Author:
Akter, Nahida

Title:
Structural and mechanistic investigations of the key esterification step in Mupirocin and Thiomarinol biosynthesis.

Structural and Mechanistic Investigations of the Key Esterification Step in Mupirocin and Thiomarinol Biosynthesis.

Nahida Akter

A thesis submitted to the University of Bristol as a part of the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

University of Bristol

School of Chemistry

Cantock's Close

Bristol, BS8 1TS

April 2023

Abstract

Polyketide synthases encode a distinct number of enzymes which catalyse different biological reactions i.e., priming, chain elongation, reduction and tailoring reactions and produce a remarkable number of structurally diverse polyketide products from bacteria, fungi, and some plants. Mupirocin and thiomarinol are polyketide antibiotics produced by *Pseudomonas fluorescens* and *Pseudo-alteromonas* sp. SANK73390, respectively. Both of these natural products contain a polyketide backbone esterified with a fatty acid derived side chain. The esterification of these two components (via the free carboxyl group of the polyketide and a hydroxyl group present on the fatty acid) represents a key biosynthetic step but the enzyme that catalyses this reaction is unknown. MupB/TmlB are potential candidates for catalysing this reaction and share 37% sequence identity with each other. The aim of this work has been the mechanistic and structural characterization of both enzymes from these pathways.

Cloning, expression, purification, and crystallization of MupB and TmlB allowed the determination of high-resolution crystal structures of both enzymes (2.4 Å and 1.5 Å respectively) which revealed a thiolase-fold common to the FabH like enzymes (KAS-III) but containing an alternate catalytic triad comprised of Cys-His-Asp rather than Cys-His-Asn typical of KAS-III enzymes. The orientation of these catalytic residues was non-canonical but matched a single example (CerJ) from the cervimycin pathway that has been proven to catalyse an esterification reaction. Extensive structural analysis revealed that MupB/TmlB possesses two alkyl channels along with a phosphopantetheine binding channel required to bind the substrates for the esterification reaction.

In vitro self-acylation assays confirmed that TmlB selectively bound branched substrates (pre-monic acid mimic, 3-methylbut-2-enoyl (MBE) although MupB non-selectively self-acylated with branched and nonbranched substrates to a low extent. Site directed mutagenesis of the proposed active site cysteine of MupB (C116A) resulted in a surprising increase in self-acylation activity, suggesting the presence of a surrogate nucleophile which is proposed to be a neighbouring serine residue (S224). A co-crystal structure of TmlB in complex with MBE was also solved to 1.59 Å that not only confirmed the presence of hydrophobic alkyl channel but suggested the potentially incorrect orientation of this substrate mimic may indicate the requirement of a more authentic polar substrate as well as cognate ACPs for a successful esterification reaction.

^{1}H-^{15}N HSQC NMR screening experiments showed that only the modular MmpE_ACP, MmpB_ACP5 and ACP7 interacted with MupB and are predicted to deliver the hydroxylated fatty acid chains to the esterification reaction. The *in vitro* acyltransferase assay of MBE loaded M/TmpE_ACP confirmed MmpE_ACP as the cognate ACP to deliver the polyketide substrate to MupB. Both MupB and TmlB efficiently hydrolysed acyl groups off the ACP to produce holo-ACP which interestingly covalently complexed with MupB/TmlB. Covalent complex formation was far more efficient with MBE present compared to mixing free proteins and may be a future route to capturing an elusive ACP:KAS-III type complex for structure determination.

Although both cognate ACPs for this reaction have been identified, extensive *in vitro* esterification assay attempts with varied substrates failed to produce any detectable esterified product in presence of acylated MupB/TmlB, producing consistent hydrolysis of 9-hydroxynonanoic acid mediated M/TmpB_ACPs to produce holo-ACPs. Nonetheless, MupB and TmlB remain strong candidates for mediating esterification but appears to suggest that both enzymes have a strict requirement for the presence of a more authentic polyketide substrate.

Acknowledgements

I feel very lucky to have spent three years of my life in Bristol, UK. During this time, I was working on a project which was beyond my chemistry knowledge but started to love the project as soon as I started to learn about this. I have been very fortunate to conduct research on this fascinating project surrounded by very talented people. Firstly, I would like to thank my supervisor Professor Matt Crump for all his guidance and support throughout my PhD, particularly during the last year at 2020 (during Covid 19 breakout) and thesis writing period. I can still recall receiving a copy of mupirocin biosynthesis pathway at our very first project meeting and I was thinking, "What have I gotten myself into"! His constant support and encouragement (especially after getting the first protein crystal) throughout the project make my PhD life easier and enjoyable. His extraordinary help and patience (specifically during my writing period after pregnancy) helps me to grow into a more diligent worker over the years. I have been very grateful for being supervised by someone who is as enthusiastic about the project as you. I would like to thank him specially for helping me during the writing period as I was mentally and physically disturbed due to Covid-19, family issues and pregnancy. I was also fortunate enough to travel to Portugal, Exeter and Warwick for conferences during my PhD for which I am very grateful.

I would also like to thank Professor Chris Willis for providing continued support and guidance to improve my knowledge on mupirocin biosynthesis pathway through our APM meetings and regular mupirocin meetings. I would like to thank specially Dr. Paul Walker for teaching me the first biological laboratory procedures regarding protein expression, purification and characterisation. I have been also grateful to Alice Parnell for helping me in protein crystallography.

I want to express my thanks to everyone involved in the mupirocin project for their assistance throughout my time in lab. A very special thanks to Dr. Ash Winter for his constant support, guidance and ideas in this work. He also provided continuous help throughout the project and helps me for planning which move the project forwards for which I am very grateful. I would also like to Dr. Matt Rowe- we started together in this lab, shared a lot of ideas, teaching and beautiful memory throughout my PhD life. Thanks for being there whenever I was homesick and/or struggling with my work in lab. Thanks to Dr. Chris Williams for your support on NMR and X-ray crystallography things which were vital to the completion of this project. I would

also like to thank Tom Simpson, Song, Luoyi and everyone else on the mupirocin project over the years who have made my work possible.

I would like to thank all members of N318 for their support over the years and for all the fun times (Lab quiz, Christmas dinners, Karaoke trips etc.). Thanks must go to Tom, Alice, Erik, Catherine, Marija, Goodwin, Lina, Amy, Nui. Thanks to undergraduate project student Toby Seddon for his contribution in this project. I would also like to thank Dr. Angus Weir for his contribution on providing the synthetic substrates used in this work, without which this project would have been impossible. Thanks to Dr. Chris Arthur and Dr. Paul Gates for their support with Mass Spectrometry throughout this work. Finally, I would also like to thank my housemates Shaini, Basiram, Samia, Sami and everyone else for being constant supportive in every aspect.

To my parents, thank you for always supporting me. The encouragement and patience you have had with me over the years has been extraordinary. This piece of work would not be possible without your persistence and sacrifice (Specially Amma) to allow me to follow this pathway into science.

Finally, the last acknowledgement must go to my husband Dr. S. M. Nizam Uddin. Thank you for your eternal support throughout the last 8 years and I know I would not be at this point without you. Words cannot express how grateful I am. And lots of love and thanks to my brave daughter Manha who lived without me for last five months so that I can complete my thesis. I'm sure both of you will now be happy more than anyone else. Thank you for being part of every step of this journey from the start of CSC scholarship to this point and I can't wait to take the next steps with you in our journey through life.

COVID-19 Statement

The last six months of the funded PhD, from March 2020 to September 2020, was significantly disrupted by COVID-19 and allowed for very little laboratory work to be completed. There are several experiments such as structure, function and mutagenesis studies were only partially completed. Subsequent financial and personal reasons dictated that I returned to Bangladesh in October 2020.

Authors Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. The work described in this thesis was carried out in the School of Chemistry, University of Bristol under supervision of M. P. Crump between October 2017 October 2022. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author and not of the University of Bristol.

Nahida Akter

April 2023

Abbreviations

AMPS	2-Acrylamide-2-methylpropanesulfonic Acid
HMG	3-Hydroxy-3-methylglutaryl
3-HP	3-Hydroxypropanoate
MBE	3-Methylbut-2-enoyl
4-HB	4-Hydroxybutyrate
6-MSA	6-Methyl Salicylic Acid
8-HO	8-Hydroxyoctanoic Acid
9-HN	9-Hydroxynonanoic Acid
AMR	Antimicrobial Resistance
ASA	Accessible Surface Area
ACP	Acyl Carrier Protein
AT	Acyl Transferase
ATP	Adenosine Triphosphate
ASU	Asymmetric Unit
bp	Base Pair
BCA	Bicinchoninic Acid
BLAST	Basic Local Alignment Search Tool
BSA	Buried Surface Area
CSP	Chemical Shift Perturbations
CoA	Coenzyme A
CID	Collision-Induced Dissociation
CV	Column Volumes
conc.	Concentration
Da	Dalton
DCC	Decarboxylative Condensation
DH	Dehydratase
DM-CHL	Desmethylsalicyl-CHL
Fast DP	Diamond Pipeline
DMSO	Dimethyl Sulfoxide
DTT	Dithiothreitol
ESI	Electrospray Ionisation
ESI-MS	Electrospray Ionisation Mass Spectrometry
ER	Enoyl Reductase
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast Performance Liquid Chromatography
FAS	Fatty Acid Synthase
FA	Formic Acid
HSQCS	Heteronuclear Single Quantum Coherence Spectroscopy
HPLC	High Performance Liquid Chromatography
IMAC	Immobilized Metal Ion Affinity Chromatography
IRMPD	Infrared Multiphoton Dissociation
Insol.	Insoluble
IleRS	Isoleucyl tRNA Synthase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
I-TASSER	Iterative Threading Assembly Refinement
KAS	Ketoacyl Synthase
KR	Ketoreductase
KS	Ketosynthase
LCMS	Liquid Chromatography Mass Spectrometry
LB	Lysogeny Broth
MS	Mass Spectrometry
MRSA	Methicillin-Resistant *Staphylococcus aureus*
MR	Molecular Replacement
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
MSA	Multiple Sequence Alignment
Mmp	Mupirocin Multifunctional Protein
NRPS	Non-Ribosomal Peptide Synthetase
NMR	Nuclear Magnetic Resonance
ORF	Open Reading Frame
OD	Optical Density
Pant.	Pantetheine
PP	Phosphopantetheine
PPTase	Phospho-Pantetheine Transferase
PKS	Polyketide Synthase
PCR	Polymerase Chain Reaction
Post-D	Post-Desalt
Post-I	Post-Induction
Pre-D	Pre-Desalt

Pre-I	Pre-Induction
PQS	Probable Quaternary Structure
PDB	Protein Data Bank
PPIs	Protein-Protein Interactions
PA	Pseudomonic Acid
RPM	Revolutions Per Minute
RMSD	Root Mean Square Deviation
SEC	Size Exclusion Chromatography
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SOC	Super Optimal broth with Catabolite repression
THP	Tetrahydropyran
TEMED	Tetramethylethylenediamine
TE	Thioesterase
Tmp	Thiomarinol Multifunctional Protein
TCEP	Tris (2-CarboxyEthyl) Phosphine
Tris	Tris(hydroxymethyl)aminomethane
TAE	Tris-acetate-EDTA
WT	Wild Type
WR	Working Reagent

Table of Contents

1. Introduction

The discovery of penicillin by Sir Alexander Fleming in 1928 from *Penicillium ascomycetous* fungi was a ground-breaking achievement for 20th century science.[1] This antibiotic has since saved millions of lives from microbial infectious diseases all over the world. Antibiotics are most often derived from natural products which are produced by microorganisms as so-called secondary metabolites during starvation. These metabolites are non-essential but are advantageous for protection and survival.[2] An antibiotic can act against bacterial infections, for example, by destroying the cell wall, inhibiting protein biosynthesis or by disrupting folic acid or nucleic acid synthesis.[3] The discovery of penicillin initiated the golden age of new antibiotic discovery between 1940-1960, amongst which 70% were derived naturally. Although medicinal and biological chemistry has driven subsequent investigation and optimization of these initial discoveries, no new class of antibiotics has been reported since 2000.[4,5] Meanwhile, the inappropriate and overuse of antibiotics has driven the widespread development of antimicrobial resistance (AMR) to nearly all available therapeutics, leading to the grim prospect of threats to healthcare, security, and economics as reported by the Review on Antimicrobial Resistance, 2016.[6] Unfortunately, whilst antibacterial resistance is increasing, the development or discovery of new antibiotics is declining. This is therefore of utmost importance. The development of new effective antibiotics might be achieved by producing potential derivatives which are clinically related to the natural antibiotics by using modern synthetic chemistry.[7] In another approach, new entities may be generated by understanding and rational engineering of antibiotic biosynthesis pathways.[8,9]

1.1 Primary and secondary metabolites

Every living organism requires a selection of primary and secondary metabolites to live and to promote survival and reproductive fitness. The primary metabolites such as amino acids, lipids, carbohydrates etc., are essential to sustain microbial lives. These are involved in primary metabolic pathways such as glycolysis and the citric acid cycle where glucose is first oxidised to release energy and produce pyruvate. The pyruvate can then be oxidised to acetyl-CoA which under the second stage of oxidation produces energy, stored as ATP.[10,11]

All these primary metabolic pathways are fundamentally similar in all organisms despite being characteristically different (i.e., movement, growth, reproduction, nutrition etc). Secondary metabolites on the other hand are biosynthesized by a certain set of organisms or cells (e.g., bacteria, fungi, plant) and are not involved directly in normal growth, development, and reproduction. For many years these natural compounds were thought to be functionless or metabolic side-products, but during the nineteenth and twentieth centuries secondary metabolites were shown to prevent the growth and replication of microbes.[3,12] Secondary metabolites can be classified based on their biosynthetic origin (Figure 1.1). For example, polyketides and fatty acids have distinct properties but are related in terms of their biosynthesis by parent organisms. This vast pool of structurally diverse bioactive organic molecules has high interest for their defensive actions against microorganism infections. These highly functionalized products can be biosynthesized from simple organic acids.[13]
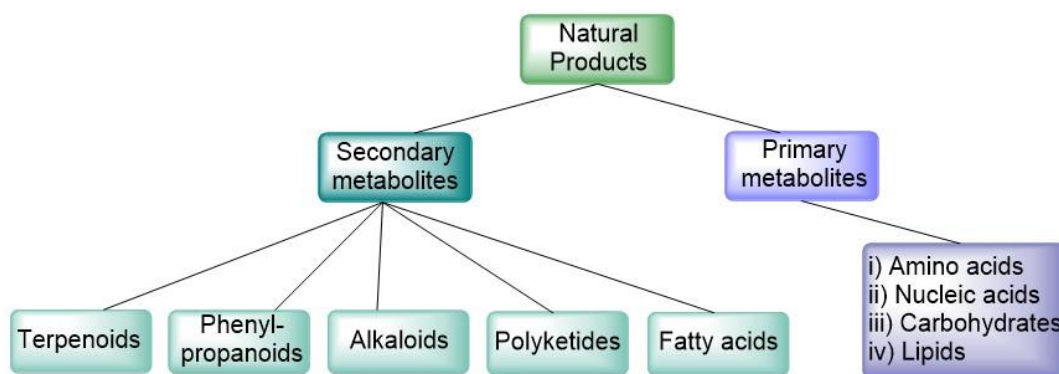
Figure 1.1. Classification of natural products based on whether they are involved in primary metabolism or not. Secondary metabolites are commonly classified based on their diversity in structure, function and biosynthesis origin.

1.2 Fatty Acids

Fatty acids are a simple chemical compound, composed of a carboxylic acid group attached to the end of a long aliphatic chain, which is either saturated or unsaturated. These are not found directly in nature but instead exists as forms of esters: triglycerides, phospholipids and cholesterol esters.[14] In living cells, fatty acids are an important dietary source which acts as fuel for muscular contraction and general metabolism. The biosynthesis of fatty acids in living cells is carried out by large multifunctional proteins, termed fatty acid synthases that follow the stepwise Claisen condensation reaction of a simple pool of start units (e.g., acetate) with extender units (e.g., malonate).[15] Some examples of fatty acids are shown in Figure 1.2.
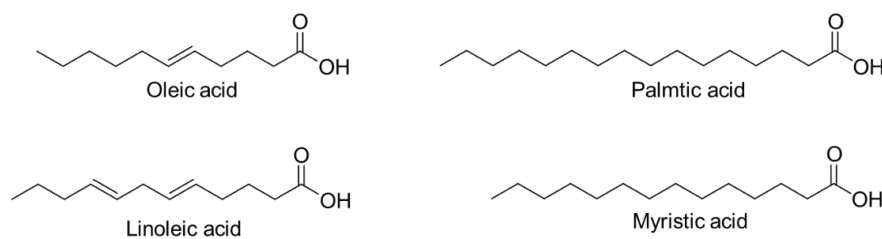
Figure 1.2. Selected molecular structures of fatty acids.

1.3 Polyketides

Polyketides are a structurally diverse and pharmacologically interesting class of secondary metabolites produced by living systems such as bacteria, fungi and plants. These compounds display a wide array of use in medical treatment as antibiotics,[16] antifungal,[17] anticancer,[18] and immunosuppressant[19] (Figure 1.3). The "ketide group" – CH$_2$-CO- was first reported by John Collie in 1907 which he determined to be the building block of the polyphenols while exploring the synthesis of orcinol from dehydroacetic acid.[20] Later, in the 1950s Arthur Birch rejuvenated Collie's findings and discussed that the aromatic polyketide 6-methyl salicylic acid (6-MSA, produced naturally by *Penicillium patulum* fungus) could be generated by repeated condensation reactions of simple acetate unit. This was further elegantly demonstrated by the incorporation of isotopically labelled precursors and subsequent analysis by chemical degradation.[21] This led to the concept of polyketide biosynthesis from simple organic molecules (i.e., acetyl-CoA, malonyl-CoA).
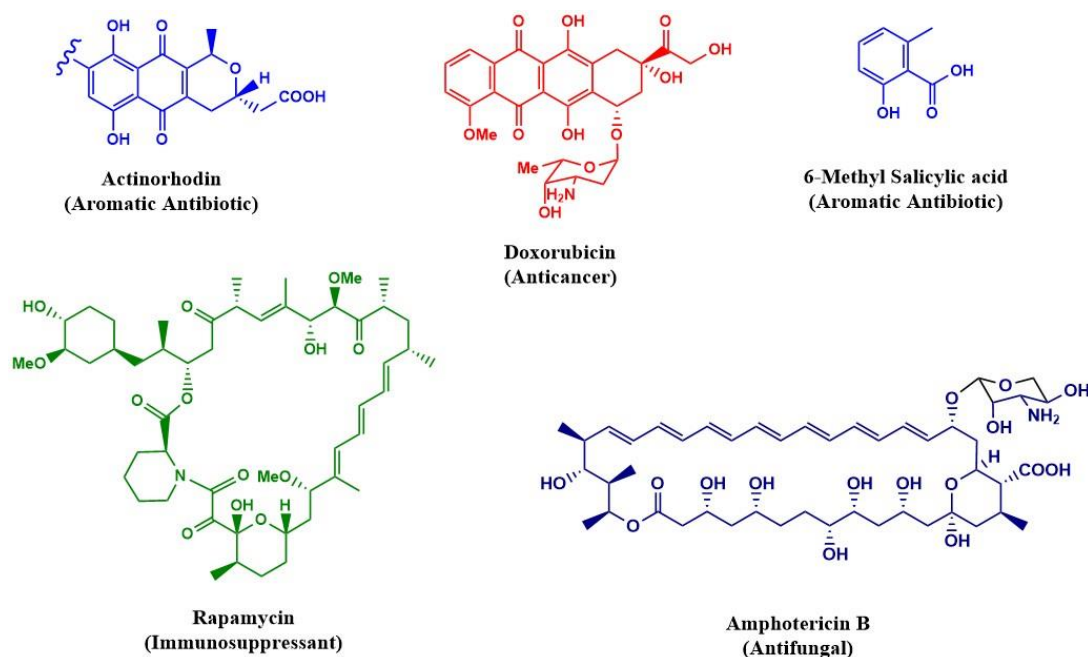
Figure 1.3. Molecular structures of some biologically active polyketides where the functions are indicated.

1.4 Biosynthesis of polyketides and fatty acids

The polyketides and fatty acids share similar biosynthetic processes which involves iterative decarboxylative Claisen-like condensation reactions. Fatty acid biosynthesis requires the FAS multienzymes whereas polyketide biosynthesis is carried out by dedicated polyketide synthases (PKS). Both processes are initiated by the transfer of an acyl-CoA starting unit to an acyl carrier protein (ACP), catalysed by an acyl transferase (AT) enzyme, and then transfer from the ACP onto the thiol group of a ketosynthase (KS) domain, (Figure 1.4).[22] The process then proceeds with the transfer of a malonyl extender unit to an ACP which is catalysed by a malonyl specific transferase enzyme (MAT or MCAT). The acyl starter unit then undergoes a decarboxylative Claisen condensation reaction with the malonyl extender unit to produce a β-keto moiety which contains an additional two carbon units. At this point the β-ketone product may further be processed reductively by a ketoreductase (KR) to form a β-hydroxyl group, which may then be dehydrated by a dehydratase (DH) to yield an α,β-unsaturated thioester. Finally, an enoyl reductase (ER) enzyme can produce a saturated alkyl chain by reducing the double bond. The required chain length of the polyketides and fatty acids can be achieved by repeating this cycle until the desired product is synthesized.
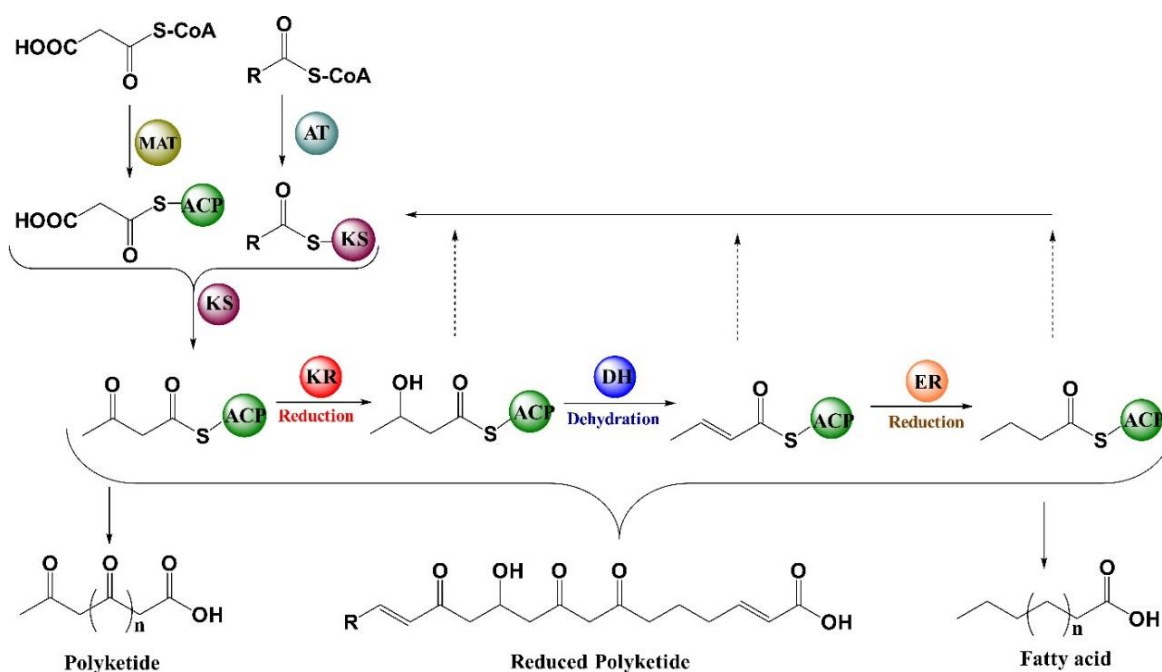
Figure 1.4. General reaction scheme depicting the fatty acids and polyketides biosynthesis. The reaction initiates with acetyl-CoA start unit and malonyl-CoA extender unit and undergoes subsequent condensation reaction. A full reductive process provides saturated fatty acid, whereas structurally diverse polyketides can be produced by the partial or full reduction process.

1.4.1 Differences between FAS and PKS

The main difference between fatty acids and polyketides biosynthesis is that PKSs may produce a diverse array of molecules by using any number of processing domains in a precisely controlled biosynthetic cycle. In PKS, the reduction steps may be used partially or completely deleted in a controlled fashion, thus generating greater structural diversity[23] where the chain termination can be achieved by the thioesterase (TE) domain. Meanwhile, in FAS biosynthesis system, the fatty acid assembly involves few choices, with the main variables being fatty acid length and degree of saturation and branching.[24–26] PKSs can also use a variety of substrates as extender units for example methyl-malonyl CoA and propionyl CoA.[27] Using different starter units can impart important structure and functions to the final polyketide. For example, β-methyl branches of the polyketide antibiotic erythromycin play an important role in binding the 50S ribosomal unit to inhibit bacterial protein synthesis.[28,29] Contrary to this, FAS typically uses malonyl CoA as a main extender unit which restricts the production of structurally diverse compounds.

1.5 Classifications of FAS and PKS

1.5.1 Classification of FAS

Based on the structural arrangements of domains, the FAS multidomain enzyme systems are mainly classified into two major groups: type I and type II (Figure 1.5). Type I FASs (typically found in animals) are comprised of a cluster of covalently linked catalytic domains, necessary for fatty acid biosynthesis that form a single polypeptide chain capable of chain extension and processing.[23,30] Structurally, type I FASs are composed to two identical polypeptide chains which create two active centres for the biosynthesis at the homodimer interface.[26,31] Substrates are shuttled as ACP bound substrates iteratively through each domain on the synthase, coupled with significant domain rearrangements that ensure access to catalytic sites of the domains.[32,33] Conversely, type II FASs (found in plants and most bacteria) are composed of separate enzymes with distinct catalytic functions and which likely exist as a non-covalent multienzyme complex.[25,34] In type II FASs the same basic reactions are carried out by individual monofunctional enzymes for biosynthesizing fatty acids.

1.5.2 Classification of PKS

Similar to FASs, PKSs are also divided into three main groups, types I, II and III based on their architecture and reaction mechanism. Type I PKSs exist as large multifunctional proteins (e.g., erythromycin synthase) whereas, type II PKSs are referred to as multi-enzyme complexes (e.g., actinorhodin polyketide synthase) where mono-functional domains exist as discrete proteins to achieve polyketide biosynthesis.[35] Type III PKSs (e.g., RppA synthase for flavolin biosynthesis) are homo-dimeric enzymes that act as condensing enzymes (Figure 1.5). Type I and type II PKSs utilises ACPs to activate the acyl CoA substrates and to channel the growing polyketide intermediates, whereas type III lack ACP and act directly on the acyl CoA substrates.[30]

1.5.2.1 Type I PKS

Type I PKSs consists of multiple FAS-like ensembles or modules that are responsible for biosynthesis a wide range of structurally diverse polyketides. This is further classified into iterative (lovastatin nonaketide synthase) and non-iterative/modular (deoxy-erythronolide B synthase, DEBS) types.[36,37] The Iterative PKS system is comprised of the repeated use of a set of catalytic domains which can be utilised multiple times in a biosynthetic pathway to achieve the desired products.[38] Each set of catalytic domains are required for catalysing 2-carbon elongation/reduction cycles where ACP repeatedly interacts with all the domains until the desired chain length has been achieved. The modular/non-iterative PKSs consists of multiple covalently linked catalytic modules, each one is required for condensation and reduction of the 2-carbon acyl starting unit prior to passing to the downstream module via KS. In modular PKSs each catalytic domain is used usually only used once although there are examples of exceptions to this rule.[27,39] For example, multiple iterations in type I modular aromatic PKSs (iPKS) yields monocyclic and higher order aromatic polyketides. Moreover, programmed iteration of a single module can be invoked by partially iterative PKSs where the polyketide intermediated passes back to the upstream KS by a downstream ACP for additional rounds of elongation.[40,41]

Further sub-division of type I modular PKSs are based on the presence of the ACP priming domain; acyl transferase (AT), in the system which is required for the supply of extender units. When multiple AT domains are embedded into the modular system it is known as cis-AT whereas, the presence of usually free-standing ATs is the hallmark of trans-AT modular type I PKSs.[13,42] The trans-AT systems have a different domain architecture compared to

their cis-AT counterparts. In each module of a cis-AT system an AT domain is covalently attached to it, responsible for loading substrates to the ACP of the same module. Whereas trans-AT systems have discrete non-covalently bound AT domains that allow the substrates access to the correct ACPs from all modules as shown in Figure 1.5. The inspection of domain order within the modules of cis-AT systems can determine the outgoing possible polyketide products, known as the co-linearity rule.[43,44] However, trans-AT systems show much complexity in domains which may feature non-elongating KS domains, domain activity in several modules or split modules.[42] Therefore, trans-AT PKSs can produce natural products, and exhibit a significantly higher structural diversity.

1.5.2.2 Type II PKS

Type II PKS consist of discrete monofunctional proteins which work together to create a multifunctional PKS complex system. This is also called the 'modular PKSs'. The monofunctional proteins in a type-II PKS are not covalently bound like the type I FAS system, rather are associated in a highly complex multi-protein megasynthase structure (Figure 1.5). A subset of three domains, including a ketosynthase (KS_α and KS_β), acyltransferase and acyl carrier protein are essential for polyketide biosynthesis and represent the minimal type II PKS system, for example the actinorhodin (act) minimal PKS. In type II PKS condensation and reduction reactions take place using individual domains of this module before dissociating apart.[45,46] In this system, the growing acyl chain is transported among the individual enzymes by ACP although no common ACP binding motifs in this enzymes have been reported yet due to the transient nature of the protein-protein interaction.[47] Therefore, understanding the mechanism underlying ACP-enzyme interactions may prove crucial for logical bioengineering of these systems to develop new antibiotics.

1.5.2.3 Type III PKS

Type III PKSs are different from the type I and type II PKSs and are constructed by single ketosynthase domains which can exist as a dimer (Figure 1.5). The chalcone-stilbene synthase enzymes are examples of type III PKS, and they are mainly found in plants.[48] Although these PKSs were initially thought to be plant specific, subsequent studies have shown that they are also present in bacteria and fungi.[49,50] Unlike other PKS systems, type III PKSs do not necessarily need an ACP for shuttling the acyl chains among the domains and rather incorporate the substrate directly from CoA.[51,52]

Type I iterative FAS e.g. mammalian

Type I iterative PKS e.g. 6-methylsalicylic acid synthase (MSAS)

Type I modular cis-AT PKS e.g. 6-deoxyerythronolide B synthase 1 (DEBS1)

Type I modular trans-AT PKS e.g. Mupirocin multifunctional protein A (MmpA)

Type II FAS e.g. *Escherichia coli*

Type II PKS e.g. Actinorhodin minimal synthase

Type III PKS e.g. Chalcone synthase (CHS)
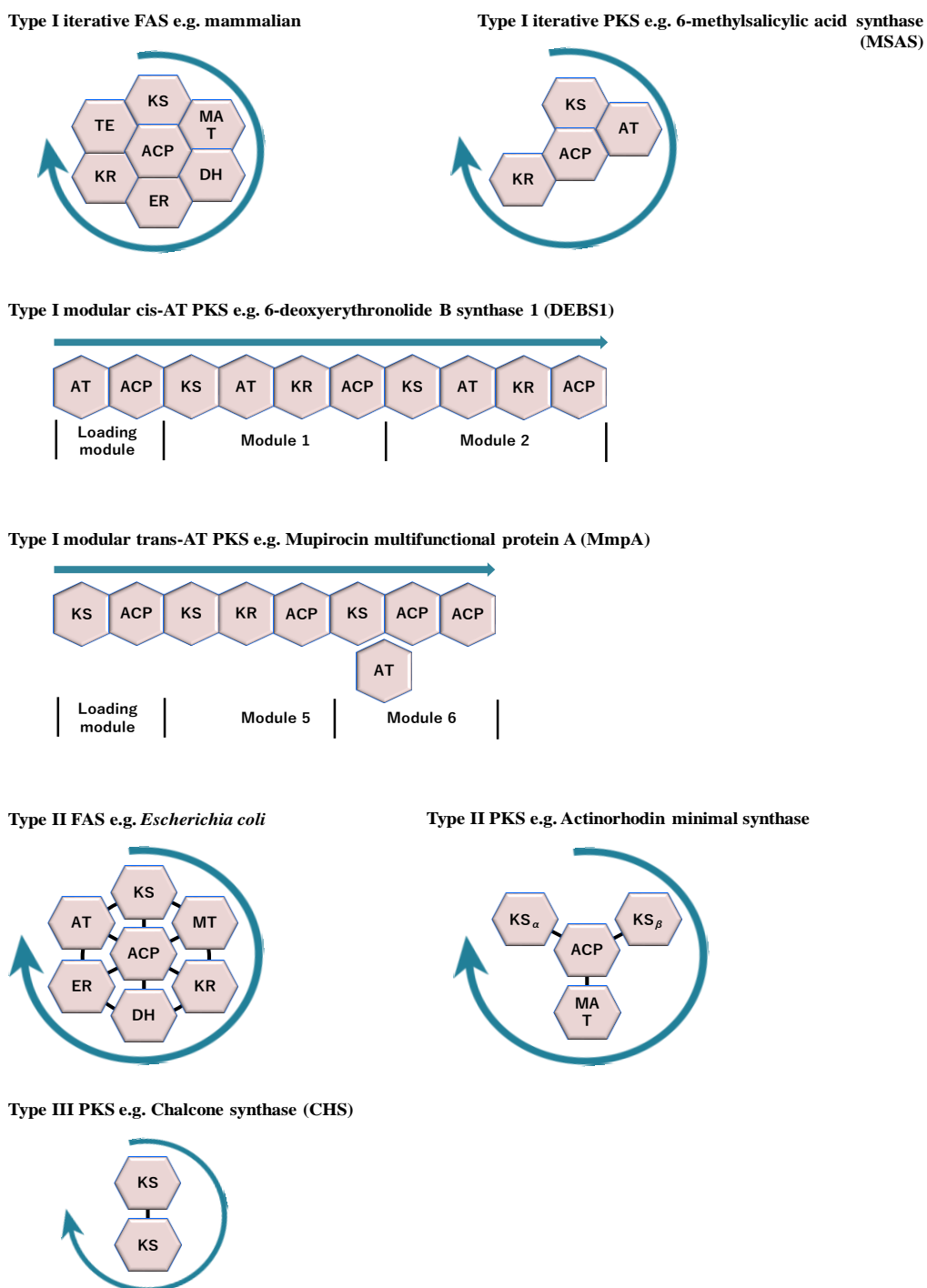
Figure 1.5. The PKSs and FASs classification diagram based on structural arrangements of different domains, are indicating by hexagonal blocks: MAT, malonyl acyltransferase; DH, dehydratase; KR, ketoreductase; ER, enoyl reductase; ACP, acyl carrier protein; TE, thioesterase.

1.6 Acyl carrier proteins

Acyl carrier proteins are essential for fatty acid and polyketide biosynthesis, as well as other discrete biological activity. These are small, globular proteins comprised of ~70-100 amino acid residues. ACPs are constructed from four α-helices (I-IV), three longer helices which are parallel to each other and one shorter helix. A hydrophobic motif (GXXS) at helix II is conserved in which the serine residue is thought to be responsible for protein activation (Figure 1.6). The ACP can be activated (apo to holo form) by post-translational addition of a CoA-derived 4′-phosphopantetheine (4′-PP) prosthetic group, which can bind the acyl/polyketide groups by forming a thioester linkage. Hence, these proteins play very important roles in shuttling the acyl moiety between the catalytic domains and direct specific protein-protein interactions.[47]
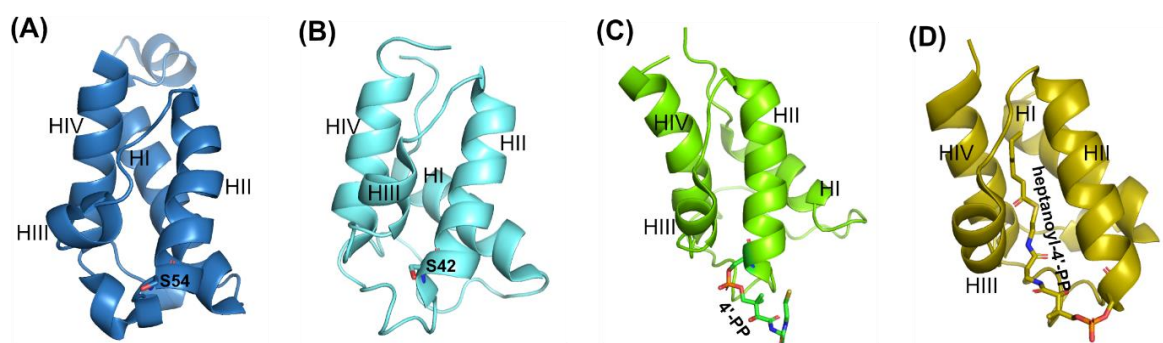
Figure 1.6. Three dimensional structures of ACPs. (A) Type I modular DEBS ACP (PDB:2JU2). (B) Type II actinorhodin ACP (PDB:2K0Y). (C) Holo-actinorhodin ACP (PDB:2K0X) where the 4′-PP chain is flapping near the binding site and (D) *E. coli* heptanoyl-ACP (PDB:2FAD) where the alkyl chain is buried inside the ACP structure. The α-helices are denoted as HI, HII, HIII and HIV.

ACPs can be found in two systems, (i) Type-I: as integral parts in multi-modular mega-synthase and (ii) Type-II: as discrete proteins in a multiprotein complex system. The Ppant arm of ACPs are observed to be in both sequestered and non-sequestered states, where type-I ACPs are not thought to sequester.[53] Di-domain and tri/tetra-domain ACPs are also discovered in many PKS and hybrid PKS/NRPS pathways and are known as tandem domains.[54,55] For example, di-domains and tri-domains are observed in the mupirocin biosynthetic pathway, where MmpA (multi-modular system) contains two domains encoding for ACPs and tri-domain ACPs from MmpB work in parallel to produce 9-hydroxynonanoic acid.[56,57]

1.7 Ketoacyl synthases (KSs)

Ketoacyl synthases (KSs), more officially known as β-ketoacyl or 3-oxoacyl synthases are the class of enzymes involved in catalysing the biosynthetic key step, the condensation reaction. The KSs are members of the thiolase superfamily and share a common catalytic fold. KSs exist as individual enzymes in type II PKSs and FASs which are essential for polyketide and fatty acid biosynthesis or may exist in type I PKSs/FASs multidomain systems.[23] They are responsible for catalysing condensation of an acyl-CoA/ACP with malonyl-CoA/ACP to produce β-ketoacyl CoA/ACP, by adding two carbon atoms in each condensation cycle to the growing acyl chain (Figure 1.4).

KSs are classified into five main families (KS1, KS2, KS3, KS4, and KS5) based on their primary structure similarities (using BLAST and MSA, Table 1.1).[58] Almost all KS1 family members are produced in bacteria, and few are found in eukaryotes. KAS III enzymes are dominant in this family, and they exist as a discrete protein in type II PKSs and FASs. These typically catalyse the first elongation step of the fatty acid biosynthesis and are thus also known as 'loading KASs'. The characteristic reaction is the decarboxylative condensation reaction to produce acetoacetyl-ACP from malonyl-ACP and acetyl-CoA. KS2s are found in eukaryotes and the 3-ketoacyl synthases are largest set of members comprising the fatty acid elongases and very long-chain fatty acid condensing enzymes. The KS3 is the largest KAS family and are present in the multidomain type I PKSs/FASs. Among many different enzymes in this family, KAS I and KAS IIs are the dominant group, and they utilise malonyl-ACP as the chain elongating agent and add to the ACP primed with a starter unit, most commonly acetyl, amongst others (Table 1.2). A large fraction of KS4 and all KS5 enzymes are produced by the eukaryotes. KS4 can be classified as chalcone synthases, stilbene synthases or type III PKSs and the KS5 family enzymes can be classified as fatty acid elongases. The typical reactions of these class of enzymes are shown in table 1.2.

Table 1.1. The list of KASs families and dominant enzymes in each family.[58]

Family	Produced by	Number of subfamilies	Examples of enzymes and genes present
KS1	Bacteria	12	KAS III - "β-ketoacyl-ACP synthase III" (FabH gene)
KS2	Eukaryotes (plants)	10	Fatty Acid Elongase (FAE) 3-ketoacyl-CoA synthase Very long-chain fatty acid condensing enzyme (VLCFA)

KS3	Bacteria and eukaryotes	14	KAS I – "3-oxoacyl-ACP synthase I" (FabB gene), KAS II – "3-oxoacyl-ACP synthase II" (FabF gene)
KS4	Bacteria and eukaryotes	10	Chalcone synthases (CHS) Stilbene synthases (STS)
KS5	Eukaryotes (animals)	11	Elongation of very long chain fatty acids (ELOVL), ELOVL1-ELOVL7 genes. Fatty acids elongation genes (ELO) ELO1-ELO6

Table 1.2. The typical reactions of the enzymes from different KAS family.[58]

Enzyme name	EC number	Characteristic reactions
β-ketoacyl-ACP synthase I	2.3.1.41	Malonyl-ACP + acyl-ACP → 3-ketoacyl-ACP + CO_2 + ACP
Naringenin-chalcone synthase	2.3.1.74	3-malonyl-CoA + 4-coumaroyl-CoA → naringenin chalcone + 3 CO_2 + ACP
Icosanoyl-CoA synthase	2.3.1.119	Malonyl-ACP + stearoyl-ACP + 2 NADPH +2 H^+ → icosanoyl-CoA + 2 $NADP^+$ + CO_2 + CoA + H_2O
β-ketoacyl-ACP synthase II	2.3.1.179	Malonyl-ACP + (Z)-hexadec-11-enoyl-ACP → (Z)-oxooctadeca-13-enoyl-ACP + CO_2 + ACP
β-ketoacyl-ACP synthase III	2.3.1.180	Malonyl-ACP + acyl-CoA → acetoacetyl-ACP + CO_2 + CoA

1.7.1 Crystal structures, catalytic triads and mechanisms.

Condensation is the characteristic reaction of the members of KASs enzymes. These enzymes are members of thiolase superfamily and share a common catalytic fold. All known crystal structures of KS1, KS3, and KS4 have a Thiolase-like fold containing a layered α-β-α-β-α tertiary structure.[59] No crystal structures of KS2 and KS5 have been reported to date. Based on the crystal structures and previous studies, catalytic triad compositions (Cys-His-Asn or Cys-His-His) are found conserved in the KS1, KS3 and KS4 families (Figure 1.7). These observations have led to the proposal that all members of the KS family employ a ping-pong kinetic mechanism, using cysteine, histidine or either histidine or asparagine as the catalytic residues.[58] For example, cysteine, histidine, and asparagine are reported to form the catalytic triad in KAS III (PDB: 1HN9) reported by Qui et.al.[60] They proposed that the C112 in the loop Nβ3-Nα3 donates a proton to the His_{244} and attacks acetyl-CoA to form self-acylated KSs. Then the malonyl-ACP interacts with H244 and N274, stabilizing a carbanion following decarboxylation which then attacks the acetyl moiety and undergoes condensation to form acetoacetyl-ACP (Figure1.8).[58]
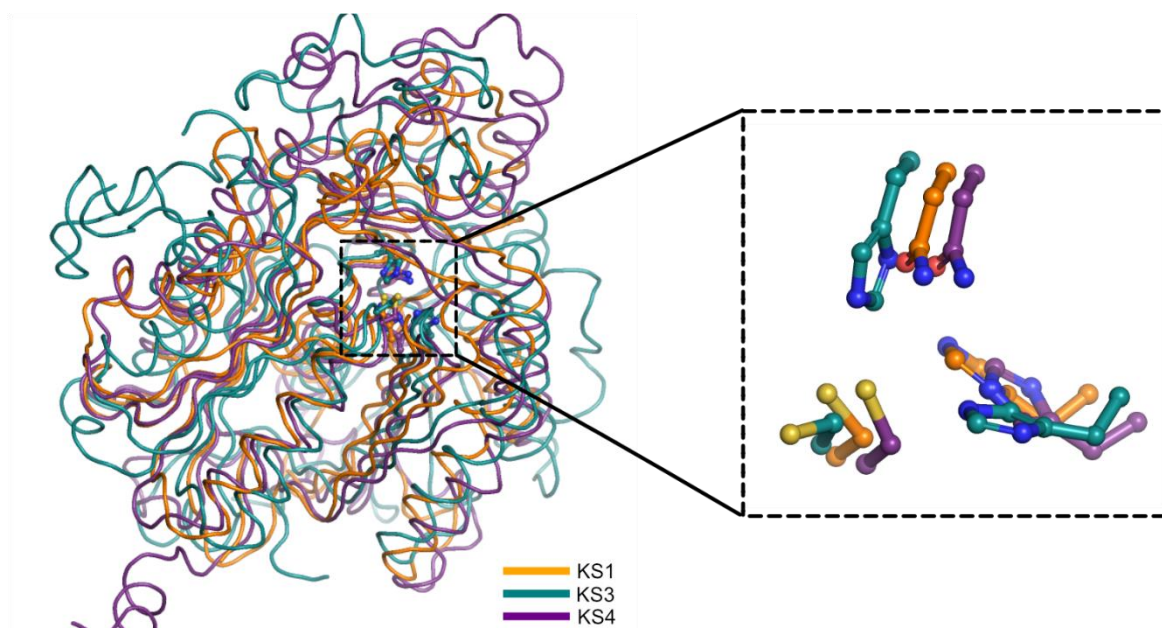
Figure 1.7. The superimposed crystal structures of KS1 (PDB: 1EBL), KS3 (PDB: 2QO3) and KS4 (PDB: 1Z1E) are shown in ribbon style (left side) and the catalytic active site orientations of the same proteins are presented in dashed line box (right side), where the residues in the bottom left are cysteine, bottom right is histidine and top are asparagine or histidine.
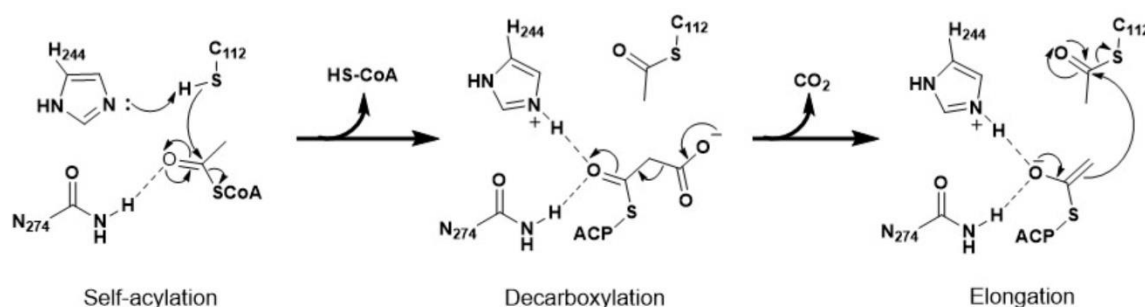
Figure 1.8. Proposed decarboxylative Claisen condensation reaction mechanism for β-ketoacyl-ACP synthase III, showing the self-acylation, decarboxylation and elongation steps to produce acetoacetyl-ACP.

1.7.2 KAS like enzymes

The KAS-like enzymes are structurally homologous to the KASs, however subtle changes to the active site residues imparts different functionality to the typical decarboxylation and condensation. Several KAS-like proteins have been reported over the past two decades, showing unexpected reaction mechanism among which, some are presented in Table 1.3. These enzymes have been shown to act as priming KSs,[61] 'reverse' esterases,[62,63] and acyltransferases.[64]

Table 1.3. List of some KAS like enzymes, have unusual catalytic triads and functions.

Proteins	Organisms	Classification	Catalytic triads	Function
ZhuH	*Escherichia coli*	KS	Cys-His-Asn	Priming KS
DpsC	*Streptomyces peucetius*	KS^S	Ser-His-Asp	Acyl transferase
CerJ	*Streptomyces tender*		Cys-His-Asp	Reverse esterase
ChlB6	*Streptomyces antibioticus*		Cys-His-Asp	Reverse esterase
CorB[65]	*Corallococcus coralloides*		Cys-His-Asn	Head-to-head condensation
PsyD[66]	*Psammocinia aff. Bulbosa*	KS^0	Cys-mutated His	Non-elongating ketosynthase

1.8 Mupirocin

Mupirocin is a natural polyketide isolated from *Pseudomonas fluorescens* NCIMB 10586 which is commonly found in plant-roots as a soil bacterium.[56,67] It has an antibiotic activity, discovered by C. Garrie in 1887[68] and almost a century later it was identified and isolated as a mixture of pseudomonic acids (PA-A to PA-D) where PA-A accounts for 90% of the mixture and antibiotic property.[69] The mupirocin molecular structure is comprised of a polyketide derived backbone, C_{17} monic acid (MA) portion esterified onto a fatty acid chain derived 9-hydroxynonanoic acid (9-HN) (Figure 1.9).
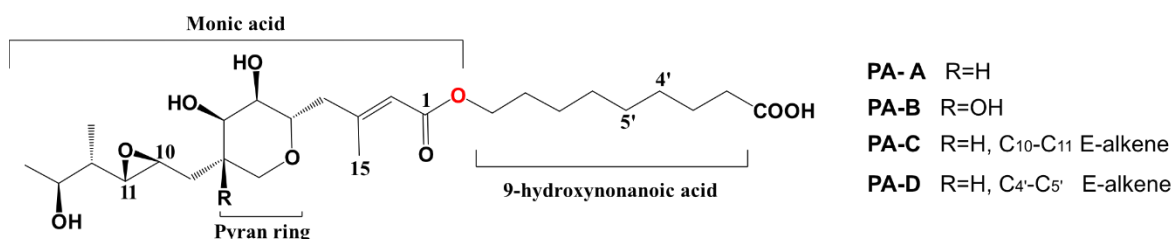
Figure 1.9. Molecular structure representing mupirocin, a mixture of pseudomonic acid (PA) A, B, C, and D, where the PKS-derived C_{17} monic acid part is linked to the 9-hydroxynonanoic acid (9-HN) via an ester linkage (red colour).

1.8.1 Clinical aspects of Mupirocin

Mupirocin is a commercially available antibiotic, active against gram-positive bacterial infections. It is mostly commonly used topically for skin infections caused by antibiotic-resistant *Staphylococcus aureus* and most significantly used to treat methicillin-resistant *Staphylococcus aureus* (MRSA) when other antibiotics are found to be ineffective. Mupirocin binds with the bacterial isoleucyl tRNA synthase (IleRS) which inhibits the bacterial protein and RNA synthesis and therefore prevents cell growth.[70] The epoxide bearing portion of the monic acid moiety of PA is thought to mimic the side chain of the isoleucine which can compete for the protein synthetic site of the bacterial enzyme. Further

the co-crystal structure of mupirocin bound IleRS (PDB:1JZS) revealed that the hydrophobic side chain of MA (C_{12}-C_{14} and C_{17}) is recognised by the Ile-specific binding pocket (P46, W518 and W558) via Van der Waals interactions. The 9-HN moiety (fatty acid) plays a complementary role in stabilising the IleRS:mupirocin complex and it adopts a conformation that prevents access to the active site (Figure 1.10).[71]

Mupirocin biological activity was reported as pH restricted and only observed to work within pH 4 to pH 9. Above or below this pH the antibiotic activity of mupirocin is abrogated due principally to opening of the epoxide ring. Further, the ester linkage can be readily hydrolysed by body fluids upon oral ingestion which completely inactivates the antibiotic.[72] The metabolic instability of mupirocin leads to poor bioavailability and is the principal reason for its restriction to topical use. The minor metabolite PA-C lacks the epoxide and is more stable *in vitro* whilst retaining comparable activity. A recently published work reported that a bioengineered mupirocin strain can produce mainly PA-C in large extent which is stable as a and efficient as PA-A.[73] Also, muta-synthesis between two different antibiotic biosynthetic pathways, mupirocin and thiomarinol produced a new family of hybrid derivatives exhibiting strong antibiotic activity.[74]
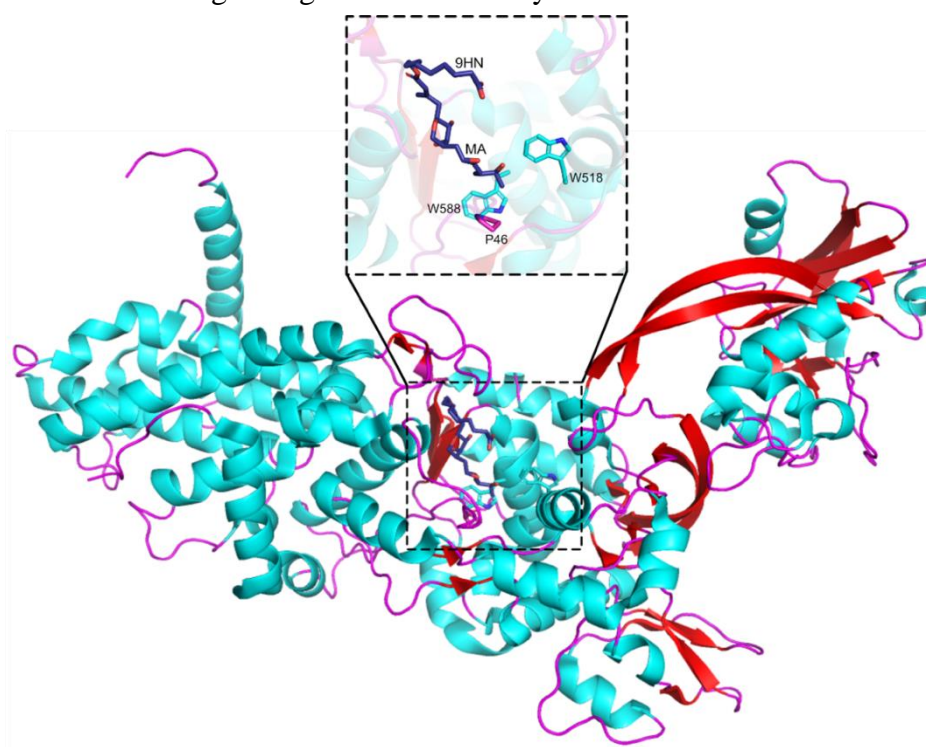
Figure 1.10. Three-dimensional (3D) structure of isoleucyl-tRNA synthetase from *Staphylococcus aureus* bound to mupirocin (purple colour), shown in stick. The mupirocin binds to the Ile-tRNA synthetic active site (shown inside the squared box with dashed line) thus inhibits the protein synthesis. The projected active site area is representing the hydrophobic pocket to co-ordinate the monic acid backbone of mupirocin. MA: Monic acid, 9HN: 9-hydroxynonanoic acid.

1.8.2 Mupirocin biosynthesis

Mupirocin is encoded as a type I *trans*-AT PKS/FAS of 75 kb gene cluster by transposon mutagenesis and subsequently analysed by gene knock-out experiments and DNA sequencing.[75,76] The gene cluster consists of 6 large open reading frames (ORFs), termed mupirocin multifunctional proteins (MmpA to MmpF) and 30 smaller ORFs (MupA-MupZ and MacpA-MacpE) are encoded which are tailoring enzymes (Figure 1.11).
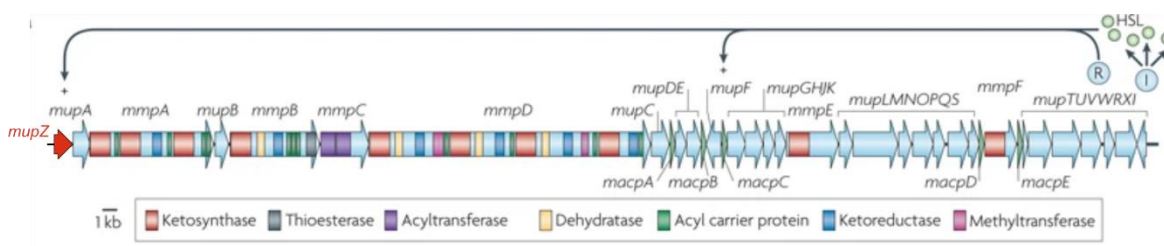
Figure 1.11. The mupirocin 'gene cluster' showing individual domains as coloured according to the key. The image is modified from C. Thomas *et.al.*[56]

Broadly, the general polyketide biosynthesis assembly mechanism is followed in mupirocin biosynthesis. An acetyl starter unit is primed onto the first KS of MmpD and malonyl-ACP undergoes a decarboxylative condensation with this complex to yield the first acetoacetyl intermediate. *In trans* acting AT enzymes are thought to prime the KS and ACP domains with acetyl and malonyl groups respectively although the details of this are poorly understood (e.g., MmpD lacks a GCN5-related N-acetyltransferase (GNAT) domain for malonyl selection/decarboxylation or a dedicated AT for unusual starter unit selection found in many other trans-AT systems).[13,77] Subsequently MmpD and MmpA further elongate a process the β-ketone which can be non-, partially or fully reduced by the actions of the KR, DH, and ERs supplemented by the actions of several *trans*-acting tailoring enzymes to produce pre-monic acid.

The proposed mechanism for mupirocin biosynthesis is represented in figure 1.12. The C_{17} pre-monic acid backbone synthesis is as stated assembled by MmpD and MmpA. The small module, MmpC is located immediately prior to MmpD in the gene cluster, and it contains two AT domains, suggesting the loading and transfer of activated acetyl starting and malonyl extender units onto MmpD. The C_{14} polyketide backbone is synthesized by six condensation reactions catalyzed by MmpD (D1-D4) and MmpA (A1-A3), where the first four are catalyzed by MmpD and two elongations in MmpA.[78,79] These are supplemented by three Cα methylations catalysed by *in-cis* methyltransferases (MT) and the action of

MupA that is predicted to perform the C6 hydroxylation. The insertion of a β-branch at C3 (MupGHJK) yields (a). Epoxidation by MmpE completes the fully tailored but non-cyclised pre-monic acid (b).[80]

The origin of the 9-hydroxynonanoic acid was initially investigated with feeding experiments using isotopically labelled (^{18}O, ^{13}C and ^{2}H) acetates to *Pseudomonas fluorescens* NCIMB 10586. These suggested priming with a three carbon starter unit to produce the elongated 9-hydroxynonanoic acid (9HN).[72] This starter unit has been shown to be 3-hydroxypropanoate (3-HP) in figure 1.12 which is produced from acetyl and malonyl CoA via the action of MacpD, MupQ and MupS. The C_3 starter unit can undergo three rounds of condensation with malonyl extender units by the joint action of MmpF and MmpB.[40] The TE domain on MmpB releases either the 9HN which is then esterified with the polyketide backbone or esterification occurs using one or all of the tri-domain ACPs present on MmpB and then the esterified product is released by the TE.

To date, several critical and non-linear mupirocin biosynthetic steps have been elucidated including THP ring formation,[81] β-branching,[82] epoxidation at C_{10}-C_{11} position.[83]. Despite this progress, the nature of the key esterification step between pre-monic acid and the fatty acid chain (unknown length) has not been proven. The major aim of this work focuses on determining the structural and mechanistic aspects of this transformation.
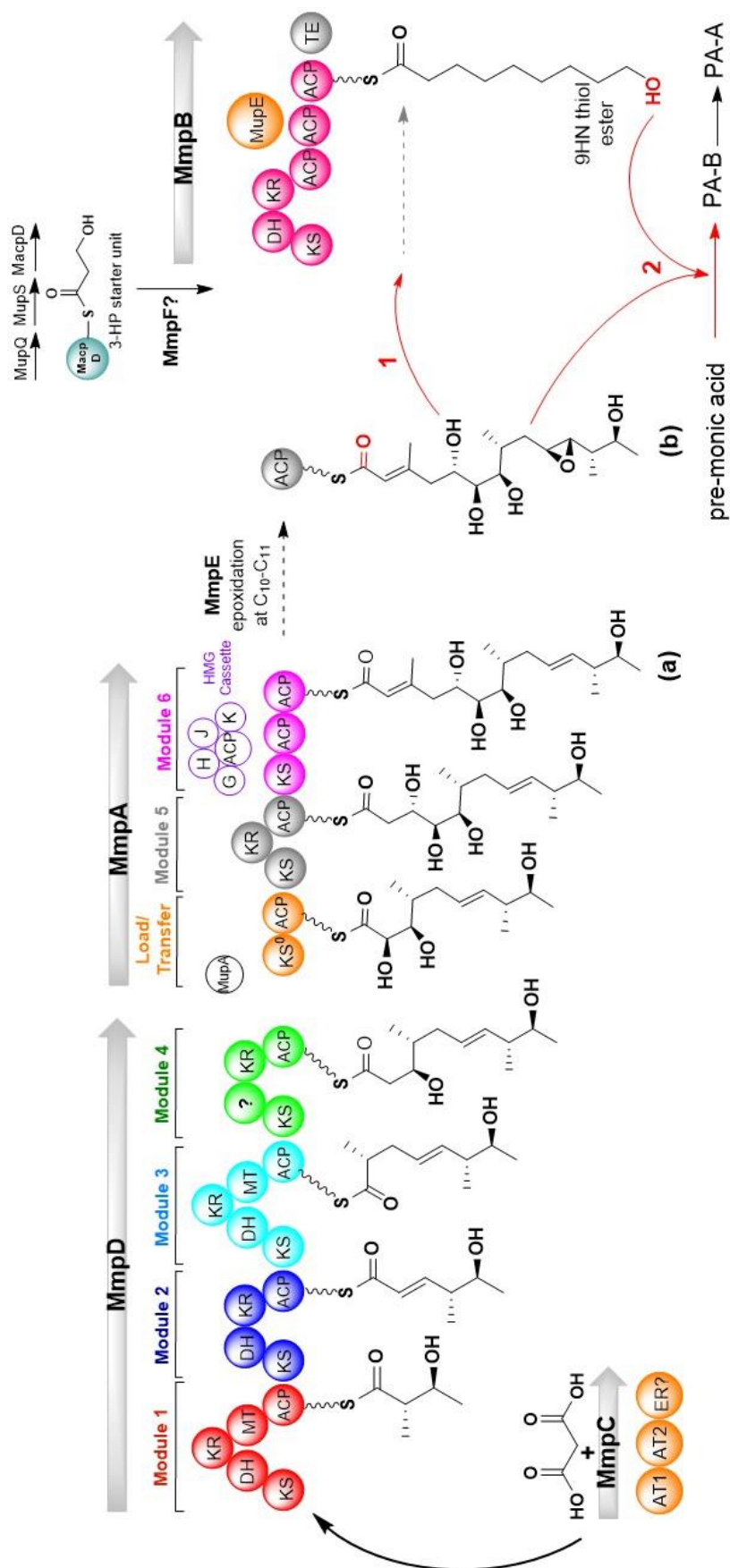
Figure 1.12. The proposed biosynthesis pathway of mupirocin. Pre-monic acid (MA) is synthesized by the decarboxylative condensation of extender units and further modified by the six modules of MmpD and MmpA. The fatty acid portion (9-HN) is synthesized iterative condensations, ketoreduction and dehydration of 3-HP starter unit which may be fully elongated to 9-HN by MmpB. Addition of 9-HN to the MA backbone can happen in two possible ways: continued extension onto MA which can be esterified with a fatty acid chain of any length in MmpB (pathway 1) or biosynthesis of MA and fully elongated fatty acid-9HN separately and followed by esterification (pathway 2).

1.9 Thiomarinol

Thiomarinol A is the major natural polyketide produced and isolated from the marine bacterium *Pseudoalteromonas* sp. SANK 73390, along with several minor compounds, that are collectively known as the thiomarinols (Figure 1.13). These exhibit antibiotic activity against Gram-positive and Gram-negative bacteria, including MRSA. They exist as unusual hybrid antibiotics, effectively two separate class of antibiotics, PA like marinolic acid and holomycin (pyrrothine), joined together through an amide bond.[84] The marinolic acid is comprised of a monic acid like polyketide backbone and 8-hydroxyoctanoic acid (8-HO), linked together by an ester linkage. Key differences to mupirocin include the replacement of the 9-HN acid by the 8-HO acid chain, the lack of the C_{10}-C_{11} epoxide in thiomarinol, an additional C_4 hydroxylation and the presence of the pyrrothine moiety.
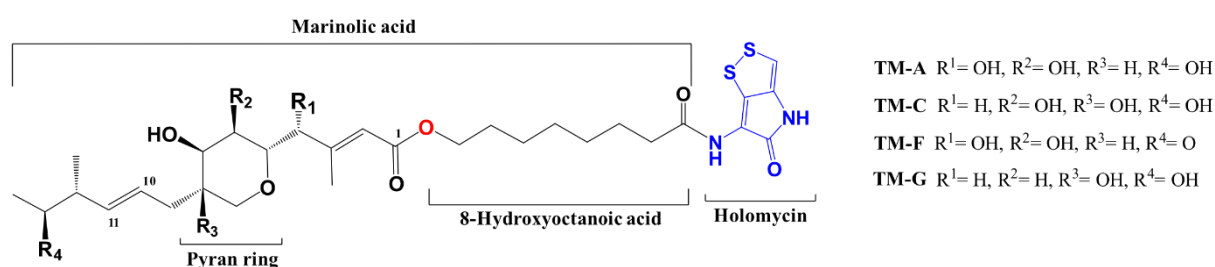
Figure 1.13. Molecular structure of thiomarinol, comprised of marinolic acid containing 8-HO bonded to the pyrrothine by an amide linkage.

1.9.1 Thiomarinol biosynthesis

The thiomarinol biosynthetic gene cluster was identified and reported D. Fukuda *et.al.*, via sequencing the total DNA of the producer organism, *Pseudoalteromonas* sp. SANK 73390.[74] One of the 273 contigs, an approximately 97 kb gene cluster encodes many genes including a hybrid gene cluster of PKSs (TmpA, C and D), a putative FAS (TmpB), and tailoring genes (TmlA-Z) which have high similarity with the mupirocin gene cluster. In addition a non-ribosomal peptide synthetase (NRPS) is found that shows high homology to holomycin biosynthetic machinery in *Streptomycin clavuligerus*.[85] Isotopic labelling and other feeding studies with wild type (WT) and mutant strains of *Pseudoalteromonas* have revealed the origins of atoms, timing of the pyrrothine addition and rationalized the occurrence of minor related products.[36]

In thiomarinol biosynthesis, a modular PKS assembles the carbon backbone and undergoes THP ring formation analogous to mupirocin biosynthesis to produce putative compound **(a)**.[55] If thiomarinol follows the same proposed model as mupirocin, a C_4 fatty acid starter unit (4-hydroxybutyrate) would be elongated to the 8-hydroxyoctanoic acid (probably catalysed by TmpF and TmpB) then esterified to the main monic acid like backbone to produce marinolic acid A **(d)**. Alternatively, as shown in Figure 1.14, fatty acid elongation may occur after esterification and the isolation of other minor metabolites from the WT and ΔNRPS strains, (b), (c), (e) and (g) might support the idea that ester-bond formation with the C_4 chain fatty acid starting unit precedes fatty acid biosynthesis (Figure 1.14). Several structural differences in the modules between thiomarinol and mupirocin systems means either scenario is a possibility although elongation then esterification has essentially been proven for mupirocin.[40] The final pyrrothine moiety is joined to the marinolic acid via an amide condensation reaction catalysed by TmlU, to yield thiomarinol A **(f)**.
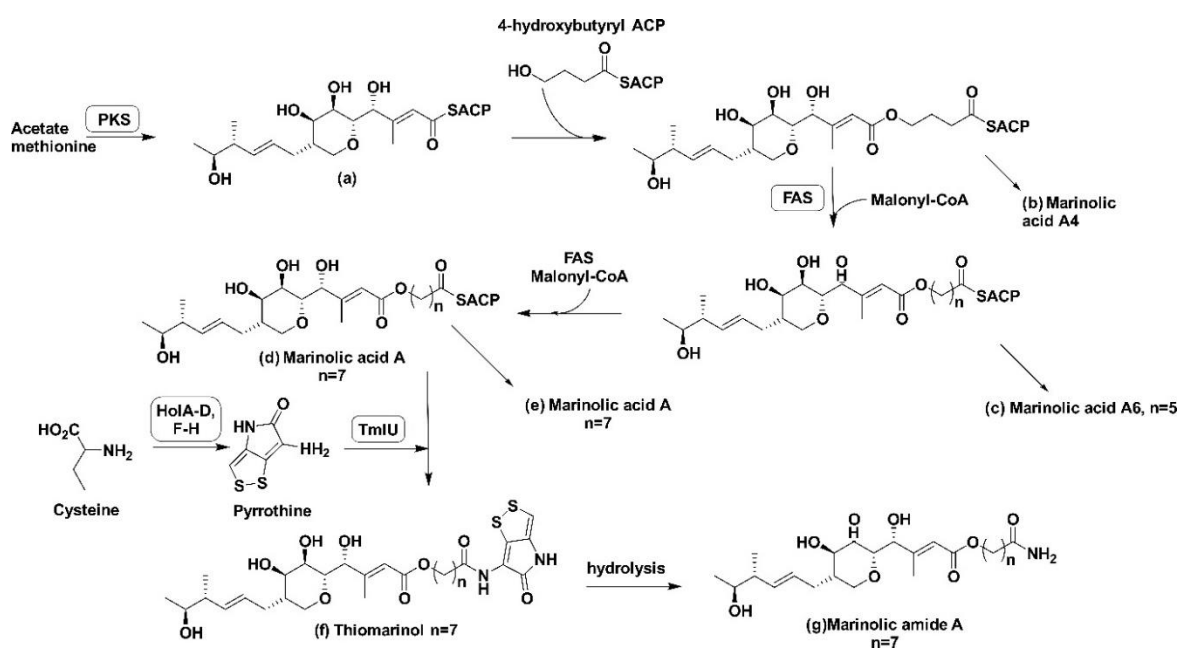
Figure 1.14: The reaction scheme for the overall biosynthesis of thiomarinol, marinolic amide and related minor metabolites in *Pseudoalteromonas*. The reaction scheme is reproduced from reference.[55]

1.10 MupB and TmlB: The putative reverse esterase

Both MupB and TmlB are encoded as discrete proteins in the mupirocin and thiomarinol biosynthesis gene clusters and share 37% amino acid sequence similarity. The mechanism for incorporation of the 9-HN and the 8-HO fatty acid moieties onto the polyketide derived pre-monic acid backbone in mupirocin and thiomarinol, respectively still remains unsolved.

Inspection of the mupirocin gene-cluster reveals that MupB is located in between the multienzyme modules MmpA and MmpB which are responsible for polyketide backbone modifications and fatty acid chain elongation from the starting unit, respectively. An *in vivo* knockout of MupB *(ΔMupB)* experiments produced the curtailed polyketide product, mupirocin H as a major product and is thought to result from stalling of the PKS and ejection of this intermediate from the MmpA module.[86] Essentially this provided no clues as to the function of MupB and pre-monic acid was not identified. A BLAST (Basic local alignment search tool) search and three-dimensional model analysis (I-TASSER) of MupB/TmlB predicts a typical FabH KAS III like fold although the normal catalytic triad, Cys-His-Asn[87] is not present and is instead substituted with Cys-His-Asp which may potentially also have a different spatial orientation. The lack of the correct motif for catalysing a decarboxylative Claisen condensation is perhaps not surprising as the pathway is equipped with all of the necessary KAS functionalities with MupB/TmlB presenting an apparent super-abundance of this functionality. The alternate active site presumably alters the typical function of this KAS like enzyme and within the last two decades, several proteins have been reported to show different functionalities other than condensation (Table 1.3). This active site architecture (Cys-His-Asp) has also been identified in the homologous KAS-like enzymes CerJ,[62] ChlB6[63] and PtmR[88] for example that also catalyse esterification reactions. In an analogous role, the polyketide portion of mupirocin/thiomarinol would be trans-acylated onto the active site cysteine of MupB/TmlB and then transferred to the terminal hydroxyl group of the fatty acid chain to complete the esterification (Figure 1.15).
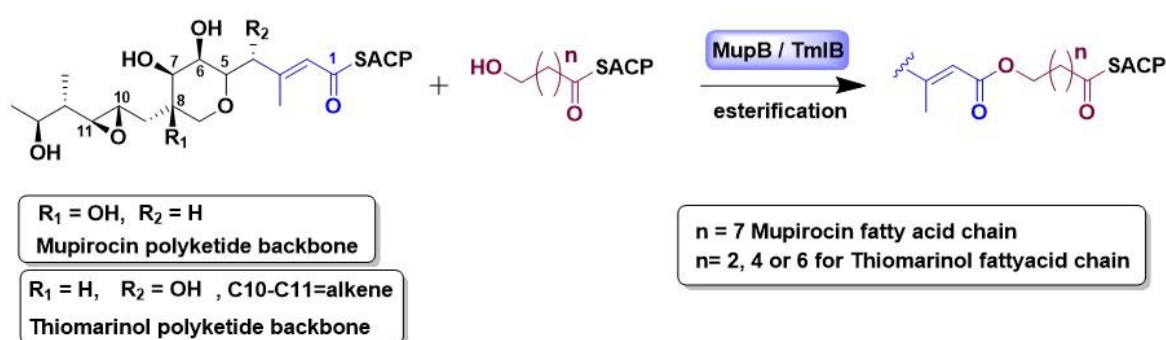
Figure 1.15. Putative esterification reaction catalyses by MupB and TmlB in mupirocin and thiomarinol biosynthetic pathway, respectively. The blue-coloured chain denotes simplified monic acid analogue, enoyl-pantetheine, and the dark-pink chain represents a fatty acid chain.

1.11 Project Aims

The investigations on both the mupirocin and thiomarinol biosynthesis pathway have spanned several decades due to their activity against bacterial resistance. However, there remains much to understand, i.e., mechanisms and selectivity of enzymatic steps, protein-protein interactions which may allow manipulation and rational engineering of the biosynthetic pathway that may offer a route to generate new novel compounds of high value. This PhD project aims to investigate the key esterification steps in mupirocin and thiomarinol biosynthesis where MupB and TmlB enzymes are proposed as potential candidates to catalyse the esterification reaction (Figure 1.15).

In this work, the expression, purification and crystallisation of key proteins MupB/TmlB has been conducted, and assays designed to reveal their structural and functional mechanisms. Through these investigations, this project has also probed the roles played by modular ACPs (M/TmpA_ACP, M/TmpE_ACP and M/TmpB_ACPs) and free-standing ACPs (M/TacpA, B). This will provide information for substrate shuttling among the catalytic domains. Broadly, the main interest of this project will be investigated as follows:

1) **Elucidating the high-resolution crystal structures of MupB and TmlB**

MupB/TmlB will be crystallised, and single crystal X-ray crystallography techniques will be used to solve their structures for the first time. The high-resolution crystal structures will be explored to understand the catalytic triads, putative substrate binding channels and electrostatic nature of the active site pockets. This may help to predict the suitable substrates and putative functional activity for the MupB/TmlB.

2) **Exploring the substrate selectivity and substrate binding to active site cysteine.**

At this stage, CoA/pantetheine derived substrates of different chain length will be used to understand the protein-ligand interactions. In addition, site directed mutagenesis and co-crystal structure with the pre-monic acid mimic will be carried out to deduce information on substrate selectivity, acyl binding pocket architecture and self-acylation mechanism.

3) **Exploring MupB/TmlB interactions with ACPs from mupirocin and thiomarinol biosynthesis pathways**

MupB and TmlB may interact with ACPs from their respective biosynthesis pathway. So, ACP derived substrates will be generated and their acyl transferase activity will be tested

via electrospray ionisation mass spectrometry (ESI-MS) techniques. Heteronuclear single quantum coherence spectroscopy (HSQC) NMR titration experiments will also be conducted to understand which ACPs might interact with MupB/TmlB.

4) Mechanistic investigation of MupB and TmlB

The putative reverse transferase activity of MupB/TmlB will be tested via *in vitro* assays, conducted in the presence of simplified mimics of pre-monic acid and M/TmpB_ACP derivatised fatty acid chains, which may help to elucidate crucial information on what controls C–O bond formation.

2. Structural studies of MupB and TmlB

The initial aims of this project involved the isolation and structural characterisation of the mupirocin and thiomarinol biosynthesis pathway enzymes, MupB and TmlB. Structural characterisation was expected to provide confirmation that both proteins possess an active site architecture consistent with a proposed role in esterification. This chapter details the successful expression, purification, and crystallisation of both MupB and TmlB. Both crystals yielded high-resolution X-ray crystallographic structure permitting a detailed examination of their three-dimensional structures, active site catalytic triads, putative substrate binding channels and enzymatic function.

2.1 Expression, purification and characterisation of MupB

The gene encoding for MupB was previously cloned from *Pseudomonas fluorescens* NCIMB 10586 genomic DNA by Dr Paul Walker and was available as an insert into a pOPINF vector.[89] The DNA was transformed and expressed in *E. coli* bacterial cells (BL21-Novagen). SDS-PAGE analysis of cell media post-induction (Post-I) yielded a visible band near 37 kDa [Post-I, Figure 2.1 (A)] suggesting that MupB had been expressed successfully (the expected mass of MupB is 35 kDa). The harvested cell pellet was lysed by sonication and the protein was extracted in the soluble supernatant as confirmed by an SDS-PAGE band at ~37 kDa [Soluble, Figure 2.1 (A)].

The MupB protein was expressed with an N-terminal His$_6$ tag from the pOPINF vector and immobilized metal ion affinity chromatography (IMAC) was used as the initial stage of the purification. MupB was eluted at an imidazole concentration of ~ 260-340 mM with a peak of 1500 mAU [Figure 2.1 (B)]. The post-IMAC fraction in SDS-PAGE gel showed an intense band of the correct ~ 37 kDa mass.

Using a preparative size exclusion chromatography (SEC) column the post-IMAC fractions were purified further and exchanged into tris buffer (50 mM tris, 100 mM NaCl, 10% glycerol, pH=7.5). The SEC column separates the molecules based on their size where the larger molecules are eluted first and the smaller molecules like salt, and imidazole eluted later with the elution of mobile phase (tris buffer). The MupB protein was eluted as a single peak and confirmed by SDS-PAGE with a distinct band at ~ 37 kDa.

The SEC fractions of MupB were mixed and spin concentrated using a 30 kDa MWCO (Molecular weight cut off) spin concentrator to a final concentration of 17 mg/mL (485 µM). A further analytical SEC analysis using 10/300 Superdex 75 analytical column suggested that the purified MupB existed as a dimeric species in solution (Figure 2.1 C). Further characterisation by Electrospray Ionisation Mass Spectrometry (ESI-MS, Figure 2.1D) yielded an observed mass of 34999 Da (expected 35000 Da). 200 µL MupB (485 µM) and the rest was aliquoted (20 µL in each aliquot) prior to flash freezing in liquid N_2 and stored at -20 °C for further use.
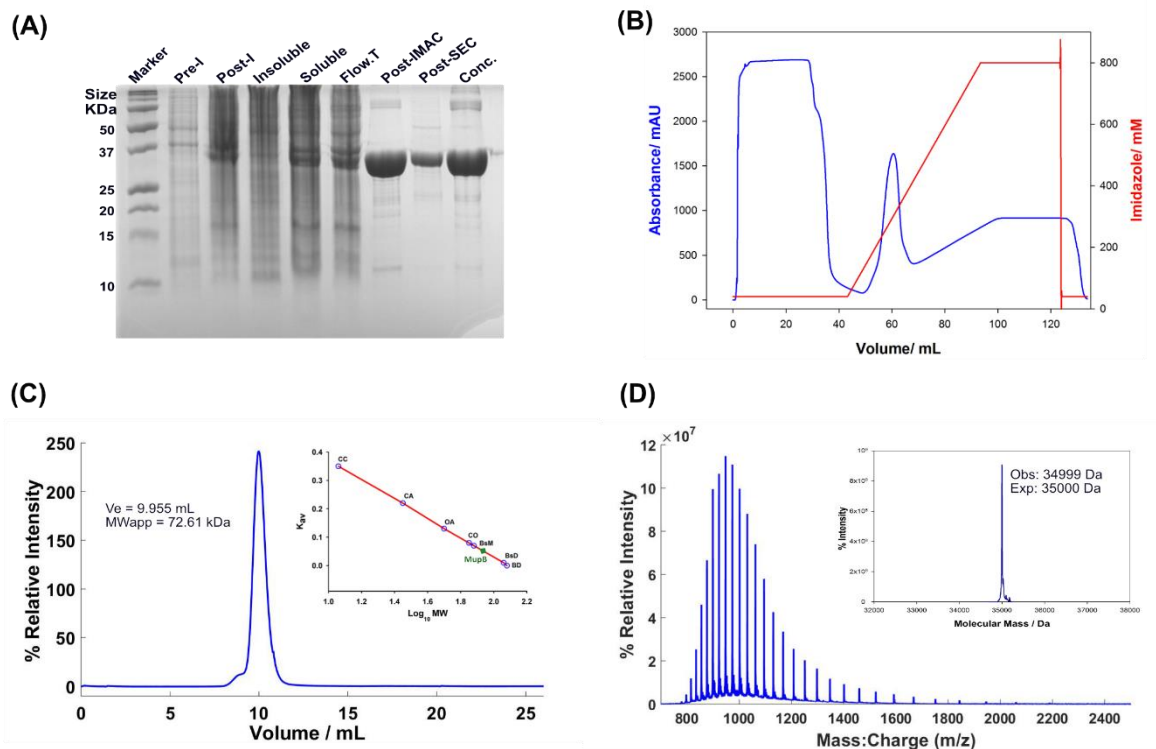
Figure 2.1: MupB protein purification and characterization. (A) SDS-PAGE gel showing successful expression is observed post induction (post-I), after further isolation (via IMAC) and purification (via SEC) with consistent observation of bands at ~ 37 kDa. (B) IMAC chromatogram reporting the elution of MupB as a single peak from an imidazole concentration gradient from 0 to 800 mM. (C) Analytical SEC of MupB eluting as a dimeric species (approximate MW:72.61 kDa), showing on an (inset) calibration curve (shown in green coloured square mark) using known standards: blue dextran (BD, 2 kDa), Cytochrome C (CC, 12.4 kDa), carbonic anhydrase (CA, 29 kDa), ovalbumin (OV, 44 kDa), bovine albumin serum monomer (BsM, 65 kDa), conalbumin (CO, 75 kDa), bovine albumin serum dimer (BsD, 130 kDa). (D) The ESI-MS spectrum of MupB.

2.2 Expression, purification and characterisation of TmlB

TmlB was cloned from the genomic DNA of *Pseudoalteromonas* sp. SANK 73390 following a general cloning procedure (Experimental 7.1). The cloning was accomplished by PCR of genomic DNA with the predesigned primers of TmlB (Table S.9.1). Agarose gel electrophoresis confirmed linearized TmlB DNA and a band at approximately 900 bps suggested successful amplification (Figure S.9.1 A). The PCR products were purified and ligated into a pOPINF expression vector. The DNA was the transformed into NEB-5α bacterial cell and successful ligation was confirmed by X-gal screening (blue-white colony screening). The white colonies were cultured, and the plasmid purified using a standard mini-prep plasmid purification procedure. Further confirmation of TmlB ligation into pPOINF plasmid was provided by sequencing (GENEWIZ ^{TM}).

The plasmid containing TmlB DNA was transformed into host *E. coli* (BL21) bacterial cells. TmlB was expressed on a 2 L scale following a standard protein expression protocol as described in the experimental 7.2. Successful expression of this protein was confirmed by the band of correct mass in SDS-PAGE gel post-induction (Post-I) sample (Figure 2.2 A). The cell pellet was lysed by sonication, extracted by centrifugation. and the supernatant was initially purified using IMAC. The IMAC chromatogram, however yielded two peaks of approximately equal intensity (Figure 2.2B).

SDS-PAGE analysis showed that the first peak contained dimeric TmlB (bands at ~ 75 kDa) and the second absorption peak (~ 985 mAU at 312 mM imidazole) contained monomeric TmlB protein showing bands at ~ 37 kDa (Figure SI.8.3 B). The IMAC fractions of the second peak were combined and exchanged with Tris Buffer (50 mM Tris, 100 mM NaCl, 10% glycerol, pH 8.0) and SEC was used to remove the residual salt and imidazole. ESI-MS analysis yielded an observed mass of 35968 Da (Expected mass 35970 Da, Figure 3.2) confirming the isolation and purification of TmlB. The sample was spin concentrated (via 30 kDa concentrator) to a concentration of 62 μM (2.23 mg/mL) and then stored at -20 °C after flash freezing in liquid N_2.
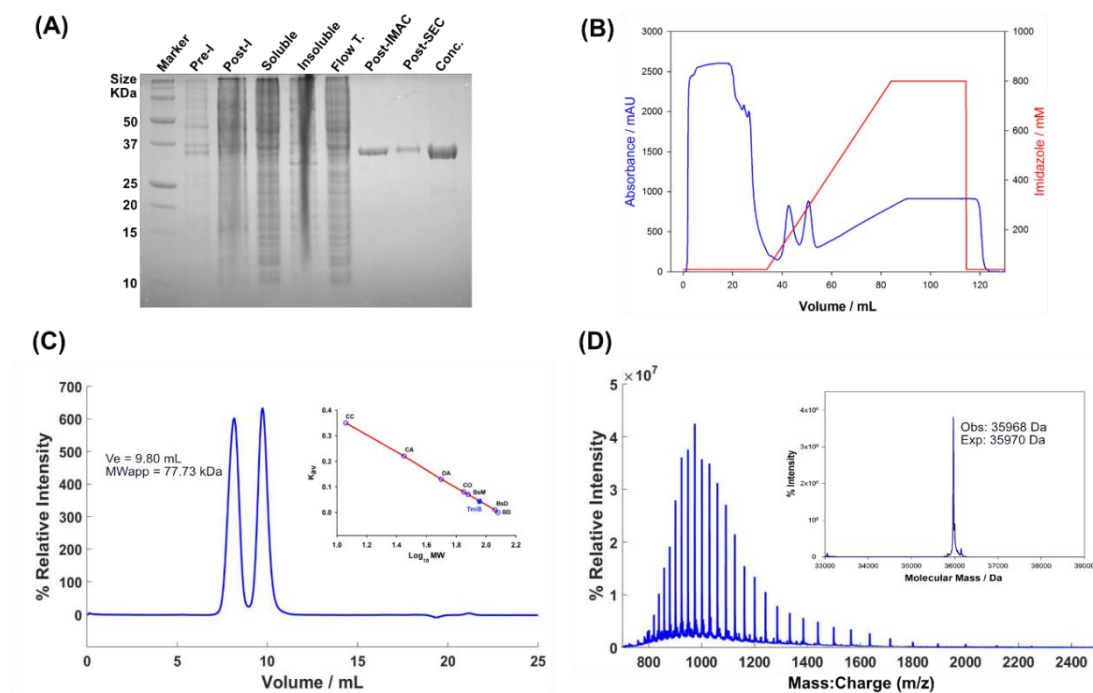
Figure 2.2: TmlB expression, purification, and characterization. (A) The SDS-PAGE gel image suggested the successful isolation and purification of TmlB as the post-IMAC, post-SEC, and concentrated samples yielded a single band at ~37 kDa. (B) IMAC chromatogram reporting the elution of TmlB from soluble supernatant where the second peak contained the protein. (C) Analytical SEC of TmlB eluting as a dimeric species (second peak). Calibration curve (inset) is as Figure 2.1. (D) The ESI-MS spectrum of TmlB.

2.3 Crystallisation and structure elucidation of TmlB

Purified TmlB (195 µM) protein was screened for crystallisation against 384 commercially available Molecular Dimension screening solutions (Morpheus, PACT primer, JCSG+ and Structure I+II) using sitting drop vapour diffusion techniques. Successful crystallisation hits were observed in three different crystallisations conditions (Table 2.1). Crystals were flash frozen and X-ray diffraction data collected at the Diamond Light Source synchrotron, UK.

Table 2.1: Successful crystallization conditions for TmlB protein.

Solution	Well-drop	Solution composition
Morpheus	A5-Drop1	30% v/w 550M_20K, 0.06M divalent complex, 0.1M MB2, pH 7.5
PACT	C11-Drop1	0.2M CaCl$_2$, 0.1M HEPES, pH 7.0, 20% w/v PEG 6K
Structure 1+2	C5-Drop1	1.4M Sodium Citrate, 0.1M HEPES, 7.5 pH

Amongst the crystals, a single crystal diffracted well and a dataset of 1800 images was collected at a wavelength of 0.9786 Å, exposure time 0.02 sec, oscillation 0.1° with cryogenic temperatures maintained throughout the data acquisition. The Auto-processing by Diamond Pipeline (Fast DP) suggested a 1.5 Å X-ray diffraction resolution limit (Table 2.2).

Table 2.2: Parameters used for data collection at diamond light source for TmlB crystals.

Parameter's name	Successful crystals condition provided best X-ray diffraction data
	30% v/w 550M_20K, 0.06M divalent complex, 0.1M MB2 7.5 pH
Cryoprotectant	N/A
Number of Images	1800
Wavelength (Å)	0.9686
Exposure time (sec)	0.020
Oscillation (°)	0.10
Beam Size (μm)	20×20
Beam Flux	1.87e+0
Fast DP Resolution (Å)	1.5

The data processing and structure determination were performed with the assistance of Dr Christopher Williams (University of Bristol). The dataset was scaled and merged into P 1 2₁ 1 using the Diamond Pipeline (Fast DP) with cell dimensions of a=52.06 b=89.29 c=66.08 A. Two molecules were observed in the asymmetric unit (ASU) with a solvent content of 39% (Figure 2.3). The structure of TmlB was determined by molecular replacement with Phaser[90] within the PHENIX suite[91] using partial models from *Aquifex aeolicus VF5* 3-oxoacyl-[acyl-carrier-protein] synthase III (2EBD) and *E. coli* FabH (3IL9). The resulting electron density maps were of sufficient quality to auto-build[92] 90% of the model. Iterative rounds of manual rebuilding and refinement in Coot[93] and phenix.refine[91,94] further improved the R-free values before the final structure was validated with MolProbity[95] and the PDB_REDO web server.[96] The molecular refinement statistics of the final TmlB crystal structure are provided in Table 2.3.
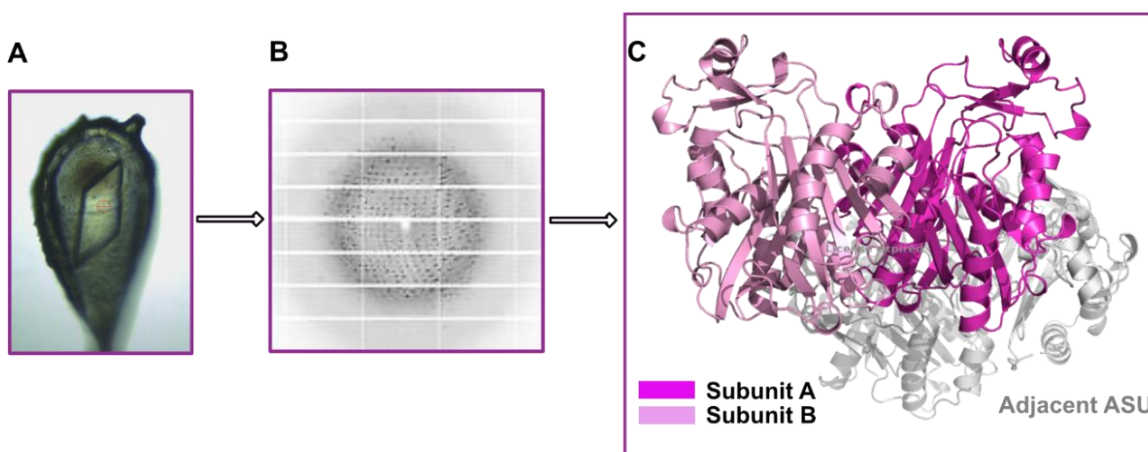
Figure 2.3: (A) The looped and frozen TmlB crystals imaged while collecting XRD data. (B) The X-ray diffraction pattern of 1.5 Å resolution of TmlB crystal and (C) The refined crystal structure of TmlB where a single ASU contain two identical monomers, thus formed a dimer within the ASU.

Table 2.3. The crystallographic data collections and refinement statistics for MupB and TmlB.

	MupB	TmlB
Crystallographic Data		
Wavelength (Å)	0.9762	0.9686
Space Group	P 1 2$_1$ 1	P 1 2$_1$ 1
Cell Dimension a,b,c (Å)	62.12, 98.54, 117.04 α=γ= 90°, β=98.78°	52.13, 89.35, 66.01 α=γ= 90°, β=106.5°
Resolution (Å)	57.84-2.379	34.17- 1.37
No. of Observation	185069 (18247)	
No. of Unique observations	55943 (5517)	120046 (11818)
Completeness (%)	99.65 (98.72)	98.84 (97.56)
I/σI	9.35 (1.15)	
R$_{merge}$	0.05982 (0.8302)	
Wilson B-factor	61.33	14.82
Refinement		
Resolution (Å)	57.84-2.379 (2.46-2.379)	34.17- 1.37 (1.419-1.37)
No. of protein atoms	8822	4919
No. of ligands atoms	7	11
Solvent	48	791
R-work	0.1928 (0.3216)	0.1473 (0.2112)
R-free	0.2412 (0.3836)	0.1633 (0.2242)
Geometry		
RMS bonds (Å)	0.011	0.012
RMS angle (°)	1.51	1.51
Ramachandran favoured (%)	96.18	97.20
Ramachandran allowed (%)	2.75	2.80
Ramachandran outliers (%)	1.07	0.00
Clash-score	6.58	4.99
Average B-factors (Å^2)	77.92	22.26
Proteins	77.96	20.43
Solvent	68.78	33.31
Ligands	83.11	48.20

PDBePISA analysis[97,98] of the final TmlB structure suggested that the observed dimeric assembly of TmlB was stable in solution based on the positive value of Standard Free Energy of Dissociation, ΔG_{diss} = 23.3 kcal/mol (Table 2.4). TmlB forms a dimer of two monomeric subunits through a dimerization interface (Figure 2.4 A) which consists of 10 hydrogen bonds and 4 salt-bridges (Table S.9.3), burying a total of 1246 Å^2 surface area (from a total of 23400 Å^2 over the complete dimer). This interface scored 0.592 in Complex Formation Significance Score, CSS (ranges from 0 to 1 as interface relevance to complex formation increases) which implies that the interface plays an essential role in protein-protein complex formation.

Table 2.4 PDBePISA analysis of TmlB dimer. Monomers were assessed for their total number of atoms, residues, total number of surface atoms and residues. Total solvent accessible area along with interface surface area are shown. The highly negative solvation energy of folding values implies that both monomers are correctly folded. One assembly (crystal split) was found to be stable at AB[NA]$_3$ composition where NA is a ligand with Accessible surface area (ASA), Buried surface area (BSA), Solvation energy effect (ΔG_{int}), Standard free energy of dissociation (ΔG_{diss}).

Interface parameter of Monomers							
Monomer	**N_{at}**	**N_{res}**	**sN_{at}**	**sN_{res}**	**Interface Area [Å^2]**	**Total Area [Å^2]**	**ΔG_{int} [kcal/mol]**
A	2370	305	1268	269	1816.5	13559.8	-292.5
B	2329	299	1261	262	1814.9	13449.6	284.4
Average:	2349	302	1264	265	1815.7	13504.9	-288.5

Stable Crystal Assemblies							
PQS No	**mm size**	**Composition**	**Stable**	**ASA [Å^2]**	**BSA [Å^2]**	**ΔG_{int} [kcal/mol]**	**ΔG_{diss} [kcal/mol]**
1	2	AB[NA]$_3$	Yes	23400	3990	-57.6	23.3

2.4 The dimer interface of TmlB

The TmlB dimer interface involves Nα2, Nβ3, Nα3, loop L3 and loop L14 secondary elements of each monomer (Figure 2.4). The major contribution to the dimer interface is through the association of the L3-L3´, Nα3-Nα3´ and Nβ3- Nβ3´ (" ´ " indicates the strands/helix/loops from the adjacent monomer of TmlB dimer) strands from the two monomers via antiparallel backbone interactions (Figure 2.4 A). Several residues from these secondary elements contribute to the dimerization via electrostatic, hydrogen bond and hydrophobic interactions (Figure 2.4 B). The dimer interface of TmlB occupies ~5.3 % of the total surface area (Figure 2.5 A). One of the major interactions between Nα3-Nα3´ involves a pair of symmetrical electrostatic interactions between K135 and D125´ at the dimer interface (Figure 2.5 B). Further interactions consist of hydrophobic packing between loops, α–helices and β-strands L14-Nα2´–Nβ3´ as well as hydrogen bonds between Nα2-L14´ and L9´-L4 loops (eg L194´ to K105, Figure 2.4 B).
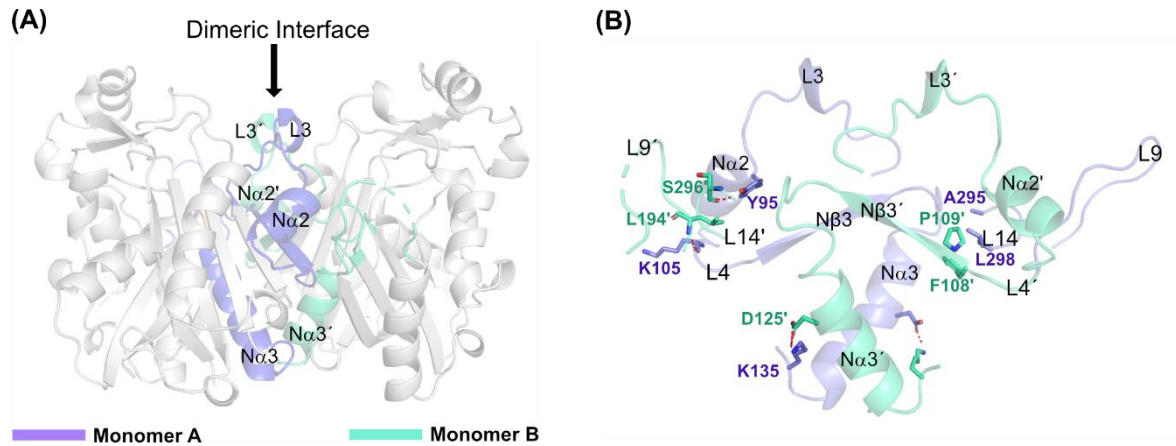
Figure 2.4. (A) The antiparallel arrangements (front view) of secondary structural elements between two TmlB monomers forming the dimeric interface (violet for monomer A and geencyan for monomer B). **(B)** All the elements from each monomer that are forming the TmlB dimeric interface are labelled as Nα2, Nα2′, Nβ3, Nβ3′, Nα3, Nα3′, loop L3, L3′, L4, L4′, L9, L9′, L14 and L14′. The major amino acid residues responsible for electrostatic interactions between K135 and D125′. Hydrogen bond and hydrophobic interactions between residues are also labelled as K105, Y95, A295, and L298 from TmlB monomer A and L194′, S296′, F108′ and P109′ from TmlB monomer B.
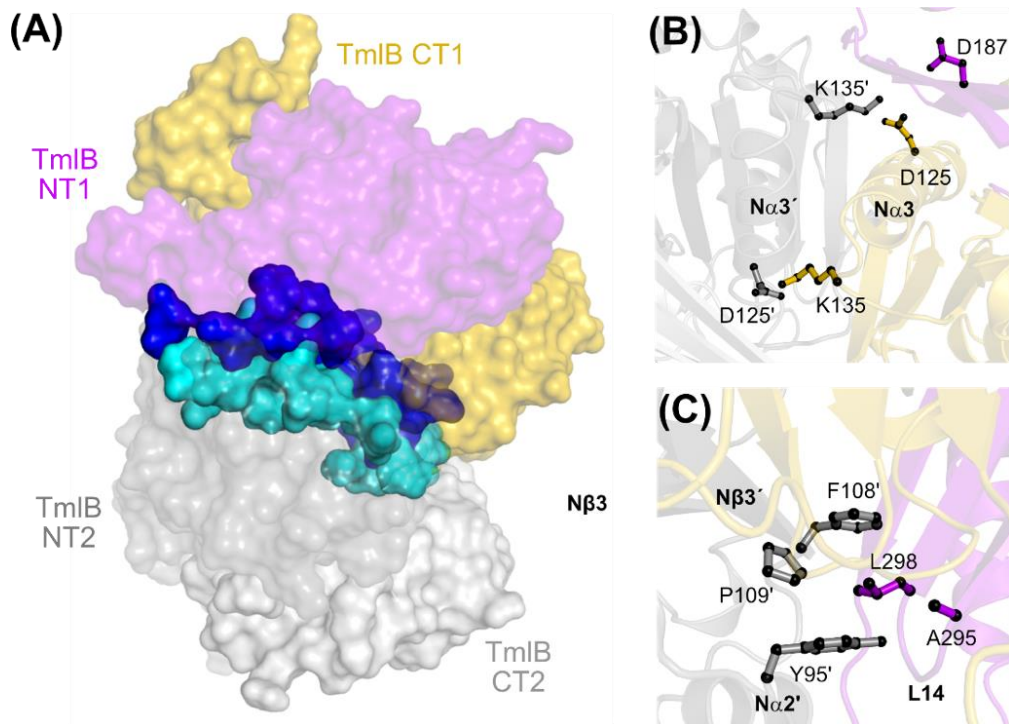
Figure 2.5 (A) The surface representation of TmlB dimer and its interface region formed by the Nα2, Nβ3, Nα3, loop L3 and loop L14 of TmlB monomer A (coloured in blue) and monomer B (coloured in cyan). **(B)** The residues showing electrostatic interactions between the Nα3- Nα3′ helices and **(C)** the hydrophobic amino acid interactions between L14-Nα2′, Nβ3′.

2.5 Crystallisation and structure determination of MupB

Purified MupB protein (485 µM) was used to perform the crystallization screening as described earlier for TmlB and several successful hits were observed after 5 days (Figure 2.6 A). The crystals were looped and cryoprotected in 20% glycerol before flash freezing. The XRD data were collected at Diamond light source synchrotron among which one crystal (0.2 M NaCl, 0.1 M HEPES pH 7, 20% w/v PEG 6K) diffracted well with resolution limits of 2.4 Å (Figure 2.6 B). The cryogenic temperature was maintained during the data acquisition and a dataset of 3600 images were collected at a wavelength of 0.9762, exposure 0.01s, oscillation 0.20° (Table 2.5). The data processing and structural determination was carried out by Dr Ash Winter and Dr Chris Williams (University of Bristol) as described in the experimental 7.5.

The diffraction data was processed into space group P 1 2_1 1 using Diamond Pipeline (Fast DP) with cell dimensions of a=62.12 b=98.54 c=117.04. The trimmed TmlB structure was used as the search model within Phaser-MR and Auto-build pipeline to determine the MupB structure. Four molecules were observed in a unit cell with a solvent content of 68% (Figure 2.6 C). Iterative rounds of manual rebuilding and refinement further improved the structure following a similar procedure described previously for TmlB. The final statistics are provided in Table 2.3.
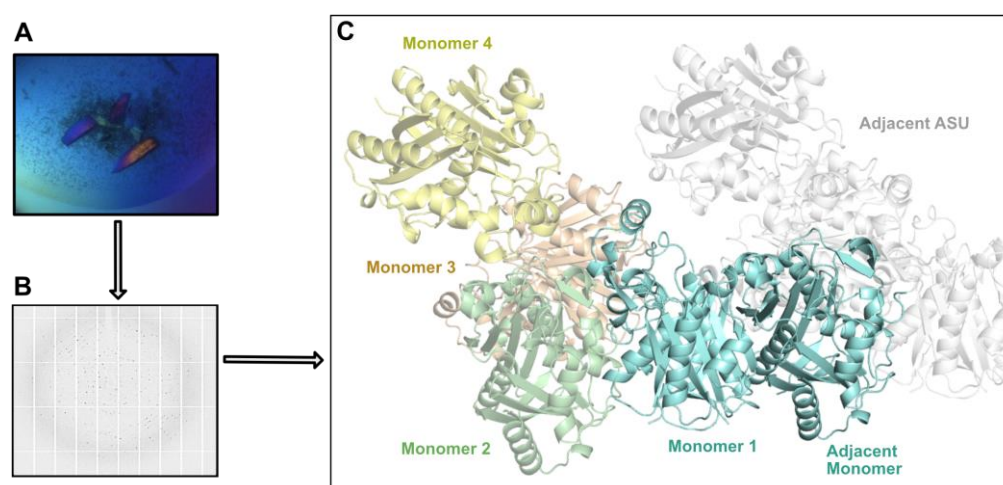
Figure 2.6. (**A**) MupB crystals imaged under cross polarised light. (**B**) X-ray diffraction pattern of 2.4 Å resolution for the MupB crystal and (**C**) ASU of MupB which contained 4 individual chains (blue, green, pink and yellow respectively), whereby dimers were formed between symmetry mates, across ASU (blue), with the remaining 3 chains not forming dimers (grey).

PDBePISA analysis suggested that MupB existed as a stable dimeric structure based on a positive ΔG_{diss} value though four monomers coexist in an asymmetric unit. One probable quaternary structure (PQS) set was formed with two assemblies each contained two monomeric units of MupB according to composition. The two different compositions of MupB dimer were found (AC and BD) to be stable in solution as $\Delta G_{diss} > 5$ kcal/mol. The negative value of solvation free energy gain on formation (ΔG_{int}) of the interface indicated the presence of a hydrophobic interface between the two monomers of MupB (Table 2.6). The hydrophobic patch is supplemented by 18 hydrogen bonds for two MupB dimer compositions (AC) and (BD) which bury 1381 Å^2 and 1364.96 Å^2 area, respectively among the total interface area (Table S.9.4).

Table 2.6: PDBePISA analysis of MupB tetramer observed within the 2.4 Å dataset. Monomers were investigated for their total number of atoms, residues, total number of surface atoms and residues. Total solvent accessible area along with interface surface area are shown. The highly negative solvation energy of folding (**ΔG_{int}**) values implies that all monomers are correctly folded. two assemblies (crystal split) were found to be stable at AB[NI]$_3$ composition where NI is a ligand with Accessible surface area (ASA), Buried surface area (BSA), Solvation energy effect (**ΔG_{int}**), Standard free energy of dissociation (**ΔG_{diss}**). (PQS-Probable Quaternary Structure, mm size- macromolecular monomeric unit, **ΔG_{int}** indicates the solvation free energy gain upon formation of the assembly, in kcal/M).

Interface parameter of Monomers							
Monomer	N_{at}	N_{res}	sN_{at}	sN_{res}	Interface Area [Å^2]	Total Area [Å^2]	ΔG_{int} [kcal/mol]
A	2233	290	1174	262	1757.8	12664.0	-255.7
B	2220	288	1180	256	1725.6	12413.4	-252.2
C	2135	277	1141	246	1766.6	12155.3	-240.6
D	2222	289	1205	258	1741.9	12743.2	-257.5
Average:	2202	286	1175	255	1747.9	12494.0	-251.5

Stable Crystal Assemblies							
PQS No	mm size	Composition	Stable	ASA [Å^2]	BSA [Å^2]	ΔG_{int} [kcal/mol]	ΔG_{diss} [kcal/mol]
1 (*)	2	AC[NI]$_3$	Yes	21240	3930	-55.6	20.8
	2	BD[NI]$_3$[OD]	Yes	21630	4010	-52.8	20.7

(*) Signs indicate that in solution not all chains appear in their original positions.

2.6 The dimer interface of MupB

According to the PDBePISA analysis, the high-resolution crystal structure of MupB is found in two different compositions, AC and BD which are stable in solution. These two dimeric compositions are identical and therefore share the same types of elements to build the characteristic dimeric interface (Figure 2.7 A). Monomers of MupB predominantly dimerise via an extensive hydrophobic interface mediated via their N-terminal cores (L3α1, L5, Nα2, Nβ3, Nα3) and C-terminal loops L9 and L14 (the tight loop between Cβ4 and Cβ5) (Figure

S.9.2). Two major contributions to the MupB dimer interface come via L3-L3′ and the L5 and Nα2′ core element interactions. L3-L3′ interact via the packing of aromatic residues F86-F86′ L5 and Nα2′ through Y94-W113′ (Figure 2.7 B). A further antiparallel backbone interaction via Nβ3-Nβ3′ also plays an important role in forming the MupB dimeric interface supplemented by a series of aromatic residues some of which are presented in figure 2.7 C. Beside these, additional hydrophobic interactions between the monomers are mediated with Nα3- Nα3′, Nα2-L9′ and Nα2-L14′via a series of hydrophobic amino acids residues (Table S.9.4).
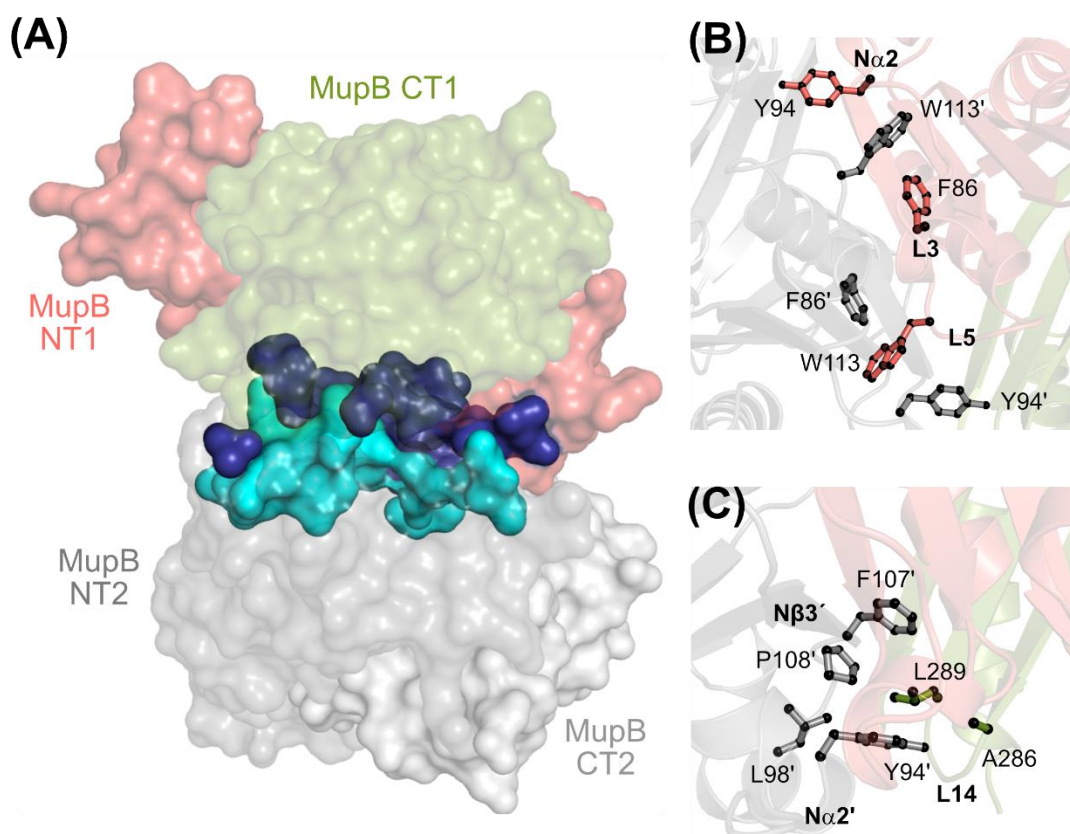
Figure 2.7. (A) The surface representation of MupB and the dimeric interface region, which is formed by the L3α1, L5, Nα2, Nβ3, Nα3, L9 and L14 structural elements of monomer A (coloured in blue) and monomer B (coloured in cyan). (B) The major aromatic residues responsible for hydrophobic interactions in MupB dimeric interface provided by L3, L5 and Nα2 elements of monomer A and its dimeric partner. (C) The other hydrophobic amino acids to generate a hydrophobic patch in the MupB dimeric interface provided by the Nα2, Nβ3, Nα3, L9 and L14 elements of one MupB monomer and its dimeric partner.

2.7 The overall structural analysis of MupB and TmlB

2.7.1 Core architecture of MupB and TmlB

The core architecture of each monomer found in both MupB and TmlB is comprised of two consecutive β-α motifs, which are termed N-terminal (β−α−β−α−β−α−β−β) and C-terminal (β−α−β−α−α−α−β−β) cores (Figure 2.8) with a high degree of structural conservation between the two homologous proteins. The N-terminal core contains five β-strands (Nβ1-Nβ5), arranged in an anti-parallel beta sheet, three α-helices (Nα1-Nα3) and loops (L1-L8). Loop L1 contains both a single α-helix (designated as L1α1) and two β-strands (L1β1 and L1β2). Loop L3 contains an additional α-helix (L3α1), which is not present within TmlB (Figure 2.8 A and C). Whereas, the C-terminal cores consist of helices Cα1-Cα3, four β-strands Cβ1-Cβ2, Cβ4-Cβ5, and lacking the third β-strand as in N-terminal core structure which is replaced by a small helix (L12α1) connecting Cα2 and Cα3 (Figure 2.8 A and C). All the helices and strands in C-terminal substructures are joined with loops, L9-L14. Features of L1 are specific to the N-terminal core and this motif acts to sterically cap the C-terminal core with the β-strands (L1β1/L1β2) acting to enforce a pseudo twofold symmetry between the antiparallel β-sheet and central α3 of each core, which one face of the β-sheet packs against. N/C α1 and α2 pack against the opposite face of the beta sheet, further stabilising this N-C core interface (Figure 2.8 B and D).
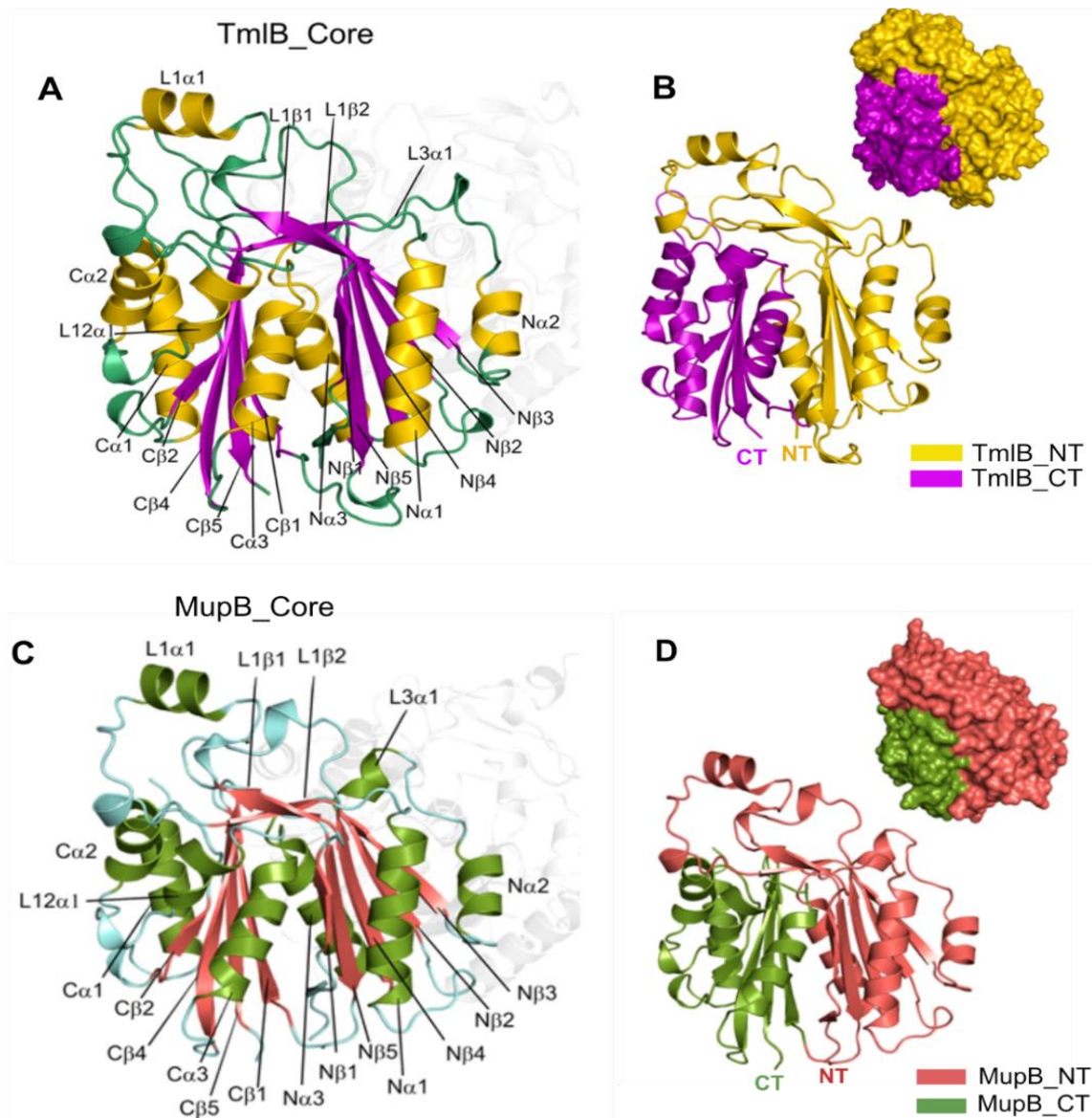
Figure 2.8. (A) The TmlB core structure (B) The TmlB monomer is comprised of N-terminal (yellow) and C-terminal (magenta) cores where the N-terminal L1β1 and L1β2 loops sterically cap the C-terminal core structure. The TmlB monomeric surface area is shown in inset. (C) MupB core structure (D) Like TmlB, MupB monomer is also built by N-terminal (red) and C-terminal (olive green) core structures, and the surface area is shown in inset.

2.7.2 The relative structures of MupB and TmlB

The high-resolution crystal structures of MupB and TmlB have revealed that both proteins are crystallized in the same space group, P 1 2_1 1 and exist as homodimers. They both adopt similar three-dimensional folds and superimposition of all the backbone residues gives a root mean square deviation (RMSD) of 1.8 Å and 0.9 Å for the dimeric and monomeric structure, respectively (Figure 2.9 (A) and (B)). A search for sequence similar non-reductant proteins with MupB and TmlB using BLAST (Basic local alignment search tool) yielded

sequences of proteins which are not yet biochemically characterised (Table S.9.5) but clearly contain FabH-like domains similar to the thiolase superfamily (Figure S.9.3). The 3D pairwise structure comparison of representative thiolase fold member proteins using DALI[99] (protein structure comparison server) confirmed that MupB and TmlB are most closely related to the KAS-III like enzymes (Table S.9.6).
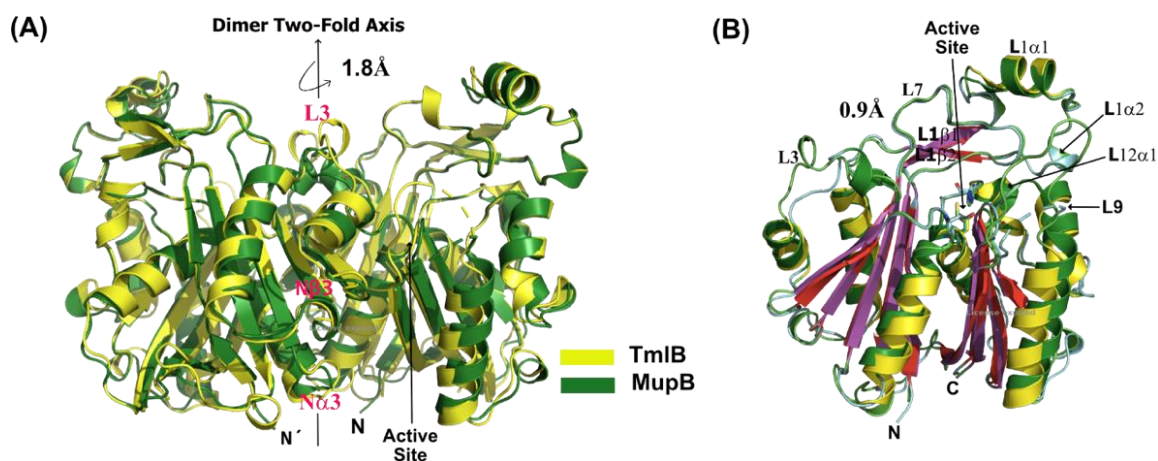
Figure 2.9. **(A)** The 3D structural alignment of MupB and TmlB gives a RMSD of 1.8 Å for the dimeric structure. The structural elements responsible for the dimeric interface are labelled in red. **(B)** The MupB (forest green) and TmlB (yellow) monomers are superimposed with a RMSD of 0.9 Å. All the loop regions are labelled which are important for substrate recognition and binding to the cavity along with the presumed active sites of MupB and TmlB.

The top ten most similar structures are listed in Table 2.7 and (essentially equivalent results for TmlB in Table S.9.7) and include a number of KAS III like proteins reported to be involved in different roles and/or functions other than Claisen condensations and are evolutionary diverged.[100][101] The top six hits include i) the recently reported ChlB3 crystal structure from the chlorothicin biosynthesis pathway in *Streptomyces antibioticus,* which acts as an acyltransferase (AT), and showed the closest resemblance with both MupB (2.2 Å) and TmlB (2.3 Å)[63], ii) DpsC, the propionyl priming ketosynthase and AT from the daunorubicin biosynthetic pathway,[64] iii) CerJ from the Cervimycin biosynthesis pathway, which transfers a malonyl unit onto a sugar moiety via C-O bond (ester) formation,[62] iv) OleA in *Xanthomas campestris,* which catalyses head to head condensation between two long-chain fatty acyl coenzyme A molecules,[102] v) PqsBC in *P. aeruginosa* PAO1, which catalyses quinolone production by condensing octanoyl-coenzyme A with 2-aminobenzoylacetate[103] and vi) Pks11 in *M. tuberculosis* that synthesizes a cyclic methyl-branched alkylpyrone.[104] Other homologs are FabH/FabH-like enzymes from different organisms and all of these enzymes catalyse Claisen like condensation. The KAS III ZhuH

is involved in starter unit priming in the biosynthetic pathway of the aromatic polyketide R1128 in *Streptomyces* sp. R1128.[61]

Table 2.7. The structural homologs search result of refined MupB crystal structure using DALI server against the whole PDB archive. Top Ten (10) protein structures those are structurally related to MupB-chain A are listed according to their Z score (Similarity score, acceptable value is > 2).

*The Dali Z-score is an optimised similarity score defined as the sum of equivalent residue-wise C_α-C_α distances among two proteins. A higher value of Z-score indicates greater similarity. **RMSD** stands for the Root-mean-square deviation, calculated between Cα-atoms of matched residues at best 3D superimposition of query and target structures. **%$_{seq}$**: % Sequence identity

PDB entry	Z-score	Cα-RMSD (Å)	%$_{seq}$	Description
7eqi-G	33.2	2.2	16	ChlB3 (AT) from *Streptomyces antibioticus*
4xsb-B	32.1	2.3	15	Butanoyl bound DpsC from *Streptomyces peucetius*
3s20-B	31.3	2.6	14	Cerulenin bound OleA from *Xanthomonas campestris*
5dwz-D	30.9	2.7	19	PqsBC (KAS III) from *Pseudomonas aeruginosa* PAO1
4jap-C	30.5	2.7	17	Pks11 from *Mycobacterium tuberculosis*
2ebd-B	29.7	2.6	15	3-oxoacyl-ACP synthase III from *Aquifex aeolicus* VF5
3t8e-B	29.5	2.6	14	Cervimycin K bound CerJ from *Streptomyces tendae*
1zow-A	29.4	2.5	14	FabH from *Staphylococcus aureus*
1mzj-A	29.2	2.6	16	ZhuH (KAS III) from *Streptomyces* sp. R1128
3il9-A	28.9	2.6	15	FabH from *Escherichia coli*

The three-dimensional structural comparison of these proteins with MupB and TmlB monomers showed that they contain the typical FabH and KAS-III like N-terminal βαβαβαββ motif. The main difference lies in the C-terminal motif of MupB/TmlB which possess a short α-helix, L12α1 connecting Cα2 and Cα3 instead of Cβ3. However, this arrangement structure is the most common motif present in the homologous proteins discussed here and the presence of Cβ3 in FabH and ZhuH is the exception. Structural alignment also shows that both MupB and TmlB share similar N- and C-terminal secondary structure core motifs to their homologs (Figure 2.10) and they associate to form dimers in a similar fashion (Figure S.9.5). However, a major difference in the construction and orientation of L3 and L9 is observed in MupB and TmlB in comparison to the homologous enzymes shown in Figure 2.10. The monomer-monomer interactions between L9-L9´, L9-L3´ and L3-L3´ play important roles in creating a substrate binding pocket that influences

substage binding specificity.[61] The L3 loop in EcFabH is the shortest (8 amino acids) of these examples whereas TmlB has a slightly longer L3 loop (14 amino acids). ZhuH, CerJ,DpsC and MupB are fairly similar (containing 12-13 amino acids). In addition, a significant difference in the L9 region is observed between the homologous enzymes. The L9 loops in ChlB3, DpsC and CerJ are comprised of two α−helices. OleA andEcFabH contain one α−helix and two β−strands, Pks11 contains two β−strands and ZhuH contains two α−helices and two β−strands to construct the loop L9. The average size of these L9 loops is constructed with 28-42 amino acid residues and encloses the region above the active site region, as shown in Figure 2.10. Compared to these protein structures, TmlB and MupB have smaller L9 loops (12-14 amino acids) without either α-helix or β-strand content (In MupB amino acid residues 187-194 are missing due to a weak electron density map). Rather than capping the active site regions via L9-L9′ interactions observed in other FabH and KAS-III enzymes, the much smaller L9 loops are tilted away from the enzyme structure, forming an outer wall on one side of the active site (clearly shaded in red for MupB and green for TmlB in figure 2.10). L3 and L3′ interact across the dimer interface but lack the L3-L9′ interactions therefore do not pack so tightly across the active site (Figure S.9.5). The result is a distinctly different and accessible region of space above the active site residues that may accommodate a larger or more complex substrate like pre-monic acid.
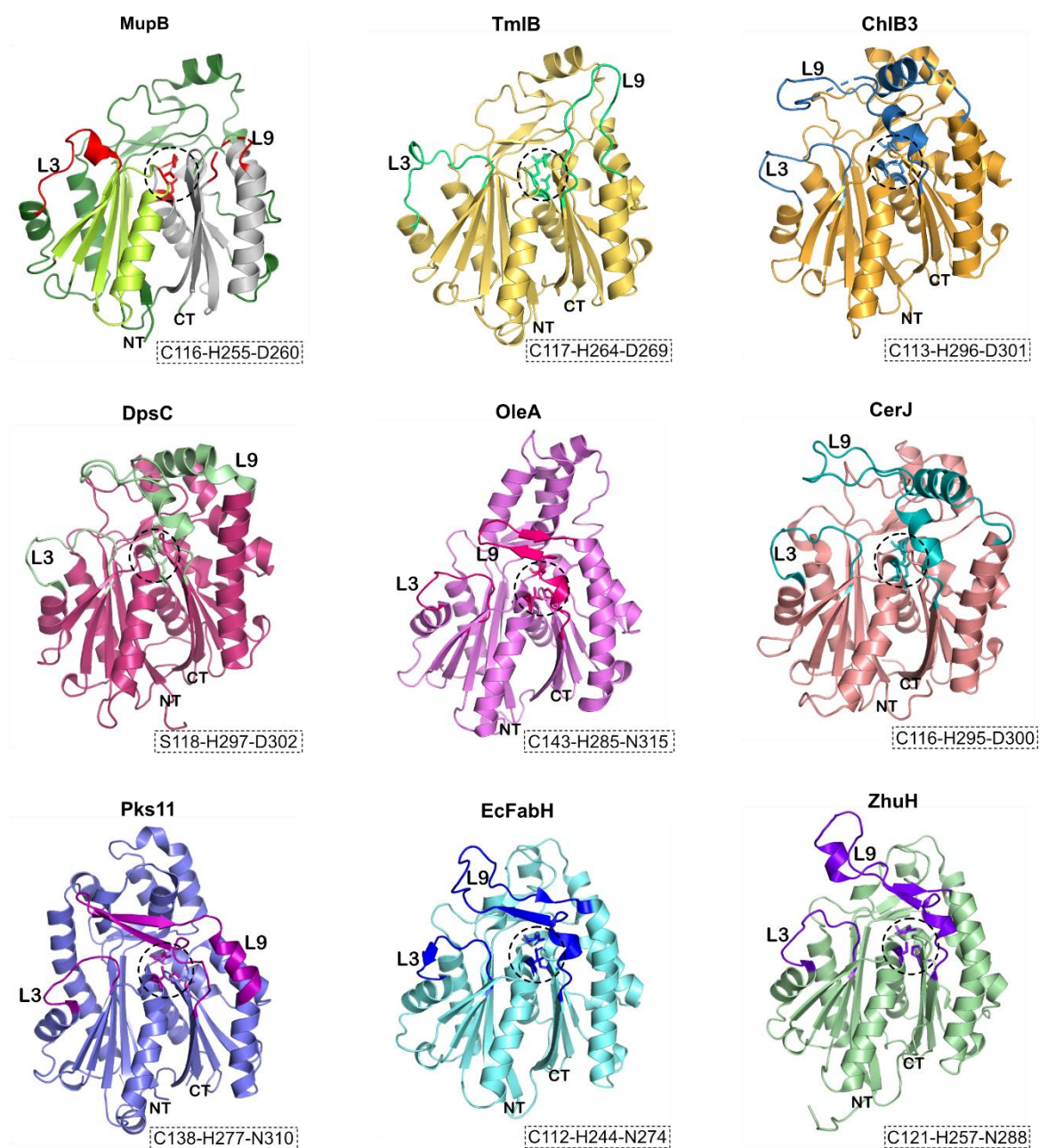
Figure 2.10. The monomeric structure comparison of MupB (forest green) and TmlB (yellow) with their structural homologs, ChlB3 (7EQI), DpsC (4XSB), CerJ (3T8E), OleA (3S20), Pks11 (4JAP), EcFabH (1EBL) and ZhuH (1MZJ). In MupB the N and C-terminal core structures have been coloured with lime green and grey, respectively to understand their construction and orientation fashions. The loops L3, L9 and active site regions (circled with black dashed line) are coloured differently than the monomers for the purpose of distinguishing easily. The active site constructing residues are shown in the black dashed line box under each monomer. An extensive difference in loop L9 has been observed, as these are larger than seen in MupB and TmlB and extended over the active site cavity in FabH and the other KS-III enzymes shown here.

2.7.3 The putative active sites of MupB and TmlB

The optimal alignment of the above discussed proteins with MupB and TmlB revealed similar active site regions (Figure 2.11). In particular, the active site cysteine (C116 and C117) residues of MupB and TmlB align very closely and are invariably positioned between the Nβ3 and Nα3 in a tight bend known as the 'nucleophilic elbow'.[105] In FabH and KAS-III enzymes (EcFabH, ZhuH, OleA, Pks11), the remaining two amino acid residues of the triad, histidine (H) and asparagine (N), appear at the C-terminal of Cβ2 and the N-terminal of Cα3. The catalytic triad residues in the KAS-III like enzymes DpsC, ChlB3 and CerJ, however, differ substantially from all known FabH and KAS-III enzymes and are formed from Ser-His-Asp (DpsC) and Cys-His-Asp (ChlB3 and CerJ) residues, respectively. In KAS-III like proteins, the histidine (H) residue sits at the C-terminal of L12α1 and the aspartic acid lies on the N-terminal of Cα3 helix (Figure 2.11). Active site loop comparison shows that the histidine residue in both MupB and TmlB replaces the active site asparagine in FabH and KAS III active site triads. The aspartic acid residues of MupB and TmlB share invariably the same position (N-terminal of Cα3) as in the KAS-III like proteins. The comparison of the four main conserved residues of representative thiolase fold member proteins, constructing the main part of the active site region revealed that the typical histidine residue in FabH and KAS-III proteins is replaced by S224 in MupB and N233 in TmlB (Table 2.8). Nevertheless, MupB and TmlB surprisingly showed that their presumed active site architectures exactly matched the position and orientation found in CerJ, DpsC and ChlB3. Therefore, TmlB and MupB are found to have non-canonical active site triads which differ from all typical FabH and KAS III enzymes, comprised of Cys-His-Asp rather than Cys-His-Asn (Figure 2.11).
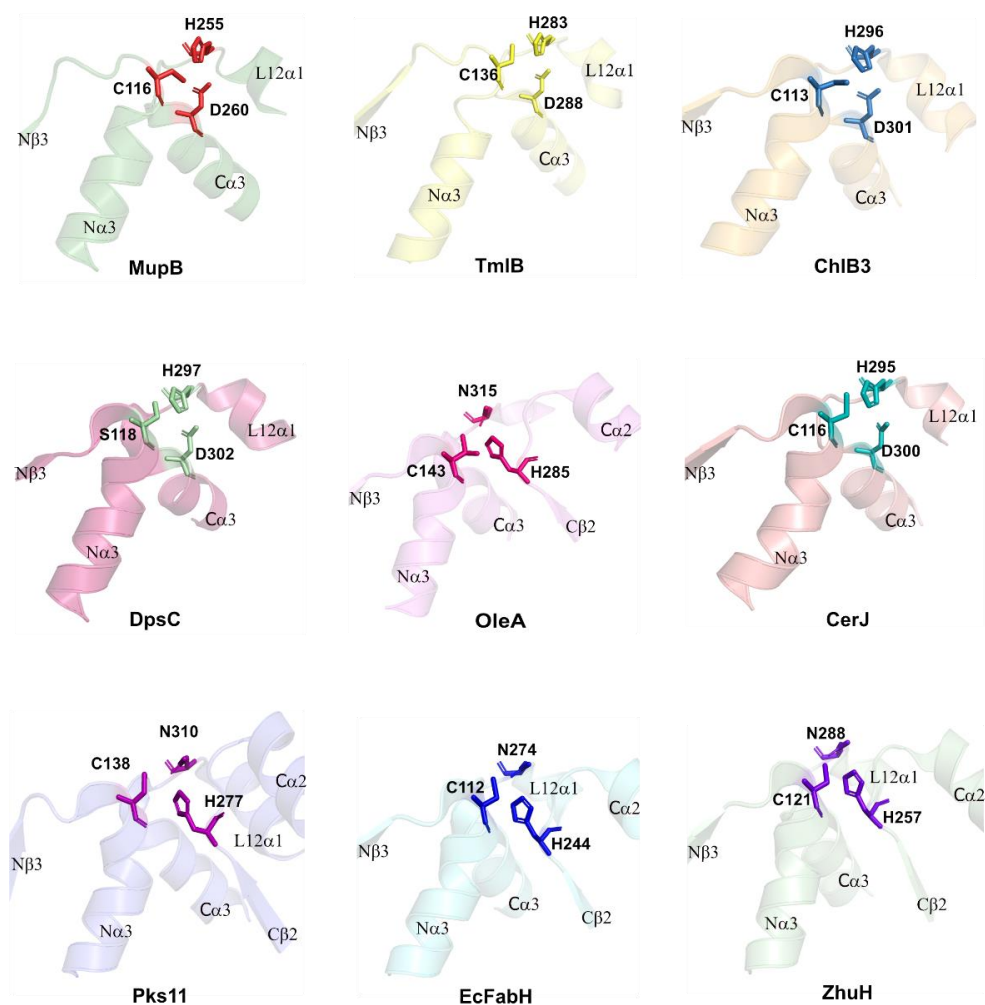
Figure 2.11. Comparison of active site building regions of MupB and TmlB with their structural homologs. The cartoon representation (60% transparent) of active site building loops, strands and helices.

Table 2.8. Comparison of the active sites of thiolase fold members. Alignment of the four conserved sites, building the main part of the of the active site of various thiolase fold members with MupB and TmlB according to the alignment of the entire amino acid sequence. Residues in gray are probably not directly involved in the catalysis. Residue numbers are adopted from the FASTA sequence of the protein database entries (PDB).

Proteins	Active site building residues			
MupB	Cys_{116}	Ser_{224}	His_{255}	Asp_{260}
TmlB	Cys_{117}	Asn_{233}	His_{264}	Asp_{269}
DpsC	Ser_{118}	Pro_{265}	His_{297}	Asp_{302}
FabH	Cys_{112}	His_{244}	Asn_{274}	Ser_{279}
CerJ (KAS III)	Cys_{116}	Pro_{264}	His_{295}	Asp_{300}
ZhuH (KAS III)	Cys_{121}	His_{257}	Asn_{288}	Ser_{293}
HMGS	Cys_{117}	His_{247}	Asn_{326}	Ser_{331}
CHS	Cys_{169}	His_{308}	Asn_{341}	Cys_{390}
THI	Cys_{126}	Asn_{352}	His_{358}	Ser_{390}

Generally, the active site cysteine (C112 in EcFabH) is crucial for the enzymatic reaction (i.e., transacylation) of condensing enzymes through the generation of the thiolate anion on the cysteine thiol group (− SH) which forms the acylated FabH/KS intermediate, whereas the histidine and asparagine are required for the decarboxylative condensation of this acylated enzyme with the incoming malonyl-ACP.[87] The active site Histidine in MupB (H255) and TmlB (H264) occupies the same relative position as the EcFabH asparagine (N274) residue, and the imidazole ring is found to be rotated at 90° with respect to the N274 amide side chain which in both cases brings a nitrogen atom (NE2) within the hydrogen bond distance (Figure 2.12) of the catalytic cysteine, which may increase the nucleophilicity of these residues. The remaining imidazole nitrogen (ND1) can form a hydrogen bond with the other active site residue, aspartic acid (D260 and D269 for MupB and TmlB, respectively) on the Cα3 helix. In both DpsC and CerJ, the nucleophilicity of the S118 and C116 is increased by the presence of residues D302 and D300 respectively which are not found in typical FabH and KAS-III proteins.[62,64] These residues, except the Cys/Ser switch in DpsC, are in the same orientation and on the same secondary structural elements as their MupB and TmlB equivalents (Figure 2.12). Therefore, the nucleophilicity of the active site cysteine residues in MupB and TmlB can also be facilitated via the His-Asp pair in addition to the benefit of the typical helix-dipole interaction generated by the Nα3 helix with the active site cysteine residue.[106] Interestingly, the deviation in catalytic triads of DpsC and CerJ proteins makes them act as acyltransferases rather than displaying ketosynthase like activity. Consequently, MupB and TmlB are highly likely to show similar activity.
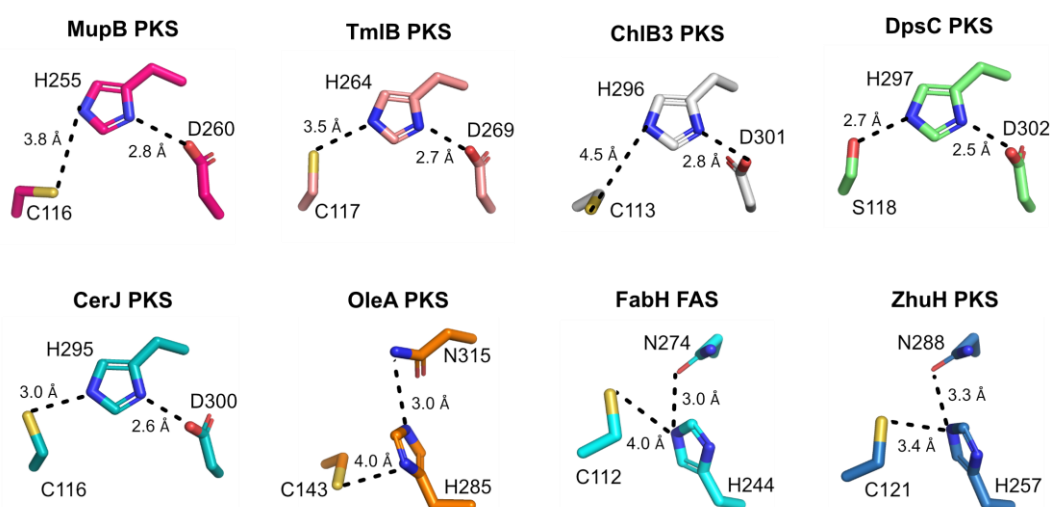
Figure 2.12 Active site triads of KAS III like enzymes: MupB, TmlB, ChlB3, DpsC, CerJ, OleA, FabH and ZhuH. The distances between residues are shown in Å.

2.7.4 Further examples of diverged KAS III sequences.

Further examples of the divergent catalytic triad are evident in enzymes where there is functional characterisation and phylogenetic similarity to MupB/TmlB but currently no structural data are reported (Figure S.9.10, Table 2.9). These include ChlB6 from the chlorothicin biosynthesis pathway in *Streptomyces antibioticus* DSM 40725 reported to catalyse the acyl transfer (esterification) of desmethylsalicyl-CHL (DM-CHL) with 6-methylsalicyl to yield 6-methylsalicyl-chlorothricin.[63,107] A ChlB6 homologue, PtmR from the pactamycin biosynthesis pathway in *Streptomyces pactum* ATCC 27456 was also reported to transfer the 6-methylsalicyl (6-MSA) from the PKS type I PtmQ_ACP to the aminocyclopentitol unit of pactamycin via an esterification reaction.[88,108–110] A few KAS-III like enzymes have also been proposed to catalyse ester formation, among which AviN from the avilamycin biosynthesis pathway in *Streptomyces viridochromogenes* Tü57 is reported to be similar to DpsC but can bind a larger substituted benzoyl substrate.

Table 2.9. 3D Comparison of active site architecture of MupB with some KAS III like enzymes that catalyses esterification and acyl transfer reactions. The table shows the conserved active site except DpsC, having serine instead of cysteine.

KAS III like Enzyme	Organism	Name of NP	% Identity to MupB	Active site triad
MupB	*Pseudomonas fluroscence*	Mupirocin	100	Cys-His-Asp
TmlB	*Pseudoaltaromonus*	Thiomarinol	37.92	Cys-His-Asp
DpsC	*Streptomyces peucetius*	Daunorubicin	14.73	Ser-His-Asp
CerJ	*Streptomyces tendae*	Cervimycin	14.02	Cys-His-Asp
ChlB6**	*Streptomyces antibioticus*	Chlorothricin	17.75	Cys-His-Asp
PtmR	*Streptomyces pactum*	Pactamycin	22.60	Cys-His-Asp
AviN	*Streptomyces viridochromogenes*	Avilamycin	16.90	Cys-Thr-Asp

* For the ChlB6, PtmR and AviN enzymes the homology models were used to compare the active site triad. The homology models were created by I-Tasser (Protein structure and function predictions).[111]

** Note ChlB3 from the same pathways is an AT with known structure that transfers 6-methylsalicyl onto the discrete ACP ChlB2. ChlB6 (no known structure) is predicted to catalyse the acyl transfer (esterification) of desmethylsalicyl-CHL (DM-CHL) with 6-methylsalicyl to yield 6-methylsalicyl-chlorothricin.
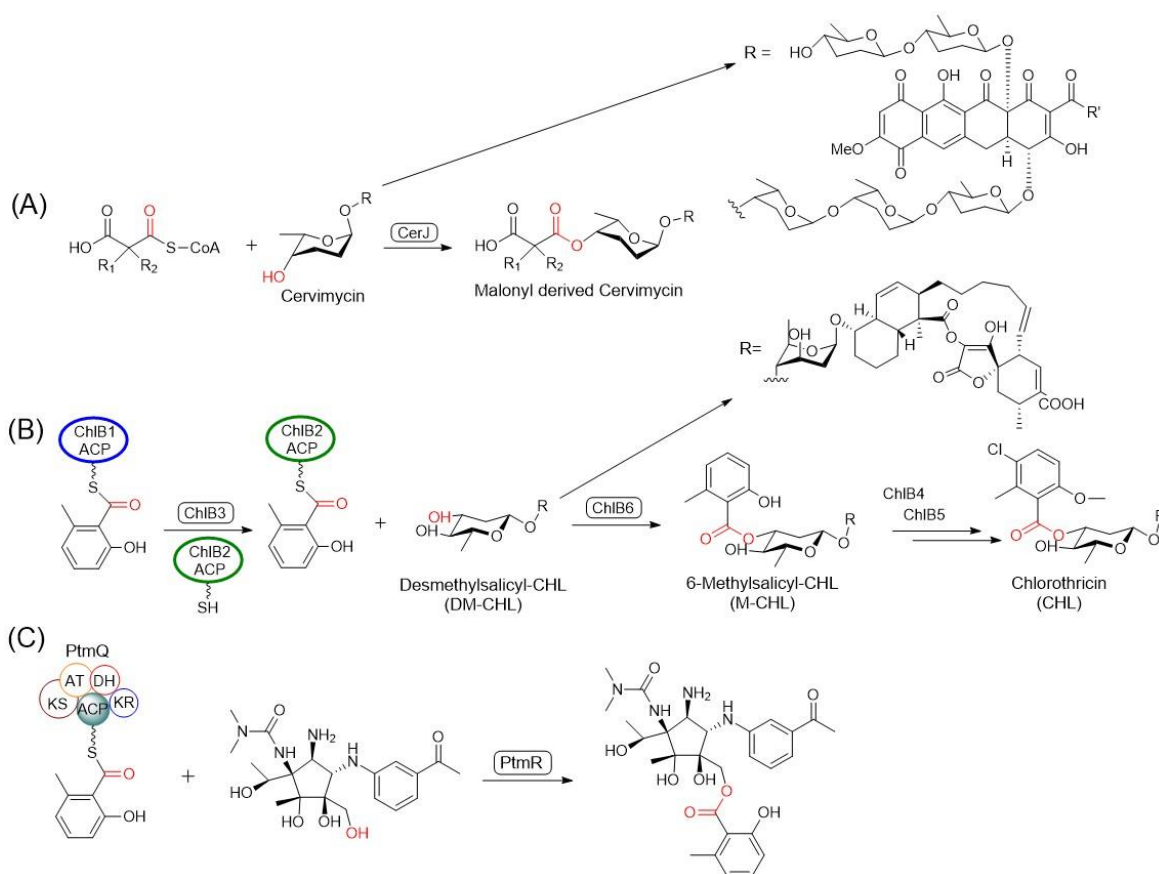
Figure 2.13. Esterification reactions catalysed by KAS III like enzymes. (A) CerJ from the cervimycin biosynthesis pathway transfers an activated malonyl unit onto the sugar residue via C–O linkage. (B) ChlB6 from the chlorothricin biosynthesis pathway catalyses ester bond formation between 5-chloro-6-methyl-O-MSA and the sugar moiety, where the MSA unit is transported by ChlB2_ACP. (C) PtmR can also catalyse ester bond formation between 6-MSA and a pactamycin derivative in the pactamycin biosynthesis pathway.

2.7.5 Putative substrate binding pockets in MupB and TmlB

2.7.5.1 Structural comparison of substrate entrance channels

Structural comparison has been carried out to understand the possible substrate entrance channel and binding pockets. It has been reported that in EcFabH and ZhuH the substrate entrance was built by four main α-helical elements which are the N-terminus L1α2, L7α2 and C-terminus Cα1, Cα2 along with the β strand L9β2.[61,87] Here, the two C-terminal helices create the main entrance, L7α2 and L9β2 form the celling of the active site pocket and L1α2 defines the side wall of the entrance. Similar substrate entrances were also reported for OleA, DpsC and CerJ. The L9β2 strand in the OleA protein has been observed to cover the top of the active site region along with L7α2 and the typical L1α2 helix is replaced by loop L1. On the other hand, the L7 loop in DpsC, CerJ and ChlB3 has no α-helix but caps the active site region along with L9α2 which replaces the typical L9β2 strand common in FabH

and KAS-III proteins. The result is that the substrate entrances of EcFabH, ZhuH, Pks11 and OleA formed with these secondary structures are narrower than those in DpsC, CerJ
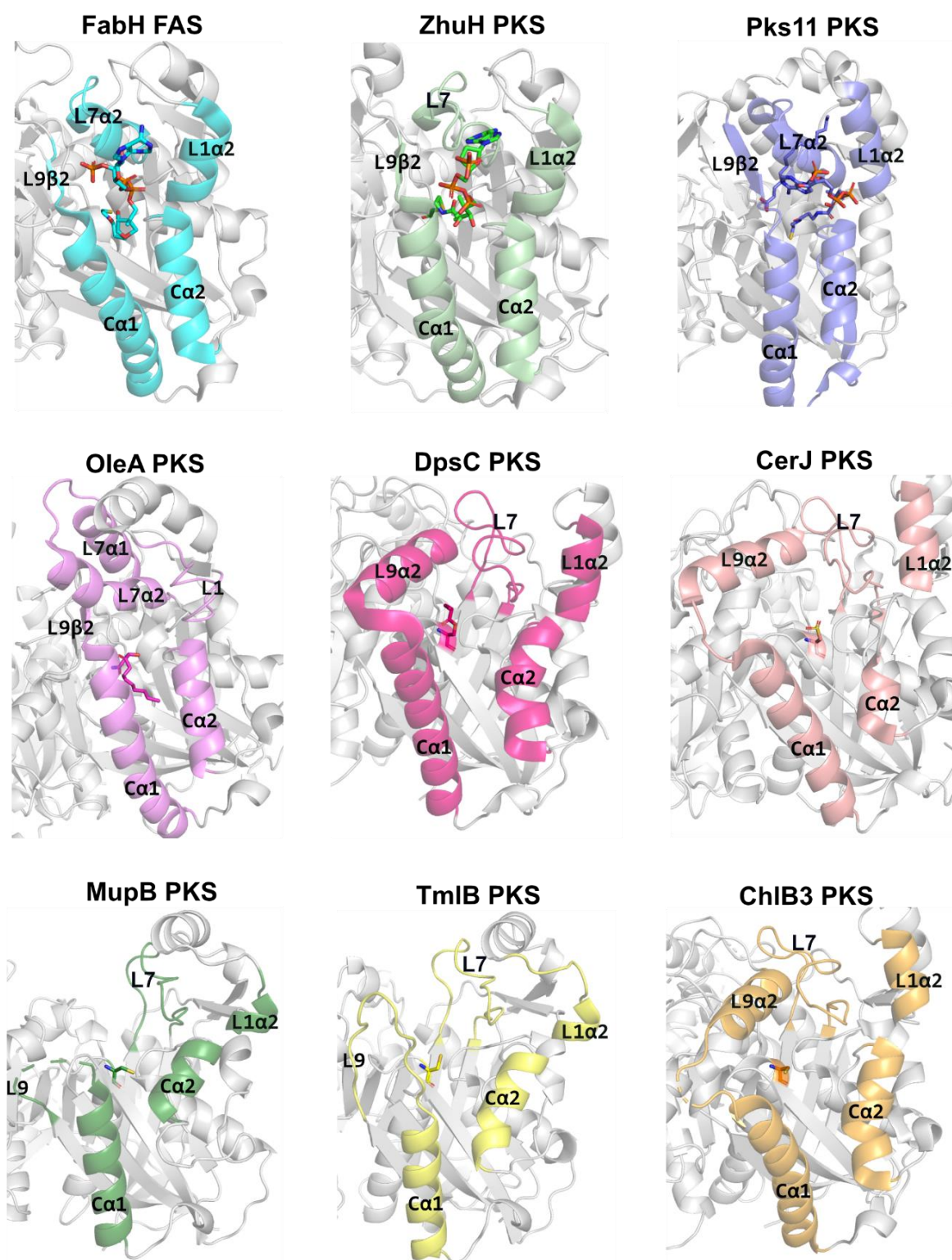
Figure 2.14: The four main α-helical elements, the N terminus L1α2, L7α2 and C terminus Cα1, Cα2 along with the β strand L9β2 responsible for forming the substance entrance to the cavity. All the proteins' secondary structure are drawn as gray cartoons. The substrate entrance building secondary elements are shaded and the active site Ser/Cys residues or bound ligands are drawn as stick models.

and ChlB3 (Figure 2.14). Structural comparison with MupB and TmlB reveals that the equivalent arrangement of these secondary structural elements is found to match more closely with the DpsC, CerJ and ChlB3 channels, although some major differences are observed in L7 and L9. Both MupB and TmlB lack any structural elements in these loops and are oriented outwards to keep the active site cavities more open. The L9 loop can indirectly enhance the flexibility of the pocket size together with L7 to accommodate a small to large molecule as well as branched chain molecules.[61,104,112] Therefore, the arrangements of loops L7 and L9 in MupB and TmlB may be naturally designed to accommodate larger and more complex moleculest like sthe pre-monic acid backbone.

2.7.5.2. Presumed CoA/pantetheine binding pockets of MupB and TmlB

To understand the presumed coenzyme A (CoA) and/or phosphopantetheine (Ppant) binding cavity, the MupB and TmlB structures were compared with CoA (shown in stick) bound 3D structures of EcFabH (cyan), ZhuH (palegreen), Pks11 (purple), CerJ (orange) and Ppant bound DpsC (hotpink) structure (Figure S.9.6). The CoA positions from EcFabH and s ZhuH occupy the proposed substrate entrance channels of MupB and TmlB and reach the active site regions as expected. However, the pyrophosphate and adenosine groups in the EcFabH, ZhuH and Pks11 structure are found to clash with the L1α1 and L7 loops of both MupB and TmlB (Figure S.9.6 A). On the other hand, the CoA and Ppant binding positions from CerJ and DpsC, respectively are well accommodated (Figure S.9.6. B). Therefore, the CoA/Ppant binding pocket architectures of CerJ and DpsC were compared to investigate the possible CoA/Pant binding residues in MupB and TmlB. It has been reported that the terminal phosphate group of the pantetheine analogue in DpsC interacts strongly with positively charged K221 and R271 (shown in black dashed line, Figure 2.15 A, B). No positively charged residues in MupB and TmlB were found in the same position. Instead, K234 and K243 in MupB and TmlB, respectively were found in the same secondary structure element, Cα2, which runs in an opposite direction compared to DpsC. In MupB the backbone amide of D196 and T197 along with K234 may form a suitable architecture to interact with the terminal phosphate group of Ppant. A similar architecture is also observed in TmlB comprising D205, T206 and K243 residues. The two methyl groups of the Ppant analogue in DpsC forms hydrophobic interactions with V327 (Cβ3) which is substituted with polar N285 and S294 in MupB and TmlB, respectively. The CoA binding pocket in CerJ is formed from several hydrophobic residues including V161, Y164, A327 that match Y155, F158 and A286 in MupB and Y157, F160 and A295 in TmlB. The CoA

in CerJ is covalently linked via a cysteine bridge with active site C116 and electrostatic interactions of R268 to the phosphate groups and hydrogen bonds to the backbone amide of the G271 and the side chain of H272. The positively charged R268 in CerJ is observed to be replaced with hydrophobic residue F227 and Y236 in MupB and TmlB. No polar residue matches with the positions of G271 and H272 and again, the sidechains of D196 and T197 for MupB and D205 and T206 for TmlB point away from the CoA (Figure 2.15 C,D). A list of common amino acid residues in MupB and TmlB that form the CoA/Ppant pockets in EcFabH, ZhuH, DpsC, CerJ is summarized in Table 2.9.
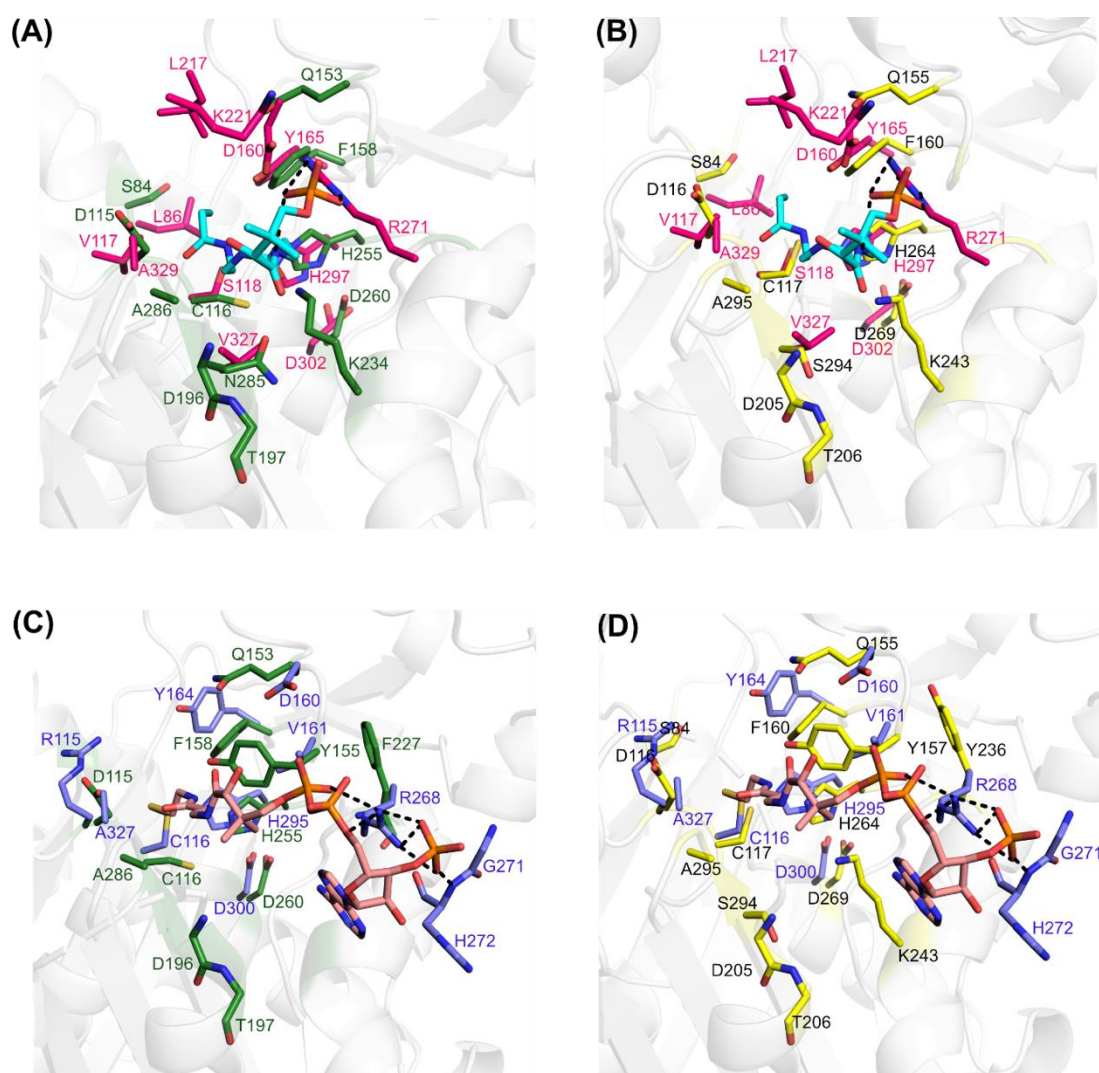
Figure 2.15. Comparison of CoA/Ppant binding residues of MupB and TmlB: DpsC active site and residues that interact with Ppant analogue (hotpink) are compared with **(A)** MupB (green) and **(B)** TmlB (yellow). CoA binding residues in CerJ (light blue) are compared with **(C)** MupB (green) and **(D)** TmlB (yellow). The polar interactions between the substrates and DpsC and CerJ are shown in black dashed line and the residues are shown in sticks

Table 2.9. CoA/Ppant binding site residues of ZhuH, DpsC and CerJ are compared with MupB and TmlB.

ZhuH	M90	A120	C121		L165	F166		N288			A320
DpsC	L86	V117	S118	D160		Y165		H297	D302	V327	A329
CerJ		R115	C116	D160	V161	Y164	L277	H295	D300	G325	A327
MupB	S84	D115	C116	Q153	Y155	F158	K234	H255	D260	N285	A286
TmlB	S84	D116	C117	Q155	Y157	F160	K243	H264	D269	S294	A295

Therefore, the CoA/Ppant binding channel comparison shows that both MupB and TmlB have the typical phosphopantetheine binding channels, but the terminal phosphate group may flip toward the positively charged residues (discussed above) of the N-termini of the Cα1 secondary elements. K234 and K243 from Cα2 in MupB and TmlB, respectively may interact with the hydroxyl group to hold the phosphopantetheine arm into the channel. Also, Y155, V230, N285 and A286 in MupB and Y157, V239, S294 and A295 in TmlB may create the hydrophobic and polar environments for the pantetheine arm to fit inside the pantetheine channels.

2.7.5.3. Presumed alkyl binding pockets of MupB and TmlB

The presumed acyl group binding pocket of MupB and TmlB was investigated by structural threading with homologous acylated enzymes (found by DALI search using 3D structure of MupB and TmlB) listed in Table 2.10. In all these structures, the acyl substrates are found to form a thioester/ester bond with active site cysteine / serine and the carbonyl oxygen of the acyl moiety occupies the oxyanion hole (OxH) formed by an arrangement of the hydrogen-bond donor amino acid residues.[113] The activation energy for the reaction is lowered due to the transition state stabilization these donors provide.

In butyryl-DpsC (PDB:4xsb), the oxyanion hole is formed by the amide protons of S118 and A329 residues whereas, L86, V117, L217 and L93′ forms the hydrophobic binding pocket for the butyryl moiety (Figure 2.16 A). By comparison S118 and A329 are replaced by C116, A286 in MupB and C117, A295 in TmlB. The alanine backbone amides are oriented away however from where they might be expected if they were to stabilise an oxyanion. The hydrophobes V117 and L86 are replaced by charged and polar residues D115, S84 in MupB and D116 and S84 in TmlB. Malonyl-CerJ (PDB:3t6s) forms an oxyanion hole using the backbone amides of C116 and A327 and also employs R115 and Y164 to bind the negatively charged carboxylate group of the malonyl group (Figure 2.16 B). The oxyanion residues of CerJ are observed to be replaced by C116, A286 in MupB and C117, A295 in TmlB. Both MupB and TmlB lack these latter residues which stabilize the

carboxylate group and instead have D115/F158 and D116/F160 respectively. The cerulenin-OleA complex has revealed the active site C143 is encompassed by two oxyanion holes, the first formed by S347, E117´ (OxH 1) and the second H285, N315 (OxH 2) and three substrate channels. OleA catalyses the head-to-head condensation of two long chain acyl-CoA molecules and one substrate channel sequentially coordinates the pantetheine moiety of these acyl-CoA substrates. OxH 1 and alkyl channel A are thought to act first in the trans-thiolation of the first fatty acid acyl-CoA onto C143 while OxH 2 and alkyl channel B act in the Claisen condensation with a second long chain fatty acyl-CoA substrate.[102,114]

Table 2.10. Acyl binding channel residues of ZhuH, DpsC, CerJ and OleA are compared with MupB and TmlB.

ZhuH	M90	T´96		C121	F166		N288			A320	
DpsC	L86	L´93	V117	S118	Y165		H297	D302		A329	
CerJ	L84	L´91	R115	C116	Y164		H295	D300	G325	A327	
OleA	V111	E´110		C143		H285	N315				S347
MupB	S84		D115	C116	F158	S224	H255	D260	N285	A286	D287
TmlB	S84		D116	C117	F160	N233	H264	D269	S294	A295	S296

Another important residue that controls the size of the acyl moiety binding pocket is the presence of large the bulky hydrophobic side chain of F87´ from the opposite monomer in EcFabH which is highly conserved in FabH enzymes and restricts binding to only small chain (C2-C3) substrates.[87] T96´ in ZhuH, L93´ in DpsC and L91´ in CerJ all play similar roles. The polar T96´ in ZhuH helps it to accommodate a branched acyl chain of four carbons, whereas the less bulky hydrophobic leucine residues in DpsC and CerJ also allow branched and nonbranched acyl substrates of four carbons. Amino acid sequence threading (Figure S.9.4) showed that both MupB and TmlB do not have any equivalent residue to match with T96´ from L3´ of the opposite monomer in ZhuH. It is reported that M90 from L3 and T96´ from L3´ in ZhuH shape the size and bottom of the acyl binding pocket (as with other KAS-III like enzymes that are discussed above where L3-L3´ is restrictive). In MupB and TmlB, L3´ is oriented well away and leaves the conserved position widely open. Therefore, the non-canonical architecture of L3-L3´ in both MupB and TmlB may accommodate larger acyl groups like pre-monic acid (Figure S.9.7).

S84, D115, C116, H255 in MupB and S84, D116, C117, H264 in TmlB may therefore be involved in forming a more polar acyl binding pocket in the immediate vicinity of the thioester bond. MupB and TmlB may also possess two alkyl channels, A and B, presumably tailored to bind (i) pre-monic acid and (ii) the 9HN moiety on the phosphopantetheine arm

of an ACP. To further explore channels A and B, 3D structures were overlaid with decane-1-thiol bound FabH (PDB:2QX1), phosphopantetheine analogue bound DpsC (PDB:4XS9) and cerulenin bound OleA (PDB:3S20) (Figure 2.17 A and 2.18 A). For MupB, the three substrates are accommodated at three different positions.
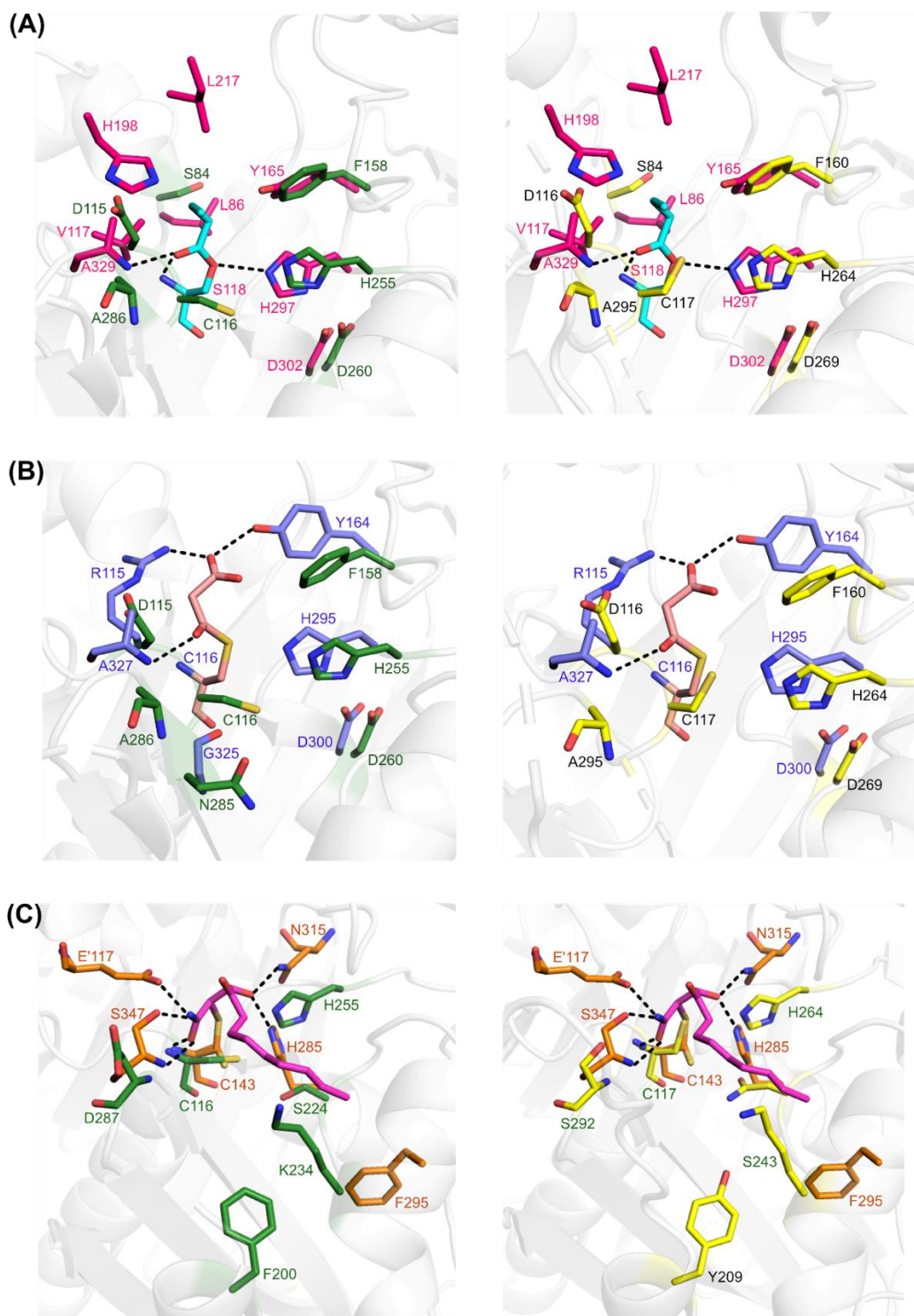
Figure 2.16. Acyl binding sites of (A) butyryl-DpsC, (B) malonyl-CerJ and (C) cerulenin-OleA are compared with MupB (green) and TmlB (yellow). The polar interactions between the substrate and oxyanion hole residues are shown in black dashed lines.

The pantetheine arm in the phosphopantetheine channel and the other two substrates (decane-1-thiol and cerulenin) occupy potential alkyl channels as proposed for OleA (Figure 2.17 A). Figure 2.17B shows two possible OxHs, A and B that can be identified for MupB. OxH A involves D115, N143 and S258 and one water molecule and suggests a putative alkyl binding channel A. OxH B is occupied by two water molecules H-bonded to S224 and H255 residues and may correspond to a second alkyl channel B (note: for Ole-cerulenin, cerulenin was presumed to be binding to the first alkyl binding site which was termed A and is equivalent to B here. For the decane-1-thiol structure, the substrate occupied what we term channel A. Essentially these are thought to be able to switch depending on the specific reaction).

The electrostatic charge map (PyMOL) shows the proposed pantetheine channel is positively charged. Alkyl channel A is mostly negatively charged and alkyl channel B is mostly hydrophobic (Figure 2.17C and D, Figure S.9.8. A). Comparing the decan-1-thiol bound alkyl channel A of FabH to MupB reveals MupB has more polar and charged residues including S84, S88, E87′, D115, E146, Q153 and S258 (Figure S.9.9) and several scaffolding hydrophobic residues along one side (Figure 2.17C).
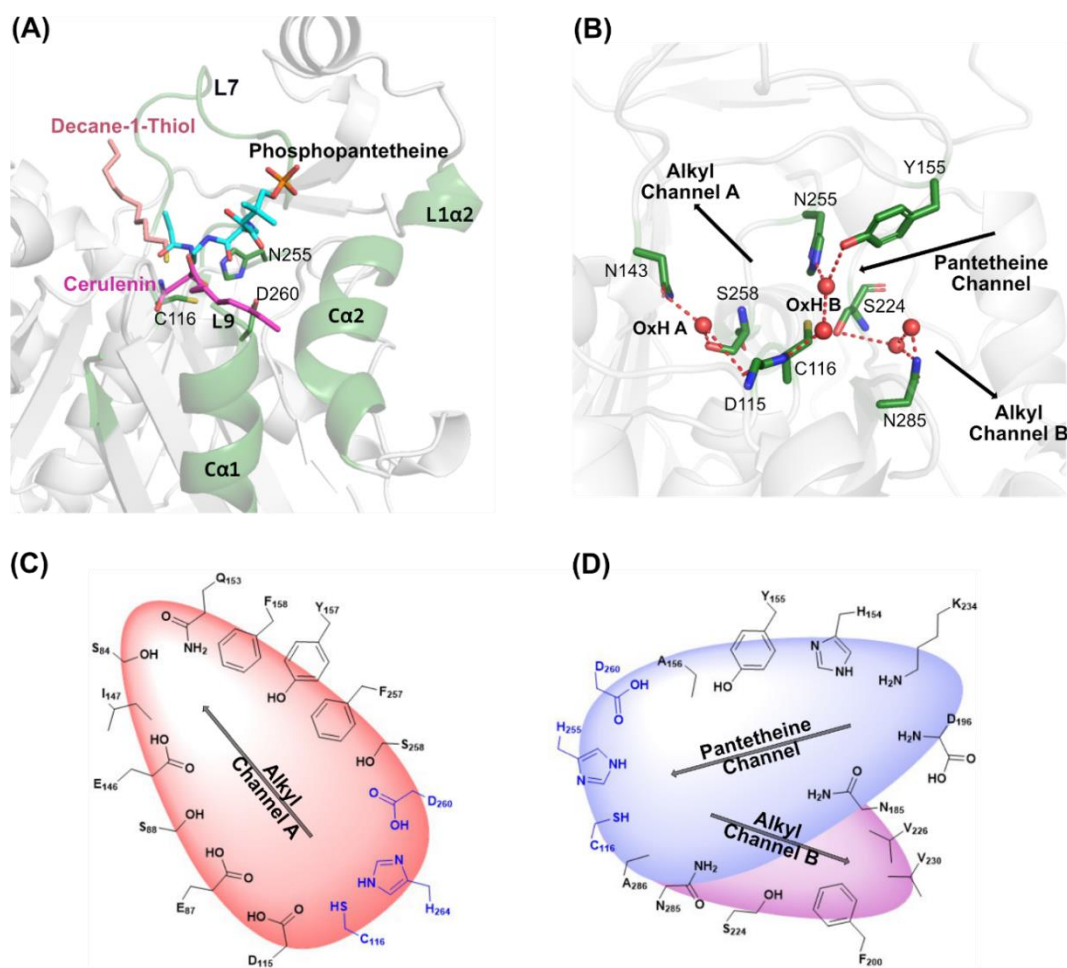
Figure 2.17. Illustration of MupB substrate binding channels. (A) MupB 3D structure overlaid with FabH, DpsC and OleA. MupB is drawn as a grey cartoon showing the active triads and the conserved secondary elements building substrate entrances into the active site are shown in green colour. For clarity, the cartoons of FabH, DpsC and OleA have been omitted and their bound ligands are represented as stick models (brown, cyan and magenta, respectively). (B) Putative assignment of MupB binding channels. Red spheres are ligand water molecules occupying OxH A and OxH B producing three possible substrate binding channels. Labelled arrows indicate the position of the alkyl and pantetheine channels in MupB. (C) The egg-shaped schematic diagram representing the alkyl channel A building residues and (D) representing the residues that may construct the pantetheine and alkyl channel B. The active site triads are shown in blue colour.

A similar architecture is found in TmlB, with identifiable alkyl channel's A and B that can again accommodate decane-1-thiol and cerulenin when superimposed with these homologous enzyme-substrate crystal structures (Figure 2.18 A). Several water molecules are also found around the C117, H264 and D269 active site triad, filling the presumed acyl binding pockets (Figure 2.18 B). By analogy to the MupB analysis, two OxHs are present. Two bound waters define OxH A which involves N119, N145 and S267 that surround one side of C117. OxH B on the other hand is defined by further bound waters on the other side of C117 and includes N233 and S294 adjacent to the catalytic H264. The three substrate binding channels display the same charge/non-polar distribution. Channel A is again lined

by negative and polar charged residues with one hydrophobic face (Figure 2.18C, S.9.8 B, S.9.9B). Conversely the pantetheine channel contains charged and polar side chains of D205, T206, K243 and alkyl channel B has a mostly hydrophobic (e.g., L235, I247, A295) environment (Figure 2.18 D).
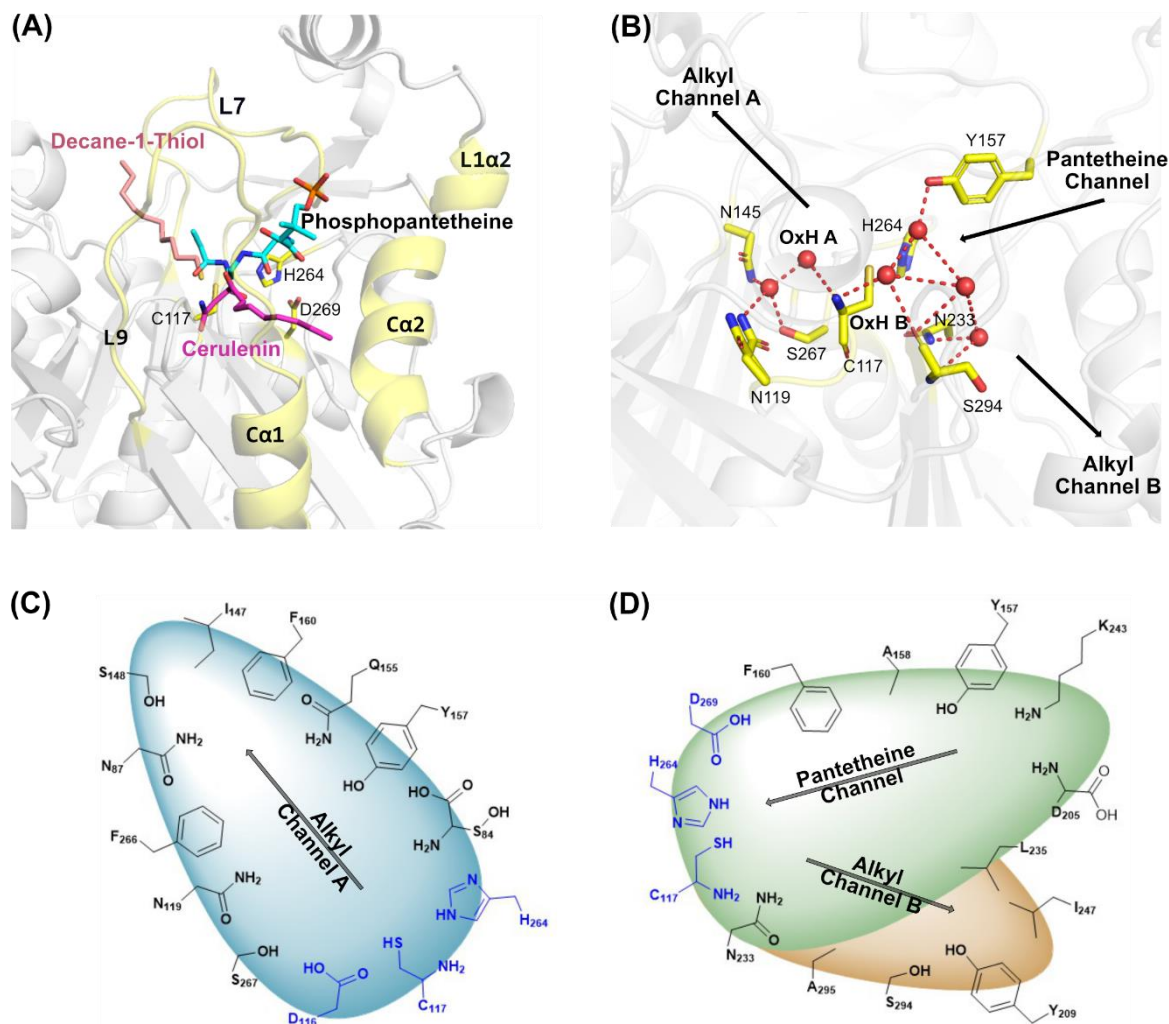
Figure 2.18. Illustration of TmlB substrate binding channels. (A) TmlB 3D structure superimposed with decane-1-thiol-FabH, Ppant-DpsC and Cerulenine-OleA and the ligands are shown in sticks. In the grey cartoon of TmlB, catalytic active sites and conserved elements forming the substrate entrance are shown in yellow colour. (B) Proposed pantetheine channel and two alkyl channels A and B are observed to be occupied with water molecules. (C) Schematic representation of TmlB alkyl channel A and (D) Egg shaped representation of possible pantetheine and alkyl binding channel B. The active site triads are shown in blue colour.

Hence, it can be proposed that the esterification may first start with self-acylation of the pre-monic acid and the protein-ligand complex will be stabilized by OxH A (D115, C116 and N143 for MupB) and occupy alkyl channel A. Once primed with an acyl group, MupB/TmlB may catalyze esterification in the presence of the appropriate hydroxylated fatty acid

delivered by a further ACP (again currently unknown). The esterified product may be generated by activation of the nucleophilic oxygen of the hydroxylated fatty acid in the presence of H255 (OxH B) which can then attack the carbonyl carbon of the acyl-adduct. The schematic illustration of the proposed reverse esterase activity of MupB has been presented in Figure 2.19.
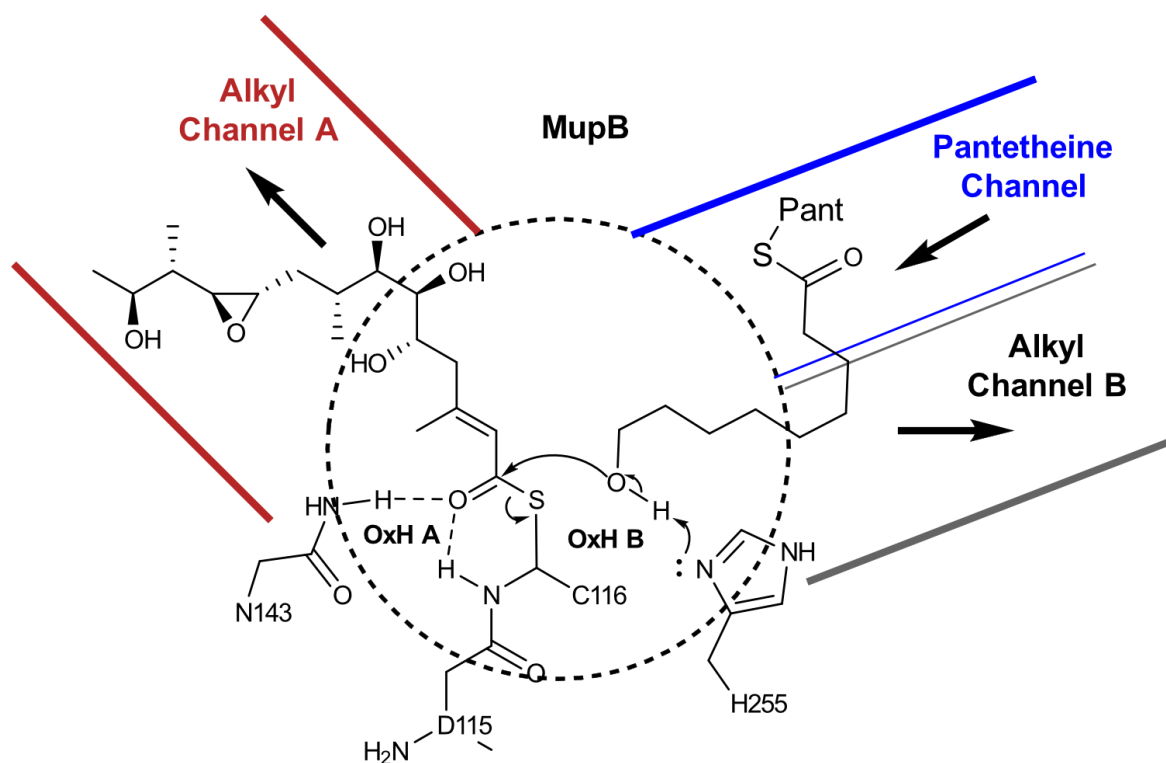
Figure 2.19. Substrate coordination and proposed esterification reaction mechanism in the active site pocket of MupB. The catalytic active site C116 along with the proposed OxH A and B are shown inside the circle (dashed-lined). The pre-monic acid bound to Ppant/ACP first enters the active site through the pantetheine channel (blue) and is then covalently linked with C116 and arranged inside the alkyl channel A. In a similar way, the 9HN chain will be delivered by another Ppant/ACP and inside the OxH B the hydroxyl oxygen atom will be activated to finally produce the esterified product.

2.8. Summary

Cloning, expression, purification, and crystallization of MupB and TmlB allowed the determination of high-resolution crystal structures in the same space group P 1 2_1 1. In an ASU, MupB has been observed to crystalize as a tetramer whereas TmlB is found as a dimer. The PDBePISA analysis confirms that both these proteins exist as stable dimeric conformations in solution. Extensive structural analysis has revealed that MupB and TmlB have conserved thiolase-folds, common to the FabH and KAS-III like enzymes but containing an alternate catalytic triad that deviates from the typical KAS-III and is comprised of Cys-His-Asp rather than Cys-His-Asn. In addition, one pantetheine binding channel and two alkyl channels have been deduced for both MupB and TmlB along with two possible oxyanion holes A and B.

3. Self-acylation studies of MupB and TmlB

Self-acylation must be the first step in the esterification reaction in the mupirocin biosynthetic pathway whereby MupB is hypothesized to be primed with pre-monic acid. This chapter details *in vitro* studies of the self-acylation reaction of MupB and TmlB with several substrates along with simplified mimics of pre-monic acid (3-methylbut-2-enoyl, MBE). This chapter also reports the high-resolution co-crystal structure of MBE-TmlB. Site-directed mutagenesis and the structural analysis of MBE-TmlB co-crystal reveals self-acylation occurs as expected at the active site cysteine of TmlB and reveals important information about the orientation of this substrate.

3.1 Self-acylation assay: MS analysis

Self-acylation reaction assays of MupB and TmlB were performed to confirm their biological activity and substrate specificity. Various chain length acyl-CoAs/pantetheines, including extended fatty acid substrates and simplified mimics of pre-monic acid, 3-methylbut-2-enoyl (MBE), were used as potential substrates due to lack of availability of the natural substrate, the C_{17} pre-monic acid backbone (Figure 3.1). All the pantetheine derivatives used in this work were supplied by Dr Angus Weir.
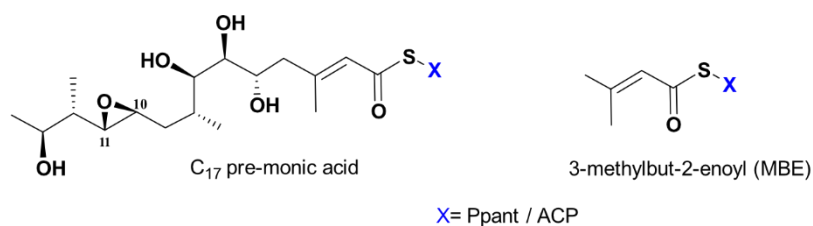
Figure 3.1. Chemical structures of pre-monic acid and its simplified derivative MBE.

3.1.1 Self-acylation assay of MupB

Purified MupB protein (30 µM) was incubated with the substrate (1 mM) for approximately 3 hours, in Tris buffer (50 mM Tris, 150 mM NaCl, pH = 7.5) and self-acylation was analysed by ESI-MS.
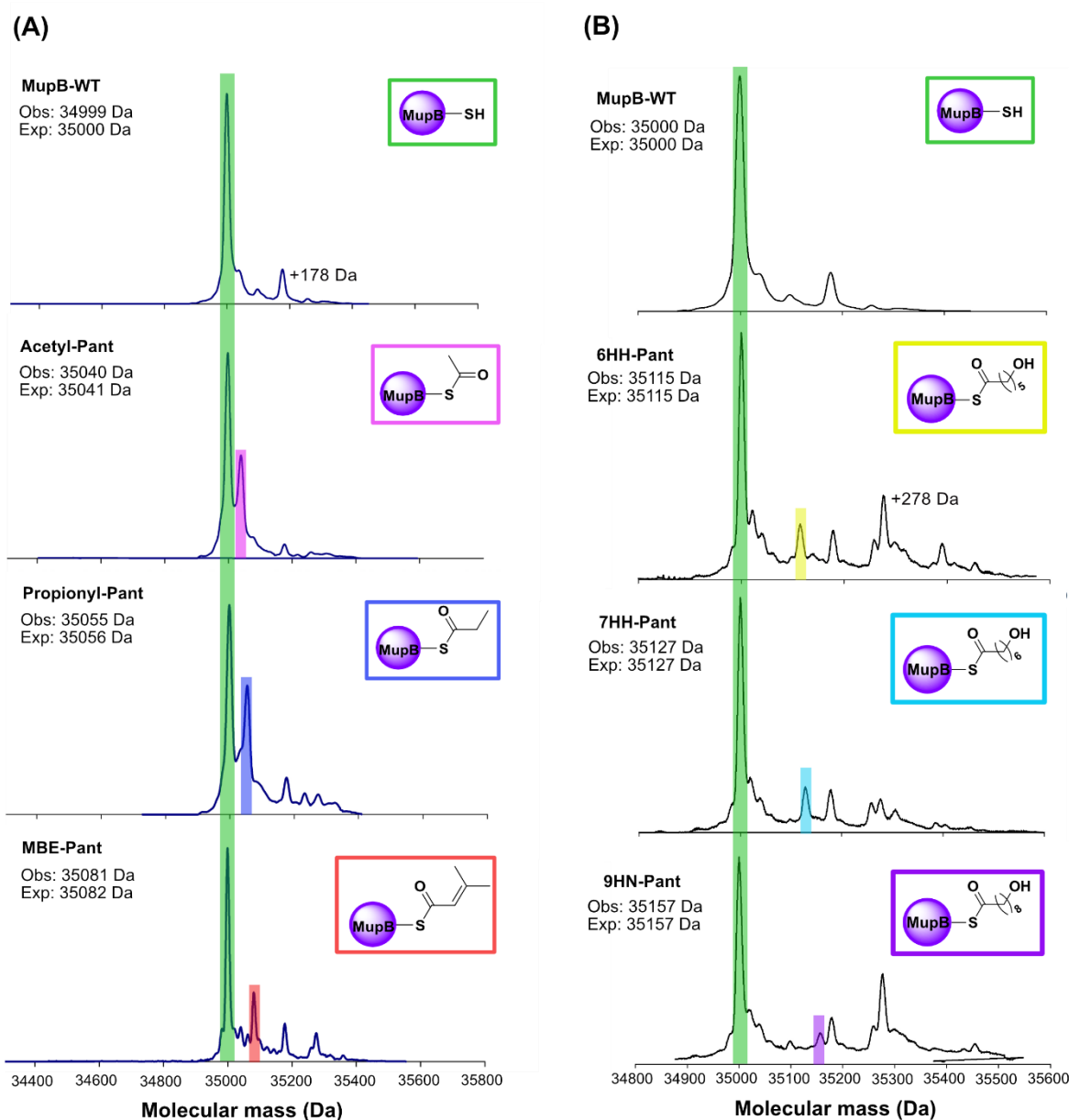
Figure 3.2. Deconvoluted ESI-MS spectra of purified MupB after incubation with different substrates. (A) Representing the self-acylated MupB with acetyl, propionyl any and MBE substrates. (B) The long fatty acid chain substrates (6HH, 7HH and 9HN) self-loading were also observed in the ESI-MS analysis.

Incubation of MupB with acetyl-, propionyl- and MBE-pantetheines showed observable self-acylation with as much ~30-40% loading with the acetyl and propionyl groups (Figure 3.2A and Table 3.1). A selection of hydroxylated long chain fatty acid pantetheine substrates 6HH, 7HH and 9HN was also tested and in this case, all loaded but with very poor efficiency. The poor selection for the longer fatty acyl substrates was expected as MupB is not expected to prime with these groups although a degree of recognition is expected as they must be selected for once MupB is correctly acylated in order to perform esterification.

3.1.2 Self-acylation assay of TmlB

TmlB (30 µM) was incubated with 1 mM of substrate for 3 hours prior to analysis with ESI-MS. The mass spectrometric analysis showed that around 80% of the TmlB was self-acylated with the MBE (Figure 3.3). Interestingly, no acylation was observed with acetyl or propionyl substrates (Table 3.1) suggesting that TmlB might be more selective for the presence of a β-branch, at least in the context of this simplified substrate.
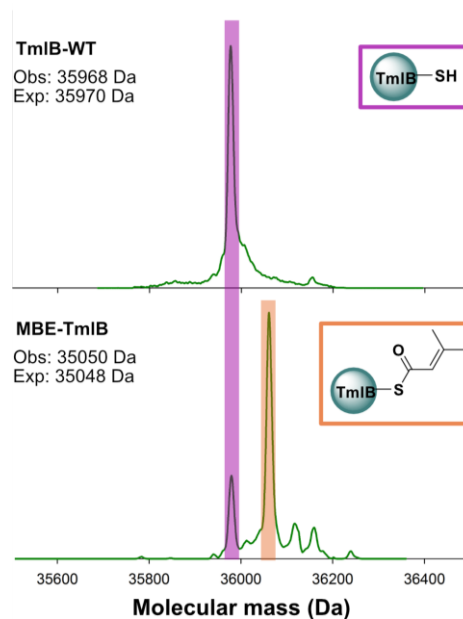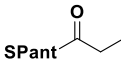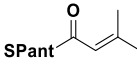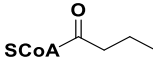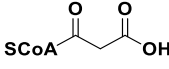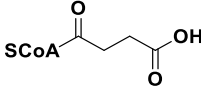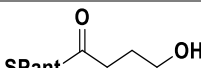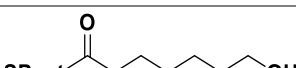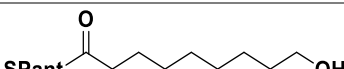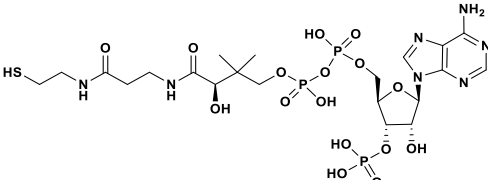
Figure 3.3. Deconvoluted ESI-MS spectra of TmlB after incubation with MBE-Pant showing ~ 80% loading of enoyl substrate into the TmlB protein.

Based on these results, it has been demonstrated that MupB has non-specific selection toward branched MBE substrate and could self-load long chain (number of C > 4) fatty acid substrates. The lack of specificity towards branched and non-branched substrates was also seen in the KAS-III and KAS-III like enzymes, ZhuH and CerJ respectively. ZhuH was reported to prefer propionyl-CoA and isobutyryl-CoA mostly as priming unit prior to elongation.[115] The malonyl transferring enzyme CerJ was found to select and esterify malonyl-CoA, methyl-malonyl-CoA, and di-methyl-malonyl-CoA substrates onto the sugar residue of cervimycin-K via an ester linkage.[62]

On the other hand, TmlB is highly selective towards the branched MBE substrate. Interestingly, MupB and TmlB do not self-load malonyl-CoA or succinyl-CoA suggesting the presence of a β- or γ-carboxyl group is selected against. Importantly however, the incomplete substrate loading onto MupB and TmlB indicates the likely necessity for the appropriate ACP for efficient substrate delivery.[116]

Table 3.1. A summary of substrates employed for self-acylation assays, their chemical structures and results of experiments with MupB/TmlB.

Substrate	Structure	Self-acylation status of MupB	Self-acylation status of TmlB
Acetyl-Pantetheine		√	×
Propionyl-Pantetheine		√	×
MBE-Pantetheine		√	√
Butyryl-CoA		×	×
Malonyl-CoA		×	×
Succinyl-CoA		×	×
4-hydroxy butyryl Pantetheine		×	×
6-hydroxy hexanoyl Pantetheine		√	×
7-hydroxy heptanoyl Pantetheine		√	×
9-hydroxy nonanoyl Pantetheine		√	×
Coenzyme A		×	×

3.2 Effects of site directed mutagenesis on self-acylation

Site-directed mutagenesis was employed to verify the role of the active site cysteine in the self-acylation reaction. MupB:C116A and TmlB:C117A mutants were successfully generated by whole plasmid PCR mutagenesis using mutagenic primers (Table S.9.1) from the pOPINF_MupB and pOPINF_TmlB parental template DNA constructs and sequenced by GENEWIZ.

The MupB:C116A mutant was expressed in T7 express cells (NEB) in an identical manner to MupB-WT (Section 2.1). Purified MupB:C116A was spin concentrated to 5 mg/L, flash

frozen in liquid N_2 and stored in -20 °C for the further use. The successful mutation and purification of MupB:C116A where the deconvolution of m/z envelope revealed the observed molecular mass of 34967 Da (expected 34968 Da) (Figure. 3.4).
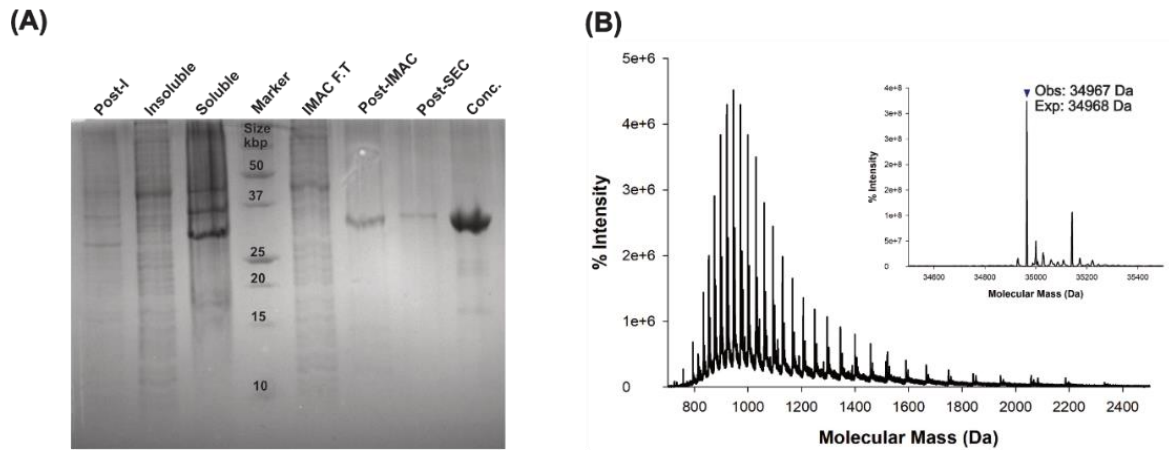
Figure 3.4. MupB:C116A mutant protein purification and characterization. (A) SDS-PAGE gel showing successful expression of MupB:C116A as soluble protein is observed in the soluble sample, further isolation (via IMAC) and purification (via SEC) samples showed bands at ~ 37 kDa. (B) The ESI-MS spectrum of MupB:C116A.

The TmlB:C117A mutant was also expressed in T7 express cells (NEB). Purification revealed that the protein was expressed as insoluble inclusion bodies of the cell pellets and was not taken on further (Figure 3.5).
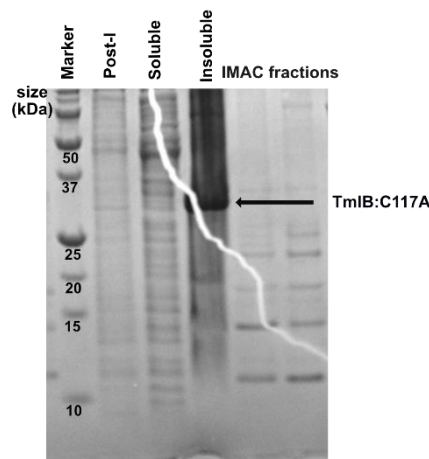
Figure 3.5: The SDS-PAGE gel image of TmlB:C117A expression and isolation. The TmlB:C117A is observed to express in the insoluble inclusion body showing the band near 37 kDa. The isolation of protein from the soluble part using IMAC showed no correct band therefore indicating no isolation of TmlB:C117A protein from the soluble part.

3.2.1 Self-acylation assay of MupB:C116A

The self-acylation assay of MupB:C116A mutant was investigated with the substrates listed in figure 3.8 (C). 30 μM of MupB:C116A protein was incubated with 1 mM of the substrate in the Tris buffer (pH 7.5) at room temperature for approximately 3 hours prior to analysis by ESI-MS (Figure 3.6). Surprisingly it was observed that MupB:C116A self-acylated with the branched substrate like MBE-pant to a similar degree as MupB-WT, despite the lack of the active site cysteine residue. Moreover acetyl, propionyl and coenzyme A also self-acylated and to a greater degree than wild type (Figure 3.2). The mechanistic basis for this observed self-acylation mechanism of MupB-WT was unclear. Examination of MupB/TmlB, FabH and KAS-III active sites had however revealed four conserved residues that form the core of the active site (Table 2.8). The three-dimensional structural comparison of these four conserved residues showed MupB to have a S224 residue in exactly the same position as H244 of EcFabH and H257 of ZhuH (Figure 3.7). Conversely, the MupB active site residue, D260, is substituted in both EcFabH and ZhuH by serine residues S279 and S293 respectively. The MupB Cys-His-Asp catalytic triad closely matches the CerJ equivalents except for this S224 which is replaced by P264 in CerJ and proposed to assist in substrate binding (Figure 3.7). The structural analysis therefore suggested that S224 might act as a surrogate nucleophile in absence of the C116 residue. This activity might only arise significantly in the absence of the natural cysteine residue which presumably either loads far more quickly and then blocks access to S224 or the activity of S224 is only uncovered once C116 is mutated, perhaps through perturbation of the active site.

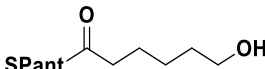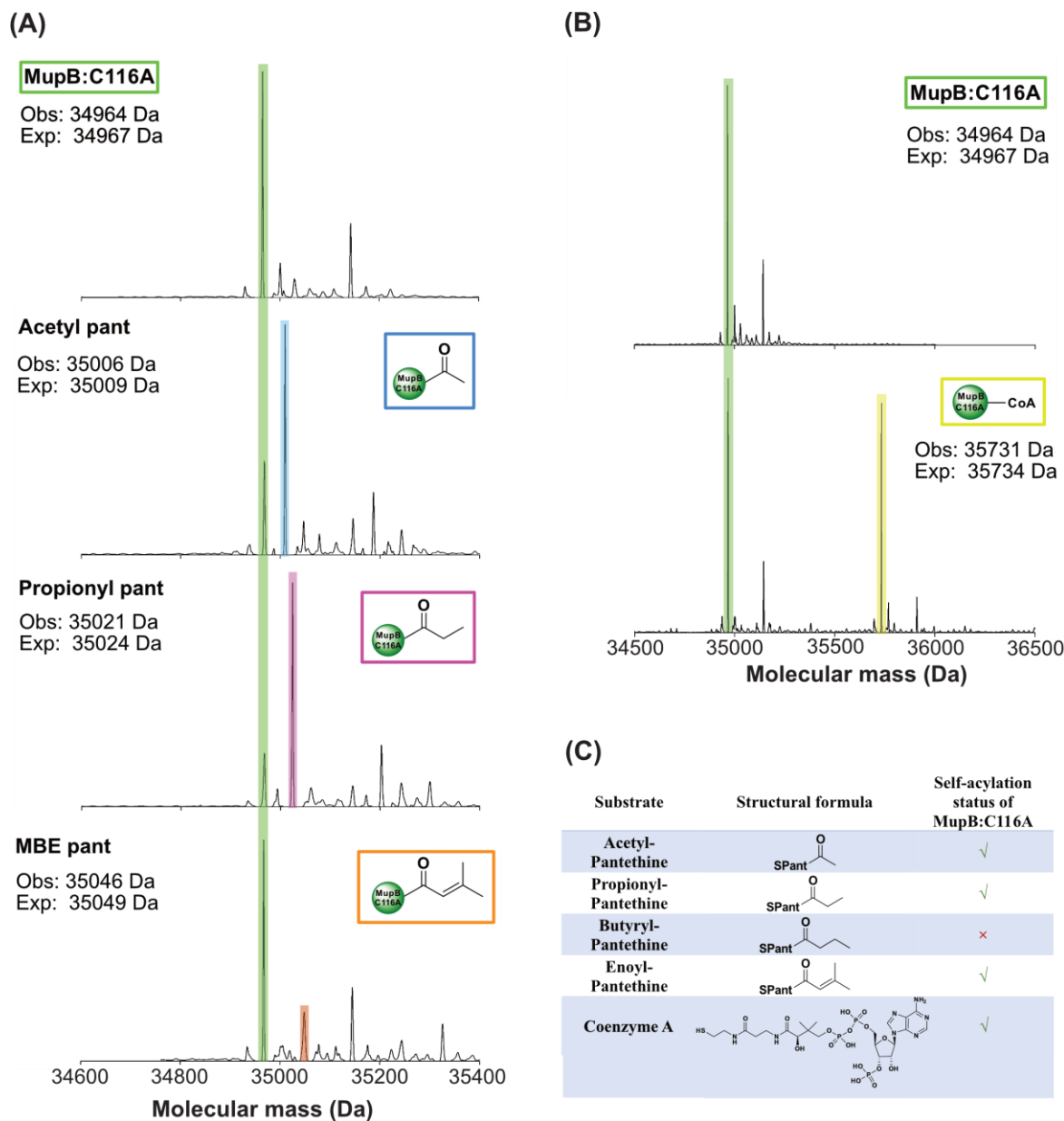**Figure 3.6.** The deconvoluted ESI-MS spectra of MupB:C116A after incubation with different substrates. (A) representing the ESI-MS spectra of MupB:C116A control in green colour and its acylated peak with acetyl (sky blue), propionyl (pink) and enoyl (orange) substrates. (B) CoA loaded MupB:C116A ESI-MS spectra is represented with yellow colour, showing 50% self-loading of CoA. In both cases the expected and observed molecular masses were mentioned in the respecting figures. (C) representing the chart of substrates used in MupB:C116A self-acylation assay along with the assay status after the reaction.
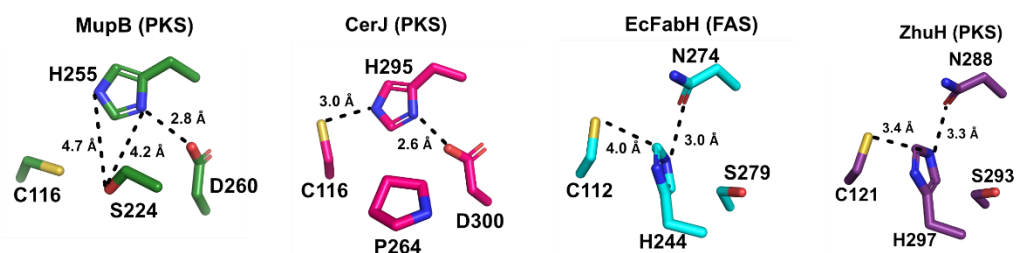
Figure 3.7. The 3D structural comparison of the four conserved residues of MupB (green) with thiolase-fold member proteins CerJ (hotpink) EcFabH (cyan), and ZhuH (violet) which build the main part of the active site.

To understand the role of S224 residue on self-acylation reaction of MupB:C116A the MupB:S224A mutant protein was tested in the *in vitro* self-acylation assay. It was of course the intention to also generate the double mutant but due to COVID-19 restrictions this was not feasible in the time available.

3.2.2 Expression, purification, and characterization of MupB: S224A

The MupB:S224A mutant gene was synthesized and subcloned into a pET151 vector (Thermo Fisher Scientific). MupB:S224A was expressed and purified in an identical manner to MupB-WT (Section 2.1). Analytical SEC confirmed that this mutant still retained a dimeric fold and ESI-MS confirmed the presence of the mutation (Figure 3.8).
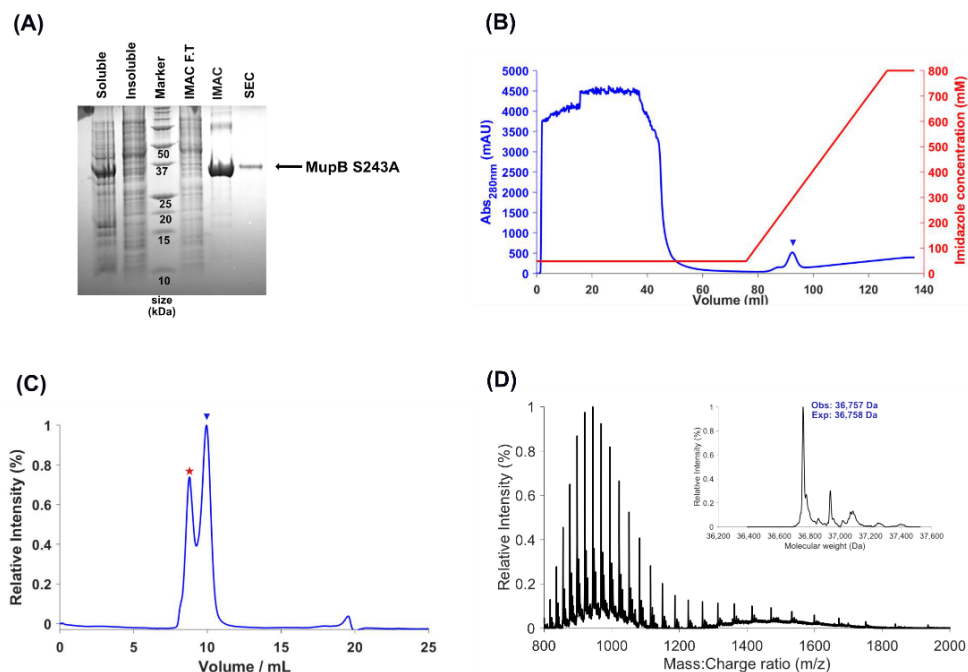
Figure 3.8: MupB:S224A mutant protein purification and characterization. (A) The SDS-PAGE gel suggested the successful expression and isolation of MupBS224A as soluble sample, IMAC and SEC further samples provided bands ~ 37KDa. (B) The IMAC chromatogram is reporting the isolation of MupB:S224A protein (marked with blue triangle) from the soluble part and (C) Analytical SEC of MupB:S224A eluting as dimeric species (blue triangle) where the first peak shows the void volume (red star). (D)The ESI-MS confirms the successful mutation of observed monomeric mass 36757 Da for MupB:S224A (expected 36758 Da).

3.2.3 Self-acylation studies of MupB:S224A

The self-acylation activity of MupB:S224A was performed under identical conditions to MupB:C116A, and again using acetyl-, propionyl- and MBE-pantetheine substrates. ESI-MS analysis after 3 hours showed that MupB:S224A could self-acylate the small chain acetyl and propionyl substrates but no significant loading of branched chain MBE substrate was observed (Figure 3.9). After 16 hours of incubation the MBE substrate showed a limited loading (Figure S.9.12).

The assay results for both MupB:C116A and MupB:S224A, demonstrate that acylation can occur in both mutants to varying degrees. Interestingly in the absence of S224, the loading of the MBE-Pant is completely abolished and loading of acetyl and propionyl groups is reduced, perhaps suggesting the serine residue may modulate the reactivity of C116. Further work with a double mutant will more precisely define the role of the S224 but C116A may have applications in providing an alternate mechanism through which to cross-link with ACPs for structural studies. Overall however, the limited substrate loading onto WT-MupB and mutant MupB may suggest a stringent requirement for the

MupB:S224A
Obs: 36,757 Da
Exp: 36,758 Da

Acetyl pant
Obs: 36,799 Da
Exp: 36,801 Da

MBE pant
Obs: 36,757 Da
Exp: 36,840 Da

Propionyl pant
Obs: 36,812 Da
Exp: 36,814 Da

Figure 3.9. Deconvoluted ESI-MS spectra for MupB:S224A (green) after 3 hours of incubation with substates. The MupB:S224A is acylated with acetyl (sky blue) and propionyl (purple) but no acylated peak is observed with MBE substrate after 3 hours. The observed and expected mass change due to acylation are mentioned in the left side of the respecting figures. (The asterisk peak at +178, is for the N-terminus gluconylation of the protein).

appropriate substrate or the need for loading from the correct ACP rather than a CoA thioester, or likely both. Repeating the self-acylation assays with a C_{17} pre-monic acid derivative or shorter truncations will therefore yield more information on the mechanism of

this reaction. Currently, various pantetheine derivatives of the C_{17} pre-monic acid are being synthesized.

3.3 Co-crystallisation of TmlB with MBE-pantetheine

To further investigate the enzymatic self-acylation mechanism and consolidate the function of the deduced active site residues, a co-crystal structure of TmlB with MBE-Pant substrate was determined. The co-crystallization was achieved by incubating purified TmlB ($310\,\mu M$) with 5 mM of MBE-Pant substrate for 3 hours and self-loading monitored by ESI-MS of this mixture [Figure 3.10]. This mixture was then directly applied to crystal screens (i.e., without any further separation) which were prepared using 384-well commercially available molecular dimension screens (Morpheus, PACT, JCSG+ and Structure 1+2, See section 6.7). A single successful hit was observed after ~ 7 days with no further crystals appearing after a further 7 days. The crystals were looped and cryoprotected using glycerol prior to freezing in liquid N_2. X-ray diffraction data was collected at the Diamond Light Source synchrotron.
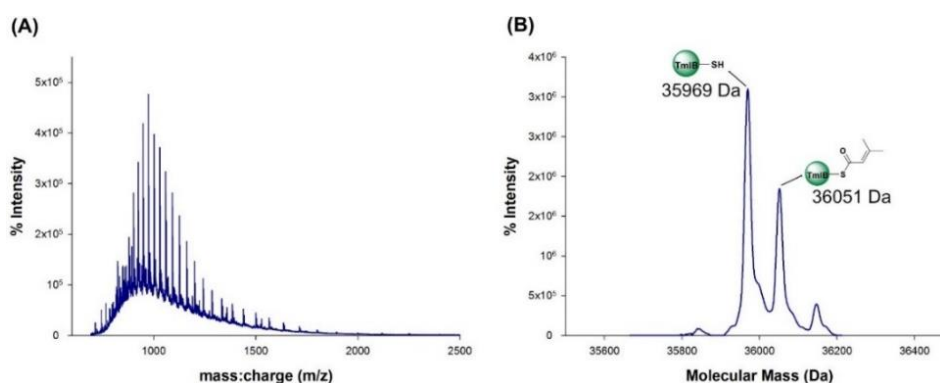
Figure 3.10. (A) is the ESI-MS charge envelop of the MBE-TmlB self-acylation reaction. (B) The narrowly spaced m/z spectra were deconvoluted to the observed mass of 36051 Da (expected 36052 Da).

A data set of 1800 images were extracted and the auto-processing by Diamond Pipeline (FAST DP) suggested a 1.59 Å overall resolution of X-ray diffraction by the MBE-TmlB co-crystal. The data processing and structural determination was done with the assistance of Dr Ash Winter (PDRA, University of Bristol) with a R-free of 0.22 and R-work of 0.18. The MBE-TmlB co-crystal structure revealed the same space group as WT-TmlB, P 1 2_1 1 with cell dimensions of a=52.13 b=89.11 c=65.89. The TmlB structure was used as a search model within the Phaser-MR and Autobuild pipeline to determine the MBE-TmlB co-crystal structure. Two molecules were observed in the asymmetric unit (ASU) with a solvent

content of 34% (Figure 3.11). The crystallographic data and refinement statistics are provided in Table 3.2.

Table 3.2. The crystallographic data collections and refinement statistics of MBE-TmlB co-crystal structure.

MBE-TmlB					
Crystallographic Data		**Refinement**		**Geometry**	
Wavelength (Å)	0.9686	Resolution (Å)	44.56-1.86 (1.89-1.86)	RMS bonds (Å)	0.0099
Space Group	P 1 2$_1$ 1	protein atoms	4721	RMS angle (°)	1.72
Unit cell a,b,c (Å)	52.13, 89.11, 65.89 α=γ= 90°, β=106.88°	ligands atoms	11	Ramachandran favoured (%)	97.65
Resolution (Å)	44.56-1.86	Solvent		Ramachandran allowed (%)	1.33
Observations	299830 (14850)	R-work	0.18	Ramachandran outliers (%)	0.00
Unique observations	48280 (2475)	R-free	0.22	Clash-score	2.66
Completeness (%)	99.7 (99.6)			Average B-factors (Å^2)	
I/σI	7.3			Proteins	29.08
R$_{merge}$	0.149			Solvent	33.89
Wilson B-factor	29.58			Ligands	33.23

PDBePISA analysis of the MBE-TmlB co-crystal structure suggested the stable dimeric assembly of TmlB based on the value of standard free energy of dissociation, ΔG_{diss} = 23.3 kcal/mol. MBE-TmlB forms a dimer of two monomeric subunits through a dimeric interface exactly the same as the TmlB-WT structure, burying a total of 4410 Å^2 surface area. The highly negative solvation energy of folding values implied that both monomers were correctly folded. One assembly (crystal split) was found to be stable at AB[CNE][NA]$_3$ composition where CNE and NA are ligands (MBE and Na$^+$) with accessible surface area (ASA), Buried surface area (BSA), Solvation energy effect (**ΔG_{int}**), Standard free energy of dissociation (**ΔG_{diss}**) presented in Table 3.3.

Table 3.3: PDBePISA analysis of MBE-TmlB dimer observed within the 1.59 Å dataset. Monomers were investigated for their total number of atoms, residues, total number of surface atoms and residues. Total solvent accessible area along with interface surface area are also shown.

Interface parameter of Monomers							
Monomer	N_{at}	N_{res}	sN_{at}	sN_{res}	Interface Area [Å^2]	Total Area [Å^2]	ΔG [kcal/mol]
A	2364	304	1257	267	1793.4	13533.3	-286.9
B	2329	299	1234	261	1796.1	13270.4	-281.7
Average:	2346	301	1245	264	1794.75	13401.9	-284.3

Stable Crystal Assemblies							
PQS No	mm size	Composition	Stable	ASA [Å^2]	BSA [Å^2]	ΔG_{int} [kcal/mol]	ΔG_{diss} [kcal/mol]
1	2	AB[CNE][NA]$_3$	Yes	23130	4410	-58.9	24.4

3.3.1 Structural analysis of MBE-TmlB

The MBE-TmlB co-crystal structure was compared and overlaid with the TmlB-WT dimeric structure using PDBeFold (Table S.9.8), indicating that subunits A and B closely matched between the two structures with RMSDs over backbone Cα atoms of 0.13 Å and 0.16 Å, respectively. Unlike the TmlB-WT structure, residues 217 to 222 on subunit B could not be refined due to lack of electron density. The co-crystal structure clearly demonstrated that the substrate MBE was incorporated into subunit A of the TmlB dimer and covalently linked to the C117 as a thioester, which was confirmed by inspection of the electron density map obtained from the rigid body refinement (Figure 3.11 A). From the electron density map, clear electron density could be observed for the carbonyl group with the adjacent covalent linkage to C117. However, electron density is weaker for both the methyl group and the $C_\alpha-C_\beta$ double bond, suggesting that these are not coordinated well by the active site pocket, hence the weak density (Figure 3.11 B). The possible reason for weak density at this region may be due to the dynamics/conformational exchange of this curtailed substrate mimic.

(A)

(B)
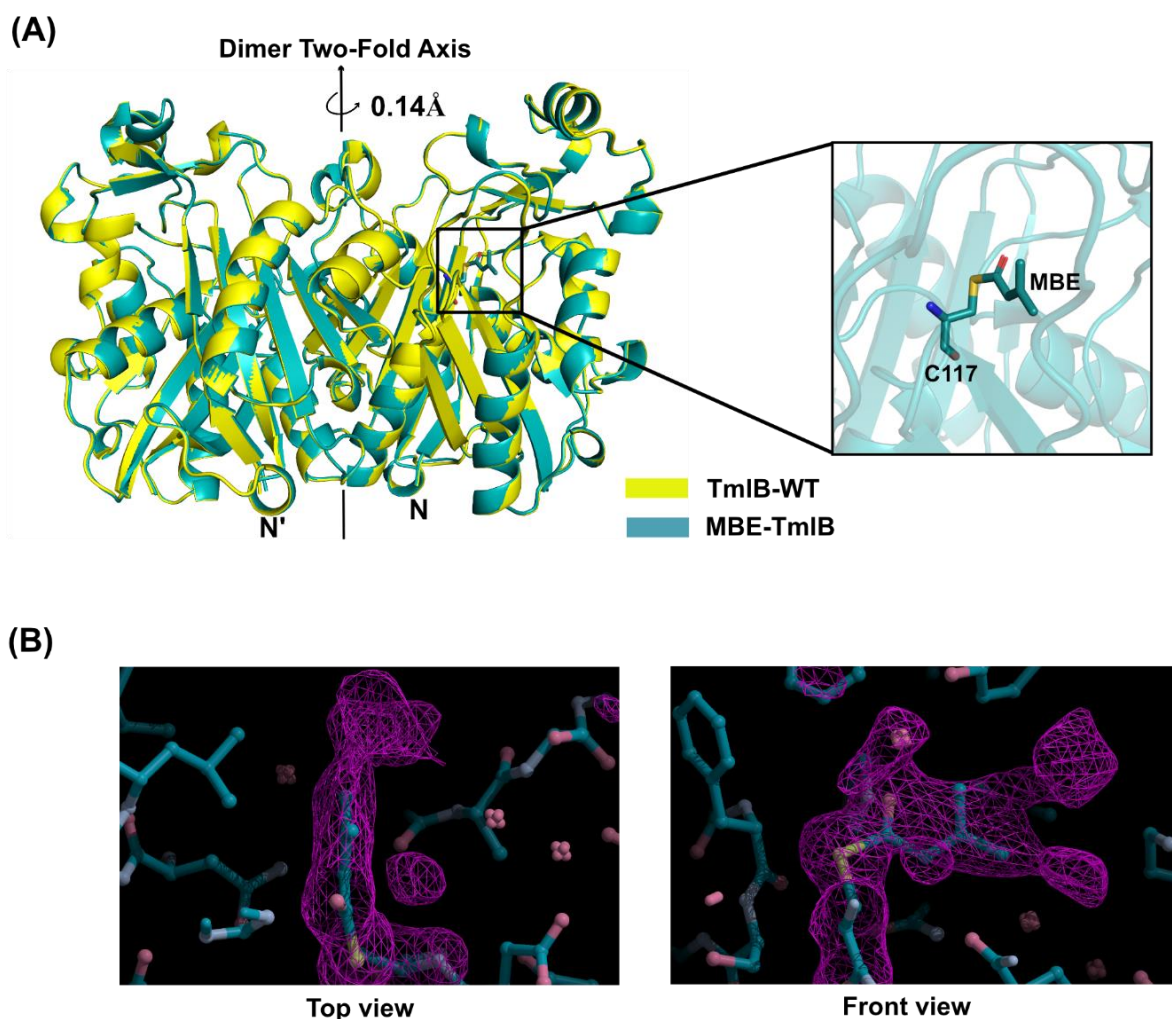
Figure 3.11. (A) The 3D secondary structural elements alignment of WT-TmlB against MBE-TmlB co-crystal with the RMSD of 0.14 Å for the dimeric structure. The square black box is enlarged to represent the covalently linked MBE substrate with the C117 residue via a cysteine bridge. The cartoon images were prepared by using PyMOL educational software. (B) The atomic model fitted the electron density map showing the top and front view of the MBE substrate bound to the C117 residue.

Further structural investigation showed the MBE substrate to be localized in a cleft formed with L5 (connecting Nβ3 and Nα3), L7, L11, the tloop connecting L12α1 and the Cα3 and C-terminus of Cβ4 (Figure 3.12 A and B). This orientation points MBE into the presumed alkyl channel B (OxH B) where the carbonyl oxygen is stabilised via electrostatic interaction between the H264 (NE2). In addition, several hydrogen and hydrophobic interactions hold the substrate into the active site cleft. A water molecule (W205) interacts with the C117 (SG) via water hydrogen bond formation (Figure 3.12 C and D) and this W205 molecule occupies the presumed alkyl channel A in TmlB (Figure 2.17). The extended architecture of the expected acyl binding pocket (alkyl channel A) shows mostly charged and polar

residues like S114, D115, N118 and N233 in the vicinity of C117 whereas the alkyl chain of MBE substrate orients such that it interacts with several hydrophobic residues i.e., Y157 and L235 (Figure S.9.13).
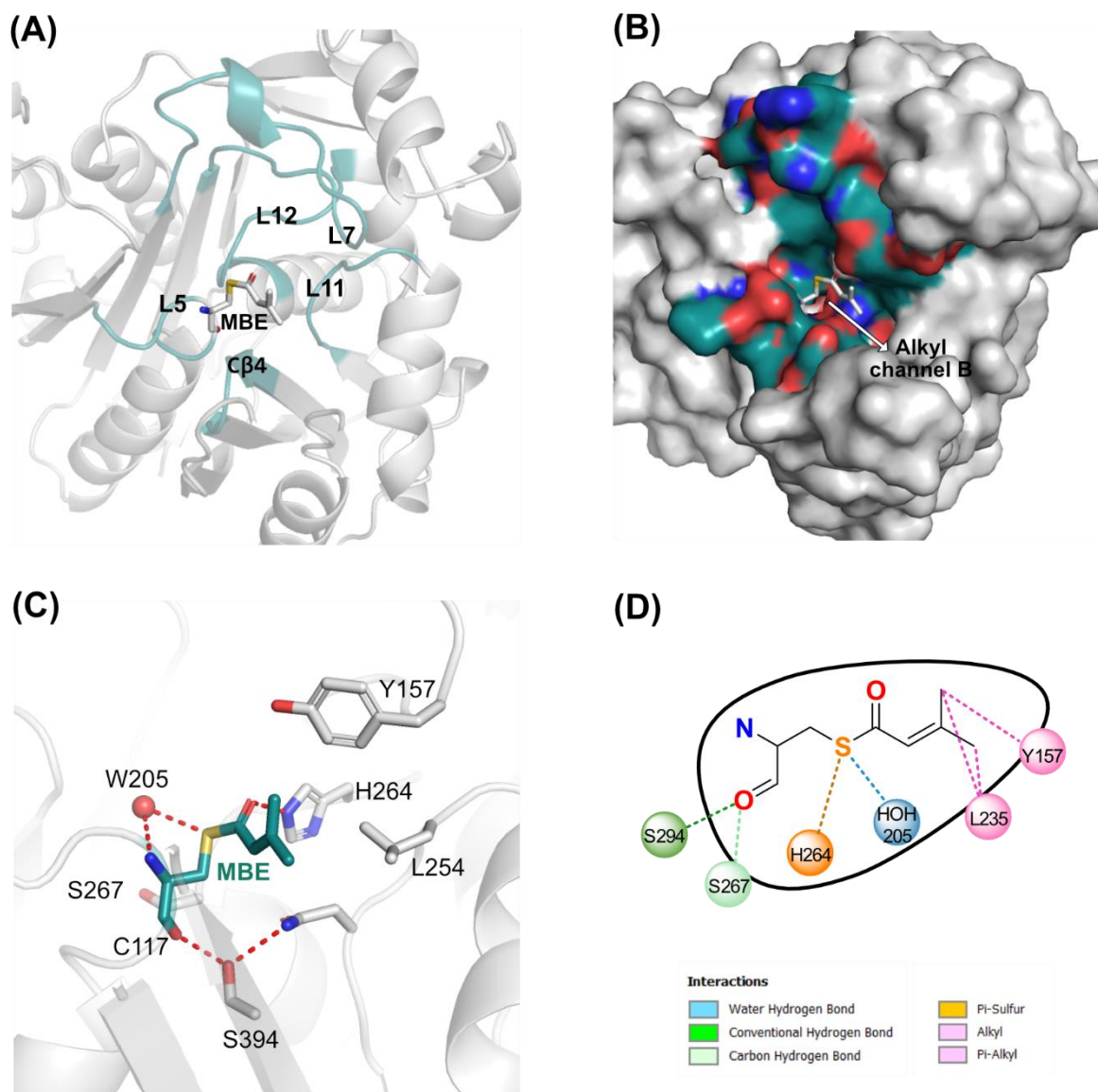
Figure 3.12. (A) The loops, creating the substrate binding pocket are shown in cyan-green colour and the MBE is shown using a white stick model. For clarity, the TmlB cartoon is shown in grey. (B) The surface model showing the substrate binding pocket and the MBE oriented toward the presumed alkyl binding channel B. (C) The 3D representation of MBE interacting residues are shown. The red dashed lines represent the polar interactions. (D) 2D representation of non-bonding interactions between MBE and residues that are responsible to stabilize the substrate. The different types of interactions are shown in different colours.

Subunit B showed no evidence of binding substrate and is instead occupied with several water molecules. No significant changes between the active site triad conformations were apparent when comparing MBE bound TmlB subunit A (Figure S.9.14) and non-ligated subunit B. Similar cases have been reported for homologs (i) CerJ covalently linked with malonic acid in subunit B only, whereas subunit A coordinated an acetate ion of crystallisation buffer[62] and (ii) C143S mutant OleA, non-covalently bound to myristic acid and myristoyl-CoA in two different channels (OleA has proposed two alkyl channel A and B) of subunit B only.[114] It has also been reported that, WT-OleA co-crystallised with cerulenin in a different space group with only one monomer in each ASU in comparison to the cerulenin (soaked) crystal structure of WT-OleA.[102] Moreover, malonyl-CoA (soaked) CerJ was reported to be crystallised as a dimeric structure in the ASU where the malonic acid was covalently bound only in one subunit, which was unexpected at the time. Therefore, the MBE incorporation into TmlB may be limited by the selection of only one monomer and may help explain the limited modification observed in acylation assays (i.e., maximum 50% could be expected) (Figure 3.10).

3.3.2 TmlB exhibits a noncanonical alkyl binding channel

A PDBeFold search of the MBE-TmlB crystal structure produced similar hits to the WT-TmlB (Table S.9.9). Among those, acetyl-FabH (PDB:4z19), propionyl-DpsC, malonyl-CerJ and cerulenin-OleA crystal structures were superimposed with MBE-TmlB (Figure 3.13). Interestingly, the MBE moiety was found to coordinate in a different orientation compared to the substrates bound to FabH, DpsC and CerJ, which differs from the typical acyl binding channel. Instead, the MBE substrate occupies the identical position of cerulenin bound to OleA which is proposed as alkyl channel B in TmlB (Figure 2.17). The carbonyl oxygen of MBE is found to orient towards the OxH 1 of cerulenin bound OleA which matches with the proposed OxH B of WT-TmlB. In chapter two (section 2.7.5.3), it has been proposed that alkyl channel A coordinates acyl substrate (pre-monic acid mimic, MBE in this case). But being mostly hydrophobic in nature MBE orients towards the hydrophobic alkyl channel B (in TmlB) (Figure 3.13). Therefore, these discussions suggests that the MBE-TmlB co-crystal structure reveals an unusual acyl binding channel like OleA. The importance of having this extra alkyl channel has already been discussed and proposed in chapter two.
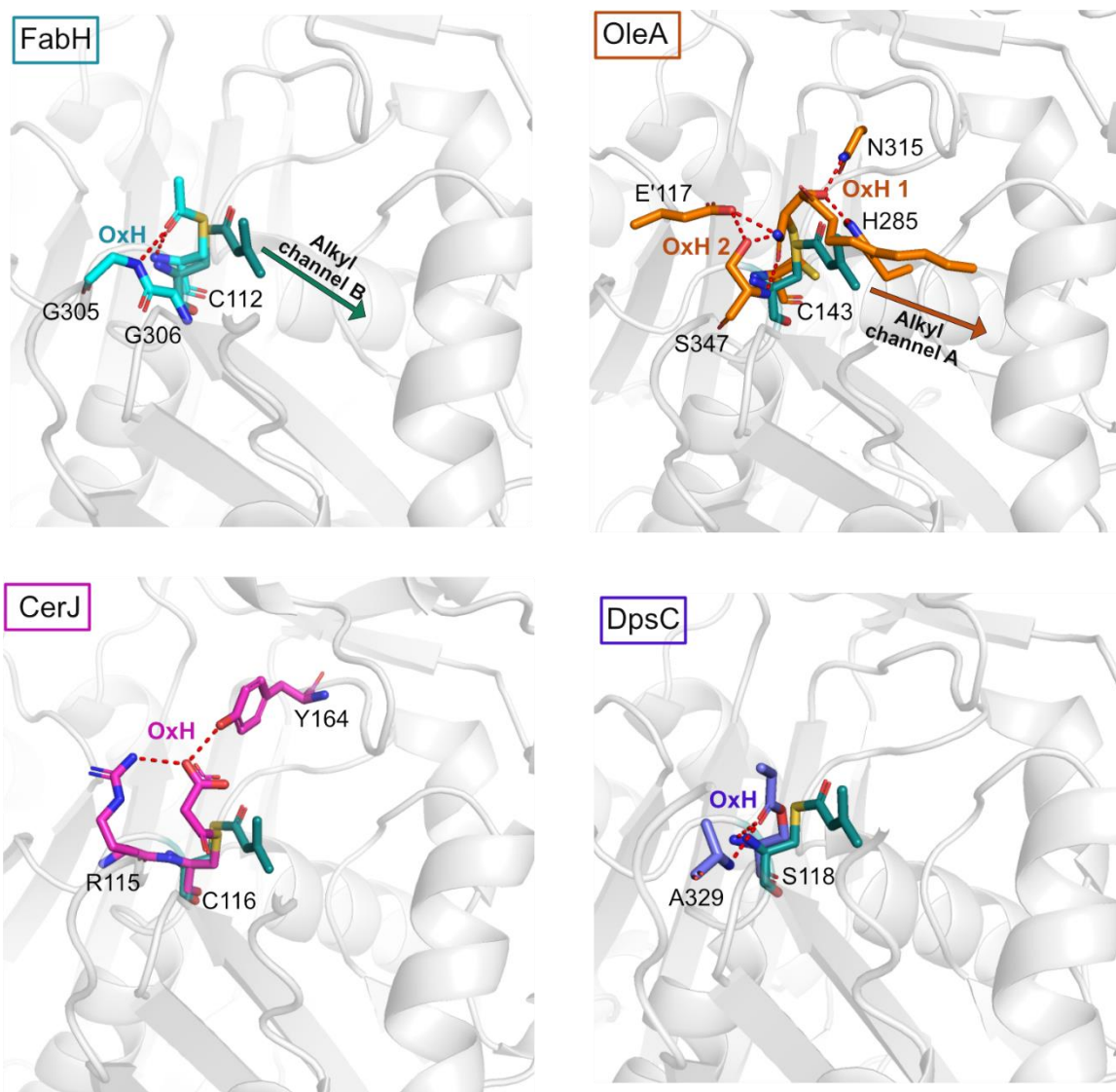
Figure 3.13. Comparison of MBE-TmlB with ligand bound homologs Ac-FabH (cyan), Cer-OleA (orange), M-CerJ (magenta) and Pr-DpsC (purple). The figure representing the ligand coordination orientation and oxyanion hole (OxH) for each enzyme where MBE incorporation into TmlB shows similarity with cerulenin in OleA. The polar interactions between ligand and OxH residues are shown in red dashed line and only the TmlB cartoon is shown in grey.

3.4 Putative mechanism of the self-acylation reaction

It has been reported that the active site cysteine residue is essential for the transacylation half reaction prior to the decarboxylation half reaction in typical FabH and KAS III enzymes.[61,87] Active site comparisons show that TmlB H264 occupies the same position as EcFabH N274, but the imidazole ring is rotated 90° relative to N274 in EcFabH. This brings the NE2 atom of H264 within hydrogen bonding distance of the C117 and may increase its nucleophilicity. The ND1 imidazole nitrogen of H264 can form a hydrogen bond with the

active site residue D269, which may further enhance the nucleophilicity of C117 residue (Figure 3.14). Similar activities were reported in CerJ and DpsC where the histidine residues of the active triad were found to increase the nucleophilicity of the active site cysteine and serine, respectively.[62,64]
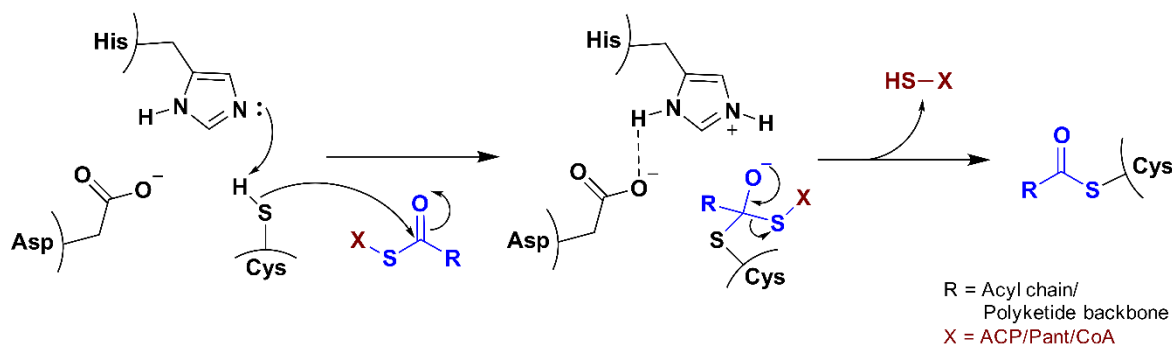
Figure 3.14. The histidine (basic) residue increases the nucleophilicity of a neighbouring cysteine residue (the nucleophile) by forming a hydrogen bond with the thiol proton of cysteine, which is further stabilized by neighbouring aspartate (acid). The activated cysteine then attacks the acyl residue/polyketide backbone (coloured in blue) and self-acylates by forming a tetrahedral intermediate and releasing the substrate delivering group CoA/Pant/ACP.

3.5 Preliminary esterification reaction assay for TmlB

The MBE self-loaded TmlB was used to further investigate the putative reverse esterase function of TmlB. TmlB was first incubated with MBE-pantetheine as described earlier and successful self-loading was confirmed by ESI-MS analysis. 4-hydroxybutyryl-pantetheine (4HB-pant) was then added to the reaction mixture and kept overnight at 4 °C. The assay progress was monitored after ~15 hours by ESI-MS and it was observed that the MBE-TmlB peak intensity decreased with respect to the control, which may suggest the production of esterified product from the simplified monic acid derivative (MBE-pant) and the fatty acid starting unit (4HB-pant) (Figure S.9.15 A). Further analysis was conducted via analytical LC-MS, which showed signals at 278, 360 and 405 Da corresponding to pantetheine, MBE-pant and the formic acid adduct of MBE-pant, but no esterification product (Figure S.9.15 B).

3.6 Summary

Self-acylation assays have shown that TmlB selectively binds branched substrate MBE, whereas MupB showed no preference towards branched or non-branched substrates although the extent of loading was poor. Unexpectedly, the cysteine to alanine mutant a MupB did not eliminate self-acylation of non-branched substrates suggesting the presence of a surrogate nucleophile, S224 near the catalytic triad. Hence, to understand the role of S224 residue in substrate binding, using a C116/S224 double mutant MupB would be an obvious future experiment. Nonetheless, a high-resolution co-crystal structure of MBE-TmlB confirmed that the active site cysteine was the site of covalent modification. A structural comparison with homologs reveals the non-canonical alkyl channel B of TmlB coordinates this short pre-monic acid mimic. Also, the preliminary esterification assay between MBE and a fatty acid starter unit (3-HP and 4-HB) suggested the requirement of appropriate substrates as well as the correct ACP partner to deliver the substrates.

4. Investigation on ACP:MupB/TmlB interaction: NMR and MS studies

This chapter focuses on the investigations of protein-protein interactions (PPIs) between the mupirocin and thiomarinol biosynthetic pathway ACPs and MupB and TmlB, respectively. As proposed, the pre-monic acid polyketide backbone is likely to be transferred from an ACP to the putative reverse-esterase enzyme MupB/TmlB in preparation for esterification with the fatty acid chain. Therefore, a combination of NMR based interaction studies and ESI-MS analysis were used to determine if PPIs could be identified between MupB/TmlB and potential ACP partners.

The Mupirocin biosynthetic pathway has five free standing ACPs, termed MacpA-E. Each ACP has been ascribed a different function and all of them are essential for the biosynthesis of mupirocin.[76,83] MacpC was reported to be involved in the introduction of β-methyl branch at the C-3 position while producing the C_{17} monic acid backbone on the MmpA module.[82] MacpD has been shown to work consecutively with both MupQ (an adenylation domain) and MupS (a malonyl-3-CoA-1 ACP reductase) to produce the 3HP starter unit for C_9 fatty acid biosynthesis.[117] MacpE has been reported to act in the latter stages of the pathway during the conversion of PA-B to PA-A.[118] More recent work has revealed the stepwise assembly of the fatty acid chain (9HN) from the starter unit 3HP by two consecutive DCC (decarboxylative condensation reaction) catalysed by MmpF and involving MacpA and MacpB.[40] This generates an MacpA bound 7-hydroxyhepatanoic acid species that undergoes a final extension step on the multidomain synthase MmpB to yield 9-hydroxynonaoic acid. However the observation of PA derivatives with curtailed fatty acid chains (e.g. Mupirocin W2 and W3)[55,119] does not rule out potential interactions of MupB with MacpA or MacpB to generate these esterified products. Similar arguments apply to the homologous thiomarinol ACPs, TacpA and TacpB.

Several cis ACPs are also potential candidates for involvement in the esterification step. A further possibility for the timing of MupB interaction is immediately after β-branch incorporation when the growing polyketide chain is attached to MmpA_3a/b ACP, the terminal di-domain ACP of the type I mega enzyme:MmpA (Introduction Fig 1.7). Therefore, MmpA_3a/b are potential candidates for transferring the polyketide chain to MupB.

Mutations of the oxidoreductase domain of MmpE generated (e.g., *MmpEΔOR/ΔMupP)* desepoxy-mupirocin P and desepoxy-PA-B (Figure 4.1) confirmed that the OR domain of MmpE is responsible for epoxidation at C_{10}-C_{11} but the precise timing of epoxidation versus esterification is unknown.[120,121]
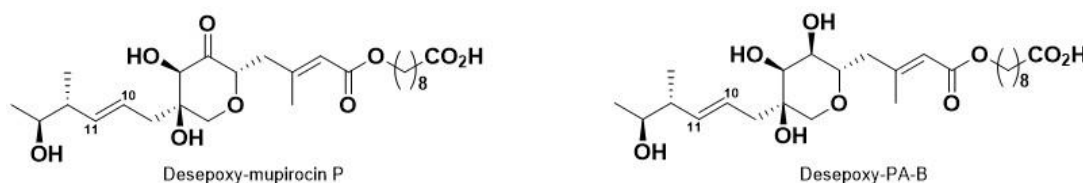
Figure 4.1. Desepoxy-mupirocin P and Desepoxy-PA-B produced from the crude extract of the *MmpEΔOR/ΔMupP* double mutant strain of *P. fluorescens* NCIMB 10586.

Bioinformatic analysis, however, of the unannotated 148 aa between the KS^0 and OR of MmpE predicted the presence of 4 ordered α-helices conserved in the phosphopantetheine (Ppant) binding ACP domain[122,123] (Figure S.9.16). Pfam (Protein families database) identified residues 29-73 as a phosphopantetheine attachment site.[124] The thiomarinol homolog, TmpE also contained an unannotated region between the KS and OR (170 aa) with a similar profile containing a phosphopantetheine attachment site. Both regions within MmpE and TmpE displayed sequence homology to ACPs found within type 1 *E. coli* ACP[125] and *S. coelicolor* AcpP[126], having the conserved Ppant binding motifs, GXXS where the serine residue forms the covalent linkage to the PPant arm (Figure 4.2 A). Homology modelling (*I-tasser*) of both those regions of MmpE and TmpE further predicted the α-helical bundle architecture to be similar to the structures of *E. coli* ACP and *S. coelicolor* AcpP (Figure 4.2 B) and these sequences were tentatively assigned as MmpE_ACP and TmpE_ACP. As there is no off-loading domain on MmpE, one hypothesis is that the ACP tethers the polyketide while C_{10}-C_{11} is epoxidized and then the polyketide is offloaded by MupB, thereby fulfilling its priming role. Indeed, the KS^0 of MmpE might act after β-branching to off-load MmpA_3a/b and a putative docking domain has been detected between MmpA and MmpE (Ash Winter, unpublished) suggesting a β-branching then epoxidation sequence. In conclusion, MmpA_3a/b, MmpE_ACP and their thiomarinol homologs TmpA_3c and TmpE_ACP were all taken into consideration for delivering the pre-monic acid to MupB and TmlB, respectively.
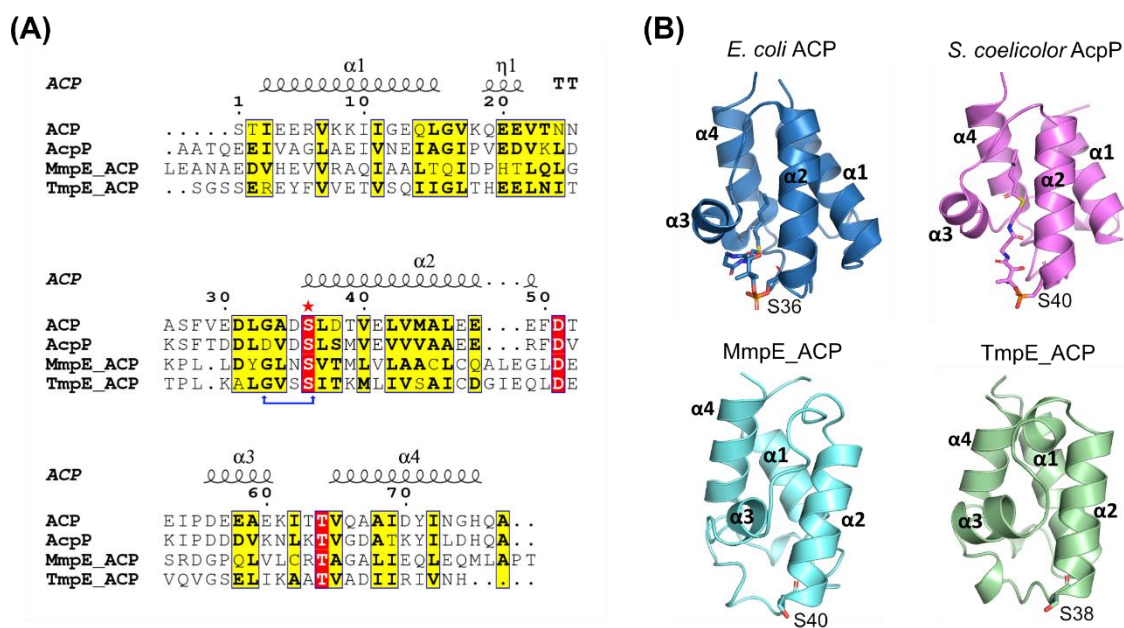
Figure 4.2. Comparison of primary sequences and homology models of MmpE_ACP and TmpE_ACP with *E. coli* ACP (PDB:2FAC) and *S. coelicolor* AcpP (PDB:2KOO). (A) Multiple sequence alignment of ACPs highlights the variability in sequence and conserved motifs. *Yellow* boxes note conserved regions with those amino acids highlighted in ***red boxe***s being completely conserved and those in ***bold font*** being mostly conserved. The ***red star*** symbol notes the highly conserved serine residue bearing the Ppant arm and the ***blue arrow*** marks the GXXS pantetheine binding motif. (B) Structural comparison of acylated *E. coli* ACP and *S. coelicolor* AcpP with the homology models (*I-tasser*) of MmpE_ACP and TmpE_ACP showing the conserved serine residue at the N-terminal of the α2 helix. The acyl-phosphopantetheine is shown in stick representation.

4.1 Expression and purification of mupirocin ACPs

4.1.1 MacpB

The gene corresponding to MacpB was cloned from the genomic DNA of *Pseudomonas fluorescens* NCIMB 10586. The genomic DNA was prepared by growing the wild-type *P. fluorescens* on an agar plate for 16 hours after which small sample cells were resuspended in sterile dH$_2$O and centrifuged prior to collection. Polymerase chain reaction (PCR) was used to amplify the *MacpB* gene using pre-designed primers (Experimental 7.1). Successfully amplified PCR products were confirmed by DNA gel-electrophoresis methods (1.5% Agarose gel). Figure 4.3 A shows bands present at approximately 250 base pairs (bp), representing the successful PCR of MacpB DNA. The PCR products were then purified following an Invitrogen quick PCR purification kit protocol. The purified MacpB DNA was ligated into the pre-linearized pOPINF vector (Carbenicillin resistant) and transformed into NEB-5α bacterial cells where successful ligation was confirmed by the X-gal screening

(Experimental 7.1). The existence of white colonies suggested ligation had occurred as expected. Four isolated white colonies (1-4) were cultured in sterile 5 mL LB media for each, and plasmids were purified using a standard miniprep plasmid purification procedure. Further confirmation of MacpB ligation into the pOPINF plasmid was provided by GENEWIZ sequencing.

The pOPINF ligated MacpB plasmid was transformed into the BL21 *E. coli* cells and protein expression was carried out on a 2 L scale via the general protein expression protocol (Experimental 7.2). The cell pellets were harvested, lysed, and isolated by centrifugation. All the pre and post induction samples, and the soluble and insoluble samples were analysed via gel electrophoresis. Although MacpB was successfully expressed as confirmed by the presence of a band in the post induction (Post-I) sample, it was only present in the insoluble pellet (Figure 4.3 B).
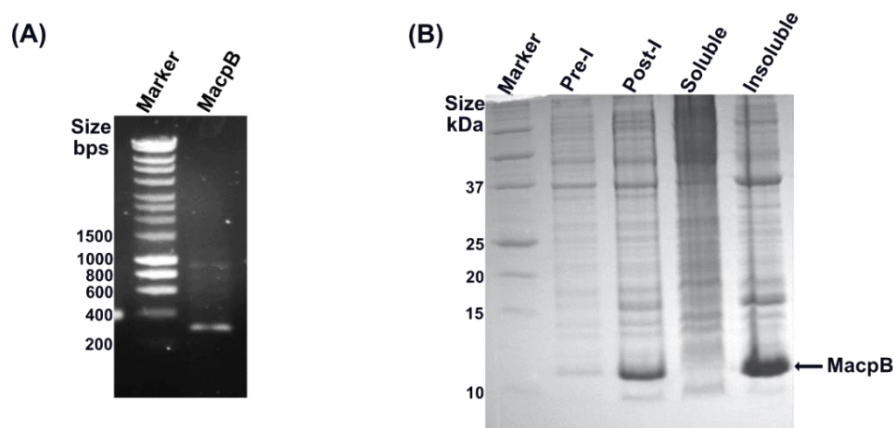
Figure 4.3. (A) An 1.5% agarose gel image showing the amplified MacpB DNA showing a band at 250 bps. (B) An SDS-PAGE gel image showing that MacpB protein expressed successfully in the insoluble cell pellet as the Post-I and Insoluble samples are showing a clear band ~ 11kDa.

Instead, the MacpB plasmid was transformed into the Arctic Express RIL cells and expressed in 2 L scale (assisted by Dr Ash Winter). Arctic cells are bioengineered to overcome the bacterial gene expression hurdle of protein misfolding and insolubility. These cells have chaperonins Cpn10 and Cpn60 from *O. antarctica*, showing high similarity with *E. coli* chaperonin which confers improved protein expression at low temperature and potentially increases the yield of soluble and folded protein.[127] Expression of MacpB using these cells (Experimental 7.2 and Figure 4.2 A) yielded a band by SDS-PAGE suggesting weak but soluble expression. Further isolation by IMAC prior to exchange into Tris buffer (25 mM Tris-HCl-, 150 mM NaCl, pH 7.5, 1 mM DTT) confirmed a band at the correct molecular mass. Analytical SEC analysis suggested MacpB was monomeric (Figure 4.2.B)

and ESI-MS gave an observed mass of 11371 Da (expected mass 11373 Da, Figure 4.4 C). The SEC fractions of MacpB protein were mixed and spin concentrated using a 5 kDa MWCO spin concentrator to a final concentration of 100 µM and flash frozen in liquid N_2 prior being stored at -20 °C for further use.
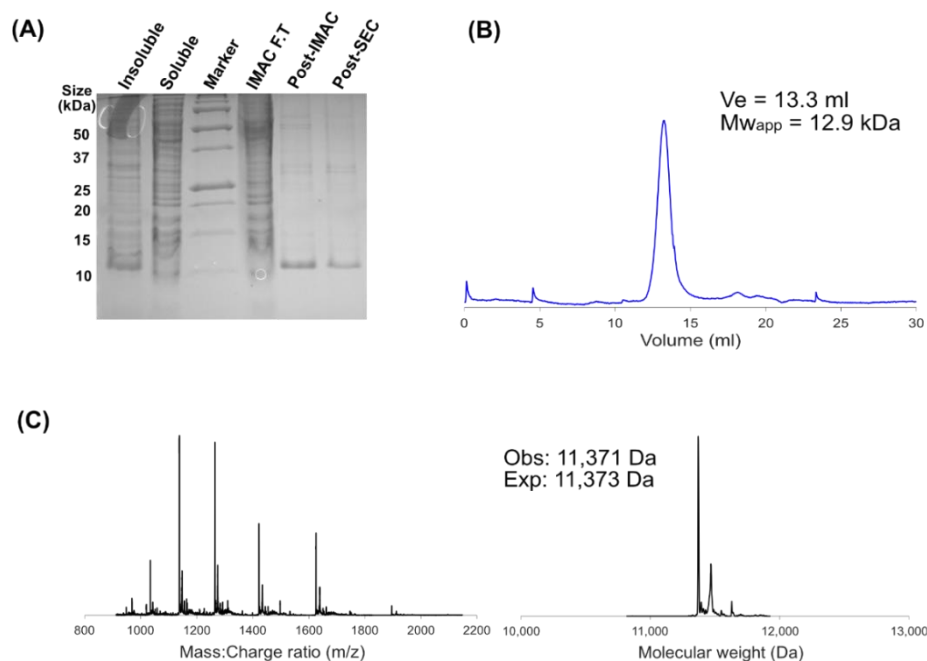
Figure 4.4. MacpB purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression and purification via IMAC and SEC. (B) The analytical SEC of MacpB eluting as a single oligomeric species at elution volume (*Ve*) 13.3 mL which indicated an approximate mass of 12.9 kDa and (C) the ESI-MS spectra of multiple charge states corresponding to the observed mass 11371 Da (expected 11373 Da).

4.1.2 MmpE_ACP

MmpE_ACP DNA was synthesized by Thermo Fisher and subcloned into the pET151-D/TOPO plasmid bearing an N-terminal His$_6$-tag. The plasmid was transformed into E. *coli* T7 Express cells and expressed at a 2 L scale. Successful soluble protein expression was confirmed by SDS-PAGE gel (Figure 4.5 A). The MmpE_ACP was isolated from the soluble supernatant using IMAC and then exchanged into Tris buffer prior to SEC. The protein was spin concentrated to 100 µM as above prior to flash freezing and stored at -20 °C for further use. Further characterizations by analytical S75 (Figure 4.5 C) showed that MmpE_ACP eluted as a single species (marked with a 'star') with the calculated molecular mass of 14 kDa. ESI-MS analysis produced the mass spectrum of MmpE_ACP (Figure 4.5 D) where the multiple charge states deconvoluted to an observed mass of 13004.0 Da (expected mass 13004.8 Da).

Figure 4.5. MmpE_ACP purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression and purification in every stepa. (B) IMAC isolation of MmpE_ACP from the supernatant showing the elution with increased imidazole concentration. (C) Analytical SEC of MmpE_ACP eluting as a single oligomeric species (marked with a 'star') on an 10/300 Superdex 75 analytical column and (D) the ESI-MS spectra of multiple charge states corresponding to the observed mass 13004 Da (expected 13004.8 Da) shown in the inset.

MacpA and MmpA_3a (the first ACP of the di-domain) were expressed and characterized successfully by Dr Matt Rowe and Dr Ash winter and were also used in this work.[128]

4.2 Expression and purification of thiomarinol ACPs

4.2.1 TacpA

The gene encoding TacpA had previously been cloned into the pOPINF vector (Dr Paul Walker) and this plasmid was transformed into host *E. coli* (BL21) cells.[117] TacpA was expressed on a 2 L scale and successful expression confirmed by SDS-PAGE gel (Figure 4.6 A). The cell pellet was lysed by sonication and extracted by centrifugation. The supernatant containing soluble protein, was purified first by using IMAC (Figure 4.6 B). TacpA eluted from 150-200 mM imidazole and the IMAC fractions of correct mass were mixed and further purified and exchanged with deionized water using size exclusion chromatography (SEC) to remove the residual salt and imidazole (Figure 4.6 C). The SEC fractions containing TacpA were mixed, freeze dried and isolated as white solid protein (20 mg). The dried protein was stored at -20 °C. ESI-MS confirmed the successful isolation and purification of TacpA (Figure 4.6 D), with an observed mass of 12378 Da (expected mass 12378 Da).
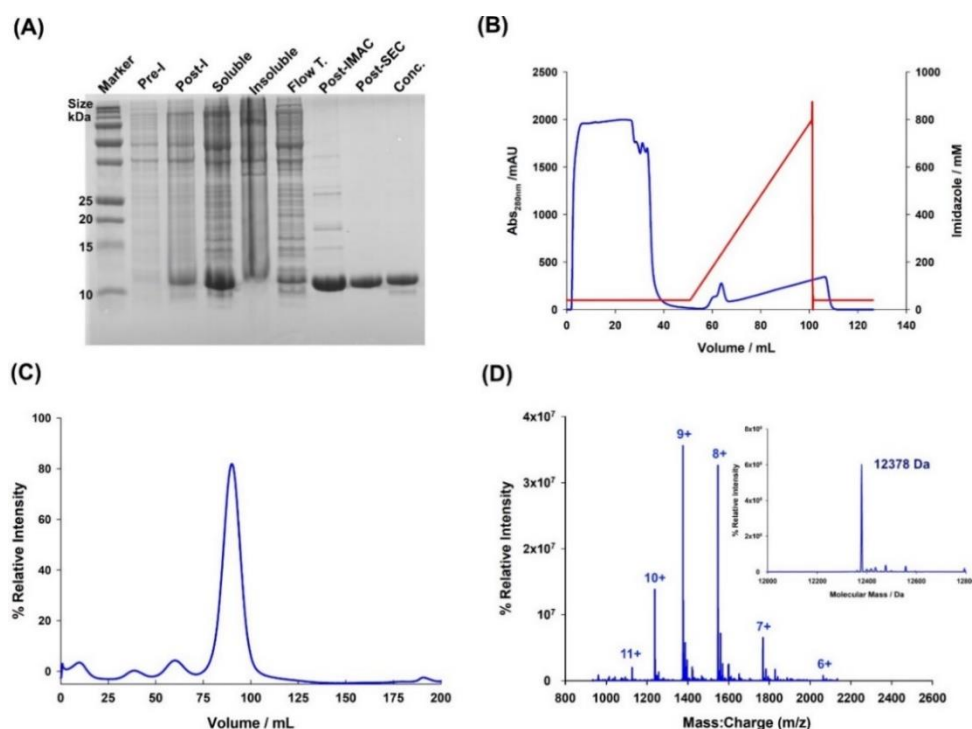
Figure 4.6. TacpA purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression and purification in every step. (B) IMAC isolation of TacpA from the supernatant showing the elution of TacpA with increased imidazole concentration. (C) SEC of TacpA eluting as a single oligomeric species on a Superdex 75 analytical column and (D) the ESI-MS spectra of multiple charge states corresponding to the observed mass 12378 Da (expected mass 12378 Da) shown in the inset.

4.2.2 TacpB

The TacpB gene was also synthesized and cloned into the pET151 plasmid bearing the N-terminal His$_6$ tag by Thermo Fischer. Successful expression TacpB protein was conducted on a 2 L scale by transformation into T7 expression cells. Protein isolation and characterization were the same as for TacpA stated above. Figure 4.7 A shows an SDS-PAGE analysis that revealed a strong band near 15 kDa in every step of the purification process, suggesting that TacpB protein successfully expressed in the soluble supernatant as monomer. The ESI-MS analysis of the purified protein confirmed the molecular weight of TacpB (observed: 16460 Da, expected: 16458 Da, Figure 4.7 D). The protein was then concentrated to 100 μM and flash frozen using liquid N$_2$ prior to storage at -20 °C.
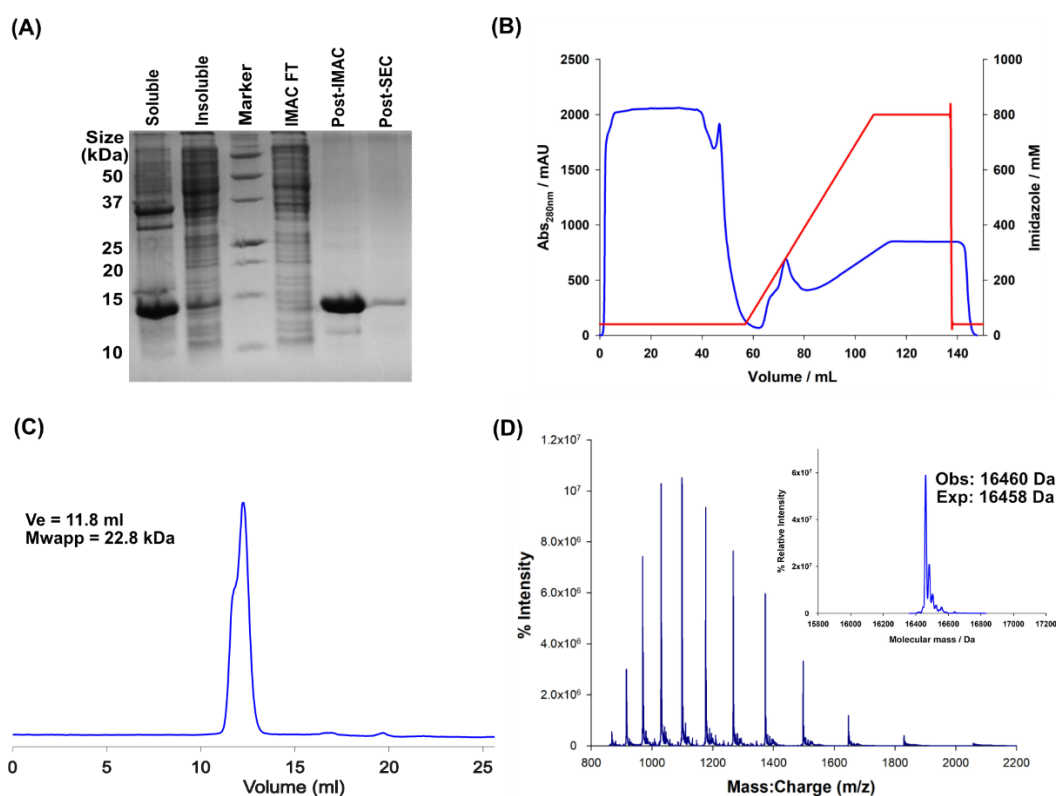
Figure 4.7. TacpB purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression as soluble protein. (B) IMAC isolation from the supernatant showing the elution of TacpB with increased imidazole concentration. (C) Analytical SEC of post-IMAC fractions of TacpB eluting as a single oligomeric species on a 10/300 Superdex 75 analytical column and (D) the ESI-MS spectra of multiple charge states corresponding to the observed mass of 16460 Da (expected mass 16458 Da) shown in the inset.

4.2.3 TmpA_3c

The thiomarinol equivalent of the mega enzyme MmpA is TmpA, which has an extra ACP (named as TmpA_3c) in the last module[74] forming a tridomain. Like TacpB, TmpA_3c was synthetically cloned into the pET151 vector by Thermo Fischer, transformed into T7 express cells and expressed on a 2 L scale. All the protein expression and purification procedures were followed as described previously. The protein was spin concentrated using a10 kDa MWCO spin concentrator to 100 μM and stored at -20 °C after aliquoted to 100 μL and flash frozen using liquid N_2. Characterization was done by gel electrophoresis (Figure 4.8 A) and ESI-MS which confirmed the successful isolation and purification of TmpA_3c (Figure 4.8 D, observed mass 16971 Da; expected mass 16970 Da).
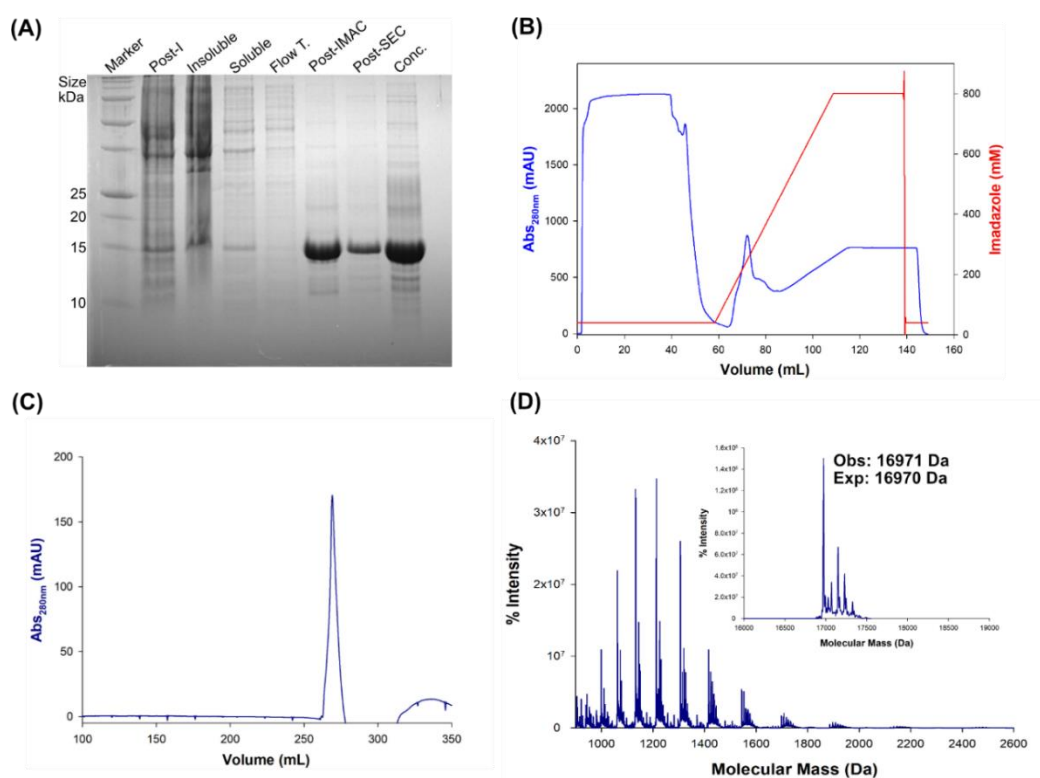
Figure 4.8. TmpA_3c purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression as soluble protein. (B) IMAC isolation from the supernatant showing the elution of TmpA_3c with increased imidazole concentration. (C) SEC of post-IMAC fractions of TmpA_3c eluting as a single oligomeric species on a 26/60 Sephacryl S100 analytical column and (D) the ESI-MS spectra of multiple charge states corresponding to the observed mass of 16971 Da (expected mass 16970 Da) shown in the inset.

4.2.4 TmpE_ACP

The TmpE_ACP DNA was also synthesized by Thermo Fischer, subcloned into pET151 vector having an N-terminal His$_6$ tag. Like the previously expressed ACPs, TmpE_ACP was also transformed into T7 express cells and expressed on a 2 L scale following the general protein expression protocol described in experimental 7.2. Successful expression of soluble TmpE_ACP was confirmed by SDS-PAGE gel (Figure 4.9 A) which showed a clear broad band at ~ 12 kDa. Further isolation and buffer exchange were done using IMAC and SEC chromatography techniques. The ESI-MS analysis of the purified protein confirmed the molecular weight of TmpA_3c (observed: 12278 Da, expected: 12278 Da) from the deconvolution of the multiple charged states shown in Figure 4.9 (D). The protein was then concentrated to 100 μM and flash frozen using liquid N$_2$ prior to storage at -20 °C.
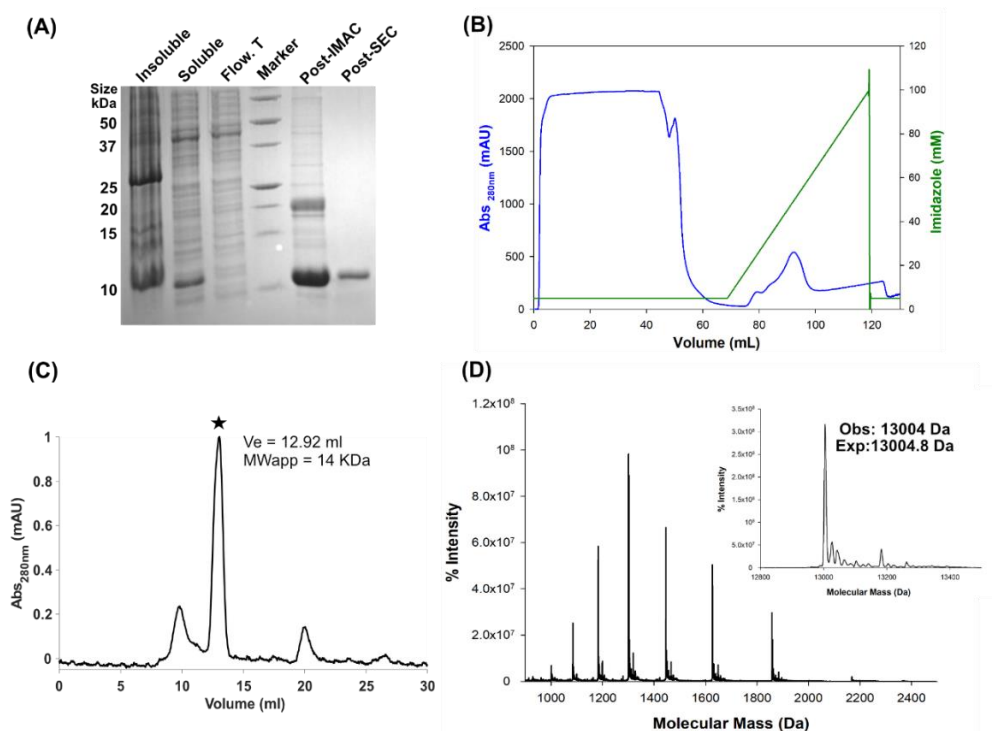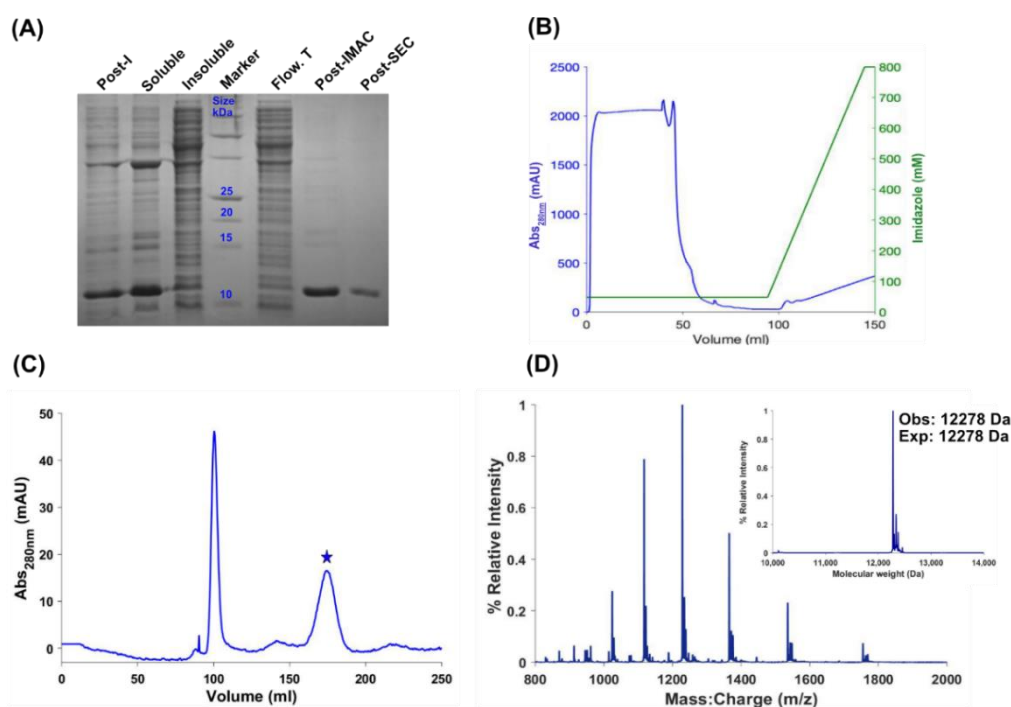

Figure 4.9. TmpE_ACP purification and characterisation. (A) SDS-PAGE gel image confirmed the successful expression as soluble protein. (B) IMAC isolation from the supernatant showing the elution of TmpE_ACP with increased imidazole concentration. (C) SEC of post-IMAC fractions of TmpE_ACP eluting as a single oligomeric species (marked with a star) on a Superdex 75 analytical column and (D) the ESI-MS spectra of multiple charge states corresponding to the observed mass of 12278 Da (expected mass 12278 Da) shown in the inset.

4.3 Interaction of MupB/TmlB with ACPs: An NMR study

4.3.1 Expression and purification ^{15}N labelled mupirocin and thiomarinol ACPs

MacpA, TacpA, TacpB, MmpA_3a, TmpA_3c, and MmpE_ACP genes were transformed into T7 express cells and expressed in M9 media supplemented with NH$_4$Cl (^{15}N, 99%), glycerol and glucose. Two different methods of single - labelled (^{15}N) protein expressions were used during this work to maximize the yield of labelled ACP. MacpA, MmpA_3a, TmpA_3c and MmpE_ACP were expressed by media exchange into M9. In this method, the seed broth was first grown in 2 L of LB media until the optical density reached 1.0. The cells were then harvested, washed thoroughly with M9 media prior to resuspension in M9 media (1L) supplemented with other nutrients. The resuspended cells were induced with 250 µM IPTG at 16 °C and left to culture overnight (200 rpm) prior harvesting. TacpA and TacpB were expressed directly in M9 media supplemented with NH$_4$Cl (^{15}N, 99%) and necessary nutrients for bacteria culture following the general M9 expression procedures in experimental 7.2. Both MacpB and TmpE_ACP failed to express soluble protein using either of these methods.[129] All the proteins were harvested by centrifugation at 6000 rpm for 10 mins and were then resuspended in column buffer A prior to storing at -20 °C.

Protein isolation and purification were performed in the same manner for all the Mup and Tml ACPs (experimental 7.3). First the cells were lysed by sonication and centrifuged at 15 K rpm for 15 mins to separate the soluble supernatant from the insoluble pellets. The soluble fraction was then purified by IMAC via a Hi-Trap 5 mL Ni^{2+} column followed by desalting in water using a SEC column before being freeze-dried for storage. The Mup_ACPs were further incubated overnight with TEV protease enzyme (0.33 µM per mg of ACP) to remove the poly-histidine tag followed by negative IMAC and another SEC purification prior to being freeze- dried. Purified proteins were characterised by SDS-PAGE to confirm the successful expression and purity of the proteins (Figure 4.10 A and B) where each lane is showing a band of the expected molecular weight of the respective ACP.
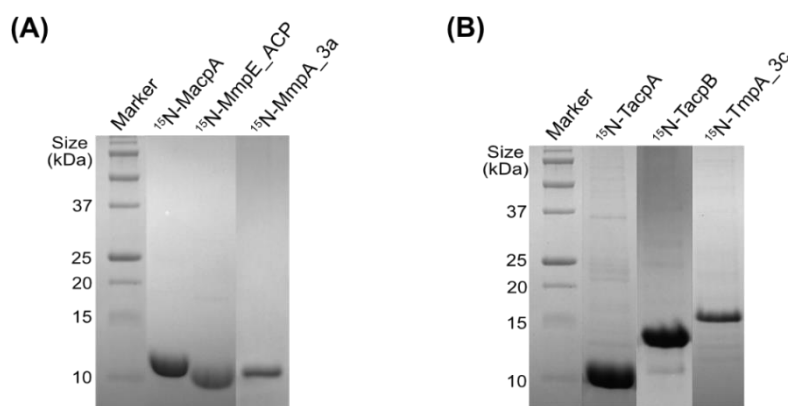
Figure 4.10. (A) SDS-PAGE gel image for purified ^{15}N labelled MacpA (expected mass: 12023 Da), MmpE_ACP (expected mass: 10097 Da) and MmpA_3a (expected mass: 11013 Da). (B) SDS-PAGE gel image of purified ^{15}N labelled TacpA (expected mass: 12517 Da), TacpB (expected mass: 16659 Da) and TmpA_3c (expected mass: 17179 Da). The images are constructed and assembled from corresponding ACP lanes that were not adjacent.

4.3.2 NMR titration of ^{15}N-ACPs with MupB and TmlB

Protein NMR is a powerful tool in structural biology where it has been applied to study and determine protein structure, dynamics and protein-protein / protein-ligand interactions.[130–133] The ^{15}N labelled ACPs have previously been used to investigate interactions between ACPs and prospective partner enzymes[117] and enable segments of antibiotic biosynthetic pathways to be mapped out prior to enzyme assay work.[40] Therefore, ^{15}N labelled ACPs from mupirocin and thiomarinol biosynthetic pathways were titrated against MupB and TmlB protein and possible PPIs followed by NMR using chemical shift perturbations. This investigation may reveal the ACP interactions with MupB/TmlB and provide clues as to which ACPs are responsible for the transfer of the substrates to the proteins.

The NMR titrations were conducted first by preparing the ACPs in 50 mM phosphate buffer (pH 8.0) from the freeze-dried proteins concentrated to 1-2 mM with 10% D_2O. The ^{15}N labelled ACPs were then incubated with MupB/TmlB at different concentration ratios and the NMR titration was monitored by heteronuclear single quantum coherence spectra (HSQC) in collaboration with Dr Ash Winter and Dr Chris Williams. The details of the NMR titration protocol are described in the experimental 7.9.

Looking first at examples of the discrete Mup and Tml ACPs, Figure 4.11 compares the ^{1}H-^{15}N HSQC control spectra of the ^{15}N labelled ACPs (coloured in blue) with the ^{1}H-^{15}N HSQC spectra after addition of two or three equivalents of MupB/TmlB (coloured in red).

Due to the insoluble nature of MacpB, the equivalent ^{15}N TacpB/TmlB pairing was used. MacpA, TacpA and TacpB all failed to show chemical shift perturbations (CSP) or exchange broadening, thereby potentially signifying no protein-protein interactions.
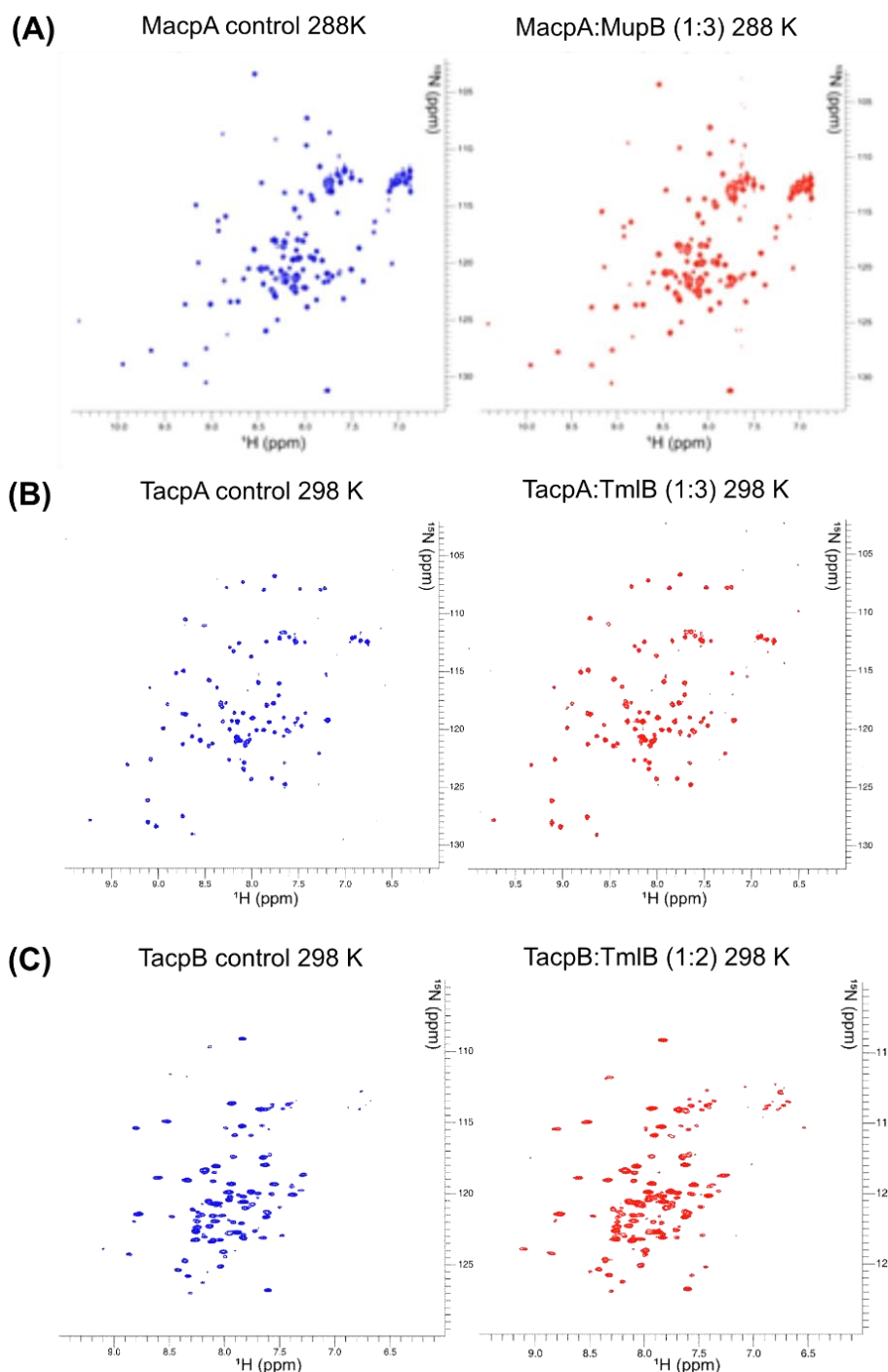
Figure 4.11. ^{1}H-^{15}N HSQC experiments to determine PPIs of free-standing Mup and Tml ^{15}N-ACPs with MupB and TmlB, respectively. (A) Control spectrum of MacpA (blue) at 288 K / pH 8.0 and in the presence of three equivalents of MupB (red). No discernible chemical shift perturbations (CSPs) or broadening of residues observed. (B) Control spectrum of TacpA (blue) at 298 K / pH 8.0 and in the presence of three equivalents of TmlB (red) and no discernible CSPs or broadening of residues observed. (C) Control spectrum of TacpB at 298 K / pH 8.0 (blue) and in the presence of two equivalents of TmlB (red) and no change in spectrum was observed.

Next, ^{15}N labelled MmpA_3a (the β-branching ACP) was titrated with MupB following the same protocols described previously, but no interactions were evident from the NMR analysis (Figure 4.12), suggesting that this ACP (or its didomain counterpart) do not act at
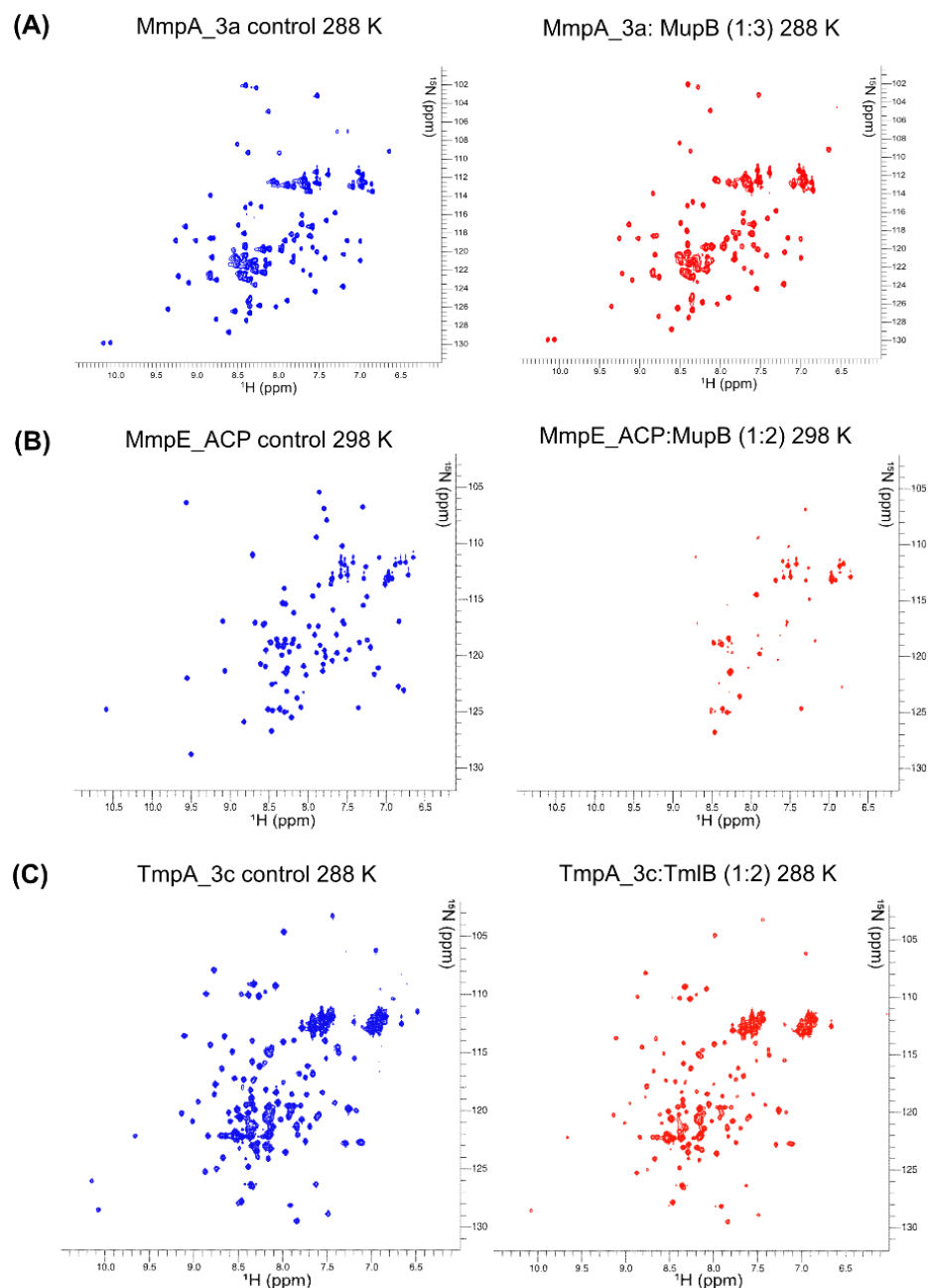
Figure 4.12. ^{1}H-^{15}N HSQC experiments to determine PPIs of modular Mup and Tml ^{15}N-ACPs with MupB and TmlB, respectively. (A) Control spectrum of MmpA_3a (blue) at 288 K / pH 8.0 and in the presence of three equivalents of MupB (red). No discernible chemical shift perturbations (CSPs) or broadening of residues observed. (B) Control spectrum of MmpE_ACP (blue) at 298 K / pH 8.0 and in the presence of two equivalents of MupB (red) and showing significant numbers of peak disappearances and (C) Control spectrum of TmpA_3c at 288 K / pH 8.0 (blue) and in the presence of two equivalents of TmlB (red) and no change in spectrum was observed.

this point in the mupirocin biosynthesis pathway. Turning to the thiomarinol equivalent, ^{15}N TmpA_3c was titrated with TmlB but again no peak broadening or CSPs were observed (Figure 4.12 A and C).

Finally, ^{15}N-labelled apo MmpE-ACP was titrated with recombinantly expressed MupB to investigate the presence of any PPIs. In this case a considerable proportion of ^{1}H-^{15}N MmpE_ACP resonances were broadened beyond detection strongly indicating that the modular MmpE_ACP interacts with MupB (Figure 4.12. B). TmpE_ACP:TmlB NMR titration was not conducted because of the insoluble nature of expression of ^{15}N labelled TmpE_ACP which would have provided valuable further validation. Nonetheless, as they share the sequence homology (30%) and model structure homology (Figure 4.2 B), it is assumed that if this interaction is real, then TmlB is likely to also interact with TmpE_ACP. The Mup/TmlB: ^{15}N-ACP NMR titration results are summarized in Table 4.1.

Table 4.1. Summary of ^{1}H-^{15}N HSQC NMR titration results between mupirocin and thiomarinol ACPs with MupB and TmlB, respectively.

Mupirocin ACPs	NMR-titration with MupB	Thiomarinol ACPs	NMR-titration with TmlB
^{15}N MacpA	No interaction	^{15}N TacpA	No interaction
^{15}N MacpB	Insoluble	^{15}N TacpB	No interaction
^{15}N MmpA_3a	No interaction	^{15}N TmpA_3c	No interaction
^{15}N MmpE_ACP	Interaction observed	^{15}N TmpE_ACP	Insoluble

4.4 3-methylbut-2-enoyl pantetheine (MBE-pant) upgrade assays

The protein-protein interactions between MupB/TmlB and ACPs from their respective pathways were next investigated by monitoring the change of mass upon transfer of substrate mimics using ESI-MS techniques. To do this required upgrading target ACPs with the MBE mimic to test transfer to MupB/TmlB.

The one-pot pantetheine upgrade and loading onto the ACPs were previously investigated by Dr Paul Walker for kalimantacin ACPs and applied here.[134] The functionalised-pantetheine derivatives can be synthesized from the commercially available D-pantothenic acid by using a chemoenzymatic method[135,136] and in this work, all the pantetheine derivatives (starting materials for the ACP upgrade) were supplied by Dr Angus Weir

(University of Bristol). The general reaction scheme for the pantetheine derivative upgrade assay and ACP loading is presented in Figure 4.13. Pantetheine or its derivatives **(a)** can be converted to their functionalised CoA derivatives **(b)** using enzymes CoaA/CoaD/CoaE (CoA mix) and required cofactors (ATP and Mg^{2+}). Then in **(c),** the apo-ACPs (inactive form of ACP) are loaded with the functionalised CoA derivatives using an appropriate phosphopantetheine transferase (PPTase), in this case MupN/TmlN.[137]

Following this method, the required pantetheine substrate, apo-ACP and CoA-enzymes were incubated with MupN in presence of $MgCl_2$ and ATP at room temperature for approximately1-6 hours as different ACPs required different time to upgrade and the MBE-ACP upgradations wereanalysed by ESI-MS. The general procedure and the reagents compositions are supplied in the experimental 7.7.
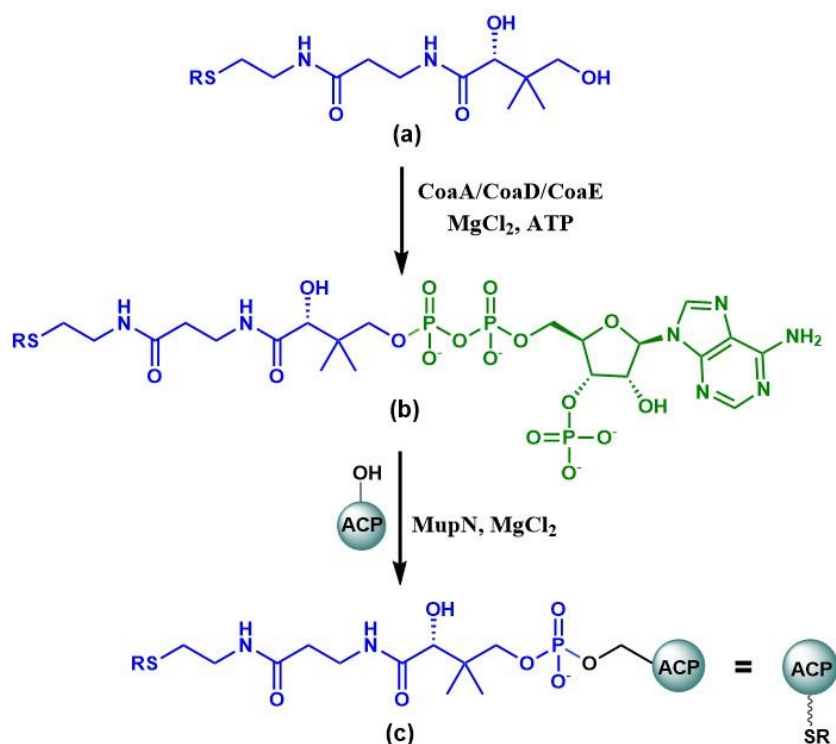
Figure 4.13: Reaction scheme to prepare the functionalized ACP (c) in a one pot, following in-situ chemoenzymatic pantetheine upgrade and ACP-loading reaction. The pantetheine derivatives **(a)** are upgraded to derivatized CoA **(b)** by CoaA/CoaD/CoaE and ATP in presence of Mg^{2+} and then the PPTase transfers the phospho-pantetheine arm [blue in (c)] to apo-ACP, thus producing the final product **(c)**. In this work, R = MBE substrate.

4.4.1 *In-vitro* enzymatic upgradation of MBE to mupirocin ACPs

The MBE upgrade assay of MacpA, MacpB and MmpA_3a were previously done by Dr Matt Rowe and used in this work. Hence, only the MmpE_ACP was upgraded for the first-time using MBE-pant in this work. 100 μM of apo-MmpE_ACP, MBE-pant and CoA upgrade enzymes were incubated with MupN in presence of $MgCl_2$ and ATP in 100 μL scale at room temperature for 3 hours prior to analysis by ESI-MS. The complete conversion of apo-MmpE_ACP (observed mass 13004 Da) to MBE-MmpE_ACP (observed mass 13427 Da) was observed as confirmed by ESI-MS and Ppant ejection assay (Figure 4.14). The general procedure and the reagent compositions are supplied in the experimental 7.7.
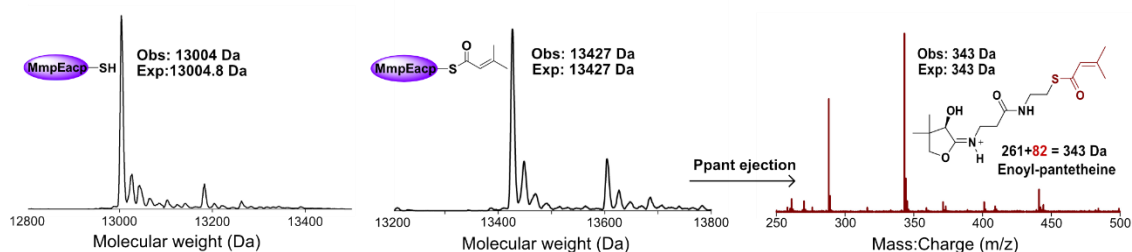
Figure 4.14. MBE-upgrade assay and loading to MmpE_ACP. The deconvoluted spectra of apo-MmpE_ACP (left side), MBE-MmpE_ACP (middle) and corresponding Ppant ejection (right side). The molecular mass changes of apo-MmpE_ACPs to MBE-MmpE_ACP was mentioned in their corresponding spectra.

The 'Ppant ejection assay' is a 'top-down' approach shown to facilitate the study of phosphopantetheine-loaded substrates by mass spectrometry.[138,139] It uses thermal activation by infrared multiphoton dissociation (IRMPD) or collision-induced dissociation (CID) to break the holo-ACP or peptidyl-ACP via selective fragmentations using collision energy leading to the ejection of a singly charged ion and a phosphorylated apo-ACP (Figure 4.15). This assay allows the molecular mass of the substrate loaded on the carrier protein to be determined accurately from the much lower mass of the ejected Ppant fragments.[140] Specifically, non-derivatised holo-ACP gives an ejected fragment of chemical formula $C_{11}H_{21}N_2O_3S^+$ showing an MS^2 peak at m/z 261.1267 (Figure 4.15), while the acylated ACP gives the Ppant peak in the MS^2 spectrum shifted by an amount equal to the mass of the acyl group.[140] This technique had been recently used to assess the functionalised mupirocin and thiomarinol ACPs as well as the enzymatic reaction mechanisms leading to fatty acid starter unit (3-HP) generation[117] and stepwise fatty acid chain (9-HN) assembly.[40]
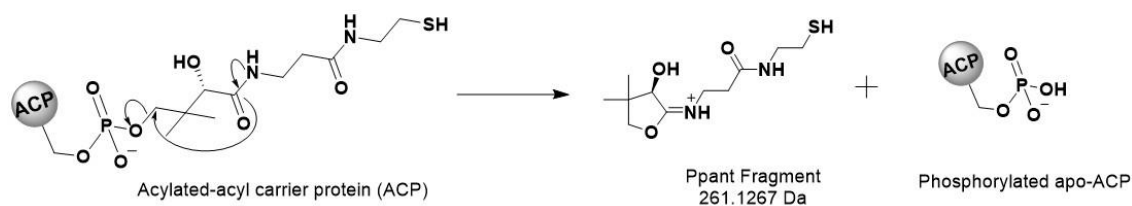
Figure 4.15. Ppant ejection assay occurs by collisional activated dissociation of an ACP, producing a singly charged Ppant ejected ion (a) and phosphorylated apo-ACP.[140]

The stability of the MBE-MmpE_ACP was monitored by MS analysis and after 24 hours at room temperature, around 30% of the enoyl group had been hydrolysed to holo-MmpE_ACP. Additionally, protein degradation was also observed at room temperature over this time frame. Storage of -20 °C slowed down this hydrolysis process, however when stored at -80 °C the MBE-MmpE_ACP was found to be stable for up to 3 months. Therefore, the acylated MmpE_ACP was exchanged into Tris buffer using a Zeba desalt column to remove excess CoAs and other cofactors, aliquoted to 50 µL and flash frozen prior to storage at -80 °C.

4.4.2 *In-vitro* enzymatic upgradation of MBE to thiomarinol ACPs

The thiomarinol free standing TacpA, TacpB and the modular TmpA_3c and TmpE_ACP were successfully upgraded with MBE-pant substrate following the one pot chemoenzymatic pantetheine upgrade mechanism as described above. Although the reaction conditions were kept same for all the thiomarinol ACP upgrade assays, they appeared to load the MBE substrate at different rates. The progress of the reactions was monitored by MS and reactions left for longer or excess reagents were added to help reactions to go to completion. For example, the MBE-TacpA conversion was slow and after 3 hours, a 50/50 mixture of apo/holo-TacpA and a fraction (~10%) of MBE-TacpA were observed. After addition of an extra equivalent of MupN and ATP and leaving for another 24 hours, ~70% of MBE-TacpA (observed mass 12800 Da) was observed along with unreacted apo-TacpA (observed mass 12378 Da, Figure 4.16 A). In the cases of TacpB, TmpA_3c and TmpE_ACP the MBE upgrading was faster than the TacpA reaction and completed in 3-4 hours. The MBE upgrade and loading to the ACPs were confirmed by MS analysis along with corresponding Ppant ejection assays for each MBE-ACP where every time the fragment ejected the molecular mass 343 Da (calculated mass 343.1267 Da) indicating the

MBE (calculate mass 82 Da) loaded pantetheine fragment (calculated mass 261.1267 Da) (Figure 4.16).
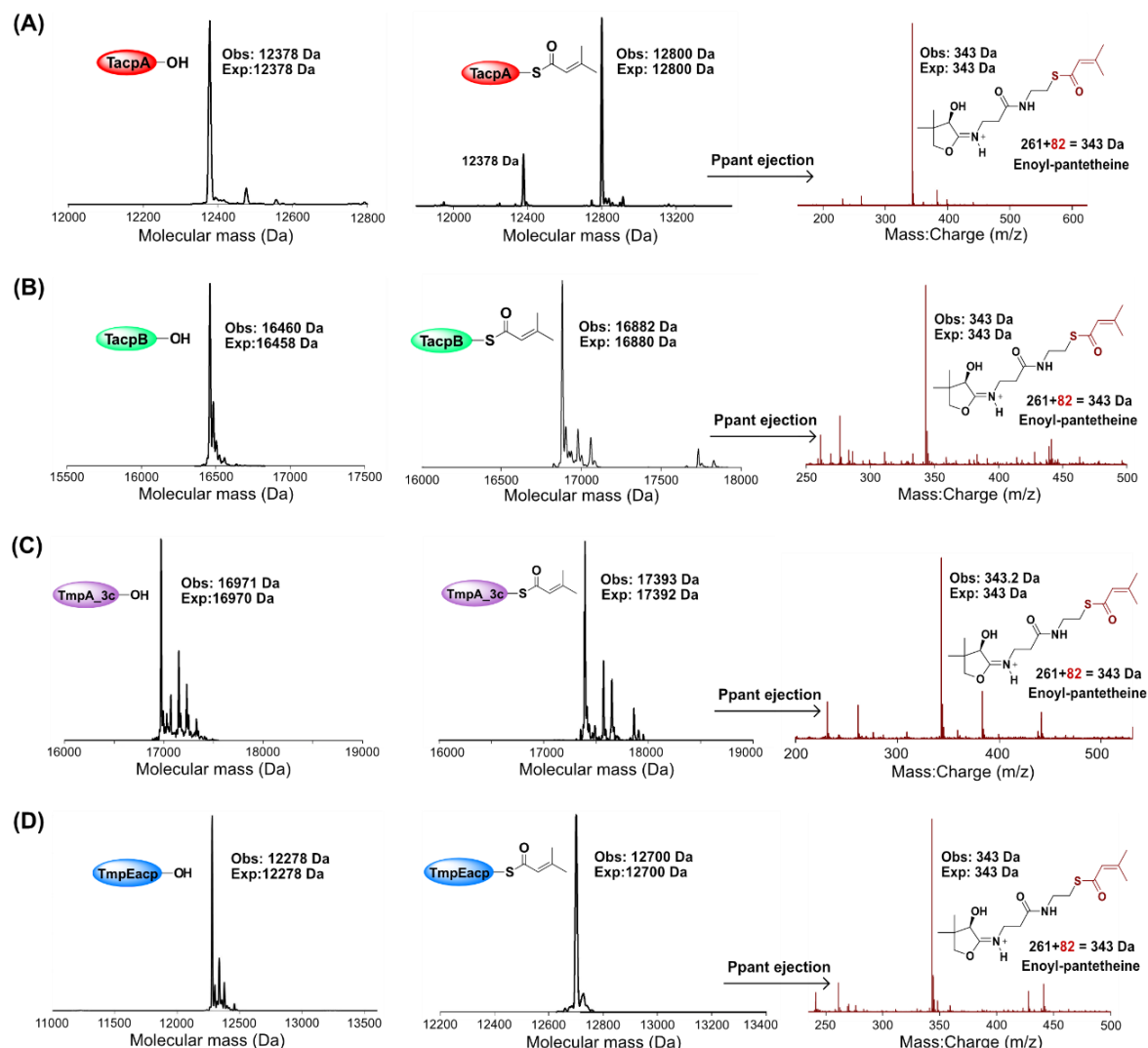
Figure 4.16. MBE upgrade assay and loading to Tml_ACPs. (A) The deconvoluted spectra of apo-TacpA (left side), MBE-TacpA (middle) and corresponding Ppant ejection (right side). (B) The deconvoluted spectra of apo-TacpB (left side), MBE-TacpB (middle) and corresponding Ppant ejection (right side). (C) The deconvoluted spectra of apo-TmpA_3c (left side), MBE-TmpA_3c (middle) and corresponding Ppant ejection (right side). (D) The deconvoluted spectra of apo-TmpE_ACP (left side), MBE-TmpE_ACP (middle) and corresponding Ppant ejection (right side). The molecular mass changes of apo-ACPs to MBE-ACPs are indicated in the corresponding spectra of Tml_ACPs.

The stability of the MBE upgraded thiomarinol ACPs was monitored by MS and after keeping at 4 °C for 24 hours, all MBE-Tml_ACPs were found to be reasonably stable (>10 % loss of derivatised ACP) except for MBE-TacpA, which precipitated. Therefore, the MBE-ACPs were aliquoted to 100 µL, after exchanging into the Tris buffer using a Zeba-desalt column and used immediately in assays.

4.5 Acyltransferase activity of mupirocin and thiomarinol ACPs

The MBE substrate transferring activity of Mup_ACPs (MacpA, MacpB, MmpA_3a, MmpE_ACP) and Tml_ACPs (TacpA, TacpB, TmpA_3c, TmpE_ACP) were investigated. Initially, the control assay was carried out by incubating MupB with an excess of MBE-pant and no ACP. Analysis by ESI-MS clearly showed the appearance of a new species alongside MupB with a mass corresponding to the derivatised enzyme (observed 35082 Da; calculated 35082 Da). Only a fraction (~ 20%) was converted and incubation under different conditions failed to improve this (Figure 4.17 A).

50-100 μM MBE-Mup_ACP samples were then incubated with 20-30 μM of MupB in Tris buffer (pH 7.5) at room temperature for ~ 3 hours before monitoring by ESI-MS. No reaction or MBE transfer was observed in the cases of MacpA, MacpB and MmpA_3a (Figure S.9.17). Conversely, efficient hydrolysis of the MBE-MmpE_ACP was observed (Figure S.9.18, no remaining signal for MBE-MmpE_ACP) although a small but identifiable species corresponding to MBE-MupB was observed (Figure 4.17 B, middle, peak coloured in pink). However, a far greater proportion of MupB:holo-MmpE_ACP complex was formed, presumably via association with holo-MmpE_ACP generated (Figure 4.17 B, right sided red colour peak). This may also indicate that the MBE substrate hydrolysed when attached to an ACP, perhaps via H_2O in the active site and for some other reason MBE-pant is less prone to this. Nonetheless, MBE-MmpE_ACP and MupB can interact and form a 'reactive' complex which also corroborates the NMR titration data.
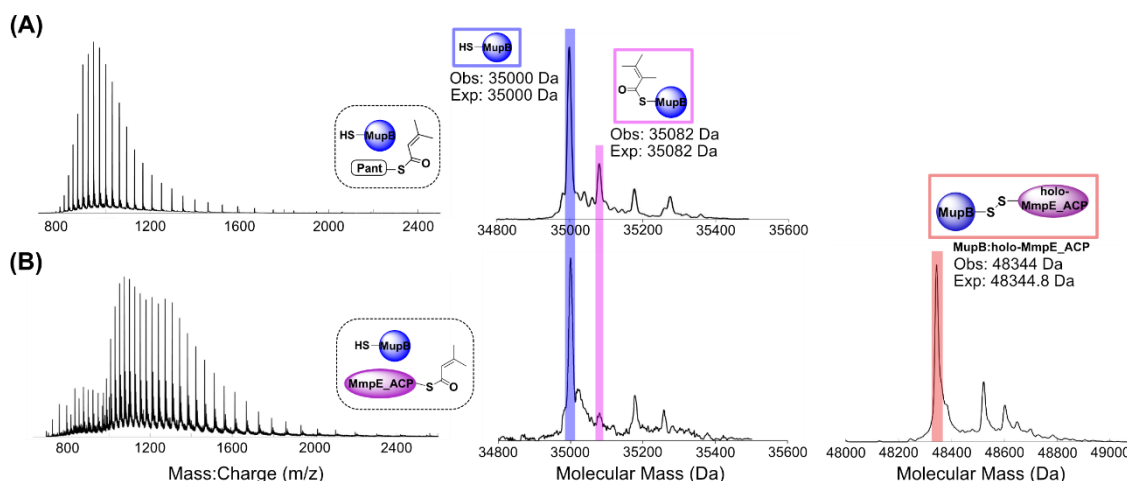
Figure 4.17. *In vitro* acyltransferase (AT) assays of MmpE_ACP and MupB. (A) ESI-MS analysis of MupB (coloured in blue) incubated with MBE-Pant (control) confirming transfer of the MBE group to MupB (coloured in pink) with an observed mass 35082 Da (expected 35082 Da). (B) ESI-MS analysis of MupB incubated with MBE-MmpE_ACP formed MBE-MupB (observed mass: 35082 Da, pink) and MupB:holo-MmpE_ACP complex (observed mass: 48344 Da, red).

The AT activity of Tml_ACPs was investigated in the same manner. Once again, the MBE derivatised TacpA, TacpB and TmpA_3c showed no transfer or non-productive hydrolysis with TmlB (Figure S.9.19) and MBE-TmpE_ACP was hydrolysed to holo-TmpE_ACP along with the TmlB:holo-TmpE_ACP complex formation (Figure 4.18 A). Unlike MupB, TmlB showed no detectable loading with MBE from this ACP (Figure 4.19 B, middle, coloured in violet) although once again TmlB could be primed directly using excess MBE-pant (Figure 4.19 B).
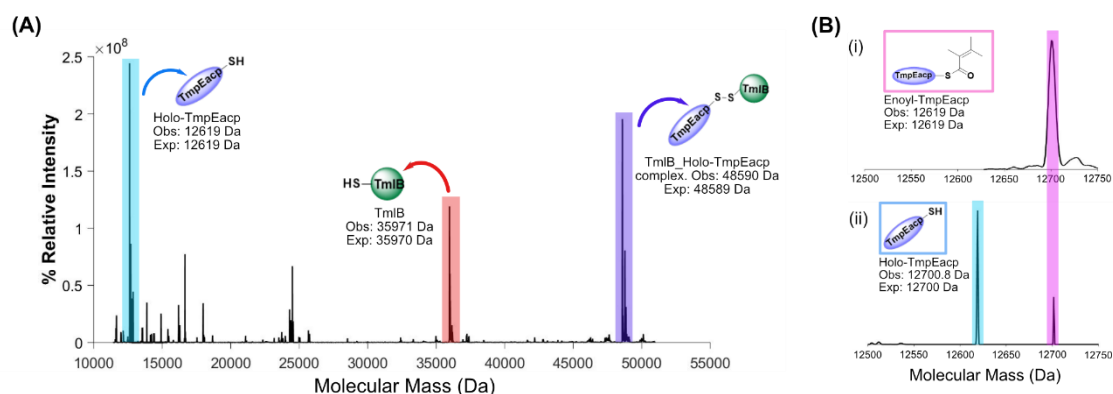
Figure 4.18. *In vitro* acyltransferase (AT) assay of TmpE_ACP. (A) Deconvoluted ESI-MS spectra of acyl transferase (AT) *in vitro* assay result of MBE-TmpE_ACP with TmlB. In presence of TmlB (observed mass: 35971 Da), MBE-TmpE_ACP (observed mass: 12700.8 Da) was observed to be hydrolysed, producing holo-TmpEacp (observed mass: 12619 Da) and also form a TmlB:holo-TmpE_ACP complex of observed mass 485920 Da was formed. (B) (i) Projected representation of deconvoluted MS spectrum of MBE-TmpE_ACP (control) and (ii) the hydrolysis of MBE-TmpE_ACP to holo-TmpEacp in the presence of TmlB, suggesting that TmpE_ACP may transfer the substrate to the putative reverse esterase TmlB.
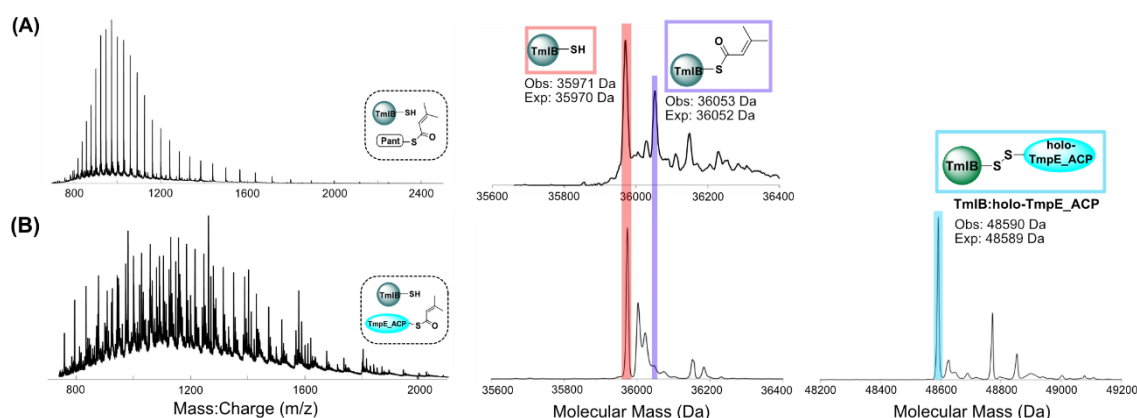
Figure 4.19. *In vitro* acyltransferase (AT) assay of TmpE_ACP. (A) Charge states and subsequent deconvolution spectrum of TmlB (coloured in red) incubated with MBE-Pant (Control), showing an MBE-TmlB peak (coloured in violet) of observed mass 36053 Da (expected 36052 Da). (B) Charge states and deconvoluted spectrum of TmlB incubated with MBE-TmpE_ACP formed the TmlB:holo-TmpE_ACP complex (observed mass: 48590 Da), coloured in cyan but no detectable MBE-TmlB peak was observed.

In summary both modular ACPs from MmpE/TmpE show an interaction with MupB/TmlB respectively with a small percentage of transfer of the MBE substrate mimic being observed in the case of the mupirocin enzymes. Loading of acyl chains onto synthase partners followed by rapid hydrolysis and loss of these derivatives has been observed elsewhere, for example in loss of acetyl priming units from the hydroxymethylglutaryl synthase of the β-branching cassette.[89,141,142] This is not unexpected on the assay timescale and is likely exacerbated by using a highly truncated substrate mimic that is not fully stabilised in the enzyme active site. The conclusion is that these ACPs are the likely donor ACPs and following epoxidation on MmpE, the pre-monic acid (or thiomarinol equivalent) is transferred to the reverse esterase enzyme.

4.6 Generation and insights of MupB/TmlB:holo-ACP complexes

The observation of crosslinked complex formation between MupB/TmlB and holo-M/TmpE_ACP was further investigated to understand their reaction extent and mechanistic insights. The complex formation was conducted by mixing holo-MmpE_ACP with partner proteins MupB in a 1:2 ratio in buffer (50 mM Tris, 100 mM NaCl, 0.5 mM TCEP, 10% glycerol, pH 8.0) and left at room temperature prior to analysis by gel electrophoresis and ESI-MS. At 30, 60, 120 and 240 mins, 10 μL samples of the reaction mixture were taken and the reaction was quenched by the addition of SDS-PAGE loading buffer. The progress of the complex formation was monitored by running the samples in an SDS-PAGE gel alongside two controls containing only MupB or holo-MmpE_ACP (Figure 4.20 A). A clear band of molecular mass ~ 48000 Da emerged in the MupB-holo-MmpE_ACP complex indicating successful complex formation although conversion was likely only a few percent. An attempt to isolate this complex compound using analytical SEC was unsuccessful possibly due to the high concentration of unreacted MupB compared to the complex. Generation of complex was also monitored by ESI-MS analysis. MupB/TmlB were incubated with their respective holo-ACPs in a 1:1 ratio into the same buffer system stated above for 3 hours. Then the crosslinked complex generated both in the mupirocin and thiomarinol systems were monitored by ESI-MS analysis. This showed the formation of MupB:holo-MmpE_ACP (Figure 4.20 B) and TmlB:holo-TmpE_ACP (Figure 4.18 C) complexes of molecular masses 48344 Da and 48590 Da (expected mass 48344.8 Da and 48589 Da), respectively. In terms of conversion efficiency, there was ~60% MupB:holo-

MmpE_ACP complex formation, whereas a very poor conversion efficiency (~20 %) was observed for TmlB (Figure S.9.20). When comparing this result with complex formation in the presence of MBE derivatised ACPs (Figure S.9.18 and 4.18 A), the conversion efficiency to complex formation for both MupB and TmlB were found to be greater than with holo-ACPs, suggesting the very interesting observation that the MBE substrate might activate the Ppant arm of holo-ACP to form a stable MupB/TmlB:holo-ACPs crosslinked complex.
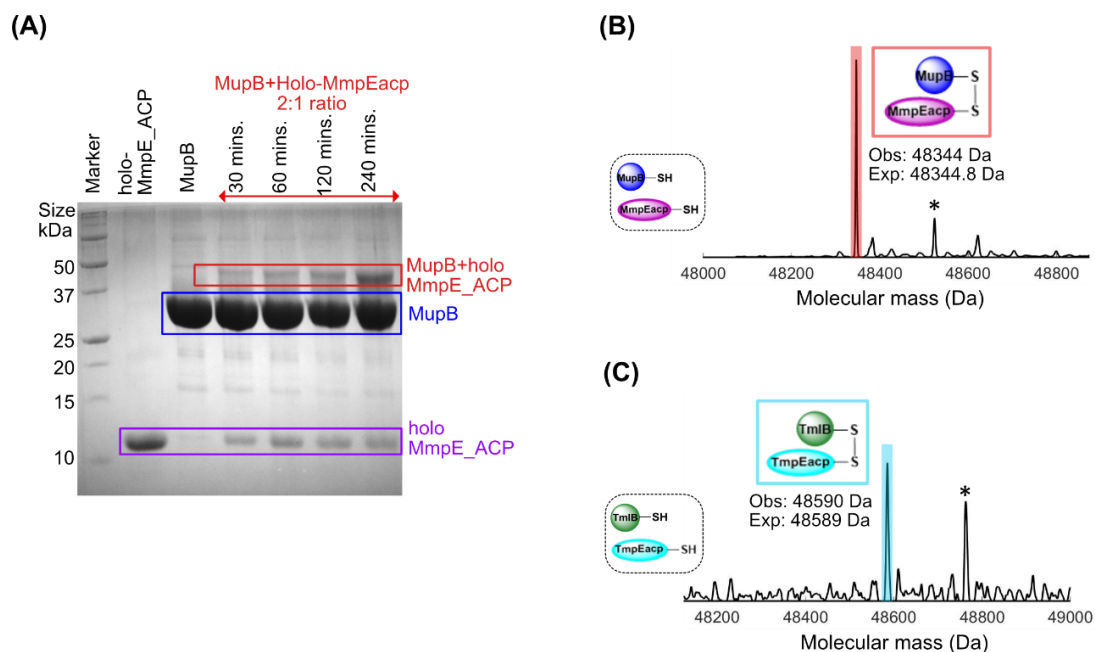
Figure 4.20. (A) Cross-linking reaction of MupB with holo_MmpE_ACP in 2:1 ratio. SDS-PAGE gel after incubationg the samples for 30, 60, 120 and 240 minutes are shown (red coloured) along with MupB (blue coloured) and holo_MmpE_ACP (violet coloured). Cross-linking between MupB and holo_MmE_ACP results in gel shifts of the MupB bands from 35 kDa to 48 kDa (MupB-holo_MmpE_ACP cross-linked complex). (B) The *in vitro* cross-linking reaction between MupB and holo_MmpE_ACP producedthe deconvoluted ESI-MS spectrum of cross-linked complex MupB-holo_MmpE_ACP (observed mass: 48344 Da, expected mass: 48344.8 Da), coloured in red. (C) The *in vitro* cross-linking reaction between TmlB and holo_TmpE_ACP produced the deconvoluted ESI-MS spectrum of cross-linked complex TmlB-holo_TmpE_ACP (observed mass: 44590 Da, expected mass: 48589 Da), coloured in cyan. The asterisk peak in (B) and (C) at +178 Da is for the N-terminus gluconylation of MupB.

Similar crosslinked complexes have been reported where modified ACPs were used to capture KSs and study protein-protein interactions which are often weak or transient and otherwise difficult to capture.[137,143] Recently, thiocyanate probed AcpP was used to gain excess to the active site of cognate KS, FabF as the unlabelled holo-AcpP struggled to form cross-linked complex. Modifying the AcpP with thiocyanate facilitated the cross-linking via disulphide bond formation between active site C163 and the AcpP-SCN sulphur atom,

releasing free CN^-.[144] Another crosslinked complex formation was reported from the disorazole biosynthetic pathway, DSZS AT-ACP1 (PDB:5Z4K) where the DSZS AT was shown to form a crosslinked complex with modified ACP.[145] In this work the authors showed that the DSZS AT S86C mutant protein was found to form the crosslinked complex with bromoacetamide probed ACP1 and proposed a cross-link reaction to generate a covalent DSZS AT-ACP1 complex (Figure S.9.21 A). A recent study of AcpP crosslinked crystal structures with KAS I type enzyme FabB (PDB:6OKC) and KAS II type enzyme FabF (PDB:6OLT) were reported by *Mindrebo et. al.*[146] They also showed that the active site cysteine of both FabB and FabF forms a covalent bond with the α-carbon of the fatty acid substrate (Figure S.9.21 B). However, to the best of our knowledge, no ACP:KAS III cross-linked complex structure has been reported to date. Therefore, the formation of crosslinked complexes between the KAS III like enzymes MupB/TmlB and holo-M/TmpE_ACPs could open a new door to understand their interactions and mechanistic insights while transporting acyl substrate.

The effect of the active site cysteine on complex formation was investigated by mixing the MupB:C116A mutant protein with holo-MmpE_ACP following the same reaction condition as described above and analysed by ESI-MS. No complex formation was observed (Figure 4.21) suggesting that the active site C116 residue in MupB is responsible for cross-linking reaction. Holo-MmpE_ACP was ~30% dimerised at the observed mass of 26688.4 Da (expected mass: 26689.6 Da) which is consistent with previous work described in this chapter. In this case the surrogate nucleophile, S224, failed to form a cross-link, most likely as this would be via an unstable S-O bond.
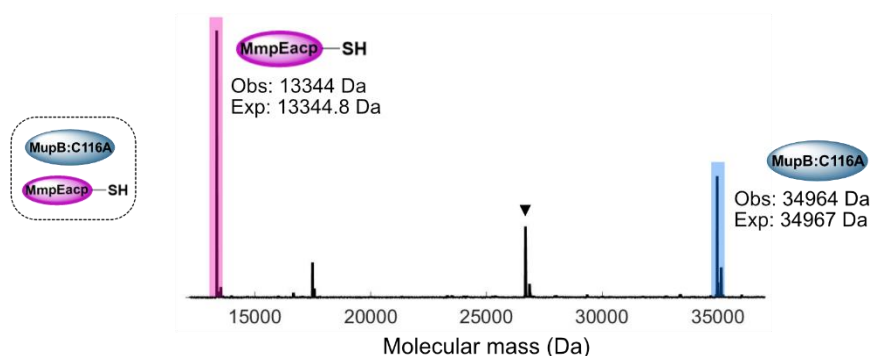
Figure 4.21. (A) The deconvoluted ESI-MS spectra for *in vitro* cross-linking reaction between the cysteine mutant MupB:C135A and holo_MmpE_ACP is showing no complex formation of expected mass 48344 Da. The holo_MmpE_ACP formed a dimer at observed mass 26688.4 Da marked with a black triangle.

4.7. Summary

The NMR titration and ESI-MS assays showed that only the modular MmpE_ACP interacts with MupB and transferred the MBE moiety to MupB albeit in very small quantities. The hydrolysis reaction was seen with TmpE-ACP and TmlB in thiomarinol system and no derivatised TmlB was detected. The formation of covalently linked complex between Mup/TmlB and holo-M/TmpE_ACP may open routes to studying KAS III – ACP recognition sites.

This KAS III:ACP interaction suggests a redesign of the previously accepted mupirocin biosynthesis pathway in such a way that MupB offloads MmpE derived substrates. The fully elongated fatty acid chain (9HN) formation has been reported in our recent work, where the final 9HN resides on the MmpB module ACP tridomain (MmpB-ACP567).[40] Hence, MmpB-ACP567 is the potential ACPs which delivers the fully elongated 9HN fatty acid chain to an acylated MupB for esterification. As MacpA and MacpB do not interact with MupB then the possibilities of delivering the 7-hydroxy heptanoyl acid (7HH) become minimised. So, the new pathway and timing of esterification reaction is proposed in the Figure 4.21.
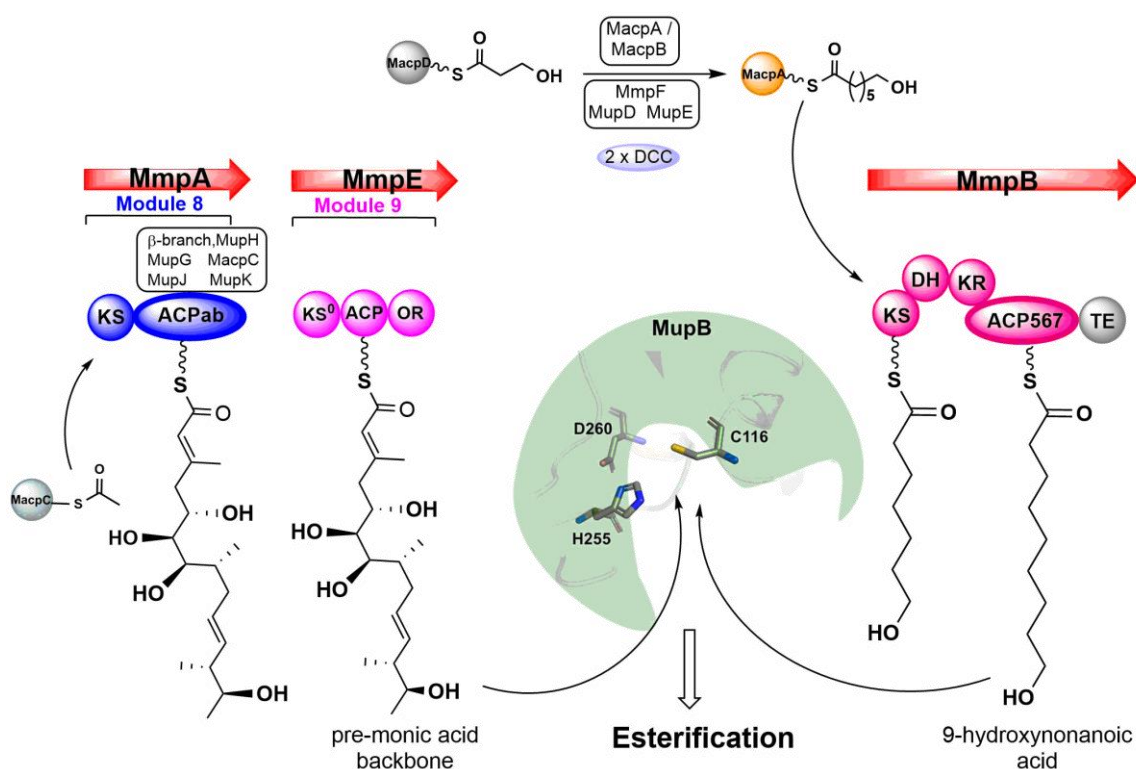
Figure 4.21. Schematic representation of the putative esterification reaction in mupirocin biosynthesis. The C_{17} pre monic acid backbone is assembled by MmpD, MmpA and MmpE. The esterification reaction might take place in the active site cavity of the putative MupB between the C_{17} monic acid backbone transported by MmpE_ACP and the fully elongated 9HN carried by MmpB_ACP567. The fatty acid elongation has been reported to take place via 2 consecutive DCC (decarboxylative Claisen condensation) reactions starting with 4HB-MacpD in the presence of MmpF, MupD and MupE, where the extender malonyl unit is delivered by MacpA and/or MacpB to produce 7HH-MacpA. The final round of DCC carried out in MmpB, producing fully elongated (9HN) fatty acid chain is then ready for esterification reaction.[40] ACP- acyl carrier protein, DH- dehydratase, OR- oxidoreductase, KS- ketosynthase, TE- thioesterase.

5. Functional investigations of MupB and TmlB

This chapter aims to investigate which partner ACP delivers the 9-HN fatty acid chain to MupB and TmlB for esterification, with a focus on the tri-domain MmpB_ACP567 and tetra-domain TmpB_ACP1-4. Previously, several PKS systems reported a requirement for an ACP for substrate delivery in the context of these alternate KAS-III like/acyl transferase family.[63,110] Amongst these the iterative PKS type I PtmQ_ACP from the pactamycin biosynthesis pathway is required to deliver 6-MSA to the reverse esterase PtmR. Interestingly this ACP shows sequence similarity (20.6 %) and conserved Ppant binding motifs with each ACP domain of the MmpB tri-domain ACP and TmpB tetra-domain ACP (Figure 5.1).
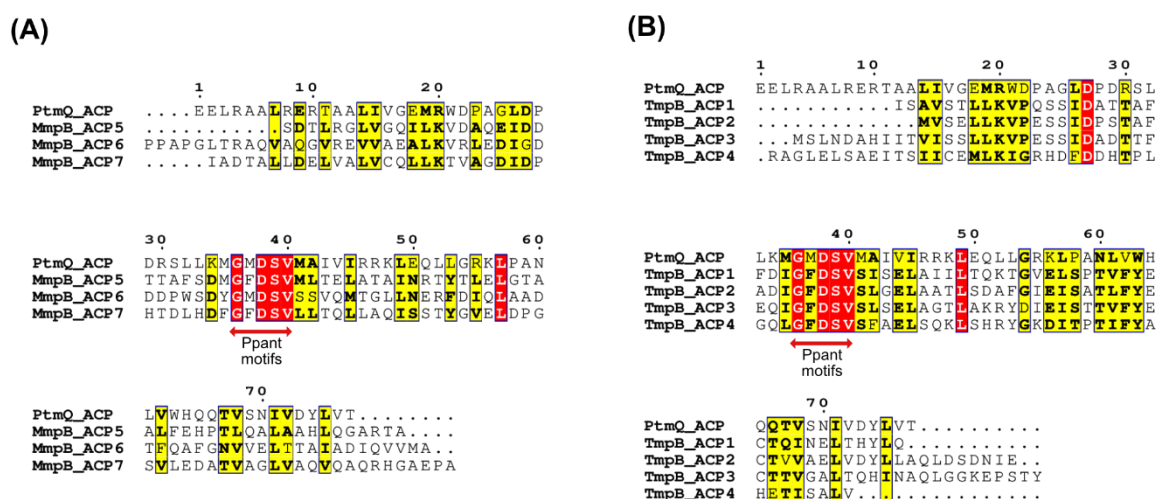
Figure 5.1. Multiple sequence alignment (MSA) of (A) individual MmpB_ACP (5,6, and 7) and **(B)** individual TmpB_ACP (1,2,3 and 4) with the pactamycin biosynthesis pathway PtmQ_ACP which delivers 6-MSA to produce the esterified product pactamycin. The red highlighted blocks represent the highly conserved residues among the ACPs. The Ppant motifs are indicated by a red arrow and the Ppant binding serine residue is marked with a blue circle.

In chapter four, it has been shown that MupB and TmlB interacted with MmpE_ACP and TmpE_ACP from their respective biosynthetic pathway. MmpE_ACP was also observed to deliver the pre-monic acid mimic, MBE substrate to the MupB (Figure 4.17). So, to identify the partner ACP responsible for delivering the fatty acid chain to the reverse esterase enzyme, the protein-protein interactions were initially investigated by titrating MupB with ^{15}N labelled ACP5 and ACP7 of tri-domain MmpB_ACP567 and followed by NMR CSP analysis. Most importantly, the mechanistic investigation of the reverse esterase activity of

MupB was also conducted in presence of the MBE substrate and a 9HN modified ACP. Similar experiments were also carried out for TmlB.

5.1 Expression and purification of MmpB_ACP567

The tri-domain MmpB_ACP567 gene had previously been cloned into a pET28a vector by Dr Xu Dong (PhD Bristol, 2011) and this plasmid was transformed into *E. coli* T7 and grown following the general protein expression protocol for LB media. The harvested cells were resuspended in Buffer A, sonicated and the soluble protein supernatant was extracted by centrifugation at rpm 15 K for 25 mins. The soluble fraction was first isolated by IMAC and then purified and exchanged in tris buffer (25 mM Tris, 150 mM NaCl, pH 7.5, 1 mM DTT) by SEC chromatography. Successful expression generated a visible band on SDS-PAGE at ~30 kDa (calculated mass 29985 Da) and was confirmed by ESI-MS (observed mass: 29982 Da). Finally, the SEC fractions of MmpB_ACP567 were mixed, and spin concentrated to 530 µM, flash frozen and stored at -20 °C (Figure 5.2 A).

To supplement experiments using the full-length tridomain construct, each separate ACP gene was produced synthetically (Thermo Fisher) and sub-cloned into a pET151-D/TOPO plasmid bearing an N-terminal His_6 tag sequence. The plasmids were transformed into *E. coli* T7 Express cells and cells grown as expressed for MmpB_ACP567. All three ACPs were soluble and gave distinct bands at the expected molecular masses by SDS-PAGE (Figure 5.2 B, C, and D). The protein isolation and final purification were conducted by IMAC and SEC in tris buffer, pH 7.5. Analytical size exclusion chromatography of purified ACPs (50 µM) confirmed that MmpB_ACP5 elutes as a single oligomeric species and MmpB_ACP6 and 7 elute as apparent dimers. Finally, the ESI-MS analysis of the multiple charge states of the ACPs produced the deconvoluted spectra of observed masses which are showed in figure 5.2, confirm the successful expression and purification of MmpB_ACP5, MmpB_ACP6 and MmpB_ACP7.
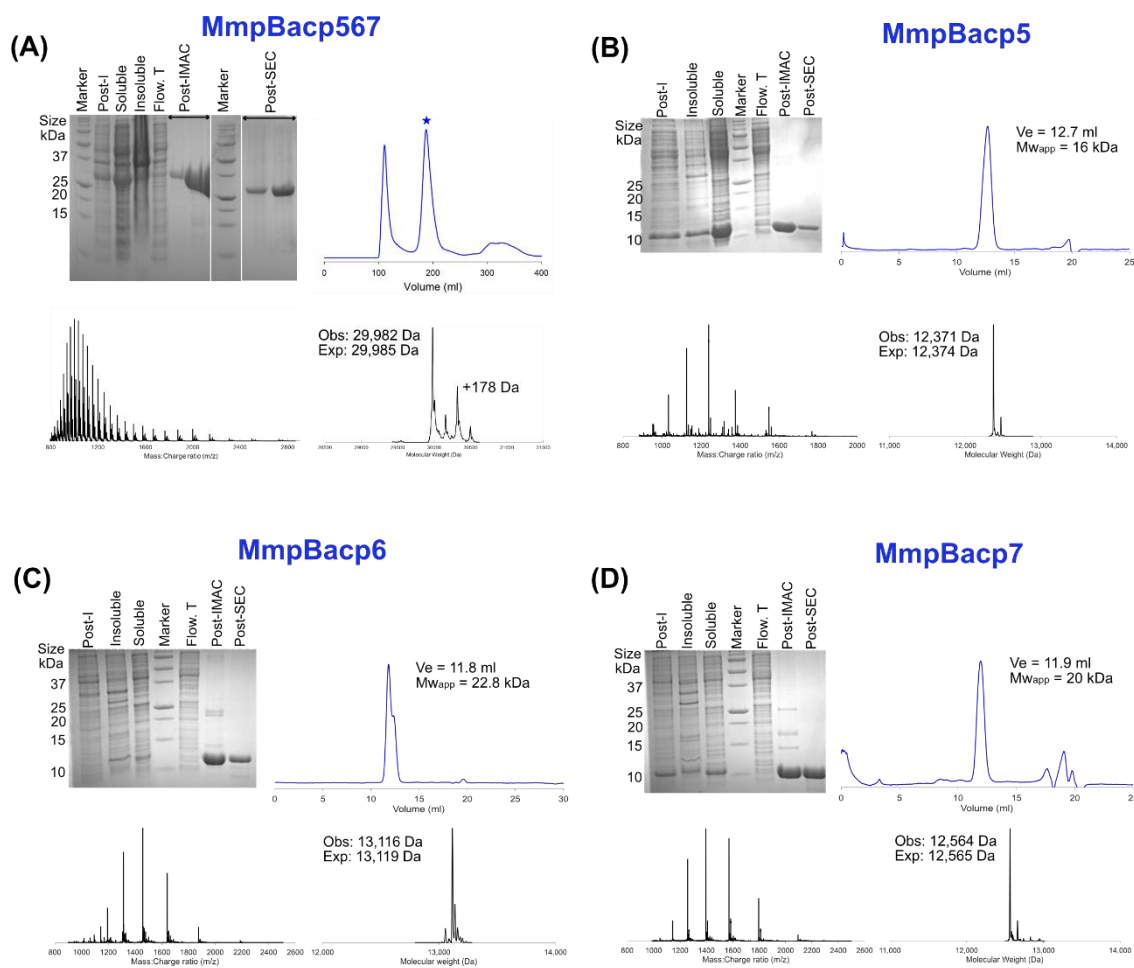
Figure 5.2. Purification followed by SDS-PAGE and ESI-MS characterisation of MmpB_ACP567 and each individual ACP of MmpB_ACP567. (A) Purification of MmpB_ACP567. SDS-PAGE following the purification of MmpB_ACP567 via IMAC and SEC. Analytical SEC chromatogram shows MmpB_ACP567 eluting as a single oligomeric species (marked with a blue star). ESI-MS yields a deconvoluted spectrum of observed mass: 29982 Da (expected:29985 Da), confirming the purity of the protein. Analytical SEC and ESI-MS analysis of B) MmpB_ACP5 C) MmpB_ACP6 and D) MmpB_ACP7.

For NMR titration studies, each of the MmpB_ACP (5,6,7) genes were transformed into T7 express cells and expressed via media exchange into M9 supplemented with NH_4Cl (^{15}N, 99%) (Experimental 7.9). The seed broth was initially cultured in 2 L of LB media according to this procedure until the optical density was at 1.0. The cells were then taken out and carefully washed in M9 media before being resuspended in M9 media (1 L) enriched with additional nutrients. Prior to
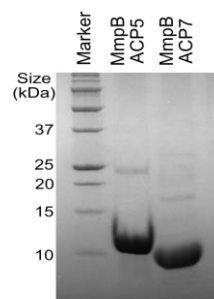
Figure 5.3. SDS-PAGE gel of purified ^{15}N-MmpB_ACP5 and ACP7 showing the bands of expected molecular weights 12523 Da and 12709 Da, respectively.

harvesting, the resuspended cells were cultured overnight (200 rpm) with 250 M IPTG at 16 °C. The proteins were isolated and purified as described above for the non-labelled proteins. MmpB_ACP6 was expressed in the insoluble fraction and could not be isolated. The isolation of ^{15}N-MmpB_ACP5 and 7 however was carried out via IMAC and exchanged into water to remove small molecules using SEC chromatography before being freeze-dried for storage at -20 °C. The purity of the expressed proteins was characterised by SDS-PAGE analysis (Figure 5.3).

5.2 Cloning, expression and purification of TmpB_ACP1-4

The Type I PKS TmpB structurally diverges from MmpB and has two extra ACPs, one in a tetra-domain ACP1-4 and another after an additional KS^0 (Figure S.9.22).[74] TmpB_ACP1-4 cloning was accomplished by PCR of *Pseudoalteromonas* sp. SANK 73390 genomic DNA with the predesigned primers of TmpB_ACP1-4 (Table S.9.1). Agarose gel electrophoresis confirmed successful amplification with a band at ~900 bps (Figure S.9.23 A) and the PCR products were later purified and ligated into the pOPINF expression vector following the general plasmid ligation protocols (Experimental 7.1). The DNA was then transformed into NEB-5α bacterial cell and successful ligation was confirmed by X-gal screening (blue-white colony screening). The white colonies were cultured, and the plasmid purified using a standard mini-prep plasmid purification procedure. Further confirmation of TmpB_ACP1-4 ligation into the pPOINF plasmid was provided by GENEWIZ. Several attempts were conducted to prepare DNA of the individual TmpB_ACPs (1-4) but were all unsuccessful. Instead, a single TmpB_ACP3 was synthesized by Thermo Fischer and subcloned into the pET151-D/TOPT plasmid bearing an N-terminal His$_6$-tag as a representative of the individual TmpB_ACPs.

Additional mini didomain constructs, TmpB_ACP12 and TmpB_ACP34 were cloned successfully from the predesigned primers (Table S.9.1) using genomic DNA. Sequences were ligated into the pOPINE vector and confirmed by X-gal screening. The plasmid culture and purification were carried out as described above and final confirmation of ligation into the pOPeINE vector was provided by GENEWIZ sequencing which showed that TmpB_ACP1-2 amplified correctly without the C-terminal His$_6$-tag.

The TmpB_ACP1-4 gene was transformed into *E. coli* T7 Express cells and expressed in 1.5 L scale following the general procedures described above. The cells were harvested,

sonicated and insoluble pellets were separated from the soluble supernatant by centrifugation. IMAC chromatography was used to isolate the soluble protein, but SDS-PAGE gel showed no visible bands for IMAC fractions and a distinct band around ~42 kDa in the insoluble sample confirms that TmpB_ACP1-4 is expressed in the insoluble pellet (Figure S.9.23 B).

Therefore, the TmpB_ACP3 and di-domain TmpB_ACP12, TmpB_ACP34 genomic DNA were used to express the protein in *E. coli* T7 Express cells following the same procedure stated above for the MmpB_ACP5/6/7. TmpB_ACP3 and TmpB_ACP34 were harvested, sonicated, and extracted from the soluble supernatant by using IMAC chromatography and finally purified by exchanging into Tris buffer by SEC chromatography. The successful expression and purity of these proteins was confirmed by SDS-PAGE and ESI-MS. TmpB_ACP3 data was not collected beyond confirmation by SDS-PAGE (Figure 5.5). The final characterisation of TmpB_ACP34 by ESI-MS gave an observed molecular mass 22011 Da (expected mass 22009 Da, Figure 5.6 B).

The TmpB_ACP12 lacked a His_6-tag and was purified using an anion exchange column, HiTrap Q FF (5 ml). Firstly, TmpB_ACP12 was exchanged into Q-Buffer A, pH 7.5 (pI = 4.12) before loading onto the Q-column which produced a negatively charged protein. Upon loading of 10 mL of TmpB_ACP12 onto the Q-column, the protein was strongly retained to the positively charged stationary phase of the column. Then TmpB_ACP12 was isolated using a NaCl gradient and was observed to elute at a NaCl concentration of ~160-200 mM with a peak of absorbance ~1500 mAU (Figure S.9.24). Final purification was achieved using SEC chromatography (25 mM Tris-HCl, 150 mM NaCl, pH 7.5). The SDS-PAGE and ESI-MS analysis confirm the successful purification of TmpB_ACP12 (Figure 5.6. A).
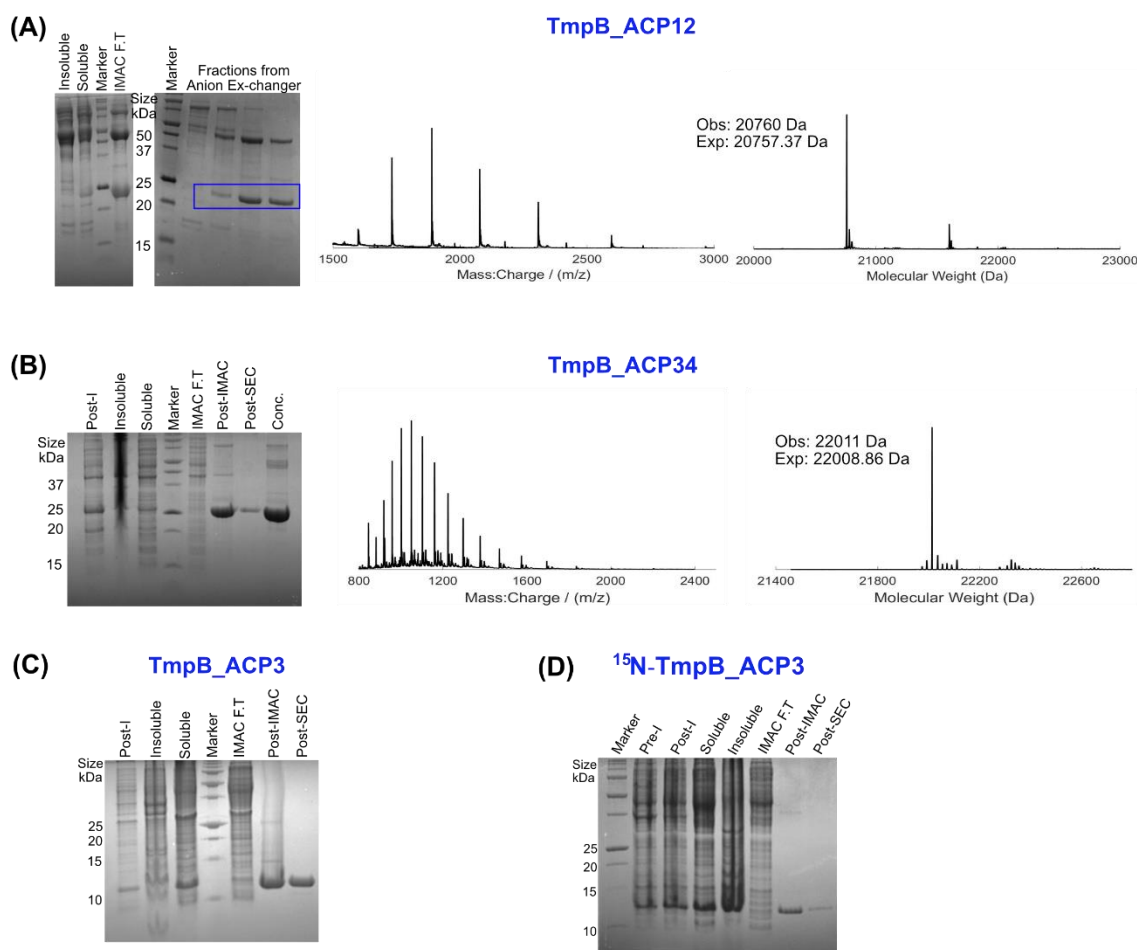
Figure 5.5. Purification followed by SDS-PAGE and characterisation of (A) TmpB_ACP12: SDS-PAGE following the purification of TmpB_ACP12 via IMAC and SEC. B) TmpB_ACP34: the ESI-MS spectrum confirms the correct molecular mass (observed: 20760 Da, expected: 20757.37 Da. C) TmpB_AC3: SDS-PAGE following the purification of TmpB_ACP34 via IMAC and SEC. ESI-MS yields a deconvoluted spectrum of observed mass: 22011 Da (expected: 22008.86 Da), confirming the purity of the protein. (D) ^{15}N-TmpB_ACP3: characterized by SDS-PAGE showing a band of expected mass around 13 kDa.

For NMR titration, ^{15}N-TmpB_ACP3 was expressed and purified via M9 media exchange following the same protocols described above for MmpB individual ACPs (5, 6 and 7) expression and purification. The successful isolation and purification were confirmed by SDS-PAGE, showing a band at ~ 13 kDa (calculated mass of ^{15}N-TmpB_ACP3: 12986 Da) depicted in Figure 5.5 D.

5.3 NMR titration of ^{15}N labelled ACPs with MupB and TmlB

Firstly, the NMR samples of ACPs were prepared in 50 mM phosphate buffer (pH 8.0) from the freeze-dried proteins concentrated up to 1-2 mM with 10% D_2O. Then one equivalent concentration of ^{15}N labelled MmpB_ACP and TmpB_ACP3 were incubated individually with three equivalent concentrations of MupB and TmlB (1:3), respectively in a 3 mm NMR tube. The PPIs were monitored by ^{1}H-^{15}N HSQC in collaboration with Dr Ash Winter.

Figure 5.6 represents the ^{1}H-^{15}N HSQC spectrum of ^{15}N-MmpB_ACPs without MupB (control, coloured in blue) with the ^{1}H-^{15}N HSQC spectrum of ^{15}N-MmpB_ACPs after addition of three equivalents of MupB (coloured in blue which are compared with the red coloured control peaks). The ^{15}N-MmpB_ACP5: MupB NMR titration showed clear peak broadening/intensity loss and CSPs (Figure 5.6 A right side red) in comparison with the control peak of ^{15}N-MmpB_ACPs without MupB (blue). The ^{15}N-MmpB_ACP7: MupB NMR titration showed mostly peak shift (Figure 5.6 B left side coloured in red) while compared with the control ^{15}N-MmpB_ACP7 titration without MupB (coloured in blue). Therefore, NMR titration suggested that both the MmpB_ACP5 and 7 appear to interact with MupB.

Figure 5.6. ^{1}H-^{15}N HSQC experiments to determine PPIs of ^{15}N-MmpB_ACP5 and 7 with MupB. (A) Control spectrum of MmpB_ACP5 at 298 K / pH 8.0 (blue) and in the presence of three equivalents of MupB (red), showing significant numbers of peak disappearances and peak shifts. (B) Control spectrum of MmpB_ACP7 at 298 K / pH 8.0 (blue) and in the presence of three equivalents of MupB (red) showing considerable peak shifts and peak broadening.
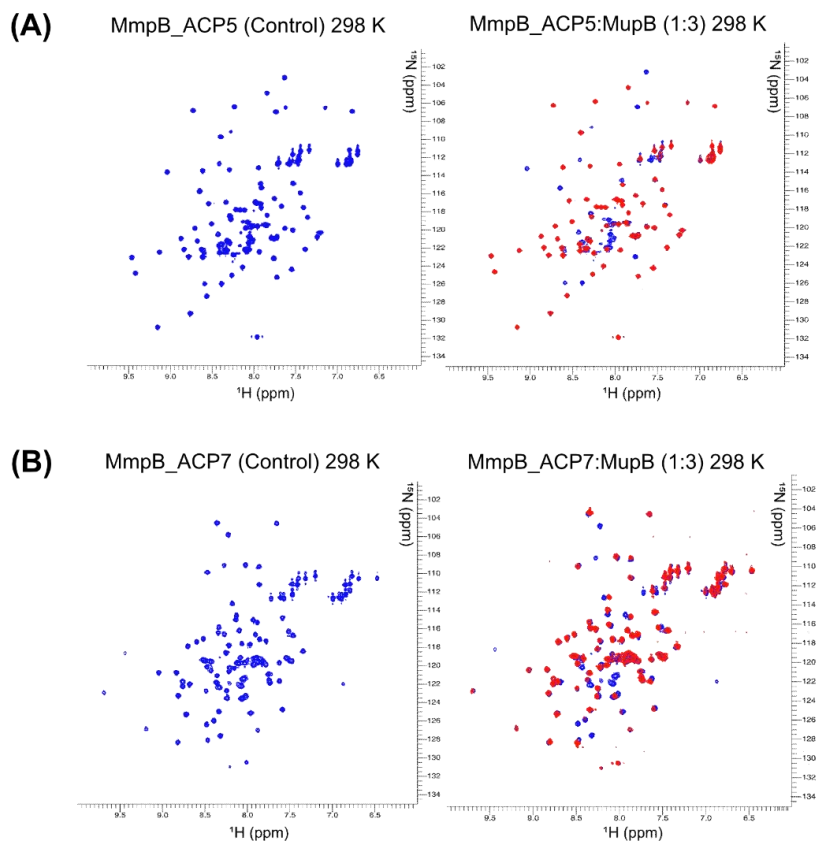
The ^{15}N-TmpB_ACP3 was also titrated with TmlB in 1:3 concentration ratio at 298 K and followed by NMR analysis (Figure 5.7). The TmpB_ACP3 was chosen as a representative of the individual ACP domains of the tetra-domain TmpB_ACP1-4. This ACP showed similar secondary structure architecture compared to the individual MmpB_ACPs (5,6,7) and contains the conserved Ppant-binding motif (GXXS) (Figure S.9.25). However, no peak shifts/disappearances were observed in the NMR titration experiment of ^{15}N-TmpB_ACP3: TmlB system (Figure 5.7), indicating that this tandem TmpB_ACP1-4 domain may contain ACPs that behave differently to their MmpB_ACP counterparts. Therefore, two possibilities might occur; first the tetra-domain TmpB_ACP1-4 might need at least two ACP domains to be active. Second only one ACP of the tetra-domain interacts with the TmlB to deliver the fatty acid chain, and it is not ACP3. These assumptions are justified later in this chapter.

Figure 5.7. ^{1}H-^{15}N HSQC experiments to determine PPIs of ^{15}N-TmpB_ACP3 with TmlB. (A) Control spectrum of TmpB_ACP3 at 298 K / pH 8.0 (blue) and in the presence of three equivalents of TmlB (red). No discernible chemical shift perturbations (CSPs) or broadening of residues were observed.
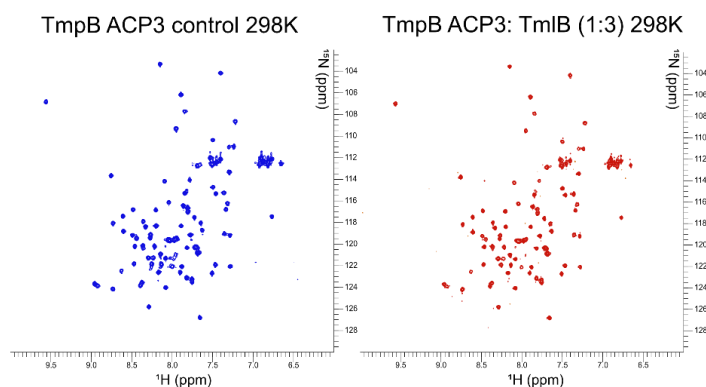
5.4 *In vitro* enzymatic upgrade of 9HN to MmpB_ACPs

9HN-Pantetheine (9HN-Pant) substrate was chemoenzymatically upgraded and loaded onto MmpB_ACP567 and individual MmpB_ACPs and conversion monitored by ESI-MS and Ppant ejection assays. Figure 5.8 shows the MmpB_ACPs completely upgraded with 9HN (middle peak, mass +497 Da) and Ppant ejection assay produced observed mass 417 Da (expected mass 418.1267 Da), indicating the 9HN (calculated mass 157 Da) loaded pantetheine fragment (calculated mass 261.1267 Da). The 9HN functionalised MmpB_ACPs were exchanged into the tris buffer (pH 7.5) via a zeba desalt column which can also remove the excess CoAs, unreacted substrate and cofactors. All the 9HN functionalised ACPs were stable at 4 °C overnight but preferred to be functionalised prior

to the esterification assay as the changes in ACP regions are crucial for esterified product analysis.
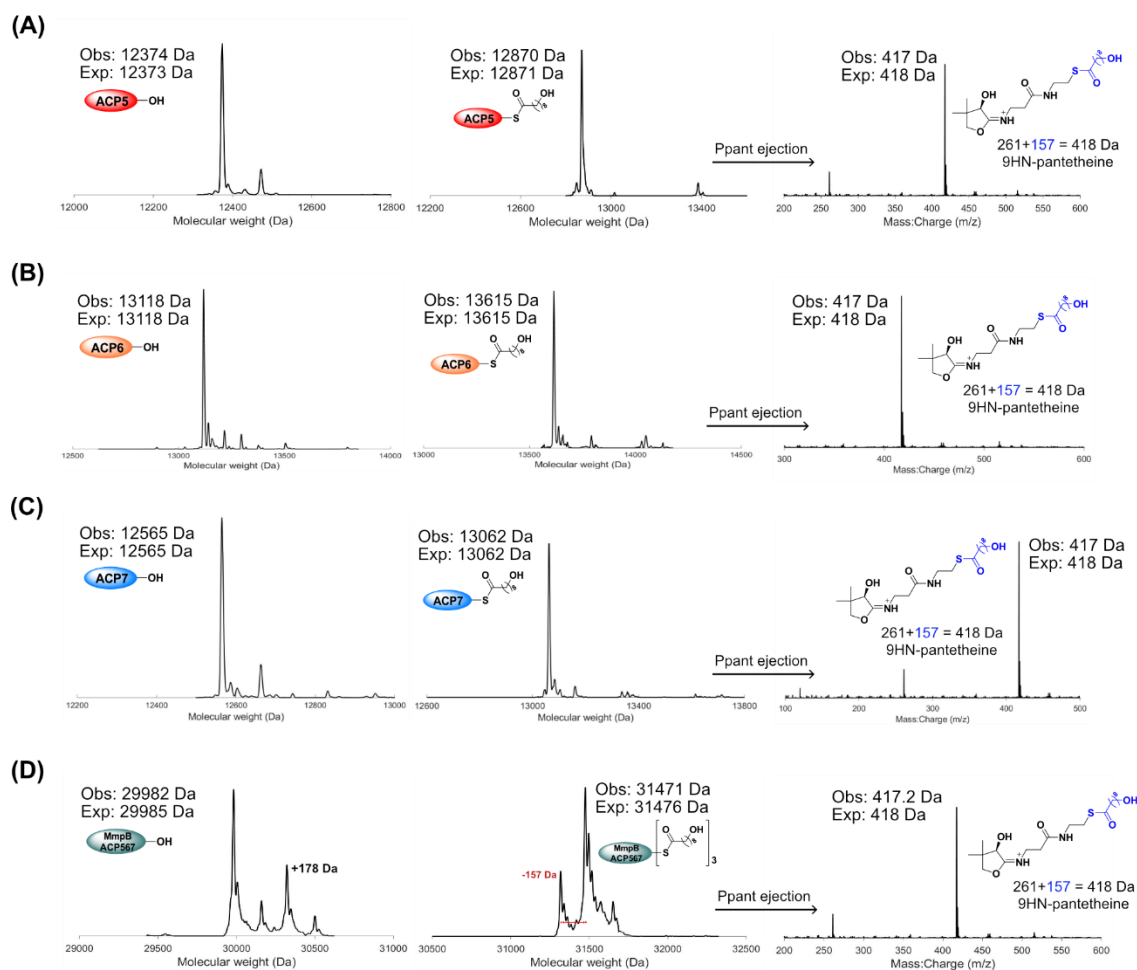
Figure 5.8. 9HN upgrade and loading to MmpB_ACPs. The deconvoluted spectra of apo-ACP (left side), 9HNACP (middle) and corresponding Ppant ejection (right side) of (A) MmpB_ACP5 (B) MmpB_ACP6 (C) MmpB_ACP7 and (D) MmpB_ACP567. The molecular mass changes of apo-ACPs to 9HN-ACPs are indicated in the corresponding spectra.

5.5 *In vitro* enzymatic upgrade of 9HN to TmpB_ACPs

Although thiomarinol bears an 8HO fatty acid side chain, 9HN-Pant substrate was used instead to functionalise the TmpB ACPs as 8HO-Pant was unavailable. Figure 5.9 (A) shows the TmpB_ACP3 functionalisation. TmpB_ACP3 was completely converted to 9HN-TmpB_ACP3 (observed mass 13336 Da). Amongst the didomain ACPs, both ACPs of TmpB_ACP12 were observed to be upgraded with 9HN efficiently, with only ~ 20% of constructs lacking the second modification. TmpB_ACP34 showed reasonable conversion but ~30% remained as unreacted apo-TmpB_ACP34 (Figure 5.9 B and C). In both cases, small spectrum of singly functionalisation (peaks less molecular masses of 157 Da) were observed and addition of more MupN and ATP had no effect. Moreover, the 9HN upgradation of these TmpB_ACPs were also confirmed by Ppant ejection assay as shown in Figure 5.9 (right side). In terms of stability, 9HN-TmpB_ACP3 was found completely stable at room temperature but the didomain functionalised constructs were observed to degrade more readily than their MmpB counterparts and were therefore prepared just prior to use in assay.
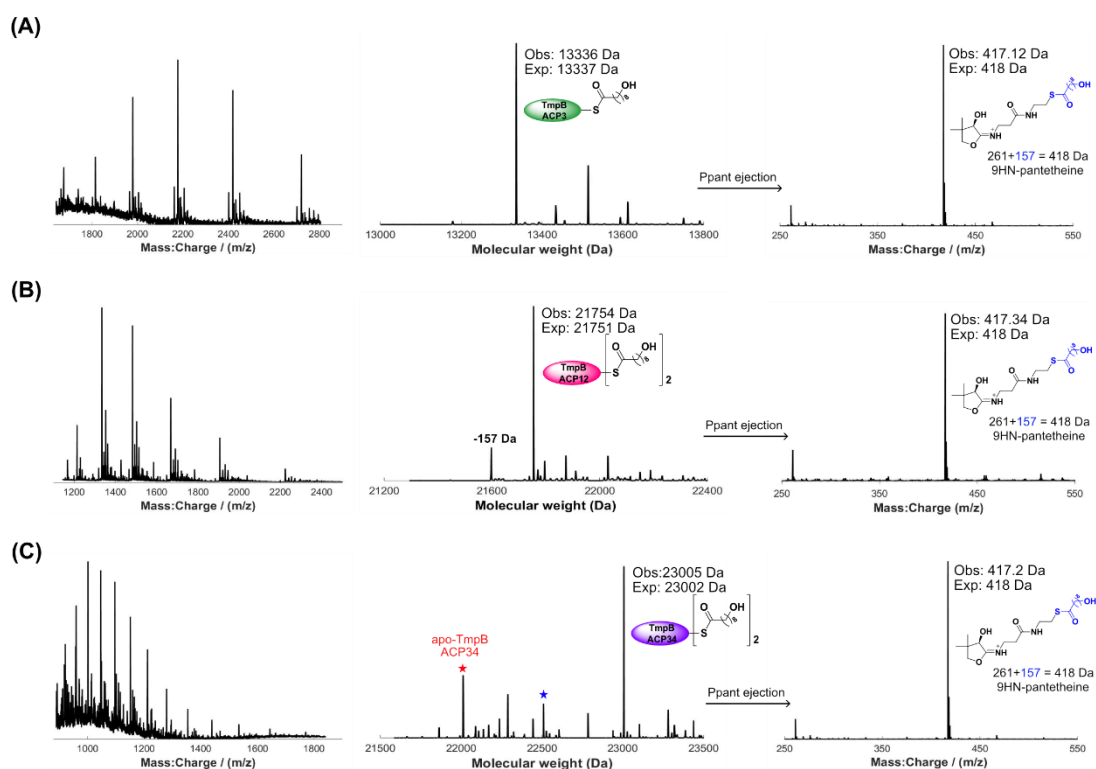
Figure 5.9. 9HN upgrade assay and loading to TmpB_ACPs. The deconvoluted spectra of multiple charge states of 9HN-ACP (middle) and their corresponding Ppant ejection (right side) of (A) TmpB_ACP3 (B) TmpB_ACP12 which has a singly 9HN upgraded spectrum at -157 Da and (C) TmpB_ACP34 (middle) with spectra of unreacted apo-TmpB_ACP34 (observed 22011 Da, expected 22008 Da) and single 9HN upgraded TmpB_ACP34 (observed 22509 Da, expected 22505 Da, marked with blue star).

5.6 Esterification reaction

5.6.1 Esterification assay of MupB

To pursue the main goal of investigating reverse esterase activity of MupB, an *in vitro* assay was conducted with the MBE and acetyl-Pant, 9HN-MmpB_ACPs and MupB. The assay was initiated via self-loading of the acetyl or MBE onto MupB by incubating 1.0 mM Ac/MBE-Pant with 50 μM MupB (50 μL reaction volume) at room temperature for ~2 hours in buffer (50 mM tris, 100 mM NaCl, pH 7.5). The self-loading was confirmed by ESI-MS analysis as described previously, typically achieving 30% for acetyl and 5-10% for the MBE substrate.

In the second step, 25 μM Ac/MBE-MupB was incubated with 25 μM of the 9HN-MmpB_ACPs at room temperature for 3 hours prior to analysis by ESI-MS. Ac/MBE-MupB was separately incubated with [9HN]$_3$-MmpB_ACP567 and individual 9HN-ACPs (5,6 and 7) of tri-domain MmpB_ACP567.

Figure 5.10 A represents the deconvoluted ESI-MS spectra of the control assay containing just MupB (without substrate) and [9HN]$_3$-MmpB_ACP567 which is compared with the MBE loaded MupB after incubation with [9HN]$_3$-MmpB_ACP567 (Figure 5.10 B). In the presence of MBE-MupB a new species of observed mass 31317 Da which is 158 Da less than the starting functionalised ACP was produced indicating a single 9HN substrate was hydrolysed off from the functionalised tridomain ACP. No peaks corresponding to the expected esterified product were observed. Similar hydrolysis of [9HN]$_3$-MmpB_ACP567 was also seen in the control assay (Figure 5.10A). An equivalent esterification assay was performed with MupB substituted with TmlB. The ESI-MS analysis again showed the production of [9HN]$_2$-MmpB_ACP567 species formed due to hydrolysis of [9HN]$_3$-MmpB_ACP567 (Figure S.9.26) but gave no productive esterification reaction.
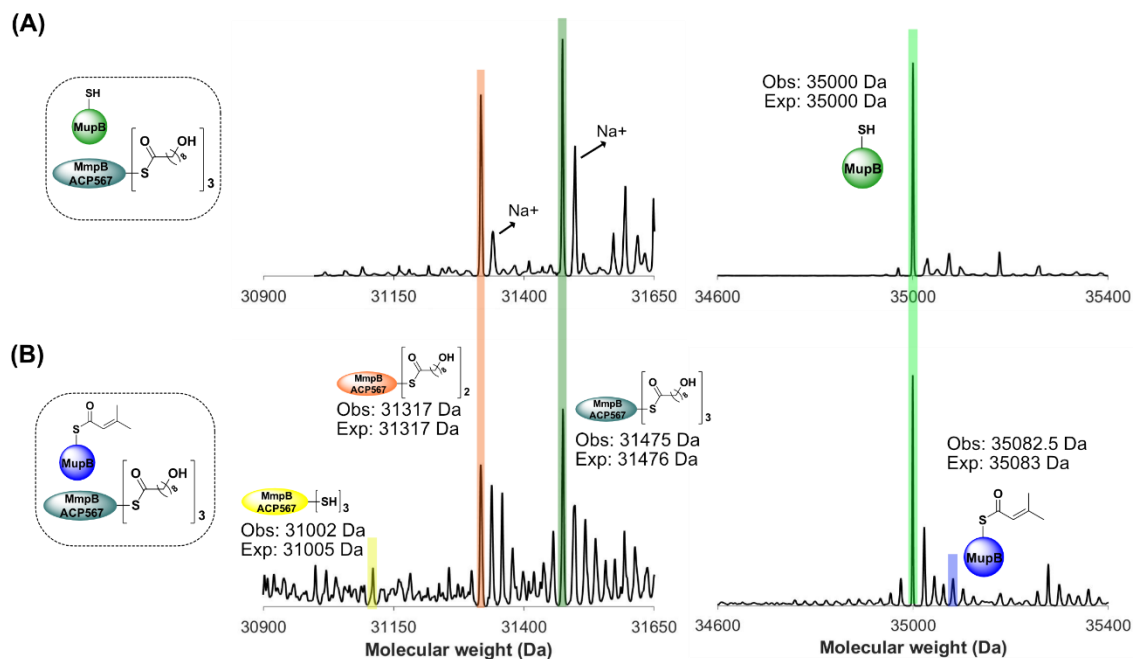
Figure 5.10. Esterification assay of MupB: (A) Control reaction between MupB and [9HN]₃-MmpB_ACP567, showing the deconvoluted spectra of [9HN]₃-MmpB_ACP567 (green) hydrolysed to produce [9HN]₂-MmpB_ACP567 (orange). A spectrum for MupB is shown with observed mass 35000 Da. **(B) MBE-MupB incubation with [9HN]₃-MmpB_ACP567 followed by ESI-MS:** the ACP region shows the conversion of holo-MmpB_ACP567 (yellow) and [9HN]₂-MmpB_ACP567 along with multiple Na-adduct spectra. Meanwhile, the MupB region shows a small spectrum of MBE-MupB along with unreacted MupB.

The PPIs between MupB and ^{15}N-ACP5 and 7 were confirmed previously by NMR titration assay, suggesting that these ACPs might be involved in transferring the 9HN fatty acid to MupB. Therefore, in vitro esterification assays were performed with Ac/MBE-MupB with 9HN-^{15}N-ACP5 and 7 (Figure S.9.27 & S.9.28; ^{15}N labelled variants were available at the time of the assay). In both cases, hydrolysis of the 9HN-^{15}N-ACP5 or -ACP7 was observed yielding the holo form. Both 9HN-^{15}N-ACP5 and -ACP7 were hydrolysed in the presence of just MupB (control).

9HN-MmpB_ACP6 was also incubated with Ac/MBE-MupB following the same reaction conditions. It was observed that in presence of MBE-MupB the 9HN-ACP6 (observed mass 13616 Da) produced a small quantity of holo-ACP6 (observed mass 13460 Da) which was then dimerized, showing the peak of observed mass 26917 Da (expected mass 26916 Da). Along with this a new deconvoluted MS peak of observed mass 27230 Da was seen which is equal to twice the mass of the 9HN-ACP6 species (expected mass 27230 Da) (Figure 5.11, middle spectrum) which likely arises from a small proportion of non-covalent dimer generated under soft-ionisation conditions and was not considered further.[147,148]
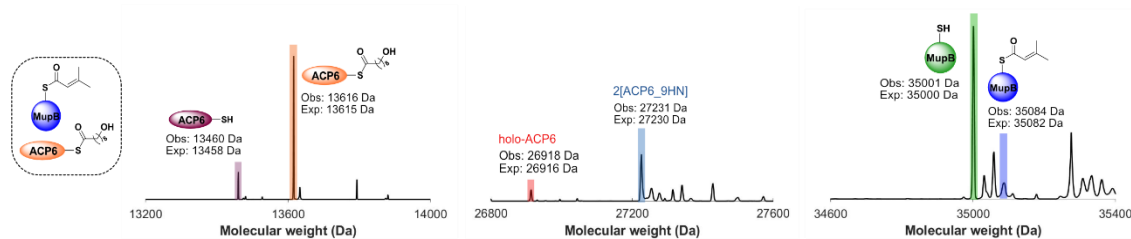
Figure 5.11. MBE-MupB incubation with 9HN-MmpB_ACP6 followed by ESI-MS: Left side the ACP region shows a small conversion tot holo-MmpB_ACP6. Middle region shows the deconvoluted spectra for holo-MmpB_ACP6 dimer (observed mass: 26918 Da) along with 9HN-MmpB_ACP6 dimer (observed mass: 27230 Da). The MupB region shows the spectrum of MBE-MupB along with unreacted MupB.

When Ac-MupB was used (Figure 5.12) hydrolysis to give holo-ACP6 was again observed (Figure S.9.29). The control reaction with just MupB and 9HN-ACP6 could not be analysed because the sample did not spray well when analysed by ESI-MS technique. Therefore, the identical reaction was conducted by adding MBE-TmlB with 9HN-ACP6. This was compared with the control reaction TmlB with 9HN-ACP6 and analysed by ESI-MS (Figure 5.13). The figure also showed the deconvoluted peaks of dimerised holo-ACP6 at 26917 Da. However, no esterified product was observed.
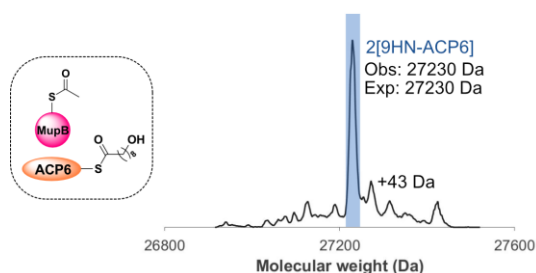
Figure 5.12. Ac-MupB incubation with 9HN-MmpB_ACP6 followed by ESI-MS: The deconvoluted spectra for the dimer of 9HN-MmpB_ACP6 (observed mass: 27230 Da).
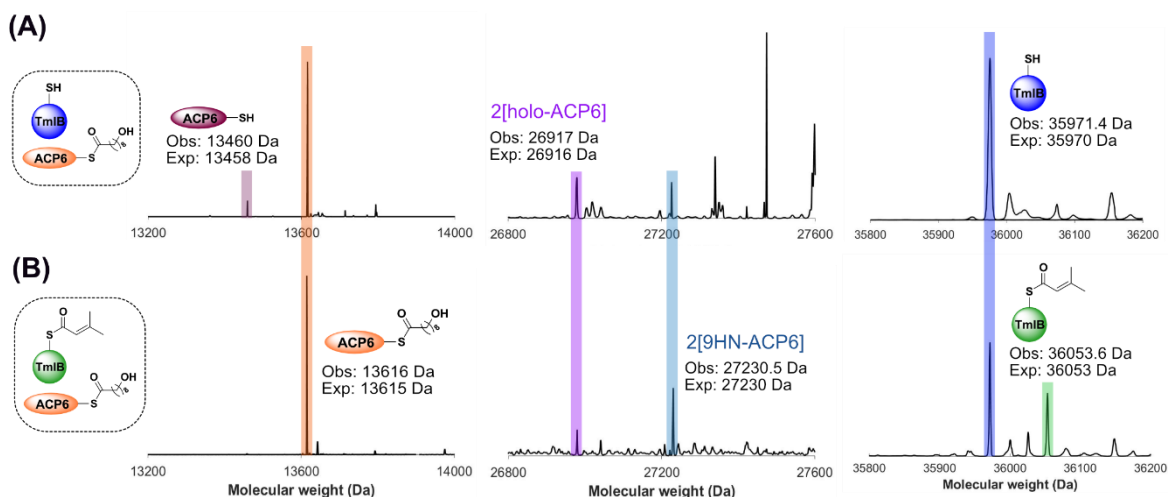
Figure 5.13. Esterification assay of TmlB. Deconvoluted ESI-MS spectra are shown for (A) Control reaction between TmlB (without substrate) and 9HN-MmpB_ACP6, showing the deconvoluted spectra of 9HN-MmpB_ACP6 (orange), hydrolysis product holo-MmpB_ACP6 (dark brown), holo-MmpB_ACP6 dimer (observed mass: 26917 Da) along with 9HN-MmpB_ACP6 dimer (observed mass: 27230 Da). (B) Equivalent spectra for MBE-TmlB incubation with 9HN-MmpB_ACP6.

Although the MmpB_ACP6 and MupB interaction could not be analysed by NMR titration, the ESI-MS analysis confirms that like the other two proteins, MmpB_ACP5 and 7, MmpB_ACP6 also interacts with MupB as well as the thiomarinol enzyme, TmlB.

Finally, it can be concluded that tri-domain MmpB_ACP567 and individual MmpB_ACP5, 6 and 7 can recognize MupB and TmlB. In these cases, the ACP can deliver the 9HN substrate into the active site cavity of MupB, but it is hydrolysed off from the Ppant arm of active ACP. This may be due to the lack of substrate stabilisation in the oxyanion hole caused by the incorrect binding orientation of the MBE substrate. MupB self-loads this substrate rather poorly suggesting that the MupB active site cavity may be selectively designed to accommodate a larger polar substrate like C_{17} pre-monic acid. Thus, the monic acid or the larger polar mimic of monic acid may enhance the esterification assay.

5.6.2 Esterification assay of TmlB with TmpB_ACPs

The *in vitro* esterification assay for TmlB was initiated by adding 30 μM of 9HN-TmpB_ACPs (ACP3, ACP12 and ACP34) with 20 μM of MBE-TmlB in separate batches in buffer (50 mM Tris, 150 mM NaCl, pH 7.5) and reacting at room temperature for ~ 3 hours.

For 9HN-TmpB_ACP3 and [9HN]$_2$-TmpB_ACP34, no changes were observed in comparison with the control assay (just TmlB with 9HN-TmpB_ACPs) (Figure S.9.30 and S.9.31) including lack of hydrolysis products. When incubated with [9HN]$_2$-TmpB_ACP12, a new observed mass of 21600 Da was detected corresponding to loss of 156 Da. This was equivalent to loss of one 9-HN molecule to yield [9HN]-TmpB_ACP12 (Figure 5.14). Hence, the results suggested that TmpB_ACP12 can recognise TmlB but TmpB_ACP3 and TmpB_34 do not although they have conserved Ppant binding motifs. TmpB_ACP3 and TmpB_ACP34 are unable to deliver the fatty acid chain into the enzymes active site cavity. This was also confirmed earlier in this chapter by NMR titration of TmlB with ^{15}N-TmpB_ACP3.
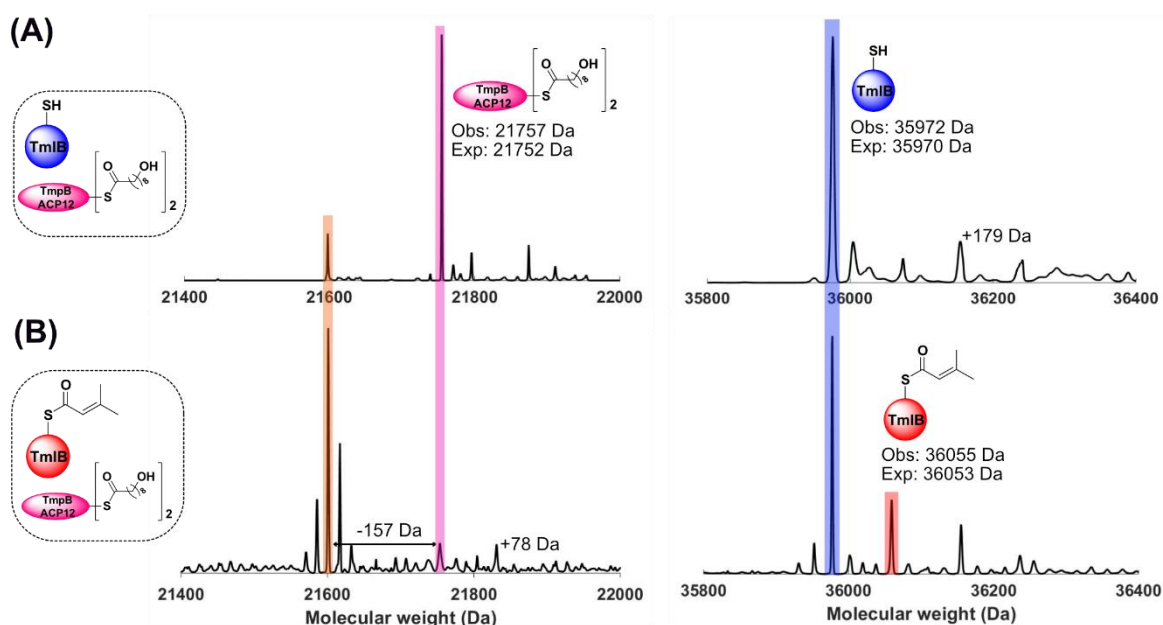
Figure 5.14. Esterification assay of TmlB: (A) Control reaction between TmlB and [9HN]$_2$-TmpB_ACP12, showing the deconvoluted spectra of [9HN]$_2$-TmpB_ACP12 (magenta) hydrolysed to produce holo-TmpB_ACP12 (orange). A spectrum for TmlB (without MBE) is shown with observed mass 35000 Da. (B) MBE-TmlB incubation with [9HN]$_2$-TmpB_ACP12 followed by ESI-MS: Middle region shows the deconvoluted spectra for holo-TmpB_ACP12. The TmlB region shows a spectrum of MBE-TmlB (red) along with unreacted WT-TmlB (blue).

5.7 Claisen condensation assay of MupB and TmlB

The lack of observed esterase activity raised the possibility that despite the contrary structural evidence, these proteins may catalyse the typical decarboxylative Claisen condensation reactions (Figure 5.15).

In this scenario, the active site cysteine (C116) of MupB would first undergo acylation and in the second, C-C bond forming step, the malonyl extender unit would be delivered to the acylated cysteine residue of MupB by an appropriate ACP (MmpE_ACP is shown as one possibility).
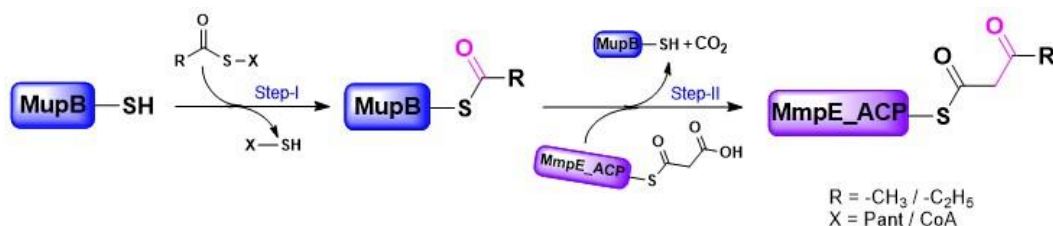
Figure 5.15. Proposed decarboxylative Claisen condensation reaction scheme of MupB. The formation of C–C bond is shown in magenta colour.

To investigate the Claisen condensation reaction, MmpE_ACP and TmpE_ACP from the mupirocin and thiomarinol biosynthesis pathways, respectively were chosen to be upgraded with a malonyl extender unit. Following the chemoenzymatic one-pot upgrading reaction conditions described in experimental section 7.7, malonyl-M/TmpE_ACPs (M-ACPs) were prepared at room temperature. Successful generation of M-ACPs were confirmed by ESI-MS after ~ 3 hours (Figure 5.16). Both M-ACPs slowly hydrolysed 30-40% of malonyl group after 24 hours at room temperature, but both were stable at – 80 °C in tris buffer (pH 7.5) for up to 3 months.
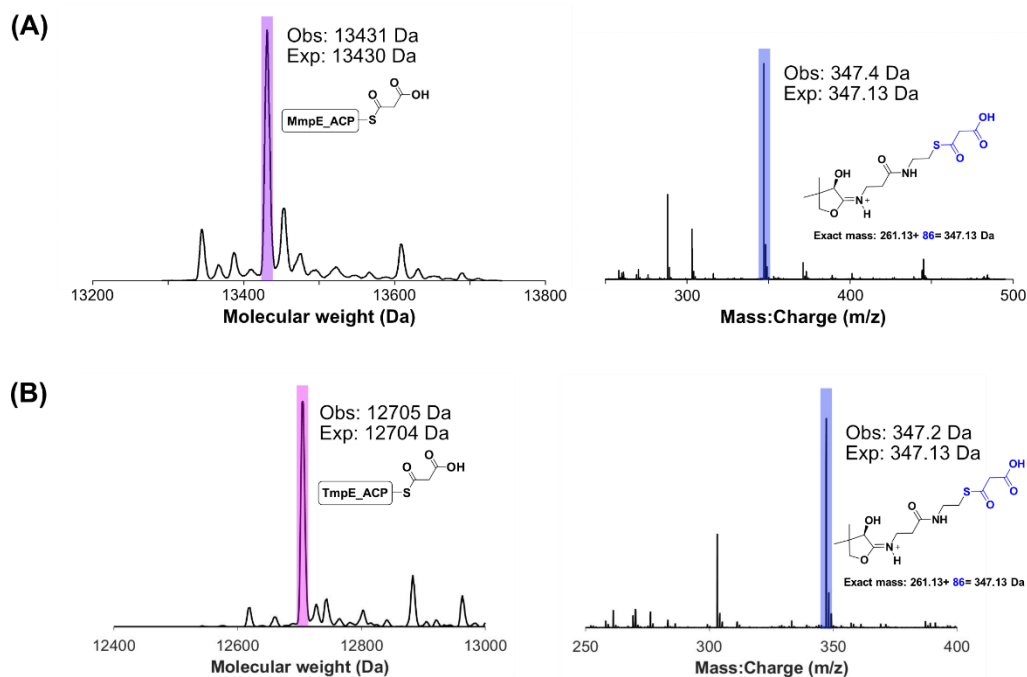
Figure 5.16. *In vitro* **enzymatic generation of M-ACPs:** The deconvoluted ESI-MS (right) and respective Ppant ejection (left side) spectra of (A)M-MmpE_ACP and subsequent Ppant ejection producing the malonyl-pantetheine fragment of observed mass 347.4 Da (expected mass 347.13 Da) (left) (B) M-TmpE_ACP spectra along with Ppant ejection of observed mass 347.2 Da.

50 μM MupB was primed with 1.0 mM of Ac/Pr-Pant for ~ 2 hours in buffer (50 mM Tris, 150 mM NaCl, pH = 7.5) at room temperature (Figure 5.17 A). As observed previously, only partial conversion could be achieved. M-MmpE_ACP (25 μM) was then added to Ac/Pr-MupB (25 μM) and incubated at room temperature for ~3 hours.
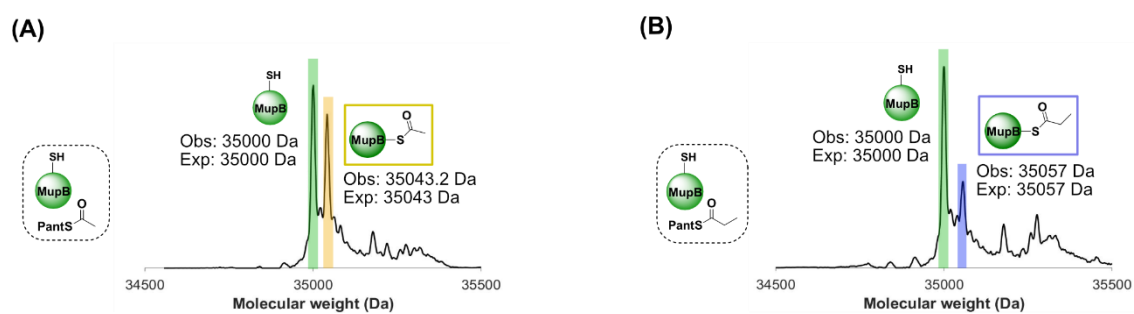
Figure 5.17. *In vitro* **translocation of Ac/Pr-substrate into MupB:** (A) Ac-Pant incubation with WT-MupB producing a spectrum of Ac-MupB (yellow) along with WT-MupB (green). (B) Pr-Pant incubated with WT-MupB and producing Pr-MupB (blue).

After 3 hours, ESI-MS analysis revealed no MmpE_ACP 4/5-carbon-β-keto thioester (for Ac and Pr, respectively) was generated (Figure 5.17 A). However, two new species of observed molecular mass 48344 Da and 48386.2 Da were observed. The formation of molecular mass 48344 Da was reported earlier in this work and indicates the generation of a cross-linked MupB:holo-MmpE_ACP complex (expected mass 48344 Da). Along with this, the observed peak of molecular mass 48386.2 Da is consistent with an increase of 42.2 Da from the MupB:holo-MmpE_ACP complex which, interestingly, may be indicating the generation of an acetyl loaded MupB:holo-MmpE_ACP complex. The equivalent reaction was observed during ESI-MS analysis after Pr-MupB incubation with M-MmpE_ACP for 3 hours at room temperature. Figure 5.18 represents the MupB:holo-MmpE_ACP complexe of observed mass 48344 Da along with the Ac/Pr loaded MupB:holo-MmpE_ACP complexes of observed masses 48386.5 Da and 48400.8 Da (expected mass 483401 Da).
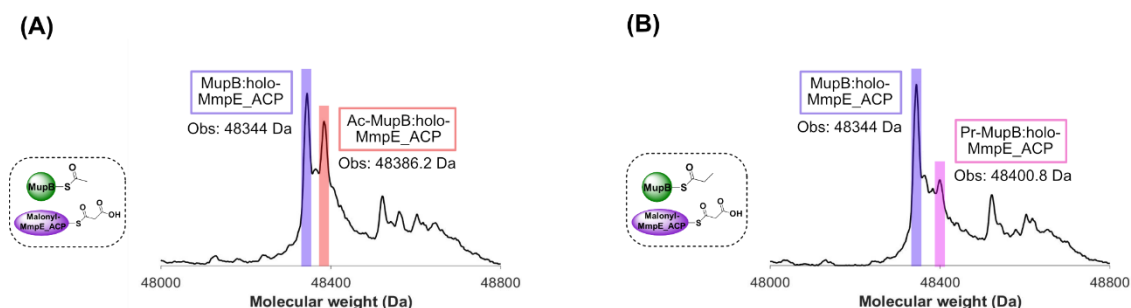
Figure 5.18. Claisen condensation assay of MupB in the presence of M-MmpE_ACP. (A) Ac-MupB incubation with M-MmpE_ACP produces the spectrum of MupB:holo-MmpE_ACP and Ac-MupB:holo-MmpE_ACP and (B) the spectrum showing the formation of MupB:holo-MmpE_ACP and Pr-MupB:holo-MmpE_ACP.

The Ac/Pr-MupB:holo-MmpE_ACP complex formation was further investigated by using a C116A mutant of MupB. The acetyl substrate was first translocated onto MupB:C116A using the S224 surrogate nucleophile (Figure 5.19) and then incubated with M-MmpE_ACPas described previously. In this case, no hydrolysis of M-MmpE_ACP was observed to produce holo-MmpE_ACP hence no crosslinked complex was formed as well (Figure 5.19). Therefore, it can be deduced that in Ac/Pr-MupB:holo-MmpE_ACP two possible events might happen (i) Ac/Pr self-loads into the surrogate active site (S224) of MupB and (ii) holo-MmpE_ACP forms a disulphide bond with the active site C116 of MupB to generate the cross-linked complex.
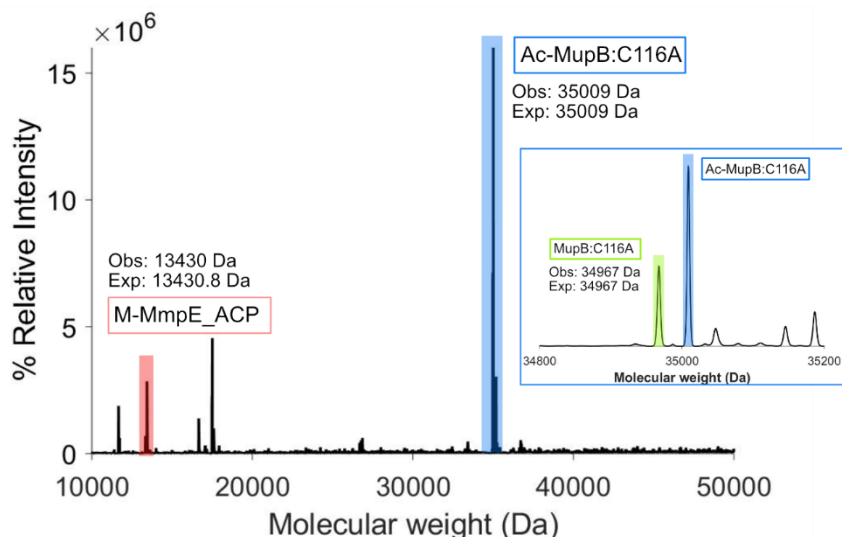
Figure 5.19. Effect MupB:C116A on Ac-MupB:holo-MmpE_ACP complex formation: Incubation of Ac-MupB:C116A with M-MmpE_ACP produced spectra of intact M-MmpE_ACP and Ac-MupB:C117A showing no crosslinked complex generation. The observed and expected molecular masses of the species are provided in the figure.

Decarboxylative Claisen condensation assay could not be examined for TmlB, because TmlB did not self-acylate with acetyl or propionyl substrates (Section 3.1.2). However, incubating TmlB with M-TmpE_ACP produced the crosslinked TmlB:holo-TmpE_ACP complex which was consistent with previous results (Section 4.6).

5.8 Summary

The work performed in this chapter has focused on investigating the functional mechanism of MupB and its thiomarinol homolog, TmlB. The proposed esterification and Claisen condensation reactions were analysed for both enzymes. NMR titrations of MmpB_ACP5 and 7 and TmpB_ACP3 with MupB and TmlB, respectively revealed that both MmpB_ACP5 and 7 showed interactions with MupB. Thes single TmpB_ACP3 domain from thiomarinol pathway showed no interaction with TmlB.

The ESI-MS experiments of esterification reactions showed that all three MmpB_ACPs can interact with MupB. But 9HN_ACP6 was found to produce an unusual product of molecular mass 27230 Da (2×9HN_ACP6). Formation of a similar product was also seen in the presence of the MupB homolog, TmlB. This leads to a future investigation where the esterification reaction will be carried out in the presence of MupB:C116A mutant protein hence can also elaborate on the effect of the active site cysteine for catalysing C-O bond

formation. None of these esterification assays produced any esterified product suggesting the lack of appropriate substrate which can bind more efficiently into the active site cavity in the proper orientation tofacilitate the esterification with the 9HN fatty acid delivered by MmpB_ACP567.

The typical Claisen condensation investigations confirm that both MupB and TmlB do not catalyse C-C bond formation. Both enzymes were observed to hydrolysethe malonyl unit to produce holo-M/TmpE_ACP which then crosslinked with MupB/TmlB, indicating that these ACPs are functionally related with MupB and TmlB. In addition, the acylated-MupB:holo-MmpE_ACP cross-linked complex is shown to be catalysed by the active site cysteine which is confirmed by the loss of formation of this complex reaction in the presence of mutant MupB:C116A.

6. Conclusions and future works

6.1 Conclusions

The structural and functional investigations presented in this thesis provide valuable information regarding the key esterification step in the mupirocin and thiomarinol biosynthesis pathways. A variety of multidisciplinary chemical and biological techniques have probed the activity of the reverse esterase enzymes, MupB and TmlB and their interactions with potential ACP partners. The MupB and TmlB enzymes and a range of ACPs from the respective biosynthetic pathways were cloned, expressed, and purified for use in functional *in vitro* assays using surrogate substrate mimics as well as interaction studies by NMR. Along with these, X-ray crystallography has also been successfully applied, yielding atomic resolution structures of MupB, TmlB and derivatised MBE-TmlB.

Analysis of the high-resolution crystal structures revealed the non-canonical active site arrangements formed by Cys-His-Asp catalytic triads in both MupB and TmlB that are homologous to similar KAS-III like enzymes that catalyse esterification reactions (Table 2.9). Two potential substrate binding channels and a pantetheine binding cavity were identified. One of the substrate channels will be required to bind and stabilize the pre-monic acid substrate while the other alkyl channel will allow the 9HN fatty acid to gain excess to the active site for catalysing the esterification reaction. Both these substrates will require the pantetheine entrance channel to deliver the substrate to the active site cavity (Figure 2.19). Hence, this speculation led to further investigations on substrate loading, substrate selectivity and real-time substrate-enzyme crystal structure determination.

Following this, *in vitro* self-acylation reactions using non-ACP bound substrates along with truncated pre-monic acid mimic (MBE) were conducted for MupB and TmlB and monitored by ESI-MS. The results showed that the thiomarinol enzyme, TmlB was highly selective towards the branched MBE substrate, whereas MupB was non-specific towards branched and non-branched substrates, but loading was generally low. Site directed mutagenesis was used to confirm the role of the active site cysteine in the self-acylation reaction. MupB:C116A was successfully cloned, expressed, and purified whereas TmlB:C117A was expressed as insoluble protein and was not pursued further. Interestingly, MupB:C116A self-loaded non branched acyl substrates and coenzyme A to a greater degree than the WT-

MupB which suggested the presence of an alternative nucleophilic residue. S224, near the active site triad, was identified as this potential surrogate. Hence, a double mutant of MupB enzymes would be an obvious choice for future work to understand the substrate binding nature of MupB. In addition to this work, a high resolution MBE-TmlB co-crystal structure confirmed that the active site cysteine was covalently bound with the MBE substrate and was found to pack into the presumed alkyl channel B (Figure 3.12). The orientation into this channel however was not expected based on homology to other systems and it was expected it should occupy channel A. It was presumed that this was due to its highly truncated nature and a lack of polar groups that would otherwise disfavour binding in this channel (which should hold the fatty acid). Infact, it is speculated that in this orientation the MBE group may act as an inhibitor which may account for the lack of observed activity in esterification assays. The complexity of the binding site and the presence of multiple substrate channels may confer a very strict substrate selectivity for these enzymes.

^{1}H-^{15}N-HSQC NMR titration of MupB and TmlB with several free standing and modular ACPs from mupirocin and thiomarinol biosynthesis pathways, respectively confirmed that only MmpE_ACP and MmpB/TmpB derived tandem ACPs interact with MupB/TmlB. A TmlB:TmpE_ACP NMR titration was not possible due to lack of soluble ^{15}N-TmpE_ACP, though it was judged that TmpE_ACP fulfils a similar role as it shares high sequence homology to its mupirocin counterpart. These results were further validated using ESI-MS to monitor the acyltransferase activity of MBE-MmpE_ACP and MBE-TmpE_ACP. In the presence of MupB/TmlB, the acyl group was observed to be efficiently hydrolysed off the ACP. Pleasingly for MBE-MmpE_ACP, the MBE substrate was observed to be transferred to MupB, albeit inefficiently compared to the rate of hydrolysis. The interesting observation was made that during the hydrolysis, a proportion of covalently linked complex between the M/TmpE_ACPs and MupB/TmlB weret generated, potentially opening a new route for studying these complexes and orientation of the phosphopantetheine chain. These findings also provide a mechanism by which the pre-monic acid is removed from the modular PKS, with MupB essentially behaving as a downstream KS.

Finally, esterification assays between 9HN mediated M/TmpB_ACPs and MBE-MupB/TmlB were conducted in this work. The ESI-MS analysis of these assays confirms that all three individual ACPs of MmpB and TmpB_ACP12 can interact with MupB and TmlB, respectively as hydrolysis of fatty acid side chains was observed only in the presence

of MupB or TmlB. However, no esterified products were observed. This was further confirmed by complementary NMR titration experiments of MupB with ^{15}N-MmpB_ACP5 and 7. MupB/TmlB did not catalyse Claisen condensation reactions which were performed to exclude this catalytic function but again hydrolysis of derivatised components was observed in this assay. The consistent hydrolytic action of MupB/TmlB strongly suggests that an incorrect substrate cannot be accommodated in the active site cavity correctly and hydrolysis occurs readily if the enzyme cannot turnover.

However, the overall conclusion regarding the key esterification reaction is that MupB and TmlB remain strong candidates for catalysing the esterification and the cognate ACPs that deliver substrates have been identified. The MBE substrate may inhibit the active site by co-ordinating in an incorrect orientation and there may be a strict requirement for the correct polyketide substrate to effectively 'prime' MupB or TmlB.

6.2 Future work

In future, producing double mutant (S224A mutant of MupB:C116A) MupB and using this variant for investigating the role of S224 residue on substrate binding into MupB would be of obvious interest. Complementary to this, co-crystal and/or crystal-soaked structures of MupB mutants with propionyl and coenzyme A substrates will further disclose the effects of C116 and S224 residues on substrate binding nature as well as structural insights into substrate binding pocket. Particularly interesting would be obtaining the correct substrate which may bind in the correct spatial orientation and prime the esterification reaction in presence of fatty acid chain. Therefore, the natural pre-monic acid or similar polar substrates mimics (larger than MBE) should be used for self-acylation assay and crystallisation which will offer greater substrate binding efficiency along with the substrate binding pocket architecture. Although this current research work confirmed that MupB/TmlB could only recognize the modular M/TmpE_ACPs and M/TmpB_ACPs from their biosynthesis pathway however, they need to be structurally characterised. This work showed that cross-linked complex formation with MupB was induced efficiently by using MBE probed MmpE_ACP, and hence could be used to trap active site cysteine of MupB for generating sufficient amount of covalently linked MupB:holo-MmpE_ACP for crystallisation. The successful crystallisation will provide the very first KAS-III:holo-ACP three dimensional structure which will provide more information to understand the elusive mechanism of ACP recognition by KAS-III type enzymes.

7. Experimental

The molecular biology protocolss described in this part, are not unique to this report (see Nahida Akter 1st and 2nd Year APM report) which is standard practice for theses in molecular biology. Therefore, this should be considered while analysing the Turnitin report.

7.1 Materials and methods

Sources of the materials

All reagents were purchased from Sigma Aldrich, Thermo Fisher Scientific and Merck Millipore. Isotopically enriched $^{15}NH_4Cl$ was purchased from Goss Scientific. Competent cells (T7 express and NEB-5α) were purchased from New England Biolabs (T7 express and NEB-5α) or Merck Millipore (Novagen BL21 (DE3)) which were used for plasmid and protein expression. The cofactors and other required plasmids for different assays were bought from Sigma Aldrich or Thermo Fisher Scientific. Kits used throughout the work on various purposes were PureLink PCR purification kit (Thermo Fisher Scientific), GenElute Plasmid Miniprep Kit (Sigma Aldrich), KAPA HiFi HotStart ReadyMix PCR kit (Roche), Q5 Mastermix (NEB)

Instruments

PCR for DNA cloning was performed with PeqSTAR 96x Universal gradient (VWR). Incubator, shaker and centrifuges (Sorvall RC-5C Plus and RC-6C plus) were used for transformation, protein expression and bacterial cell harvest. A VCX130 Vibra-cell sonicator was used to extract protein from the bacterial cell and a GE Healthcare Äkta FPLC (GE Healthcare) was used for protein purification from the soluble supernatant. DeNovix DS-11 UV-Visible spectrometer (Nano drop) was used to measure optical density (OD_{600}) of the bacterial growth media, as well as for estimation of DNA and protein concentrations. High performance liquid chromatography (Waters 2795HT HPLC) was also used for protein functional analysis. Crystallisation plates were set up for proteins using the Art-Robbins crystal Phoenix crystallisation robot.

Buffers and Solutions

The buffers and solutions recipes used throughout this work are listed below. The buffer pH was adjusted using 1M HCl or 1M NaOH where appropriate. The solutions and buffers were filtered and degassed prior to use in FPLC via a 0.2 µM cellulose nitrate filter (GE Healthcare).

Table 7.1. Reagents compositions of different buffers and solutions used in this work.

Buffer	Composition	Procedure
50× TAE	242g Trizma base, 57 mL 100% glacial acetic acid, 100 mL of 0.5 M EDTA up to 1L dH$_2$O	Agarose gel electrophoresis
Buffer A	50 mM Trizma Base, 0.5 M NaCl, 10% glycerol, pH 8.0	IMAC
Buffer B	50 mM Trizma Base, 0.5 M NaCl, 10% glycerol, 0.8 M Imidazole, pH 8.0	IMAC
Buffer C	50 mM Trizma base, 100 mM NaCl, 10% glycerol, pH 8.0	SEC
Buffer D	50 mM Na-phosphate (Na$_2$HPO$_4$+NaH$_2$PO$_4$, 100 mM NaCl, pH 8.0	SEC
Buffer E	25 mM Trizma base, 150 mM NaCl, pH 7.5	Analytical S75
Q-Buffer A	25 mM Trizma base, 50 mM NaCl, 10% glycerol, pH 7.5	Q FF column
Q-Buffer B	25 mM Trizma base, 1 M NaCl, 10% glycerol, pH 7.5	Q FF column
Anode buffer	200 mM Trizma base, pH 8.9	SDS-PAGE
Cathode buffer	100 mM Trizma base, 100 mM Tricine, 3.5 mM SDS, pH 8.25	SDS-PAGE
Gel buffer	3 M Trizma base, 10 mM SDS, pH 8.45	SDS-PAGE
Loading buffer	780 mM Trizma base, 142 mM β-mercaptoethanol, 35 mM SDS, 10% glycerol, 140 mM bromophenol blue, pH 6.8	SDS-PAGE
10% Separating solution	2.5 mL 40% acrylamide, 2.8 mL dH$_2$O, 3.33 mL Gel buffer, 50 µL 10% AMPS, 25 µL TEMED	SDS-PAGE
4% Stacking solution	0.5 mL 40% acrylamide, 3.33 mL dH$_2$O, 1.25 mL Gel buffer, 30 µL 10% AMPS, 15 µL TEMED	SDS-PAGE
Coomassie stain	2.5 g of Coomassie Blue (G250). 450 mL ethanol, 100 mL glacial acetic acid, 450 mL dH$_2$O	SDS-PAGE
Solution A	50:50 acetonitrile:dH$_2$O	ESI-MS
Solution B	0.1% formic acid (FA)	ESI-MS
Solution C	5% methanol, 0.1% formic acid	ESI-MS
Solution D	70% acetonitrile, 0.1% formic acid	ESI-MS
NaOH solution (0.5 M)	20 g NaOH in 1L dH$_2$O	IMAC column preparation
EDTA solution (0.25 M)	58.4 g EDTA in 1L dH$_2$O, pH 8.0	IMAC column preparation
NiSO$_4$ solution (100mM)	15.5 g in 1L dH$_2$O	IMAC column preparation
Ethanol solution (20%)	200 mL in 1L dH$_2$O	Column store

Media

Media was used for bacterial growth to express proteins or plasmids. All the media were sterilised by autoclaving at 121 °C, 151 psi for 15 minutes prior to use.

Table 7.2. Media and additives recipes.

Media	Ingredients
LB (Lysogeny broth)	10 g/L Tryptone, 5 g/L NaCl and 5 g/L yeast extract (5g/L), pH 7.4
M9 Minimal media	3 g/L KH$_2$PO$_4$, 1.8 g/L Na$_2$HPO$_4$, 0.5 g/L NaCl, pH 7.5
Additives for the media	
1000 x Trace Metal Elements	20 mM CaCl$_2$, 2 mM CoCl$_2$.6H$_2$O, 2 mM CuCl$_2$.2H$_2$O, 60 mM H$_3$BO$_3$.HCl, 10 mM MnCl$_2$.4H$_2$O, 2 mM Na$_2$MoO$_4$.2H$_2$O, 2 mM Na$_2$SeO$_3$.5H$_2$O, 2 mM NiCl$_2$.6H$_2$O, 2 mM ZnSO$_4$.7H$_2$O

Reagents

Reagent stock solutions were prepared in milli-Q water, and mainly used during protein expression, purification and assay reactions which are tabulated in Table 7.3. These reagents were filter sterilized by passing through a Sartorius Minit Cellulose acetate 0.2 µm syringe filter (Thermo Fisher) and stored at -20 °C. The required reagents were thawed, diluted where appropriate and stored at 4 °C prior to use.

Table 7.3. List of reagent compositions utilized in this work.

Reagent	Stock concentration	Composition
Carbenicillin	100mg/mL	1 g of carbenicillin in 10 mL dH$_2$O
Kanamycin	100 mg/mL	1 g of carbenicillin in 10 mL dH$_2$O
Ampicillin	100 mg/mL	1 g of carbenicillin in 10 mL dH$_2$O
IPTG	1M	238 mg of IPTG in 1 mL dH$_2$O
TCEP	1M	286 mg of TCEP in 1 mL dH$_2$O
DTT	1M	154 mg of DTT in 1 mL dH$_2$O
MgCl$_2$	1M	203 mg of MgCl$_2$ in 1 mL dH$_2$O
Coenzyme A	50 mM	38.3 mg of IPTG in 1 mL dH$_2$O
ATP	100 mM	50.7 mg of ATP in 1 mL dH$_2$O
Acetyl-CoA	50 mM	80.9 mg of acetyl-CoA in 1 mL dH$_2$O
Propionyl CoA	50 mM	82.3 mg of propionyl-CoA in 1 mL dH$_2$O
Butyryl-CoA	50 mM	83.76 mg of butyryl-CoA in 1 mL dH$_2$O
Malonyl- CoA	50 mM	85.35 mg of malonyl-CoA in 1 mL dH$_2$O
Succinyl-CoA	50 mM	86.76 mg of succinyl-CoA in 1 mL dH$_2$O

Sterile techniques

Where appropriate, all the glassware and media were sterilised by autoclaving at 121 °C, 151 psi for 15 minutes prior to bacterial growth. The sterile working procedures were maintained in a Microbiological safety cabinet (Medical Air Technology LTD BioMat Class II).

Preparation of Plasmids

PCR amplification: Polymerase Chain reaction (PCR), a technique to replicate repeatedly a targeted segment of DNA was used. Template genomic DNA containing DNA encoding the mupirocin and thiomarinol biosynthetic gene clusters were isolated from *Pseudomonas florescence* NCMBI 10586 and *Pseudoalteromonas* sp. SANK 73390 bacterial colonies provided by Dr Zhongshu Song (University of Bristol).

Genomic template DNA was prepared by suspending one colony into 200 µL of sterile water, mixed by pipetting and boiled for 10 minutes, before centrifugation (6000 rpm, 10 mins). Supernatant containing the genomic DNA was stored at -20 °C prior to use. Primers used in this work were designed and synthesized by IDT, UK. All the primers used in this work are listed in supplementary Table S.9.1. Adapter (normal type) incorporation into the forward and reverse primers (bold type) allow recombination into a pre-linearized pOPINF or pOPINE vector (Table S.9.1). Primer stock solutions (100 µM) were prepared using sterile distilled water and further diluted to a working concentration of 10 µM. Generally, the KAPA HiFi HotStart ReadyMix PCR Kit was used for the amplification during DNA cloning. The reagents for a single PCR reaction using KAPA ReadyMix PCR Kit are shown in Table 7.4. The PCR reaction mixture (10 µL) was transferred to 5 PCR tubes, and the PCR protocol was performed as described in table 7.5. Gradient temperature was used to optimize the appropriate annealing temperature for each primer pair in the PCR reaction.

Table 7.4. KAPA ReadyMix PCR master mix (50 µL) for a single PCR reaction.

Component	Volume /µL (in 50 µL)	Final Concentration
Sterile dH$_2$O	21	N/A
2X KAPA HiFi HotStart ReadyMix	25	1X
10 µM forward Primer	1.5	0.3 µM
10 µM Reverse Primer	1.5	0.3 µM
Template DNA (100 ng/µl)	1.0	2.0 ng/µl

Table 7.5. KAPA HiFi ReadyMix PCR Kit program details

Step	Temperature (°C)	Duration	Cycles
Initial denaturation	95	3 min	1
Denaturation	98	20 sec	
Annealing*	50-70	15 sec	15-35
Extension	72	15-60 sec/kb	
Final extension	72	1 min/kb	1
Store at	4	until further use	

*The optimal Annealing temperature for a specific primer set is likely to be different and was calculated according to the melting temperature (T_m °C) of the primers.

Agarose gel electrophoresis

PCR products were analysed by agarose gel electrophoresis using a 1% (w/v) agarose gel prepared in 1 x TAE buffer with the addition of Midori green DNA Stain (2-4 μL).[149] The gel was placed in an electrophoresis tank and immersed in 1 x TAE Buffer. A mixture of PCR product sample and 5x DNA Loading Buffer Blue (4:1, 5 μL, Bioline) was loaded in each well of the gel, and HyperLadder 1 kb (4 μL, 43 45 Bioline) was added as a marker. Electrophoresis was performed at 200 V for 20 minutes and DNA bands were detected on the resulting gel via a UV-transilluminator (Essential, UVITEQ Cambridge). Upon confirmation of the amplified DNA band of interest, samples were further purified using a Pure link PCR Purification kit according to the manufacturer's protocol.

Pre-linearizing pOPINF and pOPINE vector

To cut and linearise the vector, also called pre-linearised vector ready for recombination, 15 μL of the vector was incubated with 2 μL of 10× Fast Digest buffer, 10 units/μL KpnI and 20 units/μL HindIII (1μL) at 37 °C for 3 hours. Then the plasmid was purified using a gel extraction kit following the manufacturer's protocol and stored at -20 °C prior to use.

Plasmid ligation

The purified PCR fragments were cloned into a pre-linearized vector (pOPINE/pOPINF) using In-fusion HD Cloning Plus Kits (Sigma Aldrich). The PCR product and vector were mixed in a 1:3 standard ratio, respectively with the In-fusion enzyme. The ligation reactions were usually performed in a 10 μL volume reaction mixture following the conditions given in table 7.6.

Table 7.6. The reagents for the ligation of plasmid.

Reagents	Volume added to make 10 µL
Sterile water	5 µL
PCR product	3 µL
Pre-linearised pOPINF vector	1 µL
5× In-fusion premix	1 µL

Successful plasmid ligation was screened by transforming the ligated 2 µL of plasmid into 50 µL NEB-5α *E. coli* cells and plated on LB agar plates containing carbenicillin (100 µg/mL), X-gal (40 µg/mL) and IPTG (1 mM). The successful ligation was confirmed by the appearance of white colonies, whereas blue colonies suggested unsuccessful cloning. Isolated white colonies were resuspended into 5mL LB with 100 µg/mL carbenicillin, incubated at 37 °C overnight prior to isolation of DNA using the general protocol from thet GenElute Plasmid Miniprep Kit. The sequence of the purified DNA was analysed by GeneWiz^{TM}. The alignment of sequences was carried out by the T-Coffee Multiple sequence alignment server. All the plasmids used throughout this work are listed in supplementary Table S.9.2.

Preparation of MupB and TmlB mutant plasmids

The whole plasmid mutagenesis procedure was used to generate MupB:C116A and TmlB:C117A mutant plasmids. Following this procedure, WT-MupB and WT-TmlB were used as the template DNA and the mutagenic primers were designed from the parent WT DNA. The general PCR amplification procedure (outline above, Table 7.3) was adopted with the addition of 5% DMSO to the PCR reaction mixture using the mutagenic primers. Five 200 µL PCR tubes each containing 10 µL of the reaction mixture were heated for three minutes at 95 °C before going through 30 cycles of denaturation (98 °C, 20 seconds), annealing (temperature gradient: 64 $\pm$ 5 °C, 15 seconds), and extension (72 °C for 6.5 minutes). Following this sequence, a final extension of 10 minutes at 72 °C was used, and the sample was then stored at 4 °C. Agarose gel electrophoresis was used to confirm the success of the amplification. The amplified PCR product was digested with DpnI overnight (50 µL product, 5 µL DpnI, and 5 µL of 10×Tango Buffer) to remove the parental DNA before being purified using a PCR Purification kit. 4 µL of plasmid was transformed into 50 µL of NEB-5a cells and plated on agar plates (containing 100 µg/mL carbenicillin) following the general cell transformation protocol described in section 7.2. Plasmids were

then extracted for specific colonies and sent for sequencing which confirmed the successful point mutations, The mutant DNA sequences are listed in supplementary Table S.9.1.

7.2 Protein expression

Antibiotic agar plates

Agar powder (1.5 g) was added into 100 mL of LB media and heated gently to melt the agar powder before being autoclaved. Then the required antibiotic was added into the liquid LB-Agar media at a 1:1000 antibiotic to media ratio. This was then distributed evenly among 4-5 petri dishes and left to solidify in a sterile environment.

Cell transformation

The LB-agar plate was used to transform the bacterial cells containing the required plasmid DNA. 1 µL of the appropriate plasmid was added to 20 µL of *E. coli* cells (Novagen BL21 DE3/ T7 Express cells) and the mixture incubated on ice for 30 minutes, before being heat shocked at 42 °C for 30 seconds. Then 200 µL of SOC media was added to the mixture and incubated at 37 °C for 1 hour with shaking at 220 rpm. The cells were spread evenly on the LB-Agar plate supplemented with 100 µg/mL carbenicillin/kanamycin as appropriate and then left for ~ 16 hours at 37 °C.

General procedure of protein expression in LB

Overnight Pre-culture/Seed stock: A single isolated colony was picked and resuspended into 50 mL of LB media containing antibiotic (1:1000 dilution of 1 mg/mL stock, 50 µL) and left overnight shaking at 200 rpm at 37 °C. 2 mL of the seed broth was used to inoculate flasks containing t200 mL of LB supplemented with the appropriate antibiotic. Flasks were then incubated at 37 °C (at 200 rpm) until the OD_{600} (Optical Density) reached in an optimum (0.7-1.0) value. Cells were induced with 250 µM IPTG (isopropyl-β-D-thiogalactopyranoside) alongside additional antibiotic (100 µg/mL) and incubated at 16 °C at 200 rpm shaking overnight. The cell pellet was harvested by centrifugation (7000 RPM, 10 minutes, 5 °C) and suspended in resuspension buffer (20 mL). The cell pellet was flash frozen in liquid N_2 and stored at -20 °C until required.

To extract proteins, the cell pellets were first thawed on ice and lysed by sonication (VCX130 Vibra-Cell, Sonics) for 10-20 mins, (pulse on (5 s), pulse off (10 s)) until a viscous

and uniform liquid was observed. This suspension was centrifuged (16,000 RPM, 20 minutes, 5 °C) to separate the soluble (supernatant) and insoluble (pellet) components for further purification.

Isotopically labelled protein expression

1. **M9 expression**

M9 minimal media was used to express the isotopically ^{15}N-labelled protein. A single colony was first suspended in 10 mL of LB media supplemented with carbenicillin (100 µg/mL) and cultured at 37 °C shaking at 200 rpm for 5 h. The 10 mL pre-culture in LB was then transferred to 100 mL of M9 media supplemented with carbenicillin (100 µg/mL), NH$_4$Cl (1 g/L), glucose (2 g/L), 1 mM MgSO$_4$ and trace metal elements (100 µL) and incubated at 37 °C for 16 h to prepare the seed broth. 2 mL of seed broth was used to inoculated flasks containing 100 mL of M9 media containing the supplements stated above along with ^{15}N-NH$_4$Cl (1 g/L) and incubated at 37 °C (200 rpm) until the optical density (OD$_{600}$) of 1.0 was reached. The growing culture was then induced with 250 µM IPTG and incubated at 16 °C for 16 h prior to harvesting the cells. The protein expressed using this protocol was extracted and purified using the same pellets as for the unlabelled protein purification.

2. **M9 expression via media exchange**

The seed broth (pre-culture) was prepared as per the protocol stated in the general expression in LB media. 2 mL of the pre-culture was added to 200 mL of LB supplemented with carbenicillin and incubated at 37 °C (200 rpm) until the optical density at 600 nm (OD$_{600}$) was reached between 0.7-1.0. Then the bacterial cells were harvested by centrifugation (6000 rpm for 10 mins) and washed twice with 400 ml M9 media to remove the traces of LB media. Then the cells were resuspended in a minimum amount (20-30 ml) of M9 media and 2 mL of media was used to inoculate 100 mL M9 media containing carbenicillin (100 µg/mL), ^{15}N-NH$_4$Cl (1 g/L), glucose (2 g/L), 1 mM MgSO$_4$ and trace metal elements (100 µL) and incubated at 16 °C for 1 h prior to induction with 250 µM IPTG. The media was then incubated overnight at 16 °C (200 rpm) prior to harvesting the cells.

Arctic cell transformation and expression

Arctic Expression (DE3) cells were used to transform into and express MacpB protein. Firstly, an Eppendorf containing 100 µl cells was collected from the -80 °C freezer and

thawed on ice. 2 µL of diluted XL10-Gold β-mercaptoethanol was added to the cells and left on ice for 10 minutes prior to adding MacpB plasmid (4 µL) and incubated on ice for 30 minutes. The cells were then heat-shocked at 42 °C for 20 seconds before resting on ice for 5 minutes. 900 µL of SOC media was added to the cells under sterile conditions and left tostir (200 RPM) at 37 °C for one hour. 150 µL of cells were then transformed onto a LB agar plate containing carbenicillin antibiotic and left to incubate aat 37 °C for 16 hours until colonies were observed.

To pre-culture each isolated colony was added to 10 mL sterile LB media (supplemented with carbenicillin and gentamicin at a 1:1000 ratio) and incubated at 37 °C for 24 hours shaking at 200 RPM. 24 aliquots of pre-culture solution (10 mL) was added to 100 mL sterile LB (non-baffled flasks – no antibiotic added) and incubated with shaking at 30 °C for 3 hours. The temperature was then reduced to 10 °C prior to inducing each flask with 0.5 mM IPTG left andshaken at 200 RPM for a further 24 hours. The cells were then harvested by centrifugation (7000 RPM, 10 minutes, 5 °C) and resuspended in resuspension buffer (20 mL), frozen in liquid N_2 and stored at -20 °C until required for protein purification.

7.3 Protein purification

Immobilized Metal Ion Affinity Column (IMAC)

As all the proteins were His6-tagged an affinity column was used in this work to isolate the protein from the soluble supernatant via IMAC using a 5 ml HiTrap Affinity Column (GE Healthcare) equilibrated with 100 mM $NiSO_4$ and washed with 5 mL HPLC grade water to remove any trace amount of Ni solution that was unbound. This column was connected to the FPLC (Fast Performance Liquid Chromatography, GE Healthcare) and equilibrated with 5 CV (column volumes) Column Buffer A and B with a ratio of 95:5. The Super-loop was washed with NaOH solution and HPLC grade water prior to load protein supernatant (extracted after sonication and centrifugation) into the super-loop. The supernatant was injected into the column and non-binding proteins were washed away with 5% B buffer (5 CV). His-tagged protein was selectively eluted using an increasing imidazole concentration (6% - 100%) and collected in 1 mL fractions and immediately placed on ice.

Size exclusion chromatography (SEC)

Size-Exclusion Chromatography was performed using either a HiPrep 26/10 desalting column (50 mL) or a S75 gel filtration column (280 mL). The combined fractions from the IMAC separation were loaded into the super-loop and injected onto the column. Fractions were collected when the absorbance was observed to increase. Collection was stopped prior to elution of small molecules and salt, indicated by a peak in the conductivity.

7.4 Protein characterization

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

At different steps of protein expression and purification, samples were collected for analysis. A small amount of sample (10 µL) was mixed with gel loading buffer as the required amount (10-50 µL) and boiled (95 °C, 10 minutes) to denature the protein prior to electrophoresis. The electrophoresis gel was prepared in the laboratory following the protocol composed of SDS-PAGE (10% Separating, 5mL and 4% stacking, 2.5 mL). The gel was made up using the materials as described in table 7.7.

Table 7.7. List of reagents to prepare SDS-PAGE gel.

Materials	4% Stacking gel	10% Separating gel
40% Acrylamide	1 mL	2.5 mL
Gel buffer	2.5 mL	3.33 mL
Glycerol	-	1.3 mL
Distilled Water	3.25 mL	2.8 mL
10% AMPS (Freshly made)	50 µL	50 µL
TEMED	20 µL	10 µL

The gel was then placed in a tank containing anode buffer, and the gel tank itself contained cathode buffer. 5 µL sample was added to each separate gel well along with the SDS-marker (10 µL). The electrophoresis was then operated between 110-130V, for 50-70 minutes until the stained front had reached to the bottom of the gel. Following separation, the gel was stained using Coomassie Blue Stain for one hour and subsequently de-stained by heating while submerged in water. Protein bands were detected using a transilluminator, E-gel imager (Essential, UVITEQ Cambridge).

Electrospray Ionisation Protein Mass Spectrometry

Protein mass spectrometry (MS) was used to determine the accurate mass and characterization of protein. In this work, Electrospray Ionisation (ESI) mass spectrometer (SYNAPT G2-Si, Waters) was used.

1. **Zip-Tip sample preparation**

To prepare samples for ESI-MS analysis C4 - ZipTip^{TM} (Merck Millipore) were used. First, the Zip-Tip was activated with two 10 µL aspirations with 50% acetonitrile (reagent A), followed by equilibration (five 10 µL aspirations) with reagent B (0.1% formic acid in H$_2$O). The protein sample was loaded onto the C4 resin by 10 µL aspirations (10-20 times) into the sample, returning the sample each time into the vial. The protein sample was then desalted by aspirating 10-15 times with reagent C (5% MeOH, 0.1% formic acid in H$_2$O). The final sample was then eluted with 10 µL of reagent D (70% acetonitrile, 0.1% formic acid in H$_2$O).

2. **C4 resin protein extraction**

In some cases, when enough protein samples were available, particulate C4 resin (Jupiter 15 μ C4 300 Å Phenomenex) was used to prepare protein samples for ESI-MS analysis. This sample preparation is cost effective and provides good samples for ESI-MS analysis.

To prepare the C4 resin for protein sample loading, 5 mg of C4 resin was first resuspended in 500 µL of methanol (HPLC grade) and vortexed. The protein was then loaded onto the resin by adding 25 µL of protein sample to 25 µL of C4 resin, vortexing (for 5 min), centrifuging (at 13K rpm for 5 min), and finally discarding the supernatant. Using 0.1% formic acid (FA) in water, the resin was washed before being vortexed for 5 minutes and centrifuged as described previously. The washing step was repeatedthere. The protein sample was ultimately recovered from the resin by resuspending the resins in 10-15 µL of acetonitrile (supplemented with 0.1% FA), vortexing, and centrifuging before collecting the supernatant. Before analysing the sample by ESI-MS analysis, the centrifugation step was repeated two to three times to remove all resin particles.

The ESI-MS data for desalted proteins were acquired in positive ion mode over a mass range of 300:3000. The acquired charge states were deconvoluted using Magtran (1.03) software. The MSMS functionality was used for the Ppant ejection assays on an appropriate charge state with the addition of 0-10 V Trap CE and 0-20 V Transfer CE.

Biophysical characterisation

Analytical Size Exclusion Chromatography

Analytical SEC is normally used to estimate the molecular weight of an analyte by creating a calibration curve using known molecular weight standards. In this work a Superdex 75 10/300 GL pre-packed column (GE Healthcare) was equilibrated with 25 mM Tris, 150 mM NaCl, pH 7.5 (Buffer C) at 0.5 mL/min flow rate prior to use. The column was pre-calibrated with 1 mg/mL blue dextran (2000 kDa), bovine serum albumin dimer (130 kDa), conalbumin (75 kDa), bovine serum albumin monomer (65 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa) in buffer C and a standard calibration curve was generated by plotting analyte particle distribution coefficient (Kav) against the logarithm of the molecular weight (Log MW) and fitting the data with a linear polynomial line (y = -0.3427x + 0.7143, R^2 = 0.999) as shown in Equation 1.

$$\log MW = m \times K_{av} + C \qquad (1)$$

Here m = slope and C = intercept. The distribution coefficient, Kav can be calculated using equation 2 as described previously.[150]

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \qquad (2)$$

Where, V_e = elution volume, V_o = column void volume = 7.0 mL, V_t = geometric column volume = 26 mL. The elution volume of an analyte can be used to estimate its apparent molecular weight using this method.

Protein quantification

The total amount to proteins in a sample was determined using the two following methods:

1. *BCA protein assay kit-reducing agent compatible*

The general principle of the BCA assay kit is based on the determination of protein concentration by forming the Cu^{2+} - protein complex under alkaline conditions using bicinchoninic acid (BCA) which is followed by reduction of the Cu^{2+} to Cu^{1+} and monitored by absorbance spectrometry.

All samples were prepared in working reagent (WR) consists of 50parts of reagent A (1 %

BCA-Na$_2$, 2 % Na$_2$CO$_3$, 0.16 % Na$_2$ tartrate, 0.4 % NaOH and 0.95 NaHCO$_3$) and 1 part of reagent B (4% CuSO$_4$.5H$_2$O), in a 50:1 mixture ratio. The blank sample did not contain any protein and was used to correct the baseline in the spectrometer. 25 µL of protein sample was mixed with 1 mL of WR and incubate at 37 °C prior to measuring the absorbance at 562 nm. A standard calibration curve of BSA protein over a range of concentrations (25-2000 µg/mL) was used to calculate the unknown protein concentrations. This method of concentration measurement was particularly useful for proteins that did not contain tryptophan residues and did not show high absorbance at 280 nm.

2. **Absorbance at 280 nm assay**

Proteins exhibit absorbance under UV light, 280 nm, which is used to quantify the protein concentration. A DeNovix DS-11 UV-Vis spectrophotometer was used to measure the absorbance which calculated the concentration automatically in mg/mL using the calculated extinction coefficient (ε) (calculated from amino acid sequence analysis using ProtParam) and the Beer-Lambert law as stated**:**

$$\textbf{Protein Concentration} = A_{280} / \varepsilon l \qquad [\textit{ l=cell path length}]$$

7.5 Protein crystallization and data analysis

The sitting drop vapor diffusion method was used for crystallization screening of the MupB and TmlB enzymes. Four 96 well MRC2 crystallization plates (each well contains two drops) were set up for each protein using the ArtRobbins Crystal Phoenix^{TM} crystallization robot which screened 384 conditions using the pre-made crystallization solutions which is commercially available by the name Molecular Dimension screening solution.. Each reservoir contains 50 µL crystallization solution and the screening drops of each well were set at a 1:1 and 1:3 ratio of solution and protein. The individual crystals were looped using appropriately sized litho-loops (Molecular Dimensions), flash frozen in liquid nitrogen and sent to the Diamond Light Source lab for analysis by x-ray diffraction using beamlines I03 and I04.

Co-crystallisation was achieved by incubating the MBE-pantetheine (5mM) substrate with TmlB (~ 310 µM) in ice for 3 hours and self-acylation was confirmed ESI-MS analysis. The crystallization screenings were performed as described above using four pre-made solutions (Morpheus, PACT primer, JCSG+ and Structure I+II).

The collected diffraction images were processed and integrated using XDS and Xia2 and further analysed by Phenix.[91] Phaser -MR[151] was used for molecular replacement (MR). Manual model building and refinement were undertaken using Coot[152] and Phenix. Final models were refined using Molprobity[95] and PDB_redo server[153] and then analysed by Pymol.[154]

7.6 Holo-ACP preparation

This work involved converting apo-M/TmpE ACP to holo-M/TmpE ACP. The apo-ACP was incubated with the reagents (Table 7.8), and the mixture was allowed to sit at room temperature for 1 hour. ESI-MS was used to monitor the progress of the reaction.

Table 7.8. Reagents that are required for apo to holo conversion of ACP.

Reagent	Stock concentration	Volume added to make 100 µL reaction mixture	Final concentration
apo-ACP	> 100 µM	As required	100 µM
MgCl2	1 M	1	10 mM
CoA	100 mM	1	1 mM
MupN	15 µM	10	2 µM
TCEP	100 mM	1	1 mM

7.7 One-pot pantetheine upgrade and loading ACPs

The one-pot pantetheine upgrade assay and loading to the ACPs protocols are adopted from the literature.[155,156] The apo-ACPs were mixed with the reagents in series up to 100 µL of total reaction volume (Table 7.9) and incubated at room temperature. The upgrade progress and loading onto ACPs was monitored by ESI-MS analysis typically after 1 h but some ACP conversions required longer times (3-16 h) to achieve maximum conversion. After ~ 80% loading, excess reagents were removed by using a Zeba desalting column (7K MWCO, Thermo Fisher) and equilibrated to the reaction buffer (50mM Tris, 5 mM NaCl, pH=7.5). The protein was then concentrated using Vivaspin 20 10 kDa spin concentrated prior to storage at -20 °C.

Table 7.9. Reagents required for the one-pot pantetheine upgrading reaction of ACP.

Reagent	Stock concentration	Volume added to make 100 (µL) mixture	Final concentration
ACP			100 µM
MgCl2	1 M	1	10 mM
Pantetheine substrate	100 mM	1	1 mM

CoA mix (A/D/E)	33 µM each	3	1 µM each
MupN (PPTase)	40 µM	25	10 µM
ATP	100 mM	1.5	1.5 mM

7.8 Analytical LCMS

The crude mixture of the esterification reaction was analysed by using analytical LCMS techniques. The sample was prepared by extracting the molecules in a 50:50 mixture of acetonitrile and water supplemented with 1% formic acid, drying with N_2 gas and finally by dissolving in HPLC grade MeOH. 20 µL of the sample was injected and analysed on a Waters 2795HT HPLC system where MS coupled by LC. A Phenomenex Kinetex column (100 Å, 2.6 µ, C18, 4.6 mm × 100 mm) was used for separation from the extracted sample. HPLC grade H_2O (A) and acetonitrile (B) supplemented with 0.05% formic acid were used as the eluents at a flow rate of 1.0 mL/min using the following solvent gradient (Table 7.10).

Table 7.10. Eluent gradient run used for separating molecules from the reaction mixture.

Time/ min	Eluent
0-2	5% B
2-15	The linear gradient to 90% B
15-17	95% B
17-18	The linear gradient to 5% B
18-20	5% B

7.9 Nuclear Magnetic Resonance (NMR) titration

[15]N-labelled full-length mupirocin and thiomarinol ACPs were prepared in phosphate buffer pH 8.0 and concentrated up to 0.1-2 mM. The ^{15}N-labelled ACPs were then incubated with MupB/TmlB at different concentration ratios (1:1, 1:3 and/or 1:3) supplemented with 10% D_2O and placed into a 3mm NMR tube in collaboration with Dr Ashley Winter (University of Bristol). All protein NMR titration experiments were carried out on a Varian VNMRS 600 MHz spectrometer or a Bruker AVANCE III HD 700 MHz spectrometer by Dr Chris Williams (University of Bristol). NMR titrations with MupB/TmlB were recorded at 298 K at pH 8.0. NMR data was processed with NMRPipe [157] and spectra analysed in CcpNMR Analysis version 2.1.5[158] by Dr Ashley Winter.

8. References

1   S. Sengupta, M. K. Chattopadhyay and H.-P. Grossart, *Front. Microbiol.*, 2013, **4**, 47.

2   M. Wink, *Phytochemistry*, 2003, **64**, 3–19.

3   G. Kapoor, S. Saigal and A. Elongavan, *J. Anaesthesiol. Clin. Pharmacol.*, 2017, **33**, 300–305.

4   M. A. Fischbach and C. T. Walsh, *Science.*, 2009, **325**, 1089–1093.

5   C. T. Walsh and T. A. Wencewicz, *J. Antibiot.*, 2014, **67**, 7–22.

6   J. O'Neill, *The Review On Antimicrobial ressistance. Tackling Drug-resistant Infections Globally:Final Report And Recommendations*, 2016.

7   A. Okano, N. A. Isley and D. L. Boger, *Proc. Natl. Acad. Sci.*, 2017, **114**, E5052–E5061.

8   F. Alberti, K. Khairudin, E. R. Venegas, J. A. Davies, P. M. Hayes, C. L. Willis, A. M. Bailey and G. D. Foster, *Nat. Commun.*, 2017, **8**, 1831–1841.

9   A. M. Bailey, F. Alberti, S. Kilaru, C. M. Collins, K. de Mattos-Shipley, A. J. Hartley, P. Hayes, A. Griffin, C. M. Lazarus, R. J. Cox, C. L. Willis, K. O'Dwyer, D. W. Spence and G. D. Foster, *Sci. Rep.*, 2016, **6**, 1–11.

10   D. Voet and J. G. Voet, *Biochemistry, 4th Edition | Wiley*, 2010.

11   F. Sato and K. Matsui, in *Plant Biotechnology and Agriculture: Prospects for the 21st Century*, Elsevier, 2011, pp. 443–461.

12   H. Yoneyama and R. Katsumata, *Biosci. Biochem.*, 2006, **70**, 1060–1075.

13   E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231–316.

14   J. Beld, D. J. Lee and M. D. Burkart, *Mol. Biosyst.*, 2015, **11**, 38–59.

15   J. J. Thelen and J. B. Ohlrogge, *Metab. Eng.*, 2002, **4**, 12–21.

16   L. B. Pickens and Y. Tang, *J. Biol. Chem.*, 2010, **285**, 27509–15.

17   A. Das and C. Khosla, *Acc. Chem. Res.*, 2009, **42**, 631–639.

18   C. Carvalho, R. X. Santos, S. Cardoso, S. Correia, P. J. Oliveira, M. S. Santos and P. I. Moreira, *Curr. Med. Chem.*, 2009, **16**, 3267–85.

19   S. R. Park, Y. J. Yoo, Y.-H. Ban and Y. J. Yoon, *J. Antibiot.*, 2010, **63**, 434–441.

20   N. Collie and W. S. Myers, *J. Chem. Soc., Trans.*, 1893, **63**, 122–128.

21   A. Birch, R. Massy-Westropp and C. Moye, *Aust. J. Chem.*, 1955, **8**, 539.

22   A. C. Mercer and M. D. Burkart, *Nat. Prod. Rep.*, 2007, **24**, 750.

23   S. Smith and S.-C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041.

24   L. Zhu, J. Cheng, B. Luo, S. Feng, J. Lin, S. Wang, J. E. Cronan and H. Wang, *BMC Microbiol.*, 2009, **9**, 1–11.

25   G. Gago, L. Diacovich, A. Arabolaza, S.-C. Tsai and H. Gramajo, *FEMS Microbiol. Rev.*, 2011, **35**, 475–497.

26   S. Smith, *FASEB J.*, 1994, **8**, 1248–1259.

27   S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science*, 1991, **252**, 675–9.

28   F. Schlu È nzen, R. Zarivach, J. È rg Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath and F. Ë ois Franceschi, *Nature*, 2001, **413**, 814–821.

29   T. Tenson, M. Lovmar and M. Ehrenberg, *J. Mol. Biol.*, 2003, **330**, 1005–1014.

30   B. Shen, *Curr. Opin. Chem. Biol.*, 2003, **7**, 285–295.

31   S. Smith, A. Witkowski and A. K. Joshi, *Prog. Lipid Res.*, 2003, **42**, 289–317.

32   T. Maier, S. Jenni and N. Ban, *Science.*, 2006, **311**, 1258–1262.

33   T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322.

34   E. Schweizer and J. Hofmann, *Microbiol. Mol. Biol. Rev.*, 2004, **68**, 501–517.

35   C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Nat. Prod. Rep.*, 2007, **24**, 162–90.

36   J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.

37   A. T. Keatinge-Clay, *Cell Chem. Biol.*, 2016, **23**, 540–542.

38   D. A. Herbst, C. A. Townsend and T. Maier, *Nat. Prod. Rep.*, 2018, **35**, 1046–1069.

39   H. Chen and L. Du, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 541–557.

40   A. J. Winter, M. T. Rowe, A. N. M. Weir, N. Akter, S. Z. Mbatha, P. D. Walker, C. Williams, Z. Song, P. R. Race, C. L. Willis and M. P. Crump, *Angew. Chemie*, 2022, **134**, e202212393 (1-11).

41   N. Traitcheva, H. Jenke-Kodama, J. He, E. Dittmann and C. Hertweck, *ChemBioChem*, 2007, **8**, 1841–1849.

42   J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.

43   A. Gallo, M. Ferrara and G. Perrone, *Toxins 2013, Vol. 5, Pages 717-742*, 2013, **5**, 717–742.

44   M. Singh, S. Chaudhary and D. Sareen, *J. Biosci.*, 2017, **42**, 175–187.

45   C. W. Carreras and C. Khosla, 1998, **37**, 2084–2088.

46   C. W. Carreras, R. Pieper and C. Khosla, *J. Am. Chem. Soc.*, 1996, **118**, 5158–5159.

47   J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111–1137.

48   J. Schröder, *Nat. Struct. Biol. 1999 68*, 1999, **6**, 714–716.

49   N. Funa, Y. Ohnishi, I. Fujli, M. Shibuya, Y. Ebizuka and S. Horinouchi, *Nature*, 1999, **400**, 897–899.

50   M. Hashimoto, T. Nonaka and I. Fujii, *Nat. Prod. Rep.*, 2014, **31**, 1306–1317.

51   B. Shen and H. J. Kwon, *Chem. Rec.*, 2002, **2**, 389–396.

52   H. J. Kwon, W. C. Smith, L. Xiang and B. Shen, *J. Am. Chem. Soc.*, 2001, **123**, 3385–3386.

53   K. Finzel, D. J. Lee and M. D. Burkart, *Chembiochem*, 2015, **16**, 528.

54   E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231–316.

55   A. C. Murphy, S. S. Gao, L. C. Han, S. Carobene, D. Fukuda, Z. Song, J. Hothersall, R. J.

Cox, J. Crosby, M. P. Crump, C. M. Thomas, C. L. Willis and T. J. Simpson, *Chem. Sci.*, 2014, **5**, 397–402.

56   C. M. Thomas, J. Hothersall, C. L. Willis and T. J. Simpson, *Nat. Rev. Microbiol.*, 2010, **8**, 281–289.

57   A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399–6408.

58   Y. Chen, E. E. Kelly, R. P. Masluk, C. L. Nelson, D. C. Cantu and P. J. Reilly, *Protein Sci.*, 2011, **20**, 1659–1667.

59   W. Huang, J. Jia, P. Edwards, K. Dehesh, G. Schneider and Y. Lindqvist, *EMBO J.*, 1998, **17**, 1183–91.

60   X. Qiu, C. A. Janson, A. K. Konstantinidis, S. Nwagwu, C. Silverman, W. W. Smith, S. Khandekar, J. Lonsdale and S. S. Abdel-Meguid, *J. Biol. Chem.*, 1999, **274**, 36465–36471.

61   H. Pan, S.-C. Tsai, E. S. Meadows, L. J. W. Miercke, A. T. Keatinge-Clay, J. O'connell, C. Khosla and R. M. Stroud, *Structure*, 2002, **10**, 1559–1568.

62   T. Bretschneider, G. Zocher, M. Unger, K. Scherlach, T. Stehle and C. Hertweck, *Nat. Chem. Biol.*, 2012, **8**, 154–161.

63   X. Yi, Q. Zhao, Z. Tian, X. Jia, W. Cao, W. Liu and Q. L. He, *Chinese J. Chem.*, 2019, **37**, 821–826.

64   D. R. Jackson, G. Shakya, A. B. Patel, L. Y. Mohammed, K. Vasilakis, P. Wattana-Amorn, T. R. Valentic, J. C. Milligan, M. P. Crump, J. Crosby and S. C. Tsai, *ACS Chem. Biol.*, 2018, **13**, 141–151.

65   Ö. Erol, T. F. Schäberle, A. Schmitz, S. Rachid, C. Gurgui, M. El Omari, F. Lohr, S. Kehraus, J. Piel, R. Müller and G. M. Königd, *ChemBioChem*, 2010, **11**, 1253–1265.

66   M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chemie Int. Ed.*, 2013, **52**, 1143–1147.

67   A. T. Fuller, G. Mellows, M. Woolford, G. T. Banks, K. D. Barrow and E. B. Chain, *Nature*, 1971, **234**, 416–417.

68   A. Badder and C. Garre, 1887, **17**, 597–597.

69   F. M. Martin and T. J. Simpson, *J. Chem. Soc. Perkin Trans. 1*, 1989, 207–209.

70   P. G. Mantle and K. M. Macgeorge, *J. Chem. Soc. Perkin Trans. 1*, 1991, 255–258.

71   T. Nakama, O. Nureki and S. Yokoyama, *J. Biol. Chem.*, 2001, **276**, 47387–47393.

72   M. J. Sugden, *Biosynth. pasedudomonic acid. Univ. Bristol., 1992*.

73   L. Wang, Z. Song, P. R. Race, J. Spencer, T. J. Simpson, M. P. Crump and C. L. Willis, *Chem. Sci.*, 2020, **11**, 5221–5226.

74   D. Fukuda, A. S. Haines, Z. Song, A. C. Murphy, J. Hothersall, E. R. Stephens, R. Gurney, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *PLoS One*, 2011, **6**, 1–9.

75   B. D. Cookson, R. W. Lacey, W. C. Noble, D. S. Reeves, R. Wise and R. J. Redhead, *Lancet (London, England)*, 1990, **335**, 1095–6.

76   A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–30.

77   L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Håkansson, J. L. Smith and D. H. Sherman, *Science.*, 2007, **318**, 970–974.

78   J. Wu, S. M. Cooper, R. J. Cox, J. Crosby, M. P. Crump, J. Hothersall, T. J. Simpson, C. M. Thomas and C. L. Willis, *Chem. Commun.*, 2007, 2040.

79   S. M. Cooper, R. J. Cox, J. Crosby, M. P. Crump, J. Hothersall, W. Laosripaiboon, T. J. Simpson and C. M. Thomas, *Chem. Commun.*, 2005, 1179.

80   S. S. Gao, J. Hothersall, J. Wu, A. C. Murphy, Z. Song, E. R. Stephens, C. M. Thomas, M. P. Crump, R. J. Cox, T. J. Simpson and C. L. Willis, *J. Am. Chem. Soc.*, 2014, **136**, 5501–5507.

81   L. Wang, A. Parnell, C. Williams, N. A. Bakar, M. R. Challand, D. Kamp, T. J. Simpson, P. R. Race, M. P. Crump and C. L. Willis, *Nat. Catal.*, 2018, **1**, 968–976.

82   A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Płoskoń, P. Wattana-Amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump, *Nat. Chem. Biol.*, 2013, **9**, 685–692.

83   S. S. Gao, J. Hothersall, J. Wu, A. C. Murphy, Z. Song, E. R. Stephens, C. M. Thomas, M. P. Crump, R. J. Cox, T. J. Simpson and C. L. Willis, *J. Am. Chem. Soc.*, 2014, **136**, 5501–5507.

84   H. Shiozawa, T. Kagasaki, T. Kinoshita, H. Haruyama, H. Domon, Y. Utsui, K. Kodama and S. Takahashi, *J. Antibiot.*, 1993, **46**, 1834–42.

85   B. Li and C. T. Walsh, *Proc. Natl. Acad. Sci.*, 2010, **107**, 19731–19735.

86   J. A. Connolly, A. Wilson, M. Macioszek, Z. Song, L. Wang, H. H. Mohammad, M. Yadav, M. di Martino, C. E. Miller, J. Hothersall, A. S. Haines, E. R. Stephens, M. P. Crump, C. L. Willis, T. J. Simpson, P. J. Winn and C. M. Thomas, *Sci. Rep.*, 2019, **9**, 1–9.

87   C. Davies, R. J. Heath, S. W. White and C. O. Rock, *Structure*, 2000, **8**, 185–195.

88   M. E. Abugrain, C. J. Brumsted, A. R. Osborn, B. Philmus and T. Mahmud, *ACS Chem. Biol.*, 2017, **12**, 362–366.

89   P. David Walker, *Mechanisms for the generations structural diversity in polyketide biosynthesis. Univ. Bristol., 2019.*

90   A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.

91   P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D. Biol. Crystallogr.*, 2010, **66**, 213–221.

92   T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read and P. D. Adams, *Acta Crystallogr. D. Biol. Crystallogr.*, 2008, **64**, 61–69.

93   P. Emsley and K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.*, 2004, **60**, 2126–2132.

94   P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams, *Acta Crystallogr. D. Biol. Crystallogr.*, 2012, **68**, 352–367.

95   V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L.

W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. D. Biol. Crystallogr.*, 2010, **66**, 12–21.

96   R. P. Joosten, F. Long, G. N. Murshudov and A. Perrakis, *IUCrJ*, 2014, **1**, 213–220.

97   E. Krissinel and K. Henrick, *J. Mol. Biol.*, 2007, **372**, 774–797.

98   E. Krissinel, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 376–385.

99   L. Holm and P. Rosenström, *Nucleic Acids Res.*, 2010, **38**, W545–W549.

100   C. Jiang, S. Y. Kim and D. Y. Suh, *Mol. Phylogenet. Evol.*, 2008, **49**, 691–701.

101   R. Nofiani, B. Philmus, Y. Nindita and T. Mahmud, *Medchemcomm*, 2019, **10**, 1517–1530.

102   B. R. Goblirsch, J. A. Frias, L. P. Wackett and C. M. Wilmot, *Biochemistry*, 2012, **51**, 4138–4146.

103   S. L. Drees, C. Li, F. Prasetya, M. Saleem, I. Dreveny, P. Williams, U. Hennecke, J. Emsley and S. Fetzner, *J. Biol. Chem.*, 2016, **291**, 6610–6624.

104   K. Gokulan, S. E. O'Leary, W. K. Russell, D. H. Russell, M. Lalgondar, T. P. Begley, T. R. Ioerger and J. C. Sacchettini, *J. Biol. Chem.*, 2013, **288**, 16484–16494.

105   J. G. Olsen, A. Kadziola, P. Von Wettstein-Knowles, M. Siggaard-Andersen, Y. Lindquist and S. Larsen, *FEBS Lett.*, 1999, **460**, 46–52.

106   W. G. J. Hol, P. T. Van Duijnen and H. J. C. Berendsen, *Nature*, 1978, **273**, 443–446.

107   X. Y. Jia, Z. H. Tian, L. Shao, X. D. Qu, Q. F. Zhao, J. Tang, G. L. Tang and W. Liu, *Chem. Biol.*, 2006, **13**, 575–585.

108   T. Ito, N. Roongsawang, N. Shirasaka, W. Lu, P. M. Flatt, N. Kasanah, C. Miranda and T. Mahmud, *ChemBioChem*, 2009, **10**, 2253–2265.

109   A. A. Eida and T. Mahmud, *Appl. Microbiol. Biotechnol.*, 2019, **103**, 4337.

110   A. A. Eida, M. E. Abugrain, C. J. Brumsted and T. Mahmud, *Nat. Chem. Biol.*, 2019, **15**, 795–802.

111   W. Zheng, C. Zhang, Y. Li, R. Pearce, E. W. Bell and Y. Zhang, *Cell Reports Methods*, 2021, **1**, 100014.

112   J. N. Scarsdale, G. Kazanina, X. He, K. A. Reynolds and H. T. Wright, *J. Biol. Chem.*, 2001, **276**, 20516–20522.

113   J. Kraut, *Annu. Rev. Biochem.*, 1977, **46**, 331–358.

114   B. R. Goblirsch, M. R. Jensen, F. A. Mohamed, L. P. Wackett and C. M. Wilmot, *J. Biol. Chem.*, 2016, **291**, 26698–26706.

115   E. S. Meadows and C. Khosla, *Biochemistry*, 2001, **40**, 14855–14861.

116   T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Håkansson, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2007, **282**, 35954–35963.

117   P. D. Walker, M. T. Rowe, A. J. Winter, A. N. M. Weir, N. Akter, L. Wang, P. R. Race, C. Williams, Z. Song, T. J. Simpson, C. L. Willis and M. P. Crump, *ACS Chem. Biol.*, 2020, **15**, 494–503.

118   J. Hothersall, J. Wu, A. S. Rahman, J. A. Shields, J. Haddock, N. Johnson, S. M. Cooper, E. R. Stephens, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2007, **282**, 15451–15461.

119   A. C. Murphy, D. Fukuda, Z. Song, J. Hothersall, R. J. Cox, C. L. Willis, C. M. Thomas and T. J. Simpson, *Angew. Chemie - Int. Ed.*, 2011, **50**, 3271–3274.

120   S. S. Gao, J. Hothersall, ⊥ Ji'en Wu, A. C. Murphy, Z. Song, E. R. Stephens, C. M. Thomas, M. P. Crump, R. J. Cox, T. J. Simpson and C. L. Willis, *J. Am. Chem. Soc.*, 2014, **136**, 5501–5507.

121   S. S. Gao, L. Wang, Z. Song, J. Hothersall, E. R. Stevens, J. Connolly, P. J. Winn, R. J. Cox, M. P. Crump, P. R. Race, C. M. Thomas, T. J. Simpson and C. L. Willis, *Angew. Chemie Int. Ed.*, 2017, **56**, 3930–3934.

122   Y. M. Zhang, B. Wu, J. Zheng and C. O. Rock, *J. Biol. Chem.*, 2003, **278**, 52935–52943.

123   R. Farmer, C. M. Thomas and P. J. Winn, *PLoS One*, 2019, **14**, e0219435.

124   R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate and M. Punta, *Nucleic Acids Res.*, 2014, **42**, D222–D230.

125   A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty and A. R. Slabas, *J. Mol. Biol.*, 2007, **365**, 135–145.

126   E. Płoskoń, C. J. Arthur, A. L. P. Kanari, P. Wattana-Amorn, C. Williams, J. Crosby, T. J. Simpson, C. L. Willis and M. P. Crump, *Chem. Biol.*, 2010, **17**, 776–785.

127   L. G. Morão, L. R. Manzine, L. O. D. Clementino, C. Wrenger and A. S. Nascimento, *PLoS One*, 2022, **17**, e0271403 (1-11).

128   M. T. Rowe, *Fatty acid construction within biosynththesis of polyketide antibiotic mupirocin. Univ. Bristol.,2021*.

129   J. A. Shields, A. S. Rahman, C. J. Arthur, J. Crosby, J. Hothersall, T. J. Simpson and C. M. Thomas, *ChemBioChem*, 2010, **11**, 248–255.

130   R. Del Conte, D. Lalli and P. Turano, *Disruption of Protein-Protein Interfaces. Springer-Verlag Berlin Heidelb.*, 2013, 83–111.

131   Q. Bashir and N. Rashid, in *Applications of NMR Spectroscopy*, Bentham Science, 2020, vol. 8, pp. 121–140.

132   R. Puthenveetil and O. Vinogradova, *J. Biol. Chem.*, 2019, **294**, 15914–15931.

133   K. H. Gardner and L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 357–406.

134   P. D. Walker, C. Williams, A. N. M. Weir, L. Wang, J. Crosby, P. R. Race, T. J. Simpson, C. L. Willis and M. P. Crump, *Angew. Chemie Int. Ed.*, 2019, **58**, 12446–12450.

135   I. Nazi, K. P. Koteva and G. D. Wright, *Anal. Biochem.*, 2004, **324**, 100–105.

136   E. Strauss and T. P. Begley, *J. Biol. Chem.*, 2002, **277**, 48205–48209.

137   A. S. Worthington and M. D. Burkart, *Org. Biomol. Chem.*, 2006, **4**, 44–46.

138   N. L. Kelleher, H. Y. Lin, G. A. Valaskovic, D. J. Aaserud, E. K. Fridriksson and F. W. McLafferty, *J. Am. Chem. Soc.*, 1999, **121**, 806–812.

139   P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756–12766.

140   D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3107–3111.

141   F. P. Maloney, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Proc. Natl.*

Acad. Sci. U. S. A., 2016, **113**, 10316–10321.

142   H. M. Miziorko and M. D. Lane, *J. Biol. Chem.*, 1977, **252**, 1414–1420.

143   S. Kapur, A. Worthington, Y. Tang, D. E. Cane, M. D. Burkart and C. Khosla, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3034–3038.

144   G. A. R. Thiele, C. P. Friedman, K. J. S. Tsai, J. Beld, C. H. Londergan and L. K. Charkoudian, *Biochemistry*, 2017, **56**, 2533–2536.

145   A. Miyanaga, R. Ouchi, F. Ishikawa, E. Goto, G. Tanabe, F. Kudo and T. Eguchi, *J. Am. Chem. Soc.*, 2018, **140**, 7970–7978.

146   J. T. Mindrebo, A. Patel, W. E. Kim, T. D. Davis, A. Chen, T. G. Bartholow, J. J. La Clair, J. A. McCammon, J. P. Noel and M. D. Burkart, *Nat. Commun. 2020 111*, 2020, **11**, 1–15.

147   J. A. Loo, *Mass Spectrom. Rev.*, 1997, **16**, 1–23.

148   V. Katta and B. T. Chait, *J. Am. Chem. Soc.*, 1991, **113**, 8534–8535.

149   J. Sambrook, E. R. Fritsch and T. Maniatis, *Moleluar cloning A laboratory manual (2nd ed.). Cold Spring Harb. Lab. Press., 1989*.

150   S. Tayyab, S. Qamar and M. Islam, *Biochem. Educ.*, 1991, **19**, 149–152.

151   A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.

152   P. Emsley and K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2004, **60**, 2126–2132.

153   R. P. Joosten, F. Long, G. N. Murshudov and A. Perrakis, *IUCrJ*, 2014, **1**, 213–220.

154   L. Schrodinger, *The PyMOL Molecular Graphics System, Version 1.3 r1. PyMol., 2016*.

155   V. Agarwal, S. Diethelm, L. Ray, N. Garg, T. Awakawa, P. C. Dorrestein and B. S. Moore, *Org. Lett.*, 2015, **17**, 4452–4455.

156   R. B. Hamed, L. Henry, J. R. Gomez-Castellanos, A. Asghar, J. Brem, T. D. W. Claridge and C. J. Schofield, *Org. Biomol. Chem.*, 2013, **11**, 8191.

157   F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.

158   W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687–696.

9. Supplementary information

Table S.9.1. List of primers used in this work. The mutation sites are shown in bold.

No.	Primer	Sequence
1.	MacpB_pOPINF Forward	aagttctgtttcagggcccgatggaaatcaacgtgc
	MacpB_pOPINF Reverse	atggtctagaaagctttactagccatgcgctaccag
2.	TmlB_pOPINF Forward	aag ttc tgt ttc agg gcc cga tgg ggc aat cct gtt taca
	TmlB_pOPINF Reverse	atg gtc tag aaa gct tta cta aca tgc ctg taa taa tat ttg
3.	MupBC135A_pOPINF Forward	acatggggcgat**gct**gctaccatgctgttggca
	MupBC135A_pOPINF Reverse	tgccaacagcat**ggt**agcatcgccccatgt
4.	TmlBC136A_pOPINF Forward	gtcactgtgtcggac**gct**gctaaccttctc
	TmlBC136A_pOPINF Reverse	tgagagaaggttagc**agc**gtccgacacagt
5.	TmpB_ACP12_pOPINE Forward	aggagatataccatgattagcgcacatgcgtctatc
	TmpB_ACP12_pOPINE Reverse	gtgatggtgatgtttgttcaatattatcactatctagctgcgcta
6.	TmpB_ACP34_pOPINE Forward	aggagatataccatggaacaacgtccgcaatttatactcagtgca
	TmpB_ACP34_pOPINE Reverse	gtgatggtgatgtttttaccagcgccgaaatggtttcgtg

Table S.9.2. List of amino acid sequences of proteins used in this work.

No.	Proteins
1.	MupB
	MAHHHHHHSSGLEVLFQGPMRTAYLASTAYVLGEQAHDYRDAASFEAVCKQHSMPDFASVFGWGTY WRTTRSVGELLVESVASTLASSGLRGCDIDALVVCSSNFESGQVVDYLPLLRELQLARAFPLGVTW GDCTMLLAGLEVARAQVLAGLDNVLVVSANRIEDEAFRFQHYALFSDGAASCLVTSGRRRGFEMLG SLARSNAGLAHDPKEDDTRLFREVHEQFMHRQQINTADLEQVVCSNVFLPVLKIKEGRQGVSGTQL YVDNVTRVGHCFSADSLINLCDYQARAQQVHGGLVMLTANADGLRCQTLLQRVSDA **Calculated Mw:** 35000.52 Da (without Met)
2.	MacpA
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMNPERRNMYMEEIYTFVVSTLASSCKVQPGDIE PTTNLFADLGIDSVDFLDAVFCIEKHYDIRIPVGQWMSAVNEGNAAMTDYFVMEHFVAQIAARAAA SA **Calculated Mw:** 15042.03 Da
3.	MacpB
	AHHHHHHSSGLEVLFQGPMEINVAQQTVSESIRELKLPGWETASIETETHLQGELGIDSLHKLILL TRLQERSNFQFAQLNNEAYKFDTVGDLVNLLVAHG **Calculated Mw:** 11372.74 Da (without Met)
4.	MacpD
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTLNHQVMDQVFDQVEHQIAQVLGAKGGPLVAVEI DSRFSDLGLSSLDLATLISNLEAVYGTDPFADAVAITSIVTVADLARAYAQQGVPGPSPDPLDAQL RDLRQL **Calculated Mw:** 14922.96 Da
5.	MmpA_3a

	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMAPLAAKAAPVVPVADDECAQFLRQSLAAMLYC EPGQIRDGSRFLELGLDSVIAAQWIREINKHYQLKIPADGIYTYPVFKAFTQWVGTQLQ **Calculated Mw:** 14038.12 Da
6.	MmpE_ACP
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTLEANAEDVHEVVRAQIAALTQIDPHTLQLGKPL LDYGLNSVTMLVLAACLCQALEGLDESRDGPQLVLCRTAGALIEQLEQMLAPT **Calculated Mw:** 13004.89 Da
7.	MmpB_ACP567
	GSSHHHHHHSSGLVPRGSHMSDTLRGLVGQILKVDAQEIDDTTAFSDMGFDSVMLTELATAINRTY TLELGTAALFEHPTLQALAAHLQGARTAESQPPAPGLTRAQVA QGVREVVAEALKVRLEDIGDDDPWSDYGMDSVSSVQMTGLLNERFDIQLAADTFQAFGNVVELTTA IADIQVVMACADIERQPSAGGGVIADTALLDELVALVCQLLKTVAGDIDPHTDLHDFGFDSVLLTQ LLAQISSTYGVELDPGSVLEDATVAGLVAQVQLEHHHHHH **Calculate Mw:** 29985.51 Da
8.	MmpB_ACP5
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTRSQVDPSDTLRGLVGQILKVDAQEIDDTTAFSD MGFDSVMLTELATAINRTYTLELGTAALFEHPTLQALAAHLQGARTA **Calculated Mw:** 12373.88 Da
9.	MmpB_ACP6
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTPPAPGLTRAQVAQGVREVVAEALKVRLEDIGDD DPWSDYGMDSVSSVQMTGLLNERFDIQLAADTFQAFGNVVELTTAIADIQVVMA **Calculated Mw:** 13118.74 Da
10.	MmpB_ACP7
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTIADTALLDELVALVCQLLKTVAGDIDPHTDLHD FGFDSVLLTQLLAQISSTYGVELDPGSVLEDATVAGLVAQVQAQRHGAEPA **Calculated Mw:** 12565.13 Da
11.	MupN
	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFMDREWPLPPGDIQLWVCRDELIVAADLLAE YAAVLSDSERQRMAQFHFARHRHQYLVTRALVRHVLSMYRPGIPPQHWCFAYNDHGKPCIDPVLDP SQSLHFNVSHTDGLIVMAVSRQAIGIDVEDLQRQGDGVSVADTFFAAAERHDLQQCGAAHRAQRFF EIWTLKEAYIKARGEGLAVPLDSFAFRLGPSCGHSLQFNPPTTDARHWQFWSTLLLARYCLAVAAC HESPRARFCLRVQEVIPRVRTL DLACLCRRQAAPHTVTPPWRLPGICATPEIDLG **Calculated Mw:** 35723 Da (without Met)
12.	TmlB
	MAHHHHHHSSGLEVLFQGPMGQSCLHSAKYFLGEVERQYHEVEVFQSVLAEHGLPNMPKLFAWGGY RTSTKTIFEMAYESMKLTLEAAQVDAQDVGLVIICSSDFNPVHETQSYSQLLVSLGMKYAFPMGVT VSDCANLLSALDMTKQLIDSDKYPHIMLVTSNKISDERHRFQDYALFSDGAASIMISNKKTLNRIP EDGLNVVDSQINAQLQLKVAGEDIDDTPLYTATTQQLVDRNQITISDLSKVFSNNLYMPVITLKEG GIGVSKSQLYLDNIVRFGHCFSADSIINLADYLANS RVAAGEYFALHSSASGLRAQILLQAC **Calculated Mw: 35970.82 Da** (without Met)
13.	TacpA

	MAHHHHHHSSGLEVLFQGPMEGIYQFVQSTFSDVCNIDSEEITPTTNLFSDLGIESMDFMDVCYLI DEKYGIRIPIGEWMGRVNEGDESAADLFVLEGFVKAVSKLVEGEVA **Calculate Mw (M):** 12378.85 Da
14.	TacpB
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMPHINYISIHGERDNMSNEQASTAKALVISFID QMNRSEYRGSDIRESFELKNMLNIDSLNKLILLTRITEHLGVEFGESDITINDFTTVSDLISFVET HTRMAETHTEPS **Calculated Mw:** 16457.41 Da
15.	TacpD
	MAHHHHHHSSGLEVLFQGPMLNVQQIKNEVLDSIIQEVATILSEKLLPITHIENSTAFSELSLTSL DLAELISNLEARYEVDPFEELVAITSIVTIEDLASAYALSLSGNTEDSHDLLSDELKAIKNQVR **Calculate Mw:** 14266 Da (without Met)
16.	TmpA_3c
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTEFLRTSLANILMMGVQEIKPTRCFADFGLDSIT GMEWVKLINQTMQLGLDADIVYEHNTLERFSLYIQSLLQNKSSQQIQLIAGVLQTQSTNKTAQGPH LKVANQHKEHKQMEEAL **Calculated Mw: 16970.39 Da**
17.	TmpE_ACP
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTSGSSEREYFVVETVSQIIGLTHEELNITTPLKA LGVSSITKMLIVSAICDGIEQLDEVQVGSELIKAATVADIIRIV NH **Calculated Mw:** 12278 Da
18.	TmpB_ACP1-4
	MAHHHHHHSSGLEVLFQGPISAHASINDNDVISAVSTLLKVPQSSIDATTAFFDIGFDSVSISELA IILTQKTGVELSPTVFYECTQINELTHYLQTHMHNKIVAAVEARTEPVRTRAVSSPVNAQLIIGMV SELLKVPESSIDPSTAFADIGFDSVSLGELAATLSDAFGIEISATLFYECTVVAELVDYLLAQLDS DNIEQRPQFILSAPVEVKQTTTQSMSLNDAHIITVISSLLKVPESSIDADTTFEQIGFDSVSLSEL AGTLAKRYDIEISTTVFYECTTVGALTQHINAQLGGKEPSTYEVLPTSSLPPVHIKERAGLELSAE ITSIICEMLKIGRHDFDDHTPLGQLGFDSVSFAELSQKLSHRYGKDITPTIFYAHETISALV **Molecular weight:** 42546.03
19.	TmpB_ACP3
	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMSLNDAHIITVISSLLKVPESSIDADTTFEQIG FDSVSLSELAGTLAKRYDIEISTTVFYECTTVGALTQHINAQLGGKEPSTY **Molecular weight:** 12838.39
20.	TmpB_ACP12
	MISAHASINDNDVISAVSTLLKVPQSSIDATTAFFDIGFDSVSISELAIILTQKTGVELSPTVFYE CTQINELTHYLQTHMHNKIVAAVEARTEPVRTRAVSSPVNAQLIIGMVSELLKVPESSIDPSTAFA DIGFDSVSLGELAATLSDAFGIEISATLFYECTVVAELVDYLLAQLDSDNIEQTSPSPSLSD **Molecular weight:** 20757.37
21.	TmpB_ACP34
	TmpB_ACP34 MEQRPQFILSAPVEVKQTTTQSMSLNDAHIITVISSLLKVPESSIDADTTFEQIGFDSVS LSELAGTLAKRYDIEISTTVFYECTTVGALTQHINAQLGGKEPSTYEVLPTSSLPPVHIK

```
ERAGLELSAEITSIICEMLKIGRHDFDDHTPLGQLGFDSVSFAELSQKLSHRYGKDITPTIFYAHE
TISALVKHHHHHH
```

Molecular weight: 22008.86

Table S.9.3. Hydrogen bonds and salt bridge interactions between the monomer A and monomer B residues of the TmlB crystal structure. The monomer chain, residue and atoms involved in interactions are represented as C:RES NN [atom name] format where atom name is shown in PDB notation. The interatomic distance and buried surface area are represented by Å and Å^2, respectively.

Hydrogen Bond				
Monomer A	**BSA [Å^2]**	**Distance [Å]**	**Monomer B**	**BSA [Å^2]**
A:SER 296[O]	22.78	1.89	B:TYR 95[HH]	109.32
A:THR 112[O]	49.95	2.01	B:GLY 111[H]	48.65
A:GLY 111[O]	48.25	2.15	B:THR 112[H]	49.43
A:ASP 125[OD1]	24.93	1.97	B:LYS 135[HZ1]	88.18
A:LYS 105[O]	37.05	2.07	B:LEU 194[H]	126.49
A:LEU 294[H]	127.51	2.00	B:LYS 105[O]	36.64
A:THR 112[H]	49.95	2.09	B:GLY 111[O]	48.65
A:GLY 111[H]	48.25	2.06	B:THR 112[O]	49.43
A:LYS 135[HZ1]	67.33	2.24	B:ASP 125[OD1]	31.29
A:TYR 95[HH]	109.16	1.88	B:SER 296[O]	24.37

Salt Bridge		
MonomerA	**Distance [Å]**	**Monomer B**
A:ASP 125[OD1]	2.82	B:LYS 135[NZ]
A:ASP 125[OD2]	3.51	B:LYS 135[NZ]
A:LYS 135[NZ]	3.08	B:ASP 125[OD1]
A:LYS 135[NZ]	3.20	B:ASP 125[OD2]

Table S.9.4. Hydrophobic patch dimeric interactions for two compositions (AC and BD) of the tetrameric MupB crystal structure.

Hydrogen Bond at A+C MupB Dimer				
Monomer A	**BSA [Å^2]**	**Distance [Å]**	**Monomer C**	**BSA [Å^2]**
A:TRP 113[NE1]	201.5	3.00	C:VAL 92[O]	8.35
A:ASP 287[N]	38.4	3.48	C:TYR 94[OH]	119.51
A:SER 184[OG]	70.47	3.64	C:ALA 104[O]	13.70
A:ALA 186[N]	87.32	3.12	C:ALA 104[O]	13.70
A:SER 184[N]	70.47	2.99	C:ARG 105[O]	68.20
A:SER 184[OG]	70.47	2.67	C:ALA 106[O]	5.77
A:THR 112[N]	46.24	2.93	C:GLY 110[O]	51.55
A:ALA 106[N]	5.77	3.35	C:SER 184[OG]	70.47
A:TYR 94[OH]	119.51	2.92	C:ASP 287[O]	38.4
A:VAL 92[O]	8.35	2.93	C:TRP 113[NE1]	201.5
A:TYR 94[OH]	119.51	3.62	C:ASP 287[N]	38.4
A:ALA 104[O]	13.70	3.82	C:SER 184[OG]	70.47
A:ARG 105[O]	68.20	2.85	C:SER 184[N]	70.47

A:GLY 110[O]	51.55	3.03	C:THR 112[N]	46.24
A:ASP 115[OD2]	9.69	3.84	C:TYR 94[OH]	119.51
A:SER 184[OG]	70.47	3.35	C:ALA 106[N]	5.77
A:ALA 186[O]	87.32	3.06	C:ARG 99[NE]	36.62
A:ASP 287[O]	38.4	2.78	C:TYR 94[OH]	119.51

Hydrogen Bond at B+D MupB Dimer				
Monomer B	**BSA [Å²]**	**Distance [Å]**	**Monomer D**	**BSA [Å²]**
B:TYR 94[OH]	125.32	3.04	D:ASP 287[O]	40.33
B:TYR 94[OH]	125.32	3.74	D:ASP 115[OD2]	8.56
B:ALA 106[N]	3.07	3.90	D:SER 184[OG]	70.14
B:THR 112[N]	47.63	2.96	D:GLY 110[O]	51.52
B:TRP 113[NE1]	200.18	3.15	D:VAL 92[O]	7.37
B:SER 184[N]	70.14	2.72	D:ARG 105[O]	60.88
B:SER 184[OG]	70.14	3.76	D:ALA 104[O]	26.58
B:ASP 287[N]	40.33	3.33	D:TYR 94[OH]	125.32
B:ASP 77[O]	8.71	2.35	D:ARG 183[NH2]	55.48
B:VAL 92[O]	7.37	2.94	D:TRP 113[NE1]	200.18
B:TYR 94[OH]	125.32	3.49	D:ASP 287[N]	40.33
B:ARG 105[O]	60.88	2.79	D:SER 184[N]	70.14
B:ARG 105[O]	60.88	2.42	D:SER 184[OG]	70.14
B:GLY 110[O]	51.52	3.04	D:THR 112[N]	47.63
B:ASP 115[OD2]	8.56	3.83	D:TYR 94[OH]	125.32
B:SER 184[OG]	70.14	3.50	D:ALA 106[N]	3.07
B:ASN 185[OD1]	44.19	3.44	D:ARG 99[NH2]	0
B:ASP 287[O]	40.33	2.78	D:TYR 94[OH]	125.32

Table S.9.5. BLAST search of MupB and TmlB: First 20 hits of the BLAST search of MupB and TmlB using the NCBI database (non-reductant) sorted by their E-values.

MupB

Accession	Description	E-Value
WP_058787812.1	hypothetical protein [*Pseudomonas psychrotolerans*].	9e-139
WP_177143949.1	hypothetical protein [*Pseudomonas gingeri*].	1e-134
WP_211013248.1	hypothetical protein [*Pseudoalteromonas luteoviolacea*].	1e-68
WP_190337401.1	hypothetical protein [*Vibrio sp. S9_S30*].	1e-68
ESP90255.1	3-oxoacyl-[acyl-carrier-protein] synthase III [*Pseudoalteromonas luteoviolacea 2ta16*].	3e-67
WP_036977451.1	hypothetical protein [*Pseudoalteromonas*].	7e-67
WP_013933286.1	hypothetical protein [*Pseudoalteromonas sp. SANK 73390*].	9e-64
WP_072395680.1	hypothetical protein [*unclassified Pseudomonas*].	1e-60
WP_057008717.1	hypothetical protein [*Pseudomonas trivialis*]	2e-59
WP_055135441.1	hypothetical protein [*Pseudomonas corrugata*].	4e-58
WP_093801744.1	3-oxoacyl-ACP synthase [*Streptomyces sp. Wb2n-11*].	2e-54
WP_266942875.1	3-oxoacyl-ACP synthase [Streptomyces sp. NBC_00555].	3e-54
WP_075130984.1	3-oxoacyl-ACP synthase [*Actinophytocola xinjiangensis*].	1e-52
WP_124231856.1	3-oxoacyl-ACP synthase [*Erwinia psidii*].	3e-52
MBZ4017119.1	3-oxoacyl-ACP synthase [*Streptomyces purpurogeneiscleroticus*].	4e-52

MBV9022374.1	3-oxoacyl-ACP synthase [*Streptomycetaceae bacterium*].	2e-52
WP_228979355.1	3-oxoacyl-ACP synthase [*Streptomyces sp. DH12*].	2e-50
WP_184995921.1	3-oxoacyl-ACP synthase [*Actinoplanes digitatis*].	9e-50
WP_167473660.1	3-oxoacyl-ACP synthase [*Nocardia arthritidis*].	1e-49
MBT2402880.1	3-oxoacyl-ACP synthase [*Streptomyces sp. ISL-21*].	2e-49

TmlB

Accession	Description	E-Value
ESP90255.1	3-oxoacyl-[acyl-carrier-protein] synthase III [*Pseudoalteromonas luteoviolacea 2ta16*].	9e-177
WP_036977451.1	hypothetical protein [*Pseudoalteromonas*].	1e-176
WP_177143949.1	hypothetical protein [*Pseudomonas gingeri*].	9e-70
WP_100941003.1	hypothetical protein [*unclassified Pseudomonas*].	1e-69
WP_134924867.1	hypothetical protein [*Pseudomonas fluorescens*].	8e-67
WP_181283984.1	hypothetical protein [*Pseudomonas fluorescens group*].	1e-66
WP_253362608.1	hypothetical protein [*Pseudomonas chlororaphis*].	3e-66
WP_124432253.1	hypothetical protein [*Pseudomonas orientalis*].	2e-65
WP_072458865.1	hypothetical protein [*Pseudomonas*]	6e-65
WP_057008717.1	hypothetical protein [*Pseudomonas trivialis*]	2e-64
WP_055135441.1	hypothetical protein [*Pseudomonas corrugata*]	3e-63
WP_058787812.1	hypothetical protein [*Pseudomonas psychrotolerans*]	3e-61
WP_184995921.1	3-oxoacyl-ACP synthase [*Actinoplanes digitatis*]	2e-53
WP_124231856.1	3-oxoacyl-ACP synthase [*Erwinia psidii*].	4e-49
WP_112094664.1	3-oxoacyl-ACP synthase [*Lonsdalea populi*]	5e-48
WP_218028152.1	3-oxoacyl-ACP synthase [*Streptomyces fragilis*]	8e-48
WP_112092172.1	3-oxoacyl-ACP synthase [*Lonsdalea*]	1e-47
MBT2402880.1	3-oxoacyl-ACP synthase [*Streptomyces sp. ISL-21*]	2e-46
WP_127377363.1	3-oxoacyl-ACP synthase [*Morganella morganii*]	3e-46
WP_251049637.1	3-oxoacyl-ACP synthase [*unclassified Streptomyces*]	3e-46

Table S.9.6. Comparison of thiolase fold member proteins with MupB and TmlB. A comparison of representative members of the thiolase superfamily according to the overall 3D structural alignment with MupB and TmlB using the DALI server.

Protein	PDB entry	C-rmsd with MupB (Å)	C-rmsd with TmlB (Å)	Sequence identity (%) with MupB	Sequence identity (%) with TmlB	Function
DpsC (KS^s)	5TT4	2.3	2.5	15	13	Acyl transferase and Ketosynthase
FabH	1EBL	2.5	2.6	16	15	Claisen like condensation
CerJ (KAS III)	3S3L	2.6	2.6	14	16	O-Malonyl transferase and Reverse esterase
ZhuH (KAS III)	1MZJ	2.6	2.7	16	17	Claisen like condensation
HMGS	2F9A	2.7	3.0	11	12	Claisen like condensation
CHS	1EE0	2.8	2.9	15	11	Chalcone synthase
THI	2F2S	2.9	3.0	9	14	Claisen like condensation
FabB (KAS I)	1G5X	3.4	3.3	12	12	Claisen like condensation
FabF (KAS II)	2GFW	3.4	3.2	13	10	Claisen like condensation

Table S.9.7. The structural homologs search result of the refined TmlB crystal structure using the DALI server against the whole PDB archive. Top Ten (10) protein structures those are structurally related to the TmlB-chain A are listed according to their Z score (Similarity score, acceptable value is > 2).

* **RMSD** stands for the Root-mean-square deviation, calculated between Cα-atoms of matched residues at best 3D superimposition of query and target structures. **%$_{seq}$:** % Sequence identity

PDB entry	Z-score	Cα-RMSD (Å)	%$_{seq}$	Description
7eqi-E	32.7	2.3	13	ChlB3 (AT) from *Streptomyces antibioticus*
4xsb-B	31.6	2.5	13	Butanoyl bound DpsC from *Streptomyces peucetius*
3s20-B	31.2	2.8	12	Cerulenin bound OleA from *Xanthomonas campestris*
3t8e-A	31.1	2.6	16	Cervimycin K bound CerJ from *Streptomyces tendae*
6esz-A	30.5	2.7	16	PqsBC (KAS III) from *Pseudomonas aeruginosa* PAO1
4jap-C	30.4	2.9	14	Pks11 from *Mycobacterium tuberculosis*
2ebd-B	30.0	2.5	15	3-oxoacyl-ACP synthase III from *Aquifex aeolicus* VF5
3il9-B	29.1	2.6	15	FabH from *Escherichia coli*
1ebl-B	29.1	2.7	15	FabH from *Escherichia coli*
1mzj-A	28.7	2.7	17	ZhuH (PKS III) from *Streptomyces* sp. R1128

Table S.9.8. The 3D pairwise structural alignment result is presented here between the TmlB-WT (target) and MBE-TmlB (query) structures using the PDBeFold server. The Q-score represents the quality function of Cα-alignment (with 1 being the highest score)., which takes both alignment length and RMSD into account.

*RMSD stands for the Root-mean-square deviation, calculated between Cα-atoms of matched residues at best 3D superimposition of query and target structures. **%$_{seq}$:** % Sequence identity, **%$_{sse}$:** % of matched secondary structural elements, **N$_{align}$:** Number of matched residues or length of alignment, **N$_{res}$:** Number of query's total residues.

MBE-TmlB Chain	Q-score	RMSD (Å)	%$_{seq}$	%$_{sse}$	N$_{align}$	N$_{res}$
A	1.0	0.13	100	100	299	299
B	1.0	0.16	100	100	305	305

Table S.9.9. The structural homologs search result using the PDBe Fold server against the refined MBE-TmlB cocrystal structure. Proteins those are structurally related to MBE-TmlB_chain A are listed according to their Q-score (Quality of alignment, with 1 being the highest score).

PDB entry	Q-score	Cα-RMSD (Å)	%$_{seq}$	%$_{sse}$	Description
7eqi:H	0.50	1.92	13	78	ChlB3 (AT) from *Streptomyces antibioticus*
4z19:A	0.47	2.13	16	78	FabH from *Yersinia pestis* with acylated active site cysteine
5wgc:A	0.46	1.99	13	78	Propionyl-DpsC from *Streptomyces peucetius*
2ebd:B	0.46	2.13	16	83	3-oxoacyl-ACP synthase III from *Aquifex aeolicus* VF5
5vxe:A	0.46	1.85	13	78	Cerulenin bound OleA (E117Q) from *Xanthomonas campestris*
3il9:A	0.45	2.11	16	83	FabH from *Escherichia coli*
1ebl:A	0.45	2.17	16	83	FabH from *Escherichia coli*
3t5y:B	0.45	2.06	16	74	Malonic acid-CerJ from *Streptomyces tendae*
4jar:B	0.45	2.04	16	83	Pks11 from *Mycobacterium tuberculosis*
1mzj:B	0.43	2.27	18	78	ZhuH (KAS III) from *Streptomyces* sp. R1128

Figure S.9.1. TmlB (A) The amplified TmlB DNA showing a band ~1000 bps and (B) SDS-PAGE gel image of the TmlB IMAC fractions from the two peaks, which shows that the 2nd peak contains TmlB protein.

Figure S.9.2. (A) The structural elements of two MupB monomers (orange for monomer A and magenta for monomer C) are shown in front view. **(B)** All the elements from each monomer that form the TmlB dimeric interface are labelled as Nα2, Nα2′, Nβ3, Nβ3′, Nα3, Nα3′, loop L3, L3′, L5, L5′, L9, L9′, L14 and L14′. The hydrogen bond interactions are shown between Nα2- L14′, L9 and C-terminus of Nα2′ and L4-L9′.

Figure S.9.3. BLAST search for conserved domain using amino acid sequences of MupB and TmlB. Both proteins are found to lie in the FabH superfamily.

Figure S.9.4. Amino acid sequence alignment: The amino acid sequence and secondary structure elements of MupB and TmlB aligned with structurally homologous proteins. The active site triads are marked with a blue triangle. The nomenclature used to label the β-stands and α-helices emphasizes the internal structural repeat.
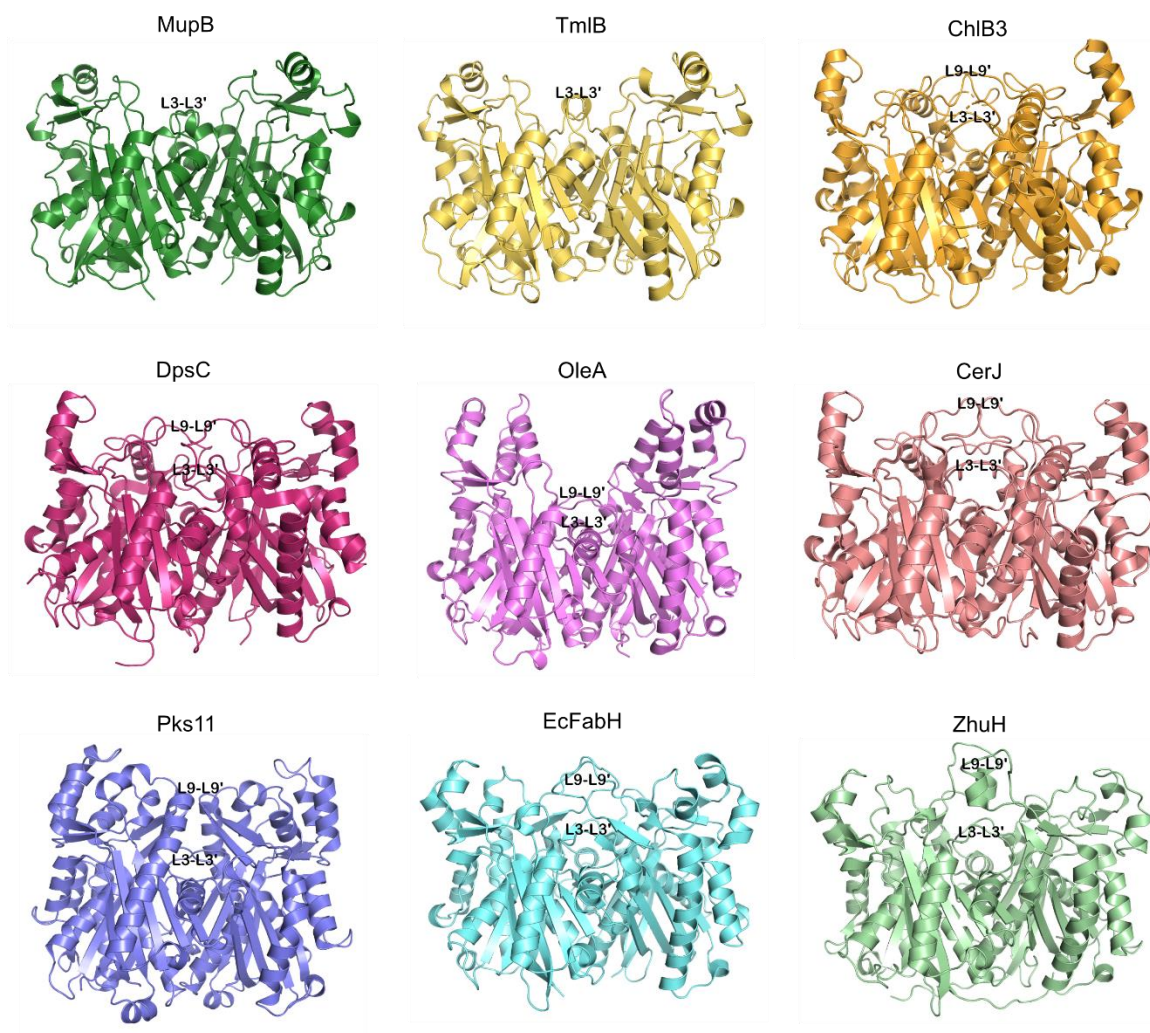
Figure S.9.5. Crystal structures of MupB and TmlB and their structural homologues in colour shades. All the structures are observed to have identical dimeric conformation of their identical monomers. The main differences are observed between the architecture and arrangements of the loops and other secondary elements and not the core structures. The position of loops L3-L3´ and L9-L9´ are shown here.
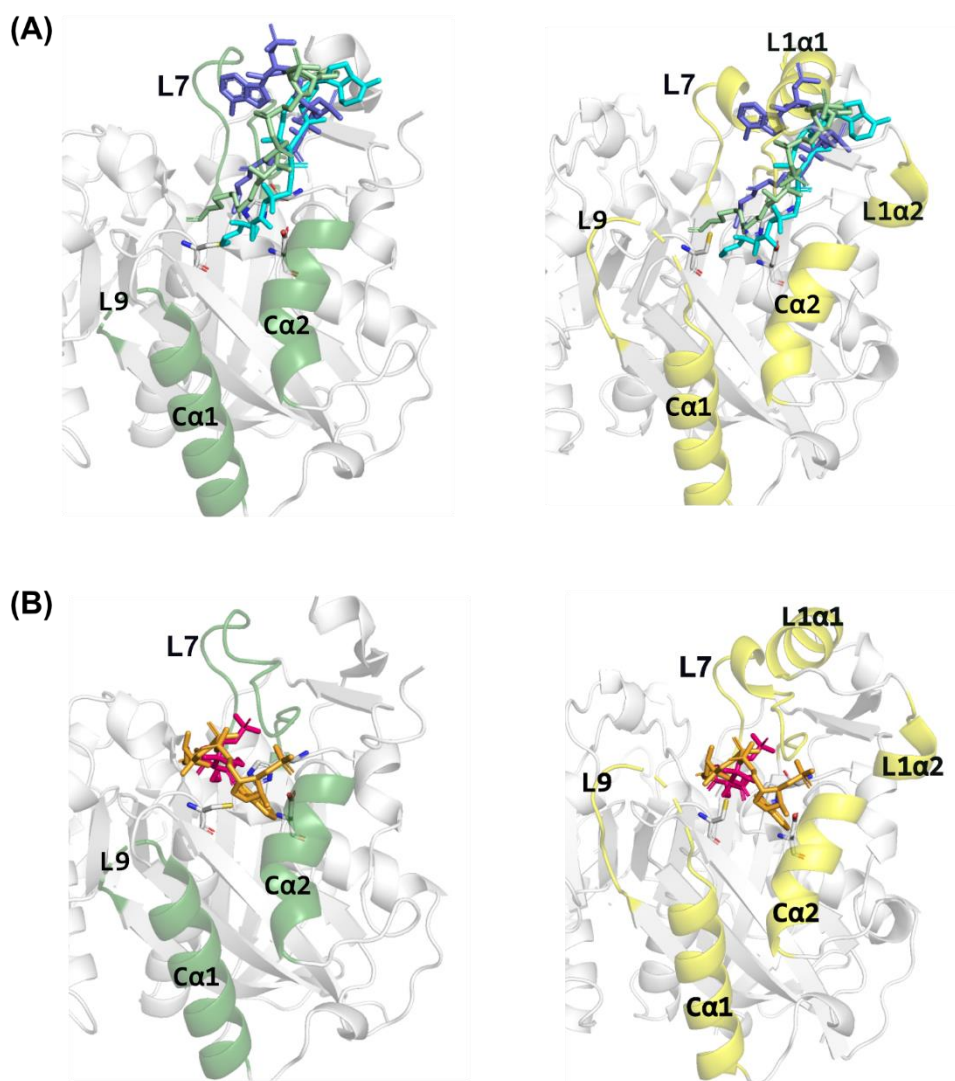
Figure S.9.6. Structural comparison of the coenzyme A and phosphopantetheine entrance to the active site cavity of MupB (green) and TmlB (yellow) with (A) EcFabH and ZhuH where the CoAs are shown as sticks in cyan and purple, respectively (B) DpsC and CerJ where the phosphopantetheine from DpsC is shown in hot pink and CoA from CerJ is shown in orange. The substrate entrance channels of MupB are coloured in green and yellow, respectively. Only the corresponding tertiary structure of MupB and TmlB monomers are shown here in grey cartoon.
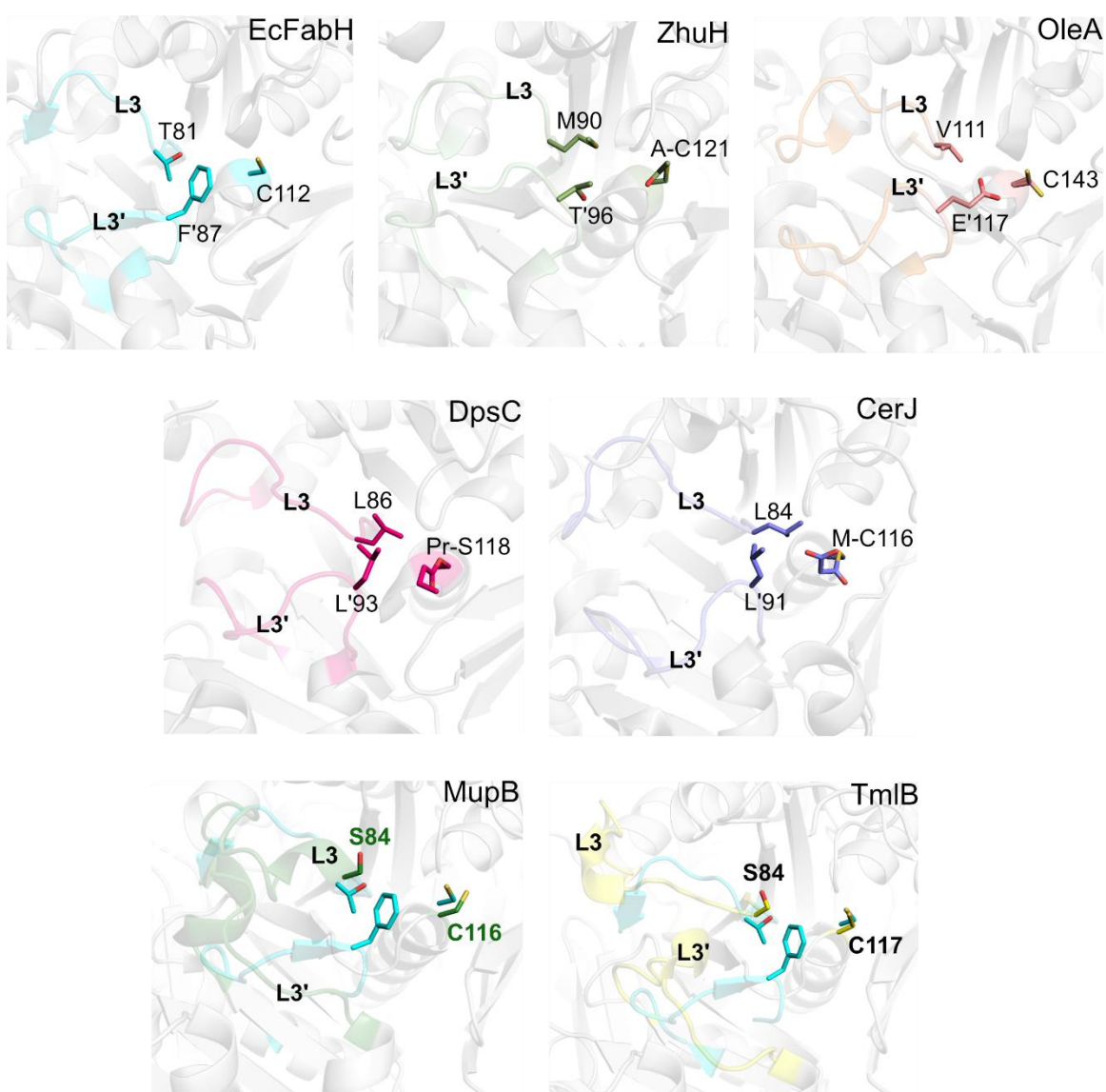
Figure S.9.7. Structural features L3-L3´ of EcFabH (Cyan), ZhuH (olive green), OleA (brown), DpsC (hot pink) and CerJ (purple) is compared with MupB (green) and TmlB (yellow) L3-L3´ loops. It has been observed that S84 from L3 in both MupB and TmlB matched with the all the homologs the L3 residue for example T81 from EcFabH. But The L3´ loops from both MupB and TmlB are seen to flip backward compared to the L3´ loop of EcFabH.
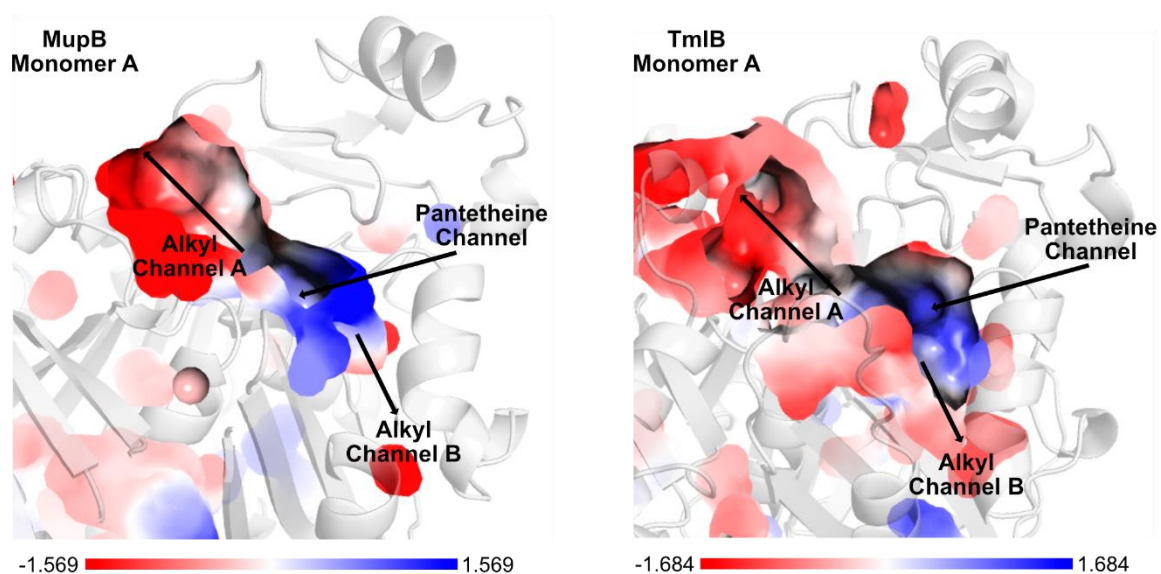
Figure S.9.8. Calculated electrostatic charge map of putative active site cavities of (A) MupB and (B) TmlB. The positive and negative charge states are shown in bule and red colours. The putative active cavity and electrostatic charge state are produced by using Pymol educational software.
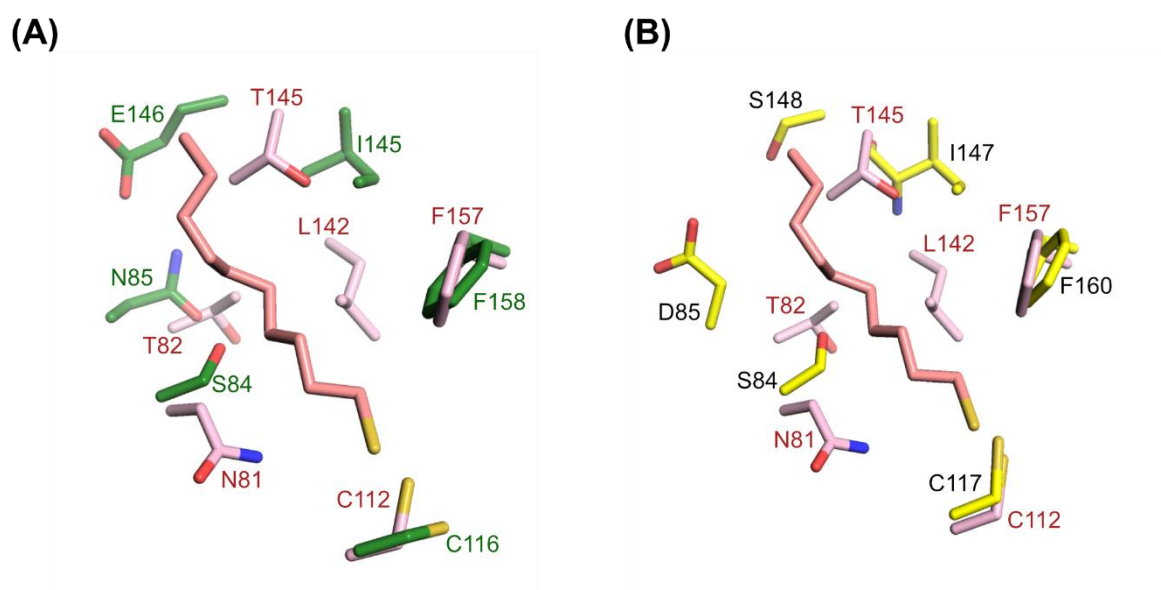
Figure S.9.9. Alkyl binding residue comparison of decane-1-thiol-FabH (PDB:2QX1) with (A) MupB and (B) TmlB). Only those FabH alkyl binding residues are shown here which matched similar positions when while comparing with MupB and TmlB.

Figure S.9.10. Phylogenetic analysis of MupB, TmlB and other ketosynthases (KSs). The Neighbor-Joining method was used for tree construction. The tree is represented in a traditional rectangular way (inset shows the radial distribution of the phylogenetic tree).

```
                 1        10        20        30        40        50
                 .         .         .         .         .         .
MupB    .....MRTA..YLASTAYVLGEQAHDYRDAASFEAVCKQHSMPDFASVFGWGTYW....RTTR
TmlB    .....MGQS..CLHSAKYFLGEVERQYHEVEVFQSVLAEHGLPNMPKLFAWGGYR....TSTK
ChlB6   .....MKVSGIHVWSIGLYHATQ.....WMSPDEARAAGLCDDEFPE.GDLKGV...QIAPGI
DpsC    MSVPQGGAPGDLYVAGCGVWLPP......PVTTEQALAAGHCDRRLASSTRMLSV...AVADKE
CerJ    .....MRWENLFVSGVAAWLPP......LSTAQDAVMAGLLDPARSKLRGIESVTVASDAEED
PtmR    .....MRTPGIFLAATGAYLPE......RTSVEDAVRRGWYERERMESCGWRNV...AVADGI
AvinN   ....MQIADLYIGALGVFVPP......CVSVEWAVERGLYPAEEAEAHELGGV...AIAGDV


                 60        70        80        90        100       110
                 .         .         .         .         .         .
MupB    SVGELLVESVASTLASSGLRGCDIDALVVCSSNFESGQ.VVDYLPLLRELQLARAFPLGVTWG
TmlB    TIFEMAYESMKLTLEAAQVDAQDVGLVIICSSDFNPVHETQSYSQLLVSLGMKYAFPMGVTVS
ChlB6   PPVDMAVAAAREAIGRAGVAPAGIDLLVHAPCFWNGPEEWSPAGYLLRELGCPDG.SGQVLDQ
DpsC    TPAEMAALAAQTALDRSGVAPAHVDLVLHASLYFQGHHLWAPSSYVQRVAVGNRC.PAMEVRQ
CerJ    APPRMAARAARAALGRGDVDPADVSLVLHSSLWFQGIDLWPAASYVAHEAVGRHV.PAFGLAQ
PtmR    SAPDMAVAAVRQAVARSGLSRDDIDLLVHSCAYHQGPDGWSAPHYILRATLGTPV.PALTVEQ
AvinN   PPPEMALRAAQQAVKRWGGSPTEFDLLLYASTWHQGPDGWPPQAYLQRHLVGGDM.LALEIRQ


                 120       130       140       150       160       170
                 .         .         .         .         .         .
MupB    DCTMLLAGLEVARAQVL.AGLD.NVLVVSA.NRIE...DEAFRFQ.HYALFSDGAASCLVTSG
TmlB    DCANLLSALDMTKQLIDSDKYP.HIMLVTS.NKIS...DERHRFQ.DYALFSDGAASIMISNK
ChlB6   GCNGMLAALEMAAGWLTLRGGDATALLTTATMVDGSPTIDRWRSAGYGIAIGDGAGAIVLGRR
DpsC    VSNGGMAALELARAYLLAAP.DRVAALITTGDRMHPGPGFDRWSSD.PGTVYADGGTALVLSRQ
CerJ    RCNGGMGAIELAGAYLGSGIGAGHAALLTTGDRFAGPRIDRWNSV.DVTMYGDGAAALVLSTR
PtmR    GCNAFLAALEMATQYLLCAP.TRSGAVVSAADNFGAPSVDRWHAH.RDSVLADAGAAVVLSKR
AvinN   GCNGVFSALELAVGYLQADP.NRTSALIVAADNYGTPLIDRWRMG.PGFIGGD.ALPPWCCKR
        ★


                          180       190
                          .         .
MupB    RRR.......GFEMLGSLARSNAGL.AH.....................DP.....KEDDT
TmlB    KTLNRIP...EDGLNVVDSQINAQLQLKVA......................GE......DIDDT
ChlB6   PGLAEISAVNSTTVPEMEQFHRGALPLIEPTPPLRPKVDVLARARENAALGAQNGLELQRLQI
DpsC    GGFARLRSLVTVSEPVLEGMHRGGHPFGPPSPEEQRAVDLDAHKRAYVAEAGSSFS..VARVS
CerJ    DGFARVLSTATGVDNSLEILARGDEPFAPHPVEPSPVADLGTRTVRGAELADLPDL..THRYI
PtmR    SGWAELRAVESVSLPQFEILNRGHAPIFPPALTLGKKLDMNEHLEAMVAELGPRASEIVEEYG
AvinN   PGFARLCSVASKGLPEIESLHRGDEPLFPPSITRGRATDFSARIGQQFATRSPAFVAMAE.IQ


                 200       210       220       230       240       250
                 .         .         .         .         .         .
MupB    RLFREVHEQFMHRQQINTADLEQVVCSNVFLPV.LKIKEGRQGVSGTQLYVDNVTRVGHCFSA
TmlB    PLYTATTQQLVDRNQITISDLSKVFSNNLYMPV.ITLKEGGIGVSKSQLYLDNIVRFGHCFSA
ChlB6   QAYQKAMARTLDDVGIGPEDLAKVLFAHVGGAQTDAIVMQQLGLPLAKSTWDHGRSIGHVGAS
DpsC    AGQEEALTGALEAAGAGLDDISRVVLPHMGWRRLSAAYFNKWHIQPERTTWEFGRRTGHLGGG
CerJ    DLLVAAKTQALEDAGTAIEDIAHAVIPVSR.RGTGHELHDLLGLPDERTSWAYGRTTGHVGAG
PtmR    ASGTKLVEQVASEAGITVPDLSHVLHLGAATTDFLDSHLRPMGLDASLGSVEFFRDVGHAGAA
AvinN   DHMMDVAERALAGAGIGMADVARVSFMNYHAKWSSSAEWRPGGCRCPGPRGNSAGRSATCGAS
                                                                    ★


                 260       270       280       290       300
                 .         .         .         .         .
MupB    DSLINLCDYQARAQQVHGGLVMLTANADGLRCQTLLQRVSDA................
TmlB    DSIINLADYLANSRVAAGEYFALHSSASGLRAQILLQAC...................
ChlB6   DHVYDLEHLITSGELSEGDRLLLVCGTAGFHITSLVLTVGDLSPYERYQRDT.......
DpsC    DPIAGFDHLVGSGRLAPGELCLLVSVGAGFSWSCAVVELLERPSWAAAPAAR.......
CerJ    DQYAGLAHLVENALVQPGDRVLLFGGGAGYTCTAAVVEILRMPAQCMR..........
PtmR    DVGIHLDHLACSGRLAAGDHLLMLSAGPGLMITAAVVTVLETPASHTAAPADDEGSGAW
AvinN   DHLLSMEHIVRTGELAPGDHVLQLATAPGLVVSSAVLQVLESPDWDE...........
        ★
```
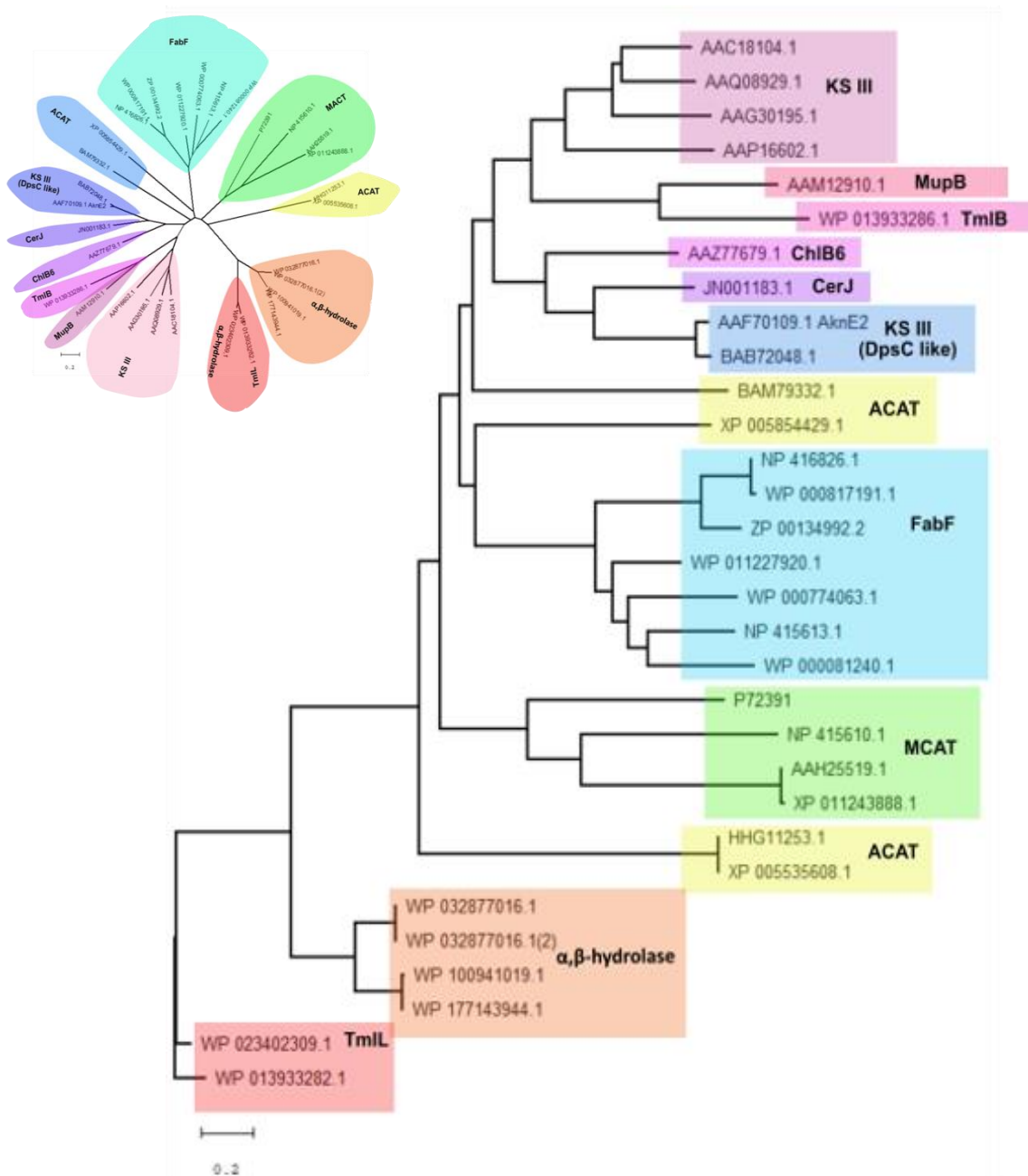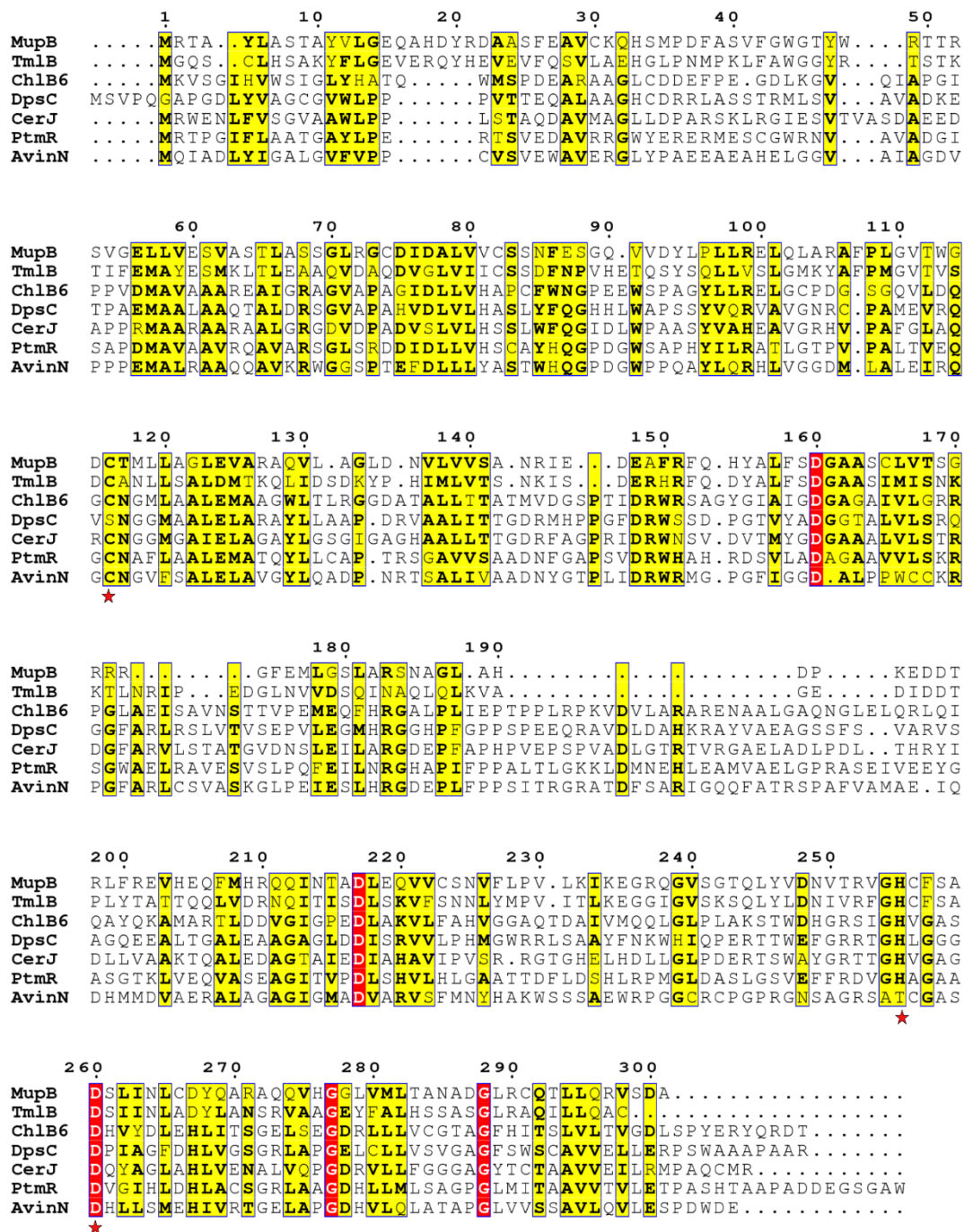
Figure S.9.11. Multiple sequence alignment (MSA) of MupB with reverse esterase enzymes CerJ, ChlB6, PtmR and AviN along with the thiomarinol homologue TmlB. The DpsC amino acid sequence was also compared with these enzymes as it shows extensive similarities with these reverse esterase enzymes but acts as an acyltransferase.

Figure S.9.12. MBE self-loading onto MupB:S224A mutant protein has been observed after overnight incubation (~16 hours) with MBE-Pant.

MupB:S224A

Obs: 36,757 Da
Exp: 36,758 Da
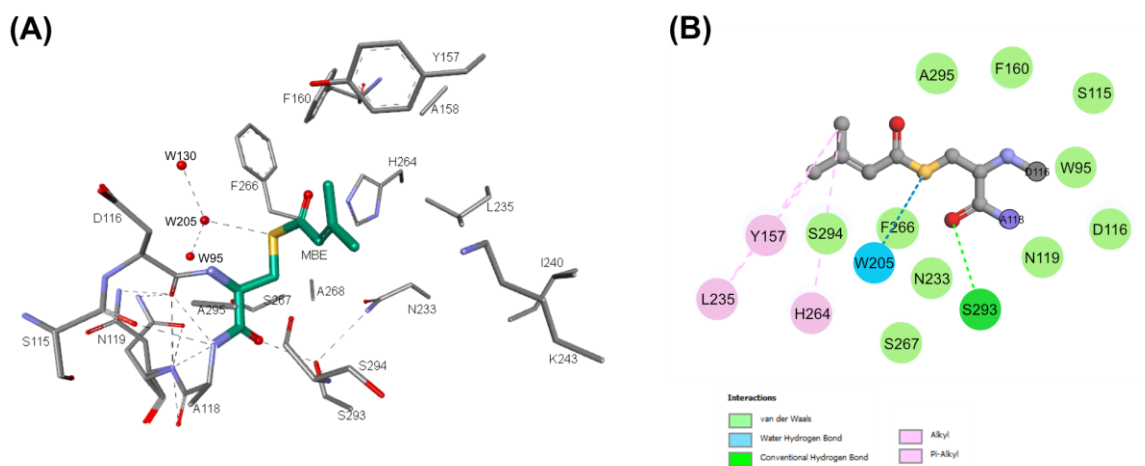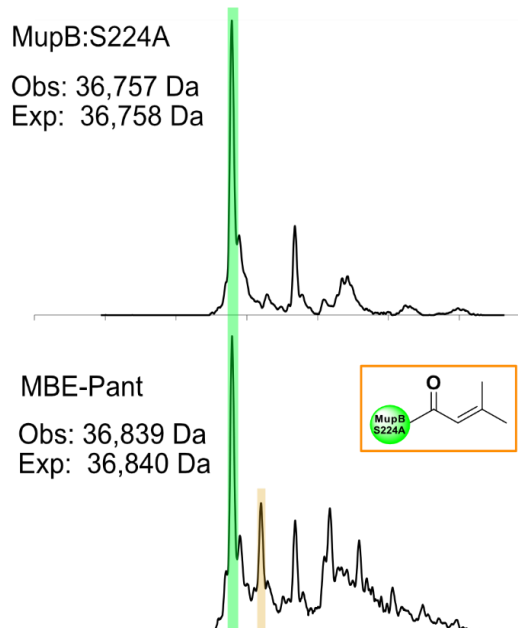
MBE-Pant

Obs: 36,839 Da
Exp: 36,840 Da

(A)

(B)

Figure S.9.13. (A) 3D-representation of the MBE substrate binding pocket architecture (extended) for the MBE-TmlB co-crystal structure. The hydrogen bonding network has been shown usings black dashed line. (B) 2D-representation of the extended substrate binding pocket. Different non-bonding interactions are shown in different colours.
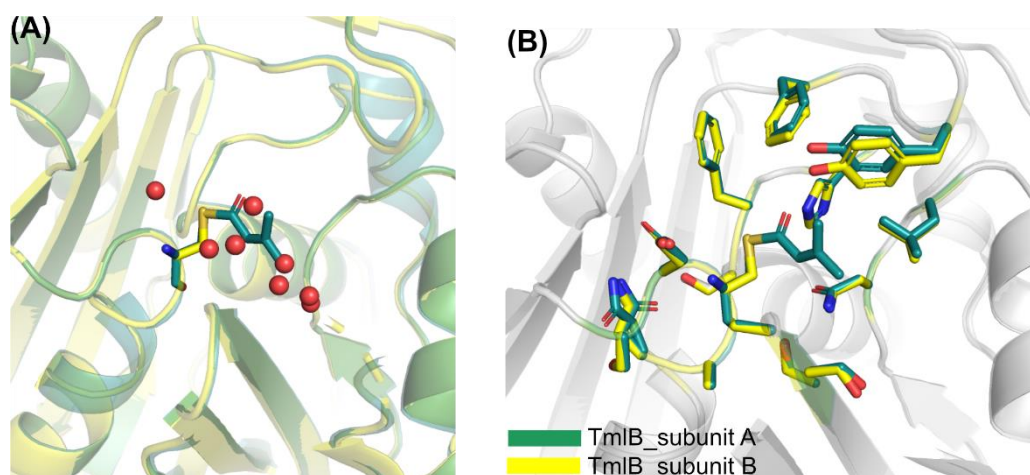
Figure S.9.14. (A) 3D-structural comparison of two identical monomers of TmlB where water molecules (red spheres) in TmlB_subunit B occupy the same position as the MBE substrate in the TmlB_subunit A. (B) Active site residues showt no changes between two subunits.
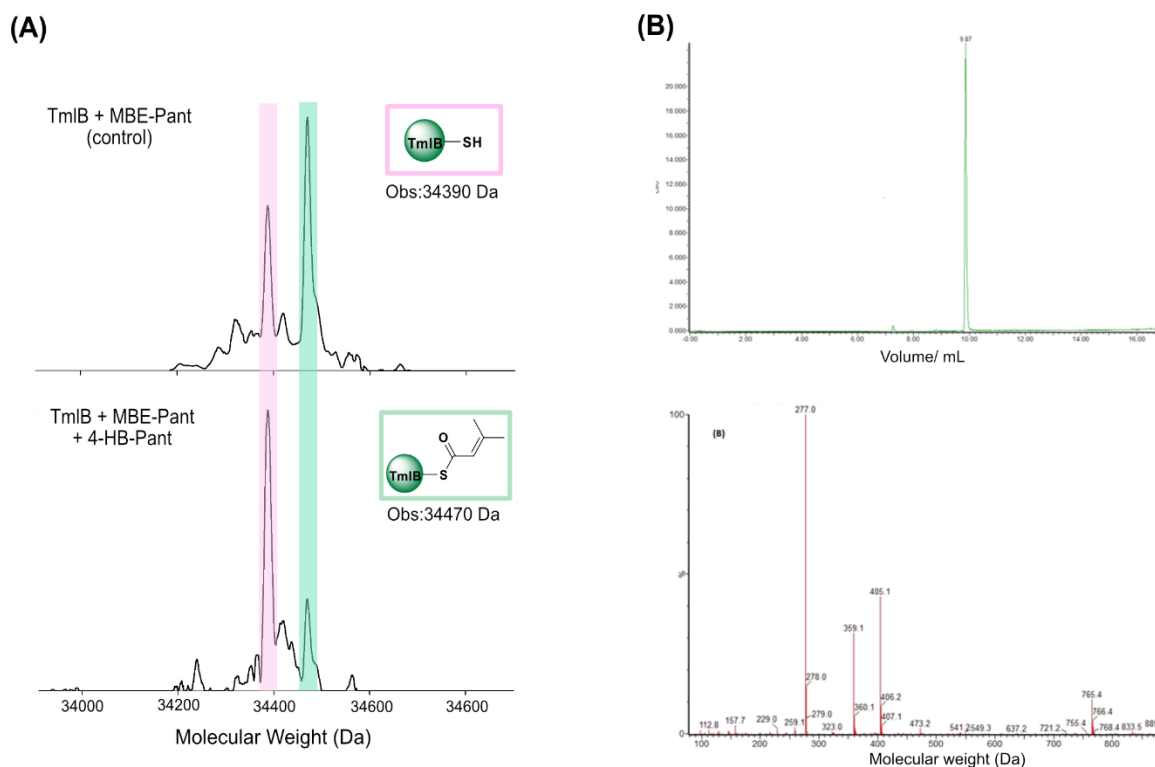

Figure S.9.15. (A) ESI-MS spectra for the esterification reaction of TmlB (does not contain the His$_6$ tag). MBE-TmlB peak intensity decrease is observed after incubating WT-TmlB with MBE-Pant. (B) Analytical LC-MS analysing of the esterification reaction. All the compounds are observed to elute as a single ELSD (Evaporative Light Scattering Detector) spectrum at 10 mL which using negative ESI, reported the extractions of the fragments of MBE-Pant (360 Da), pantetheine (277 Da) and a formic acid adduct of MBE-Pant (405 Da).
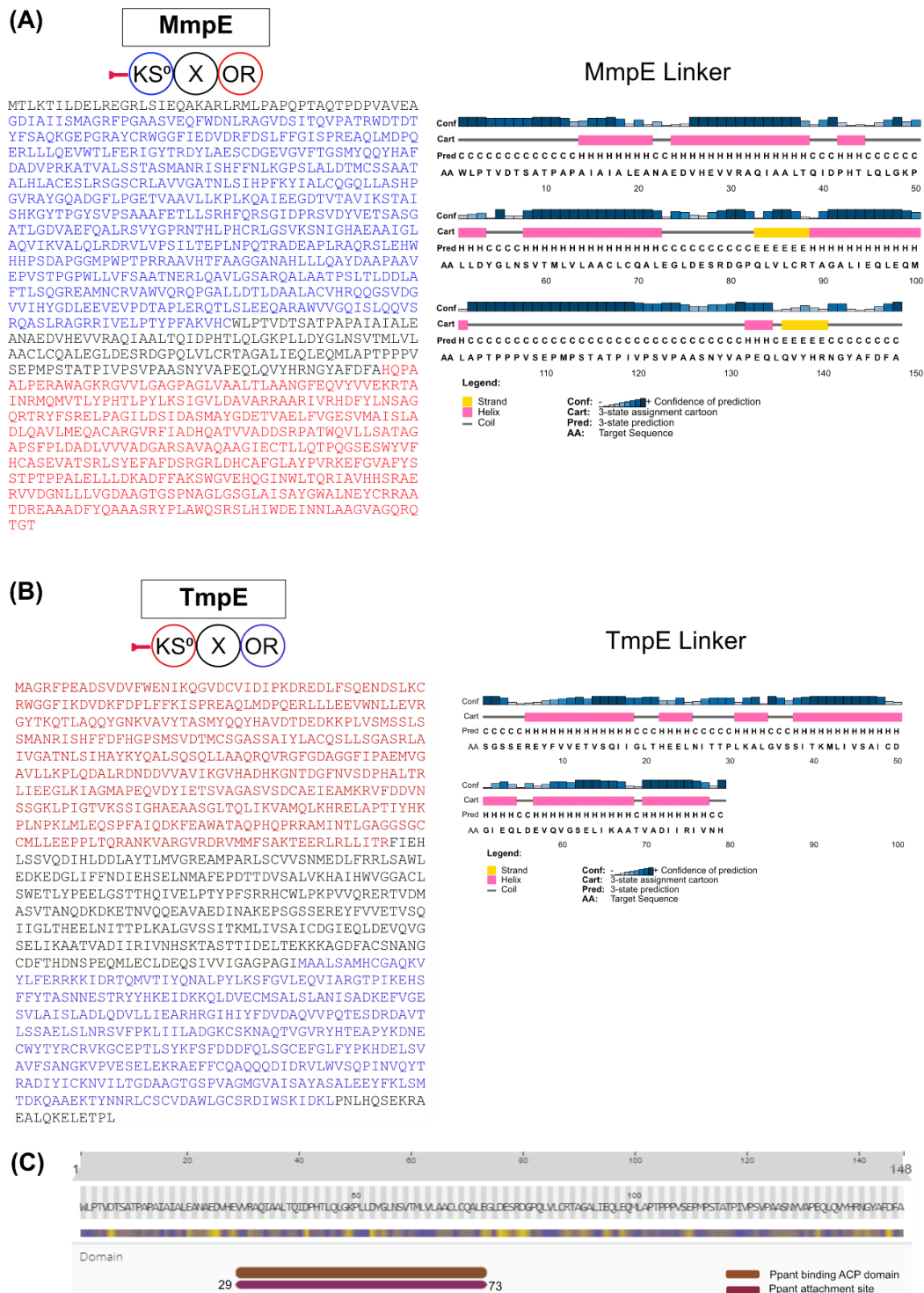
(A)

MmpE

KS⁰ X OR

MTLKTILDELREGRLSIEQAKARLRMLPAPQPTAQTPDPVAVEA
GDIAIISMAGRFPGAASVEQFWDNLRAGVDSITQVPATRWDTDT
YFSAQKGEPGRAYCRWGGFIEDVDRFDSLFFGISPREAQLMDPQ
ERLLLQEVWTLFERIGYTRDYLAESCDGEVGVFTGSMYQQYHAF
DADVPRKATVALSSTASMANRISHFFNLKGPSLALDTMCSSAAT
ALHLACESLRSGSCRLAVVGATNLSIHPFKYIALCQGQLLASHP
GVRAYGQADGFLPGETVAAVLLKPLKQAIEEGDTVTAVIKSTAI
SHKGYTPGYSVPSAAAFETLLSRHFQRSGIDPRSVDYVETSASG
ATLGDVAEFQALRSVYGPRNTHLPHCRLGSVKSNIGHAEAAIGL
AQVIKVALQLRDRVLVPSILTEPLNPQTRADEAPLRAQRSLEHW
HHPSDAPGGMPWPTPRRAAVHTFAAGGANAHLLLQAYDAAPAAV
EPVSTPGPWLLVFSAATNERLQAVLGSARQALAATPSLTLDDLA
FTLSQGREAMNCRVAWVQRQPGALLDTLDAALACVHRQQGSVDG
VVIHYGDLEEVEVPDTAPLERQTLSLEEQARAWVVGQISLQQVS
RQASLRAGRRIVELPTYPFAKVHCWLPTVDTSATPAPAIAIALE
ANAEDVHEVVRAQIAALTQIDPHTLQLGKPLLDYGLNSVTMLVL
AACLCQALEGLDESRDGPQLVLCRTAGALIEQLEQMLAPTPPPV
SEPMPSTATPIVPSVPAASNYVAPEQLQVYHRNGYAFDFAHQPA
ALPERAWAGKRGVVLGAGPAGLVAALTLAANGFEQVYVVEKRTA
INRMQMVTLYPHTLPYLKSIGVLDAVARRAARIVRHDFYLNSAG
QRTRYFSRELPAGILDSIDASMAYGDETVAELFVGESVMAISLA
DLQAVLMEQACARGVRFIADHQATVVADDSRPATWQVLLSATAG
APSFPLDADLVVVADGARSAVAQAAGIECTLLQTPQGSESWYVF
HCASEVATSRLSYEFAFDSRGRLDHCAFGLAYPVRKEFGVAFYS
STPTPPALELLLDKADFFAKSWGVEHQGINWLTQRIAVHHSRAE
RVVDGNLLLVGDAAGTGSPNAGLGSGLAISAYGWALNEYCRRAA
TDREAAADFYQAAASRYPLAWQSRSLHIWDEINNLAAGVAGQRQ
TGT

MmpE Linker

(B)

TmpE

KS⁰ X OR

MAGRFPEADSVDVFWENIKQGVDCVIDIPKDREDLFSQENDSLKC
RWGGFIKDVDKFDPLFFKISPREAQLMDPQERLLLEEVWNLLEVR
GYTKQTLAQQYGNKVAVYTASMYQQYHAVDTDEDKKPLVSMSSLS
SMANRISHFFDFHGPSMSVDTMCSGASSAIYLACQSLLSGASRLA
IVGATNLSIHAYKYQALSQSQLLAAQRQVRGFGDAGGFIPAEMVG
AVLLKPLQDALRDNDDVVAVIKGVHADHKGNTDGFNVSDPHALTR
LIEEGLKIAGMAPEQVDYIETSVAGASVSDCAEIEAMKRVFDDVN
SSGKLPIGTVKSSIGHAEAASGLTQLIKVAMQLKHRELAPTIYHK
PLNPKLMLEQSPFAIQDKFEAWATAQPHQPRRAMINTLGAGGSGC
CMLLEEPPLTQRANKVARGVRDRVMMFSAKTEERLRLLITRFIEH
LSSVQDIHLDDLAYTLMVGREAMPARLSCVVSNMEDLFRRLSAWL
EDKEDGLIFFNDIEHSELNMAFEPDTTDVSALVKHAIHWVGGACL
SWETLYPEELGSTTHQIVELPTYPFSRRHCWLPKPVVQRERTVDM
ASVTANQDKDKETNVQQEAVAEDINAKEPSGSSEREYFVVETVSQ
IIGLTHEELNITTPLKALGVSSITKMLIVSAICDGIEQLDEVQVG
SELIKAATVADIIRIVNHSKTASTTIDELTEKKKAGDFACSNANG
CDFTHDNSPEQMLECLDEQSIVVIGAGPAGIMAALSAMHCGAQKV
YLFERRKKIDRTQMVTIYQNALPYLKSFGVLEQVIARGTPIKEHS
FFYTASNNESTRYYHKEIDKKQLDVECMSALSLANISADKEFVGE
SVLAISLADLQDVLLIEARHRGIHIYFDVDAQVVPQTESDRDAVT
LSSAELSLNRSVFPKLIILADGKCSKNAQTVGVRYHTEAPYKDNE
CWYTYRCRVKGCEPTLSYKFSFDDDFQLSGCEFGLFYPKHDELSV
AVFSANGKVPVESELEKRAEFFCQAQQQDIDRVLWVSQPINVQYT
RADIYICKNVILTGDAAGTGSPVAGMGVAISAYASALEEYFKLSM
TDKQAAEKTYNNRLCSCVDAWLGCSRDIWSKIDKLPNLHQSEKRA
EALQKELETPL

TmpE Linker

(C)

Figure S.9.16. Bioinformatic analysis and identification of a missing ACP. (A) AA sequence of modular MmpE, PsiPred of unannotated X region (148 AA) showing four ordered α helices. (B) AA sequence of modular TmpE, PsiPred of the unannotated X region (148 AA) showing four ordered α helices. (C) Pfam analysis of MmpE Linker amino acids (148 AA) using InterPro bioinformatics tool, indicated that the MmpE linker has a conserved Ppant attachment site (29-73 AA).
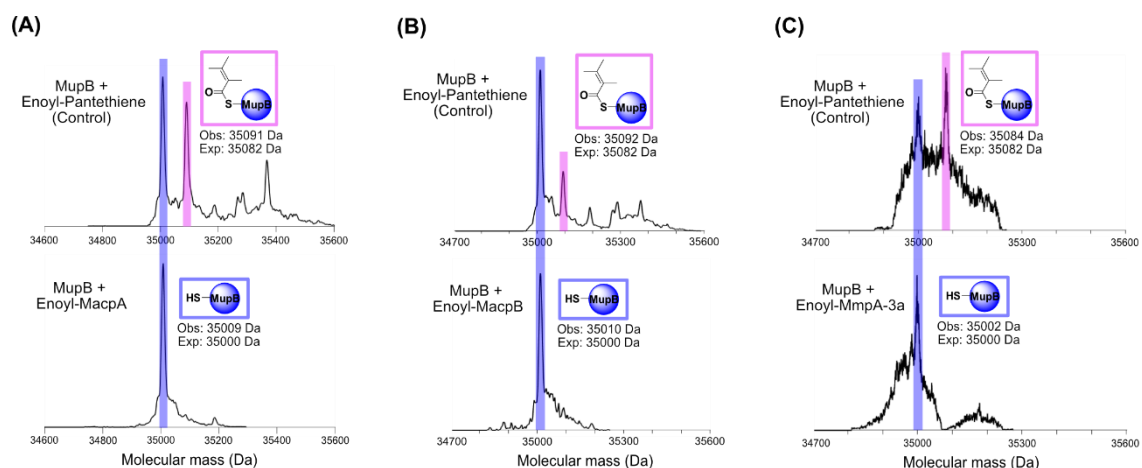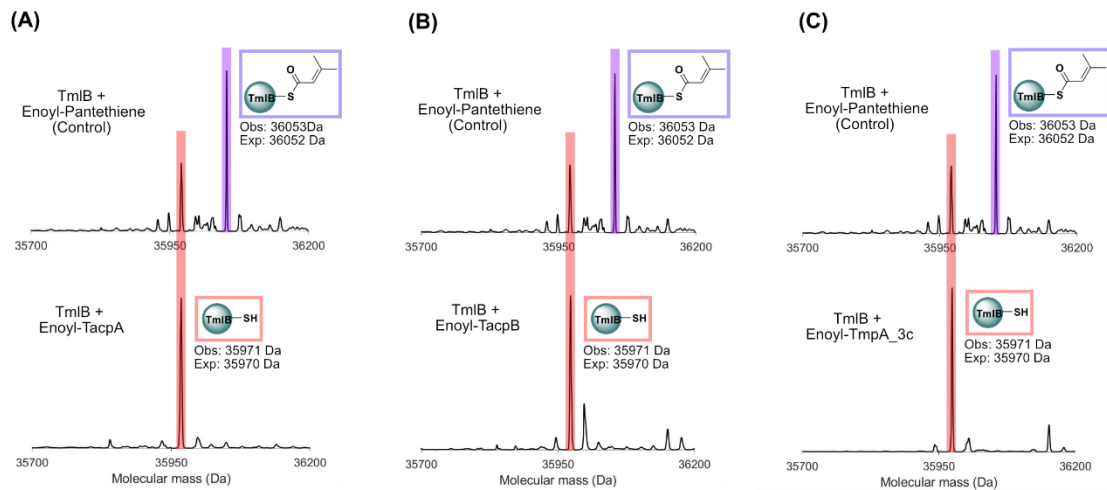
Figure S.9.17. *In vitro* acyltransferase (AT) assay of Mup_ACPs. (A) Deconvoluted ESI-MS spectrum of MupB (coloured in blue) incubated with E-MacpA (bottom) and compared with control MupB with E-Pantetheine (top), showing an E-MupB peak (coloured in pink) of observed mass 35091 Da (expected 35082 Da). (B) Deconvoluted spectrum of MupB incubated with E-MacpB showed no enoyl transfer and (C) Deconvoluted spectrum of MupB incubated with E-MmpA_3a also showed no enoyl transfer. All the control reactions are show at the top with AT assay reactions at the bottom.

Figure S.9.18. The deconvoluted ESI-MS spectra of acyl transferase (AT) *in vitro* assay result between MBE-MmpE_ACP and MupB. In the presence of MupB (observed mass: 35000 Da), MBE-MmpE_ACP has been observed to be completely hydrolysed, producing holo-MmpE_ACP which then forms a MupB-holo_MmpE_ACP cross-linked complex of observed mass 48344 Da.

Figure S.9.19. *In vitro* acyltransferase (AT) assay of Tml_ACPs. (A) Deconvoluted ESI-MS spectrum of TmlB (coloured in red) incubated with E-TacpA (bottom) and compared with control TmlB with E-Pantetheine (top), showing an E-TmlB peak (coloured in violet) of observed mass 36053 Da (expected 36052 Da). (B) Deconvoluted spectrum of TmlB incubated with E-TacpB showed no enoyl transfer and (C) deconvoluted spectrum of TmlB incubated with E-TmpA_3c also showed no enoyl transfer. All the control reactions are shown at the top with AT assay reactions at the bottom.
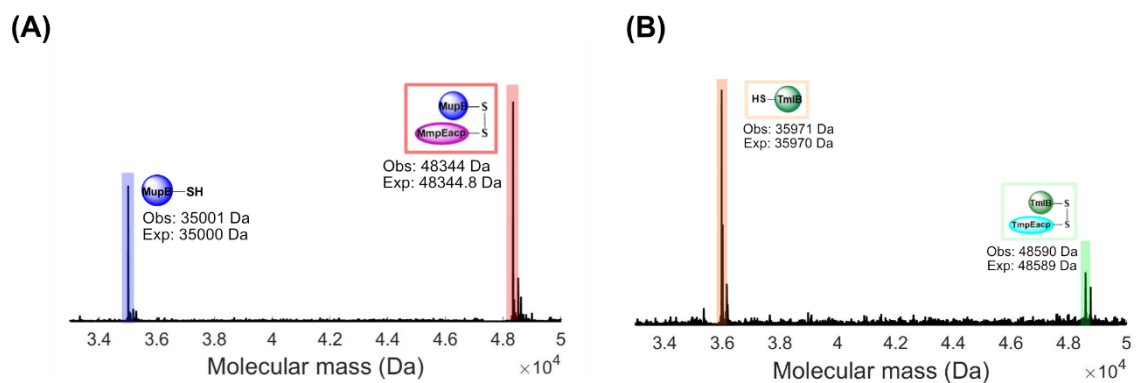
Figure S.9.20. Crosslinked complex generation of (A) MupB and holo-MmpE_ACP and (B) TmlB and holo-TmpE_ACP. In both cases MupB and TmlB remained unreacted to a large extent.
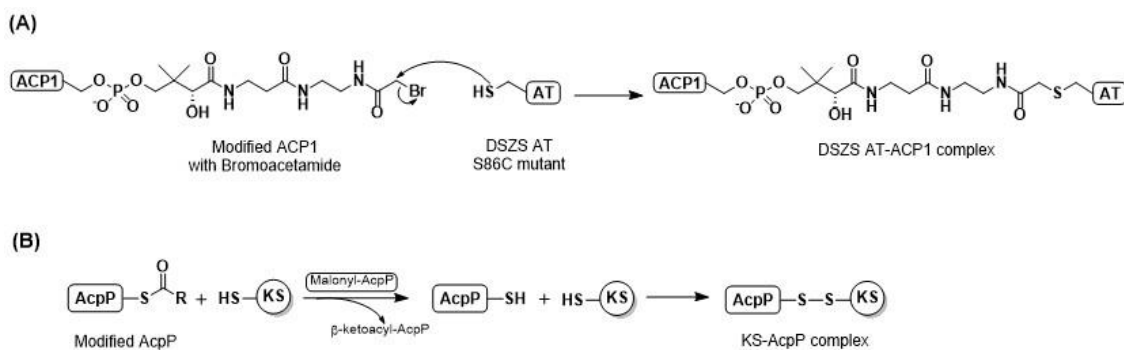
Figure S.9.21. (A) Proposed reaction mechanism of the cross-linking reaction to generate a covalent DSZS AT-ACP1 complex in the disorazole biosynthetic pathway. (B) Proposed KS-AcpP complex formation mechanism demonstrating the production of holo_AcpP while producing β-ketoacyl-AcpP, which is then mixed with the partner KS proteins (FabB and FabF) to generate a cross-linked complex.
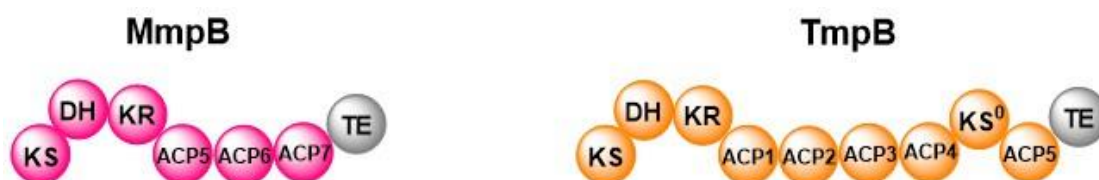
Figure S.9.22. Model representation of MmpB and TmpB mega-enzymes showing that TmpB has one extra KS and two extra ACPs than MmpB.
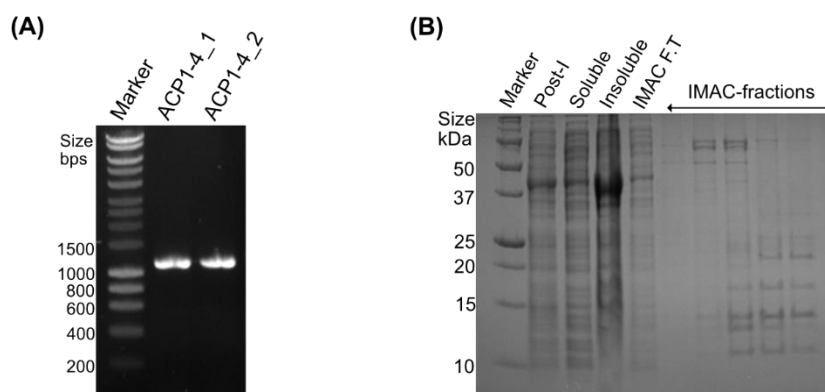
Figure S.9.23. (A) 1% agarose gel representing successfully cloned linear PCR products of TmpB_ACP1-4 showing bands of the correct mass (~ 1200 bps). (B) Purification of TmpB_ACP1-4 followed by SDS-PAGE gel. The insoluble sample has a thick band ~ 40 kDa which indicates that TmpB_ACP1-4 expressed in the insoluble pellet (expected mass: 42546 Da).
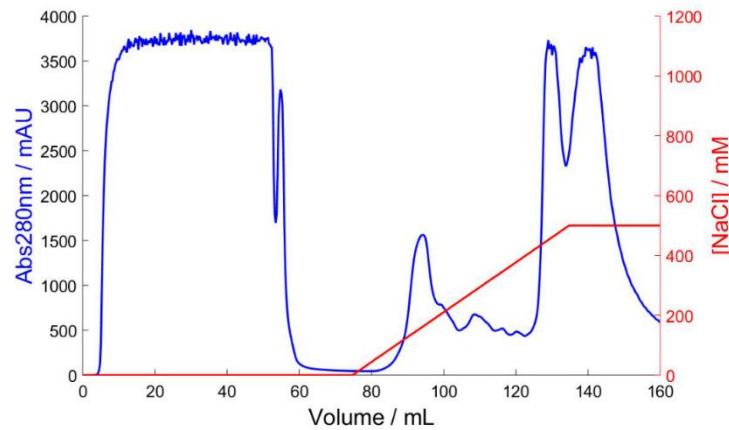
Figure S.9.24. Q FF (5 mL) column chromatography of TmpB_ACP12 protein isolation using a NaCl gradient. TmpB_ACP12 elutes at 160-200 mM NaCl concentration with an absorption peak of ~ 1800 mAU at 280 nm.

(A)

TmpB_ACP1:

ISAVSTLLKVPQSSIDATTAFFDIGFDSVSISELAIILTQKTGVELSPTVFYECTQINELTHYLQ
HHHHHHH------------HHHH---HHHHHHHHHHHHHH-----HHHH----HHHHHHHH-

TmpB_ACP2:

MVSELLKVPESSIDPSTAFADIGFDSVSLGELAATLSDAFGIEISATLFYECTVVAELVDYLLAQLDSDNIE
-HHHHH------------HHHH---HHHHHHHHHHHHHH-----HHHHH----HHHHHHHHHHH-------

TmpB_ACP3:

MSLNDAHIITVISSLLKVPESSIDADTTFEQIGFDSVSLSELAGTLAKRYDIEISTTVFYECTTVGALTQHINAQLGGKEPSTY
--HHHHHHHHHHHHH------------HHHH---HHHHHHHHHHHH------HHHHH----HHHHHHHHHH---------

TmpB_ACP4:

RAGLELSAEITSIICEMLKIGRHDFDDHTPLGQLGFDSVSFAELSQKLSHRYGKDITPTIFYAHETISALV
---HHHHHHHHHHHHH------------HHH----HHHHHHHHHHHHHH-----HHHH----------

MmpB_ACP5:

SDTLRGLVGQILKVDAQEIDDTTAFSDMGFDSVMLTELATAINRTYTLELGTAALFEHPTLQALAAHLQGARTA
HHHHHHHHHH------------HHH----HHHHHHHHHHHH------HHHH----HHHHHHHHH----

(B)

Figure S.9.25. (A) Secondary structure prediction of individual TmpB_ACP (1-4) compared with MmpB_ACP5 where TmpB_ACP1-3 show structural similarities with MmpB_ACP5. (B) MSA of TmpB-ACP1-4 with MmpB_ACP5 shows the highly conserved residues boxed in red and shares the same Ppant binding architecture.
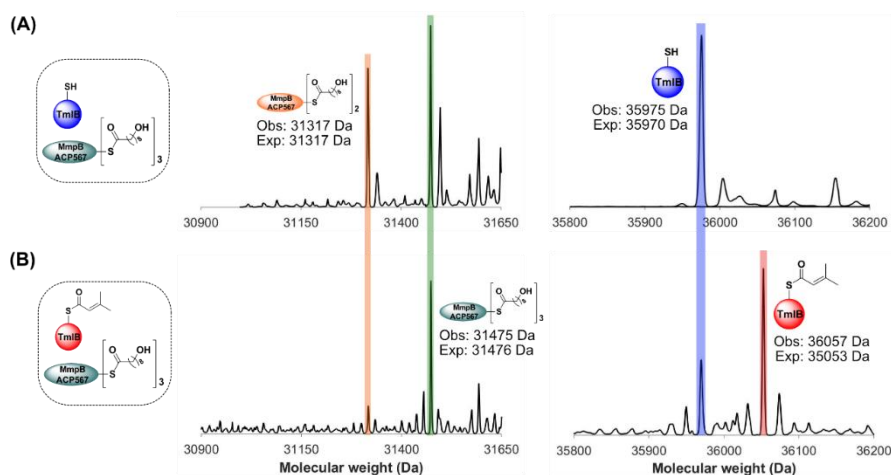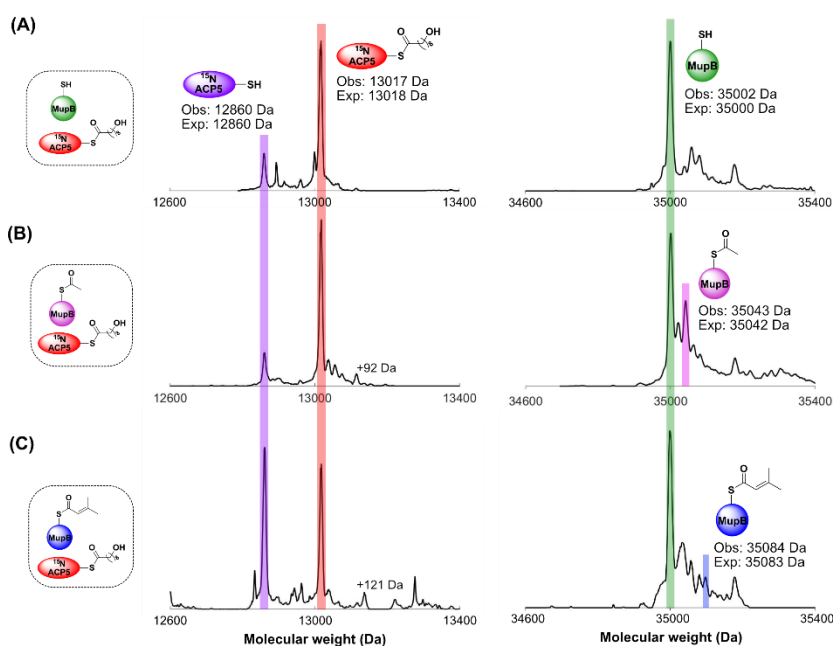
Figure S.9.26. Esterification assay of TmlB: (A) Control reaction between TmlB and [9HN]$_3$-MmpB_ACP567, showing the deconvoluted spectra of [9HN]$_3$-MmpB_ACP567 (green) hydrolysed to produce [9HN]$_2$-MmpB_ACP567 (orange). A spectrum for TmlB indicated an observed mass of 35975 Da. (B) E-TmlB incubation with [9HN]$_3$-MmpB_ACP567 followed by ESI-MS: ACP region shows the conversion of holo-MmpB_ACP567 (yellow) and [9HN]$_2$-MmpB_ACP567. The TmlB region shows a spectrum of E-TmlB (red) along with unreacted WT-TmlB (blue).

Figure S.9.27. Esterification assay of MupB with 9HN-^{15}N-MmpB_ACP5: (A) Control reaction between MupB and 9HN-MmpB_ACP5, showing the deconvoluted spectra of 9HN-MmpB_ACP5 (red) hydrolysed to produce holo-MmpB_ACP5 (violet). A spectrum for MupB (green) is shown with an observed mass 35002 Da. (B) Ac-MupB incubation with 9HN-MmpB_ACP5 followed by ESI-MS: ACP region shows the conversion of holo-MmpB_ACP5 and glycerol adduct at +92 Da. The MupB region shows a spectrum of Ac-MupB (magenta) along with unreacted WT-MupB (blue. (C) E-MupB incubation followed by ESI-MS: ACP region show the conversion of holo-MmpB_ACP5 and tris adduct at +121 Da. The MupB region shows a spectrum of E-MupB (blue) along with unreacted WT-MupB.
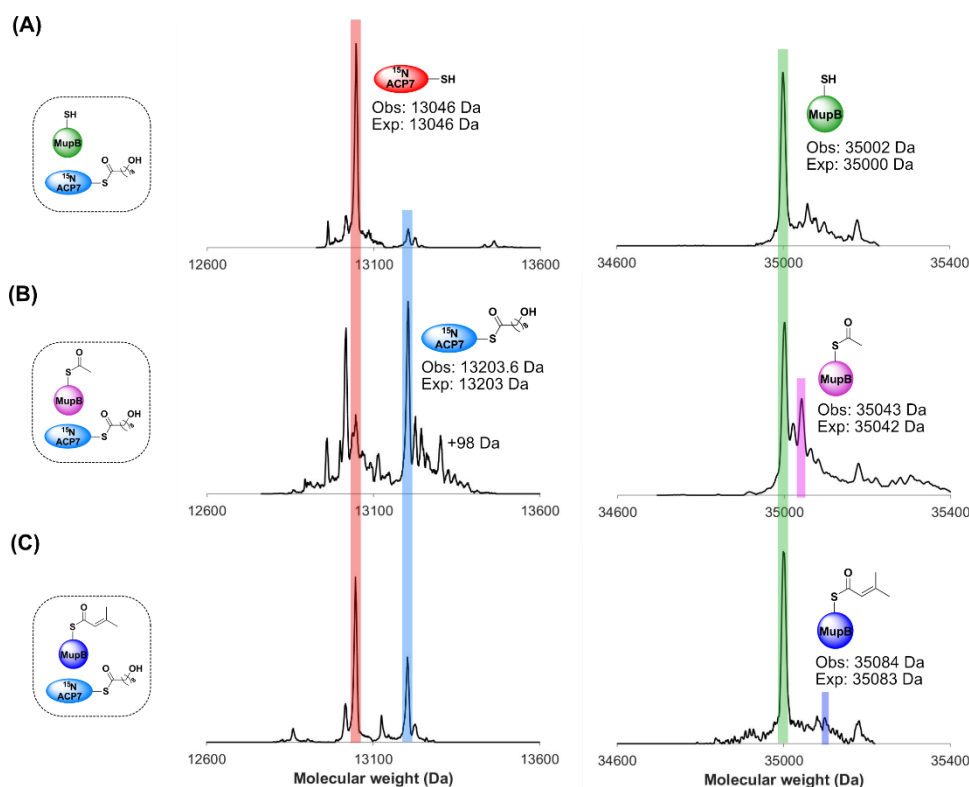
Figure S.9.28. Esterification assay of MupB with 9HN-^{15}N-MmpB_ACP7: (A) Control reaction between MupB and 9HN-MmpB_ACP7, showing the deconvoluted spectra of 9HN-MmpB_ACP (sky blue) hydrolysed to produce holo-MmpB_ACP7(red). A spectrum for MupB (green) indicated an observed mass of 35002 Da. (B) Ac-MupB incubation with 9HN-MmpB_ACP7 followed by ESI-MS: ACP region shows the conversion of holo-MmpB_AC75 and a phosphate adduct at +98 Da. The MupB region shows a spectrum of Ac-MupB (magenta) along with unreacted WT-MupB (blue. (C) E-MupB incubation followed by ESI-MS: ACP region is showing the conversion of holo-MmpB_ACP5. The MupB region shows a spectrum of E-MupB (blue) along with unreacted WT-MupB.
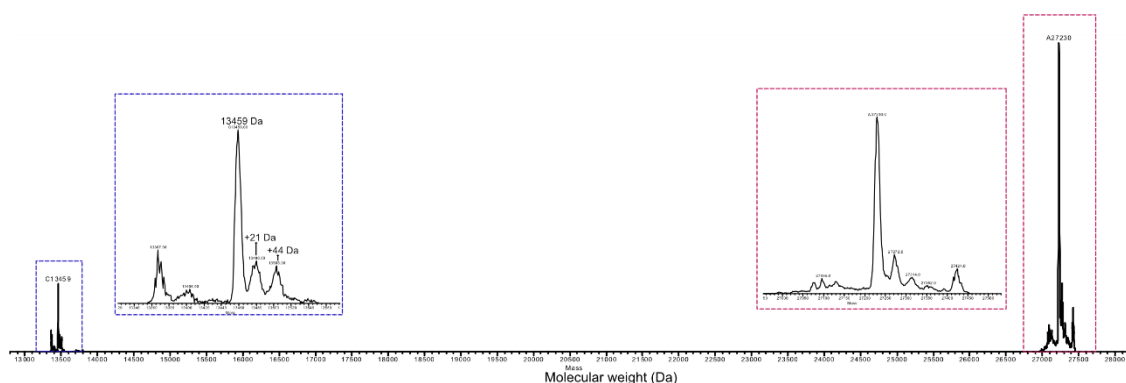

Figure S.9.29. Ac-MupB incubation with 9HN-MmpB_ACP6 followed by ESI-MS: The deconvoluted spectra for the dimer of 9HN-MmpB_ACP6 (observed mass: 27230 Da) and holo-MmpB_ACP6 (observed mass: 13459 Da) along with Na-adducts.
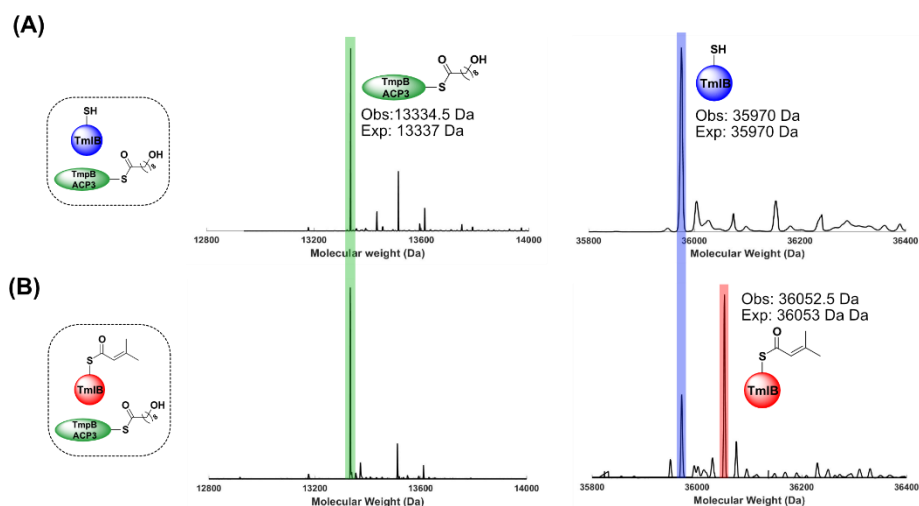
Figure S.9.30. Esterification assay of TmlB: (A) Control reaction between TmlB and 9HN-TmpB_ACP3, showing the deconvoluted spectra of 9HN-TmpB_ACP3 (green). The spectrum for TmlB (without MBE) indicated an observed mass of 35970 Da. (B) MBE-TmlB incubation with 9HN-TmpB_ACP3 followed by ESI-MS: The ACP region does not show any change in the spectrum. The TmlB region shows a spectrum of MBEE-TmlB (red) along with unreacted WT-TmlB (blue).
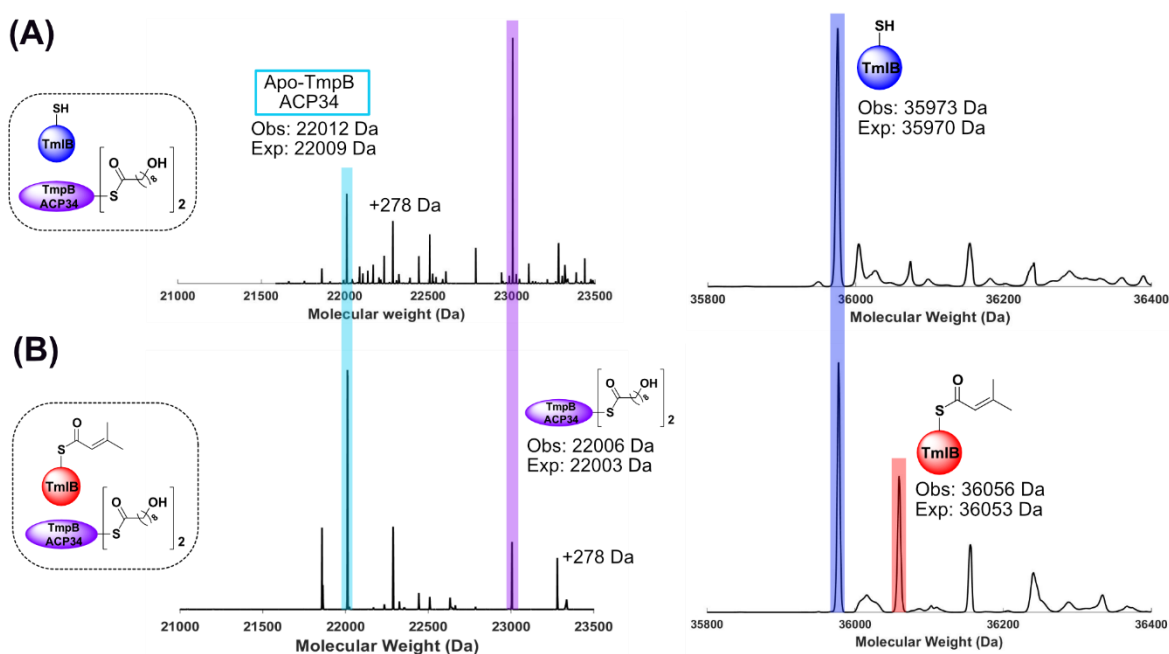

Figure S.9.31. Esterification assay of TmlB: (A) Control reaction between TmlB and [9HN]$_2$-TmpB_ACP34, showing the deconvoluted spectra of 9HN-TmpB_ACP34 (violet), apo-TmpB_ACP34 (cyan) and a spectrum for TmlB (without MBE) indicated an observed mass of 35973 Da. (B) MBE-TmlB incubation with [9HN]$_2$-TmpB_ACP34 followed by ESI-MS: The ACP region shows the conversion into apo-TmpB_ACP34. The TmlB regionshows a spectrum of MBE-TmlB (red) along with unreacted WT-TmlB (blue).