture using a QIAprep Spin Miniprep Kit (Qiagen) and then stored at -20 ^{o}C.

2.2.9 Big dye DNA sequencing

The sequencing reactions contains 5 μL of plasmid DNA, 1 μL of primer (10 μM stock), 2 μL of Terminator 3.1 buffer and 2 μL of Big Dye master mix version 3.1 were prepared on ice, which were down in pairs (one with the forward primer and one with the reverse, the sequence information of these primers is shown in **Table 2.8**). After the thermal cycles (25 cycles of 95 ^{o}C for 30 s, 50 ^{o}C for 20 s and 60 ^{o}C for 4 min) in a PCR machine, the DNA sample was send for sequencing using the Sanger dideoxy chain termination method by the Genepool at the Ashworth Laboratories, the University of Edinburgh.

Sequencing Primers	Plasmid	Sequence (5'-3')
pET forward	pET	TTAATACGACTCACTATAGGG
pET reverse	pET	CTAGTTATTGCTCAGCGGT
pGEM T7 forward	pGEM	TAATACGACTCACTATAGGG
pGEM T7 reverse	pGEM	ATTTAGGTGACACTATAGAA

Table 2.8 Sequencing Primers used for DNA in pET and pGEM vectors

2.2.10 Transformation of *E. coli*

2.2.10.1 Heat shock

Chilled plasmid DNA (2 µl) was gently mixed with chemical competent *E. coli* cells (25 µl) and incubated on ice for 25 min. After 40 s heat shocking at 42 ^{o}C, the cells were then placed on ice for a further 2 min. 100 µl of preheated SOC media was added to the cells which were then agitated at 37 ^{o}C, 200 rpm for 1 h. This culture (25 µl for plasmid transformation, the entire mixture for cloning steps, e.g. mutagenesis or ligations) was evenly spread on LB agar plates supplemented with appropriate antibiotic using sterile techniques. Plates were then incubated at 37 ^{o}C overnight.

2.2.10.2 Electroporation

Chilled plasmid DNA (2 µl) was added to electrocompetent *E. coli* cells (50 µl) in a 0.2 cm gap electrocuvette (BioRad) and electroporated with a single pulse of 1.7 kV, set to a capacitance of 25 µF and a resistance of 200 Ω. 1 mL of preheated SOC media was immediately added to the cells which were then agitated at 37 ^{o}C, 200 rpm for 1 h. This culture (100 µl) was evenly spread on LB agar plates supplemented with appropriate antibiotic using sterile techniques. Plates were then incubated at 37 ^{o}C overnight.

2.3 Protein expression

2 L of LB broth of *E. coli* BL21 (DE3) transformed with the resulting plasmid were grown until the OD_{600} reached 0.6-0.8 and then induced with 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 30 ℃ for 4 h. The cells were then harvested by centrifugation (25mins, 3,500rpm) and stored at -80 ^{o}C until needed for protein purification. And this protein expression step was monitored by running a SDS-PAGE gel.

2.4 Protein isolation

All purification steps were performed at 4 ℃ or on ice.

2.4.1 Purification of *B. subtilis* PCAS/NTH

2.4.1.1 Cell lysis by cell disruptor

Cells contains *B. subtilis* PCAS/NTH were resuspended in buffer 1A (**Section 2.1.6**, **Table 2.4**) and lysed with a cell disruptor at 30 kPSI (Constant Systems Ltd). The crude lysate was cleared by centrifugation (45,000 *g*, 4 ℃, 30 min) and filtered through a 0.22 μm filter.

2.4.1.2 The 1st nickel column

The supernatant after the centrifugation step was then loaded onto a 5 mL His-trap (GE Healthcare) column attached to an AKTA system. Nickel column belongs to affinity column; it retains His_6 tagged protein on the stationary phase of the column.

The column was then washed with the same buffer 1A without the protease inhibitor tablet and DNase. The PCAS/NTH was eluted with an increasing imidazole gradient from 25mM (buffer 1B) to 300 mM (buffer 1C) over 20 column volumes (100 ml). Eluted PCAS/NTH was dialysed against 1L of buffer 1D for 1h.

2.4.1.3 Overnight TEV protease cleavage and the 2nd nickel column

The His_6 tag was cleaved by incubation with TEV protease (mass ratio 1:15) overnight. The uncleaved protein and protease were then removed by passage through a second HisTrap column in buffer 1E which contains 20 mM imidazole.

2.4.1.4 Size exclusion chromatography (SEC)

Finally, the PCAS/NTH sample was loaded onto a gel filtration Superdex 200 16/60 column (GE Healthcare) in buffer 1F and eluted at 1 mL/min monitoring with UV-vis $A_{280\ nm}$. The individual components loaded on it are retained by the stationary phase differently and separate from each other while they are running at different speeds through the column with the eluent according to their different size. Each step of the purification was monitored by SDS-PAGE and the mass of the purified PCAS/NTH was confirmed by LC ESI-MS analysis.

2.4.2 Purification of TEV protease

2.4.2.1 Cell lysis by sonication

Cell pellets containing TEV protease were resuspended in buffer 2A, lysed with sonication using a sonicator and then ultracentrifuged at 20,000 rpm for 25 min.

2.4.2.2 Nickel resin

The cell free extract was incubated with 2 mL of equilibrated Ni-NTA resin for 1 hour at 4 ^{o}C. Then the protein was washed with buffer 2B (30 mM imidazole) and eluted with buffer 2C (300 mM imidazole). Elution fractions were pooled and dialysed against 1 L of buffer 2D for 1 h and then buffer 2E for another 1 h. Then TEV protease was diluted to 1 mg mL^{-1} and stored as 1 mL aliquots at -80 ^{o}C.

2.4.3 Purification of *B. subtilis* PCAS/NHis

2.4.3.1 Cell lysis by sonication

Around 2 g of cell pellets containing *B. subtilis* PCAS/NHis were resuspended in 25 ml of buffer 3A. Once fully resuspended, the cells were lysed with sonication using a sonicator fitted with a blunt tip for 15 min (30 s on / 30 s off, on ice) and then ultracentrifuged at 20,000 rpm for 25 min.

2.4.3.2 Nickel column

The cell free extract was then filtered and loaded onto a pre-packed 1 mL His-trap (GE Healthcare) column pre-equilibrated with buffer 3A. Then PCAS/NHis was

eluted from the column with an increasing imidazole gradient from 10 mM (buffer 3A) to 1 M (buffer 3B) over 30 column volumes (30 ml) and monitored by UV at 280 nm. Elution fractions of the UV 280 nm peak were pooled together (5mL) and dialysed two times against 1 L of buffer 3C for at least 2 h.

2.4.3.3 Size exclusion chromatography (SEC)

The filtered enzyme was then loaded onto a gel filtration Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with buffer 3C. Protein was eluted from the column with buffer 3C, and the elution fractions containing a high purity of PCAS/NHis were pooled together.

2.4.4 Purification of *B. subtilis* PCAS/UT

2.4.4.1 Cell lysis by sonication

Cell pellets containing *B. subtilis* PCAS/UT were resuspended in buffer 4A and then lysed with sonication. The crude lysate was cleared by centrifugation at 20,000rpm for 25min.

2.4.4.2 Anion exchange chromatography

After 0.22 μm filtering, the cell free extract was loaded onto a HiLoad^{TM} 16/10 Q Sepharose^{TM} High Performance column (20 mL, GE Healthcare) pre-equilibrated with buffer 4B. The protein was eluted by a gradient from 0% to 50% of buffer 4C/buffer 4B (100 mM to 500 mM NaCl) over 15 column volumes.

2.4.4.3 Phenyl HP column

Next, the eluted PCAS/UT was dialysed against 1L of buffer 4D for 1h and loaded onto a HiTrap Phenyl HP hydrophobic interaction column (1 mL, GE Healthcare) pre-equilibrated with buffer 4D. The protein was eluted by a decreasing ammonium sulphate gradient from from 1 M (buffer 4D) to 0 (buffer 4B) over 30 column volumes.

2.4.4.4 Size exclusion chromatography (SEC)

The fractions containing PCAS/UT were pooled, concentrated to 5 mL and then purified by a 16/60 superdex 200 column (GE Healthcare) with buffer 4E.

2.4.5 Purification of *E. coli* AONS/CHis

2.4.5.1 Cell lysis by sonication

Cell pellets containing *E. coli* AONS/CHis were resuspended in buffer 5A, lysed with sonication using a sonicator and then ultracentrifuged at 20,000 rpm for 25 min.

2.4.5.2 Nickel resin

The cell free extract was incubated with 2 mL of equilibrated Ni-NTA resin for 1 hour at 4 ^{o}C. Then the protein was washed with buffer 5B (10 mM imidazole) and eluted with buffer 5C (300 mM imidazole). Elution fractions were pooled together and dialysed two times against 1 L of buffer 5D for at least 2 h.

2.4.5.3 Size exclusion chromatography (SEC)

The filtered protein was then loaded onto a gel filtration Superdex 200 16/60 column (GE Healthcare) and eluted from the column with buffer 5D. The elution fractions containing a high purity of AONS/CHis were pooled together.

2.4.6 Purification of *C. amycolatum* BioWF

2.4.6.1 Cell lysis by sonication

Cell pellets containing *C. amycolatum* BioWF were resuspended in buffer 6A, lysed with sonication and then ultracentrifuged at 20,000 rpm for 25 min.

2.4.6.2 Nickel resin

The cell free extract was incubated with 2 mL of equilibrated Ni-NTA resin for 1 hour at 4 ^{o}C. Then the protein was eluted with 10 mM, 50 mM, 150 mM, 200 mM, 300 mM imidazole (buffer 6B-6G). Fractions containing BioWF were pooled and dialysed two times against 1 L of buffer 6H for at least 2 h.

2.4.6.3 Size exclusion chromatography (SEC)

The filtered protein was then loaded onto a gel filtration Superdex 200 16/60 column (GE Healthcare) and eluted from the column with buffer 6H.

2.5 Protein analysis

2.5.1 SDS-PAGE gel

Protein samples were denatured by adding SDS sample buffer (5x) and boiling for 10 min at 100 ^{o}C. 10 µl of the protein samples were loaded onto the SDS-PAGE gel along with low molecular weight marker (LMW marker, GE Health Care) and then run in TGS running buffer (x1) at a constant voltage of 200 volts for 50 min. After this, the gel was stained with coomassie blue at 37 ^{o}C with gently shaking for about one hour followed by destaining with distilled water using microwave for 15 min, or with InstantBlue ultrafast protein stain (Sigma-Aldrich) at room temperature for 15 min without destain step. The SDS-PAGE gel composed of two layers - running and stacking gels and was parepared as shown in **Table 2.9**.

Components	15% running gel	4% stacking gel
dH2O	5.7 mL	2.9 mL
40% acrylamide	6 mL	0.75 mL
1.5 M Tris-HCl, pH 8.8	4 mL	-
0.5 M Tris-HCl, pH 6.8	-	1.25 mL
SDS (10% w/v)	150 µL	50 µL
APS (50 mg/ml)	300 µL	100 µL
TEMED	20 µL	5 µL

Table 2.9 Preparation of acrylamide gels for SDS-PAGE analysis

2.5.2 Determination of protein concentrations using Absorbance, A_{280}

Protein concentrations were determined using the UV-vis absorption at 280 nm. This was done on a Varian's Cary 50 UV-vis spectrophotometer. The Beer-Lambert's law

(**Equation 2.1**) was used for calculation and the extinction coefficient (ε) was determined using the Vector NTI software.

$$A=\varepsilon cl$$

Equation 2.1 Beer-Lambert's law. *A*= absorbance at 280 nm, *ε*= extinction coefficient ($M^{-1}cm^{-1}$), *c*= protein concentration (M), *l*= the path length (1 cm).

2.6 Protein characterization

2.6.1 Monitoring of acyl-CoA thioester bond formation

Measurements were made by a continuous UV-vis spectrophotometric assay which is described by Ploux in 1992 [40]. This assay is based on monitoring the formation of the acyl-CoA thioester bond at 230 nm. The reaction volume was 200 µL and contained 100 mM TrisHCl pH 8.5, 200 mM NaCl, 10 mM $MgCl_2$, 0.4 mM CoASH, 0.4 mM ATP and 0.2 µM PCAS enzyme were incubated at 30 ℃. The reactions were started by the addition of 1.5 mM pimelic acid and the 230 nm absorbance was recorded for 30m min on a BioTek Synergy HT plate reader with Costar 96-well UV-transparent plates. The rate over the first 10 min was analysed using the Michaelis-Menten model and a nonlinear regression fit on GraphPad to give values of K_M and k_{cat}.

2.6.2 Pyrophosphate production assay

The PCAS activity was also determined by monitoring the release of pyrophosphate (PP_i) via a two-enzyme coupled assay. The final volume of the reactions was 200 µL and contained 50 mM TrisHCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM

triscarboxyethylphosphine (TCEP), 1 mM CoASH, 1 mM ATP, 0.5 mM 7-methyl-6-thioguanosine (MESG, Berry & Associates), 0.03 U of inorganic pyrophosphatase (PPase from baker's yeast, Sigma), 1 U of bacterial purine nucleoside phosphorylase (PNP, Sigma) and 0.1 μM PCAS enzyme (or up to 3μM for several mutants) were pre-incubated at 30 ℃ for 15 min. The reaction was initiated by adding 1.5 mM pimelic acid (or substrates of varying carbon chain length of dicarboxylic acids or fatty acids). The increase in absorbance at 360 nm resulting from the enzymatic conversion of MESG to 7-methyl-6-thioguanine was monitored over 30 min on a BioTek Synergy HT plate reader with Costar 96-well UV-transparent plates. The data from the first 10 minutes was analysed using the Michaelis-Menten model and a nonlinear regression fit on GraphPad gave values of K_M and k_{cat}.

2.6.3 High performance liquid chromatography (HPLC) assay

The formation of acyl-CoA esters was detected by an HPLC method. The reactions contained 25 mM TrisHCl pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.2 mM TCEP, 1 mM CoASH, 1 mM ATP, 5 μM PCAS and 1.5 mM pimelic acid (or suberic acid (DC8), heptanoic acid (FA7) for mutants Y211F) were incubated at 30 ℃ in a final volume of 700 μL. The reactions were terminated at varying time points by incubation with acetonitrile (volume ratio 1:2) on ice for 5 min to precipitate the protein. The samples were clarified by centrifugation, diluted with distilled water (5 fold diluted finally) and filtered through a 0.45 μm filter.

2.6.3.1 HPLC assay by using C8 reverse phase column

10 µL of the sample was then injected onto Luna 5u C8(2) RP-HPLC column (150x4.60mm, Phenomenex) and separated by a gradient from 5% to 55% acetonitrile (0.1% TFA, v/v)/water (0.1% TFA, v/v) for 25min at 260nm. Where necessary the peak containing the acyl-CoA product was collected, freeze dried and subject to electrospray ionisation mass spectrometry (ESI-MS).

2.6.3.2 HPLC assay by using C18 reverse phase column

10 µL of the sample was then injected onto Luna 5u C18(2) RP-HPLC column (100A, 250x4.60mm, Phenomenex), and eluted with water (0.1% TFA (trifluoroacetic acid), v/v) for 5 min followed by a 15 min gradient from 5% to 55% acetonitrile (0.1% TFA, v/v)/water (0.1% TFA, v/v) which was maintained for 2 min at 260nm.

2.7 *In vivo* complementation assay for *B. subtilis* PCAS

2.7.1 Lysogenization of *E. coli* ΔBioH mutation strain

E. coli ΔBioH mutation cells ordered from Coli Genetic Stock Centre (CGSC# 12019) were made λ (DE3) lysogens using a commercial kit from Merck.

2.7.2 Making electrocompetent Cells

E. coli ΔBioH overnight culture was back diluted to OD_{600} of 0.1 in 100 mL of LB broth with kanamycin (50 µg/mL) and shook at 37 ℃ until the OD_{600} reached around 0.6. After chilling on ice for one hour, the cells were harvested by centrifugation and then

washed with pre-chilled filter sterilized distilled water for three times. The final volume (about 2 mL) was stored in 20 % glycerol as 100 µL aliquots at -80 ℃.

2.7.3 Preparation of M9 Minimal media

M9 Minimal media was prepared carefully as shown in **Table 2.10**. Appropriate antibiotic of the required concentration was added on cooling.

Stock	Volume
M9 salts stock (5x) contains 6.78 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl and 1 g NH_4Cl were dissolved in deionized water to a final volume of 200mL and sterilized by autoclaving.	40mL
1M $MgSO_4$ sterilze by autoclaving	0.4mL
0.5M $CaCl_2$ sterilze by autoclaving	0.04mL
3g of agar in 150 mL of deionized water sterilze by autoclaving	150mL
20% w/v glucose sterilze by filter	10mL

Table 2.10 Components for preparing M9 Minimal media

2.7.4 *In vivo* complementation assay

Plasmid of *Bacillus subtilis* pimeloyl-CoA synthetase (PCAS, *BioW*/pET22b) was used to transform *E. coli* ΔBioH (DE3) electrocompetent cells and then spread on M9 minimal media agar plates supplemented with IPTG (0.1mM), pimelic acid (0.1mM), ampicillin (100 µg/ml) and kanamycin (50 µg/ml) or plates supplemented with biotin (1 µM), ampicillin (100 µg/ml) and kanamycin (50 µg/ml). The same was done with the empty vector (pET22b) as a negative control. And all the plates were incubated at 37 ^{o}C overnight.

2.8 Native mass spectrometry

The PCAS sample was desalted twice over P6 Micro Bio-Spin columns (Bio-Rad) into 200 mM ammonium bicarbonate and further diluted to approximately 20 mM (10 mM dimer). Native mass spectra were acquired before and after addition of 1.0 mM MgATP, 1.5 mM pimelic acid and 0.5 mM CoASH. Samples were infused by nano-electrospray ionisation (nano-ESI) using a TriVersa NanoMate (Advion) set at 1.6 kV and 0.8 psi pressure. Electrospray was coupled to a Synapt G2 Q-TOF (Waters) with the source set at 80 ^{o}C, backing pressure increased to 5 mbar, and sampling cone voltage of 150 V. Spectra were averaged over 4 min of acquisition data and presented without further data processing. Peak *m/z* annotations were extracted from smoothed and centroided using Waters MassLynx V4 software.

2.9 Site-directed and site-saturation mutagenesis of PCAS

The *B. subtilis* PCAS mutants Y199F, Y211F, R213A, R11A and R13A were constructed according to the overlapping primer site-directed mutagenesis method [70] with the primers pairs shown in **Table 2.11**. SDM reactions contains template DNA (100 ng), forward primer (1.5 μL of 10 μM stock), reverse primer (1.5 μL of 10 μM stock), dNTP (1 μL of 10 mM stock), Pfu 10x buffer (5 μl), turbo cloned Pfu polymerase (1 μl) and filter sterilized water to obtain a final volume of 50 μL were prepared and placed in the thermo cycles (**Table 2.12**). After this, DpnI (1 μl) and CutSmart buffer (6 μl) were added to the reaction mixture and incubated for 4.5 h at 37 ^{o}C to digest the template DNA. Then 2 μL of the newly synthesized DNA with point mutation was used to transform C2987 competent cells (50 μL) and grow on LB plates

overnight at 37 ^{o}C. Next day, colonies were picked, send for sequencing and then based on the results PCAS mutants Y199F, Y211F, R213A, R11A and R13A were taken forward for analysis. In the meantime, the PCAS mutants R227E and R227K were generated by saturation mutagenesis with the primers shown in **Table 2.11** using the same method as above.

Primers	Sequence (5'-3')
Y199F forward	ATGCTGGTCGGACGATCCGGAT**TTC**ATAACAGGCTATGT TGCGGGTAAGAAAATGGGCTAT
Y199F reverse	CTTACCCGCAACATAGCCTGTTAT**GAA**ATCCGGATCGTCC GACCAGCATAATTCTGCAACG
Y211F forward	ATGTTGCGGGTAAGAAAATGGGC**TTC**CAGCGTATTACAG CAATGAAAGAATACGGGACTG
Y211F reverse	CTTTCATTGCTGTAATACGCTG**GAA**GCCCATTTTCTTACC CGCAACATAGCCTGTTATG
R213A forward	GCGGGTAAGAAAATGGGCTATCAG**GCA**ATTACAGCAATG AAAGAATACGGGACTGAAG
R213A reverse	CCCGTATTCTTTCATTGCTGTAAT**TGC**CTGATAGCCCATT TTCTTACCCGCAACATAG
R11A forward	GAAGAAACTTTTTATAGTGTC**GCT**ATGAGGGCTTCAATG AATGGATCTCATGAAGACGGCGG
R11A reverse	GAGATCCATTCATTGAAGCCCTCAT**AGC**GACACTATAAA AAGTTTCTTCTTCCATGG
R13A forward	GAAGAAACTTTTTATAGTGTCAGAATG**GCT**GCTTCAATG AATGGATCTCATGAAGACGGCGG
R13A reverse	GAGATCCATTCATTGAAGC**AGC**CATTCTGACACTATAAA AAGTTTCTTCTTCCATGG
227X forward	ATACGGGACTGAAGAGGGCTGC**RMA**GTCTTTTTTATTGA TGGCTCCAATGATGTAAACA
R227X reverse	GGAGCCATCAATAAAAAAGAC**TKY**GCAGCCCTCTTCAGT CCCGTATTCTTTCATTGCTG

Table 2.11 Site-directed and site-saturation mutagenesis primers, where R = A/G, M = A/C, K = G/T, Y = C/T. Bases in bold are the sites of mutagenesis.

Steps	Cycle	Temperature (^{o}C)	Time
Initial Denaturing	1	95	2 min
Denaturing	30	95	30 s
Annealing		55	30 s
Extension		72	6.5 min
Final Extension	1	72	10 min
Hold at 4 ^{o}C			

Table 2.12 Polymerase Chain Reaction (PCR) protocol

All PCAS point mutant proteins were overexpressed with 0.5 mM IPTG at 20 ℃ overnight. The purification protocol is the same as the His_6 tagged wild type protein with TEV protease removal of the His_6 tag. The mass of each PCAS mutant was confirmed by LC electrospray ionization mass spectrometry (ESI-MS).

Chapter III. Characterisation of Bacillus subtilis pimeloyl-CoA synthetase (PCAS)

3.1 Cloning of *B. subtilis* pimeloyl-CoA synthetase (PCAS)

The full-length, codon-optimised *BioW* gene (*Bacillus subtilis (strain 168)*, UNIPROT code: P53559) was cloned into three expression plasmids to give three different recombinant forms of PCAS; one with a non-cleavable N-terminal His_6 tagged PCAS cloned in pET-28a plasmid (named as PCAS/NHis, **Section 3.1.1**, **Section 3.2.3**), a pEHISTEV plasmid clone which gives an N-terminal tobacco etch virus (TEV) protease-cleavable His_6 tag PCAS (named as PCAS/NHT, **Section 3.1.2**, **Section 3.2.1**) [69] and an un-tagged PCAS in pET-22b plasmid (named as PCAS/UT, **Section 3.1.3**, **Section 3.2.4**) . The basic information of these three versions of PCAS was shown in **Table 3.1**.

Name	Number of amino acids	Molecular mass*	His_6 tag	Vector
PCAS/NHis	278 aa	31662.5 Da	Uncleaverbale N-terminal His_6 tagged	pET-28a
PCAS/NTH (before TEV protease cleavage)	283 aa	32596.5 Da	Cleaverbale N-terminal His_6 tagged	pEHISTEV
PCAS/NTH (after TEV protease cleavage)	260 aa	29628.3 Da	Un-tagged	pEHISTEV
PCAS/UT	258 aa	29499.2 Da	Un-tagged	pET22b

Table 3.1 Information of *B. subtilis* pimeloyl-CoA synthetase (PCAS) in different expression systems studied in this project. * Molecular mass shown in this table was predicted by using the ExPASy-ProtParam tool.

3.1.1 Cloning of PCAS/NHis

The DNA and amino acid sequence of wild type *Bacillus subtilis* PCAS is shown in **Appendix 5.1**. Inherent BamHI restriction site GGA TCC (Gly Ser) and NdeI restriction site GCATATGAA (the restriction site is underlined, Ala Tyr Glu) were mutated out of the wild type sequence as GGC TCC (still translated as Gly Ser) and GCGTATGAA (the restriction site is underlined, still translated as Ala Tyr Glu), at the same time, NdeI and BamHI restriction sites were attached at the two ends (**Appendix 5.2**). This codon-optimised *BioW* gene was purchased from GenScript in a pUC57 vector and then cloned into pET-28a vector via NdeI and BamHI (**Fig. 3.1**). There is a stop codon before the BamHI site means the protein will be only N-terminal His_6 tagged. This cloning work was done by MChem project student Andrew Piper.

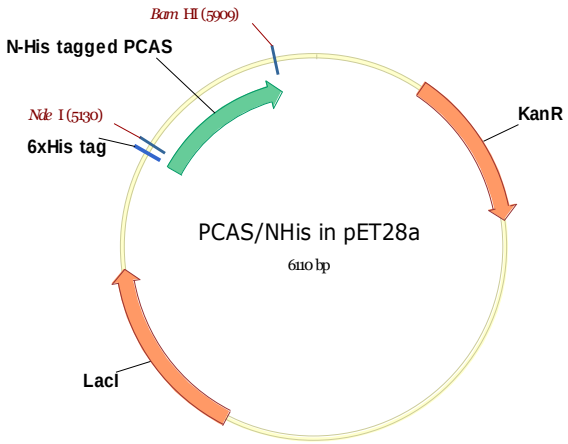


Fig. 3.1 Plasmid map of *Bacillus subtilis* PCAS/NHis in pET28a vector

3.1.2 Cloning of PCAS/NTH and PCAS/UT

The *BioW* gene was amplified from the PCAS/NHis/pET28a plasmid using a forward primer containing the Nco1 restriction site and cloned into the pEHISTEV expression

vector via NcoI and BamHI sites. This gives a translated PCAS protein with an N-terminal tobacco etch virus (TEV) protease-cleavable His_6 tag (**Appendix 5.3, Fig. 3.2 the left panel**). Additionally, *BioW* gene was also cloned into a pET22b vector from PCAS/NHis/pET28a plasmid via the Nde1 and Xho1 restriction sites to obtain an un-tagged protein i.e. PCAS/UT (**Appendix 5.4**, **Fig.3.2 the right panel**).

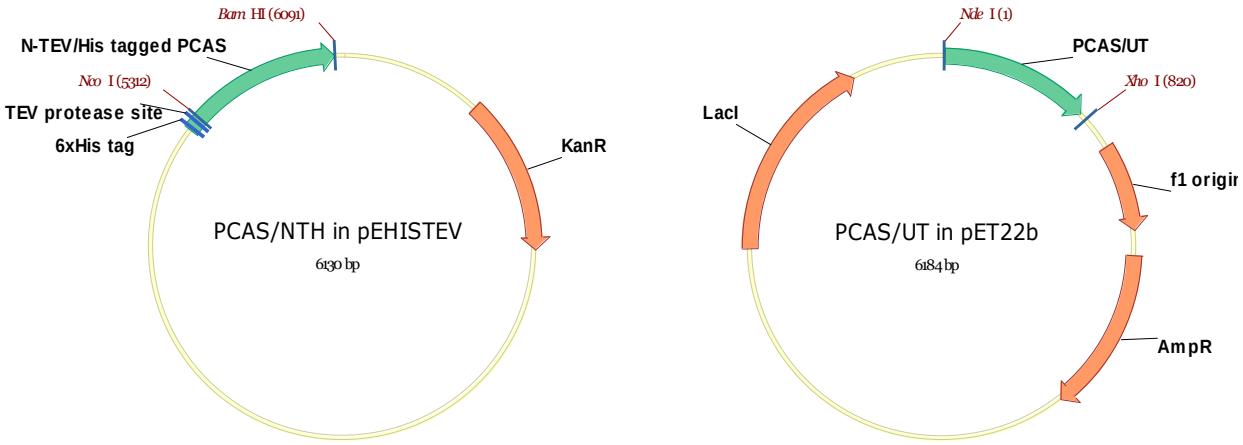


Fig. 3.2 Plasmid maps of **left**) *Bacillus subtilis* PCAS/NTH in pEHISTEV vector, **right**) *Bacillus subtilis* PCAS/UT in pET22b vector

3.2 Protein Expression and Isolation

3.2.1 Isolation of PCAS/NTH

The N-terminal TEV protease cleavable His_6 tagged PCAS/HTH form was expressed in BL21 (DE3) *E. coli* strain after induction with 0.1 mM IPTG and incubating at 30 ^{o}C for 4 hours. Using the pEHISTEV expression system, allows for removal of the His_6 tag after the initial nickel affinity chromatography isolation by TEV protease cleavage, and results in the production of a native, “tag-free” PCAS. The purification

steps including an initial HisTrap™ FF 5mL Ni:Sepharose column firstly isolated the N- terminal His_6/TEV-tagged PCAS from the cell free extract (CFE), then the His_6 tag was cleaved by incubating with TEV protease (enzyme: protease ratio=15:1) overnight at 4 ^{o}C, followed by removal of the His_6 tag and TEV protease from the sample through a second HisTrap™ FF 5mL Ni:Sepharose column. Further purification by Superdex 200 size-exclusion chromatography gave an asymmetric peak at 73.4 mL corresponding to the theoretical mass of 57,276.6 Da (**Fig. 3.3**), the mass of the purified PCAS/NTH was calculated using a calibration curve of known molecular weights as shown in **Appendix 5.8**). Compared to subunit *Mr* of around 30 kDa determined by SDS-PAGE gel (**Fig. 3.4**), the active PCAS/NTH was found to be a homodimer. And the final yield is ~21.8 mg per 1L of culture.

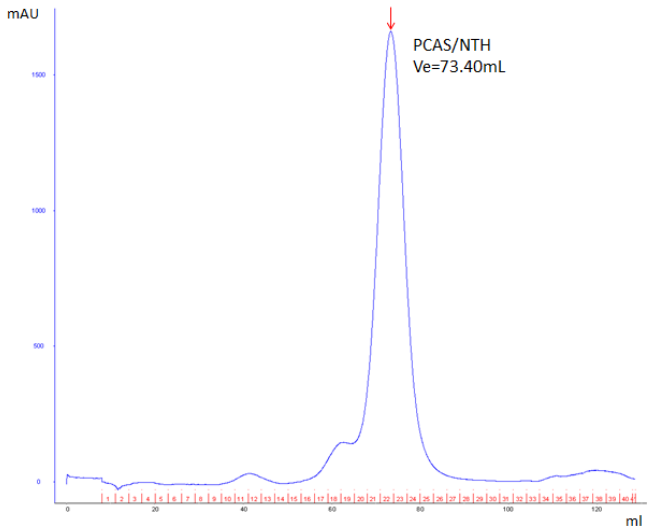


Fig. 3.3 AKTA trace from S200 gel filtration column (A_{280nm}) of PCAS/NTH which gives a peak at 73.4 mL. Based on the calibration curve show in **Appendix 5.8**, it corresponds to the theoretical mass of ~57 kDa, this means PCAS/NTH is a homodimer.

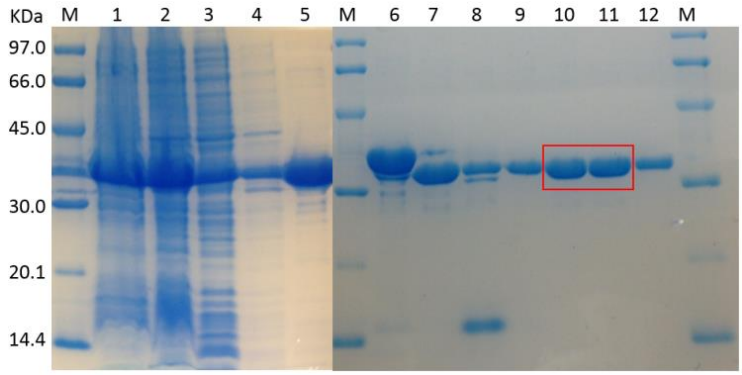


Fig. 3.4 SDS-PAGE gel showing the purity of PCAS/NTH. M: LMW Marker, Lane 1: Cell pellet, Lane 2: Cell lysis, Lane 3: Ni resin flow through, Lane 4: Ni resin wash, Lane 5: Ni resin elution, Lane 6: Before TEV protease cleavage, Lane 7: Second Ni resin flow through, Lane 8: Second Ni resin elution, Lane 9-12: Superdex 200 fractions 21-24. It showed a band at ~30 kDa which corresponding to the PCAS/NTH monomer.

The predicted mass of PCAS/NTH calculated by ExPASy-ProtParam tool is 29628.3 Da. As shown in **Fig. 3.5**, analysis by LC-ESI-MS showed an ion envelope ranging from 38+ to 21+ charge states. We picked the most abundant peak in the mass spectrum at 1059.17 Da with a charge state of 28+, so the deconvoluted mass is [1059.17 x 28] - 28 (H+) = 29628.8 Da ($C_{1304}H_{2025}N_{367}O_{400}S_{12}$) which is in very good agreement with the predicted mass of 29628.3 Da.

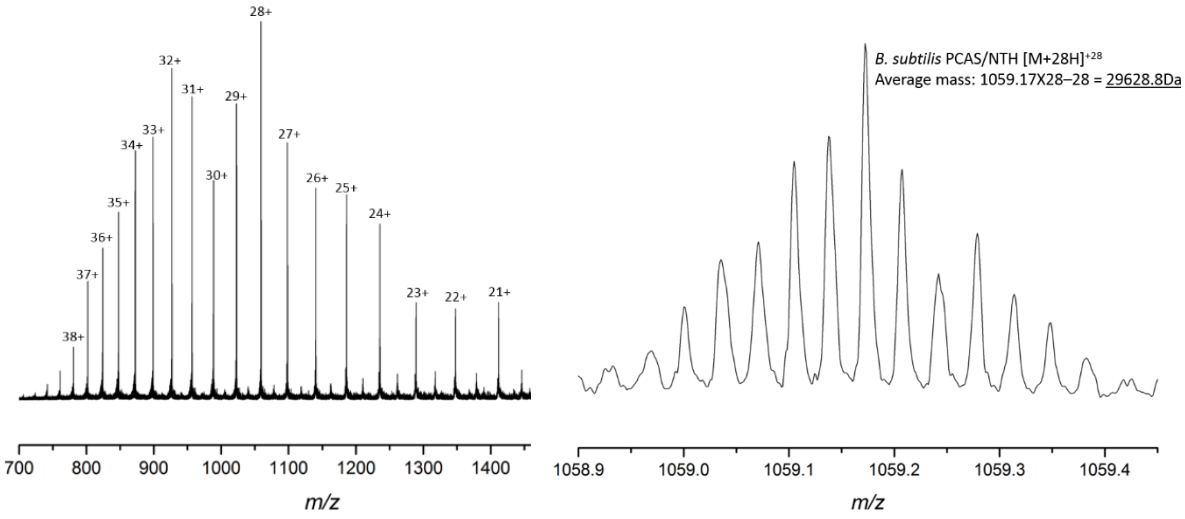


Fig. 3.5 The LC-ESI-MS of *B. subtilis* PCAS/NTH. The **left** panel is the ion envelope of the + 38 to +21 charge states and the **right** panel shows the observed isotope distribution ($C_{1304}H_{2025}N_{367}O_{400}S_{12}$) of the +28 charge state.

3.2.2 TEV protease

His_6 tagged TEV protease was expressed in Rosetta (DE3) *E. coli* strain after induction with 0.4 mM IPTG and incubating for >16 hours at 20 ^oC and then isolated from cell free extract (CFE) by nickel resin eluting with increasing imidazole gradient. The elution fractions were combined based on the SDS/PAGE analysis (**Fig. 3.6**) and this method typically yielded ~15.9 mg TEV protease per 1L LB media.

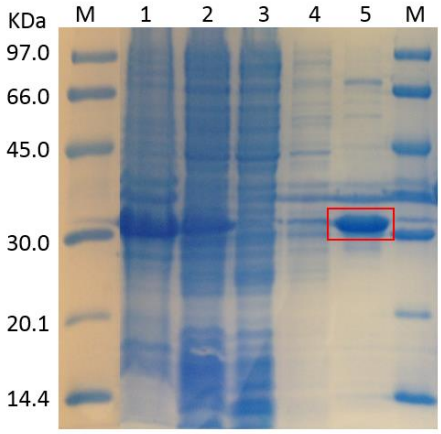


Fig. 3.6 SDS-PAGE gel showing the purity of TEV protease. M: LMW Marker, Lane 1: Cell pellet, Lane 2: Cell lysis, Lane 3: Ni resin flow through, Lane 4: Ni resin wash, Lane 5: Ni resin elution

3.2.3 Isolation of PCAS/NHis

Full-length N-terminal His_6 tagged PCAS, here named as PCAS/HNis, was expressed in BL21 (DE3) *E. coli* after induction with 0.1 mM IPTG and incubating at 30 ^{o}C for 4 hours, and isolated from cell free extract (CFE) by HisTrap™ FF 1mL Ni:Sepharose column eluting with increasing imidazole gradient (**Fig. 3.7 A)**. The elution fractions were then analyzed by SDS/PAGE and showed a band at ~32 kDa which corresponding to the PCAS/NHis monomer (**Fig. 3.8)**. Further purification by Superdex 200 size-exclusion chromatography gave a peak at 72.2 mL corresponding to the theoretical mass of 60819.3 Da (**Fig. 3.7 B**), the mass of the protein was calculated using a calibration curve of known molecular weights as shown in **Appendix 5.8**). Compared to subunit *Mr* of around 32 kDa determined by SDS-PAGE gel, the active PCAS was found to be a homodimer. These processes typically yielded ~18 mg L^{-1} of growth media.

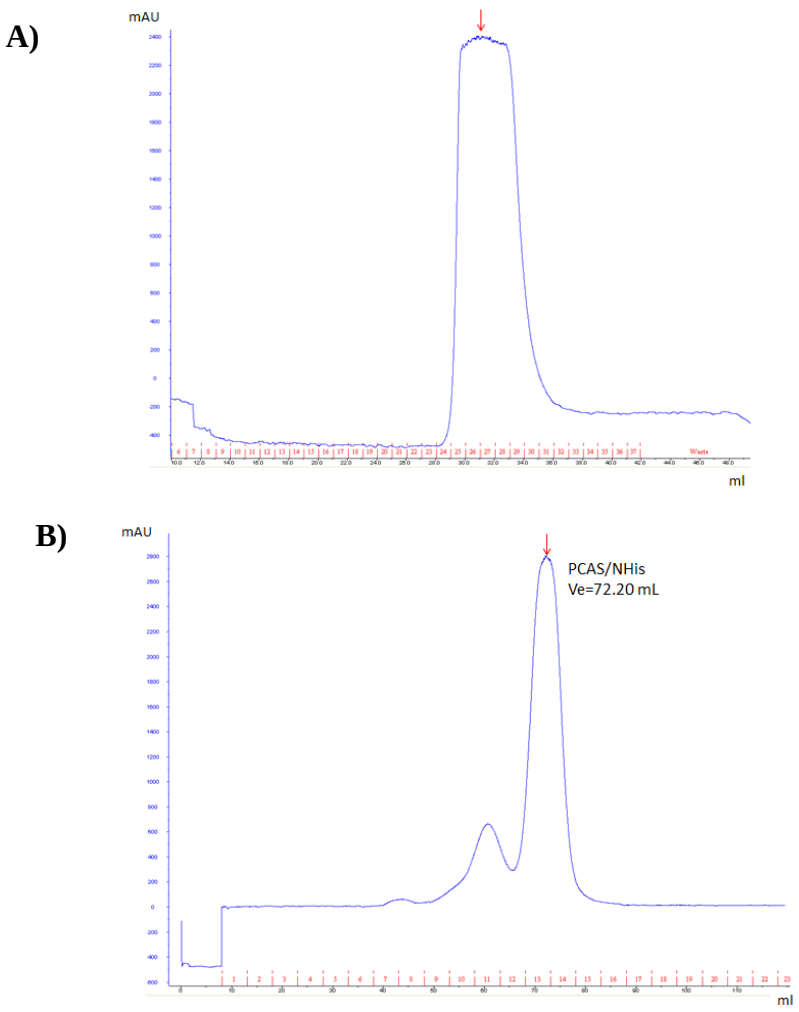


Fig. 3.7 AKTA traces of **A)** HisTrap FF chromatography **B)** superdex 200 column (A_{280nm}) of PCAS/NHis. Purification by S200 gave a peak at 72.2 mL corresponding to the theoretical mass of ~61 kDa, thus PCAS/NHis was found to be a homodimer.

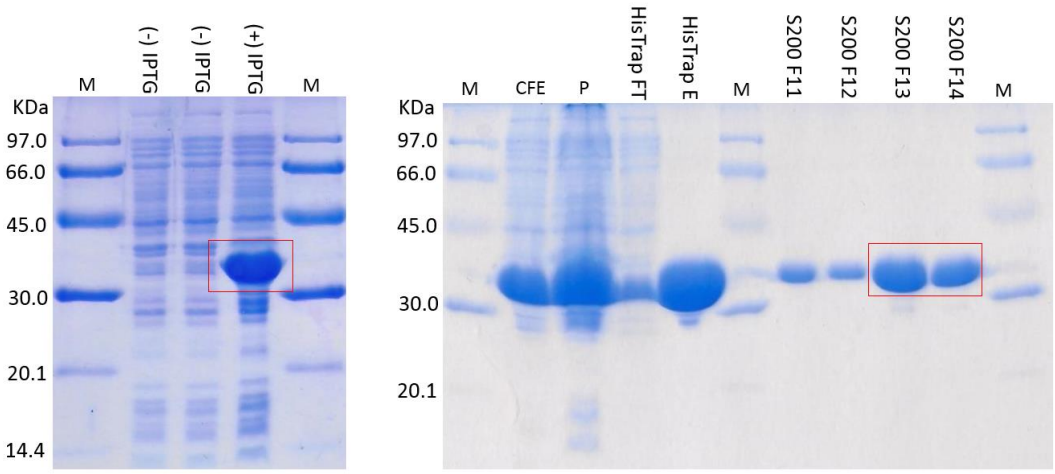


Fig. 3.8 SDS-PAGE gel showing the expression and purification steps of PCAS/NHis. They both showed a band at ~32 kDa which corresponding to the PCAS/NHis monomer.

The mass of PCAS/NHis calculated by ExPASy-ProtParam tool website is 31662.5 Da. Analysis of the PCAS/NHis form by LC-ESI-MS (**Fig. 3.9**) showed an ion envelope ranging from 33+ to 21+ charge states. We used the most abundant ion with 26+ charge state (1213.73 Da) to calculate the experimental mass. The deconvoluted mass is [1213.73 x 26] - 26 (H+) = 31531.0 Da ($C_{1389}H_{2149}N_{403}O_{422}S_{13}$). Considering the loss of N-terminal methionine (31531.0+131=31662.0 Da), the experimental mass matches the predicted mass.

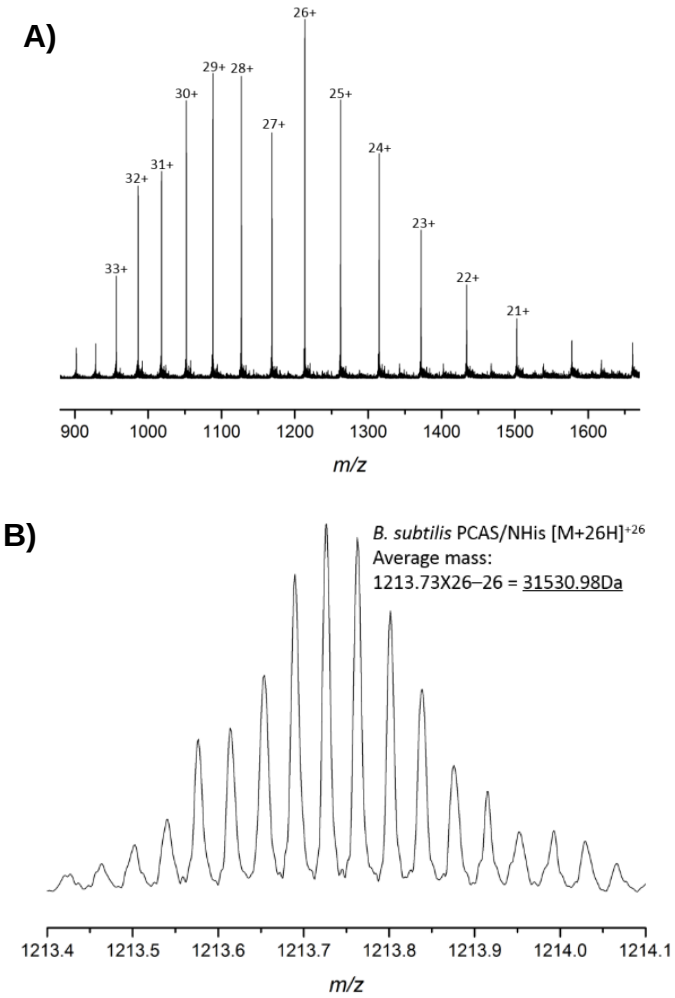


Fig. 3.9 The LC-ESI- MS of PCAS/NHis. **A)** The ion envelope of the + 33 to +21 charge states. **B)** The observed isotope distribution of the +26 charge state ($C_{1389}H_{2149}N_{403}O_{422}S_{13}$).

3.2.4 Isolation of PCAS/UT

Un-tagged PCAS (PCAS/UT) was also isolated in this study involved Q Sepharose ion exchange column, Phenyl column and S200 gel filtration chromatography (**Fig. 3.10**). For S200, elution at 74.1 mL corresponding to the native *Mr* of 53,641.2 Da which was calculated using a calibration curve as shown in **Appendix 5.8**, compared

to subunit *Mr* determined by SDS-PAGE gel (**Fig. 3.11**), PCAS/UT was found to be a homodimer as well. The yield is ~ 2.0 mg per 1L growth media.

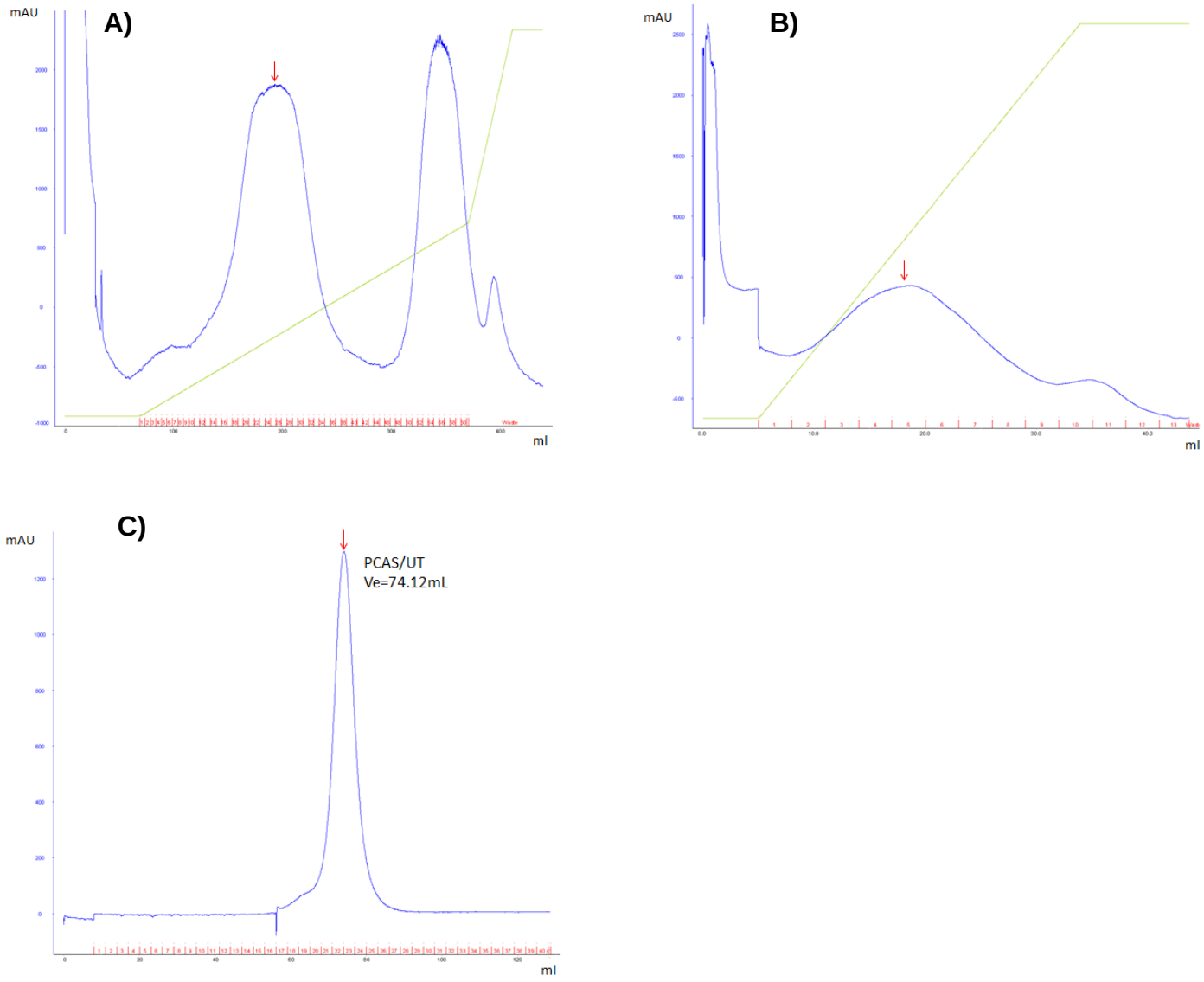


Fig. 3.10 AKTA traces of **A)** Q Sepharose ion exchange column **B)** Phenyl column **C)** superdex 200 column (A_{280nm}) of PCAS/NHis. Based on the calibration curve show in **Appendix 5.8**, V_e of 74.1 mL corresponds to the theoretical mass of ~54 kDa, thus PCAS/NHis is a homodimer.

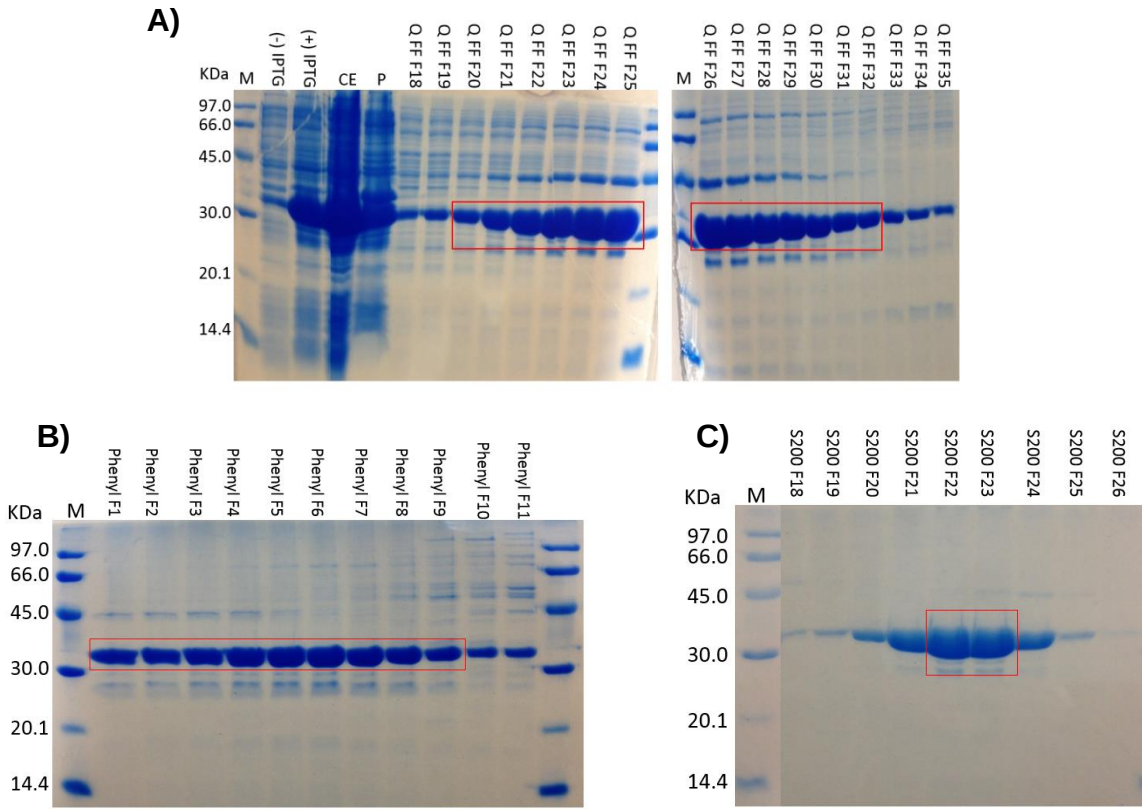


Fig. 3.11 SDS-PAGE gel monitoring the purification steps of PCAS/UT. **A**) Cell lysis and Q Sepharose ion exchange column, **B**) Phenyl column, **C**) S200 gel filtration chromatography. They all showed a band at ~32 kDa which corresponding to the PCAS/UT monomer.

ExPASy-ProtParam tool predicts the mass of PCAS/UT is 29499.2 Da. LC-ESI-MS (**Fig. 3.12**) showed an ion envelope with a charge state distribution of 44+ to 21+. We selected the most abundant ion at 820.46 Da with a charge state of 36+ to work out the experimental mass [820.46 x 36] - 36 (H+) = 29500.6 Da ($C_{1299}H_{2018}N_{366}O_{397}S_{12}$) which is in good agreement with the predicted mass.

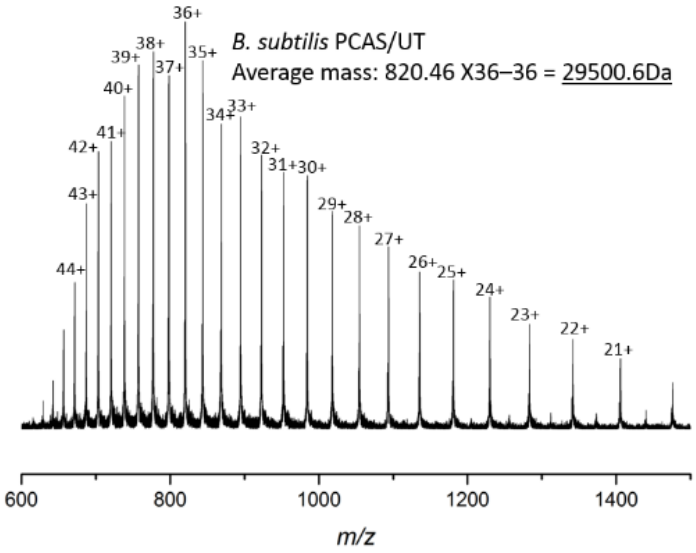


Fig. 3.12 Ion envelope of the + 44 to +21 charge states of the PCAS/UT.

3.3 Protein Analysis

Three *in vitro* assay methods along with one *in vivo* assay have been designed for the analysis of PCAS enzyme activity. Each one takes advantage of the chemistry of the reaction, either by monitoring the formation of the acyl-CoA thioester product either by UV-vis spectroscopy or mass spectrometry or by coupling the other product pyrophosphate (PPi) in a coupled assay. The *in vivo* assay requires the use of a mutant *E. coli* strain that requires pimeloyl-CoA for growth. The first one I optimised is ENZCHECK, which is a commercially-available two-enzyme (purine nucleoside phosphorylase pyrophosphorylase (PNP) and inorganic pyrophosphorylase (PPase)) coupled assay, which links pyrophosphate (PPi) formation in the first half of the reaction to the generation of a colorimetric product detectable at 360 nm [71, 72] (**Fig. 3.13**). Moreover, to validate the coupled assay, we also monitored acyl-CoA thioester

bond formation at 230 nm [40] and observed pimeloyl-CoA product formation by HPLC and mass spectrometry [43].

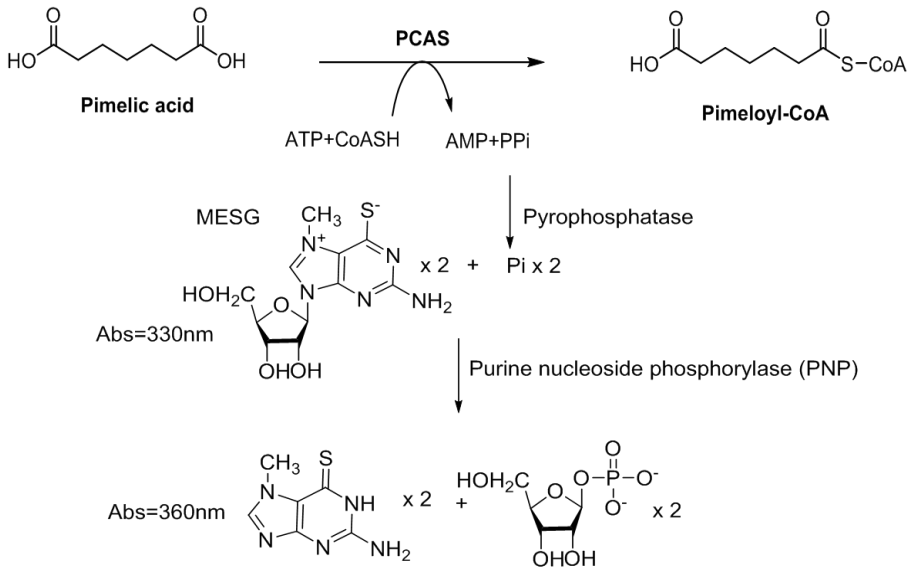


Fig. 3.13 Basis of the pyrophosphate (PPi) production assay. Pimelic acid is converted to pimeloyl-CoA by PCAS. The PPi released is hydrolysed to two molecules of phosphate (Pi) by pyrophosphatase (PPase). Purine nucleoside phosphorylase pyrophosphorylase (PNP) catalyses the reaction between the released Pi and 7-methyl-6-thioguanosine (MESG) to give the product 7-methyl-6-thioguanine (which absorbs at 360 nm, $\varepsilon 360 = 11{,}000\ M^{-1}cm^{-1}$) and ribose phosphate.

3.3.1 pH profile of PCAS activity

The optimal pH for PCAS/NTH enzyme was determined by using the A_{230nm} UV-vis assay under V_{max} conditions. In this assay, a range of pH values from pH 7.4-10.5 were tested by using 50mM His/HCl buffer pH adjusted to 7.4-8.7 and 50mM

glycine/NaOH buffer pH adjusted to 8.4-10.5. The pH profile showed a bell-shaped curve and PCAS/NTH was found to be active over the full testing pH range with the highest activity at pH 8.7 (**Fig. 3.14**).

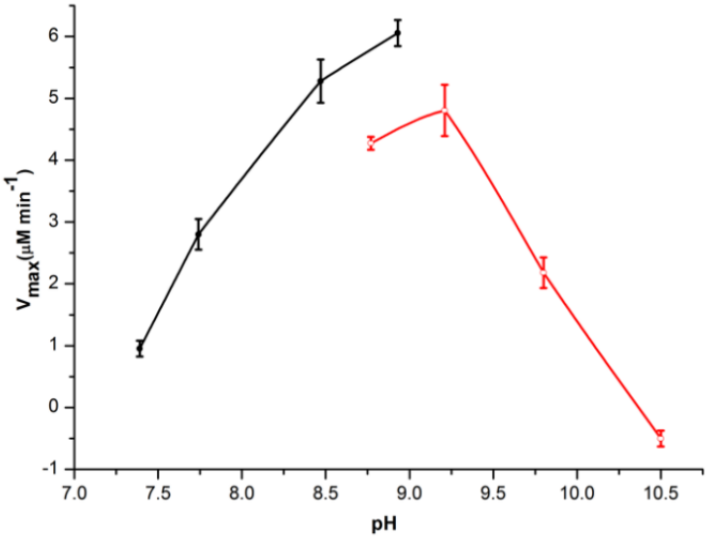


Fig 3.14 pH profile plot of PCAS/NTH using the A_{230nm} UV assay

Black: 50mM His/HCl buffer, Red: 50mM Glycine/NaOH buffer.

3.3.2 Kinetic study of PCAS

3.3.3.1 Kinetic study of PCAS using the A_{230nm} UV-vis assay

The faster A_{230nm} UV-vis assay, which monitors the acyl-CoA thioester bond formation, was first used to determine the kinetic values of four recombinant versions of PCAS including PCAS/NTH, uncleaved PCAS/NTH (the same as PCAS/NTH but without TEV protease cleavage step, i.e. His_6/TEV-tagged PCAS), PCAS/NHis and PCAS/UT. Typical saturation Michaelis-Menten kinetics were observed for all four versions of PCAS (**Fig. 3.15, Fig. 3.16, Fig. 3.17 and Fig. 3.18**). As shown in **Table 3.2**, 1) the K_M values of PCAS/NTH for pimelic acid, CoASH, ATP, $MgCl_2$ are

respectively 127.3±13.8 μM, 189.8±16.8 μM, 242.0±24.6 μM and 5.5±1.1 mM with an average k_{cat} of 0.73 s^{-1}; 2) the K_M values of uncleaved PCAS/NTH for pimelic acid, CoASH, ATP, $MgCl_2$ are respectively 110.4±12.6 μM, 162.8±4.6 μM, 214.3±20.2 μM and 5.8±1.7 mM with an average k_{cat} of 0.55 s^{-1}; the K_M of PCAS/NHis for pimelic acid, CoASH, ATP, $MgCl_2$ are respectively 181.8±26.8 μM, 151.8±23.0 μM, 191.3±12.9 μM, 3.0±0.3 mM with an average k_{cat} of 0.52 s^{-1}; 4) the K_M of PCAS/UT for pimelic acid, CoASH, ATP, $MgCl_2$ are respectively 76.0±7.9 μM, 262.9±106.5 μM, 153.0±13.0 μM and 3.0±0.5 mM with an average k_{cat} of 0.68 s^{-1}. All four versions of PCAS show similar k_{cat}.

Enzyme	PCAS/NTH (after TEV protease cleavage)				Enzyme	PCAS/NTH (before TEV protease cleavage)			
Substrate	Pimelic acid	CoASH	ATP	$MgCl_2$	Substrate	Pimelic acid	CoASH	ATP	$MgCl_2$
K_M (μM)	127.28±13.81	189.80±16.77	242.00±24.62	5.46±1.11 mM	K_M (μM)	110.38±12.61	162.75±4.60	214.33±20.19	5.75±1.71 mM
k_{cat} (s^{-1})	0.59	0.82	0.78	0.74	k_{cat} (s^{-1})	0.44	0.62	0.56	0.56
K_{cat}/K_M ($M^{-1}s^{-1}$)	4635.45	4320.34	3223.14	135.53	K_{cat}/K_M ($M^{-1}s^{-1}$)	3986.23	3809.52	2612.79	97.39
Average k_{cat} (s^{-1})	**0.73**				Average K_{cat} (s^{-1})	**0.55**			
Enzyme	**PCAS/Nhis**				Enzyme	**PCAS/UT**			
Substrate	Pimelic acid	CoASH	ATP	$MgCl_2$	Substrate	Pimelic acid	CoASH	ATP	$MgCl_2$
K_M (μM)	181.77±26.87	151.82±23.00	191.25±12.92	2.95±0.29 mM	K_M (μM)	76.00±7.87	262.85±106.46	153.00±13.04	3.01±0.47 mM
k_{cat} (s^{-1})	0.47	0.58	0.57	0.47	k_{cat} (s^{-1})	0.58	1.36	0.76	0.69
K_{cat}/K_M ($M^{-1}s^{-1}$)	2585.69	3820.31	2980.39	159.32	K_{cat}/K_M ($M^{-1}s^{-1}$)	7631.58	5174.05	4967.32	229.24
Average k_{cat} (s^{-1})	**0.52**				Average k_{cat} (s^{-1})	**0.68**			

Table 3.2 Kinetic parameters of **four versions of PCAS** using the A_{230nm} UV-vis assay

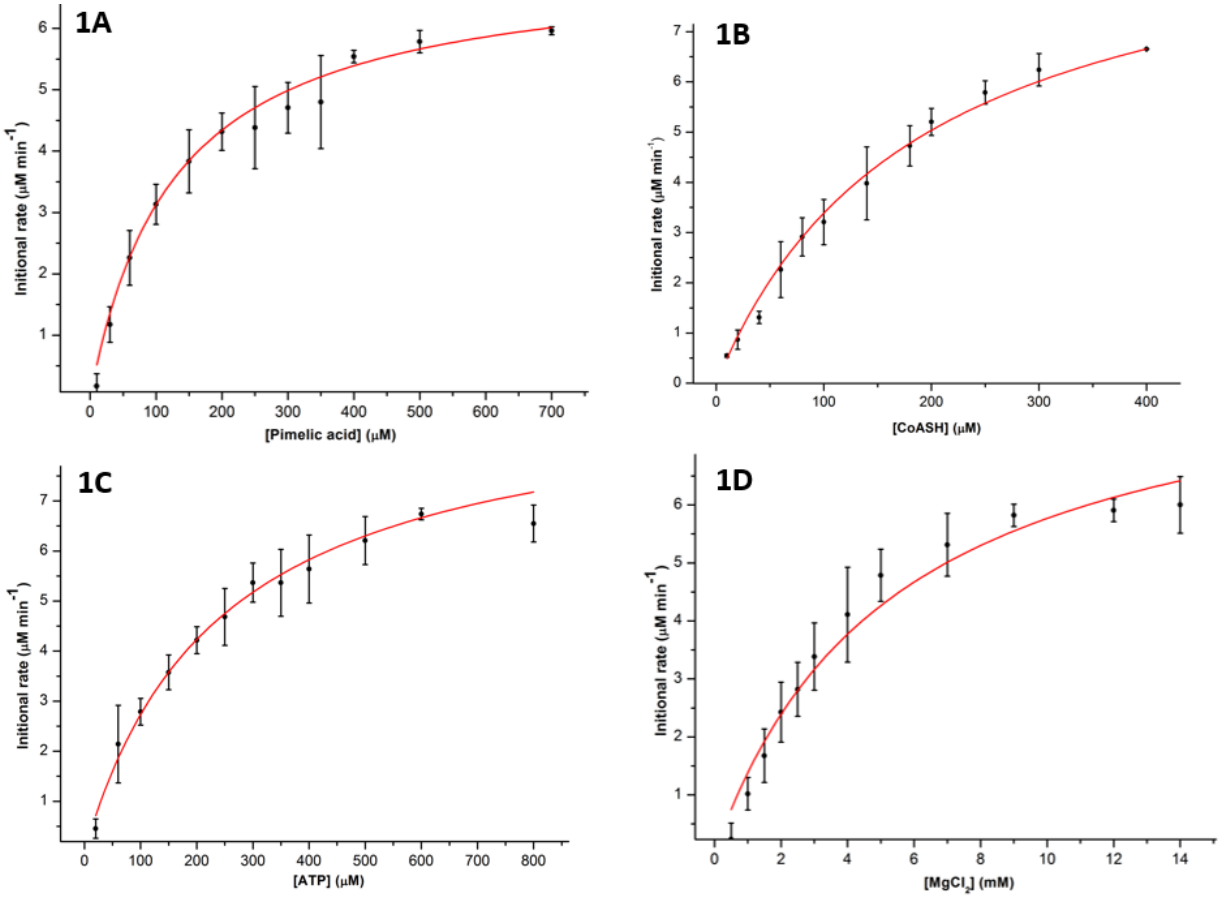


Fig. 3.15 shows Michaelis-Menten plot of the kinetic study of **PCAS/NTH** using the A_{230nm} UV-vis assay, it was fitted to a plot of the reaction velocity against the concentrations of different substrates which was drawn by Origin software. **1A)** K_M for pimelic acid is 127.3±13.8 μM, **1B)** K_M for CoASH is 189.8±16.8 μM, **1C)** K_M for ATP is 242.0±24.6 μM, **1D)** K_M for $MgCl_2$ is 5.5±1.1 mM. The average k_{cat} is 0.73 s^{-1}.

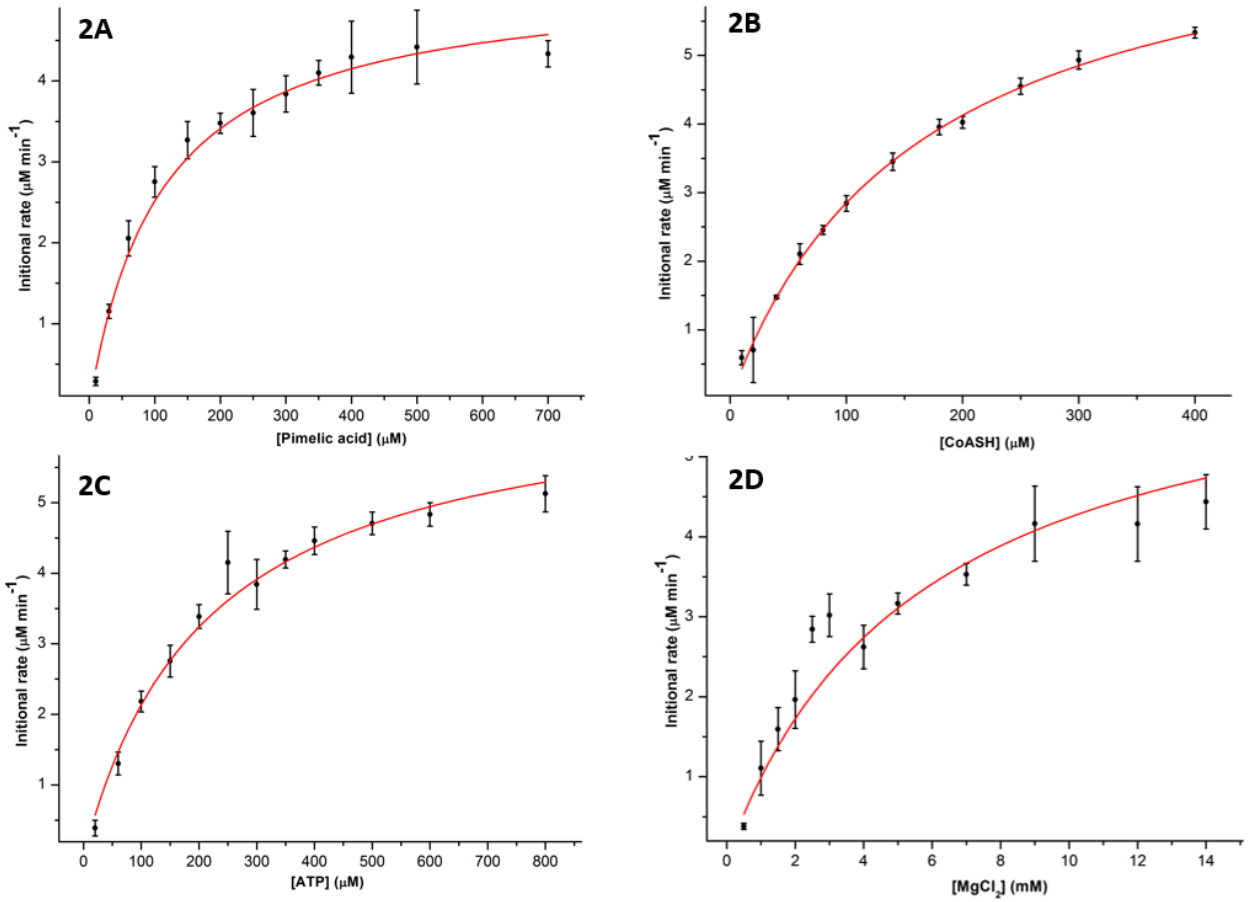


Fig. 3.16 Michaelis-Menten plot of the kinetic study of **uncleaved PCAS/NTH** using the A_{230nm} UV-vis assay which was drawn by Origin software. **2A)** K_M for pimelic acid is 110.4±12.6 μM, **2B)** K_M for CoASH is 162.8±4.6 μM, **2C)** K_M for ATP is 214.3±20.2 μM, **2D)** K_M for $MgCl_2$ is 5.8±1.7 mM. The average k_{cat} is 0.55 s^{-1}.

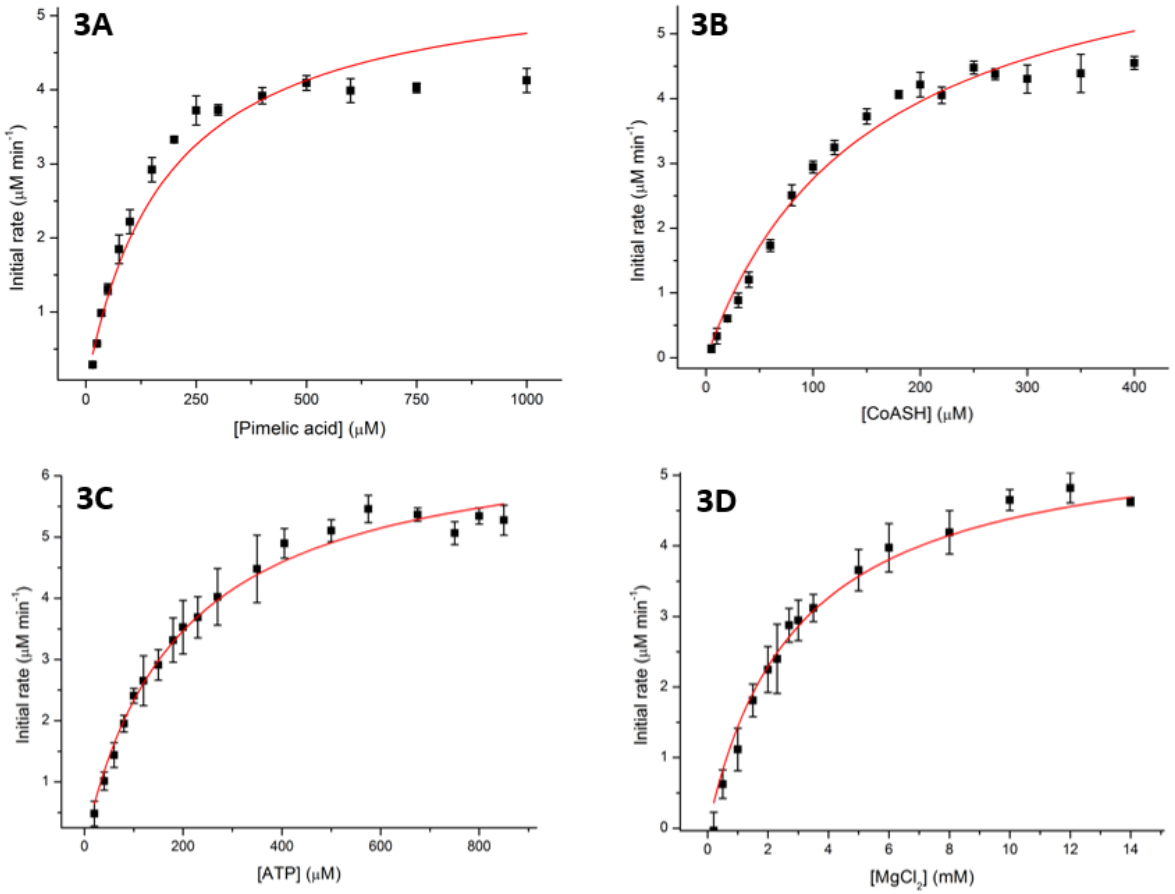


Fig. 3.17 Michaelis-Menten plot of the kinetic study of **PCAS/NHis** using the A_{230nm} UV-vis assay which was drawn by Origin software. **3A)** K_M for pimelic acid is 181.8±26.8 μM, **3B)** K_M for CoASH is 151.8±23.0 μM, **3C)** K_M for ATP is 191.3±12.9 μM, **3D)** K_M for $MgCl_2$ is 3.0±0.3 mM. The average k_{cat} is 0.52 s^{-1}.

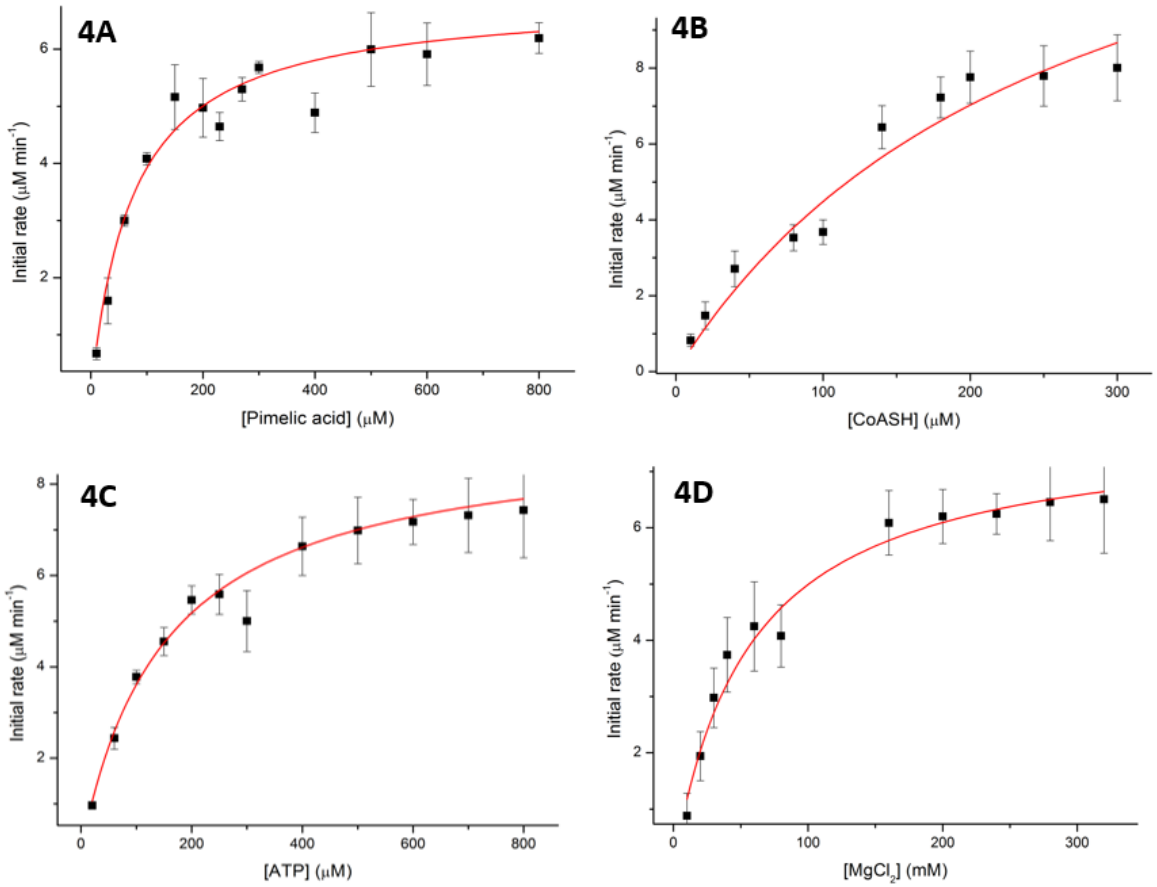


Fig. 3.18 Michaelis-Menten plot of the kinetic study of **PCAS/UT** using the A_{230nm} UV-vis assay which was drawn by Origin software. **4A)** K_M for pimelic acid is 76.0±7.9 μM, **4B)** K_M for CoASH is 262.9±106.5 μM, **4C)** K_M for ATP is 153.0±13.0 μM, **4D)** K_M for $MgCl_2$ is 3.0±0.5 mM. The average k_{cat} is 0.68 s^{-1}.

3.3.3.2 Kinetic study of PCAS using the pyrophosphate production assay

Additionally, kinetics of PCAS/NTH was also validated by the more accurate two enzyme coupled assay. In my hands, the PPi production assay proved to be more useful than the A_{230nm} UV-vis assay due to lower background absorbance when all of the substrates and enzyme were present. Michaelis-Menten kinetics were observed (**Fig. 3.19**), the K_M values for pimelic acid and ATP are respectively 70.5±6.8 μM and 299.6±3 μM with k_{cat} of 0.63 s^{-1} for pimelic acid and 0.44 s^{-1} for ATP (**Table 3.3**). These enzyme activity parameters I observed are substantially higher than previously reported by Manandhar and Cronan (~0.63 s^{-1} for pimelic acid compared with 8.4 x 10^{-5} s^{-1}) [43], but comparable with that observed by Ploux et al for the related PCAS from *B. sphaericus* determined with the A_{230nm} UV-vis assay (K_M of pimelic acid, ATP and CoASH are 145 μM, 170 μM and 33 μM with an average k_{cat} of 0.4 s^{-1}) [40].

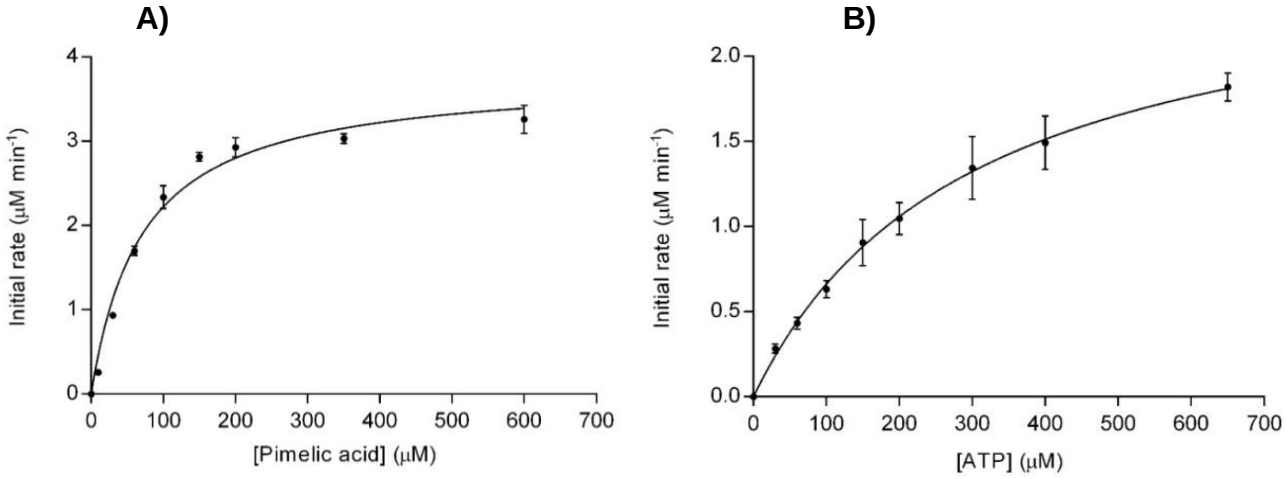


Fig. 3.19 Michaelis-Menten plot of the kinetic study of **PCAS/NTH** using the pyrophosphate production assay. It was fitted to a plot of the reaction velocity against the concentrations of pimelic acid or ATP which was drawn by GraphPad software. **A)** K_M for pimelic acid is 70.5±6.8 μM, **B)** K_M for ATP is 299.6±37.6 μM.

	Pyrophosphate production assay		A_{230nm} UV-vis assay	
Parameter	**Pimelic acid**	**ATP**	**Pimelic acid**	**ATP**
K_M (μM)	70.5 ± 6.8	299.6 ± 37.6	127.3 ± 13.8	242.0 ± 24.6
k_{cat} (s^{-1})	0.63 ± 0.02	0.44 ± 0.03	0.59	0.78
k_{cat} / K_M (M^{-1} s^{-1})	8,956.20	1,472.00	4635.45	3223.14

Table 3.3 Kinetic parameters of **PCAS/NTH** using the pyrophosphate production assay comparing with the A_{230nm} UV-vis assay.

3.3.3 Substrate specificity assay

Different chain lengths of dicarboxylic acid and fatty acid were tested with PCAS/NTH by using the PPi production assay under V_{max} conditions including dicarboxylic acids - glutaric acid (C5 dicarboxylic acid), apidic acid (C6 dicarboxylic acid), pimelic acid (C7 dicarboxylic acid), suberic acid (C8 dicarboxylic acid), azelaic acid (C9 dicarboxylic acid) and fatty acids - heptanoic acid (C7 monocarboxylate), octanoic acid (C8 monocarboxylate). The results show PCAS is highly specific for pimelic acid with k_{cat} values of <1% compared to the C7 substrate for all dicarboxylic acid chain lengths tested. And importantly, only trace activity was detected with both C7 monocarboxylate and C8 monocarboxylate (as shown in **Fig. 3.20** and **Table 3.4**).

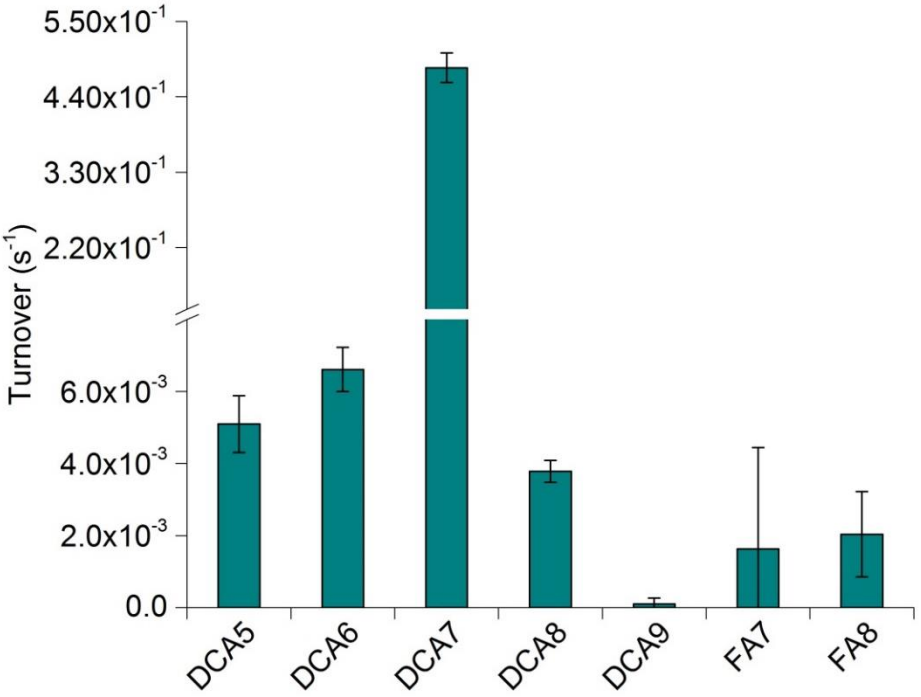


Fig. 3.20 Assay to determine the chain length specificity of PCAS/NTH with dicarboxylic acids and fatty acids. The pyrophosphate release assay was used to determine enzyme activity and is based on turnover (k_{cat}, s^{-1}) in the presence of PCAS/NTH enzyme (0.1 μM), different carboxylic acids (dicarboxylic acid, DCA, mono fatty acid, FA, 1.5 mM), ATP (1.0 mM) and CoASH (1.0 mM).

Substrates		Turnover (s^{-1})
C5 dicarboxylic acid	Glutaric acid	0.0051
C6 dicarboxylic acid	Adipic acid	0.0066
C7 dicarboxylic acid	**Pimelic acid**	**0.48**
C8 dicarboxylic acid	Suberic acid	0.0038
C9 dicarboxylic acid	Azelaic acid	0.0001
C7 fatty acid	Heptanoic acid	0.0016
C8 fatty acid	Octanoic acid	0.0020

Table 3.4 Turnover values (k_{cat}) of the dicarboxylic fatty acid substrates of different chain lengths for PCAS/NTH determined using the PPi assay.

3.3.4 High Performance Liquid Chromatography (HPLC) analysis of acyl-CoA formation

The production of pimeloyl-CoA catalysed by both PCAS/NTH (after TEV protease cleavage, un-tagged PCAS) and PCAS/NHis (N-His_6 tagged PCAS) was confirmed by HPLC assay and mass spectrometry.

3.3.4.1 HPLC assay by using C8 reverse phase column

Due to the lack of a commercial pimeloyl-CoA standard, both the enzyme-catalysed reaction (PCAS/NHis) and negative control were performed under the same conditions and then loaded on to a C8 RP-HPLC column. In **Fig. 3.21**, the ATP and CoASH peaks can be easily assigned based on available standards and AMP shows a similar elution time from the C8 column to ATP due to their similar structures. In the negative control, the ATP peaks remains constant, this means there was very little degradation. The CoASH peaks change suggesting some interconversion between oxidised/reduced forms. In contrast, for the enzyme catalysed reaction, the ATP and CoASH peaks decrease along with an increasing AMP peak (shown as a shoulder due to the large peak size of ATP). As well as these changes, there is a new peak **(D)** eluting at 10.9 min which grew in size between 10 min and 4 hours, which we tentatively assigned as pimeloyl-CoA and later confirmed as such by LC-MS analysis.

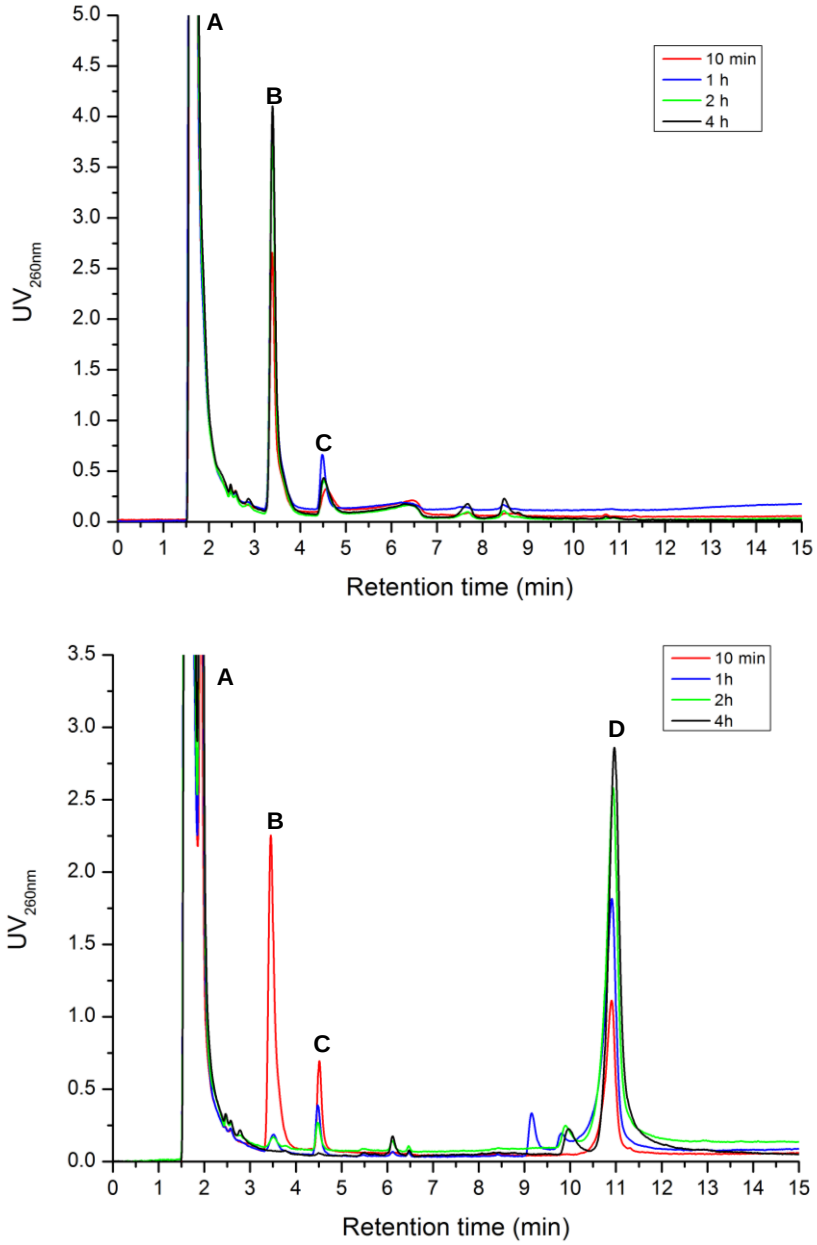


Fig. 3.21 HPLC results for **PCAS/NHis** on C8 column. **Upper**: negative control (**Peak A** (~1.9 min) is ATP; free CoASH splits into two **peaks B** (3.5 min) and **C** (4.5 min) – reduced and oxidised forms; the extra small peaks are impurities). **Lower**: PCAS/HNis catalysed reaction (**Peak A** is a mixture of ATP and AMP; free CoASH splits into two peaks **B** and **C**; **peak D** (10.9 min) is the final product pimeloyl-CoA (the extra small peaks are unknowns).

The production of pimeloyl-CoA was confirmed by LC-MS, and a clear peak at 910.2544 Da (1+) corresponding to pimeloyl-CoA ($C_{28}H_{46}N_7O_{19}P_3S$, predicted mass: 910.1855 Da, **Fig. 3.22** in red colour and **Fig. 3.23**) was observed. In principle, the different between the predicted mass and the practical mass should be less than 3 ppm, but our error is more than this due to the lack of instrument calibration. After purification and lyophilisation, pimeloyl-CoA was obtained as white powder and was then confirmed by 1H NMR (**Fig. 3.24**, **Appendix 5.10**).

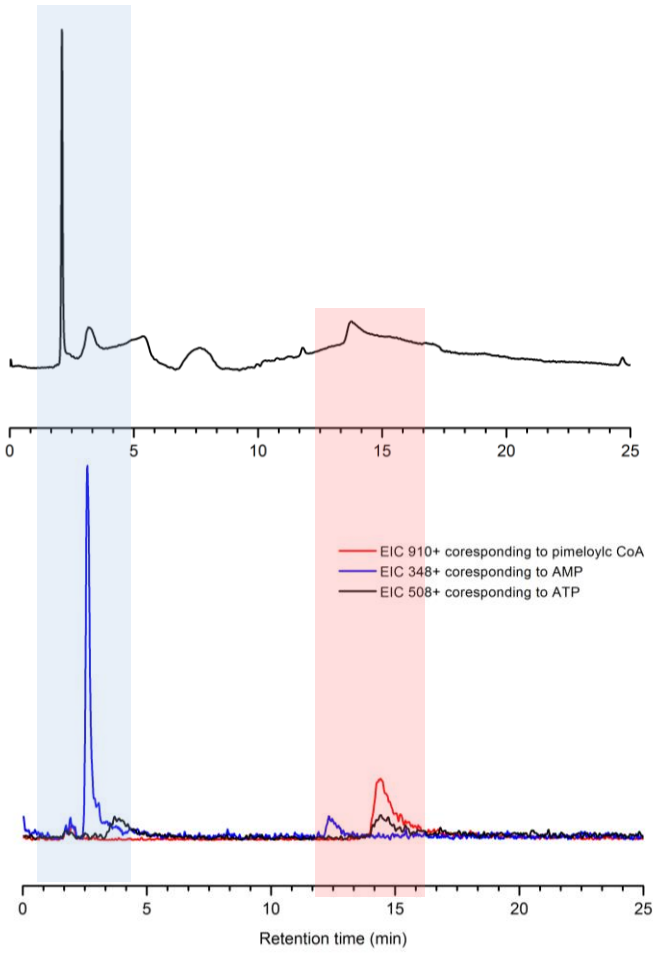


Fig. 3.22 LC-MS of PCAS/NHis catalysed reaction. **Upper**: the UV-vis A_{260nm} chromatogram, **Lower**: the EIC chromatogram (EIC 910+ corresponding to pimeloyl-CoA is in red; EIC 348+ corresponding to AMP is in blue; EIC 508+ corresponding to ATP is in black).

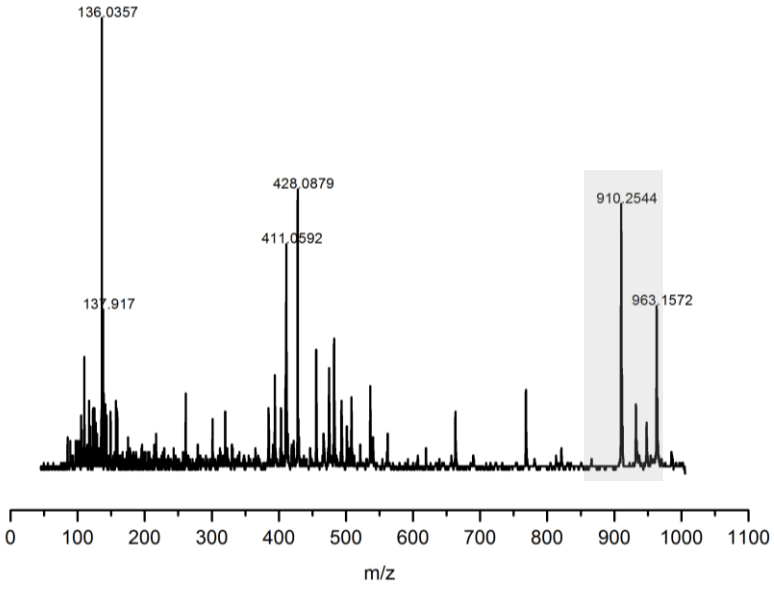


Fig. 3.23 The mass spectrum of the fractions eluting from the C8 column at 14.3-14.5 min.

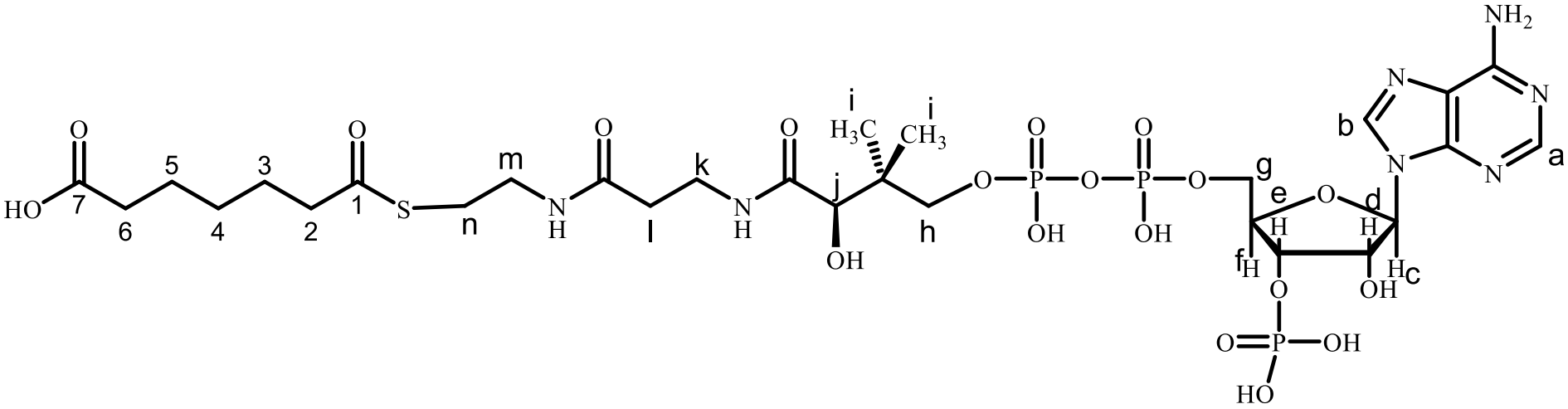


Fig. 3.24 1H NMR (1H, 500Hz,) of pimeloyl-CoA produced by PCAS/NHis: 8.62 (b, 1H, s); 8.37 (a, 1H, s); 6.18 (c, 1H, d); 4.56 (f, 1H, m); 4.21 (g, 2H, m); 4.00 (j, 1H, s); 3.81 (h, 1H, m); 3.55 (h,1H,m); 3.41 (k, 2H, t); 3.30 (m, 2H, t); 2.95 (n, 2H, t); 2.57 (1, 2H, t); 2.39 (2, 2H, t); 2.31 (6, 2H, t); 1.54 (3 and 5, 4H, m); 1.25 (4, 2H, m); 0.89 (I, 3H,s); 0.76 (I, 3H,s).

3.3.4.2 HPLC assay by using C18 reverse phase column

An optimised HPLC assay using C18 RP-HPLC column was also performed to validate the formation of acyl-CoA by PCAS/NTH. The ATP, AMP and CoASH peaks can be easily assigned based on available standards (**Fig. 3.25 A** and **B)**. This column gave much better separation of ATP (3.2 min) and AMP (4.2 min), the peak of CoASH splits into two peaks 13.3 min and 14.8 min refer to the reduced and oxidised forms. For the enzyme catalysed reaction (**Fig. 3.25 C)**, the ATP (3.2 min) and CoASH (13.3 min and 14.8 min) peaks decrease along with an increasing AMP (4.2 min) peak. As well as these changes, there is a new peak eluting at 16.2 min, which we tentatively assigned as pimeloyl-CoA and later confirmed as such by LC-MS analysis. This suggested that the PCAS/NTH is indistinguishable from the NHis tagged version of PCAS.

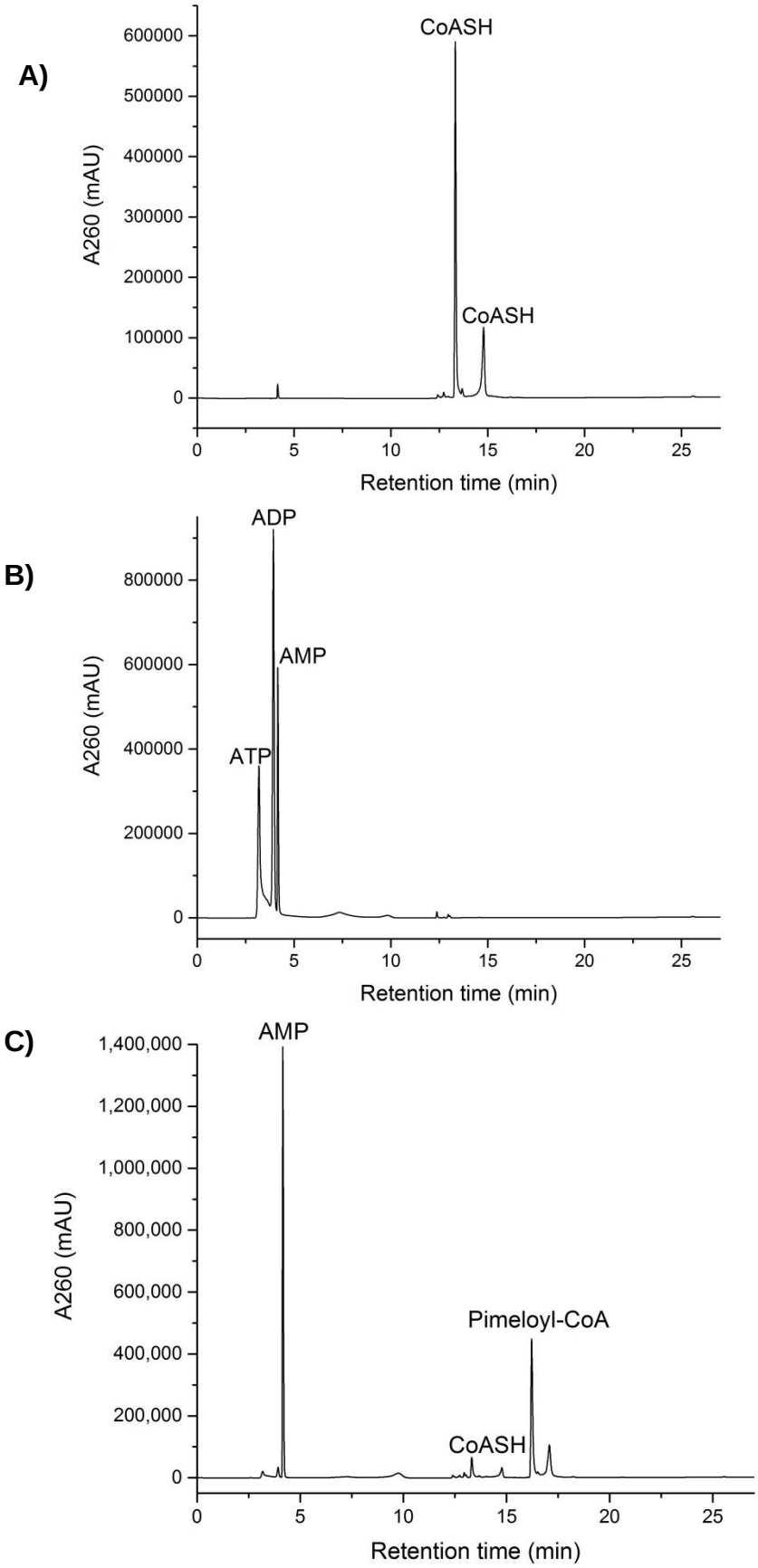


Fig. 3.25 HPLC analysis of standards and **PCAS/NTH** reaction on C18 column. **A)** Standard CoASH (13.3 min and 14.8 min), **B)** Standards ATP (3.2 min), ADP

(3.9 min), AMP (4.2 min) mixture, **C)** PCAS/NTH catalysed reaction by incubation of the enzyme with ATP, CoASH and pimelic acid shows formation of pimeloyl-CoA.

3.3.5 *In vivo* complementation assay

This method was first generated by Bower and Pero [73], they found the expression of *Bacillus subtilis* PCAS in an *E. coli* ΔBioC ΔBioH mutation strain with the supplementation of pimelic acid allows cell growth in the absence of biotin. In 2010, Lin and Cronan [74] repeated this with an *E. coli* ΔBioC mutation strain and obtained the same results. In my hands, an *E. coli* ΔBioH mutant strain was directly purchased from the Coli Genetic Stock Centre (CGSC) website (CGSC# 12019) and performed as follows. M9 minimal media agar plates supplemented with both 100 μg/ml ampicillin and 50 μg/ml kanamycin were used for mutation strain growth. The strain itself carried the kanamycin resistance gene, and PCAS/pET22b and pET22b empty vector plasmids provided the ampicillin resistance. Biotin was added at 1 μM as positive control. All the plates were incubated at 37 ^{o}C overnight. As expected, expression of the *Bacillus subtilis* PCAS in ΔBioH strain allowed cell growth on biotin-free plates with the supplementation of 0.1mM pimelic acid and 0.1mM IPTG, whereas no detectable growth was seen for the empty vector (**Fig. 3.26**). This is because pimeloyl-CoA, an important building block for biotin biosynthesis, beyond being produced by the long BioC-BioH fatty acid elongation pathway, can also be synthetized by the PCAS one step reaction (as shown in **Fig. 3.27**), thus PCAS can bypass the BioH/BioC gene knockout strain when supplied with pimelic acid.

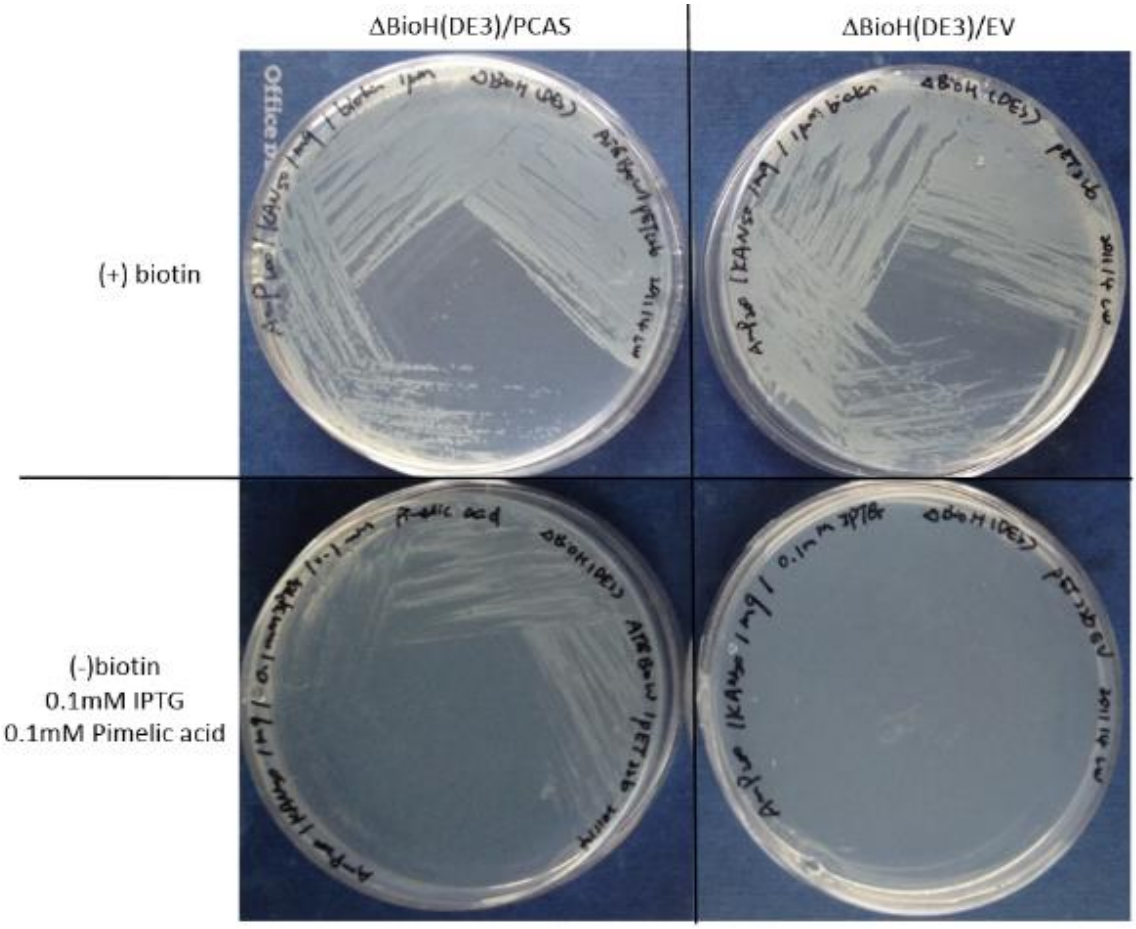


Fig. 3.26 The results of PCAS *in vivo* complementation assay. Expression of PCAS in *E. coli* ΔBioH mutation strain allowed cell growth on biotin-free plates with the supplementation of pimelic acid, whereas no detectable growth was seen for the empty vector.

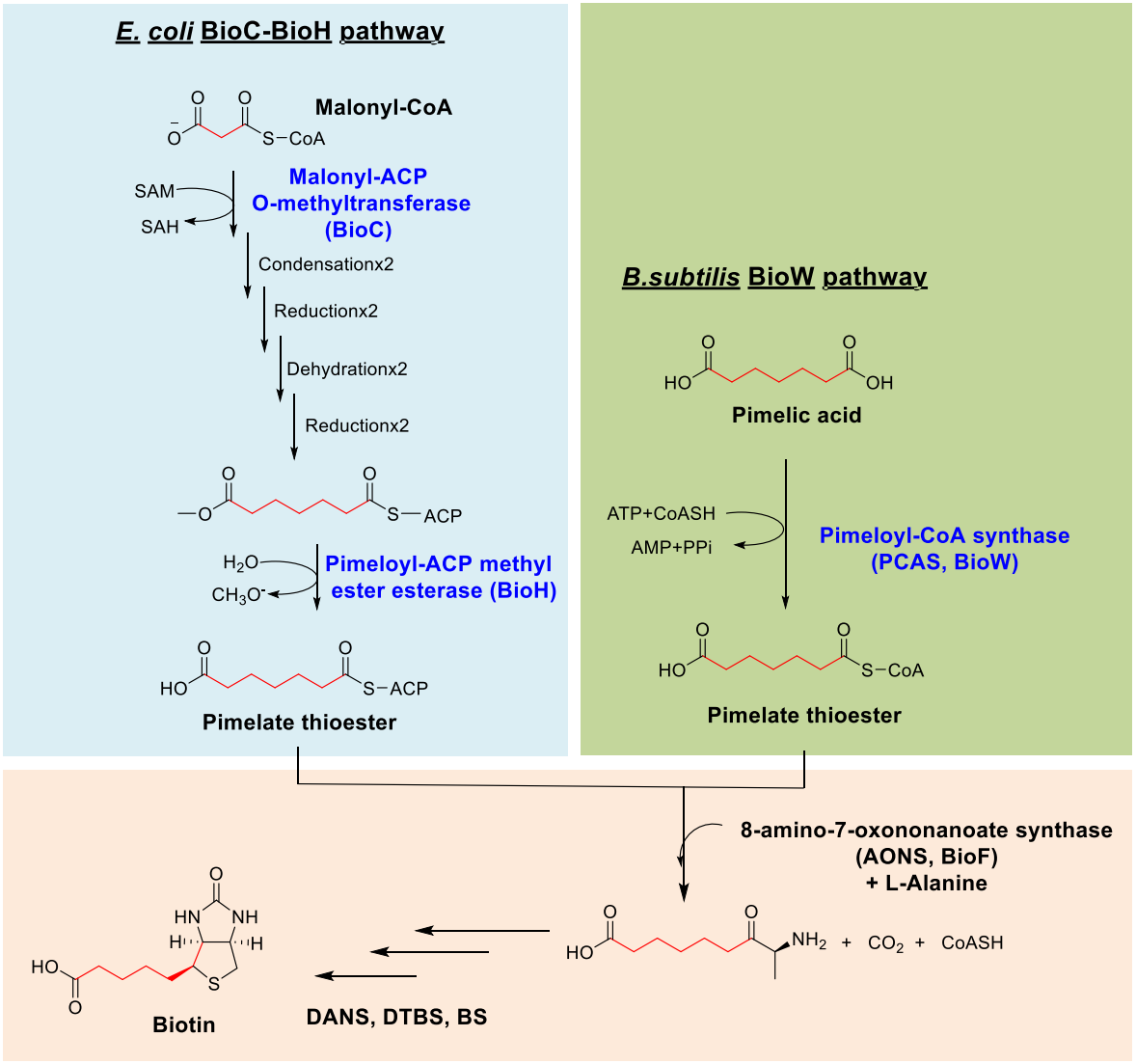


Fig. 3.27 Biotin biosynthesis pathway in *E. coli* and *B. subtilis.* Pimelate thioester, an essential building block for biotin biosynthesis, beyond being produced by the BioC-BioH fatty acid elongation pathway (blue part), can also be synthetized by the PCAS one step reaction (green part).

3.4 X-ray Crystallization and Structure Determination of PCAS

3.4.1 Crystallization

The X-ray crystal structural study of *B. subtilis* PCAS was carried in collaboration with Professor James Naismith and Dr. Lucile Moyni éat the University of St. Andrews. My role was to optimize the expression, purification and characterization of all PCAS

constructs and to screen crystallization conditions. Once any crystals were formed I also optimized the growth of diffraction-quality crystals suitable for structural studies. For all PCAS constructs (PCAS/NHis, PCAS/UT, PCAS/NTH), only PCAS/NTH at concentration of 19.6 mg mL^{-1} with 5.0 mM pimelic acid, 1.0 mM $MgCl_2$ and 5.0 mM AMP-PNP crystallised using the sitting drop method. After one week, crystals were observed in the well solution containing 30 % PEG 400, 0.1 M HEPES pH 7.5. And based on this condition, optimisation screens were prepared by varying the percentage of PEG 400 (24% - 35%) and varying the pH values of the HEPES buffer (pH 7.0, 7.5, 8.0). Different incubation temperatures (4 ^{o}C, 25 ^{o}C) were also tried. Crystals of PCAS/NTH appeared at 4 ^{o}C after a week from a hanging drop of 1 µl of protein solution (19.6 $mg.ml^{-1}$, 5.0 mM pimelic acid, 1.0 mM $MgCl_2$ and 5.0 mM ATP or AMP-PNP) with 1 µl of reservoir solution containing 26 % PEG 400, 0.1 M HEPES pH 7.5 in vapor diffusion with reservoir (**Fig. 3.28**). Then these crystals were sent to diamond by Dr. Lucile Moyni é.

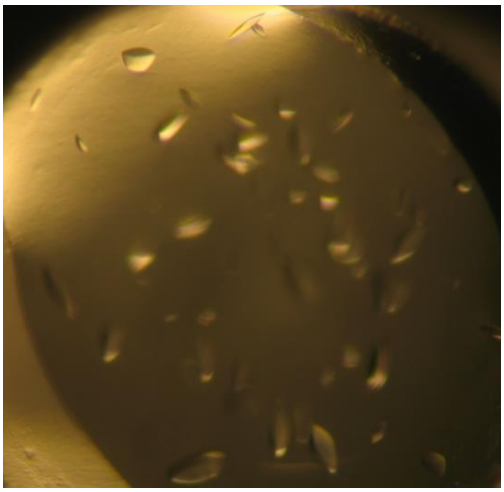
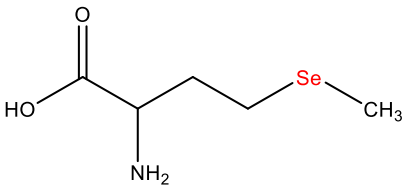


Fig. 3.28 Crystals of PCAS/NTH (19.6 mg ml^{-1}) with 5.0 mM pimelic acid, 1.0 mM $MgCl_2$ and 5.0 mM AMP-PNP.

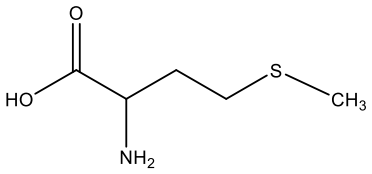
3.4.2 Phase problem

When transferring the X-ray diffraction data to the real structure image, the intensities of the scattered waves can be used to imagine slicing in all directions, but leaves a phase problem [75, 76]. The most popular method to solve this is called molecular replacement, which obtains the phase information from the crystal of a homologous molecule or the same molecule with different substrates bound. Unfortunately, there is no structural information of the whole *BioW* superfamily. Instead, heavy-atom substitution was used to overcome the phase problem in this project. Heavy atoms can be incorporated either directly into the protein target using seleno-methione wherever there is a Met residue [75]. Alternatively, heavy atoms (usually metals) can be soaked into the crystals [75].

3.4.2.1 Seleno-methionine (SeMet) labelled PCAS/NTH



Selenomethionine



Methionine

Seleno-methionine (SeMet) labelled PCAS/NTH was expressed in the *E. coli* methionine auxotroph strain B834 (DE3) after induction with 0.1 mM IPTG and incubating at 25 ^{o}C for >16 hours with SeMet containing growth media. The enzyme was purified using the method as the wild type PCAS/NTH but with reducing reagent (DTT, BME or TCEP) present in all steps. Purification by superdex 200 size-exclusion chromatography gave asymmetric peak at 76.5 mL corresponding to the theoretical mass of 59795.5 Da (**Fig. 3.29 A**, **Appendix 5.9**). Compared to subunit *Mr* of around 30 kDa determined by SDS-PAGE gel (**Fig. 3.29 D**), the active PCAS/NTH was found

to be a homodimer, the same as the wild type. The final yield is 1.3 mg per 1L of growth culture.

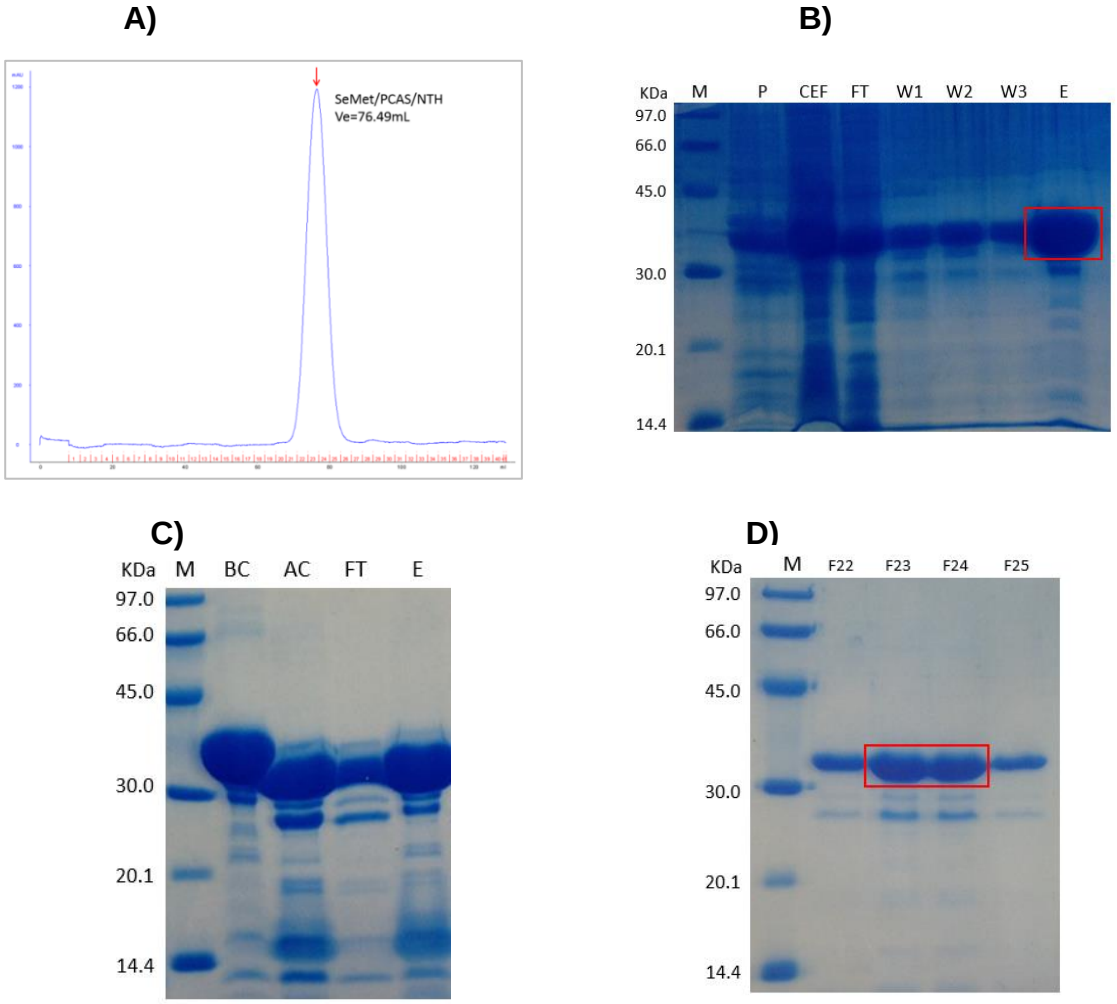


Fig. 3.29 A) AKTA trace from superdex 200 column (A_{280nm}); SDS-PAGE gel showing the purity of SeMet/PCAS/NTH during its isolation steps **B)** Cell lysis and the 1st Ni resin, **C)** TEV protease overnight cleavage and the 2nd Ni resin, **D)** superdex200 gel filtration column

Due to the incorporation of seleno-methionine, the quality of the mass spectrum is quite poor. In **Fig. 3.30** (right panel), other than the theoretical isotope envelope for the wild type enzyme, only one single broad peak was shown. This means the result is not good enough for calculating the exact mass. However, by easily comparing with the

two predicted masses at +36 charge state (The line in blue presents the unlabelled PCAS/NTH mass whereas the red line shows the fully labelled PCAS/NTH mass at +36 charge state), we can still confirm that the PCAS/NTH enzyme is fully labelled with SeMet.

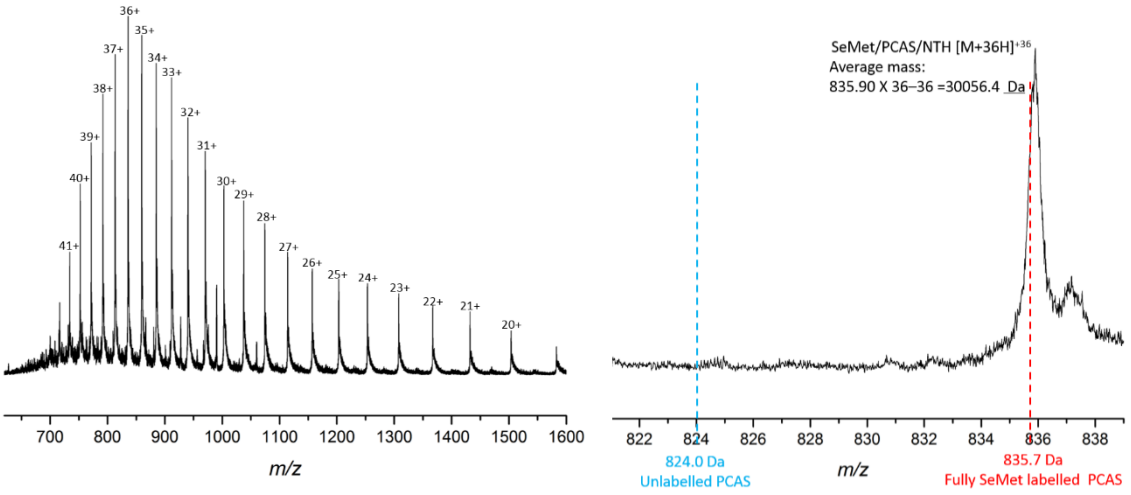


Fig. 3.30 The LC-ESI-MS results of SeMet/PCAS/NTH. The **left** panel shows the ion envelope (in black dots) of the + 41 to +20 charge states of the SeMet/PCAS/NTH. The **right** panel is the zoom in details of the +36 charge state.

Unfortunately, the same crystallization condition as the wild type enzyme gives no diffraction-quality crystals of the SeMet labelled PCAS/NTH. At the meantime, general screening under different conditions (varying buffers, pH, temperature and incubation time) was also tried, but no eligible crystal was obtained as well.

3.4.2.2 Heavy metal soaking

In calibration with Dr. Lucile Moyni é, crystals getting from **Section 3.4.1** (Crystals of PCAS/NTH with pimelic acid, $MgCl_2$ and ATP) were soaked in 50 mM K_2PtCl_4 for approximately 15 minutes and back-soaked in mother liquor containing pimelic acid,

ATP and $MgCl_2$. Then these heavy metal labelled crystals were sent to diamond as well.

3.4.3 Data Collection and Structural Solution

This part was carried out by Prof. Naismith and Dr. Lucile Moynié. Data were collected at the beamlines IO2, IO3 and IO4 at the Diamond light source Oxfordshire. Data were processed with XIA2 [77-81] or HKL2000 [82]. Phases were determined by SAD (single-wavelength anomalous diffraction) and an initial model built using the Crank pipeline [83] as implemented in CCP4i2 [84] and the structure refined with REFMAC5 and adjusted with COOT [85] and Autobuild program of PHENIX [86].

Subsequent structures were solved by molecular replacement with the program PHASER [87] using the coordinates from the experimental structure. Final models were adjusted with COOT [85] and refinement was carried out using REFMAC in the CCP4 program suite with NCS restraints and TLS parameters [88]. Final refinement statistics are given in **Table 3.5**. Coordinates and topologies of ligands were generated by PRODRG [89]. The quality of all structures were checked with MOLPROBITY [90]. Figures were drawn using PYMOL [91].

Table 3.5 Crystallographic data and refinement statistics

Ligands	K_2PtCl_4 Pimeloyl-adenylate/PPi	AMP-PNP/pimelic acid	Pimeloyl-adenylate/PPi	Pimelic acid/CoASH
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	50.61 78.36 166.70	49.59, 77.90, 166.0	50.12, 78.11, 165.82	50.59, 78.36, 166.43
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)[a]	83.35-2.44 (2.50-2.44)	55.33 – 2.04 (2.09 – 2.04)	50.0-2.34 (2.38-2.34)	70.89-2.25 (2.37-2.25)
Rmerge	0.082 (0.681)	0.059 (0.552)	0.092 (0.419)	0.057 (0.757)
I/σ(I)	23.3 (3.3)	17.3 (2.7)	21.4 (3.36)	18.8 (2.0)
Completeness (%)	99.9 (99.9)	99.2 (92.2)	99.5 (94.2)	99.8 (99.7)
Aver. redundancy	15.2 (10.5)	6.4 (4.1)	4.2 (2.9)	6.4 (5.1)
Z'	2	2	2	2
Refinement				
Resolution (Å)	83.35-2.44	41.53-2.04	35.35-2.34	57.11-2.25
No. reflections	24163	39459	26785	30569
R_{work} / R_{free}	0.200/0.242	0.203/0.232	0.199/0.234	0.217/0.2518
No. atoms				
Protein	3987	4066	4057	3979
Water molecules	18	72	99	21
Ligand/ion	92	88	88	60
B-factors (Å^2)				
Protein	56.97	41.21	41.73	59.37
Chain B-factor range	52.17(B)-61.78(A)	39.57(A)-42.85(B)	38.05(A)-45.42(B)	53.27(B)-65.47(A)
Water	44.81	36.05	32.06	53.94
Ligand/ion	52.03	34.2	31.29	83.14
Ligand B-factor range	46.89(B)-57.17(A)	33.95(B)-34.45 A)	30.90(B)-31.68(A)	78.83(A)-87.46(B)
R.m.s deviations				
Bond lengths (Å)	0.011	0.01	0.011	0.012
Bond angles ( )	1.236	1.309	1.315	1.329
Ramachandran plots summery (%)[c]	98.6/1.4/0	100/0/0	99.2/0.8/0	98.3/1.7/0

[a] Numbers in parentheses represent statistics in highest resolution shell.

[b] The lowest and highest B-factor calculated by averaging all protein and all inhibitor atoms in each chain in the asymmetric unit. The chain giving rise to the lowest and highest values are given in parentheses.

[c] % of residues in most favoured regions/allowed regions/disfavoured).

3.4.4 Structural analysis

3.4.4.1 Overall structure of PCAS

The asymmetric unit contains two monomers of the *B. subtilis* PCAS protein (**Fig. 3.31a**). Analysis with the PISA server at EBI [92] which analyses the oligomeric stability, identifies that the dimer observed in the asymmetric unit is stable burying approximately 900 $Å^2$ of surface area per monomer. The dimeric structure is thus consistent with native mass spectrometry (**Section 3.5**) and gel filtration experiments (**Section 3.2.1**). Each monomer consists of two domains, a small 70 residue N-terminal domain and a larger C-terminal domain that comprises the rest of the sequence (**Fig. 3.31b**). The N-terminal domain has a three stranded anti-parallel β-sheet (β1-3) with an α-helix lying on one face. The dimer interface is confined to the N-terminal domain that brings together the β3 strands of each monomer such the dimer has a 6 stranded antiparallel β-sheet. In addition to these main chain hydrogen bonds there are a few additional polar and hydrophobic interactions that involve the α-helix (α1). Superimposing monomers reveals that the relative orientation of domains differs within each monomer suggesting the linkage between the domains is flexible. The C-terminal domain has a central, predominantly anti-parallel 7 stranded β-sheet, each face of the sheet has two α-helices and one of the helices (α5) stacks against a small two stranded parallel β-sheet (β4 and13) (**Fig. 3.31b**). The domain has a fifth helix (α4) that packs against the loop of the C-terminus. Analysis with PDB fold at the EBI does not identify other enzymes with this fold, suggesting the structure represents an extension to the family of enzyme known to catalyse the adenylation reaction [45]. This family currently contains three classes (I-III) and so we propose that PCAS is the first representative of Class IV within this family (**Fig. 3.32**).

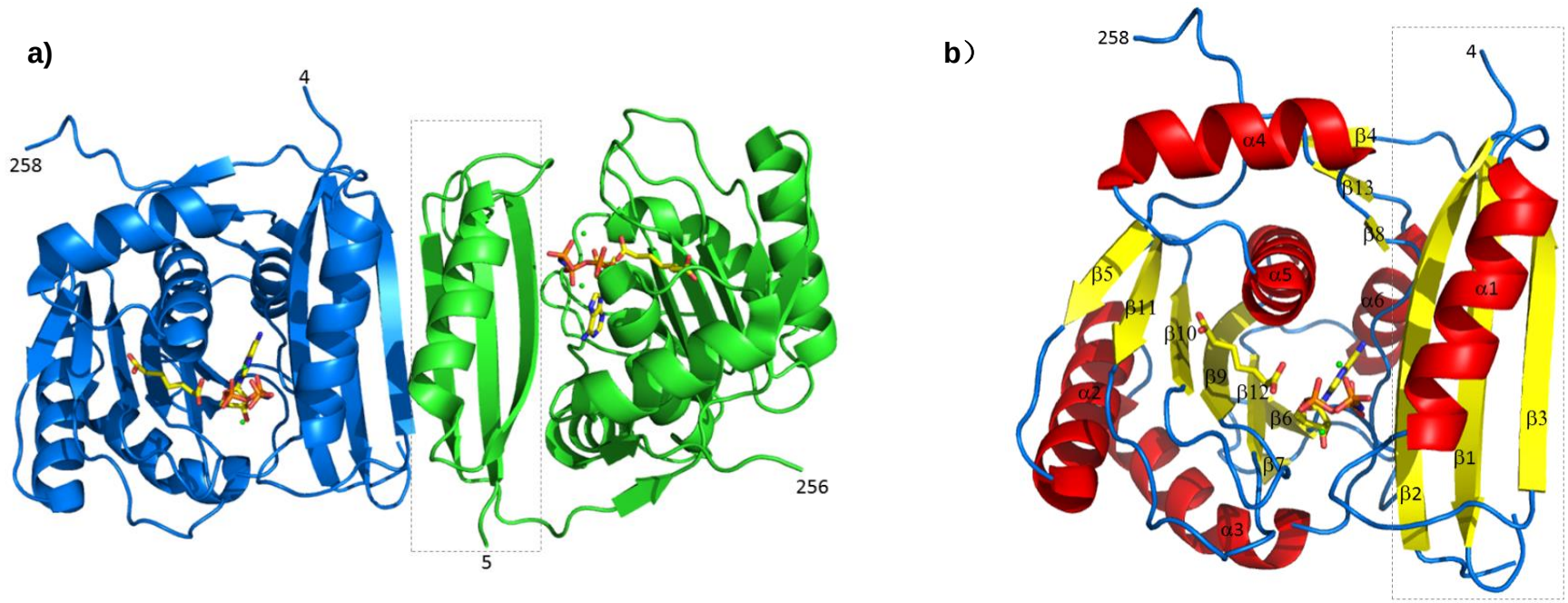


Fig. 3.31 X-ray crystal structure of *B. subtilis* PCAS. **(a)** Overall structure of the PCAS dimer. The AMP-PNP and pimelic acid molecules are shown as sticks (carbon atoms in yellow, oxygen in red, nitrogen in blue and phosphate in orange). The A and B subunits of the protein are colored in green and blue, respectively. The N-terminal domain is highlighted by a dashed box. **(b)** Overall structure of PCAS monomer. Secondary structure elements are labelled and coloured in red (α-helices), yellow (β-sheets) and blue (loops). The AMP-PNP and pimelic acid molecules are shown as sticks (carbon atoms in white with other atoms as a), Mg2+ ions are shown as green spheres.

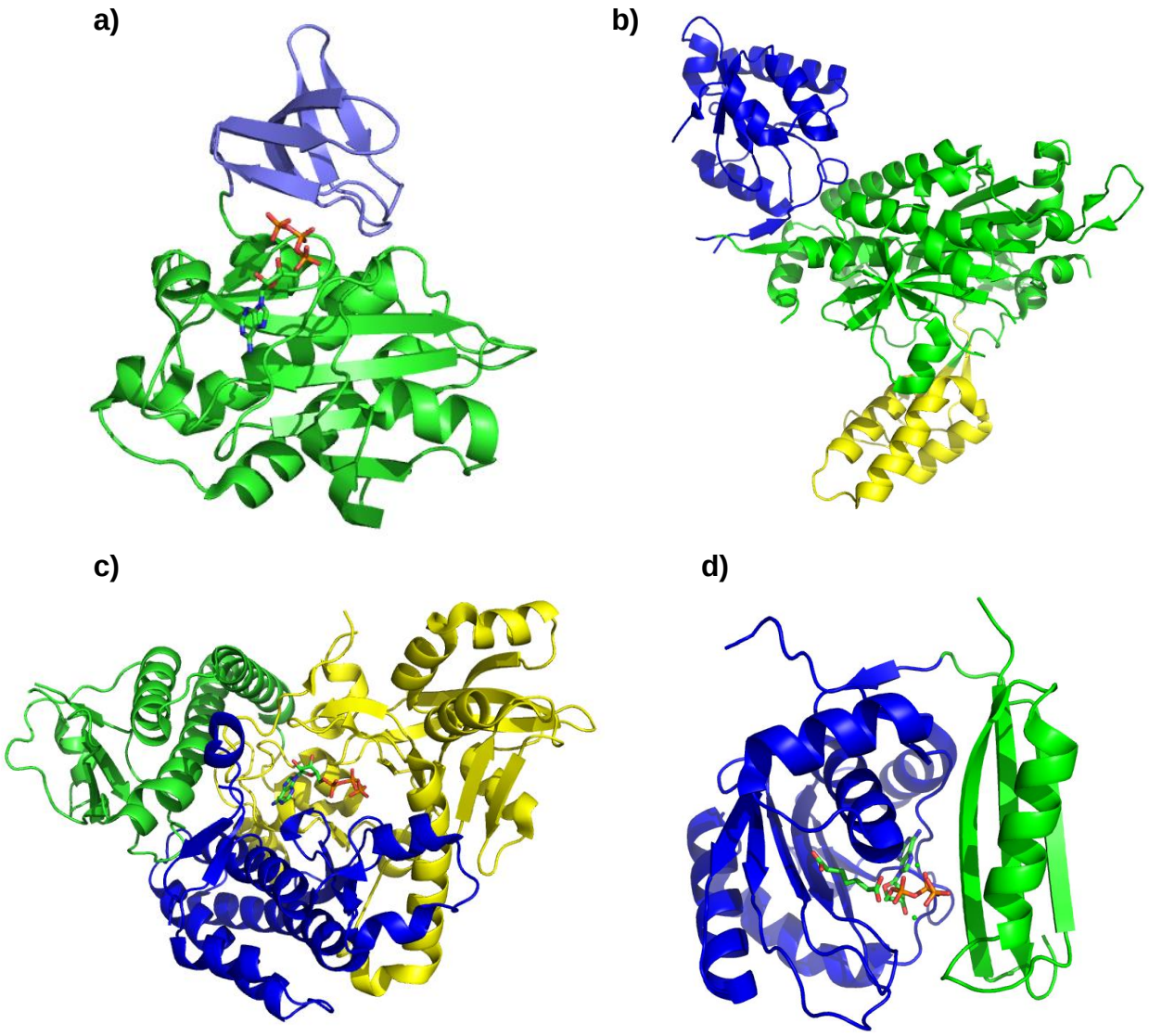


Fig. 3.32 Structures of classes I, II, III of adenylating enzymes and the proposed class IV PCAS. **(a)** Biotin protein ligase (BPL) from *Aquifex aeolicus* in complex with ATP (sticks) (PDB code: 3efs) [16] , the N-terminal domain is in green and C-terminal domain in blue. **(b)** Human glycyl-tRNA synthetase (PDB code: 2pme) [51], the central catalytic domain is in green and the anticodon-binding domain in blue, insertion I domain in blue. **(c)** Achromobactin synthetase protein D (AcsD) from *Pectobacterium chrysanthemi* in complex with ATP (sticks) (PDB code: 2W02) [47], the N-terminal domain is in green, central domain in yellow and C-terminal domain in blue. **d)** Pimeloyl-CoA synthetase (PCAS) from *Bacillus subtilis* (PDB code: 5flg), the N-terminal domain is in green and C-terminal domain in blue.

3.4.4.2 ATP binding site of PCAS

The enzyme recognises three substrates catalysing two reactions, both around the same carboxylate of pimelic acid, which, in the first step, reacts with ATP and in the second step, with CoASH. The structures of two complexes (AMP-PNP plus pimelic acid and pimeloyl-adenylate plus PPi) were obtained in this study, and their electron density map was shown in **Fig. 3.33**. Both co-complexes locate the active site almost entirely within the C-terminal domain and identify the residues in contact with ATP and pimelic acid.

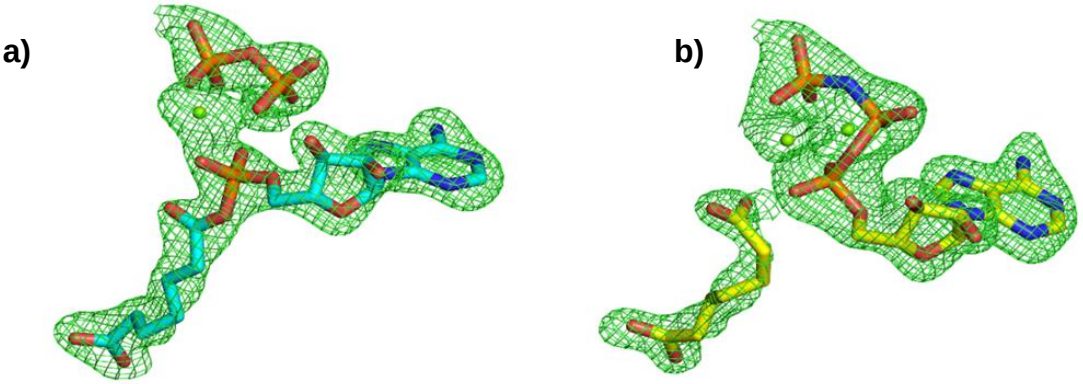
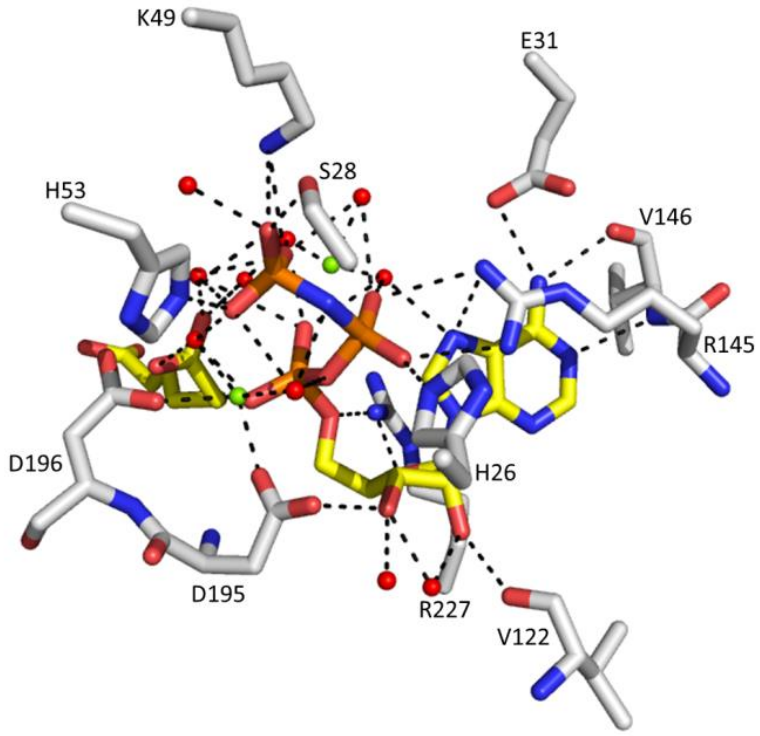


Fig. 3.33 F_o-F_c electron density omit map at 3 σ around the pimeloyl-adenylate (a) and around AMP-PNP and pimelic acid (b). Carbon atoms are coloured in cyan (pimeloyl-adenylate) and yellow (AMP-PNP/pimelic acid). Oxygen atoms are in red, nitrogen in blue, phosphate in orange and Mg^{2+} ions are shown as green spheres. This figure was generated by collaborator Dr. Lucile Moynié.

In both complexes the adenosine and ribose rings, the β and γ phosphates and the two Mg^{2+} ions are located in essentially identical positions making the same contacts with the protein (**Fig. 3.34** and **Fig. 3.35**). The adenosine ring sits in a pocket with the side chain of Arg145 stacking on its face; the ring itself makes two hydrogen bonds to the main chain at Val146, a hydrogen bond to Glu31 and to a network of water molecules. The ribose ring hydrogen bonds to Arg227 (**Fig. 3.35b**), the main chain of Val122 and to the network of water molecules (**Fig. 3.34**).

a)



b)

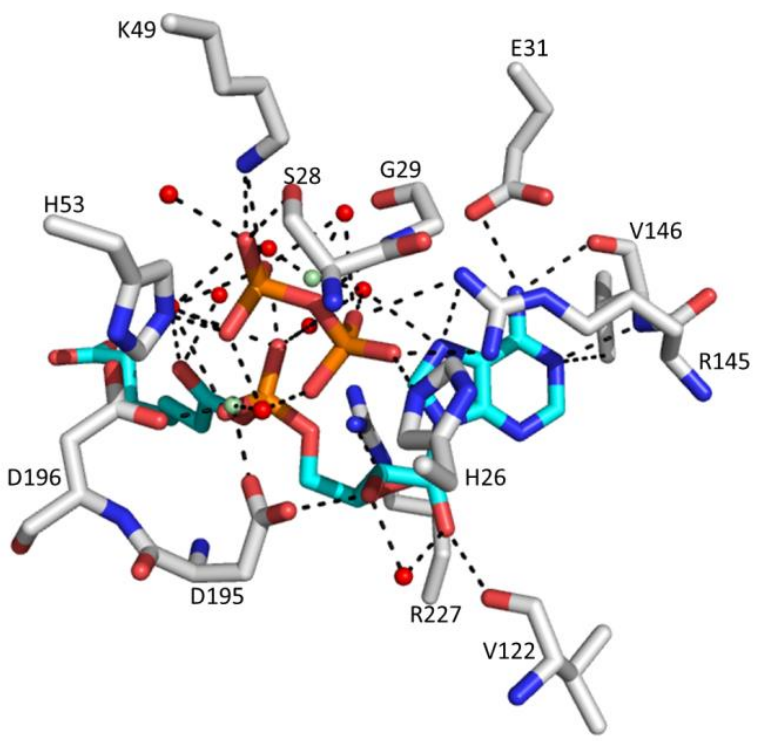


Fig. 3.34 Binding sites of AMP-PNP and pimelic acid **(a)** and pimeloyl-adenylate and PPi **(b)**. Carbon atoms are coloured in cyan (pimeloyl-adenylate) and yellow (AMP-PNP/pimelic acid). Hydrogen bonds are shown as black dashes lines. Oxygen atoms are in red, nitrogen in blue, phosphate in orange and Mg^{2+} ions are shown as green spheres.

Comparing the atoms at the α phosphate in both complexes reveals that the phosphorous atom has shifted less than 1 Å and the three oxygen atoms that remain are unaffected by bond breaking/making and have shifted less than 0.2 Å (**Fig. 3.35a, b**). Two of these oxygen atoms are each bound to a different Mg^{2+} ion, the third oxygen atom which connects to the ribose ring is salt bridged to Arg227 (**Fig. 3.35b**).

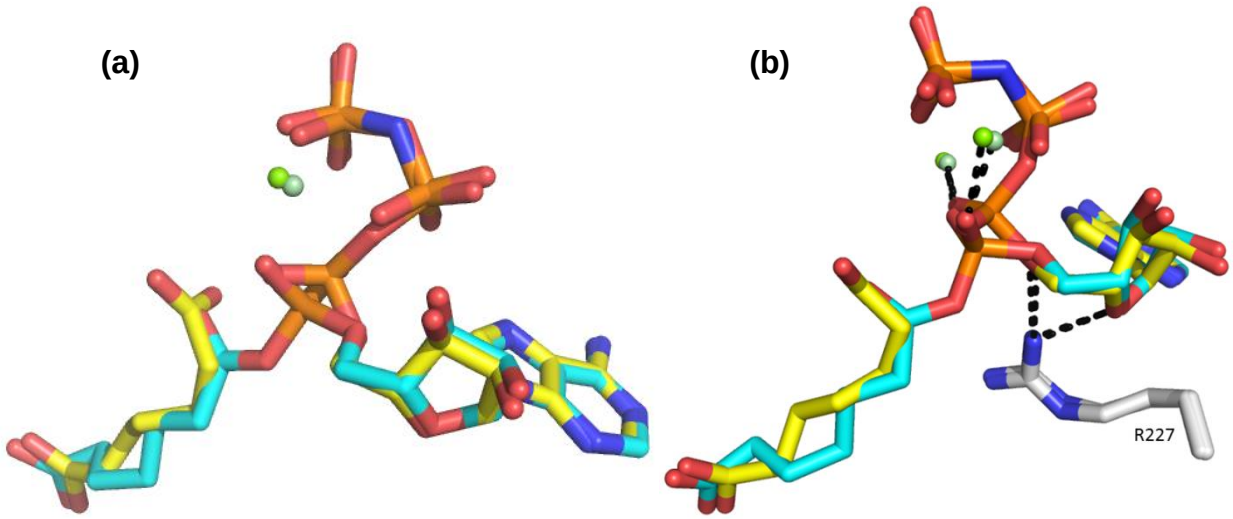


Fig. 3.35 Superposition of the pimeloyl-adenylate/PPi and AMP-PNP/pimelic acid complexes. For clarity, the protein has been removed. The colour scheme is the same as **Fig. 3.34**. **(a)** The superposition allows a detailed model for the pentavalent phosphorous transition state to be deduced, **(b)** shows the details of the three oxygen atoms which bonded to the α phosphate.

Combining the complexes reveals that the α oxygen-phosphorous bond which connects to the β-phosphate makes an angle of 180° with the oxygen atom that links the α phosphorous to the carboxylate of pimelic acid (**Fig. 3.35a**). From this, a model of the penta-covalent bipyramidal transition state can be constructed.

3.4.4.3 Active site of PCAS

In the enzyme complex with pimeloyl-adenylate and PPi, as shown in **Fig. 3.36b**, the carboxyl adenylate bond sits between Ser194 on one side and on the other side two water molecules (W1 and W2) which are hydrogen bonded to each other. The side chain of Ser194 is hydrogen bonded to the main chain of the protein. The hydrogen bond network of the water molecules is extensive and differs between monomers. Two water molecules are positioned close to the adenylate ester group but neither is ideally positioned to attack the carbonyl. One water molecule (W1) is bound to the guanidyl group Arg227 which would be predicted to lower the nucleophilicity of the water; the other water (W2) that is hydrogen-bonded to the first water would require some re-arrangement to be correctly positioned. In the adenylate complex Arg170 is hydrogen bonded to the pimelic acid carboxylate oxygen atom that is not linked to the phosphorous. In contrast, in the enzyme complex with AMP-PNP and pimelic acid, the carboxylate group is in a different orientation and not aligned for attack at the α phosphate (**Fig. 3.36a**). The carboxylate makes a different salt bridge to Arg170 and hydrogen bonds to Ser194 as well as waters molecules ligated to the Mg^{2+} ions. The β and γ phosphate groups bind two Mg^{2+} ions, the β phosphate makes interactions with the N-terminal domain side chains of His26 and Arg145 and the γ phosphate makes contacts with the side chains of Lys49 and His53. The Mg^{2+} ions both have octahedral coordination and are ligated by three phosphate oxygen atoms. One magnesium ion has three bound water molecules and the other magnesium ion the side chains of Asp195, Asp196 and one water complete the coordination shell.

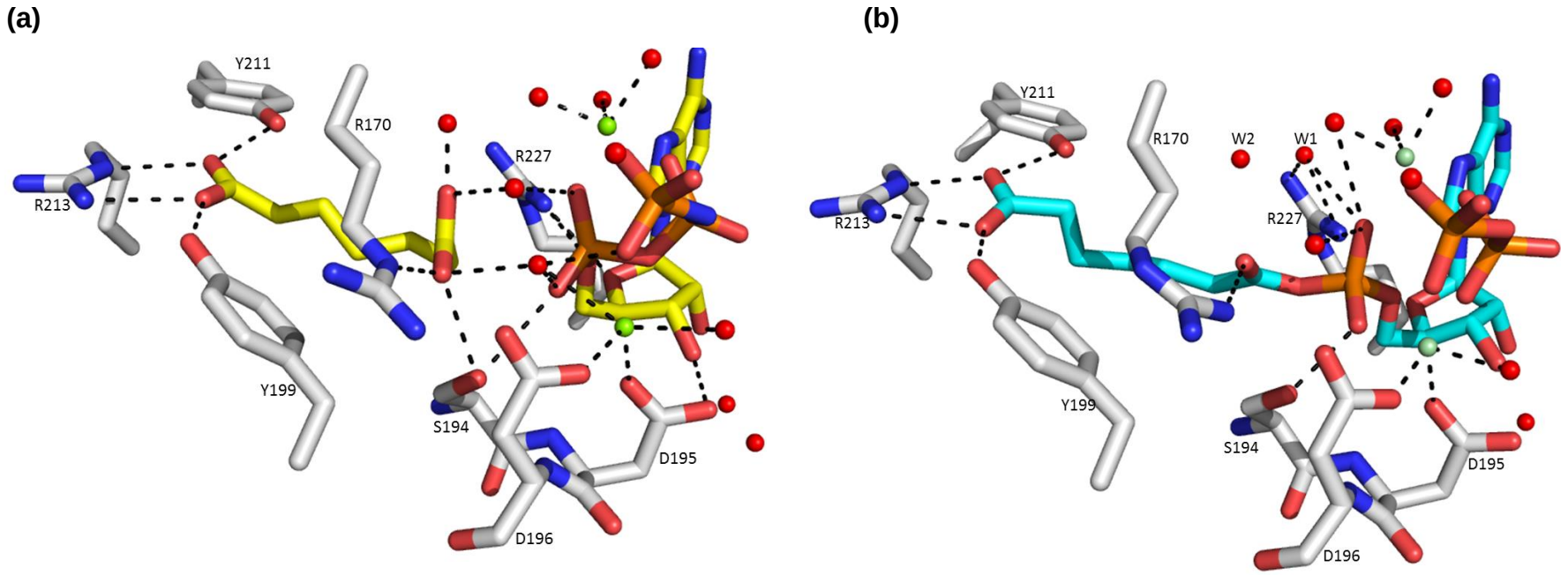


Fig. 3.36 Binding sites of AMP-PNP and pimelic acid **(a)** and pimeloyl-adenylate and PPi **(b)**. For more clarity, only the polar interactions have been represented. The colour scheme is the same as **Fig. 3.34**. In **(b)**, the pimeloyl-adenylate is stable, despite two water molecules (W1 and W2) being positioned close to the carbonyl group of adenylate ester. The hydrogen bonding network and positioning of waters inhibits their ability to attack the ester.

3.4.4.4 Pimelic acid binding site of PCAS

Pimelic acid binds in an elongated tunnel (**Fig. 3.37**) and the carboxylate group at the C7 (ω) position, that does not react with ATP is anchored in both structures by a salt bridge to Arg213, as well as forming hydrogen bonds to two tyrosines side-chains (199 and 211, **Fig. 3.36**). There are some small differences in the precise conformation of the alkyl chains between the two structures that arise from the different location of the reactive carboxylate (C1, α) in the two subunits. Despite this, in both structures the alkyl chain makes contacts (< 4 Å) along its length with the enzyme such that any substitution of hydrogens with larger heteroatoms would result in clashes.

3.4.4.5 Coenzyme A binding site of PCAS

A crystal of the enzyme with CoASH and pimelic acid has density for two pimelic acid molecules but only one CoASH molecule (**Fig. 3.38**). There is however no reliable density for the pantethine arm (only the adenine, ribose and three phosphates clearly defined). CoASH is bound close to the dimer interface where it interacts the strands of the N-terminal domain. The adenine ring makes a hydrogen bond to Glu66 from the other subunit, whilst the 3-phosphate is bound by a cluster of three Arg residues (13,123,145). A surface plot (**Fig. 3.37**) of the complex reveals a highly positively charged tunnel, formed by residues Arg13, His26, Glu31, Arg123 and Arg145, that reaches from the CoA binding sites to the pimelic acid. The 5-pyrophosphate group does not make any hydrogen bonds. Due to the two fold non-crystallographic axis were a second molecule of CoA bound the terminal 5′ phosphates of the CoA would be closer than 4.0 Å.

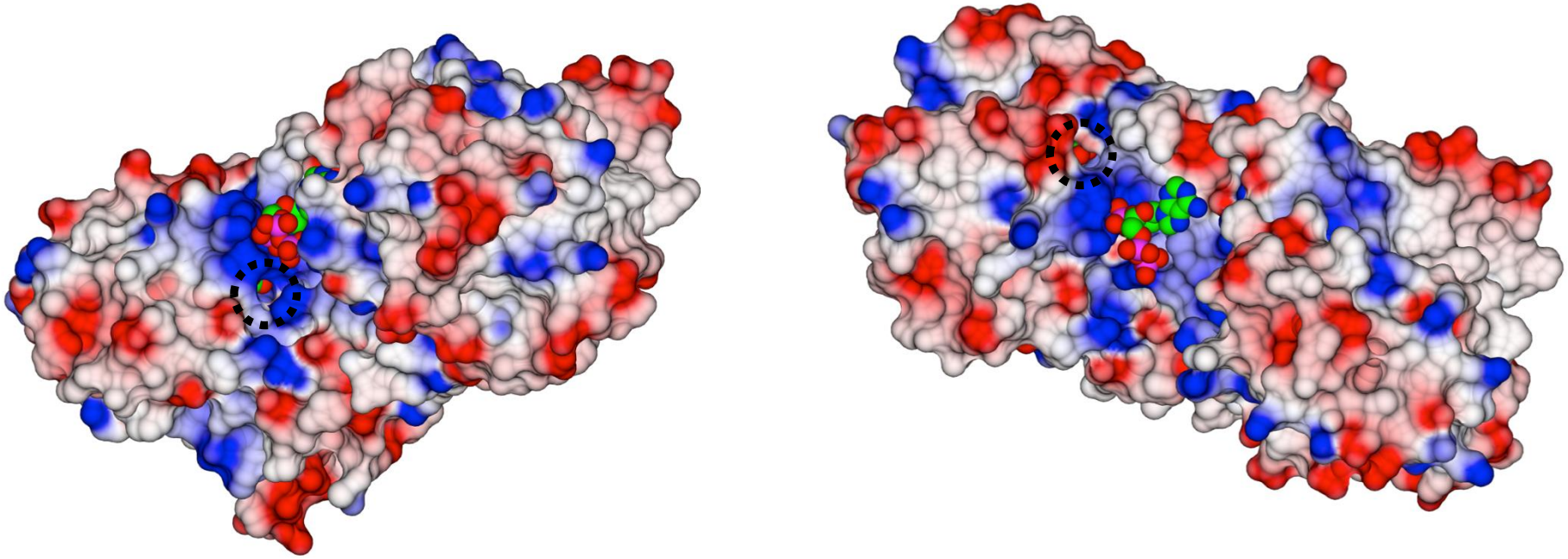


Fig. 3.37 Two views of the electrostatic surface of the PCAS with the ordered portion of CoA shown as spheres. There are two possible routes to the active site denoted with dashed circle. Positively charged regions are shown in blue and negatively regions are in red. Oxygen atoms are in red, nitrogen in blue, phosphate in pink. This figure was generated by collaborator Dr. Lucile Moynié.

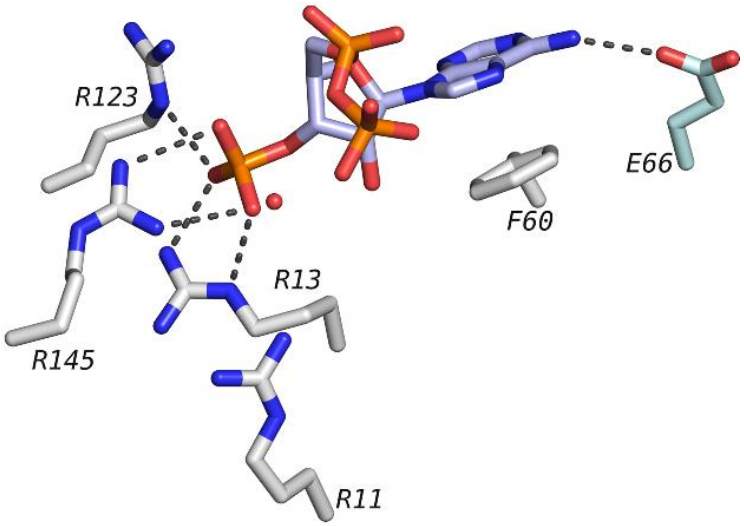


Fig. 3.38 Binding site of the adenosine 3', 5'-diphosphate group of CoA. Residues within 4.0 Å of the CoA are displayed. Carbon atoms of the CoA are in light blue and residues of A and B subunits are in light cyan and white, respectively. Other colour scheme is the same as **Fig. 3.34**. This figure was generated by collaborator Dr. Lucile Moynié.

3.5 Native mass spectrometry analysis of PCAS

Native ESI-MS mass spectrometry analysis can be used to analyse proteins in their folded state and can also be applied to enzymes to detect bound ligands and intermediates [93]. When PCAS was exchanged into an ammonium bicarbonate buffer and sprayed under native conditions we observed charge state distributions consistent with monomeric and dimeric protein (**Fig. 3.39a**). Deconvolution of the observed +14 charge state is consistent with the mass of the PCAS dimer in the ligand-free form (boxed in **Fig. 3.39a**, expanded in **Fig. 3.39b**, observed *m/z* = 4233.9, theoretical *m/z* = 4233.6).

When PCAS was mixed with pimelic acid and MgATP, then immediately applied to a spin-column to remove excess reagents, we observed two species, a dominant mass of *m/z* = 4304 consistent with the formation of the PCAS dimer: 2 x pimeloyl-adenylate complex with 2:2 stoichiometry (mass difference 980.7 Da, **Fig. 3.39c**) and a smaller peak with a mass consistent with a single pimeloyl-adenylate bound (*m/z* = 4268.9). When PCAS was mixed with pimelic acid, MgATP and CoASH we observed an increase in the relative abundance of the PCAS dimer with only one pimeloyl-adenylate bound (*m/z* = 4269.0, **Fig. 3.39d**) and of the ligand-free PCAS (*m/z* = 4234.4). Under these conditions, we only observed ions consistent with the binding of the pimeloyl-adenylate and failed to detect complexes with either pimelic acid, MgATP or PP_i bound. This strongly suggests that the PCAS/adenylate intermediate complex is highly stable and is broken down in the presence of CoASH.

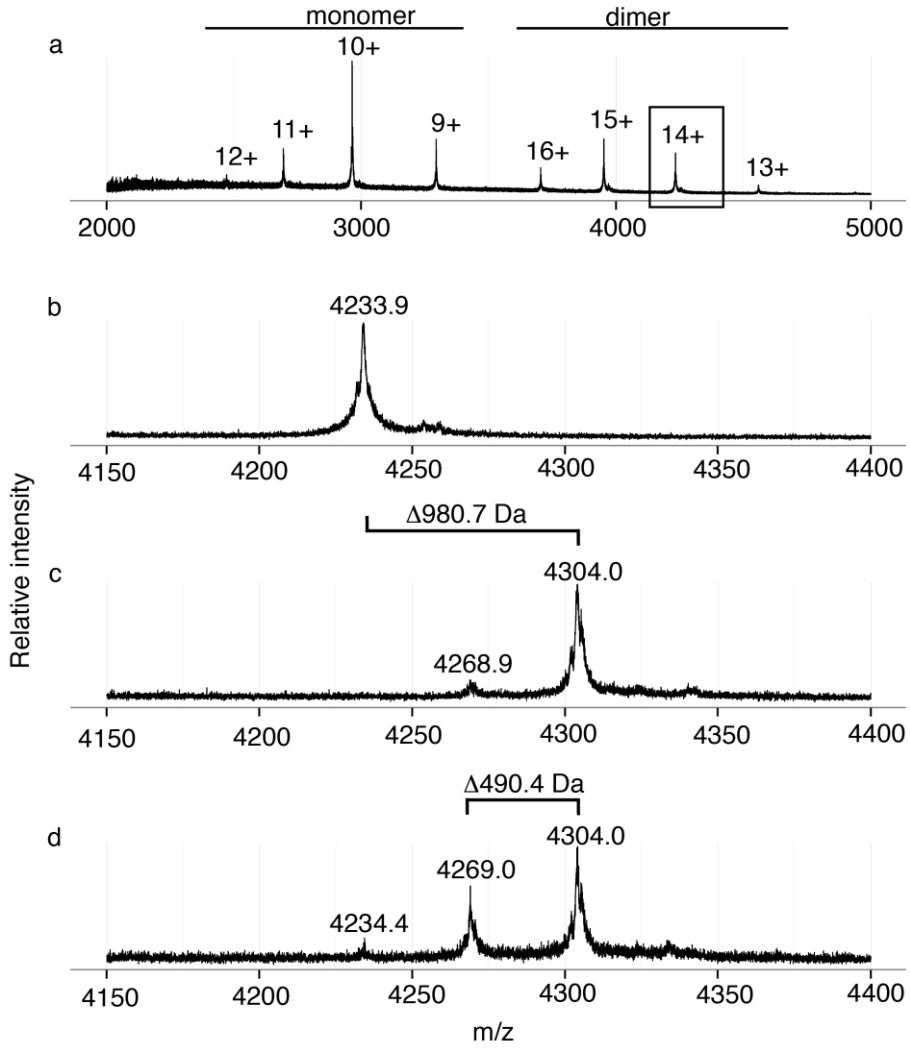


Fig. 3.39 Native mass spectrometry analysis of PCAS. **(a)** PCAS is observed as a monomer (charge states 12+ to 9+) and a dimer (charge states 16+ to 13+) by native mass spectrometry when ionised from 200 mM ammonium bicarbonate buffer. The 14+ charge state dimer species [boxed in **(a)** and expanded in **(b)**], has observed *m/z* = 4233.9 and is in good agreement with the theoretical *m/z* = 4233.6. **(c)** Incubation of PCAS with 1.0 mM pimelic acid and 1.0 mM MgATP led to the formation of a 14+ ion with *m/z* = 4304.0. This mass difference (980.7 Da) is consistent with non-covalent binding of two molecules of pimeloyl-adenylate (theoretical mass 489.4 Da). **(d)** Addition of 0.5 mM CoASH to the incubation described in **(c)** led to an increase in the intensity of the ion with *m/z* = 4269.0 which is consistent with one molecule of pimeloyl-adenylate.

3.6 Site directed mutagenesis of residues potentially involved in substrate binding and catalysis.

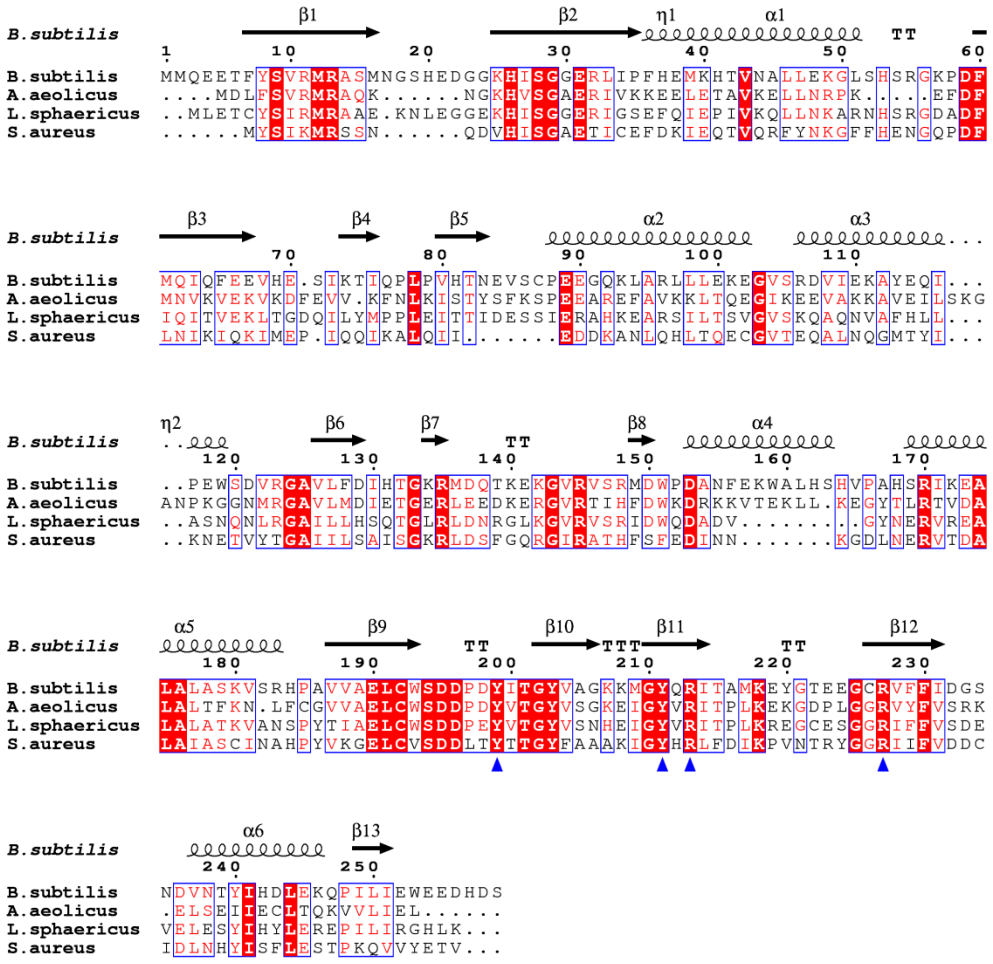


Fig. 3.40 Sequence alignment of bacterial PCAS. The sequences used are from *Bacillus subtilis* (UNIPROT code: P53559, BIOW_BACSU), *Aquifex aeolicus* (O67575, BIOW_AQUAE), *Lysinibacillus sphaericus* (P22822, (IOW_LYSSH), and *Staphylococcus aureus* (P67549, BIOW_STAAM). The residues targetted for engineering the fatty acid specificity (Tyr199, Tyr211, Arg213, Arg227) are marked with blue triangles. Figure produced using ESPript 3 [94].

The X-ray structure of the *B. subtilis* PCAS with bound ligands and adenylate intermediate revealed residues in the active site potentially involved in substrate specificity and catalysis. This structural data combined with a multiple sequence alignment of PCAS isoforms from various organisms (**Fig. 3.40**) guided the selection of residues for site directed mutagenesis.

```
  0   GAMEEETFYS VRMRASMNGS HEDGGKHISG GERLIPFHEM KHTVNALLEK
 50   GLSHSRGKPD FMQIQFEEVH ESIKTIQPLP VHTNEVSCPE EGQKLARLLL
100   EKEGVSRDVI EKAYEQIPEW SDVRGAVLFD IHTGKRMDQT KEKGVRVSRM
150   DWPDANFEKW ALHSHVPAHS RIKEALALAS KVSRHPAVVA ELCWSDDPDY
200   ITGYVAGKKM GYQRITAMKE YGTEEGCRVF FIDGSNDVNT YIHDLEKQPI
250   LIEWEEDHDS 259
```

Fig. 3.41 The sequence of the full length clone and recombinant *B. subtilis* PCAS/NTH after TEV cleavage to remove the His_6 tag is shown. The PCAS/NTH numbering begins at Ala1 and ends with Ser259. Residues in the active site are in red (Tyr199, Tyr211, Arg213 and Arg227).

As shown **Fig. 3.41**, four residues were chosen (Y199, Y211, R213, R227) to mutate in an effort to probe their role in substrate recognition and the catalytic mechanism. We were especially keen to understand the origin of the specificity towards the unusual C7 dicarboxylic acid substrate pimelate, since PCAS ultimately controls the structure of the final biotin vitamin. Seven PCAS mutants, including PCAS Y199F, Y211F, R213A, R227E, R227K, R11A and R13A were selected as targets and standard site directed mutagenesis was used to prepare the mutant PCAS genes in the pEHISTEV plasmid. Each mutant was expressed in BL21(DE3) and, gratifyingly, the solubility was similar to the wild type enzyme. Each mutant purified in a similar manner to the wild type PCAS/NTH which gives un-tagged protein in the end (with equivalent yields). Analysis by S200 size exclusion chromatography showed all mutants are homodimers (**Fig. 3.42**, **Table 3.6**). And all of them were later confirmed by LC-ESI-MS and the data was shown in **Table 3.7**.

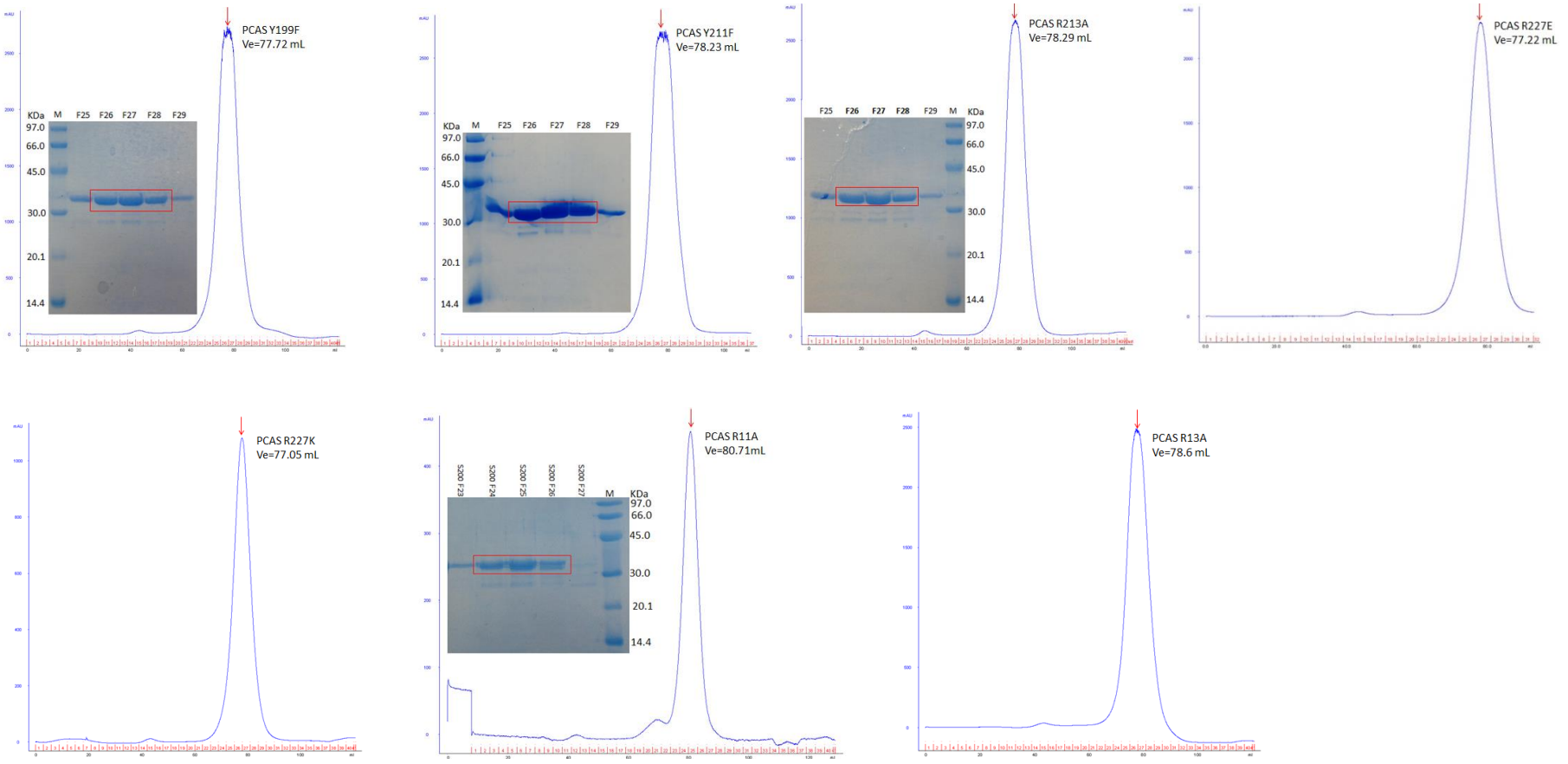


Fig.3.42 UV-vis $A_{280\ nm}$ traces from the AKTA purifier of the superdex 200 size-exclusion chromatography purification and SDS-PAGE gel showing the purification steps of PCAS mutants (Gels for mutants R227E, R227K and R13A are missing). All seven S200 chromatography gave a single asymmetric peak, and the *Ve* of Y199F, Y211F, R213A, R227E, R227K, R11A, R13A are 77.7 mL, 78.2 mL, 78.3 mL, 77.2 mL 77.1 mL, 80.7 mL, 78.6 mL separately.

Mutants	Native *Mr determined* by S200 column (kDa)	Subunit *Mr determined* by SDS-PAGE gel (kDa)	Native *Mr*/Subunit *Mr*
Y199F	52.6 *	~30	1.8 - dimer
Y211F	49.9 *	~30	1.7 - dimer
R213A	49.6 *	~30	1.7 - dimer
R227E	55.4 *	~30	1.8 - dimer
R227K	56.4 *	~30	1.9 - dimer
R11A	48.2 **	~30	1.6 - dimer
R13A	60.8 **	~30	2.0 - dimer

Table 3.6 shows the native *Mr* of seven PCAS mutants determined by S200 column (the results labelled with * were calculated using a calibration curve shown in **Appendix 5.9** and ** used the one shown in **Appendix 5.8**), compared to the subunit *Mr* determined by SDS-PAGE gel (**Fig. 3.42**), all the mutants were found to be a homodimer.

Mutants	Predicted molecular weight (Da) using ExPASy-ProtParam tool	LC-ESI-MS results (Da)	Formula
Y199F	29612.3	29611.5	$C_{1304}H_{2025}N_{367}O_{399}S_{12}$
Y211F	29612.3	29611.4	$C_{1304}H_{2025}N_{367}O_{399}S_{12}$
R213A	29543.2	29542.5	$C_{1301}H_{2018}N_{364}O_{400}S_{12}$
R227E	29601.2	29599.9	$C_{1303}H_{2020}N_{364}O_{402}S_{12}$
R227K	29600.3	29600.2	$C_{1304}H_{2025}N_{365}O_{400}S_{12}$
R11A	29543.2	29547.2	$C_{1301}H_{2018}N_{364}O_{400}S_{12}$
R13A	29543.2	29546.5	$C_{1301}H_{2018}N_{364}O_{400}S_{12}$

Table 3.7 shows the predicted mass of PCAS mutants using the ExPASy ProtParam tool and the LC-ESI-MS results. All seven LC-ESI-MS results showed an ion envelope, and the most abundant ion was chosen to work out the experimental mass. As expect, all of them are in good agreement with the predicted mass.

3.7 Catalytic activity and substrate specificity of PCAS mutants

The enzyme activity (turnover) was measured using the PPi-based coupled assay with saturating concentrations of the acid, MgATP and CoASH. Mutants R227E and R227K showed reduced PCAS turnover ($k_{cat,}$ s^{-1}) with the natural substrate pimelic acid (DCA7) by around 20 fold (~4% activity remaining) (**Fig. 3.43**). The R213A mutation reduced PCAS activity >99% compared with wild type and was effectively inactive. In contrast, mutant Y199F retained 55% activity and similarly, mutant Y211F retained 36% activity.

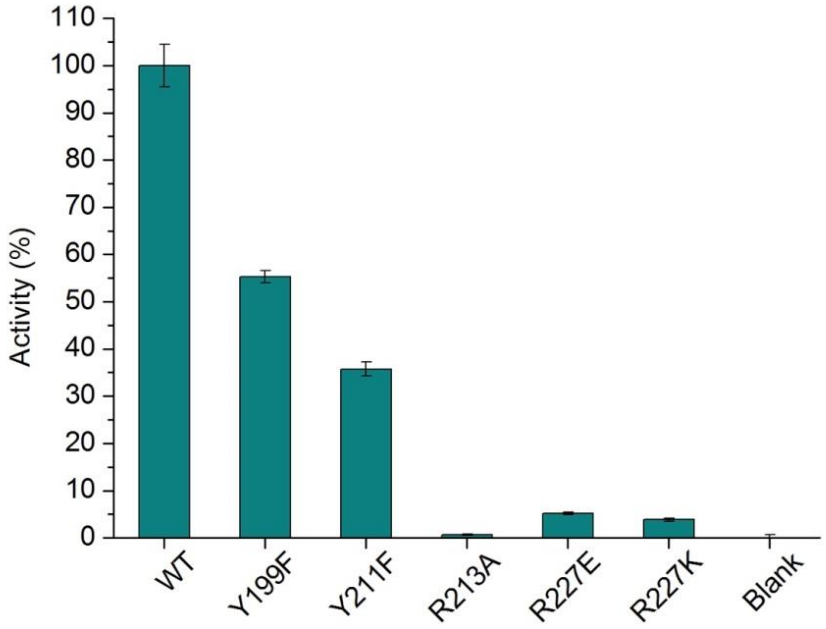


Fig. 3.43 Activity of wild type PCAS and active site mutants. The pyrophosphate (PPi) release assay was used to determine enzyme activity and is based on turnover (k_{cat}, s^{-1}) in the presence of PCAS enzyme (0.1 μM, WT and mutant), pimelic acid (1.5 mM), ATP (1.0 mM), CoASH (1.0 mM). The value for WT PCAS (0.48 s^{-1}) was set at 100% and the PCAS mutant values are shown relative to WT.

The wild type PCAS displays high specificity for pimelic acid (DCA7) with much reduced turnover observed (<1% k_{cat}) with both shorter and longer acyl chain length substrates (DCA5, DCA6, DCA8, DCA9, **Fig. 3.44**). The acceptance by PCAS of slightly longer chain length dicarboxylic acid substrates (up until C9) was first observed by Cronan using a radioactive assay [43]. Both Y199F and Y211F mutants displayed reduced turnover compared to the WT PCAS with shorter chain length substrates DCA5 and DCA6 (**Fig. 3.44**). In contrast, these mutants displayed increased activity relative to wild type PCAS with the longer substrates DCA8 and DCA9 (**Fig. 3.44**).

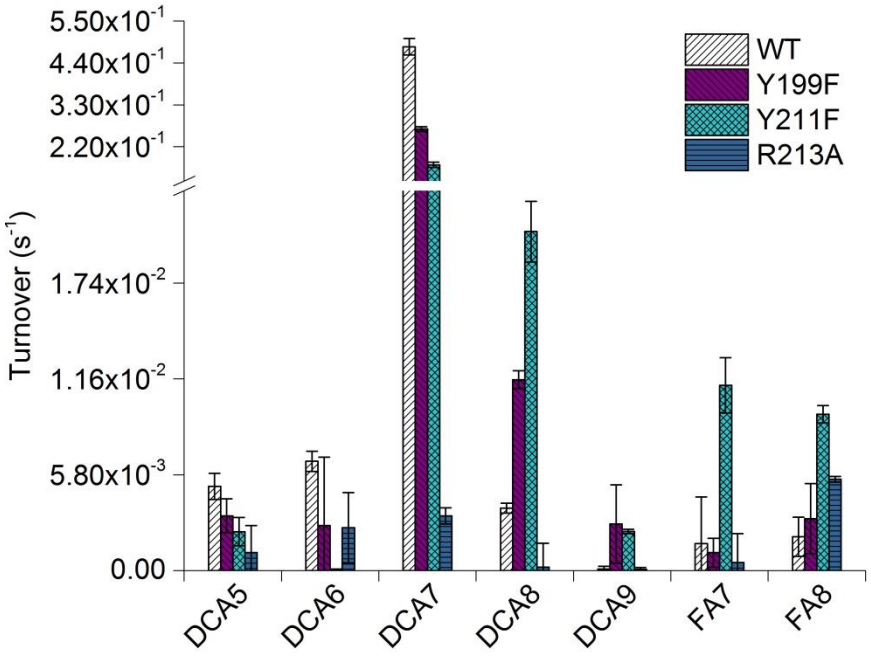


Fig. 3.44 Assay to determine the chain length specificity of PCAS and designed mutants. The turnover (k_{cat}, s^{-1}) of each PCAS WT enzyme and mutants Y199F, Y211F and R213A (0.1 μM) in the presence of different carboxylic acids (dicarboxylic acid, DCA, mono fatty acid, FA, 1.5 mM), ATP (1.0 mM) and CoASH (1.0 mM).

To reduce errors and increase the yield of the reaction we increased the enzyme concentration. Using an enzyme concentration of 3.0 μM the Y199F mutant resulted in a two-fold improved activity with DCA8 but no improvement with DCA9 over wild type (**Fig. 3.45**). In contrast, the Y211F mutant displayed ~4 fold increased activity with the DCA8 substrate and ~3 fold increased activity with the DCA9 substrate relative to the wild-type PCAS (**Fig. 3.46**).

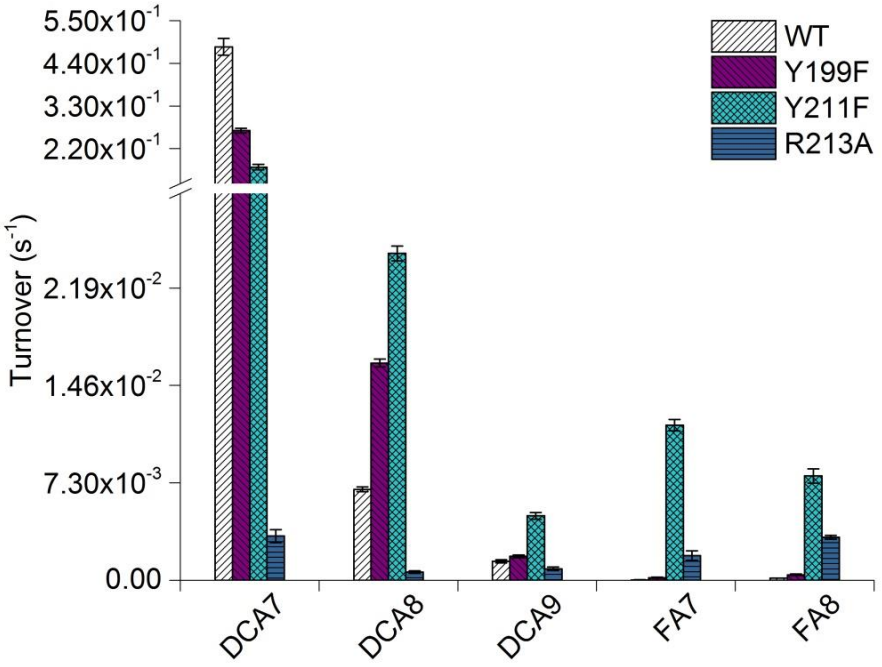


Fig. 3.45 The PCAS Y211F mutant displays broadened fatty acid specificity. The turnover (k_{cat}, s^{-1}) of each PCAS enzyme, WT and mutants Y199F, Y211F and R213A (3.0 μM) in the presence of different carboxylic acids (dicarboxylic acid, DCA, mono fatty acid, FA, 1.5 mM), ATP (1.0 mM) and CoASH (1.0 mM).

To generate sufficient quantities of acyl-CoA product we set up a reaction containing PCAS (either wild type or mutants Y199F and Y211F), DCA8, MgATP and CoASH that was incubated for 5 hours at 30 ^{o}C. The production of the DCA8-CoA (suberoyl-CoA) product was confirmed by C18-RP HPLC analysis (**Fig. 3.45**). The

HPLC analysis of the WT PCAS showed that it used up the ATP and CoASH substrates and produced both AMP and produced a new peak at 17.2 min that was confirmed to be DCA8-CoA (suberoyl-CoA). Based on this peak the two mutants produced slightly more acyl-CoA product which is in agreement with their increased turnover observed using the PPi assay (**Fig. 3.44**) above.

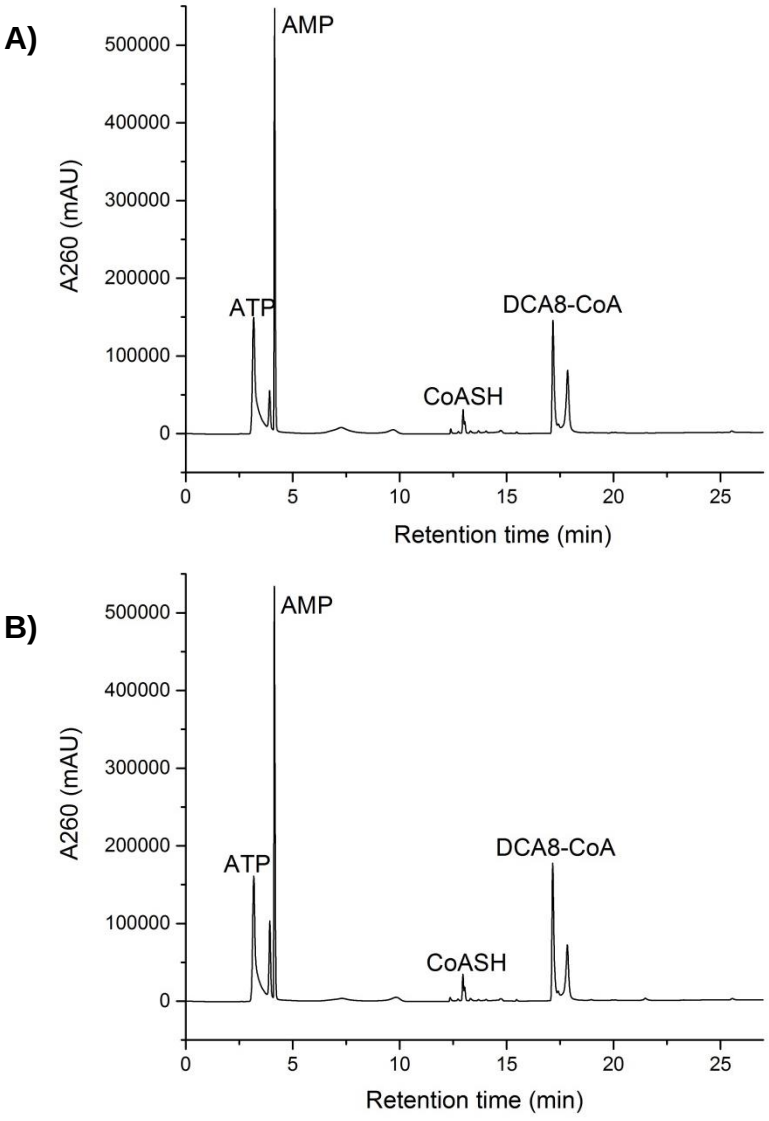


Fig. 3.46 C18 RP-HPLC analysis of **A)** wild type, **B)** Y211F PCAS which were incubated with the DCA8 (suberic acid), ATP and CoASH substrates for 5 hours at 30 ^{o}C before analysis by HPLC. For both figures, the DCA8-CoA (suberoyl-CoA) product elutes at 17.2 min.

The PCAS crystal structure with CoASH bound (albeit without clear density for the phosphopantetheine part) identified the side-chains of Arg11 and Arg13 as potentially involved in binding the phosphate groups of this substrate. To test this we generated mutants at these two sites (R11A and R13A). Both mutants were essentially inactive in the PCAS assay with pimelic acid (**Fig. 3.47**, <1.0 % compared with WT).

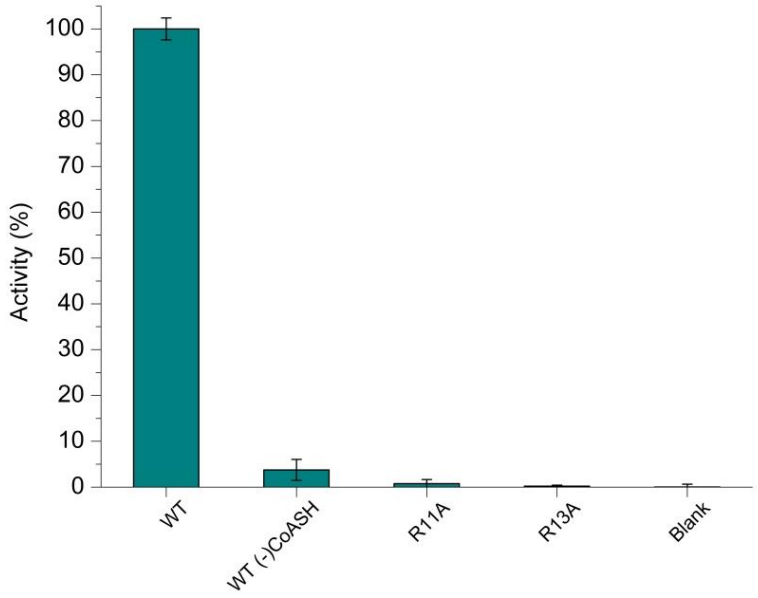


Fig. 3.47 Kinetic analysis of the PCAS R11A and R13A mutants with DCA7 (pimelic acid) as a substrate. The data was obtained using the PPi assay and the WT activity is included for comparison.

3.8 An engineered PCAS displays novel activity

The wild-type PCAS enzyme gave no HPLC detectable product when incubated with the mono fatty acid substrates FA7 and FA8 at both low (0.1 μM) and high (3.0 μM) enzyme concentrations respectively (**Fig. 3.44** and **Fig. 3.45**).

Similarly, the Y199F mutant was also inactive with FA7 and FA8 (**Fig. 3.45**). In contrast, the Y211F displayed activity with both monocarboxylic acid substrates at 0.1 μM concentration (**Fig. 3.44**) and at the higher concentration (3.0 μM) the Y211F

mutant displayed a k_{cat} of 0.012 s^{-1} and 0.008 s^{-1} with FA7 and FA8 respectively (**Fig. 3.45**).

Indeed, turnover was such for this PCAS Y211F mutant that we could determine the K_M for FA7 (529.0 ± 79.7 μM) and ATP (479.4 ± 81.5 μM) using the PPi coupled assay (**Fig. 3.48** and **Table. 3.8**).

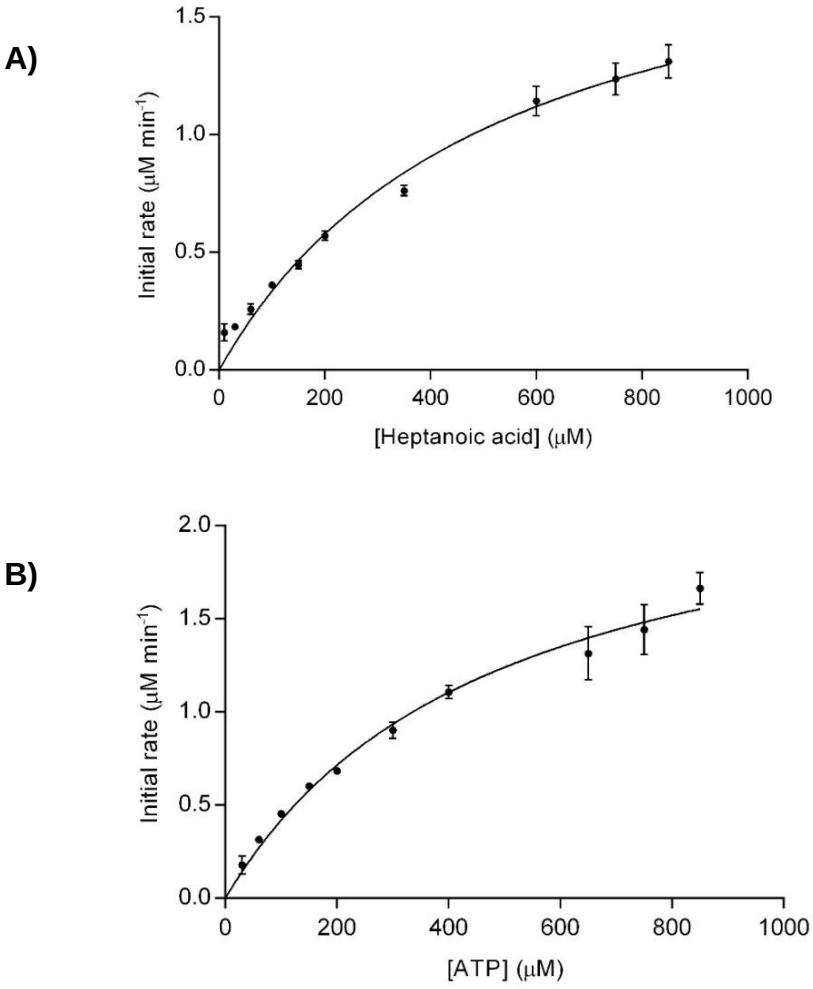


Fig. 3.48 Michaelis-Menten plot of the kinetic study of Y211F PCAS for substrate heptanoic acid (**A**) and ATP (**B**) using the pyrophosphate production assay.

	Y211F		WT
Parameter	**Heptanoic acid**	**ATP**	**Heptanoic acid**
K_M (μM)	529.0 ± 79.7	479.4 ± 81.5	
k_{cat} (s^{-1})	0.0140 ± 0.0011	0.0162 ± 0.0013	0.0016
k_{cat}/K_M (M^{-1} s^{-1})	26.2	33.8	

Table. 3.8 Kinetic parameters of Y211F PCAS for substrate heptanoic acid and ATP using the pyrophosphate production assay.

We set up a "synthetic reaction" by incubation of the Y211F enzyme (5.0 μM) with saturating concentrations of heptanoic acid (FA7), MgATP and CoASH for 5 hours at 30 ^{o}C and produced the FA7-thioester which was purified by HLPC (**Fig. 3.49 B**). For comparison, the wild-type type PCAS produced no detectable product under the same conditions (**Fig. 3.49 A**). It is worth noting that despite producing no detectable product with pimelic acid, we found that the PCAS R213A mutant did give rise to detectable amounts of the FA7 and FA8 products (**Fig. 3.45**).

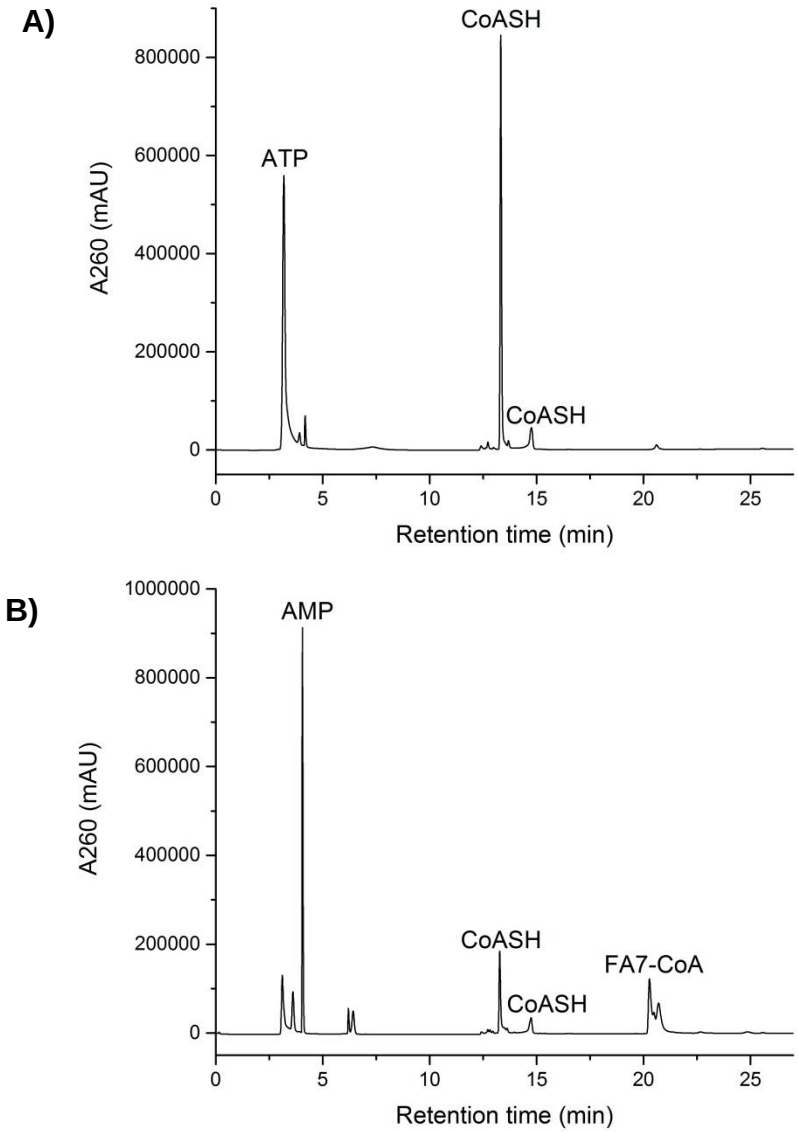


Fig. 3.49 Synthesis of hepatonyl-CoA (FA7-CoA) by the engineered PCAS Y211F. **A)** WT PCAS and **B)** Y211F mutant was incubated with the FA7, ATP and CoASH substrates for 5 hours at 30 ^{o}C before analysis by HPLC.

B) For the Y211F mutant catalysed reaction the ATP (3.2 min) and CoASH (13.3 min and 14.8 min) peaks decrease along with an increasing AMP (4.2 min) peak. As well as these changes, the FA7-CoA heptanoyl-CoA product elutes at 20.3 min., in contrast to **A)** where detectable product was observed using the WT PCAS.

3.9 Discussion

In *Bacillus subtilis*, pimeloyl-CoA synthetase (PCAS, EC 6.2.1.14, UNIPROT code: P53559, 29 kDa) is the first enzyme in the biotin biosynthetic pathway and acts as a highly specific substrate selection gate ensuring the integrity of the carbon chain in biotin synthesis. The PCAS reaction is highly specific to pimelic acid, mono acids and different chain lengths are either not processed at all or are multiple orders of magnitude worse substrates [43]. Whilst this ensures that only the correct vitamin is made *in vivo* rather than potentially toxic analogues, it represents an unfortunate limitation. The ability to generate acyl-CoA thioesters from simple short chain fatty acid precursors through biocatalytic methods would be extremely valuable (The 2016 price list is shown in **Table 3.9**) since they are used as substrates in studies of important metabolic enzymes such as those involved in fatty acid, polyketide, sphingolipid and lipid A/lipopolysaccharide (LPS) biosynthesis [95-98].

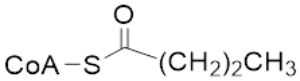
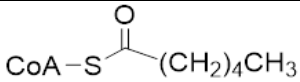
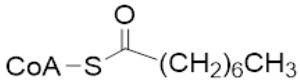
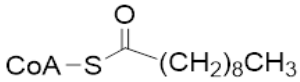
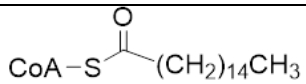
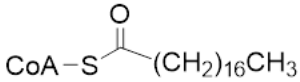
Acyl-CoA thioesters	Chemical structures	Weight	Prices on Sigma
Butyryl coenzyme A		5 mg	£118.0
Hexanoyl coenzyme A		5 mg	£108.5
Octanoyl coenzyme A		5 mg	£114.5
Decanoyl coenzyme A		5 mg	£110.0
Palmitoyl coenzyme A		5 mg	£124.5
Stearoyl coenzyme A		5 mg	£102.5

Table 3.9 The price list of varying carbon chain length of acyl-CoA thioesters from Sigma

For the purposes of enzyme design/evolution, if the proposed adenylate forming mechanism is correct, elongation/changing of the enzyme substrate should be achieved by expanding/contracting the substrate binding pocket. In this study we have chosen to work with the PCAS from *Bacillus subtilis* due to its small size comparing to other acyl-CoA/ACP ligases (as shown in **Table 3.10**, 259 aa, predicted molecular weight 29,630 Da). Meanwhile, most naturally occurring fatty acids have an even number of carbon atoms, but since PCAS uses odd-numbered carbon substrates, this makes PCAS quite unique.

Enzyme	Organism	Number of amino acid
Pimeloyl-CoA synthetase (PCAS)	*Bacillus subtilis*	259 aa
Acyl-acyl carrier protein synthetase (AasS)	*Vibrio harveyi*	554 aa
Long-chain-fatty-acid-CoA ligase	*Thermus thermophilus*	541 aa
Acetyl-coenzyme A synthetase	*Salmonella typhimurium*	652 aa

Table 3.10 The sizes of different acyl-CoA/ACP ligases

Up until our project there had been surprisingly few detailed studies of any PCAS enzyme. The only published report is the early work from 1992 by Ploux et al on the PCAS enzyme from *Bacillus sphaericus* [40]. However, during the early stages of this project, Manandhar and Cronan [43] published a paper reporting the characterisation of a *BioW* gene product encoding the PCAS from the *Bacillus subtilis* strain 168 which is 100% identical to the *B. subtilis* PCAS used here. They had reported that attachment of a His_6 tag at either end of recombinant *B. subtilis* PCAS destroyed its ability to function in biotin complementation assays in *E. coli*. Therefore they carried out their studies on an un-tagged recombinant form of the *B. subtilis* PCAS. Incubation of purified PCAS with [α-^{32}P]-ATP provided evidence for an adenylate intermediate whilst HPLC and mass spectrometry established formation of the pimeloyl-CoA

product. In contrast to Manandhar and Cronan, we found that a His_6 tag positioned at the N-terminus of *B. subtilis* PCAS allowed the isolation of milligram amounts of active and pure PCAS. We used HPLC to show that the isolated His_6 tag form of the enzyme was able to convert pimelic acid, MgATP and CoASH to pimeloyl-CoA and AMP (**Fig. 3.21**).

As discussed in the introduction, the generation of the key acyl-pimeloyl thioester intermediate varies between microorganisms. In *B. subtilis* PCAS acts as the dedicated entry point to the biotin pathway. The PCAS enzyme must be highly specific for the odd chain, di-acid substrate pimelic acid or else a range of biotin derivatives with altered chain lengths would be observed in PCAS-containing species. To investigate the origin of this specificity our kinetic data on the isolated *B. subtilis* PCAS shows that the enzyme turns over pimelic acid (DCA7) at least 100 times faster than either the C6 (DCA6) or C8 (DCA8) di-acids. To provide molecular insight into this kinetic observation we determined the structure of PCAS and a series of co-complex crystal structures. The solution of the X-ray structure was particularly challenging because there were no structures in the Protein Data Bank (PDB) of high enough sequence homology to use as a template to solve the structure by molecular replacement. To include a heavy atom for phasing we prepared selenomethione-labelled PCAS but unfortunately this failed to provide diffraction-quality crystals. In the end, we resorted to Pt soaks and these derivatives gave crystals that diffracted and allowed a structure solution by my collaborators Prof. Jim Naismith and Dr. Lucile Moyni é at the University of St. Andrews. The *B. subtilis* PCAS has a novel fold for an adenylase, with two domains. The smaller N-terminal domain which forms the dimer interface is responsible for binding the adenosine and three phosphates of CoASH. The larger C-terminal domain binds both the ATP and pimelic acid. A combination of structures

allowed us to visualise the pentavalent phosphorus transition state and identify the key residues and ions that stabilise it. The constellation of residues that surrounds the transition state is different to that seen before in other enzymes that generate an adenylate intermediate, but the chemical basis is unsurprisingly similar; positive charges. This is a recurring theme in enzymes which break or make phosphoester bonds and a striking example of convergent evolution controlled by the requirement to stabilise a particular transition state [45].

Our structures reveal that fatty acid specificity is controlled by a cluster of residues that bind to and fix the position of the ω- carboxylate of pimelic acid whilst the carbon chain sits in a hydrophobic tunnel. Only when the intervening seven carbon atoms adopt an extended conformation will the other (α) carboxylate group be appropriately positioned to attack the α-phosphate of ATP required to generate the pimeloyl-adenylate intermediate. Since the tunnel is restricted in volume and the ω position fixed, additional carbons will result in clashes with ATP and shorter chains will not reach it. We identify this as the origin of the chain length selectivity of PCAS.

The enzyme proceeds in two distinct chemical steps with the first being the formation of the pimeloyl-adenylate. Superposition of the ligand-bound PCAS structures with AMP-PNP and pimeloyl-adenylate reveals that the geometry of the incoming nucleophilic oxygen atom of the carboxylate and the phosphate is ideal for addition to form the pentavalent phosphorus transition state. The side chains of residues Arg227 and Arg170 will stabilise the negative charge on the carboxylate activating it as a nucleophile. In the AMP-PNP analog structure, we identify two Mg^{2+} ions that will bridge the α and γ phosphates of ATP and in doing so each Mg^{2+} binds to a terminal oxygen atom of the α phosphate thereby stabilising the negative charge on the

transition state. In the PCAS:pimeloyl-adenylate:PP_i complex the two Mg^{2+} ions bind to and, we presume, stabilise the negatively charged PP_i leaving group.

The biochemical data show that only tiny amount of PPi realise (and thus activity in our coupled assay) in the absence of CoASH (**Fig. 3.47**). This may due to the fact that in the absence of CoASH, all PCAS is captured by the adenylates intermediate thus there is no more enzyme to turn over the reaction. And this suggestion is confirmed by native mass spectrometry as well (**Fig. 3.49c**). Adenylates are usually (and in fact have been selected by evolution to be) highly reactive, in solution such a species would be expected to react almost instantly. This was exemplified in studies of the enzyme biotin protein ligase (BPL) that generates the biotin-adenylate intermediate that is used to transfer biotin on to the biotin carboxy carrier proteins (BCCPs) of carboxylases. Mutation of key, conserved active site arginine residues within BPL did not abolish activity, rather released the adenylated biotin into solution where it modified lysine residues on the protein surface in a promiscuous manner [16, 99]. The PCAS assay does have a detectable background rate without CoASH suggesting water can hydrolyse the adenylate but at a much slower rate than in free solution. The structure reveals that the adenylate although bound in the enzyme is not shielded from solvent. Water molecules are positioned close to the carboxylate group of the adenylate in the crystal structure. We suggest that the positioning and hydrogen bonding of the water molecules means they cannot rapidly hydrolyse the adenylate and this accounts for the very unusual stability of the adenylate intermediate.

We have been unable to obtain a co-complex with CoASH but site directed mutagenesis and electron density in a soaked crystals suggest a binding site. The structure possesses a tunnel from the surface to the catalytic site that is consistent with the pantothenate arm reaching in to attack the adenylate. The thiol group must displace

at least one of the water molecules at the adenylate. The much higher nucleophilicity of the thiol compared to water means that it will be sufficiently reactive to decompose the adenylate even in the presence of Arg227. It has been proposed that adenylating enzymes of ANL family undergo large conformational changes during their catalytic cycle, the C-terminal domain rotates in relation to the N-terminus in the presence of CoASH [44, 48]. We did not observe any such large conformational changes in the class IV PCAS but we cannot exclude such a mechanism. Taking the biochemical and structural data together we can propose a chemical mechanism for the PCAS-catalysed reaction (**Fig. 3.50**).

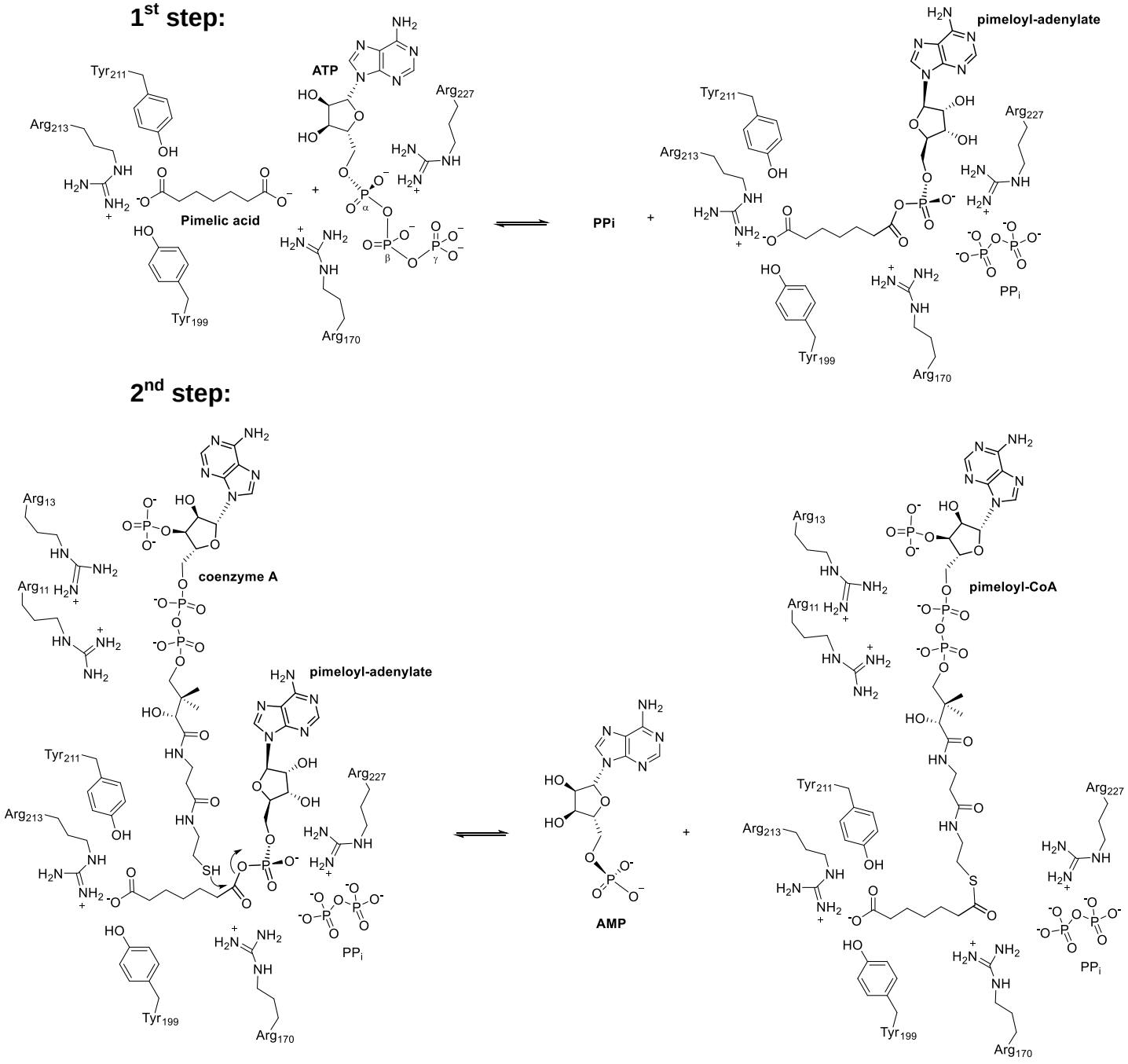


Fig. 3.50 Schematic of the PCAS active site and proposed mechanism for the forming of the adenylate intermediate (1st step) and CoASH attack on pimeloyl-adenylate (2nd step). The key residues in the C-terminal domain involved in pimelic acid recognition (Tyr199, Tyr211 and Arg213) are highlighted, as are those involved in pimeloyl-adenylate binding (Arg170, Arg227). The N-terminal residues involved in CoASH binding (Arg11 and Arg13) are included, as is the pyrophosphate (PPi) derived from ATP. We propose formation of the pimeloyl-CoA thioester product by attack of the adenylate intermediate by the thiol of the incoming coenzyme A.

Thioesters are high value substrates used as important building blocks for organic synthesis and chemical biology for example in native chemical ligation [100]. Specific coenzyme A thioesters are valuable substrates for enzyme assays and mechanistic probes [101]. They are chemically synthesised in an analogous way to the enzymatic process; typically by acid activation (e.g. as acid anhydrides or acid chlorides), then reaction with an appropriate thiol. The activated acids are often labile under aqueous conditions lowering isolated yields. The di-acids have to be protected at one carboxylate to direct the synthesis of di-acid thioester targets. More recently methods such as direct iron-catalysed coupling of aldehydes with sulfur surrogates [102] and methyl acyl phosphates in water have been developed [103]. However, as an alternative to these chemical methods, a "green" biocatalytic synthesis of simple fatty acid thioesters would be highly desirable and the relatively simple, catalytic domain of PCAS provided an ideal template to engineer an enzyme with modified substrate specificity. We reasoned that the chain length specificity displayed by PCAS is set by the residues at the ends of the tunnel that interact directly with the two carboxylic acid groups. The C7 end is defined by the arginine and two tyrosine (Y199 and Y211) residues so we made mutants that trimmed back H-bond donors and side chains that mediate electrostatic interactions. We removed the tyrosine hydroxyl at Tyr211 by mutation to phenylalanine and found that the enzyme retained activity with pimelic acid. However, as we predicted, the PCAS Y211F mutant gained the ability to synthesise heptanoyl-CoA from (mono-carboxylic) heptanoic acid, MgATP and CoASH converting about 50 % of the substrate. We saw no evidence that the adenylate intermediate is released into solution (hydrolysis in the absence of CoASH or heptanoic acid modification of protein).

Wild-type PCAS displays exquisite substrate specificity but our structural and engineering results have shown that single active site residues can be changed such that the mutant enzyme retains activity and can be used to prepare non-cognate acyl-thioesters. PCAS is a relatively small enzyme, with a few residues defining the active site, so it offers great scope to expand the range of thioesters that can be synthesised. We identified Arg213 in the active site and mutagenesis of this residue (R213A) inactivated the enzyme with pimelic acid as a substrate. However, this mutant displayed activity with C8 monocarboxlic acids and suggests that this site can be further exploited and engineered to accept longer chains. Saturation mutagenesis at specific sites, combined with the application of the tools of directed evolution and a high-throughput assay for PCAS would allow its full synthetic potential to be explored [104].

After my viva, Manandhar and Cronan [43] published a second paper reporting the crystal structure of PCAS from *Aquifex aeolicus* which is 39.3 % identical to the *B. subtilis* PCAS used here. The enzyme activity parameters they observed with HPLC assay are comparable with our results which were determined by pyrophosphate production assay and A_{230nm} UV-vis assay (0.75 s^{-1} for pimelic acid compared with 0.63 s^{-1} and 0.59 s^{-1}) (**Table 3.11**).

PCAS	***B. subtilis***	***B. subtilis***	***B. sphaericus***	***B. subtilis***	***A. aeolicus***
Project	This project	This project	Ploux's paper [40]	Cronan's 2013 paper [43]	Cronan's 2017 paper
Assay	Pyrophosphate production assay	A_{230nm} UV-vis assay	A_{230nm} UV-vis assay	TLC assay	HPLC assay
K_M (μM) for pimelic acid	70.5 ± 6.8	127.3 ± 13.8	145	19.8 ± 0.6	10.7 ± 0.96
k_{cat} (s^{-1})	0.63 ± 0.02	0.59	0.4	8.4×10^{-5}	0.75 ± 0.002
k_{cat} / K_M (M^{-1} s^{-1})	8,956.20	4635.45	2,758.62	4.24	70,093.46

Table 3.11 Comparing of the kinetic parameters of PCAS.

They determined the structures of *A. aeolicus* PCAS in complex with pimelate/AMP-PCP and CoASH/AMP at 2.55 Å and 2.25 Å resolution respectively. The same with our *B. subtilis* PCAS, the asymmetric unit contains two monomers. And the overall fold of the monomer is similar with our *B. subtilis* PCAS (**Fig. 3.51a**), which consists of two domains, a small 55 residues N-terminal domain that contains three anti-parallel β-sheet with a single α-helix and a larger C-terminal domain of around 200 residues (**Fig. 3.51b**).

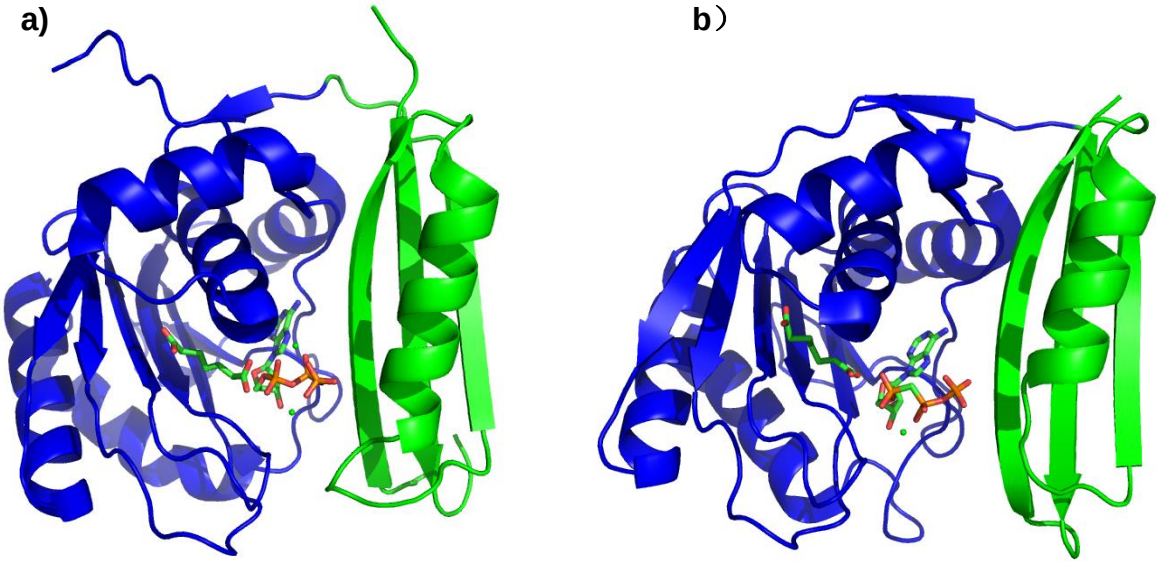


Fig. 3.51 Overall X-ray crystal structure of the PCAS monomer. **(a)** *B. subtilis* PCAS (PDB code: 5flg) **(b)** *A. aeolicus* PCAS (PDB code: 5TV8). The N-terminal domain is in green and C-terminal domain in blue.

As shown in **Fig. 3.52**, they reported that Tyr187, Tyr199 and Arg201 (which corresponding to Tyr199, Tyr211 and Arg213 in our *B. subtilis* PCAS, **Fig. 3.53**) engage the ω carboxylate of the substrate pimelic acid, and the α carboxylate interacts with Arg159, Ser182, and Arg215 (corresponding to Arg170, Ser194, and Arg227 in *B. subtilis* PCAS, **Fig. 3.53**). They believe that Arg159 (Arg170 in *B. subtilis* PCAS) serves as a "ruler" to measure the correct length of substrate (C7- pimelic acid), thus

the misalignment of Arg159 upon formation of the non-cognate adenylates intermediate could trigger the hydrolytic cleavage. And they reported that this proofreading activity can be abolished by a single active site mutation (R159A).

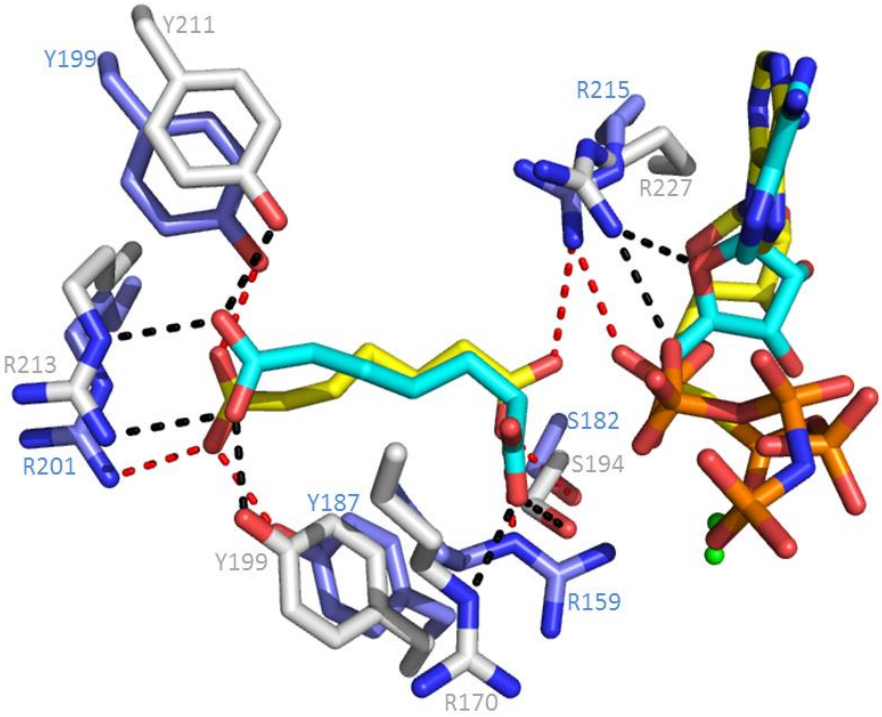


Fig. 3.52 Superposition of the active sites of *B. subtilis* PCAS (PDB code: 5flg) and *A. aeolicus* PCAS (PDB code: 5TV8). For clarity, the protein has been removed. Carbon atoms of active sites of *B. subtilis* PCAS are coloured in grey and the substrate (AMP-PNP and pimelic acid) in cyan. Carbon atoms of *A. aeolicus* PCAS are in light blue and the substrate (AMP-PCP and pimelic acid) in yellow. Hydrogen bonds are shown as black dashes lines (*B. subtilis* PCAS) and red dashes lines (*A. aeolicus* PCAS). Oxygen atoms are in red, nitrogen in blue, phosphate in orange and Mg^{2+} ions are shown as green spheres.

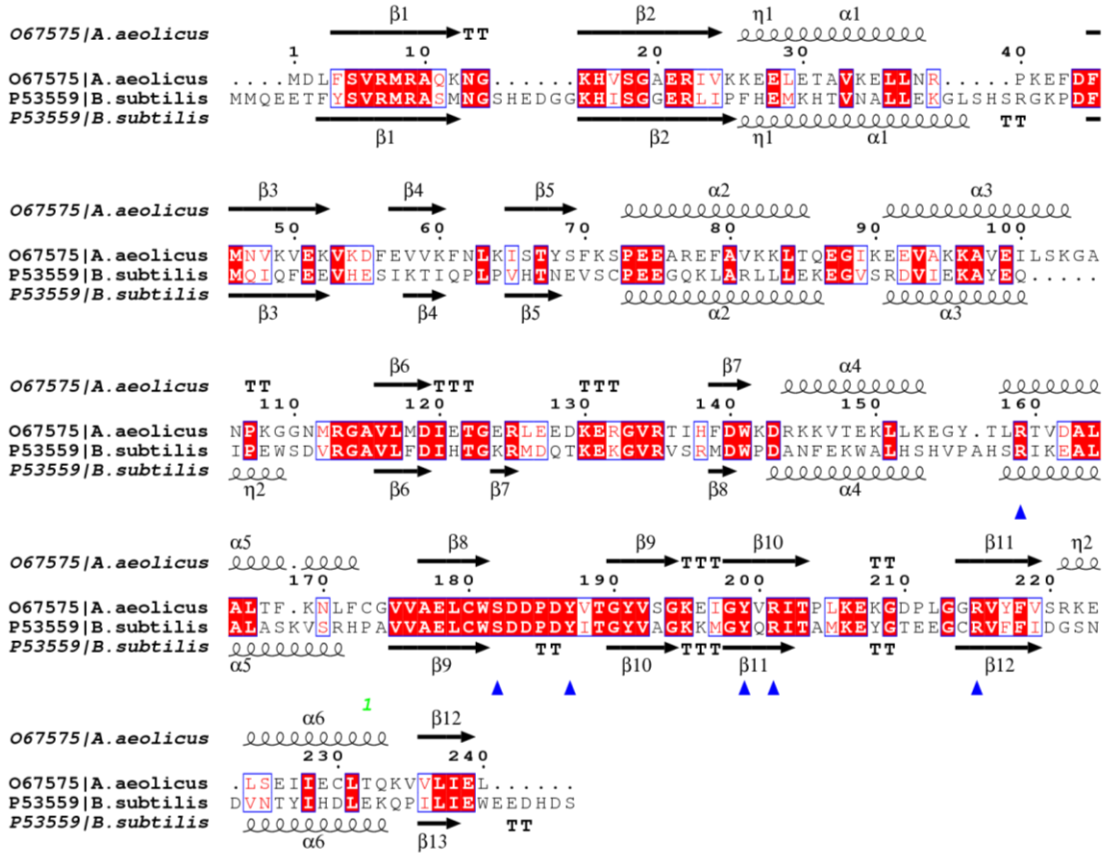


Fig. 3.53 Sequence alignment of *Aquifex aeolicus* (O67575, BIOW_AQUAE) with*Bacillus subtilis* (UNIPROT code: P53559, BIOW_BACSU). The residues Tyr187, Tyr199, Arg201, Arg159, Ser182, and Arg215 (corresponding to Tyr199, Tyr211 and Arg213, Arg170, Ser194, and Arg227 in *B. subtilis* PCAS) are marked with blue triangles. Figure produced using ESPript 3 [94].

Chapter IV. Studies of *E*scherichia *coli* 8-amino-7-oxononanoate synthase (AONS) and a *Corynebacterium amycolatum SK46* BioWF fusion

4.1 Characterisation of *E*scherichia *coli* 8-amino-7-oxononanoate synthase (AONS)

4.1.1 Cloning of *E*scherichia *coli* AONS

The PLP-dependent *E. coli* 8-amino-7-oxononanoate synthase (AONS, Uniprot code: P12998) had been studied previously in Edinburgh by Baxter, Campopiano and colleagues. This enzyme catalyses the condensation of L-alanine with pimeloyl-CoA to generate the biotin intermediate AON. In a series of papers they reported the results of a combination of X-ray crystallography, UV-vis spectroscopy, kinetics and site-directed mutagenesis to study PLP-bound intermediates. The 3D structure (PDB code: 1BS0) of the holo-AONS was the first of any member of the AOS family and a catalytic mechanism was proposed which could also be applied to others members of the family. The recombinant *E. coli* AONS used in these studies was expressed using a pET-derived plasmid in its native form i.e. without a His_6-tag at either the N- or C-termini. Milligram quantities of AONS were isolated in three steps including ammonium sulphate precipitation, phenyl-Sepharose and Q-Sepharose anion exchange columns [21, 22]. Affinity chromatography is now used routinely for quick protein

purification, so in this project, the full-length *BioF* gene was cloned into two expression plasmids (pET28a and pET22b) to obtain recombinant proteins being tagged at the N-terminus (pET28a) or C-terminus (pET22b) with a His_6 tag.

4.1.1.1 Cloning of AONS/NHis

The full-length *BioF* gene from plasmid AONS/pET16b, which was generated by the previous PhD student in our lab to obtain un-tagged AONS, was cloned into pET28a vector via NdeI and BamHI sites (**Fig. 4.1**). This gives a translated AONS protein with an N-terminal His_6 tag. The DNA and amino acid sequence of AONS/NHis are shown in **Appendix 5.5**.

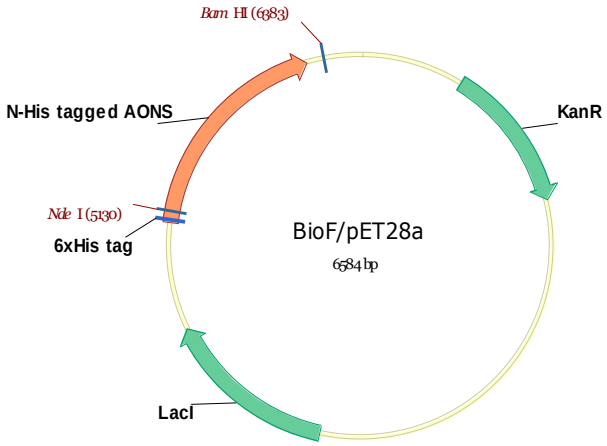


Fig. 4.1 Plasmid map of *E. coli* AONS/NHis in pET28a vector

4.1.1.2 Cloning of AONS/CHis

The full-length *BioF* gene was amplified from the AONS/pET28a plasmid using a forward primer containing NdeI site and a reverse primer containing XhoI site without stop codon, and then cloned into the pET22b vector via NdeI and XhoI restriction sites to obtain a C-terminal His_6 tagged protein (**Fig. 4.2**). The DNA and amino acid sequence of AONS/CHis are shown in **Appendix 5.6**.

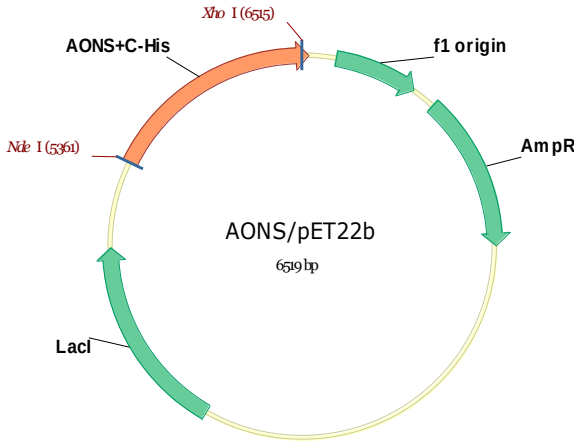


Fig. 4.2 Plasmid map of *E. coli* AONS/CHis in pET22b vector

4.1.2 Isolation of *Escherichia coli* AONS

4.1.2.1 Isolation of AONS/NHis

In the nickel resin purification step, SDS-PAGE analysis showed that most of AONS/NHis flowed directly through the column and only little enzyme was obtained in the elution fractions. This lack of binding to the resin may be because the N-terminal His_6 tag was buried inside the protein and is thus not accessible to the Ni resin.

4.1.2.2 Isolation of AONS/CHis

The C-terminal His_6 tagged AONS was expressed in BL21 (DE3) *E. coli* strain containing plasmid AONS/pET22a after induction with 0.1 mM IPTG and incubation at 30 ^{o}C for 4 hours and isolated from cell free extract (CFE) by incubation with Ni resin. This form of the enzyme did bind and was eluted from the resin with 300 mM imidazole. Analysis of the fractions by SDS-PAGE showed a a strong band at ~42 kDa which corresponds to the AONS monomer (**Fig. 4.4)**. Further purification by Superdex 200 size-exclusion column gave a peak at 72.2 mL corresponding to the theoretical mass of 93,438.1 Da (**Fig. 4.3**, the calculation was shown in **Appendix 5.9**). Compared

to subunit *Mr* of ~42 kDa determined by SDS-PAGE gel, AONS was found to be a homodimer. This method typically yielded ~21.4 mg AONS per litre of LB media.

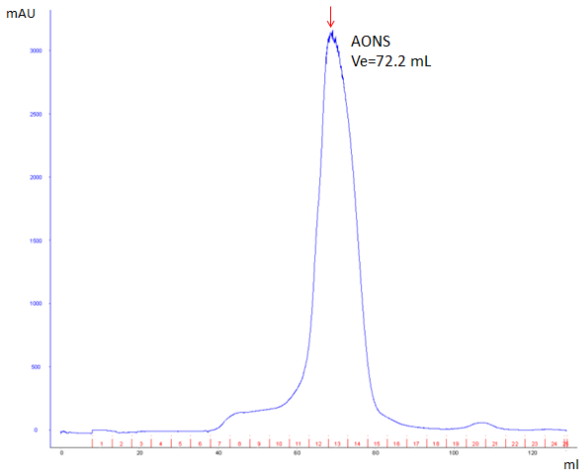


Fig. 4.3 AKTA trace of S200 column (A_{280nm}) of *E. coli* AONS which gives a peak at 72.2 mL. Based on the calibration curve shown in **Appendix 5.9**, it corresponds to the theoretical mass of ~93 kDa, compared to subunit *Mr* of ~42 kDa determined by SDS-PAGE gel, AONS was found to be a homodimer.

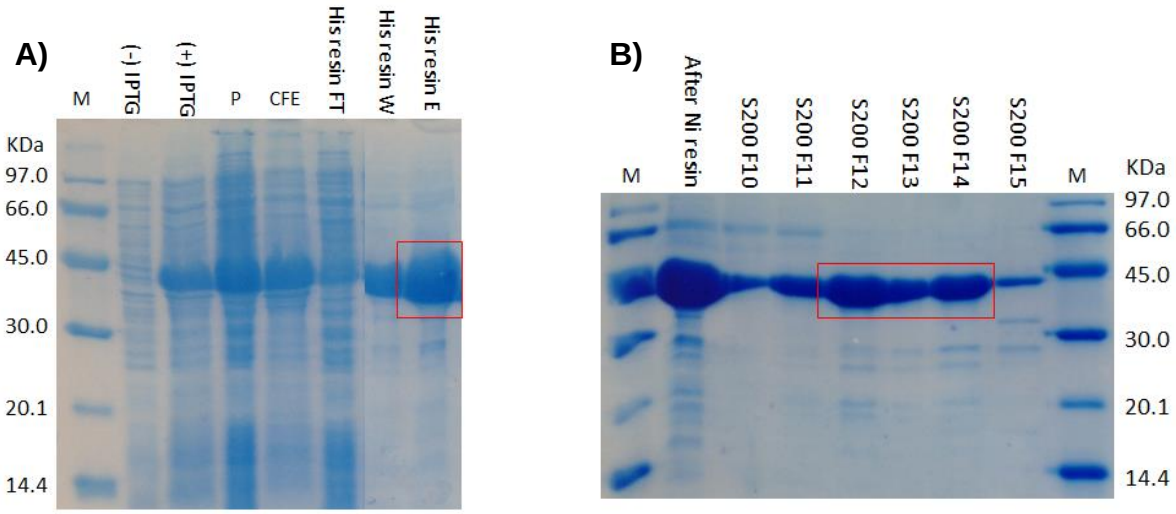


Fig. 4.4 SDS-PAGE gel showing the purification steps of AONS including **A)** protein expression, cell lysis and the Ni resin **B)** S200 gel filtration column. They both showed a band at ~42 kDa which corresponding to the AONS monomer.

4.1.3 L-Alanine binding to AONS

The purified AONS/CHis was dialysed against fresh PLP and excess PLP removed by a spin filter. The UV-vis spectrum (**Fig. 4.6 A**) showed the absorbance maxima at 330 and 425 nm typical of a PLP enzyme and matched well with the un-tagged AONS reported by Baxter and colleagues [22]. This suggested that the His_6 tag did not impact on the ability of the enzyme to bind PLP. Furthermore, the binding of the L-alanine substrate to the recombinant AONS/CHis was measured by titration of L-alanine and monitoring of the changes at the two wavelengths. The UV-vis spectrum gives insight into two internal aldimine forms of the bound PLP cofactor at equilibrium. Titration of L-alanine shifts the equilibrium in favour of external aldimine forms (**Fig. 4.5**).

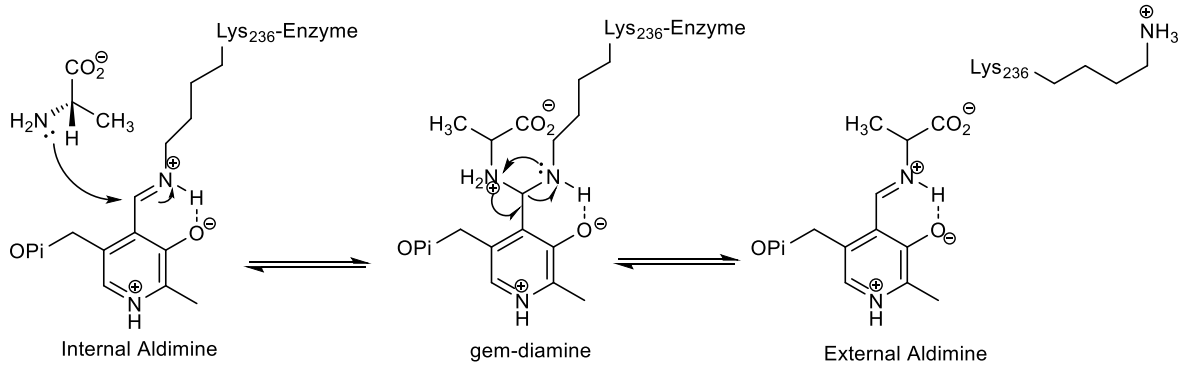


Fig. 4.5 PLP binging in *E. coli* AONS. The first step is binding of L-Ala via a transaldimination reaction to form a germinal diamine, then secondly displacement of the lysine residue of AONS to form the external aldimine.

By plotting the concentration-dependent changes at 425 nm we determined a K_d for L-alanine of 0.71 ± 0.05 mM which matches well with that determined previously by Baxter and colleagues (0.9 ± 0.1 mM) [22].

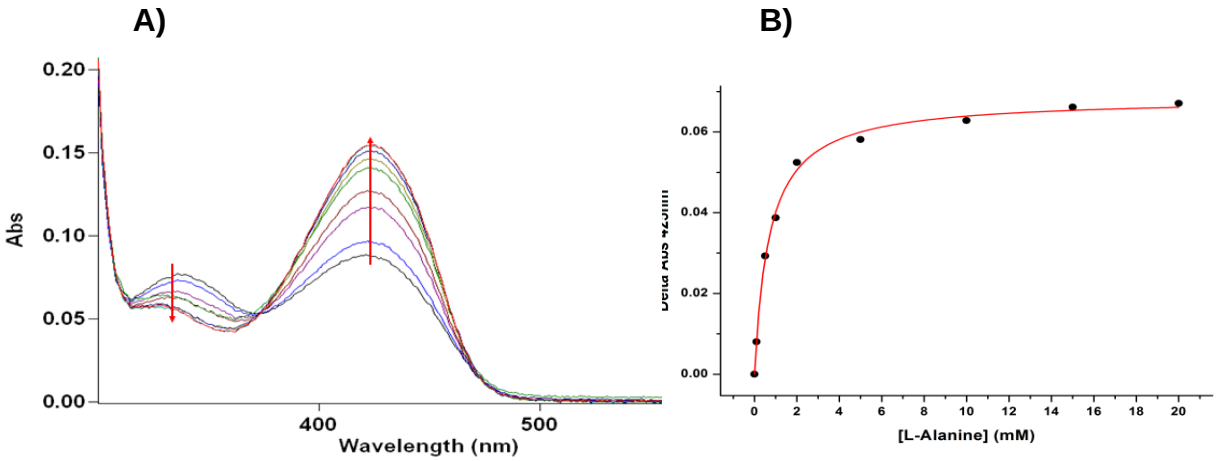


Fig. 4.6 Spectroscopic data of the K_d of AONS/CHis for amino acid substrate L -Alanine, with arrows indicating peak increase or decrease. **A)** UV-vis spectrum for AONS/CHis with L-Ala, **B)** Plot of L-Ala concentration against the absorbance at 425 nm.

4.2 Studies of the *Corynebacterium amycolatum SK46* BioWF fusion

The amino acid sequence alignment of the N-terminal part and the C-terminal part of *Corynebacterium amycolatum SK46* BioWF fusion with *B. subtilis* PCAS (**Fig. 4.7**) and *E. coli* AONS (**Fig. 4.8**) shows high conservation of the key catalytic residues in the active site separately, this suggests that a common protein function and reaction mechanism.

Fig. 4.7 Amino acid sequence alignment of *C. amycolatum SK46* BioWF and *B. subtilis* PCAS (UNIPROT code: P53559). The black triangle highlights the highly conserved residues (Tyr199, Tyr211, Arg213, Arg227) of PCAS. Figure produced using ESPript 3 [94].

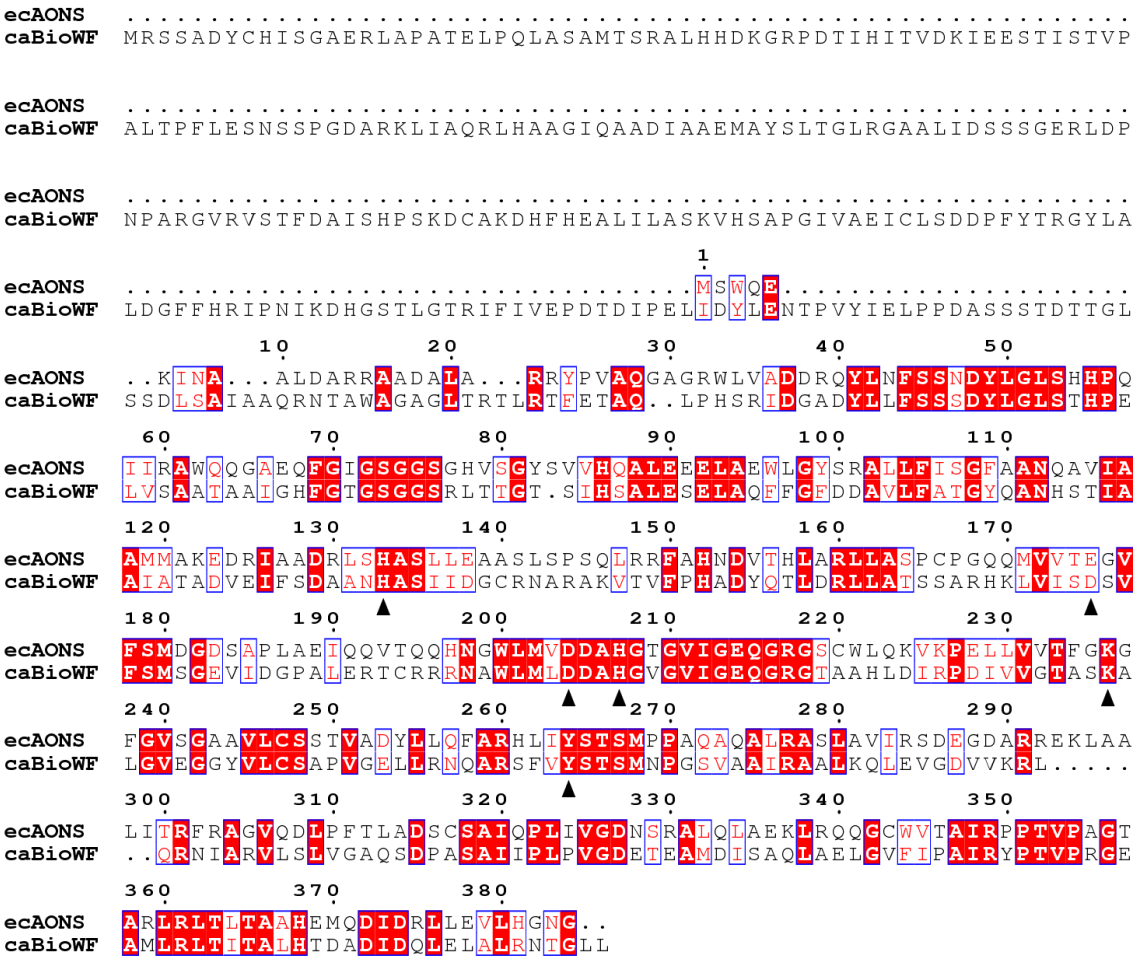


Fig. 4.8 Amino acid sequence alignment of *C. amycolatum SK46* BioWF and *E. coli* AONS (Uniprot code: P12998). The black triangle highlights the active sites (His133, Glu175, Asp204, His207, Lys236, Tyr264) of AONS [39]. Figure produced using ESPript 3 [94].

The *C. amycolatum* BioWF is predicted to be composed of two parts as shown in **Fig. 4.9** - the N-terminal PCAS domain is in green colour and the C-terminal AONS domain is in blue colour, which are linked together by nine amino acids shown in black.

```
  1 MRSSADYCHI SGAERLAPAT ELPQLASAMT SRALHHDKGR PDTIHITVDK
 51 IEESTISTVP ALTPFLESNS SPGDARKLIA QRLHAAGIQA ADIAAEMAYS
101 LTGLRGAALI DSSSGERLDP NPARGVRVST FDAISHPSKD CAKDHFHEAL
151 ILASKVHSAP GIVAEICLSD DPFYTRGYLA LDGFFHRIPN IKDHGSTLGT
201 RIFIVEPDTD IPELIDYLEN TPVYIELPPD ASSSTDTTGL SSDLSAIAAQ
251 RNTAWAGAGL TRTLRTFETA QLPHSRIDGA DYLLFSSSDY LGLSTHPELV
301 SAATAAIGHF GTGSGGSRLT TGTSIHSALE SELAQFFGFD DAVLFATGYQ
351 ANHSTIAAIA TADVEIFSDA ANHASIIDGC RNARAKVTVF PHADYQTLDR
401 LLATSSARHK LVISDSVFSM SGEVIDGPAL ERTCRRRNAW LMLDDAHGVG
451 VIGEQGRGTA AHLDIRPDIV VGTASKALGV EGGYVLCSAP VGELLRNQAR
501 SFVYSTSMNP GSVAAIRAAL KQLEVGDVVK RLQRNIARVL SLVGAQSDPA
551 SAIIPLPVGD ETEAMDISAQ LAELGVFIPA IRYPTVPRGE AMLRLTITAL
601 HTDADIDQLE
```

Fig. 4.9 The amino acid sequence of recombinant *C. amycolatum SK46* BioWF fusion.

4.2.1 Cloning of Corynebacterium amycolatum SK46 BioWF fusion

During our bioinformatic searches for BioW/PCAS homologs we discovered an unusual BioW in the genome of *Corynebacterium amycolatum SK46* which was part of a fused pair with a BioF gene encoding an AONS homolog. This was of particular interest since the two enzymes catalyse consecutive steps in biotin biosynthesis. The PCAS enzyme generates a pimeloyl-CoA product which is the substrate for the AONS enzyme. We were interested to discover if the two enzymes interacted in any way and study how the substrates/products pass between the two components of the fusion.

Our initial goal was to express and purify recombinant *C. amycolatum* PCAS/AONS fusion from *E. coli.* The full-length *BioWF* gene from *C. amycolatum SK46* (Uniprot code: E2MUP3) was directly purchased from GenScript in a pUC57 vector and then cloned into pET-28a vector via NdeI and BamHI restriction sites (**Fig. 4.10**) which

gives N-terminal His_6 tagged BioWF fusion. The DNA and amino acid sequence is shown in **Appendix 5.7** .

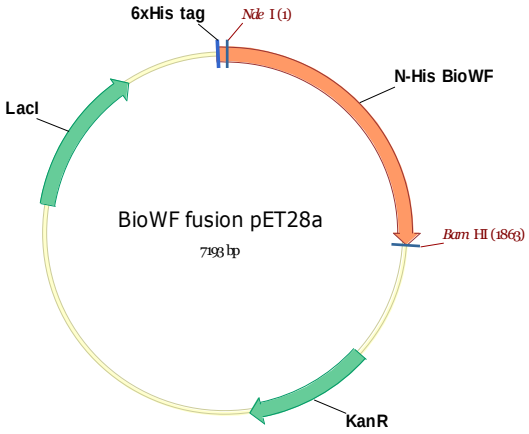


Fig. 4.10 Plasmid map of *C. amycolatum SK46* BioWF fusion in pET28a vector

4.2.2 Expression and Isolation of *C. amycolatum SK46* BioWF fusion

Different protein expression conditions were tried with BL21 (DE3) and BL21 (DE3) Gold *E. coli* strains. Finally, the optimal expression of soluble BioWF was obtained from the BL21 (DE3) Gold *E. coli* strain induced with 0.5 mM IPTG and grown at 20 ^{o}C overnight (**Fig. 4.11**). This expression system resulted in the appearance of a clear band at ~67 kDa on the SDS-PAGE gel which matches the predicted mass of the fusion.

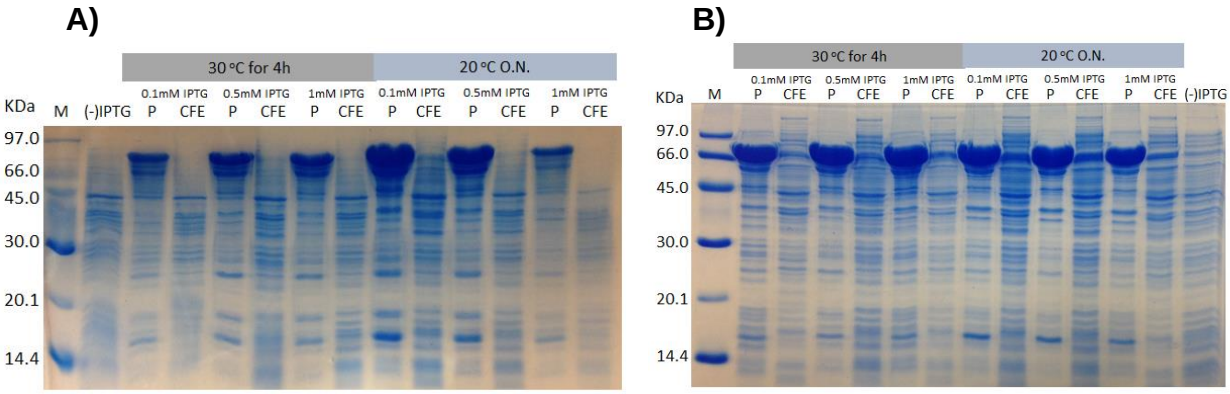


Fig. 4.11 Expression test of *Corynebacterium amycolatum SK46* BioWF fusion in **A)** BL21 (DE3) *E. coli* strain and **B)** BL21 (DE3) gold *E. coli* strain with different conditions.

Full-length N-terminal His_6 tagged BioWF was expressed in BL21 (DE3) gold *E. coli*, and isolated from cell free extract (CFE) by incubation with Ni resin then eluting with an increasing imidazole concentration gradient. The elution fractions were then analyzed by SDS-PAGE and showed a band at ~67 kDa which corresponding to the BioWF monomer (**Fig. 4.13**). Further purification by Superdex 200 size-exclusion column gave a peak at 64.4 mL corresponding to the theoretical mass of 130928.7 Da (**Fig. 4.13**, the calculation was shown in **Appendix 5.8**). Compared to the subunit *Mr* of ~67 kDa determined by SDS-PAGE gel, BioWF was found to be a dimer. The methods typically yielded ~11.3 mg of protein per litre of LB media.

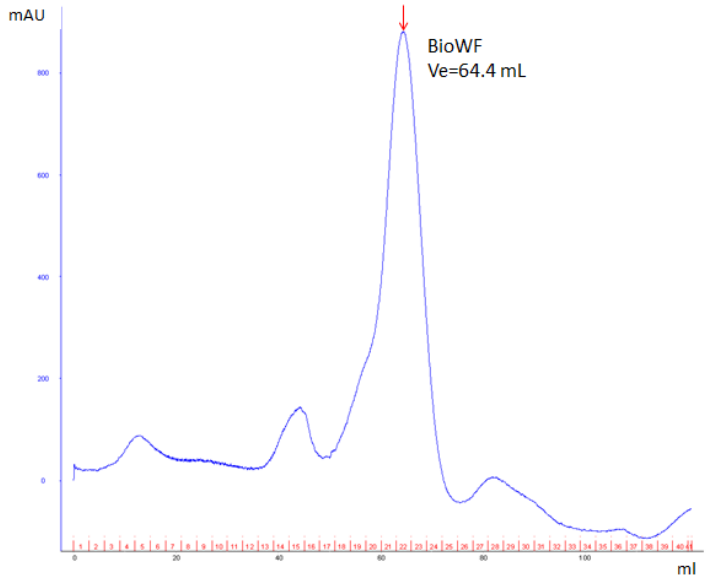


Fig. 4.12 AKTA trace of S200 column (A_{280nm}) of BioWF which gives a peak at 64.4 mL. Based on the calibration curve shown in **Appendix 5.8**, it corresponds to the theoretical mass of ~131 kDa, compared to subunit *Mr* of ~67 kDa determined by SDS-PAGE gel, BioWF was found to be a dimer.

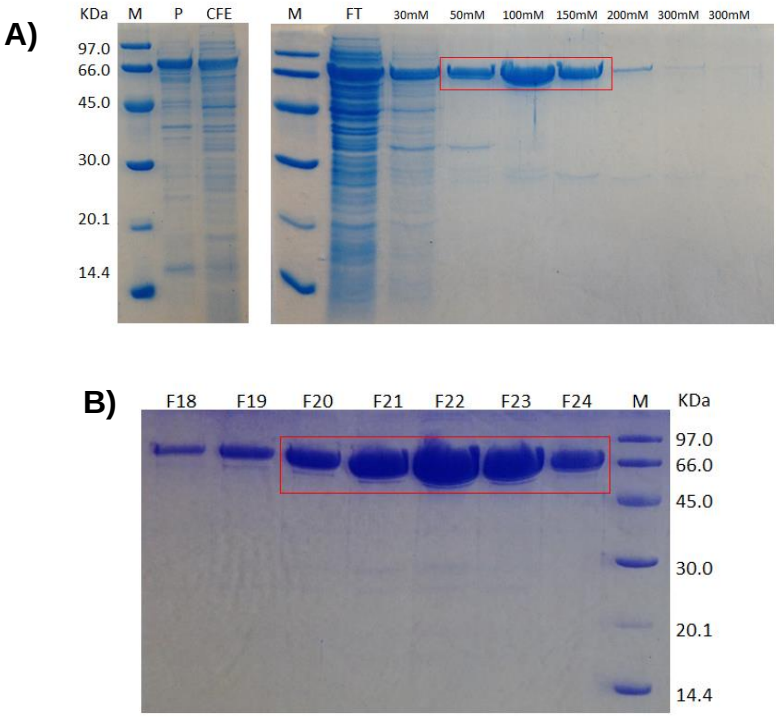


Fig. 4.13 SDS-PAGE gel showing the purification steps of BioWF including **A)** Cell lysis and Ni resin **B)** S200 gel filtration column.

4.2.3 Enzyme activity of *Corynebacterium amycolatum SK46* BioWF fusion

It was important to discover if the BioWF fusion had expressed in a soluble form such that both the PCAS and AONS domains had folded into catalytically active enzymes. A relatively straightforward analysis was to determine if the C-terminal AONS domain of the fusion contained the expected PLP cofactor. UV-vis analysis shown below (**Fig. 4.14 left panel**) showed that the protein had the characteristic spectrum similar to that observed with the *E. coli* AONS previously. Two maxima were observed at 330 and 420 nm. A titration with L-alanine was carried out and this led to changes in the UV-vis spectrum suggesting that L-Ala had bound, the concentration-dependent changes at 420 nm were plotted and a K_d of 1.80 ± 0.16 mM was determined

(Compared to the K_d of 0.71 ± 0.05 mM for *E. coli* AONS). This was strong evidence that the AONS part of the fusion bound the PLP cofactor and the L-alanine substrate,

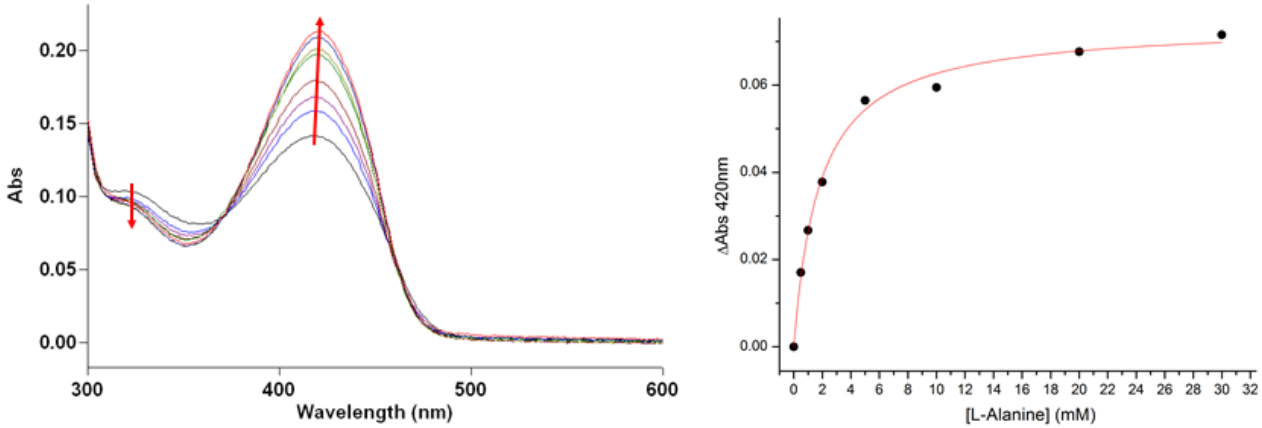


Fig. 4.14 Spectroscopic data of the K_d of BioWF for amino acid substrate L-Alanine, with arrows indicating peak increase or decrease. **left**) UV-vis spectrum for BioWF with L-Ala, **right**) Plot of L-Ala concentration against the absorbance at 420 nm.

4.3 Discussion and future work

The PLP-dependent enzyme AONS catalyses the condensation of pimeloyl-CoA with L-alanine to give 8-amino-7-oxononanoate (AON) in the first committed step of biotin biosynthesis. We used genome mining to identify a putative bi-functional enzyme with an N-terminal PCAS domain fused to C-terminal AONS domain in the organism *Corynebacterium amycolatum*. Recombinant PCAS/AONS was expressed and purified from *E. coli* and initial studies suggest that it forms a functional fused dimeric enzyme. Unfortunately I did not have time to complete my studies of the *C. amycolatum* BioWF fusion since the work with the *B. subtilis* PCAS enzyme moved to a structural and functional conclusion. Future work on the BioWF fusion should include be to determine if both the PCAS and AONS domains are catalytically active individually

for their respective reactions with their substrates. The kinetics of each enzyme will be interesting. If would also be interesting to analyse if the PCAS reaction is CoASH or ACP-dependent and to discover if the pimeloyl-thioester product of the PCAS domain is the substrate for the AONS domain. Future structural studies should also determine the architecture of the dimeric fusion protein namely, are the two domains arranged in a head-to-head or head-to-tail orientation. A key feature to observe is the linking sequence between the two domains and to determine the structure and dynamics of this part that links the two catalytic components.

Conclusion and future work

The generation of the key acyl-pimeloyl thioester intermediate in biotin biosynthesis varies between microorganisms. In *Bacillus subtilis* pimeloyl-CoA synthetase (PCAS, EC 6.2.1.14, UNIPROT code: P53559) is the first enzyme in the biotin biosynthetic pathway and acts as a highly specific substrate selection gate ensuring the integrity of the carbon chain in biotin synthesis. The PCAS catalysed reaction is composed of two steps; the first involves activation of pimelic acid (C7 dicarboxylic acid) using ATP to give an enzyme-bound, acyl-adenylate intermediate and pyrophosphate (PPi), and in the second step, this pimeloyl-adenylate reacts with coenzyme A (CoASH) to form pimeloyl-CoA. During the early stages of this project, Manandhar and Cronan [43] published a paper reporting the characterisation of PCAS from the *B. subtilis* strain 168 which is 100% identical to the *B. subtilis* PCAS used here. They had reported that attachment of a His_6 tag at either end of recombinant *B. subtilis* PCAS destroyed its ability to function in biotin complementation assays in *E. coli*. In contrast to Manandhar and Cronan, we found that a His_6 tag positioned at the N-terminus of *B. subtilis* PCAS allowed the isolation of milligram amounts of active and pure PCAS. Enzyme activity and kinetic constants were measured using reverse-phase HPLC and enzyme coupled spectroscopic assays. These revealed the enzyme is highly specific to pimelic acid, mono acids and different chain lengths are either not processed at all or are multiple orders of magnitude worse substrates [43].

In collaboration with colleagues at The University of St. Andrews various commercial and in-house screens were used to obtain diffraction-quality crystals suitable for X-ray crystallography. This also included the generation of seleno-methionine (SeMet) labelled PCAS, as well as heavy-metal soaks. Structures of *B. subtilis* PCAS in complex with the substrate pimelic acid and the pimeloyl-adenylate intermediate and

product PP_i were determined at 2.04 Å and 2.34 Å resolution respectively. The *B. subtilis* PCAS displays a novel 3D fold and defines a new class (Class IV) in the adenylating enzyme superfamily that catalyse acetylation-orming reactions with a two domain architecture. The smaller N-terminal domain which forms the dimer interface is responsible for binding the adenosine and three phosphates of CoASH. The larger C-terminal domain binds both the ATP and pimelic acid. The enzyme uses an internal ruler composed of a number of conserved arginine residues (Arg213, Arg227 and Arg170) to select the correct dicarboxylic acid substrate. Thioesters are high value substrates used as important building blocks for organic synthesis and chemical biology. They are chemically synthesised but the activated acids are often labile under aqueous conditions lowering isolated yields. As an alternative to these chemical methods, a "green" biocatalytic synthesis of simple fatty acid thioesters would be highly desirable. The relatively simple, catalytic domain of PCAS provided an ideal template to engineer an enzyme with modified substrate specificity. The X-ray structural study in this thesis allowed rational engineering of the PCAS active site to generate mutants with altered substrate specificity. Mutant PCAS Y211F was shown to synthesise both heptanoyl (C7) and octanoyl (C8) mono carboxylic acid-CoA and C8 dicarboxylic-CoA products highlighting the synthetic potential of PCAS.

We also identified Arg213 in the PCAS active site. Mutagenesis of this residue (R213A) inactivated the wild type enzyme when assayed using the DCA7 pimelic acid substrate but it displayed activity with C8 monocarboxlic acids. This suggests that this site can be further exploited and engineered to accept longer chains. Saturation mutagenesis at specific sites within PCAS, combined with the application of the tools of directed evolution and a high-throughput assay for PCAS would allow its full synthetic potential to be explored [104].

The PCAS pimeloyl-CoA product is the substrate for the next enzyme in the biotin pathway, a pyridoxal 5'phosphate (PLP)-dependent 8-amino 7-oxononanoate synthase (AONS) that catalyses the condensation of pimeloyl-CoA with L-alanine to give AON. We used genome mining to identify a putative bi-functional enzyme with an N-terminal PCAS domain fused to C-terminal AONS domain in the organism *Corynebacterium amycolatum*. An expression plasmid with the recombinant *C. amycolatum* BioWF (PCAS/AONS fusion) was prepared and used to purify the target protein from *E. coli*. Initial studies suggest that it bound both the PLP cofactor and the L-alanine substrate associated with the AONS domain. In the future, the activity of the N-terminal PCAS domain and the C-terminal AONS domain of *C. amycolatum* BioWF fusion should be studied with their individual substrates. It will be interesting to investigate if the PCAS/AONS fusion can convert pimelic acid to AON in the presence of ATP, CoASH and L-alanine and determine the kinetics of each step. It would also be interesting to analyse if the pimeloyl-thioester product of the PCAS domain is an acyl-CoA or acyl-ACP and determine which is a substrate for the AONS domain. It would also be of interest to incorporate the *B. subtilis* PCAS Y211F mutation that broadens the fatty acid specificity into the *C. amycolatum* PCAS/AONS fusion and determine if it could be used generate novel AON analogues in a single enzyme-catalysed step.

References

1. Tong, L. Structure and function of biotin-dependent carboxylases. *Cellular and molecular life sciences : CMLS* **70**, 863-891 (2013).
2. Zempleni, J. Uptake, localization, and noncarboxylase roles of biotin. *Annual review of nutrition* **25**, 175-196 (2005).
3. Choi, J.S. et al. Plumbagin as a new natural herbicide candidate for Sicyon angulatus control agent with the target 8-amino-7-oxononanoate synthase. *Pesticide Biochemistry & Physiology* **103**, 166-172 (2012).
4. Zempleni, J., Wijeratne, S.S. & Hassan, Y.I. Biotin. *BioFactors (Oxford, England)* **35**, 36-46 (2009).
5. Lin, S. & Cronan, J.E. Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* **7**, 1811-1821 (2011).
6. Cronan, J.E. & Lin, S. Synthesis of the alpha,omega-dicarboxylic acid precursor of biotin by the canonical fatty acid biosynthetic pathway. *Curr Opin Chem Biol* **15**, 407-413 (2011).
7. Wildiers, E. Nouvelle substance indispensable au développement de la levure. *La Cellule* **18**, 313-316 (1901).
8. Streit, W.R. & Entcheva, P. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl Microbiol Biotechnol* **61**, 21-31 (2003).
9. Boas, M.A. The Effect of Desiccation upon the nutritive properties of egg-white. *The Biochemical journal* **21**, 712-724.711 (1927).
10. György, P.G. The curative factor (vitamin H) for egg white injury, with particular reference to its presence in different food stuffs and in yeast. *J. Biol. Chem.* **131**, 733–744 (1939).
11. Allison, F.E., Hoover, S.R. & Burk, D. A respiration coenzyme. *Science (New York, N.Y.)* **78**, 217-218 (1933).
12. Kogl F, T.B.Ü.d.B.-P. Darstellung von krystallisiertem Biotin aus Eigelb. *Z. Physiol Chem* **242**, 43-73 (1932).
13. V, D.U.V., Dittmer, K., Hague, E. & Long, B. The growth-stimulating effect of biotin for the diphtheria bacillus in the absence of pimelic acid. *Science (New York, N.Y.)* **96**, 186-187 (1942).
14. Harris, S.A., Wolf, D.E., Mozingo, R. & Folkers, K. Synthetic biotin. *Science (New York, N.Y.)* **97**, 447-448 (1943).
15. DeTitta, G.T., Edmonds, J.W., Stallings, W. & Donohue, J. Molecular structure of biotin. Results of two independent crystal structure investigations. *Journal of the American Chemical Society* **98**, 1920-1926 (1976).
16. Tron, C.M. et al. Structural and functional studies of the biotin protein ligase from Aquifex aeolicus reveal a critical role for a conserved residue in target specificity. *Journal of molecular biology* **387**, 129-146 (2009).
17. Lane, M.D., Rominger, K.L., Young, D.L. & Lynen, F. The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-biotin. II. investigation of the reaction mechanism. *The Journal of biological chemistry* **239**, 2865-2871 (1964).
18. Knowles, J.R. The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* **58**, 195-221 (1989).

19. Attwood, P.V. & Wallace, J.C. Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc Chem Res* **35**, 113-120 (2002).
20. Cronan, J.E. & Lin, S. Synthesis of the alpha,omega-dicarboxylic acid precursor of biotin by the canonical fatty acid biosynthetic pathway. *Current Opinion in Chemical Biology* **15**, 407-413 (2011).
21. Alexeev, D. et al. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *Journal of molecular biology* **284**, 401-419 (1998).
22. Webster, S.P. et al. Mechanism of 8-amino-7-oxononanoate synthase: spectroscopic, kinetic, and crystallographic studies. *Biochemistry* **39**, 516-528 (2000).
23. Eisenberg, M.A. & Star, C. Synthesis of 7-oxo-8-aminopelargonic acid, a biotin vitamer, in cell-free extracts of Escherichia coli biotin auxotrophs. *Journal of bacteriology* **96**, 1291-1297 (1968).
24. Del Campillo-Campbell, A., Kayajanian, G., Campbell, A. & Adhya, S. Biotin-requiring mutants of Escherichia coli K-12. *Journal of bacteriology* **94**, 2065-2066 (1967).
25. Lemoine, Y., Wach, A. & Jeltsch, J.M. To be free or not: the fate of pimelate in Bacillus sphaericus and in Escherichia coli. *Mol Microbiol* **19**, 645-647 (1996).
26. Akatsuka, H., Kawai, E., Sakurai, N. & Omori, K. The Serratia marcescens bioH gene encodes an esterase. *Gene* **302**, 185-192 (2003).
27. Sanishvili, R. et al. Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from Escherichia coli. *The Journal of biological chemistry* **278**, 26039-26045 (2003).
28. Lin, S., Hanson, R.E. & Cronan, J.E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nature chemical biology* **6**, 682-688 (2010).
29. Cryle, M.J. Selectivity in a barren landscape: the P450(Biol)-ACP complex. *Biochemical Society Transactions* **38** (2010).
30. Lin, S. & Cronan, J.E. Closing in on complete pathways of biotin biosynthesis. *Molecular Biosystems* **7**, 1811-1821 (2011).
31. Stok, J.E. & De Voss, J. Expression, purification, and characterization of BioI: a carbon-carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in Bacillus subtilis. *Arch Biochem Biophys* **384**, 351-360 (2000).
32. Lin, S., Hanson, R.E. & Cronan, J.E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nature Chemical Biology* **6**, 682-688 (2010).
33. Cryle, M.J. & De Voss, J.J. Carbon-carbon bond cleavage by cytochrome p450(BioI)(CYP107H1). *Chem Commun (Camb)*, 86-87 (2004).
34. Cryle, M.J. & Schlichting, I. Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15696-15701 (2008).
35. Agarwal, V., Lin, S., Lukk, T., Nair, S.K. & Cronan, J.E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17406-17411 (2012).
36. Kack, H., Sandmark, J., Gibson, K., Schneider, G. & Lindqvist, Y. Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes. *Journal of molecular biology* **291**, 857-876 (1999).

37. Porebski, P.J. et al. Structural characterization of *Helicobacter pylori* dethiobiotin synthetase reveals differences between family members. *The FEBS journal* **279**, 1093-1105 (2012).
38. Berkovitch, F., Nicolet, Y., Wan, J.T., Jarrett, J.T. & Drennan, C.L. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science (New York, N.Y.)* **303**, 76-79 (2004).
39. Alexeev , D. et al. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, Acyl-CoA- condensing enzyme *J.Mol.Biol.* **284**, 401-419 (1998).
40. Ploux, O., Soularue, P., Marquet, A., Gloeckler, R. & Lemoine, Y. Investigation of the first step of biotin biosynthesis in Bacillus sphaericus. Purification and characterization of the pimeloyl-CoA synthase, and uptake of pimelate. *The Biochemical journal* **287 (Pt 3)**, 685-690 (1992).
41. Binieda, A. et al. Purification, characterization, DNA sequence and cloning of a pimeloyl-CoA synthetase from Pseudomonas mendocina 35. *The Biochemical journal* **340 (Pt 3)**, 793-801 (1999).
42. Ploux, O., Soularue, P., Marquet, A., Gloeckler, R. & Lemoine, Y. Investigation of the 1st step of biotin biosynthesis in bacillus-sphaericus - purification and characterization of the pimeloyl-coa synthase, and uptake of pimelate. *Biochemical Journal* **287**, 685-690 (1992).
43. Manandhar, M. & Cronan, J.E. Proofreading of noncognate acyl adenylates by an acyl-coenzyme a ligase. *Chem Biol* **20**, 1441-1446 (2013).
44. Gulick, A.M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol* **4**, 811-827 (2009).
45. Schmelz, S. & Naismith, J.H. Adenylate-forming enzymes. *Curr Opin Struct Biol* **19**, 666-671 (2009).
46. Schmelz, S. et al. Structural basis for acyl acceptor specificity in the achromobactin biosynthetic enzyme AcsD. *Journal of molecular biology* **412**, 495-504 (2011).
47. Schmelz, S. et al. AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nature chemical biology* **5**, 174-182 (2009).
48. McElroy, W.D., DeLuca, M. & Travis, J. Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid utilization are compared. *Science (New York, N.Y.)* **157**, 150-160 (1967).
49. Bar-Tana, J. & Rose, G. Studies on medium-chain fatty acyl-coenzyme a synthetase. Enzyme fraction I: mechanism of reaction and allosteric properties. *The Biochemical journal* **109**, 275-282 (1968).
50. Fulda, M., Heinz, E. & Wolter, F.P. The fadD gene of *Escherichia coli K12* is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family. *Mol Gen Genet* **242**, 241-249 (1994).
51. Xie, W., Nangle, L.A., Zhang, W., Schimmel, P. & Yang, X.L. Long-range structural effects of a Charcot-Marie-Tooth disease-causing mutation in human glycyl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9976-9981 (2007).
52. Bornscheuer, U.T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185-194 (2012).
53. Wang, M., Si, T. & Zhao, H.M. Biocatalyst development by directed evolution. *Bioresource Technology* **115**, 117-125 (2012).

54. Turner, N.J. Directed evolution drives the next generation of biocatalysts. *Nature Chemical Biology* **5**, 568-574 (2009).
55. Zaks, A. Industrial biocatalysis. *Current Opinion in Chemical Biology* **5**, 130-136 (2001).
56. Zhao, H.M., Chockalingam, K. & Chen, Z.L. Directed evolution of enzymes and pathways for industrial biocatalysis. *Current Opinion in Biotechnology* **13**, 104-110 (2002).
57. Hudlicky, T. & Reed, J.W. Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. *Chemical Society Reviews* **38**, 3117-3132 (2009).
58. Reetz, M.T. Directed evolution of enantioselective enzymes: an unconventional approach to asymmetric catalysis in organic chemistry. *Journal of Organic Chemistry* **74**, 5767-5778 (2009).
59. Wandrey, C., Liese, A. & Kihumbu, D. Industrial biocatalysis: past, present, and future. *Organic Process Research & Development* **4**, 286-290 (2000).
60. Francis, J.C. & Hansche, P.E. Directed evolution of metabolic pathways in microbial populations .1. modification of acid-phosphatase ph optimum in s-cerevisiae. *Genetics* **70**, 59-& (1972).
61. Khurana, J., Singh, R. & Kaur, J. Engineering of *Bacillus* lipase by directed evolution for enhanced thermal stability: effect of isoleucine to threonine mutation at protein surface. *Molecular Biology Reports* **38**, 2919-2926 (2011).
62. Zhao, H.M. & Arnold, F.H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Engineering* **12**, 47-53 (1999).
63. Giver, L., Gershenson, A., Freskgard, P.O. & Arnold, F.H. Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 12809-12813 (1998).
64. Moore, J.C. & Arnold, F.H. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology* **14**, 458-467 (1996).
65. Kawata, T. & Ogino, H. Enhancement of the organic solvent-stability of the LST-03 lipase by directed evolution. *Biotechnology Progress* **25**, 1605-1611 (2009).
66. Baxter, S. et al. An improved racemase/acylase biotransformation for the preparation of enantiomerically pure amino acids. *Journal of the American Chemical Society* **134**, 19310-19313 (2012).
67. Reetz, M.T., Wang, L.W. & Bocola, M. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie-International Edition* **45**, 1236-1241 (2006).
68. Lutz, S. & Patrick, W.M. Novel methods for directed evolution of enzymes: quality, not quantity. *Current Opinion in Biotechnology* **15**, 291-297 (2004).
69. Liu, H. & Naismith, J.H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102-111 (2009).
70. Liu, H. & Naismith, J.H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).
71. Katano, H., Tanaka, R., Maruyama, C. & Hamano, Y. Assay of enzymes forming AMP+PPi by the pyrophosphate determination based on the formation of 18-molybdopyrophosphate. *Anal Biochem* **421**, 308-312 (2012).

72. Katano, H., Watanabe, H., Takakuwa, M., Maruyama, C. & Hamano, Y. Colorimetric determination of pyrophosphate anion and its application to adenylation enzyme assay. *Anal Sci* **29**, 1095-1098 (2013).
73. Bower, S. et al. Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *Journal of bacteriology* **178**, 4122-4130 (1996).
74. Lin, S. & Cronan, J.E. The BioC O-methyltransferase catalyzes methyl esterification of malonyl-acyl carrier protein, an essential step in biotin synthesis. *The Journal of biological chemistry* **287**, 37010-37020 (2012).
75. Taylor, G.L. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66**, 325-338 (2010).
76. Taylor, G. The phase problem. *Acta Crystallogr D Biol Crystallogr* **59**, 1881-1890 (2003).
77. Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190 (2010).
78. Zhang, Z., Sauter, N.K., van den Bedem, H., Snell, G. & Deacon, A.M. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *J Appl Crystallogr* **39**, 112-119 (2006).
79. Sauter, N.K., Grosse-Kunstleve, R.W. & Adams, P.D. Robust indexing for automatic data collection. *J Appl Crystallogr* **37**, 399-409 (2004).
80. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26**, 795-800 (1993).
81. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D* **62**, 72-82 (2006).
82. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymol., Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).
83. Skubak, P. & Pannu, N.S. Automatic protein structure solution from weak X-ray data. *Nature communications* **4**, 2777 (2013).
84. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235-242 (2011).
85. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
86. Terwilliger, T.C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* **64**, 61-69 (2008).
87. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
88. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
89. Schuttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-1363 (2004).
90. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
91. Schrodinger, LLC (2010).
92. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774-797 (2007).

93. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Curr Opin Struct Biol* **23**, 252-260 (2013).
94. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-324 (2014).
95. Finzel, K., Lee, D.J. & Burkart, M.D. Using modern tools to probe the structure-function relationship of fatty acid synthases. *Chembiochem* **16**, 528-547 (2015).
96. Merrill, A.H., Jr. Sphingolipid and glycosphingolipid metabolic pathways in the era of sphingolipidomics. *Chem Rev* **111**, 6387-6422 (2011).
97. Raetz, C.R., Reynolds, C.M., Trent, M.S. & Bishop, R.E. Lipid A modification systems in gram-negative bacteria. *Annu Rev Biochem* **76**, 295-329 (2007).
98. Weissman, K.J. The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **11**, 660-670 (2015).
99. Choi-Rhee, E., Schulman, H. & Cronan, J.E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Sci* **13**, 3043-3050 (2004).
100. Dawson, P.E., Muir, T.W., Clark-Lewis, I. & Kent, S.B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).
101. Mishra, P.K. & Drueckhammer, D.G. Coenzyme A analogues and derivatives: synthesis and applications as mechanistic probes of coenzyme A ester-utilizing enzymes. *Chemical reviews* **100**, 3283-3310 (2000).
102. Huang, Y.T., Lu, S.Y., Yi, C.L. & Lee, C.F. Iron-catalyzed synthesis of thioesters from thiols and aldehydes in water. *The Journal of organic chemistry* **79**, 4561-4568 (2014).
103. Pal, M. & Bearne, S.L. Synthesis of coenzyme A thioesters using methyl acyl phosphates in an aqueous medium. *Organic & biomolecular chemistry* **12**, 9760-9763 (2014).
104. Renata, H., Wang, Z.J. & Arnold, F.H. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angewandte Chemie (International ed. in English)* **54**, 3351-3367 (2015).

Chapter V. Appendix

Appendix 5.1 The DNA and amino acid sequence of wild type *B. subtilis* PCAS (Uniprot code: P53559)

DNA sequence of recombinant *B. subtilis* PCAS

```
  1 ATGATGCAAG AAGAAACTTT TTATAGTGTC AGAATGAGGG CTTCAATGAA
 51 TGGATCTCAT GAAGACGGCG GAAAGCATAT ATCCGGCGGA GAACGGCTTA
101 TTCCTTTCCA TGAGATGAAG CATACAGTCA ATGCTTTATT AGAAAAAGGG
151 TTATCCCATT CAAGAGGAAA ACCTGATTTT ATGCAAATTC AATTTGAAGA
201 GGTACATGAA TCGATAAAAA CCATTCAGCC ATTGCCTGTG CATACGAATG
251 AAGTGAGCTG CCCGGAAGAA GGACAAAAGC TTGCCCGATT GTTATTGGAA
301 AAAGAAGGCG TTTCACGAGA CGTGATTGAA AAAGCATATG AACAAATCCC
351 TGAATGGTCA GATGTCAGGG GTGCGGTGTT GTTTGATATT CATACAGGCA
401 AGCGAATGGA TCAAACAAAA GAAAAAGGGG TGCGGGTCTC CAGAATGGAT
451 TGGCCGGACG CTAATTTTGA AAAATGGGCG CTTCACAGTC ACGTGCCAGC
501 TCATTCAAGA ATAAAAGAGG CCCTTGCGCT CGCTTCAAAG GTAAGCCGGC
551 ACCCGGCAGT CGTTGCAGAA TTATGCTGGT CGGACGATCC GGATTACATA
601 ACAGGCTATG TTGCGGGTAA GAAAATGGGC TATCAGCGTA TTACAGCAAT
651 GAAAGAATAC GGGACTGAAG AGGGCTGCCG AGTCTTTTTT ATTGATGGAT
701 CCAATGATGT AAACACGTAC ATACATGACC TGGAGAAGCA GCCTATTTTA
751 ATAGAGTGGG AGGAAGATCA TGACTCATGA
```

(Inherent BamHI restriction site is in blue, Inherent NdeI restriction site is in green start codon and stop condon are underlined.)

Amino acid sequence of recombinant *B. subtilis* PCAS

```
  1 MMQEETFYSV RMRASMNGSH EDGGKHISGG ERLIPFHEMK HTVNALLEKG
 51 LSHSRGKPDF MQIQFEEVHE SIKTIQPLPV HTNEVSCPEE GQKLARLLLE
101 KEGVSRDVIE KAYEQIPEWS DVRGAVLFDI HTGKRMDQTK EKGVRVSRMD
151 WPDANFEKWA LHSHVPAHSR IKEALALASK VSRHPAVVAE LCWSDDPDYI
201 TGYVAGKKMG YQRITAMKEY GTEEGCRVFF IDGSNDVNTY IHDLEKQPIL
251 IEWEEDHDS 259
```

Appendix 5.2 The DNA and amino acid sequence of codon-optimised *B. subtilis* PCAS/NHis (N-terminal His_6 tagged, in plasmid pET-28a)

DNA sequence of recombinant *B. subtilis* PCAS/NHis

```
-56 ATGGGCAGCA GCCATCATCA TCATCATCAC AGCAGCGGCC TGGTGCCGCG
 -6 CGGCAGCCAT ATGCAAGAAG AAACTTTTTA TAGTGTCAGA ATGAGGGCTT
 44 CAATGAATGG ATCTCATGAA GACGGCGGAA AGCATATATC CGGCGGAGAA
 94 CGGCTTATTC CTTTCCATGA GATGAAGCAT ACAGTCAATG CTTTATTAGA
144 AAAAGGGTTA TCCCATTCAA GAGGAAAACC TGATTTTATG CAAATTCAAT
194 TTGAAGAGGT ACATGAATCG ATAAAAACCA TTCAGCCATT GCCTGTGCAT
244 ACGAATGAAG TGAGCTGCCC GGAAGAAGGA CAAAAACTTG CCCGATTGTT
294 ATTGGAAAAA GAAGGCGTTT CACGAGACGT GATTGAAAAA GCGTATGAAC
344 AAATCCCTGA ATGGTCAGAT GTCAGGGGTG CGGTGTTGTT TGATATTCAT
394 ACAGGCAAGC GAATGGATCA AACAAAAGAA AAAGGGGTGC GGGTCTCCAG
444 AATGGATTGG CCGGACGCTA ATTTTGAAAA ATGGGCGCTT CACAGTCACG
494 TGCCAGCTCA TTCAAGAATA AAAGAGGCCC TTGCGCTCGC TTCAAAGGTA
544 AGCCGGCACC CGGCAGTCGT TGCAGAATTA TGCTGGTCGG ACGATCCGGA
594 TTACATAACA GGCTATGTTG CGGGTAAGAA AATGGGCTAT CAGCGTATTA
644 CAGCAATGAA AGAATACGGG ACTGAAGAGG GCTGCCGAGT CTTTTTTATT
694 GATGGCTCCA ATGATGTAAA CACGTACATA CATGACCTGG AGAAGCAGCC
744 TATTTTAATA GAGTGGGAGG AAGATCATGA CTCATGAGGA TCC
```

(Mutated inherent BamHI restriction site is in blue, Mutated NdeI restriction site is in green, NdeI restriction site is in purple, BamHI restriction site is in red, N-terminal His_6 tag in orange, start codon and stop condon are underlined)

Amino acid sequence of recombinant *B. subtilis* PCAS/NHis

```
-18 MGSSHHHHHH SSGLVPRGSH MQEETFYSVR MRASMNGSHE DGGKHISGGE
 32 RLIPFHEMKH TVNALLEKGL SHSRGKPDFM QIQFEEVHES IKTIQPLPVH
 82 TNEVSCPEEG QKLARLLLEK EGVSRDVIEK AYEQIPEWSD VRGAVLFDIH
132 TGKRMDQTKE KGVRVSRMDW PDANFEKWAL HSHVPAHSRI KEALALASKV
182 SRHPAVVAEL CWSDDPDYIT GYVAGKKMGY QRITAMKEYG TEEGCRVFFI
232 DGSNDVNTYI HDLEKQPILI EWEEDHDS
```

(N-terminal His_6 tag is in orange)

Appendix 5.3 The DNA and amino acid sequence of codon-optimised *B. subtilis* PCAS/NTH (N-terminal TEV protease-cleavable His_6 tagged, in pEHISTEV plasmid)

DNA sequence of recombinant *B. subtilis* PCAS/NTH

```
-71 ATGTCGTACT ACCATCACCA TCACCATCAC GATTACGATA TCCCAACGAC
-21 CGAAAACCTG TATTTTCAGG GCGCCATGGA AGAAGAAACT TTTTATAGTG
 29 TCAGAATGAG GGCTTCAATG AATGGATCTC ATGAAGACGG CGGAAAGCAT
 79 ATATCCGGCG GAGAACGGCT TATTCCTTTC CATGAGATGA AGCATACAGT
129 CAATGCTTTA TTAGAAAAAG GGTTATCCCA TTCAAGAGGA AAACCTGATT
179 TTATGCAAAT TCAATTTGAA GAGGTACATG AATCGATAAA AACCATTCAG
229 CCATTGCCTG TGCATACGAA TGAAGTGAGC TGCCCGGAAG AAGGACAAAA
279 ACTTGCCCGA TTGTTATTGG AAAAAGAAGG CGTTTCACGA GACGTGATTG
329 AAAAAGCGTA TGAACAAATC CCTGAATGGT CAGATGTCAG GGGTGCGGTG
379 TTGTTTGATA TTCATACAGG CAAGCGAATG GATCAAACAA AAGAAAAAGG
429 GGTGCGGGTC TCCAGAATGG ATTGGCCGGA CGCTAATTTT GAAAAATGGG
479 CGCTTCACAG TCACGTGCCA GCTCATTCAA GAATAAAAGA GGCCCTTGCG
529 CTCGCTTCAA AGGTAAGCCG GCACCCGGCA GTCGTTGCAG AATTATGCTG
579 GTCGGACGAT CCGGATTACA TAACAGGCTA TGTTGCGGGT AAGAAAATGG
629 GCTATCAGCG TATTACAGCA ATGAAAGAAT ACGGGACTGA AGAGGGCTGC
679 CGAGTCTTTT TTATTGATGG CTCCAATGAT GTAAACACGT ACATACATGA
729 CCTGGAGAAG CAGCCTATTT TAATAGAGTG GGAGGAAGAT CATGACTCAT
779 GAGGATCC
```

(Mutated inherent BamHI restriction site is in blue, Mutated NdeI restriction site is in green, NcoI restriction site is in purple, BamHI restriction site is in red, N-terminal TEV protease-cleavable His_6 tag in orange, TEV protease site in grey, start codon and stop condon are underlined)

Amino acid sequence of recombinant *B. subtilis* PCAS/NTH before TEV protease cleavage

```
-23 MSYYHHHHHH DYDIPTTENL YFQGAMEEET FYSVRMRASM NGSHEDGGKH
 27 ISGGERLIPF HEMKHTVNAL LEKGLSHSRG KPDFMQIQFE EVHESIKTIQ
 77 PLPVHTNEVS CPEEGQKLAR LLLEKEGVSR DVIEKAYEQI PEWSDVRGAV
127 LFDIHTGKRM DQTKEKGVRV SRMDWPDANF EKWALHSHVP AHSRIKEALA
177 LASKVSRHPA VVAELCWSDD PDYITGYVAG KKMGYQRITA MKEYGTEEGC
227 RVFFIDGSND VNTYIHDLEK QPILIEWEED HDS
```

(N-terminal TEV protease-cleavable His_6 tag is in orange, TEV protease site in grey)

Amino acid sequence of recombinant *B. subtilis* PCAS/NTH after TEV protease cleavage

```
  0 GAMEEETFYS VRMRASMNGS HEDGGKHISG GERLIPFHEM KHTVNALLEK
 50 GLSHSRGKPD FMQIQFEEVH ESIKTIQPLP VHTNEVSCPE EGQKLARLLL
100 EKEGVSRDVI EKAYEQIPEW SDVRGAVLFD IHTGKRMDQT KEKGVRVSRM
150 DWPDANFEKW ALHSHVPAHS RIKEALALAS KVSRHPAVVA ELCWSDDPDY
200 ITGYVAGKKM GYQRITAMKE YGTEEGCRVF FIDGSNDVNT YIHDLEKQPI
250 LIEWEEDHDS
```

Appendix 5.4 The DNA and amino acid sequence of codon-optimised *B. subtilis* PCAS/UT (un-tagged PACS, in plasmid pET22b)

DNA sequence of recombinant *B. subtilis* PCAS/UT

```
  4 CATATGCAAG AAGAAACTTT TTATAGTGTC AGAATGAGGG CTTCAATGAA
 54 TGGATCTCAT GAAGACGGCG GAAAGCATAT ATCCGGCGGA GAACGGCTTA
104 TTCCTTTCCA TGAGATGAAG CATACAGTCA ATGCTTTATT AGAAAAAGGG
154 TTATCCCATT CAAGAGGAAA ACCTGATTTT ATGCAAATTC AATTTGAAGA
204 GGTACATGAA TCGATAAAAA CCATTCAGCC ATTGCCTGTG CATACGAATG
254 AAGTGAGCTG CCCGGAAGAA GGACAAAAAC TTGCCCGATT GTTATTGGAA
304 AAAGAAGGCG TTTCACGAGA CGTGATTGAA AAAGCGTATG AACAAATCCC
354 TGAATGGTCA GATGTCAGGG GTGCGGTGTT GTTTGATATT CATACAGGCA
404 AGCGAATGGA TCAAACAAAA GAAAAAGGGG TGCGGGTCTC CAGAATGGAT
454 TGGCCGGACG CTAATTTTGA AAAATGGGCG CTTCACAGTC ACGTGCCAGC
504 TCATTCAAGA ATAAAAGAGG CCCTTGCGCT CGCTTCAAAG GTAAGCCGGC
554 ACCCGGCAGT CGTTGCAGAA TTATGCTGGT CGGACGATCC GGATTACATA
604 ACAGGCTATG TTGCGGGTAA GAAAATGGGC TATCAGCGTA TTACAGCAAT
654 GAAAGAATAC GGGACTGAAG AGGGCTGCCG AGTCTTTTTT ATTGATGGCT
704 CCAATGATGT AAACACGTAC ATACATGACC TGGAGAAGCA GCCTATTTTA
754 ATAGAGTGGG AGGAAGATCA TGACTCATGA GGATCCGAAT TCGAGCTCCG
804 TCGACAAGCT TGCGGCCGCA CTCGAG
```

Amino acid sequence of recombinant *B. subtilis* PCAS/UT

```
  2 MQEETFYSVR MRASMNGSHE DGGKHISGGE RLIPFHEMKH TVNALLEKGL
 52 SHSRGKPDFM QIQFEEVHES IKTIQPLPVH TNEVSCPEEG QKLARLLLEK
102 EGVSRDVIEK AYEQIPEWSD VRGAVLFDIH TGKRMDQTKE KGVRVSRMDW
152 PDANFEKWAL HSHVPAHSRI KEALALASKV SRHPAVVAEL CWSDDPDYIT
202 GYVAGKKMGY QRITAMKEYG TEEGCRVFFI DGSNDVNTYI HDLEKQPILI
252 EWEEDHDS
```

(Mutated inherent BamHI restriction site is in blue, Mutated NdeI restriction site is in green, NdeI restriction site is in purple and XhoI restriction site is in red, N-terminal TEV protease-cleavable His_6 tag in orange, TEV protease site in grey, start codon and stop condon are underlined)

Appendix 5.5 The DNA and amino acid sequence of recombinant *E. coli* AONS (N-His$_6$ tagged, in pET28a, Uniprot code: P12998)

DNA sequence of recombinant *E. coli* AONS /NHis

```
 -59 ATGGGCAGCA GCCATCATCA TCATCATCAC AGCAGCGGCC TGGTGCCGCG
  -9 CGGCAGCCAT ATGAGCTGGC AGGAAAAAAT CAACGCGGCG CTCGATGCGC
  41 GGCGTGCTGC CGATGCCCTG CGTCGCCGTT ATCCGGTGGC GCAAGGAGCC
  91 GGACGCTGGC TGGTGGCGGA TGATCGCCAG TATCTGAACT TTTCCAGTAA
 141 CGATTATTTA GGTTTAAGCC ATCATCCGCA AATTATCCGT GCCTGGCAGC
 191 AGGGGGCGGA GCAATTTGGC ATCGGTAGCG GCGGCTCCGG TCACGTCAGC
 241 GGTTATAGCG TGGTGCATCA GGCACTGGAA GAAGAGCTGG CCGAGTGGCT
 291 TGGCTATTCG CGGGCACTGC TGTTTATCTC TGGTTTCGCC GCTAATCAGG
 341 CAGTTATTGC CGCGATGATG GCGAAAGAGG ACCGTATTGC TGCCGACCGG
 391 CTTAGCCATG CCTCATTGCT GGAAGCTGCC AGTTTAAGCC CGTCGCAGCT
 441 TCGCCGTTTT GCTCATAACG ATGTCACTCA TTTGGCGCGA TTGCTTGCTT
 491 CCCCCTGTCC GGGGCAGCAA ATGGTGGTGA CAGAAGGCGT GTTCAGCATG
 541 GACGGCGATA GTGCGCCACT GGCGGAAATC CAGCAGGTAA CGCAACAGCA
 591 CAATGGCTGG TTGATGGTCG ATGATGCCCA CGGCACGGGC GTTATCGGGG
 641 AGCAGGGGCG CGGCAGCTGC TGGCTGCAAA AGGTAAAACC AGAATTGCTG
 691 GTAGTGACTT TTGGCAAAGG ATTTGGCGTC AGCGGGGCAG CGGTGCTTTG
 741 CTCCAGTACG GTGGCGGATT ATCTGCTGCA ATTCGCCCGC CACCTTATCT
 791 ACAGCACCAG TATGCCGCCC GCTCAGGCGC AGGCATTACG TGCGTCGCTG
 841 GCGGTCATTC GCAGTGATGA GGGTGATGCA CGGCGCGAAA AACTGGCGGC
 891 ACTCATTACG CGTTTTCGTG CCGGAGTACA GGATTTGCCG TTTACGCTTG
 941 CTGATTCATG CAGCGCCATC CAGCCATTGA TTGTCGGTGA TAACAGCCGT
 991 GCGTTACAAC TGGCAGAAAA ACTGCGTCAG CAAGGCTGCT GGGTCACGGC
1041 GATTCGCCCG CCAACCGTAC CCGCTGGTAC TGCGCGACTG CGCTTAACGC
1091 TAACCGCTGC GCATGAAATG CAGGATATCG ACCGTCTGCT GGAGGTGCTG
1141 CATGGCAACG GTTAATAAAC AAGCCATTGC AGCGGCATTT GGTCGGGCAG
1191 CCGCACACTA TGAGCAACAT GCAGATCCTC TAGAGTCGAC CTGCAGGCAT
1241 GCAAGCTTCG AGGATCC
```

(NdeI restriction site is in purple, BamHI restriction site is in red, N-terminal His$_6$ tag in orange, start codon and stop condon are underlined)

Amino acid sequence of recombinant *E. coli* AONS/NHis

```
-19 MGSSHHHHHH SSGLVPRGSH MSWQEKINAA LDARRAADAL RRRYPVAQGA
 31 GRWLVADDRQ YLNFSSNDYL GLSHHPQIIR AWQQGAEQFG IGSGGSGHVS
 81 GYSVVHQALE EELAEWLGYS RALLFISGFA ANQAVIAAMM AKEDRIAADR
131 LSHASLLEAA SLSPSQLRRF AHNDVTHLAR LLASPCPGQQ MVVTEGVFSM
181 DGDSAPLAEI QQVTQQHNGW LMVDDAHGTG VIGEQGRGSC WLQKVKPELL
231 VVTFGKGFGV SGAAVLCSST VADYLLQFAR HLIYSTSMPP AQAQALRASL
281 AVIRSDEGDA RREKLAALIT RFRAGVQDLP FTLADSCSAI QPLIVGDNSR
331 ALQLAEKLRQ QGCWVTAIRP PTVPAGTARL RLTLTAAHEM QDIDRLLEVL
381 HGNG
```

(N-terminal His_6 tag is in orange)

Appendix 5.6 The DNA and amino acid sequence of recombinant *E. coli* AONS (C-terminal His_6 tagged, in pET22b)

DNA sequence of recombinant *E. coli* AONS/CHis

```
 -59 CATATGAGCT GGCAGGAGAA AATCAACGCG GCGCTCGATG CGCGGCGTGC
  -9 TGCCGATGCC CTGCGTCGCC GTTATCCGGT GGCGCAAGGA GCCGGACGCT
  41 GGCTGGTGGC GGATGATCGC CAGTATCTGA ACTTTTCCAG TAACGATTAT
  91 TTAGGTTTAA GCCATCATCC GCAAATTATC CGTGCCTGGC AGCAGGGGGC
 141 GGAGCAATTT GGCATCGGTA GCGGCGGCTC CGGTCACGTC AGCGGTTATA
 191 GCGTGGTGCA TCAGGCACTG GAAGAAGAGC TGGCCGAGTG GCTTGGCTAT
 241 TCGCGGGCAC TGCTGTTTAT CTCTGGTTTC GCCGCTAATC AGGCAGTTAT
 291 TGCCGCGATG ATGGCGAAAG AGGACCGTAT TGCTGCCGAC CGGCTTAGCC
 341 ATGCCTCATT GCTGGAAGCT GCCAGTTTAA GCCCGTCGCA GCTTCGCCGT
 391 TTTGCTCATA ACGATGTCAC TCATTTGGCG CGATTGCTTG CTTCCCCCTG
 441 TCCGGGGCAG CAAATGGTGG TGACAGAAGG CGTGTTCAGC ATGGACGGCG
 491 ATAGTGCGCC ACTGGCGGAA ATCCAGCAGG TAACGCAACA GCACAATGGC
 541 TGGTTGATGG TCGATGATGC CCACGGCACG GGCGTTATCG GGGAGCAGGG
 591 GCGCGGCAGC TGCTGGCTGC AAAAGGTAAA ACCAGAATTG CTGGTAGTGA
 641 CTTTTGGCAA AGGATTTGGC GTCAGCGGGG CAGCGGTGCT TTGCTCCAGT
 691 ACGGTGGCGG ATTATCTGCT GCAATTCGCC CGCCACCTTA TCTACAGCAC
 741 CAGTATGCCG CCCGCTCAGG CGCAGGCATT ACGTGCGTCG CTGGCGGTCA
 791 TTCGCAGTGA TGAGGGTGAT GCACGGCGCG AAAAACTGGC GGCACTCATT
 841 ACGCGTTTTC GTGCCGGAGT ACAGGATTTG CCGTTTACGC TTGCTGATTC
 891 ATGCAGCGCC ATCCAGCCAT TGATTGTCGG TGATAACAGC CGTGCGTTAC
 941 AACTGGCAGA AAAACTGCGT CAGCAAGGCT GCTGGGTCAC GGCGATTCGC
 991 CCGCCAACCG TACCCGCTGG TACTGCGCGA CTGCGCTTAA CGCTAACCGC
1041 TGCGCATGAA ATGCAGGATA TCGACCGTCT GCTGGAGGTG CTGCATGGCA
1091 ACGGTCTCGA GCACCACCAC CACCACCACT GA
```

(NdeI restriction site is in purple, XhoI restriction site is in red, C-terminal His_6 tag in orange, start codon and stop condon are underlined)

Amino acid sequence of recombinant *E. coli* AONS/CHis

```
  1 MSWQEKINAA LDARRAADAL RRRYPVAQGA GRWLVADDRQ YLNFSSNDYL
 51 GLSHHPQIIR AWQQGAEQFG IGSGGSGHVS GYSVVHQALE EELAEWLGYS
101 RALLFISGFA ANQAVIAAMM AKEDRIAADR LSHASLLEAA SLSPSQLRRF
151 AHNDVTHLAR LLASPCPGQQ MVVTEGVFSM DGDSAPLAEI QQVTQQHNGW
201 LMVDDAHGTG VIGEQGRGSC WLQKVKPELL VVTFGKGFGV SGAAVLCSST
251 VADYLLQFAR HLIYSTSMPP AQAQALRASL AVIRSDEGDA RREKLAALIT
301 RFRAGVQDLP FTLADSCSAI QPLIVGDNSR ALQLAEKLRQ QGCWVTAIRP
351 PTVPAGTARL RLTLTAAHEM QDIDRLLEVL HGNGLEHHHH HH
```

(C-terminal His_6 tag is in orange)

Appendix 5.7 The DNA and amino acid sequence of recombinant *Corynebacterium amycolatum SK46* BioWF fusion (N-terminal His_6 tagged, in plasmid pET28a, Uniprot code: E2MUP3)

DNA sequence of recombinant *C. amycolatum* BioWF/NHis

```
 -59 ATGGGCAGCA GCCATCATCA TCATCATCAC AGCAGCGGCC TGGTGCCGCG
  -9 CGGCAGCCAT ATGCGCTCGT CCGCTGATTA CTGTCACATC TCGGGTGCTG
  41 AACGCCTGGC CCCGGCTACC GAACTGCCGC AACTGGCATC CGCAATGACC
  91 AGCCGTGCCC TGCATCACGA TAAAGGCCGC CCGGACACCA TCCATATCAC
 141 GGTTGATAAA ATCGAAGAAT CCACCATCTC AACGGTCCCG GCGCTGACCC
 191 CGTTTCTGGA AAGCAACAGC TCTCCGGGCG ATGCTCGCAA ACTGATTGCG
 241 CAGCGTCTGC ACGCGGCCGG TATTCAAGCA GCTGATATCG CGGCCGAAAT
 291 GGCCTATAGC CTGACCGGCC TGCGTGGTGC AGCTCTGATC GATAGTTCCT
 341 CAGGTGAACG TCTGGACCCG AATCCGGCCC GCGGTGTTCG TGTCAGCACC
 391 TTTGATGCTA TTTCGCATCC GAGCAAAGAT TGCGCAAAAG ACCATTTCCA
 441 CGAAGCCCTG ATCCTGGCAT CGAAAGTTCA CAGCGCCCCG GGCATTGTCG
 491 CAGAAATCTG TCTGTCTGAT GACCCGTTTT ATACCCGTGG CTACCTGGCG
 541 CTGGATGGCT TTTTCCATCG CATTCCGAAC ATCAAAGACC ACGGCAGTAC
 591 CCTGGGTACG CGTATTTTCA TCGTGGAACC GGATACCGAC ATTCCGGAAC
 641 TGATCGATTA TCTGGAAAAT ACGCCGGTTT ACATTGAACT GCCGCCGGAT
 691 GCCTCGAGCT CTACCGACAC CACGGGTCTG AGTTCCGATC TGAGCGCAAT
 741 TGCCGCACAG CGTAATACCG CCTGGGCAGG TGCAGGTCTG ACCCGCACGC
 791 TGCGTACCTT TGAAACGGCG CAACTGCCGC ATAGTCGCAT TGATGGCGCC
 841 GACTATCTGC TGTTCTCATC GAGCGATTAC CTGGGTCTGT CCACCCATCC
 891 GGAACTGGTG TCAGCAGCTA CCGCGGCCAT TGGCCACTTT GGCACGGGTT
 941 CTGGCGGTAG TCGTCTGACC ACGGGTACCA GCATCCATTC AGCTCTGGAA
 991 TCGGAACTGG CGCAGTTTTT CGGCTTTGAT GACGCTGTCC TGTTCGCGAC
1041 GGGTTATCAA GCGAACCATA GCACCATTGC AGCTATCGCT ACGGCGGATG
1091 TTGAAATTTT TAGCGACGCG GCCAACCACG CCAGCATTAT CGATGGTTGC
1141 CGCAATGCCC GTGCAAAAGT GACCGTTTTC CCGCATGCGG ATTACCAGAC
1191 CCTGGACCGT CTGCTGGCCA CGTCTAGTGC ACGCCACAAA CTGGTGATTT
1241 CCGATTCAGT GTTTAGCATG AGCGGCGAAG TGATTGATGG TCCGGCCCTG
1291 GAACGTACCT GCCGTCGCCG TAATGCCTGG CTGATGCTGG ATGACGCACA
1341 TGGCGTCGGT GTGATTGGCG AACAGGGCCG TGGTACCGCA GCTCACCTGG
1391 ATATTCGTCC GGACATTGTG GTTGGTACCG CATCCAAAGC ACTGGGTGTC
1441 GAAGGCGGTT ATGTGCTGTG TTCAGCACCG GTTGGTGAAC TGCTGCGCAA
1491 CCAGGCGCGT TCTTTTGTGT ACTCTACCAG TATGAATCCG GGCAGTGTTG
1541 CGGCCATTCG TGCAGCTCTG AAACAGCTGG AAGTGGGTGA TGTCGTGAAA
1591 CGCCTGCAAC GTAATATTGC TCGTGTTCTG TCCCTGGTCG GCGCCCAGTC
1641 TGATCCGGCT AGTGCGATTA TCCCGCTGCC GGTTGGTGAT GAAACCGAAG
1691 CCATGGACAT TTCTGCTCAA CTGGCGGAAC TGGGCGTCTT CATTCCGGCC
1741 ATCCGTTATC CGACCGTGCC GCGTGGTGAA GCAATGCTGC GTCTGACCAT
1791 CACGGCACTG CACACCGATG CCGACATTGA CCAACTGGAA CTGGCGCTGC
1841 GTAACACGGG CCTGCTGTAA GGATCC
```

(NdeI restriction site is in purple, BamHI restriction site is in red, N-terminal His_6 tag in orange, start codon and stop condon are underlined)

Amino acid sequence of recombinant *C. amycolatum* BioWF/NHis

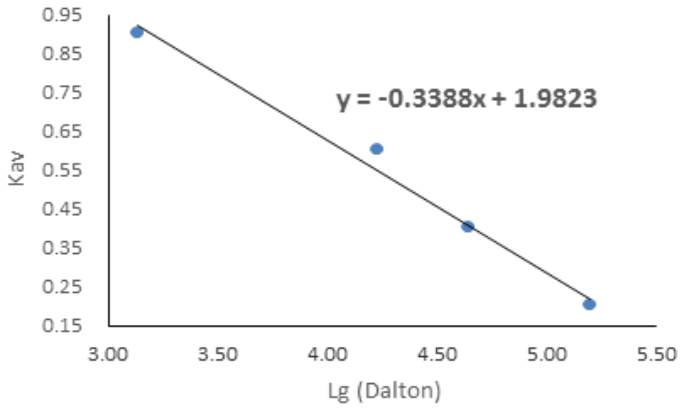
```
-19 MGSSHHHHHH SSGLVPRGSH MRSSADYCHI SGAERLAPAT ELPQLASAMT
 31 SRALHHDKGR PDTIHITVDK IEESTISTVP ALTPFLESNS SPGDARKLIA
 81 QRLHAAGIQA ADIAAEMAYS LTGLRGAALI DSSSGERLDP NPARGVRVST
131 FDAISHPSKD CAKDHFHEAL ILASKVHSAP GIVAEICLSD DPFYTRGYLA
181 LDGFFHRIPN IKDHGSTLGT RIFIVEPDTD IPELIDYLEN TPVYIELPPD
231 ASSSTDTTGL SSDLSAIAAQ RNTAWAGAGL TRTLRTFETA QLPHSRIDGA
281 DYLLFSSSDY LGLSTHPELV SAATAAIGHF GTGSGGSRLT TGTSIHSALE
331 SELAQFFGFD DAVLFATGYQ ANHSTIAAIA TADVEIFSDA ANHASIIDGC
381 RNARAKVTVF PHADYQTLDR LLATSSARHK LVISDSVFSM SGEVIDGPAL
431 ERTCRRRNAW LMLDDAHGVG VIGEQGRGTA AHLDIRPDIV VGTASKALGV
481 EGGYVLCSAP VGELLRNQAR SFVYSTSMNP GSVAAIRAAL KQLEVGDVVK
531 RLQRNIARVL SLVGAQSDPA SAIIPLPVGD ETEAMDISAQ LAELGVFIPA
581 IRYPTVPRGE AMLRLTITAL HTDADIDQLE LALRNTGLL
```

(N-terminal His_6 tag is in orange)

Appendix 5.8 Calibration curve of the Old superdex200 column

OLD S200

Standars	Ve	Vo	Vt	Mr (Dalton)	Kav	Lg (Dalton)
gamma-globulin	60.80	46.00	120.00	158,000.00	0.20	5.20
ovalbumin	75.60	46.00	120.00	44,000.00	0.40	4.64
myoglobin	90.40	46.00	120.00	17,000.00	0.60	4.23
vitamin B12	112.60	46.00	120.00	1,350.00	0.90	3.13



Kav=-0.3388Lg(M_r)+1.9823
Kav=(Ve-Vo)/(Vt-Vo)

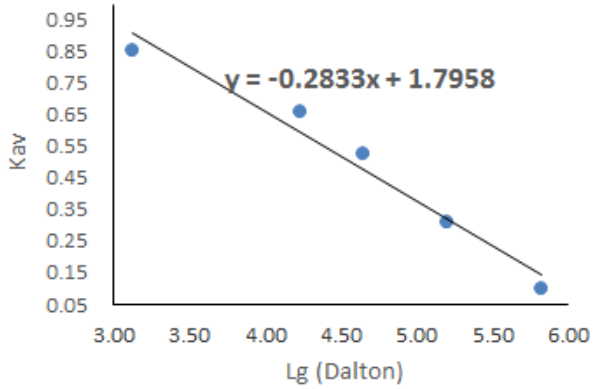
Mr=10exp {[(Ve-46)/(120-46)-1.9823]/-0.3388}

Sample	Ve	Vo	Vt	Kav	Xsample (lg)	Mr (Da)	subunit Mr (Da) form Vector NTI	Mr/subunit Mr
PCAS/NTH	73.40	46.00	120.00	0.37	4.76	57,276.61	29,626.54	1.9
PCAS/Nhis	72.20	46.00	120.00	0.35	4.81	63,961.81	31,660.69	2.0
PCAS/UT	74.12	46.00	120.00	0.38	4.73	53,641.19	23,522.07	2.3
BioWF	64.40	46.00	120.00	0.25	5.12	130,928.65	67,884.09	1.9
PCAS R11A	80.71	46.00	120.00	0.47	4.68	48,215.17	29,541.43	1.6
PCAS R13A	78.60	46.00	120.00	0.44	4.78	60,789.91	29,541.43	2.1

Appendix 5.9 Calibration curve of the New superdex200 Column

NEW S200

Standars	Ve	Vo	Vt	Mr (Dalton)	Kav	Lg (Dalton)
vitamin B12	108.65	41.94	120.00	1,350.00	0.85	3.13
myoglobin	93.31	41.94	120.00	17,000.00	0.66	4.23
ovalbumin	83.09	41.94	120.00	44,000.00	0.53	4.64
gamma-globulin	66.35	41.94	120.00	158,000.00	0.31	5.20
Thyroglobulin	49.88	41.94	120.00	670,000.00	0.10	5.83



$Kav=-0.2833Lg(M_r)+1.7958$

$Kav=(Ve-Vo)/(Vt-Vo)$

$Mr=10^{\exp\{[(Ve-41.94)/(120-41.94)-1.7958]/-0.2833\}}$

Sample	Ve	Vo	Vt	Kav	Xsample (lg)	Mr (Da)	subunit Mr (Da) form Vector NTI	Mr/subunit Mr
SeMet/PCAS/NTH	76.49	41.94	120.00	0.44	4.78	59,795.50	30,049.20	2.0
PCAS Y199F	77.72	41.94	120.00	0.46	4.72	52,591.18	29,610.64	1.8
PCAs Y211F	78.23	41.94	120.00	0.46	4.70	49,871.34	29,610.64	1.7
PCAS R213A	78.29	41.94	120.00	0.47	4.70	49,560.75	29,541.43	1.7
PCAS R227E	77.22	41.94	120.00	0.45	4.74	55,401.64	29,599.47	1.9
PCAS R227K	77.05	41.94	120.00	0.45	4.75	56,391.02	29,598.53	1.9
AONS/CHis	72.20	41.94	120.00	0.39	4.97	93,438.12	42,657.00	2.2

Appendix 5.10 1H NMR (1H, 500Hz) of pimeloyl-CoA produced by PCAS/NHis reaction