Author:
Rowe, Matthew T

Title:
Fatty acid construction within the biosynthesis of the polyketide antibiotic mupirocin

Fatty acid construction within biosynthesis of the polyketide antibiotic mupirocin.

Matthew Thomas Rowe

A thesis submitted to the University of Bristol in accordance with 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

October 2021

Abstract

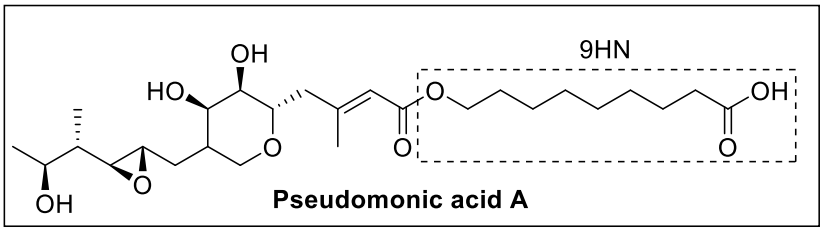
The biosynthesis of polyketide natural products utilises sets of enzymes known collectively as a polyketide synthases (PKSs). These highly orchestrated assemblies control a series of concatenated condensation and reductive cycles in addition to specific tailoring modifications. The precise order and mechanisms of these individual steps can be challenging to decipher, especially in *trans*-AT PKS systems which are typified by the action of many non-covalently attached interacting domains (*trans*-acting domains). The polyketide antibiotic mupirocin consists of a mixture of pseudomonic acids with pseudomonic acid A making up over 90% of the mixture. The mupirocin biosynthetic pathway consists of a combination of type I multifunctional *trans*-AT PKS and *trans*-acting tailoring domains, including five free-standing acyl carrier proteins (ACPs) encoded for within a complex 75 kb gene cluster. Understanding the role of each enzyme and ACP within this complex biosynthetic system may hold the key to deciphering the complete pathway.

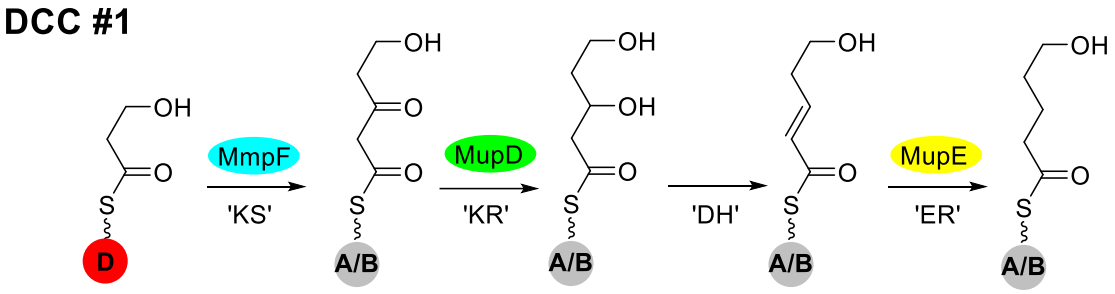
The mixture of pseudomonic acids which comprise mupirocin consist of a polyketide-like monic acid moiety esterified to a fatty acid-like 9-hydroxynonanoic acid (9HN) tail. In previous work, the initial steps of 9HN biosynthesis were shown to begin with the construction of 3-carbon alcohol moiety, 3-hydroxypropanoate (3HP) attached to mAcpD. This unusual starter unit is proposed to undergo three rounds of KS-catalysed condensation and reduction to produce 9HN. In this work, the enzymes responsible for these key condensation and tailoring reactions were cloned, expressed and purified. A strategy that employed ^{1}H-^{15}N HSQC NMR screening experiments to identify interacting partners for the ACPs provided initial evidence that the KS domain, MmpF, interacted with mAcpD. Subsequent *in-vitro* reconstitution and mass spectrometry (MS) analysis of the MmpF-catalysed condensation with 3HP-mAcpD confirmed that MmpF catalysed the first condensation reaction (DCC #1) of 9HN biosynthesis. The two ACPs with unassigned roles in the pathway (mAcpA and mAcpB) were found to deliver malonyl extender units for this reaction. Downstream enzyme MupD was able to reduce the resultant 5-hydroxy-3-oxopentanoyl thioester to the corresponding 3,5-dihydroxypentanoyl thioester intermediate, confirming its role as a KR domain.

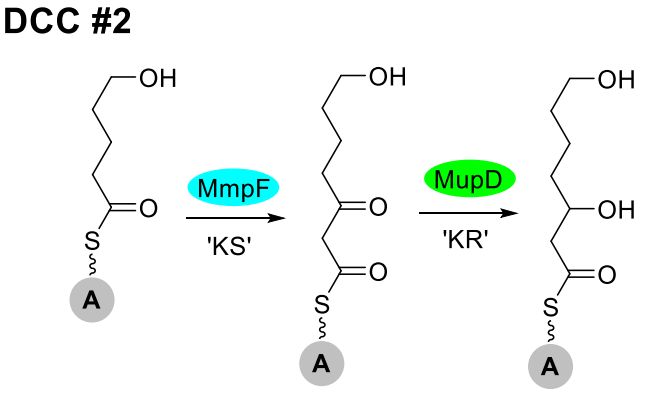
The putative medium chain dehydrogenase domain MupE was confirmed to be an ER domain through *in-vitro* assays with ACP bound mimics of the proposed natural substrate. Both MupD and MupE were found to have strong selectivity for mAcpB substrates. A candidate for the missing dehydratase activity could not be identified using *in vitro* assays and purified DH domains from the PKS.

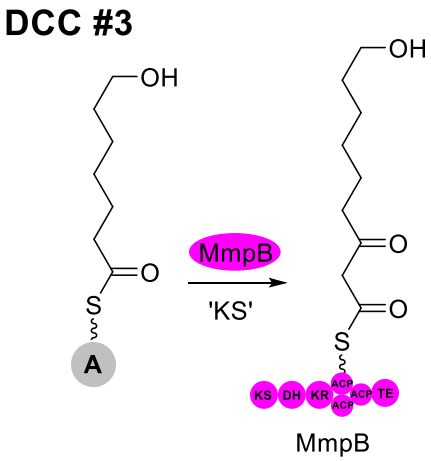
The role of the KS domain from the multidomain module MmpB was tested through *in vitro* assays with authentic substrates. MmpB_KS was found to accept 7-hydroxyheptanoate (7HH) and 6-hydroxyhexanote (6HH) most efficiently. In addition, MmpB_KS was able to catalyse elongation of 7HH utilising malonyl delivered only by the individual ACP domains from MmpB_ACP (not mAcpA or mAcpB). This confirms that MmpB_KS catalyses the final condensation reaction (DCC #3) in 9HN

biosynthesis. Only mAcpA was found to deliver 7HH during this condensation reaction. The ability of MmpF and MmpB_KS to catalyse the remaining 5-carbon to 7-carbon condensation reaction (DCC #2) was tested using 5-hydroxypentanoate (5HP) substrate. Only MmpF, utilising malonyl delivered by mAcpA was able to turnover condensation of 5HP confirming that MmpF catalyses the first and second condensation reactions with the final condensation catalysed by MmpB_KS.









Acknowledgements

I have been very lucky to live in Bristol for 8 years during my undergraduate and postgraduate studies. The School of Chemistry at Bristol has been a fantastic community that has allowed me to make the most of my time as a student and a researcher. I have been very fortunate to conduct research on a project that I have found genuinely fascinating surrounded by very talented people. Firstly, I would like to thank my supervisor Matt for his guidance and support throughout. I still remember our first meeting on the project – being handed a copy of the mupirocin pathway (which I still have today) and thinking "what have I got myself in for"! I have appreciated your constant support throughout all the challenging times – even picking me up from the hospital after my knee surgery! Project-wise, I have been very grateful to have been supervised by someone as passionate about the project as you. Many overrun meetings debating the sequence of events in fatty acid biosynthesis on the whiteboard in your office are a happy memory of this. 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 Paul R and Alice P for your support during my SWBio DTP second rotation project. During this time, I was able to work in C101 (although mainly in the play-pen) and was able to experience life in Biochemistry Department. I was able to travel up to Diamond Light Source and get some nice diffraction patterns (even if it did mean I missed the 2018 World Cup Final). I would also like to thank Chris Willis for being a further port of call if problems arose and for providing continued support and guidance through our regular mupirocin meetings.

I would like to thank everyone on the mupirocin project for their support throughout my time in the lab. Specifically, I would like to thank Ash for your constant support and for being a great sounding board for ideas. You also provided continuous help throughout the project and vital planning which drove the project forwards for which I am very grateful. I would like to thank Nahida – I remember meeting you for the first time in the postgraduate lounge and knowing we were going to get on immediately. Thanks for being there to share ideas and being a constant positive presence in the lab. I would like to thank Angus, certain parts of this project simply would not have happened without your support on the synthetic side of things. Again, thank you for being a great person to bounce ideas off too. The nickname 'Hawaii-5-Rowe' was less desired, but I suppose it is better than Snowflake. Thank you to Paul W for showing me the ropes in N318, guiding me through my Master's project and helping to persuade me I was good enough to do a PhD! 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 (pub trips, Christmas dinners etc.). Thank you to Chris Williams for your constant support and for putting up with my silly questions throughout my time in the lab. Your support on the NMR and X-Ray Diffraction side of things were vital to the completion of the project. Thanks to Chris Arthur, Paul Gates

and John Crosby for providing support with Mass Spectrometry throughout my time in Bristol. Thanks to Burnham and Keith for the entertainment throughout my project. The Karaoke trips, snow days and after work pub trips (lab whiskey!) were a constant source of happiness throughout the PhD. I still vigorously maintain that the steak taco we had circa 2019 was the best dinner I ate that year. Thank you to Erik for constant history chats although I still maintain that Belgium does indeed exist. Thanks for teaching me very complicated board games and very complicated computer games. Thanks to Marija for always being there for a 6pm motivational chat when required. Thanks to everyone else; Alice B, Matt G, Lina, Amy M, Elise W, Nui, Annabel P and the many project students over the years for making the lab such an enjoyable place to be.

I would like to thank the SWBio DTP and BBSRC for funding my research. I have thoroughly enjoyed being part of this cohort of students and it has given me a further support network that was vital to my studies.

Finally, I need to thank everyone else who has supported me throughout my project. Thanks to my family for your unwavering support throughout my project. Thanks to my flatmate Ashley for the great accommodation and pep-talks! And finally, thank you to my partner Lucy – it has been an amazing 5 years. Thank you for being my rock and listening to me ramble incessantly about the role of various proteins within the pathway. I honestly don't think I could have done this without your support so thank you. We've been fortunate to have had some great times together and I hope they continue long into the future.

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 the supervision of Professor M. P. Crump between October 2017 and June 2021. 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.

Matthew Thomas Rowe

October 2021

Abbreviations

3HP – 3-Hydroxypropanoate

4HB – 4-Hydroxybutyrate

5HP – 5-Hydroxypentanoate

6HH – 6-Hydroxyhexanoate

7HH – 7-Hydroxyheptanoate

9HN – 9-hydroxynonanoic acid

ACP – Acyl Carrier Protein

ArCP – Aryl Carrier Protein

AT – Acyltransferase

ATP – Adenosine Tri-Phosphate

BGC – Biosynthetic Gene Cluster

BLAST – Basic Local Alignment Search Tool

BSA – Bovine Serum Albumin

CoA – Coenzyme A

CSP – Chemical Shift Perturbation

Da – Dalton

DH – Dehydratase

DIOA – 2-deoxyscylloinosamine

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

ER – Enoylreductase

ESI-MS – Electrospray Ionisation Mass Spectrometry

FAAL – Fatty Acyl: Adenylate Ligase

FAS – Fatty acid Synthase

FDA – Food and Drug Administration

FPLC – Fast Protein Liquid Chromatography

GTP – Guanidine triphosphate

HMG – 3-hydroxy-3-methyl-glutaryl

HPLC – High Performance Liquid Chromatography

HSQC – Heteronuclear Single Quantum Coherence Spectroscopy

HTH – Head-to-Head

HTT – Head-to-Tail

I-TASSER – Iterative Threading ASSEmbly Refinement

IleRS – Isoleucyl tRNA Synthetase

IMAC – Immobilized Metal ion Affinity Chromatography

IPTG – Isopropyl β-D-1-thiogalactopyranoside

JTT – Jones-Taylor-Thornton

KAS – Keto-acyl Synthase

KR – Ketoreductase

KS – Ketosynthase

LB – Lysogeny Broth

MDR – Medium Chain Dehydrogenase

Mmp – Mupirocin Multifunctional Protein

MRSA – Methicillin-resistant *Staphylococcus aureus*

MT – Methyl transferase

NADH – Nicotinamide Adenine Dinucleotide

NADPH – Nicotinamide Adenine Dinucleotide Phosphate

NEB – New England Biolabs

NMR – Nuclear Magnetic Resonance

NRPS – Non Ribosomal Peptidal Synthase

ORF – Open Reading Frame

PA – Pseudomonic Acid

Pant. – Pantetheine

PCR – Polymerase Chain Reaction

PCP – Peptidal carrier protein

PDB – Protein Data Bank

PKS – Polyketide Synthase

Ppant – Phosphopantetheine ejection assay

PPtase – Phosphopantetheinyl transferase

Post-D. – Post Desalt

Post-I. – Post Induction

RPM – Revolutions per minute

SDR – Short Chain Dehydrogenase

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC – Size Exclusion Chromatography

SNac – N-Acetylcysteamine

TCEP – Tris(2-CarboxyEthyl)Phosphine

TE – Thioesterase

TEV – Tobacco Etch Virus

Tmp – Thiomarinol multifunctional protein

Tris – tris(hydroxymethyl)aminomethane

WT - Wild type

Table of Contents

1 Introduction

1.1 Natural Products

The group of chemical compounds produced by living systems (most notably plants, bacteria and fungi) are known collectively as natural products. This intriguing class of compounds contain unique bioactivity and chemical diversity. Throughout human history, natural products have been hugely important in terms of their medicinal properties.[1] The discovery of pollen clusters from a 60,000 year old Neanderthal burial site in northern Iraq may provide the earliest recorded use of natural products for medicinal purposes.[2] The first evidence of human exploitation of natural products was during ancient Mesopotamia some 4,600 years ago. A clay tablet detailing the use of approximately 1,000 plants and plant derived-substances with medicinal properties has been dated to this time period. [3] Oils of the *Cedrus* and *Cupressus sempevirens* species (cedar and cypress) included in this text are still used today for their insecticidal and anti-fungal properties. [4]

The use of plant-based medicinals is documented by nearly every major human civilisation. The "Ebers Papyrus" is an ancient pharmaceutical record detailing over 700 compounds used by the ancient Egyptians in 1500 BC. The Chinese *Materia Medica* is the earliest text detailing the use of plant-based remedies in Ancient China. In the ancient Western world, it was the ancient Greeks that contributed most to the rational development of herbal medicines. Further to this, the isolation of morphine and strychnine from plant extracts in the early 1800s contributed to the idea of a 'pure compound acting as a specific drug'. This paved the way for the first commercial natural product, morphine, sold by apothecary Heinrich Emanuel Merck – whose descendants would later become founders of the eponymous pharmaceutical company. [5]

The use of natural products in the pharmaceutical industry; either directly (such as morphine), or as a template for the development of novel pharmaceuticals (like salicin as a template for aspirin) is widespread. In the early 1900's approximately 80% of all medicines were obtained from plant sources. [6] A recent analysis of FDA (U.S. Food and Drug Administration) approved drugs has concluded that today this figure is likely to be around 60%.[7] The comprehensive use of natural products as medicines has fuelled research into their structure and function. [8]

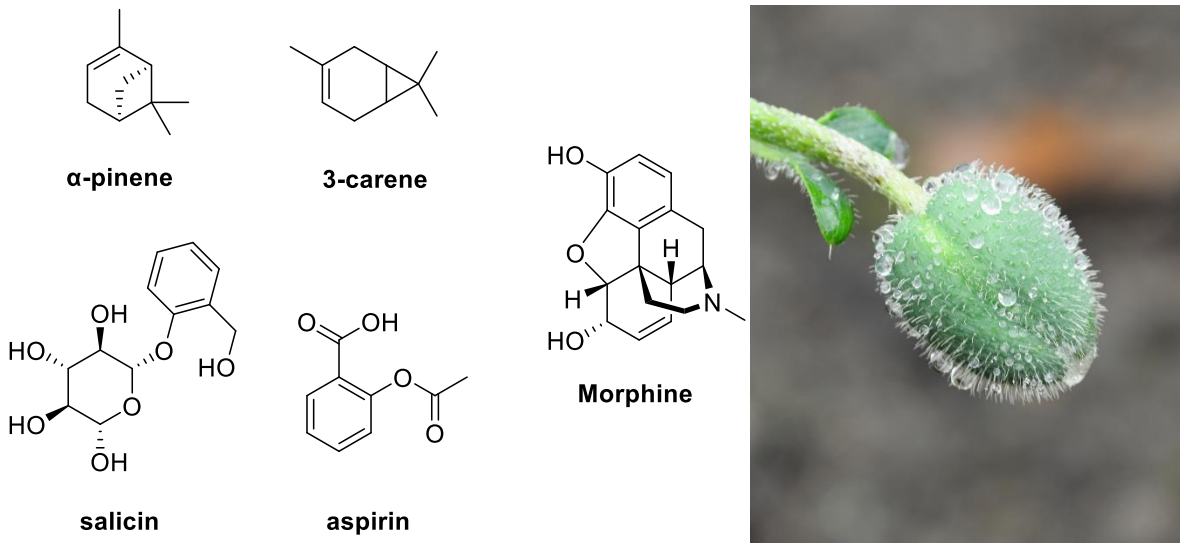


Figure 1: The chemical structures of selected pharmaceutically relevant natural products. α-pinene and 3-carene are two most abundant compounds found in cypress oil. Morphine is isolated from opium poppy (pictured) [9]*. salicin was first isolated from willow bark and was modified to produce the painkiller aspirin.*

Natural products are characterised into two distinct compound classes: primary and secondary metabolites. Primary metabolites are compounds involved in primary metabolism such as the biosynthesis and breakdown of many carbohydrates, proteins, fats and nucleic acids. Primary metabolic pathways include glycolysis: oxidation of glucose to obtain adenosine triphosphate, (ATP) and the Krebs' cycle; the oxidation of acetyl-CoA to produce Guanosine triphosphate (GTP) and other important metabolites.[10,11] Despite the extremely varied characteristics of living organisms, primary metabolic pathways are fundamentally conserved in all organisms, demonstrating their essential nature. Secondary metabolites are a separate group of compounds which comparatively are not essential for the growth, development, or reproduction of an organism. Rather than playing an essential role, these secondary metabolites have evolved to assist the survival of the organism, typically in response to an environmental cue or to act as a defence mechanism. [12] The term secondary metabolite is often used synonymously with natural product to denote pharmacologically active compounds.[13]

One explanation for the potent biological effects of secondary metabolites is long term co-evolution within biological communities. Interacting organisms that have evolved in close proximity will have produced the first natural products to interfere with the growth and proliferation of competing species. As the production of these compounds is advantageous, they became a trait on which natural selection could act, and were retained and improved throughout the course of evolution. [1] For example, a plant species predated by a bacterium will gain a clear evolutionary advantage if it can produce a compound with antibiotic activity. This biological arms race between different species has caused the abundance of biological activity observed in natural product secondary metabolites.[14]

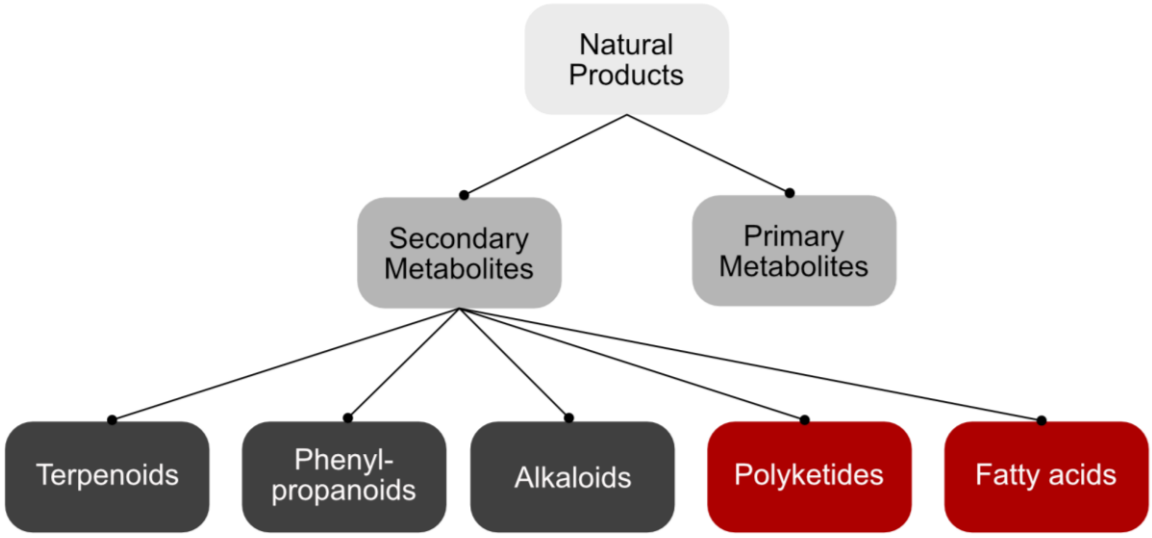


Figure 2: The different classes of natural products. Natural products can be split into primary and secondary metabolites depending on whether the molecule is involved in primary metabolism or not. Secondary metabolites can be further split into classes based on chemical structure and biomolecules utilised in their construction.

The classification of secondary metabolites is further split into different classes based upon their biosynthetic origin. Polyketides and fatty acids are two distinct but related classes of secondary natural products. Structurally, these classes of compounds are divergent. Fatty acids consist of long aliphatic chains containing a typically saturated carbon backbone. Conversely, a more functionally diverse and typically oxidised structure is observed in polyketides. The structures of various polyketides (daunorubicin, cervimycin C and actinorhodin) and fatty acids (palmitic acid and eicosapentaenioc acid) are shown in Figure 3. Despite differing structures, polyketides and fatty acids are related by the way they are constructed by the parent organism. Both classes of compounds are produced in a 'bottom up' approach utilising condensation of simple biomolecules to give the chemical structure required. Both fatty acids and polyketides are biosynthesised using related protein architecture from similar biomolecules despite their often vastly different structures. [15]

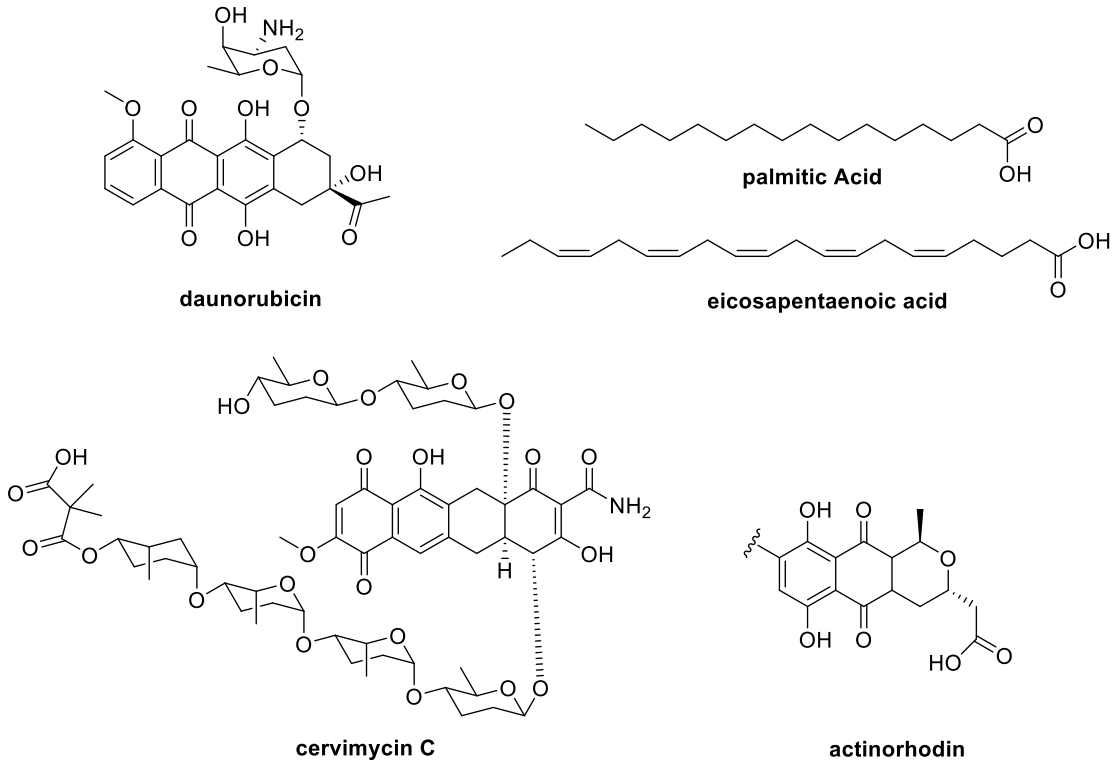


Figure 3: Selected structures of polyketides and fatty acids. Polyketides typically have an oxidised structure compared to fatty acids. The three polyketide structures shown here (daunorubicin, cervimycin C and actinorhodin) all show strong antibacterial activity. Note: single monomer of actinorhodin shown – this compound exists as a dimer.

1.2 Biosynthesis of Polyketides and Fatty Acids.

Within the genome of a polyketide/fatty acid-producing organism lies the biosynthetic gene cluster (BCG). This section of DNA contains all the genes required to synthesise the precise chemical structure of the polyketide/fatty acid from readily available cell components. On transcription and translation of this region of DNA, a series of proteins are assembled by the cell. [11] These proteins act like machines in a production line, taking simple starter units; malonyl-CoA and its derivatives, and condensing them to form the carbon backbone of the natural product.

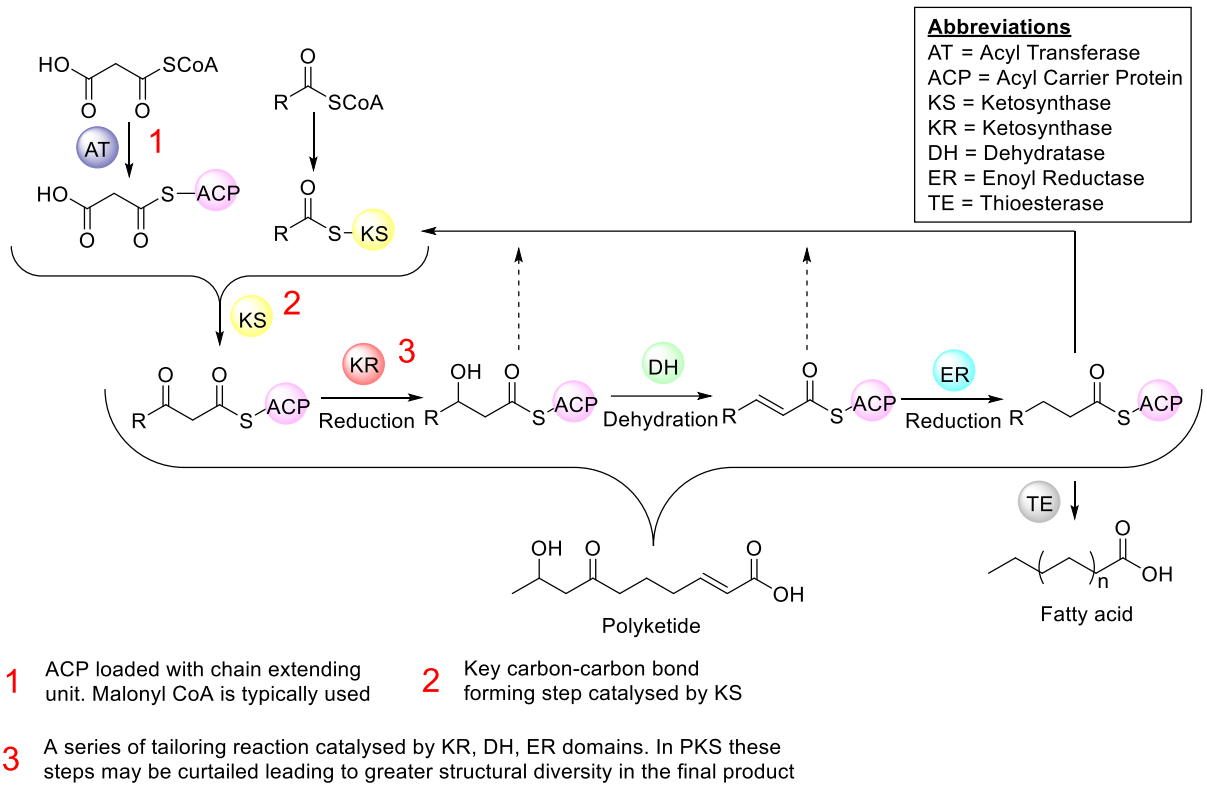


Figure 4: Depiction of the general steps in fatty acid and polyketide biosynthesis. Malonyl CoA is transferred to an acyl carrier protein (ACP) via action of acyl transferase (AT) domain (Step 1). ACP delivers a malonyl to KS domain loaded with acetate (or acyl group) which catalyses decarboxylative Claisen condensation to form a β-keto ester (Step 2). Resulting product undergoes reduction and dehydration steps resulting in a fully or partially saturated acyl chain (Step 3). The resulting acyl moiety will remain attached to the ACP and can undergo multiple rounds of chain extension until the required acyl chain length is reached. At this point the acyl moiety is removed from the ACP via the action of thioesterase (TE) domain. Within polyketide biosynthesis the final reductive steps may be curtailed leading to greater structural diversity of the final compound giving rise to their more varied chemical structure.

The biosynthesis of polyketides and fatty acids is initiated by loading of a chain extending unit: typically malonyl-CoA (although other units can be used) [16], onto an acyl carrier protein (ACP). This is catalysed by acyl transferase (AT) domain (Figure 4) to produce ACP bound malonyl. The malonyl group is covalently bound to the ACP by a thioester bond at the end of the phosphopantetheine arm of the ACP. Following this, a ketosynthase (KS) domain will condense a starter acyl unit with the ACP bound malonyl. This condensation reaction is known as a decarboxylative Claisen condensation (DCC) and occurs with the loss of CO_2 to produce a β-keto ester product. The DCC is the key carbon-carbon bond forming step during polyketide and fatty acid biosynthesis. The first round of extension will use acetyl CoA as an extender unit – further rounds will use the growing acyl chain. The resulting β-keto ester is successively reduced to β-hydroxyl by a ketoreducase (KR) domain, an α,β-unsaturated carbonyl by a dehydratase (DH) domain and a fully saturated acyl chain by an enoylreductase (ER). The acyl chain remains attached to the ACP throughout these tailoring steps. Once the first round of condensation-reduction has taken place, the resultant acyl chain can be transferred to the KS domain and the cycle

can repeat. It is through this cycle of condensation-reduction that the carbon backbone of the polyketide/fatty acid is constructed. When the required acyl chain length has been reached a thioesterase (TE) domain will hydrolyse the thioester bond releasing the resultant carboxylic acid. [17] Within fatty acid biosynthesis, these collective enzymes make up the fatty acid synthase (FAS) and within polyketide biosynthesis the polyketide synthase (PKS). [16]

1.2.1 Catalytic Domains involved in biosynthesis

Biosynthesis of polyketides occurs through the use of a series of canonical catalytic domains. These domains have been briefly described above. This section seeks to give more specific information about the AT and KS domains which are vital for biosynthesis to take place. The role of the ACP and each of the tailoring domains (KR, DH and ER) are described elsewhere in this thesis.

1.2.1.1 The role of the Acyl Transferase Domain

This enzyme completes the transfer of the chain extending group (typically malonyl) as is shown in Figure 4. Transfer occurs with an ACP which has been activated by addition of a phosphopantetheine prosthetic group (more on this in section 1.9). The interaction of the ACP with the AT domain has been shown to trigger conformational change of the AT domain leading to transacylation of the extender unit to the terminal thiol of the ACP utilising catalytic residues of a serine and a histidine.[18] The hydroxyl group of the serine side chain acts as a nucleophile attacking thiolester carbon of the acyl-CoA moiety. Subsequent attack of the terminal thiol of holo-ACP at the same position results in transfer of the acyl moiety to the ACP.

The incorporation of AT domains that transfer non-malonyl CoA moieties can occur for a subtype of PKS domains in which the AT domain is not covalently bound to the other modules (so called *trans*-AT PKS – discussed in more detail later). These PKS systems have been shown to incorporate more exotic acyl-CoAs such as ethyl-malonyl CoAs, methoxymalonyl CoAs and succinyl CoAs.[19,20]

Within biosynthetic pathways, AT domains are required to show specificity for the correct extender unit from a large pool of potential compounds. Therefore, a subtype of AT domains exist which contain specific motifs that will preferentially accept malonyl-CoA. These are known as malonyl-CoA:ACP Transacylase domains or MCAT. A specific subclass of AT domains that transfer a methyl-malonyl group to their cognate ACPs also exist. Further specificity may be afforded to AT domains by the presence of an acyl hydrolase (AH) domain typically preceeding the AT domain within the BGC. These AH domains comprise an evolutionary distinct group of AT domains whose role is to remove incorrect intermediates from the active sites of AT domains thereby ensuring only transfer of the correct intermediate occurs.

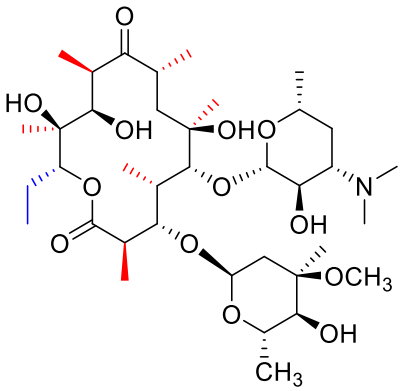
1.2.1.2 The role of the Ketosynthase domain

The KS domain will catalyse the key carbon-carbon bond forming reaction which underpins the biosynthesis of polyketides and fatty acids. Ketosynthases catalyse a condensation reaction that occurs with the loss of a CO_2 moiety. The conserved cysteine residue (activated by a conserved histidine residue) will become acylated with the growing polyketide chain. Extension of this polyketide then occurs by interaction with an α-carboxyacyl-ACP is facilitated by another conserved histidine. The active site of KS enzymes therefore typically consist of a cysteine residue and two histidine moieties.

A subclass of KS enzymes exists in which the active site cysteine residue is replace by a catalytically inactive residue. Despite lacking the ability to condense malonyl units, these KS enzymes nevertheless play an essential role in biosynthesis by catalysing the translocation of acyl chain intermediates. KS^0 domains are very selective for their required substrate and have been shown to act to ensure only the correct intermediate is passed on to the ACP to undergo further condensations. [21] This helps to ensure that only the correct natural product is produced by the biosynthetic system at work.

1.2.2 The Differences between FAS and PKS

Despite a similar protein architecture utilised in the biosynthesis of fatty acids and polyketides, there are key differences between these two systems. FAS systems typically utilise malonyl CoA as the main extender unit, this gives the non-branched saturated structure typical of fatty acids. Contrary to this, PKS have a greater propensity to accept extender units other than malonyl CoA, including methyl-malonyl CoA, propionyl CoA and others. The use of methyl-malonyl CoA is observed in the biosynthesis of the polyketide antibiotic erythromycin **1**. The structure of erythromycin consists of a 14-carbon macrolide ring, whereby a series of β-methyl branches are incorporated via methyl-malonyl CoA extender units.[22] This gives a series of β-methyl branches in the final structure (shown in red in Figure 5). These methyl groups are vital for erythromycin mode of action, being present in many macrolide antibiotics including clarithromycin and roxithromycin. The β-methyl branches are required to increase binding affinity of Erythromycin to its target 50S ribosomal subunit, thereby inhibiting protein synthesis.[23,24] Incorporation of unusual starter units in PKS systems such as methyl-malonyl CoA leads to greater structural diversity observed in the final product.



erythromycin

1

Figure 5: Structure of the polyketide antibiotic erythromycin. Use of methyl-malonyl CoA extender units during biosynthesis leads to the incorporation of methyl groups observed in red. A propionyl starter unit is initially loaded onto the KS giving the propionyl group shown in blue. PKS systems can modify the starter unit used to influence the chemical structure assembled.

As well as the propensity of PKS to incorporate more varied starter units, further differences between fatty acid and polyketide biosynthesis are observed. Within FAS, the action of KR, DH and ER typically all occur sequentially to produce the fully saturated acyl chain. Within PKS systems, these final reductive steps may be curtailed at specific extension rounds leading to greater structural diversity (Figure 3). [25]

1.3 Type I and Type II PKS

Polyketide synthases can be classified based on the structure of their constituent catalytic domains. Type I PKS refers to linearly arranged and covalently fused catalytic domains within large multifunctional enzymes. [26] These are typically found in eukaryotes and bacteria. Type II PKS consist of discrete protein domains that are not covalently bound together. The growing acyl chain will remain covalently bound to ACP and interaction of ACP with cognate successive protein partners progresses the biosynthetic sequence. During this process, the ACP has been shown to sequester the growing acyl chain. In type II systems, where the ACP-bound substrate exists in a pool of potential interacting partners, substrate sequestration serves to protect the acyl chain from unwanted catalytic activity. [27] Protein-protein interactions between the ACP and its cognate partner are required to release the substrate into the active site of each enzyme. Substrate sequestration does not occur in type I systems as the ACP and catalytic domains are close enough to interact directly. [28] Type I and type II distinctions

are similarly made with FAS systems also although the structures of these megasynthase domains are not always the same.

The domain organisation of the type I PKS/FAS megasynthases have been extensively studied. In the infancy of understanding the structure of PKS/FAS megasynthase modular architecture, two models were created to describe the ability of ACP to interact with all catalytic domains in a module. These were known as the head-to-tail (HTT) model and the head-to-head (HTH) model. An overview of these is shown in Figure 6. HTT model consists of two megasynthase strands fully extended aligned in an anti-parallel manner. The ACP of one strand can interact with domains from both megasynthases.[29] The HTH model instead consists of a coiled subunit in which the ACP can interact mainly with domains from the same polypeptide strand – whilst also being capable of interaction with KS and AT of the different strand.

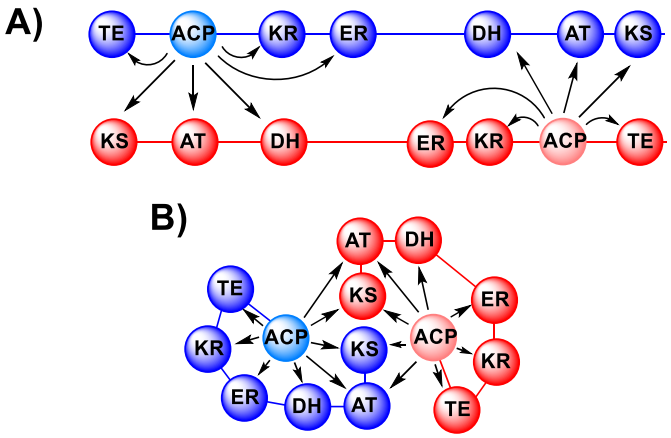


Figure 6: Figure showing two hypothesised domain organisations of PKS/FAS megasynthase. A) Head-to-tail organisation of megasynthase B) Head-to-head organsiation of megasynthases. Head-to-head can interact with all domians within same strand as well as AT and KS domains of second strand. Head-to-tail cannot interact with all domains of same strand.[25]

To confirm which model was correct a series of mutations were made to test whether HTT or HTH is taking place *in vivo*. These mutations confirmed that HTH model must be taking place by definitively proving that ACPs act only with other domains in the same module (HTT would require ACPs to interact with domains of alternative modules as shown in Figure 6). [30,31]

This HTH model was proven further when the structure of Type I FAS was solved by electron microscopy and crystallographic techniques. The structure of a porcine Type I FAS was solved to a resolution of 4.5 Å resolution in 2006, subsequently being improved to 3.2 Å. [32,33] The domain architecture consisted of an intertwined dimer with two lateral semi-circular reaction chambers, each containing a full set of catalytic domains required for fatty acid elongation (Figure 7). Electron density for the ACP and TE domains was not elucidated implying increased mobility of these regions. Sequence analysis and unassigned electron density is consistent with the ACP domain being located at the end of

the ‘arm’ – next to the KR domain as depicted. This arrangement will allow the ACP to interact with the required catalytic domains as shown in Figure 7.

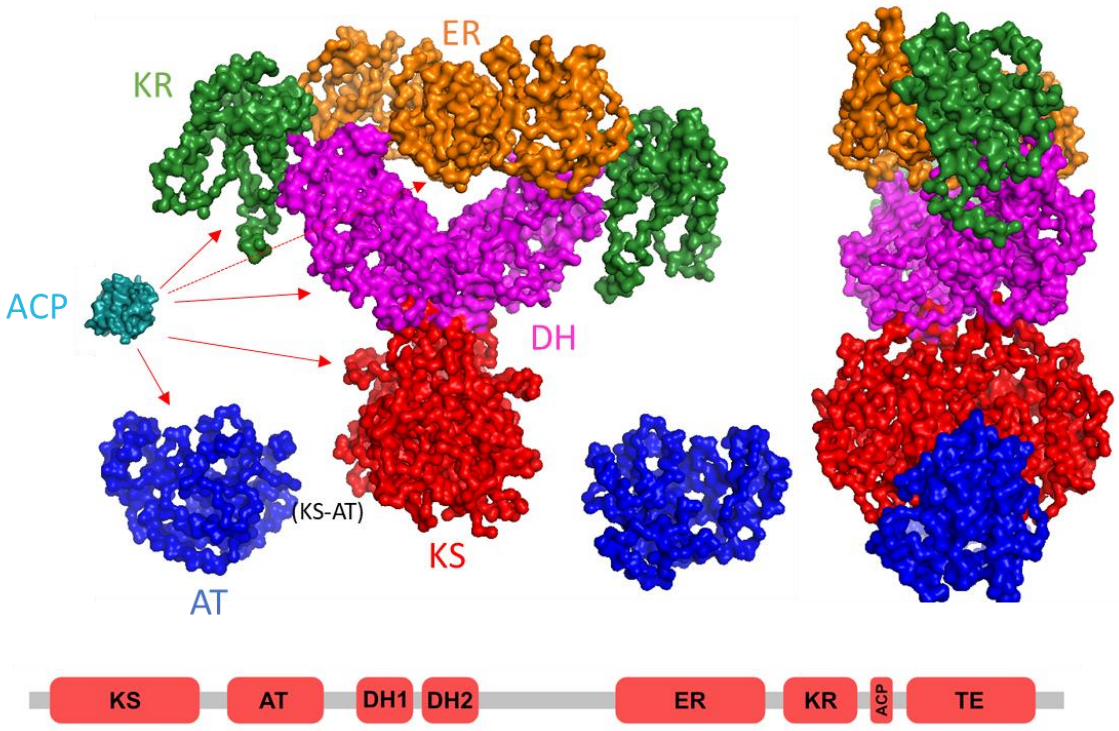


Figure 7: 4.5 Å Crystal structure of type I Porcine fatty acid synthase. The main catalytic domains are shown. KS = Red, AT = Blue, DH = Magenta, ER = Orange and KR = Green. The KS and AT domains are covalently linked via a KS-AT docking domain (not shown). The ACP and TE domains remain undefined due to proposed flexibility. It is postulated that the ACP domain will be located on the terminus of the KR domain allowing interaction with all domains.[32] The side on view shows the domains from the direction of ACP attack. Linear organisation of FAS is shown below.

Following this, the structure of the KS-AT di-domain of deoxyerythronolide B synthase (DEBS) was solved at 2.7 Å resolution. [34] This showed remarkable similarity to the FAS KS-AT domain confirming the similarity between the biosynthetic machinery of FAS and PKS.

Unlike the type I FAS system discussed above, type II PKS are not covalently bound together. Instead, these systems can associate to form a highly structured multi-protein megasynthase which is capable of catalysing condensation and reduction cycles before dissociating apart again. [35] Polyketide assembly by type II PKS systems is notoriously difficult to study due to transient nature of protein-protein interaction. This implies that the ACP plays a crucial role in shuttling the growing acyl chain among the individual enzymes within the megasynthase complex. [36] Despite this, no common ACP binding motif has been found in enzymes known to interact with ACPs. [36] Understanding the mechanism for ACP-partner interaction may prove very important for rational engineering of these systems to produce novel antibiotics. [37]

As well as the type I (covalently bound domains) and type II (discrete domains) PKS classification described above; PKS can be further classified as iterative and non-iterative systems. Iterative systems are characterised by the repeated use of the same catalytic domains to achieve biosynthesis. Each iterative PKS contains one set of all the catalytic domains required for the condensation and reduction sequence. The ACP repeatedly interacts with all the domains within a module to perform the 2-carbon elongation/reduction cycle until the required acyl chain length has been achieved. Non-iterative domains consist of multiple covalently bound modules, each one responsible for the condensation and reduction of a single 2-carbon acyl moiety before passing on to the KS of the next module. Non-iterative systems are relatively very large consisting of multiple modules covalently bound together. Each module contains all the domains required to catalyse and reduce a single 2-carbon moiety. Due to this, the core structure of the outgoing polyketide can often be predicted based on analysis of the domains which make up each module. This is known as the co-linearity rule and has been previously exploited for the discovery of natural products. [38,39]

1.4 Further Classification of PKS domains

Further classification of PKS type I systems can be achieved based on presence of the ACP priming domain; AT, within each module. PKS systems consisting of modules containing AT domains are classified as *cis*-AT PKS synthases and PKS lacking a covalently bound AT are known as *trans*-AT synthases. In *trans*-AT systems, the priming of the ACP is undertaken by a non-covalently bound AT which will associate, transfer malonyl to the correct ACP, and dissociate again. [20]

Further to this, the domain architecture of *trans*-AT systems are different to their *cis*-AT counterparts. Within cis-AT systems each module contains a covalently attached AT domain that is responsible for loading the ACP of the same module. Trans-AT systems instead incorporate a discrete AT domain that is able to transfer extender units to ACPs from all modules as shown in Figure 8B. The

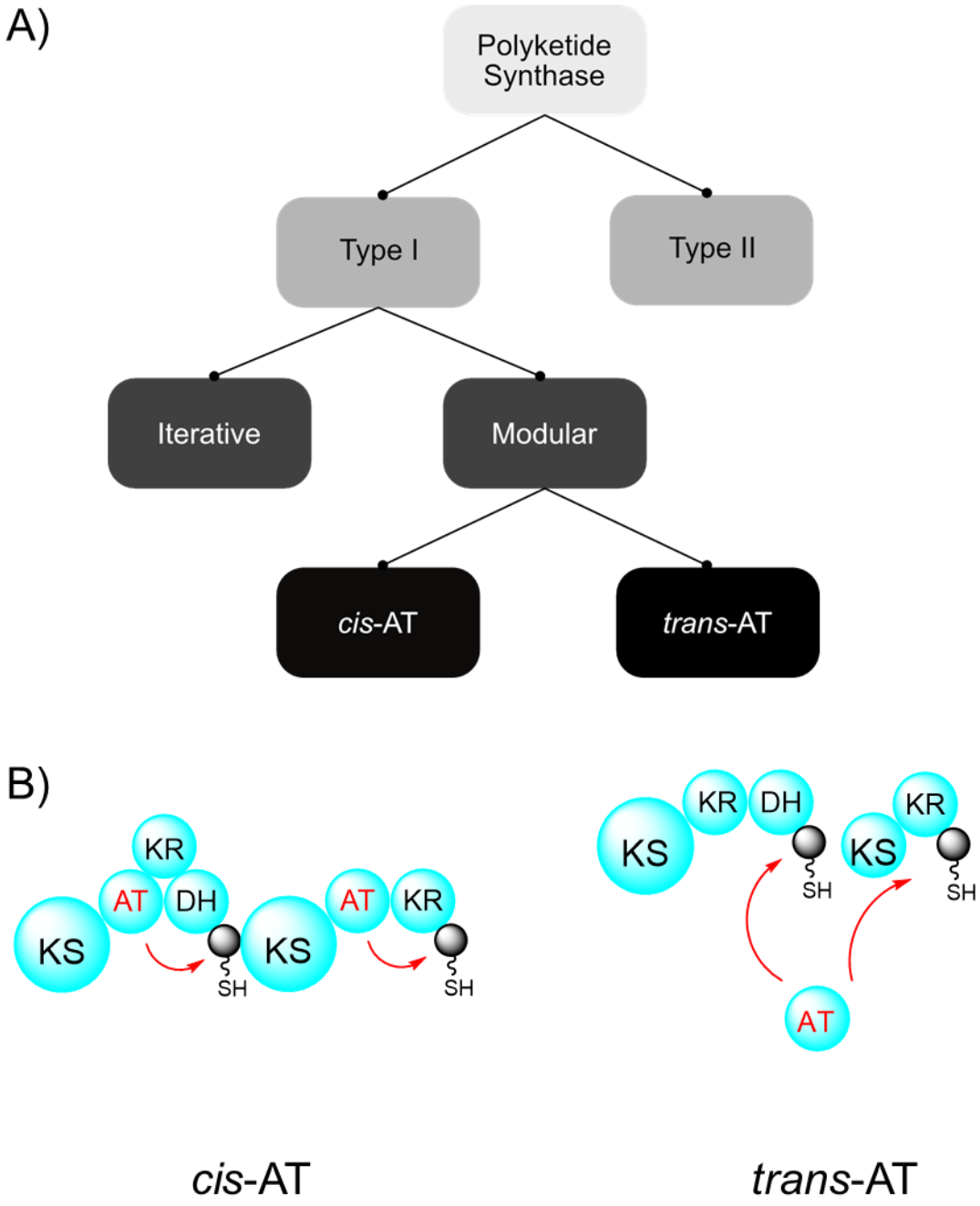


Figure 8: A) The classification of polyketide synthase systems. B) The typical domain organisation of cis-AT and trans-AT systems.

PKS systems of the *trans*-AT type are remarkably common – a 2014 report found that approximately 38% of all modular bacterial PKSs are *trans*-AT type. [40] Not only this, *trans*-AT gene clusters are particularly common in less well-studied bacteria. Studying *trans*-AT systems therefore indicates a significant potential for the discovery of novel natural products with pharmacological promise.

Numerous peculiarities exist in *trans*-AT systems when compared to *cis*-AT counterparts. Systems of *trans*-AT typically have much greater architectural diversity within modules when compared to *cis*-AT systems. For example, whereas *cis*-AT systems typically utilise the core biosynthetic machinery of KS, AT, ACP, KR, DH, ER and MT domains – much greater domain complexity is observed for *trans*-AT systems. Such unique domain architectures may include non-elongating KS domains, domain action in different modules or split modules (modules divided across multiple proteins). [20] This leads to much greater structural diversity in the outgoing natural products of *trans*-AT systems.

The incorporation of unique and often irrational modular architecture in *trans*-AT systems results in regular breaking of the co-linearity rule (as defined above). The breakdown of this rule for *trans*-AT systems means there is a lack of understanding about how these systems operate.

PKS systems of the *cis*-AT and *trans*-AT type also differ in their mode of evolution. *cis*-AT systems evolved by gene duplication of individual modules followed by specificity of each domain.[41] The KS domains of *cis*-AT systems therefore phylogenetically clade with other KS domains within the same biosynthetic cluster. In contrast, *trans*-AT PKS genes are thought to have assembled through horizontal gene transfer between bacteria. [42] The KS domains of *trans*-AT PKS systems therefore phylogenetically clade with KS domains of similar substrate specificity. This can provide a valuable information about the biosynthesis of *trans*-AT systems when co-linearity fails.

1.5 Structure and Clinical Relevance of Mupirocin

Mupirocin is a polyketide antibiotic produced by *Pseudomonas fluorescens*, a Gram-negative soil bacterium. Although the antibiotic activity of this rod-shaped bacterium had been known since 1887, the structure of mupirocin was not elucidated until nearly a century later. [43,44] The structure of mupirocin was found to consist of a mixture of compounds, known as pseudomonic acids A-D.[45] The general structure of the pseudomonic acids were found to consist of a pyran ring containing heptaketide moiety (monic acid) esterified to a 9-carbon fatty acid derived moiety (9-hydroxynonanoic acid) (Figure 9) **2**. Pseudomonic acid A contains an epoxide group at C-10 and was found to make up over 90% of the mixture of mupirocin. [46] Pseudomonic acid B (PA B) makes up 8% of the mixture and consists of a hydroxyl group at C-8 and pseudomonic acid C (PA C) was found to make up a majority of the remaining 2%, uniquely having an E-alkene at C-10. A very small amount of pseudomonic acid D was also observed, containing an alkene with an E configuration at the C-4 position of the fatty acid chain.

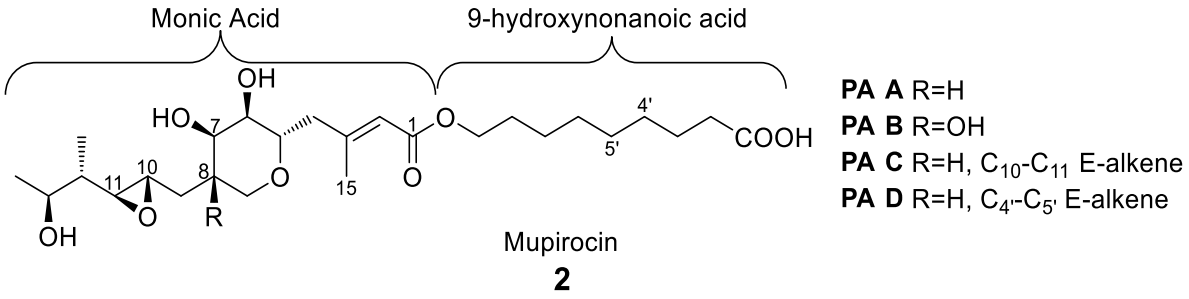
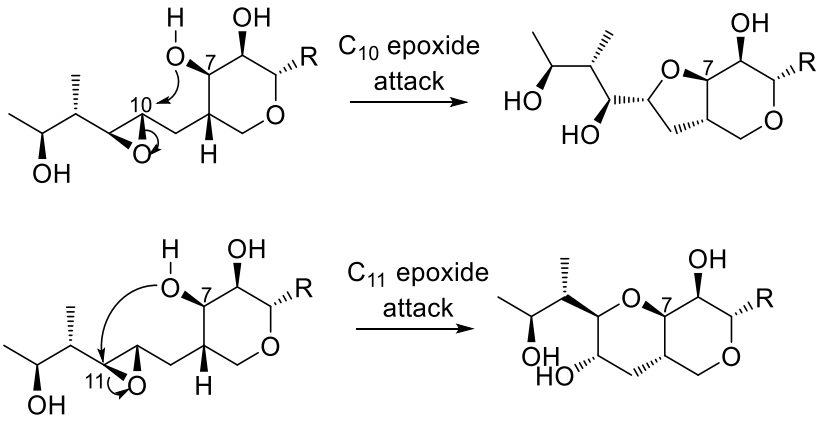


Figure 9: Structure of polyketide antibiotic mupirocin is a mixture of pseudomonic acids. The main metabolite – making up 90% of the mixture is pseudomonic acid A. Two distinct regions of this compound are visible, a polyketide derived monic acid and a fatty-acid derived 9-hydroxynonanoic acid moiety connected by an ester bond.

Mupirocin has attracted much interest due its alluring antibiotic activity, specifically against gram-positive bacteria such as Methicillin-resistant *Staphylococcus aureus* (MRSA). [47] For many years, mupirocin was sold as an ointment for the topical treatment of bacterial skin infections [48] and as an antibacterial nasal spray. [49] The clinical use of mupirocin has been limited due to the instability of the main metabolite PA A. The ester bond that links the monic acid and fatty acid region of mupirocin is rapidly hydrolysed within the body when orally ingested, rendering it inactive. In addition, strong binding to blood serum also contributes to a limited active concentration in the body.[50] Outside the pH range 4-9, attack of the C-7 hydroxyl group into the epoxide at C-10 or C-11 position irreversibly produces two cyclic ethers which are no longer biologically active (Scheme 1). The minor metabolite PA C contains a double bond instead of the epoxide at C-10 and is consequently more stable.



Scheme 1: The intramolecular attack of C-7 hydroxyl into C-10 and C-11 positions forming two inactive cyclic ethers. This typically occurs outside of the pH range 4-9 and leads to the poor oral bioavailability of mupirocin. R group constituted the fatty acid chain of mupirocin.

One of the major goals of the early work on this pathway was to produce a strain of *P. flourescens* capable of producing a larger amount of the stable metabolite PA C. [51] Recently published work by this group has shown optimisation of the bacteria fermentation conditions was successful in producing a

substantial increase in the amount of Pseudomonic acid C. [52] This has the potential for industrial-scale production of PA C and its use as a novel antibiotic.

1.6 Antibiotic Activity of Mupirocin

The origin of the antibiotic activity of mupirocin is due to potent inhibition of the eubacterial isoleucyl-tRNA synthetase (IleRS) class of enzymes. These enzymes are used to affix isoleucine to its cognate tRNA during bacterial protein synthesis. The mode of action of mupirocin was confirmed by co-crystallisation of mupirocin with *Thermus thermophilus* IleRS. [53] The C-14 methyl terminus of monic acid was found to mimic the hydrophobic side chain of L-isoleucine allowing binding to the active site of the protein. The pyran ring and C1-C3 enone region were found to affix to the ATP-binding pocket of IleRS. The 9-hydroxynonanoic acid (9HN) chain of mupirocin also plays a significant role in binding – replacement of this moiety with an ethyl chain led to 10% binding affinity when compared to the full 9HN.[54] The 9HN chain was found to sit in a hydrophobic groove on the surface of IleRS, stabilising the resulting IleRS-mupirocin complex.

The reversible nature of the binding interaction is consistent with mupirocin being a competitive inhibitor of IleRS. When expression of the gene cluster is turned on, to ensure survival of the mupirocin producing organism, the biosynthetic cluster contains an alternative IleRS, MupM, which is resistant to mupirocin. Whereas eubacterial isoleucyl-tRNA synthase are members of class I aminoacyl-tRNA synthetases (aaRSs), MupM is a member of class II which has a distinct catalytic domain architecture and different motif for ATP binding. [55] The antibiotic activity of mupirocin is therefore non-existent against class II IleRSs. Analysis of strains of MRSA with high-level resistance to mupirocin have identified a new type of class II IleRS therefore circumventing mupirocin inhibition.

1.7 Polyketide Antibiotic Thiomarinol

Thiomarinol is a polyketide antibiotic with a comparable structure to mupirocin produced by marine bacterium *Pseudoalteromonas sp.* SANK 73390. [56] Like mupirocin, thiomarinol is a mixture of multiple different metabolites, each with a related structure. The structure of the major metabolite, thiomarinol A **3** is a hybrid of two independently active species; the monic acid and 8-carbon fatty acid chain resembling mupirocin coupled with the pyrrothine moiety found in holomycin (Figure 10).[57] As well as showing excellent activity against gram-positive bacteria, thiomarinol is able to show strong activity against many gram-negative bacteria also. Antibiotic activity of thiomarinol consists of inhibition of bacterial protein synthesis due to mupirocin-like core in conjunction with pyrrothine moiety which facilitates antibiotic activity by inhibition of RNA synthesis.[58] This hybridity of two active warheads

covalently bound together leads to a much more potent antibiotic than the constituent parts alone.[59] The structural similarity of mupirocin and thiomarinol implies a biosynthetic pathway with shared features.

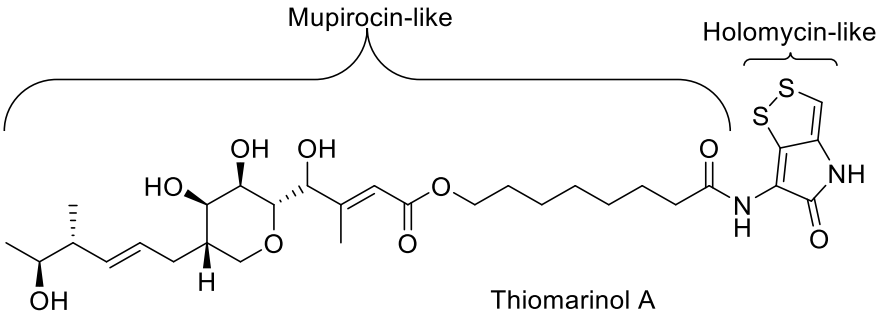


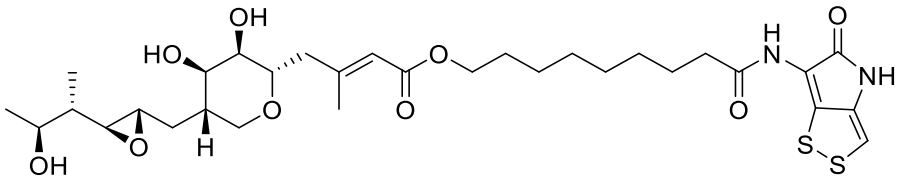
Figure 10: Structure of the polyketide antibiotic thiomarinol A showing the holomycin-like pyrrothine and Mupirocin-like core. The hybridity of these two antibiotics is responsible for the increased antibiotic activity of thiomarinol compared to mupirocin.

1.8 Importance of Studying Biosynthetic Pathways

Biosynthetic pathways constitute nature's way of synthesising complex molecules. The potency of many molecules produced from living systems is a result of the many years of evolution acting to hone the selectivity and biological activity. Despite this, the compounds that have been exploited so far only represent the tip of the iceberg when compared to the chemical diversity represented in nature.[60] There is clearly still a lot of work to be done to realise the potential of natural products as pharmaceutical drugs.

There are many compelling arguments to study biosynthetic systems. As has been previously noted, there exists a gap in knowledge about how PKS (particularly *trans*-AT PKS systems) function to produce their respective natural products. By understanding the complex logic of these remarkable biosynthetic factories, a more robust set of rules may be developed which will allow accurate prediction of chemical structure from functional assignment of domains alone. This will exponentially speed up the time taken to identify new polyketides from genome alone.

Increased understanding of specific PKS systems is also a very valuable goal. The rational modification of specific catalytic domains may lead to the production of entirely new natural products. The production of mupirocin pyrrothine amide **4** is an example of this (Figure 11). [56] This novel antibiotic was produced by utilising a protein from the biosynthesis of thiomarinol to load pyrrothine onto pseudomonic acid A. This produced an entirely new antibiotic with strong antibiotic activity. This kind of rational protein engineering can only be achieved with considerable understanding of the underlying biosynthetic pathway.



Mupirocin pyrrothine amide

4

Figure 11: Chemical structure of novel antibiotic mupirocin pyrrothine amide. This antibiotic was produced by transfer of thiomarinol pyrrothine moiety to the fatty acid chain of pseudomonic acid A.

Biosynthetic pathways are also an important source of novel catalytic proteins. The study of the abyssomycin C biosynthetic pathway led to the discovery of a new type of catalytic domain known as a Diels-Alderase enzyme. [61] This is the one of the first examples of a natural protein having the ability to catalyse the chemistry nobel-prize winning Diels-Alder reaction. Biocatalysts capable of unlocking new and efficient Diels-Alder reactions would be a major achievement. Work is currently being done to evaluate and improve the scope of this remarkable enzyme. [62]

Increasing our knowledge of the inherent logic which governs the functioning of these systems is a long-term goal of this project. This knowledge should be exploited to pave the way for rational engineering of these systems to produce novel compounds. [63] This is especially prevalent due to the lack of new natural product based antibiotics coming to market and the subsequent human and economic effect of this. [64,65,66] The selectivity of an ACP to interact with multiple protein partners may be key to understanding polyketide biosynthesis.

1.9 The Role of the ACP in Biosynthetic Systems

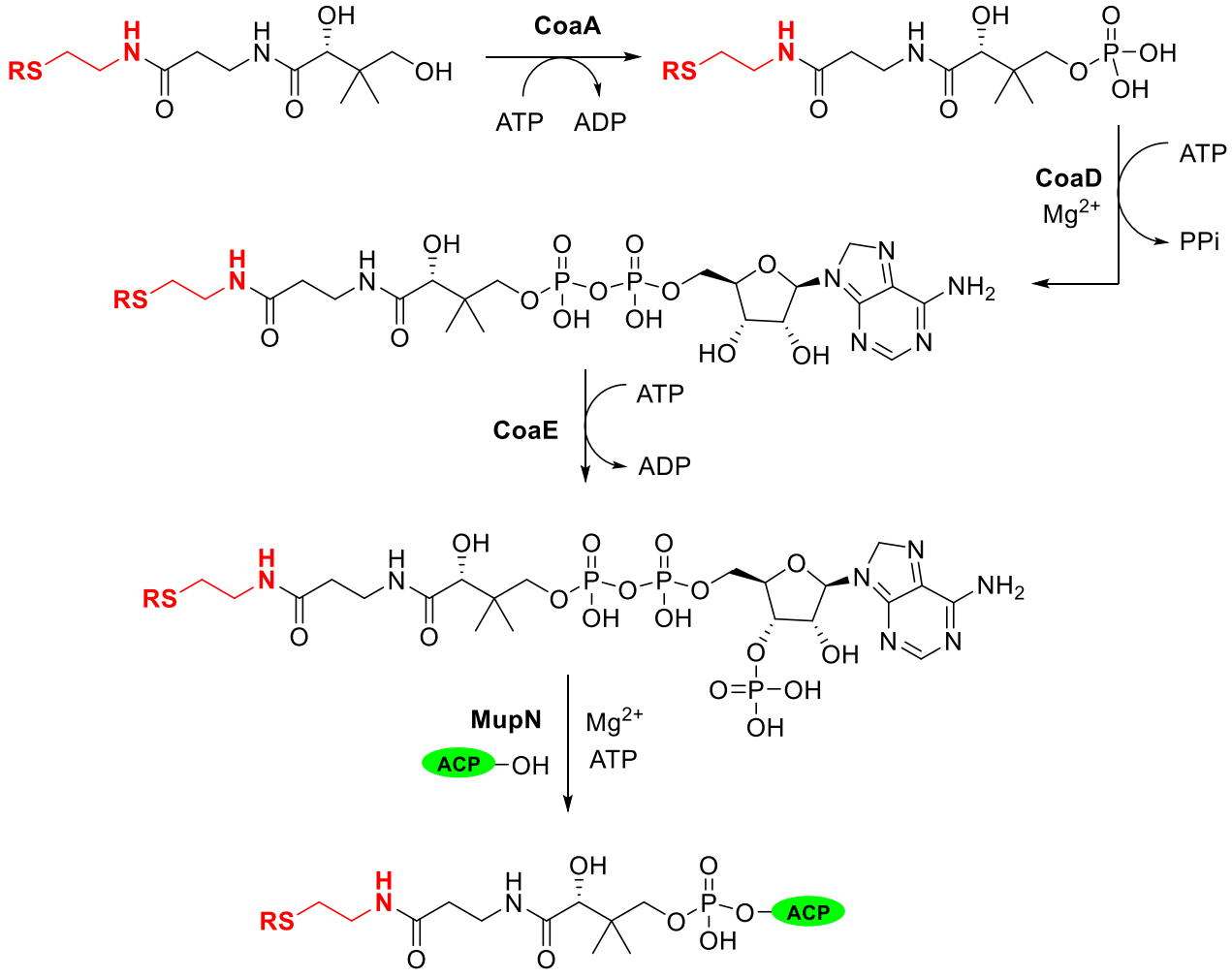
Acyl carrier proteins (ACPs) play a vital role in shuttling the growing acyl chain to the correct catalytic domains during biosynthesis. The ACP is a small, acidic protein typically consisting of less than 100 amino acid residues. ACPs adopt a common tertiary structure consisting of four α-helices (I-IV) enclosing a hydrophobic core. ACPs within type II systems have been found to 'sequester' and stabilise the growing acyl chain in the hydrophobic core during biosynthesis.

Post-translational modification of ACPs is required to convert the ACP from the inactive *apo* form to the active *holo*-ACP. The transfer of the phosphopantetheine arm from coenzyme A (CoA) to a conserved serine residue found at the base of helix II results in the formation of active *holo*-ACP. The phosphopantetheine arm is typically 20 Å in length and acyl intermediates are tethered to the terminal thiol allowing interaction with the KS and other catalytic domains while still attached to the ACP.

The transfer of phosphopantetheine to the ACP is catalysed by the phosphopantetheinyl transferase (PPTase) class of enzymes. [67] This superfamily of PPTases can be split into two groups, group I (AcpS-

type) and group II (Sfp-type). The characteristics and interactions of both groups have been well characterised. [68] Sfp-type PPTase enzymes have been shown to be promiscuous in terms of the ACP being modified or acyl moiety used. [69,70] The endogenous PPTase enzyme of mupirocin, MupN, is an Sfp-type PPTase. Work from Shields *et al* has shown MupN can functionalise mupirocin ACPs during the *in vitro* reconstitution of the mupirocin biosynthetic pathway producing active *holo*-ACP from *apo*-ACP.[71]

A synthetic approach can be taken to produce pantetheine synthetic intermediates with variable acyl groups on the terminal thiol. [72] This functionalised pantetheine substrate can be 'upgraded' to CoA moiety utilising enzymes from the biosynthesis of Coenzyme A (CoaA, CoaD and CoaE). [73] The corresponding CoA intermediate can be transferred onto the desired ACP in a one-pot reaction, utilising the promiscuous nature of MupN. [74] This process is outlined in Scheme 2.

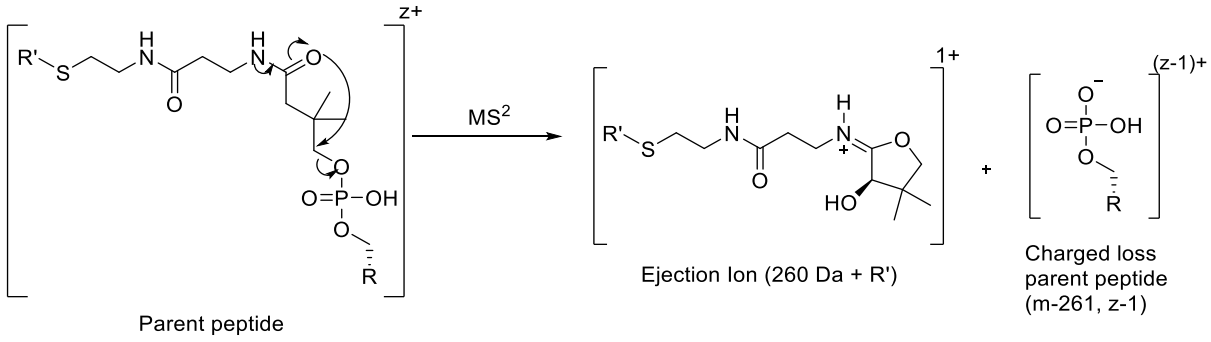


Scheme 2: A diagram showing the loading of ACP with synthetic pantetheine intermediate – a key reaction that underpins the work done in this project. A synthetic method has been developed which allows synthesis of pantetheine moiety with various groups on the terminal thiol (shown in red). This pantetheine analogue is converted to corresponding CoA analogue utilising enzymes from the biosynthesis of coenzyme A. Subsequent CoA analogue can be loaded onto apo-ACP *utilising the PPTase enzyme MupN.*

This approach to produce functionalised ACPs allows studying of enzymatic reactions of PKS pathways *in vitro*. Previous biosynthetic investigations have involved using SNAc derivatised substrates to probe biosynthetic pathways.[75,76] Although SNAcs are small and easily accessible by synthetic means, these do not full replicate the flexible phosphopantetheine arm of the ACP. SNAc moieties are also not bound to an ACP therefore negating the potential protein-protein interactions between ACP and catalytic domain that are often integral to the correct delivery of substrate. [77] The gold-standard for replicating these reactions *in vitro* is by using functionalised ACPs. By using protein mass spectrometry to monitor these ACPs, *in vitro* assays can be conducted with functionalised ACPs and their proposed cognate partners. This will allow greater understanding of the roles of different ACPs in the mupirocin biosynthetic pathway.

The small size (< 15 KDa) and relative thermal and ionic stability of type II bacterial ACPs encourages their study by Electrospray Ionisation Mass Spectrometry (ESI-MS). As the ACP stabilises the growing acyl chain during biosynthesis – changes to the acyl chain (after interaction with a catalytic domain) can be observable through mass changes of the ACP. This can provide valuable insight into the function of *trans*-acting domains within biosynthetic pathways.

The identity of the acyl species located on the terminal thiol of prosthetic arm of ACPs can be analysed with greater resolution by utilising the phosphopantetheine (Ppant) ejection assay. The Ppant ejection assay is a technique that relies on collisionally induced dissociation tandem mass spectrometry of an ACP to relieve a characteristic ejected ion and phosphorylated *apo*-ACP (Scheme 3). The acyl moiety on the terminal thiol of this ejected ion is not affected by this fragmentation and can be analysed at higher mass accuracy. This technique means that analysis of the acyl group on the terminal thiol of ACPs is very easily achieved. This makes the study ACPs by mass spectrometry a very attractive method for the characterisation of biosynthetic pathways.



Scheme 3: The mechanism of further fragmentation of ACPs and subsequent structure of ejection ion. This allows spectrometric analysis at greater resolution than that achieved for proteins. Z indicates the charge of the protein species undergoing fragmentation. The charge of the ejection ion will always be 1, leading to the mass to charge ratio being equal to the mass of the fragmented species.

1.10 Biosynthesis of Mupirocin

1.10.1 Type I PKS and Formation of Monic Acid

The 74,000 Bp (base pair) biosynthetic cluster of mupirocin has been identified and fully sequenced. [78] Within this genome are over 35 open reading frames (ORF), six of these encode proteins which have been found to have multiple independent catalytic domains; designated mupirocin multifunctional proteins (Mmp) A-F. As the biosynthetic cluster of mupirocin has been elucidated, [78] knowledge of the precise chemical steps required to form this clinically useful antibiotic has been an attractive goal.[48]

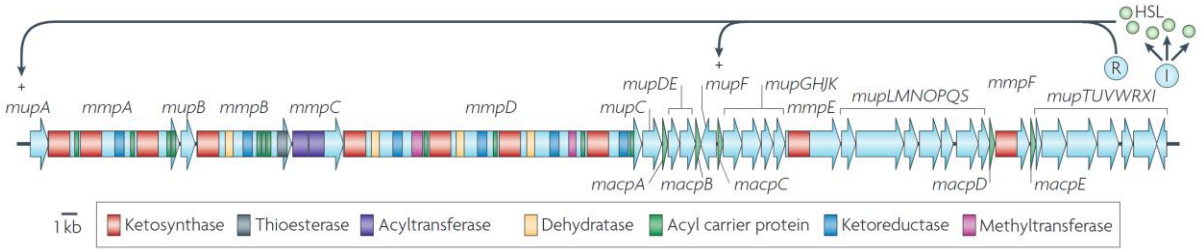


Figure 12: A diagram showing the genetic organisation of the mupirocin biosynthethetic cluster. This image was reproduced from reference 46.

The biosynthetic pathway of mupirocin contains many constituent parts. MmpA and MmpD encode for non-iterative *trans*-AT modular type I PKS systems responsible for the condensation of six molecules of malonyl CoA with one molecule of acetyl CoA to produce the 14-carbon backbone of monic acid. The multifunctional enzyme domain MmpC contains two AT domains and is thought to provide the AT domains required to load ACPs with a malonyl group during biosynthesis. MmpC also contains an ER domain which is thought to be *trans*-acting throughout the formation of the carbon backbone and potentially in the formation of the fatty acid chain. This process is shown in Figure 13.

Once construction of the carbon backbone has been achieved, a series of enzymes act to form the β-branch at C-3 position. These enzymes are MupG, MupH, MupJ, MupK and mAcpC. MupH homologues typically exist in a 'HMG cassette' and are responsible for the incorporation of branched enone using malonyl-CoA as a source of carbon. [79] Subsequent transfer to the KS^0 domain of MmpE produces intermediate **5** (Figure 13). KS^0 domains are non-condensing KS domains typically utilised in the transfer of the acyl intermediate between modules of the PKS. Formation of the C-10 epoxide occurs while covalently attached to MmpE forming intermediate **6**. Following this, it is hypothesised that the tetrahydropyran ring of mupirocin is closed utilising MupW and MupT. Attempts to show this transformation via whole cell bioassays were initially unsuccessful, instead forming the 5-membered ring. Re-analysis of the biosynthetic cluster of mupirocin found an additional protein, MupZ, not originally annotated, whose job was to preferentially form the 6-membered ring. [80,81] The metabolite produced after ring closure shares striking similarity to monic acid but for alcohol functionality at C-8, **7**.

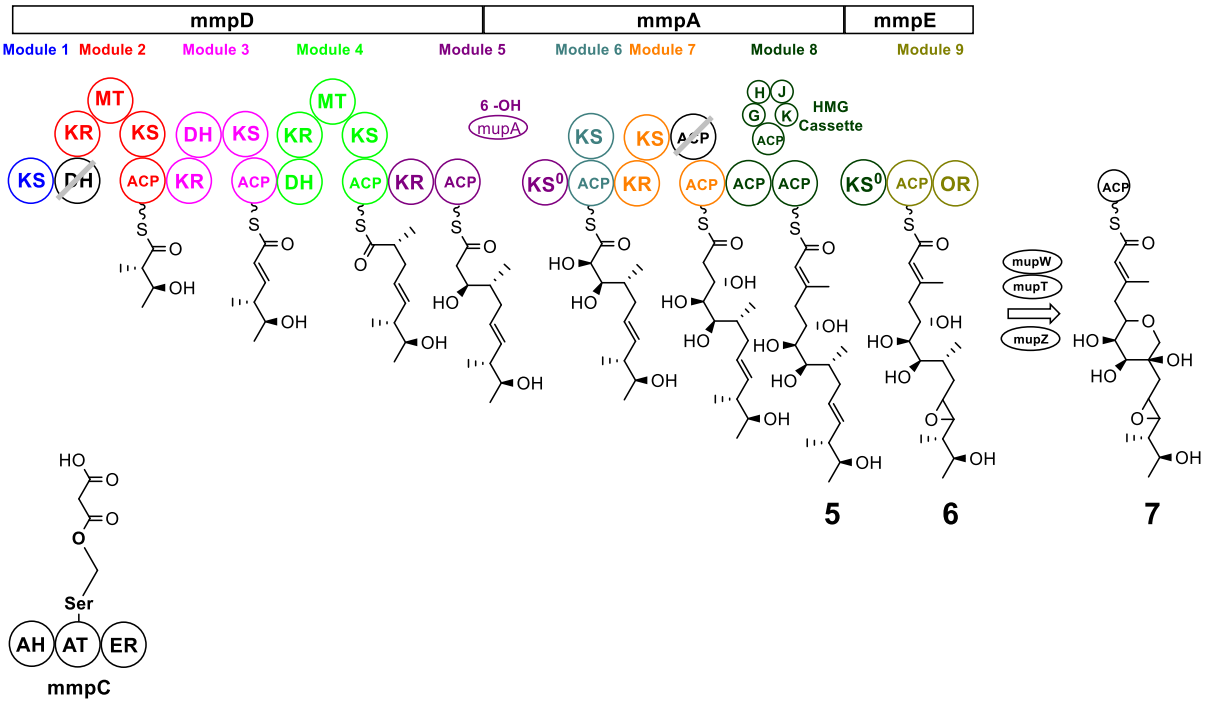
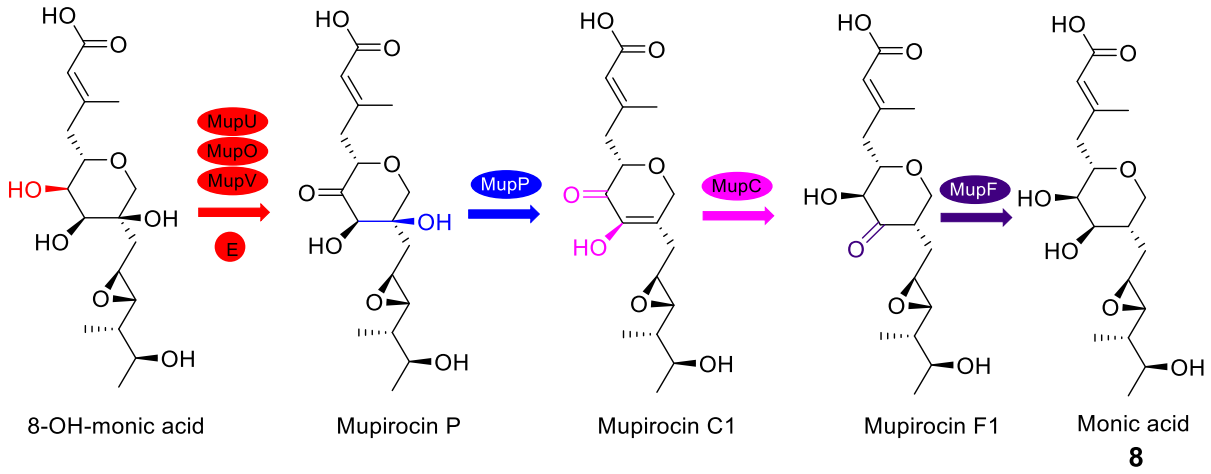


Figure 13: The type 1 trans-AT PKS biosynthesis of the monic acid region of mupirocin. A total of 9 modules are responsible for 7 decarboxylative Claisen condensations to produce the C-8 OH containing monic acid intermediate. Mupirocin is a trans-AT, trans-ER system with AT and ER functionality catalysed by trans-acting mmpC. Three mupirocin multifunctional proteins (mmp) mmpD, mmpA and mmpE are responsible polyketide construction. Certain domains are present in mmpD and mmpA but do not have a proposed function leading to a breakdown in the co-linearity rule. Hydroxylation in C-6 position is proposed to be catalysed by trans-acting mupA. Formation of β-branch occurs between mmpA and mmpE catalysed by HMG Cassette of MupGHJK and mAcpC. Ring closure step is catalysed by mupW, mupT and mupZ to produce monic acid with C-8 hydroxyl.

The removal of the C-8 alcohol to form the main metabolite PA A occurs *via* the action of 7 *trans*-acting tailoring enzymes and an ACP (Scheme 4). This was confirmed by selected mutations of these enzymes (MupC, MupF, MupO, MupP, MupU, MupV and mAcpE) showing subsequent lack of PA B to PA A conversion.[81] Once the C-8 alcohol has been removed the required monic acid moiety **8** has been produced.



Scheme 4: A scheme showing the biochemical steps required to convert PA B to PA A. MupU, MupO, MupV and mAcpE are required to convert C-6 hydroxyl to ketone. MupP is required to eliminate C-8 hydroxyl. MupC is required to convert C-6 ketone, C-7 hydroxyl to C-6 hydroxyl, C-7 ketone functionality. MupF is required to convert the C-7 ketone to C-7 hydroxyl. This produces monic acid which is esterified to 9HN to produce PA A.

1.10.2 Biosynthesis of 9-hydroxynonanoic acid

Comparatively less is known about the biosynthesis of the fatty acid region of mupirocin when compared to construction of monic acid region **8**. To understand more, the feeding of isotopally enriched precursors to growing cultures of *P. fluorescens* was conducted. [82] Initial feeding of [1-^{13}C,2-$^{2}H_3$]-acetate (Figure 14) to growing cultures of *P. fluorescens* was conducted and subsequent labelling pattern elucidated. [83] This evidence was consistent with the formation of a 3-carbon starter unit which then undergoes three rounds of FAS chain extension to give 9-hydroxynonanoic acid. This 3-carbon starter unit has been shown to be 3-hydroxypropanoate (3HP) **10** produced from acetate derived malonyl CoA **9** via the action of three proteins; mAcpD, MupQ and MupS (Scheme 5). 3HP will be produced attached to terminal thiol of mAcpD so the natural product will be 3HP-mAcpD **10$_D$**. The functions of these proteins were initially proposed based on similarity to proteins found in the biosynthetic cluster of difficidin. [84]

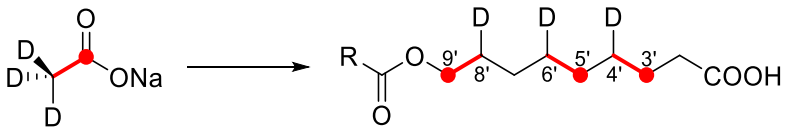
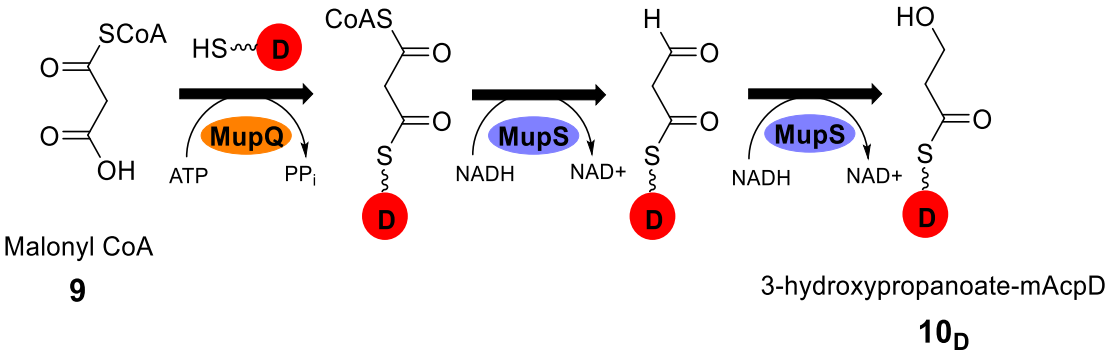


Figure 14: Double labelled acetate ([1 -^{13}C,2- $^{2}H_3$]-acetate) was fed into growing cultures of P. fluorescens. Isolation of the main metabolite PA A and analysis of the labelling pattern through NMR analysis was conducted. Above shows the labelling pattern observed for the 9HN chain of PA A. This is consistent with an acetate derived 3-carbon moiety which undergoes a further 3 rounds of fatty acid chain extension to produce the 9HN. This provides the basis for the proposed biosynthesis of 9HN.



Scheme 5: The biosynthesis of 3-hydroxypropanoate starter unit; the first step in the formation of 9-hydroxynonanoic acid. First, the carboxylic acid group of malonyl CoA is activated with ATP and loaded on to holo-mAcpD. The bis-thioester intermediate produced then undergoes two rounds of MupS-catalysed reduction to produce the ACP bound 3-carbon moiety: 3HP.

Similar isotope feeding studies involving the structurally related polyketide antibiotic thiomarinol also gave strong evidence for a homologous pathway via the action of tAcpD, TmlQ and TmlS to give a 4-hydroxybutyrate (4HB) from succinyl CoA starter unit. [85]

The fatty acid chain lengths in mupirocin and thiomarinol are different. Mupirocin consists of a 9-carbon fatty acid chain whereas thiomarinol consists of an 8-carbon chain. The construction of each chain is proposed to occur in a similar way. In the mupirocin system, a 3-carbon alcohol is produced (via the action of MupQ/S and AcpD), this undergoes three rounds of fatty acid chain extension to produce the 9-carbon chain length required. In thiomarinol a 4-carbon alcohol is produced (via the action of TmlQ/S and AcpD) which undergoes a further 2 round of fatty acid chain extension to produce 8-carbon fatty acid chain.

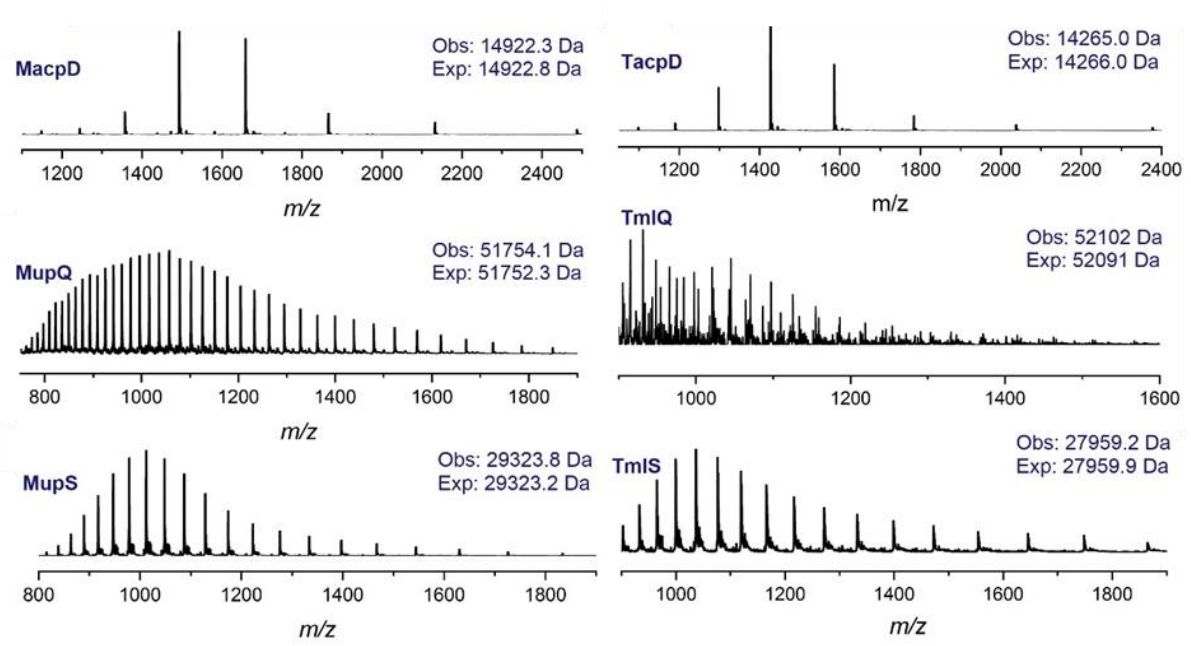


Figure 15: ESI-MS of the charge states of purified proteins required for formation of alcohol intermediate in mupirocin and thiomarinol.[86]

The roles of Q, S and AcpD were confirmed by the expression and purification of these proteins and conduction of *in vitro* assays for mupirocin and thiomarinol systems.[87] MupQ/S, TmlQ/S, mAcpD and tAcpD were all expressed and purified to yield soluble protein (Figure 15). Assays with these proteins was successful in producing the novel 3HP **10** and homologous 4HB as confirmed by electrospray ionisation mass spectrometry (ESI-MS). Formation of 3HP and 4HB constitute the first step in the formation of the fatty acid regions of mupirocin and thiomarinol, respectively.

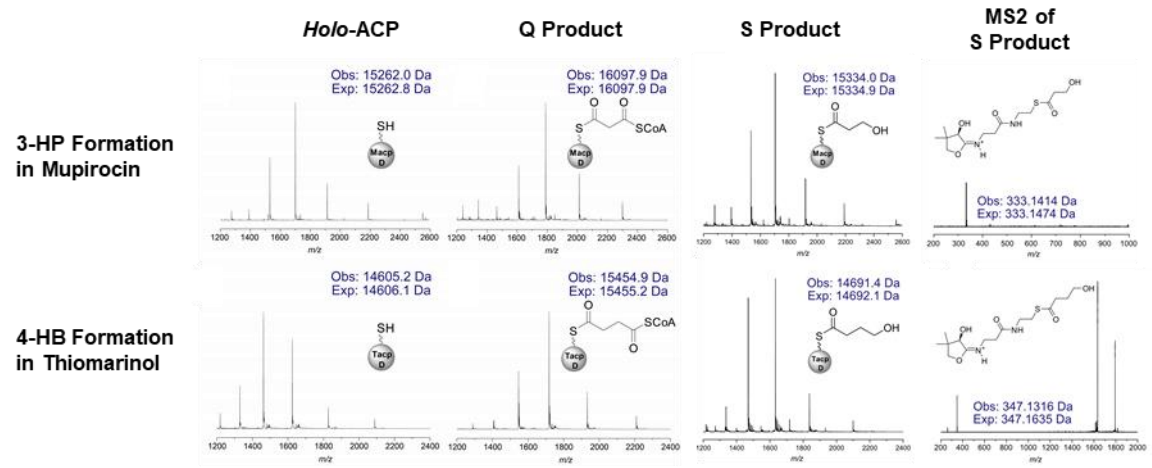


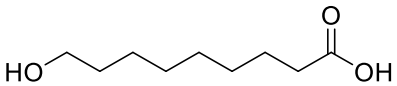
Figure 16:A figure showing the ESI-MS spectra of each stage of 3HP/4HB production. The production of 3HP and 4HB was monitored by ESI-MS. AcpD charge states were monitored at each stage of the reaction. Successful production of holo-AcpD was observed for ACPs in both pathways by utilising the PPTase enzyme MupN. Subsequent addition of Q and substrate (malonyl CoA/ succinyl CoA for mupirocin/ thiomarinol respectively) and cofactors ($MgCl_2$ and ATP) led to formation of bis-thioester product (Q product). Subsequent reduction using S and NADH produced 3HP/4HB for mupirocin and thiomarinol respectively. Further fragmentation of S product confirmed formation of 3HP/4HB respectively.

Within the biosynthesis of mupirocin, the remaining biosynthetic steps in 9HN formation are not known. This PhD project seeks to elucidate these biosynthetic steps and fully characterise the proteins involved. The overabundance of *trans*-acting catalytic domains – most of which have been proven to be vital for normal PA A production – provides a challenge for elucidation of the *in vivo* biosynthetic steps. Characterisation of these catalytic domains and testing their functions through protein mass spectrometry will be conducted. Heteronuclear single quantum coherence spectroscopy (HSQC) NMR experiments will provide useful information about which ACPs interact with each catalytic domain. With this information, as well as knockout studies and bioinformatic analysis of mupirocin genes, it is hoped the exact chemical steps of 9HN formation can be elucidated.

In addition to understanding the individual biochemical steps involved in fatty acid formation – this project also seeks to understand the vital role played by *trans*-acting ACPs in this pathway. The ability of each ACP to complement each other will be tested. This will provide information about the selectivity of the pathway and the role of the ACP within this. Confirming the role of ACPs and tailoring enzymes in fatty acid construction will allow a route to the biosynthesis of 9-hydroxynonanoic acid to be constructed.

2 Project Aims

The precise biosynthetic steps that generate the 9HN **11** component of mupirocin are not yet known. This PhD project seeks to elucidate these mechanisms and fully characterise the proteins involved. The apparent overabundance of *trans*-acting catalytic domains – most of which have been proven to be vital for normal PA A production – makes this challenging. In this work, the cloning, expression and purification of key proteins implicated in the biosynthesis of 9HN will be conducted. Heteronuclear single quantum coherence spectroscopy (HSQC) NMR experiments will provide preliminary information about which domains interact with mAcpD, the protein which carries the 3HP starter unit. This will inform the design of *in vitro* assays with authentic substrates and analogues, with enzymatic reactions monitored by protein mass spectrometry. With this information, as well as knockout studies and bioinformatic analysis of mupirocin genes, it is hoped the exact chemical steps of 9HN formation can be elucidated.



9-hydroxynonanoic acid **11**

*Figure 17: The chemical structure of the fatty acid region of mupirocin, 9-hydroxynonanoic acid **11**.*

In addition to understanding the individual biochemical steps involved in fatty acid formation – this project also seeks to understand the vital role played by stand-alone ACP domains in this pathway. The ability of each ACP to complement each other during key biosynthetic steps will be tested. This will provide information about the selectivity of the pathway and the role of the ACP within this.

Broadly the specific aims of this project can be split into multiple sections:

1) Further work regarding the formation of 3HP

Isotopically enriched feeding studies and *in vitro* assays have shown that 9HN biosynthesis starts with the formation of 3HP and further studies of the mechanism of 3-hydroxypropanoate (3HP) formation will be undertaken. These studies will seek to show the production of an aldehyde intermediate during the MupS catalysed reduction. In conjunction with Dr Williams, the NMR structure of mAcpD will be solved and the interaction between mAcpD with each of MupQ and MupS will be tested through HSQC NMR titrations. In addition, the role of the C-terminal extension will be tested through production of mAcpD mutant lacking this part of the amino acid sequence.

2) Elucidating the identity of enzymes that may interact with mAcpD and perform the first modification of the 3HP group

Three rounds of fatty acid chain extension are expected to occur to convert 3HP to 9HN. The enzyme candidate that first interacts with 3HP-mAcpD will be identified by ^{1}H-^{15}N HSQC NMR titrations of mAcpD with the KS or KS-like enzymes MmpF, MmpB_KS and MupB. Following this, *in vitro* assays will seek to experimentally verify the subsequent steps in 9HN biosynthesis using authentic substrates.

3) Understanding the identity of enzymes responsible for tailoring of 9HN moiety.

Each fatty acid extension step involves a condensation and reductive cycle through the action of a KR, DH and ER. The enzymes responsible for these reductive steps will be elucidated and the selectivity of these enzymes with natural and non-natural ACP bound substrates will be tested. The ability of these enzymes to accept the correct substrates attached to non-cognate ACPs will also be tested.

4) Elucidating the identity of KS enzyme responsible for each DCC reaction during construction of 9HN.

The final aim is to map out each KS catalysed elongation reaction within the biosynthesis of 9HN. Synthetic pantetheine analogues will be attached to cognate and non-cognate ACPs and their function with each KS domain will be tested. Functional studies will take the form of *in vitro* assays with purified proteins as monitored by protein mass spectrometry. The identity of each ACP to deliver a malonyl moiety during KS-catalysed elongation will also be tested. This will allow further understanding of the roles of all ACPs within the biosynthesis of 9HN.

3 Results and Discussion

The biosynthetic cluster of mupirocin contains many ACP domains. As well as containing ACPs covalently tethered as part of complex type 1 PKS systems many discrete ACP domains are observed. The nomenclature for these ACPs is based on their position in the mupirocin genetic cluster. Five of these *trans*-acting ACPs have been observed. These are identified as mupirocin ACPs (mAcp) followed by A,B,C,D and E.

3.1 Mechanistic and Structural Investigation of 3HP Formation

3.1.1 Mupirocin ACPs contain C-terminal extensions

Amino acid sequence analysis of the mupirocin biosynthetic gene cluster (BGC) has shown there is an apparent overabundance of ACPs. Until recently, of the five discrete ACPs, only mAcpC had been assigned a definitive role. mAcpC had been previously identified as part of the β-branching cassette involved in C-3 β-methyl incorporation and specifically acts as the donor ACP for acetyl delivery to the HMG Cassette.[79] Recently it has been shown by the Crump group that mAcpD is involved in generation of the 3HP fatty acid starter unit, in conjunction with tailoring domains MupQ and MupS. Sequence analysis of mAcpD has revealed that it contains a C-terminal extension in addition to the classic four helix bundle present in most ACPs (Figure 18). [36] This unusual extension has also been shown to be important for function. [71] A further ACP, mAcpA, also shares this unusual C-terminal extension. The C-terminal extension of mAcpD is composed of predominantly hydrophobic and negatively charged residues whereas this region of mAcpA predominantly consists of alanine residues, interspersed with isoleucine, arginine and serine. Despite this structural homology, previous work has shown that mAcpA and the thiomarinol homolog tAcpA are unable to complement mAcpD/tAcpD respectively in the formation of 3HP or 4HB.[86] This chapter will focus on further mechanistic and structural investigations of the role of mAcpD, and specific work undertaken as part of the preparation of the work published in 2019.

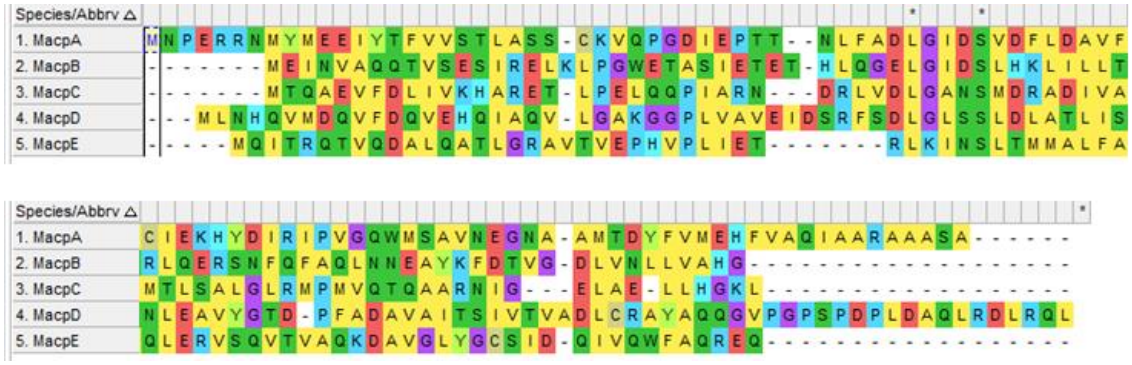
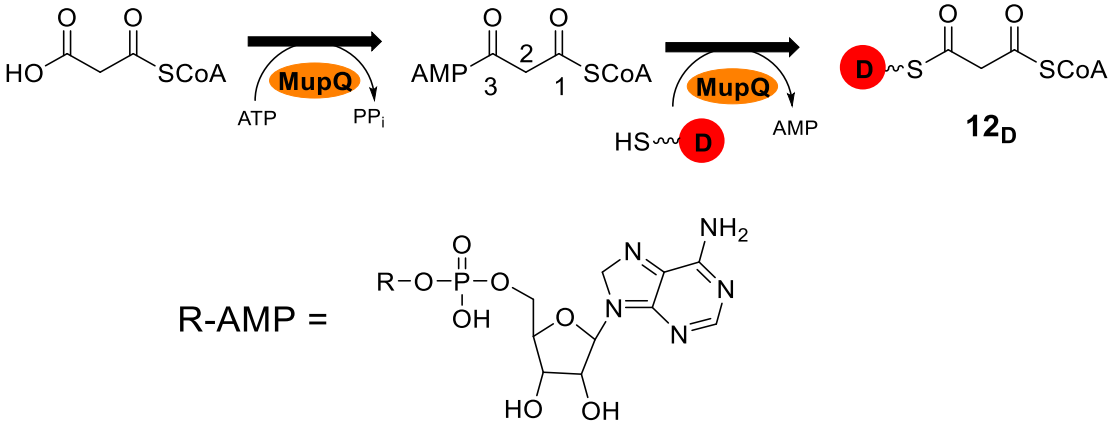


Figure 18: A figure showing the alignment of the 5 stand-alone ACPs within the mupirocin cluster. mAcpA and mAcpD both contain a longer than normal C-terminal extension as can be observed in the above alignment.

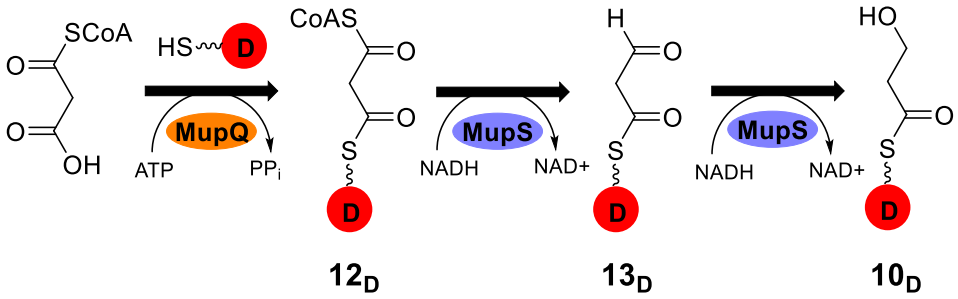
3.1.2 Aldehyde formation during 3HP biosynthesis

The formation of 3HP-mAcpD through action of MupQ and MupS was confirmed by *in vitro* assays in previous work (see Introduction).[86] MupQ shows homology to Fatty Acyl: Adenylate ligase (FAAL) di-domain, AmpG, observed in ambruticin biosynthesis. [88] MupQ functions by initially catalysing activation of malonyl CoA with ATP to form an adenyl monophosphate ester. This activated intermediate is then covalently attached to the terminal thiol of *holo*-mAcpD yielding bis-adenylate intermediate attached to mAcpD **12$_D$** as shown in Scheme 6.



Scheme 6: A scheme showing mechanism of action of adenylation domain MupQ. Malonyl CoA is activated at C-3 position by addition of AMP moiety from ATP. The activated acyl group is subsequently transferred to the terminal thiol of holo-mAcpD at the activated C-3 position. 'D' label represents holo-mAcpD. The chemical structure of the AMP cofactor is shown.

Following formation of bis-adenylate **12$_D$**, addition of MupS (and NADH) reduces this to the 3HP-mAcpD moiety **10$_D$**.[86] During this two-stage reduction, only **10$_D$** was initially observed by ESI-MS but was presumed to form via aldehyde intermediate **13$_D$**. This is shown in Scheme 7. In this work, attempts were made to confirm the stepwise reduction and to curtail the reductive cycle to isolate and identify **13$_D$** via ESI and phosphopantetheine ejection assay.



Scheme 7: Reaction scheme showing the conversion of malonyl CoA to mAcpD bound 3HP ***10D****. Reduction of bis-thioester intermediate* ***12D*** *is expected to occur in a two-step reduction via an aldehyde* ***13D****.*

To observe **13D**; the amount of NADH was reduced in the final MupS reductive step. MupQ intermediate **12D** (100 µM) was prepared as previously described from *holo*-mAcpD.[86] Once formed, the reaction mixture was split into three separate assays (S1, S2 and S3) containing NADH concentrations of 1000, 300 and 100 µM respectively. Each assay was initiated by the addition of 5 µM MupS and monitored by ESI-MS. After 90 minutes, Assay S1 and S2 showed varying ratios of the non-reduced ACP bound substrate **12D** (observed; 16097 Da, expected; 16098 Da) and **10D** (observed; 15334 Da, expected; 15335 Da) but no **13D**. The final assay S3 was also checked after 90 minutes. This showed multiple species. The major species was MupQ bis-adenylate **12D** but a smaller amount of a new species with a mass of 15331.8 Da was also observed (Figure 19). This is consistent with the expected mass of **13D** (expected; 15332.9 Da). Further fragmentation of this species gave the characteristic ejection ion of 331.14 Da confirming the formation of aldehyde intermediate **13D** (expected; 331.13 Da) not 3HP product (expected 333.15 Da). These results confirmed that conversion of malonyl CoA **9** to mAcpD-3HP **10D** proceeded *via* mAcpD bound aldehyde intermediate **13D**.

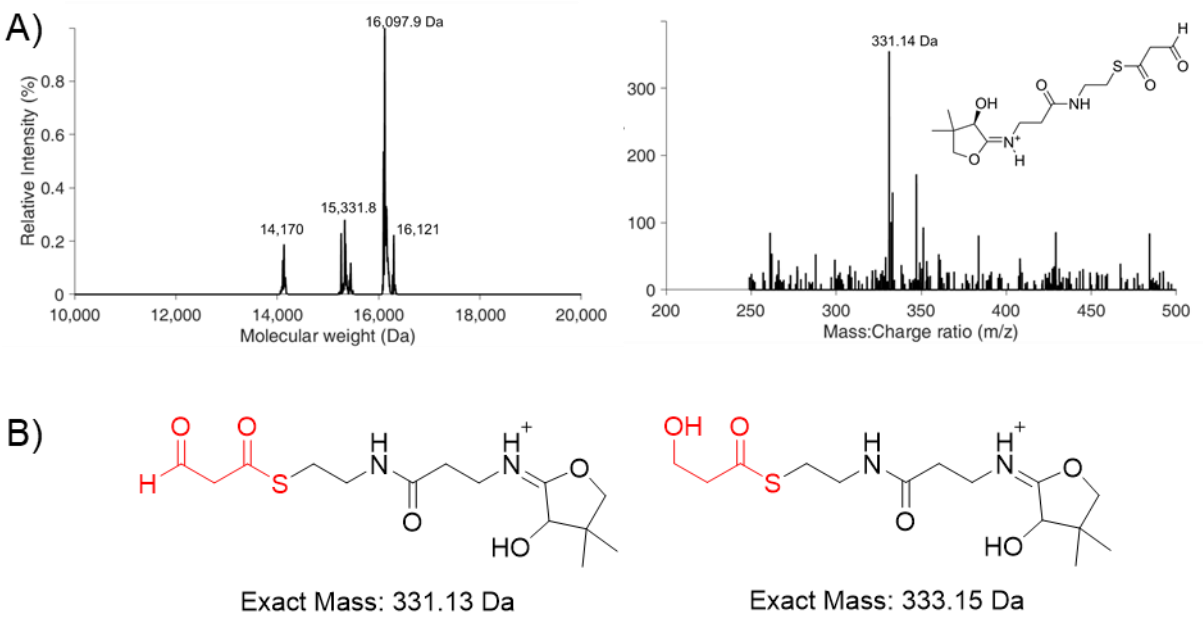


*Figure 19: Curtailed reduction assay S3 after 90 minutes. A) The deconvolution of Assay S3 (left). Pantetheine ejection of 15331.8 Da species (right). B) The chemical structure and expected masses of the Ppant ions of aldehyde **13** and 3HP **10**.*

3.1.3 mAcpD structural characterisation by NMR spectroscopy

Genetic analysis of stand-alone ACPs contained within mupirocin BGC was conducted by Shields *et al.* [71] To investigate the importance of the C-terminal extension of mAcpD, the ACP with the C-termini deleted (ΔmAcpD) and cloned into a plasmid containing all other BGC genes. When compared to a plasmid containing wild type (WT) mAcpD, BGC with mutant mAcpD only produced 6% titre of mupirocin.[71] This confirmed the importance of C-terminal extension of mAcpD. Work was conducted to understand the role of this unusual structural feature.

To probe the structure of the C-terminal domain of mAcpD, structural characterisation by solution-state NMR was conducted. A ^{13}C and ^{15}N-isotopically labelled sample of mAcpD was expressed and purified to allow application of triple resonance assignment methods.

A plasmid containing mAcpD was available and was used to express ^{13}C,^{15}N-mAcpD in a modified growth protocol. This involved supplementing minimal phosphate buffer with ^{13}C-glucose and ^{15}N-ammonium chloride in which *E-coli* was grown. Detailed growth protocol is given in Methods. After expression, the protein was purified by IMAC followed by size-exclusion chromatography using an S75 column to isolate pure ^{13}C,^{15}N-mAcpD in the *apo*-form as confirmed by an SDS-PAGE gel (Figure 20–SEC band). The polyhistidine tag was removed from ^{13}C,^{15}N-mAcpD by incubation with TEV protease overnight (0.33 µM TEV per mg of mAcpD) followed by negative IMAC and further SEC purification to yield pure cleaved ^{13}C,^{15}N-mAcpD in the *apo*-form as confirmed by SDS-PAGE gel (Figure 20A - Cleaved band) and ESI-MS (Observed: 12414.9 Da, Expected for 100% incorporation: 12432.2 Da, Figure 20B). The ESI-MS mass observed was consistent with 97.4% incorporation of $^{13}C/^{15}N$ into mAcpD. The yield of the protein was 3.5 mg/L.

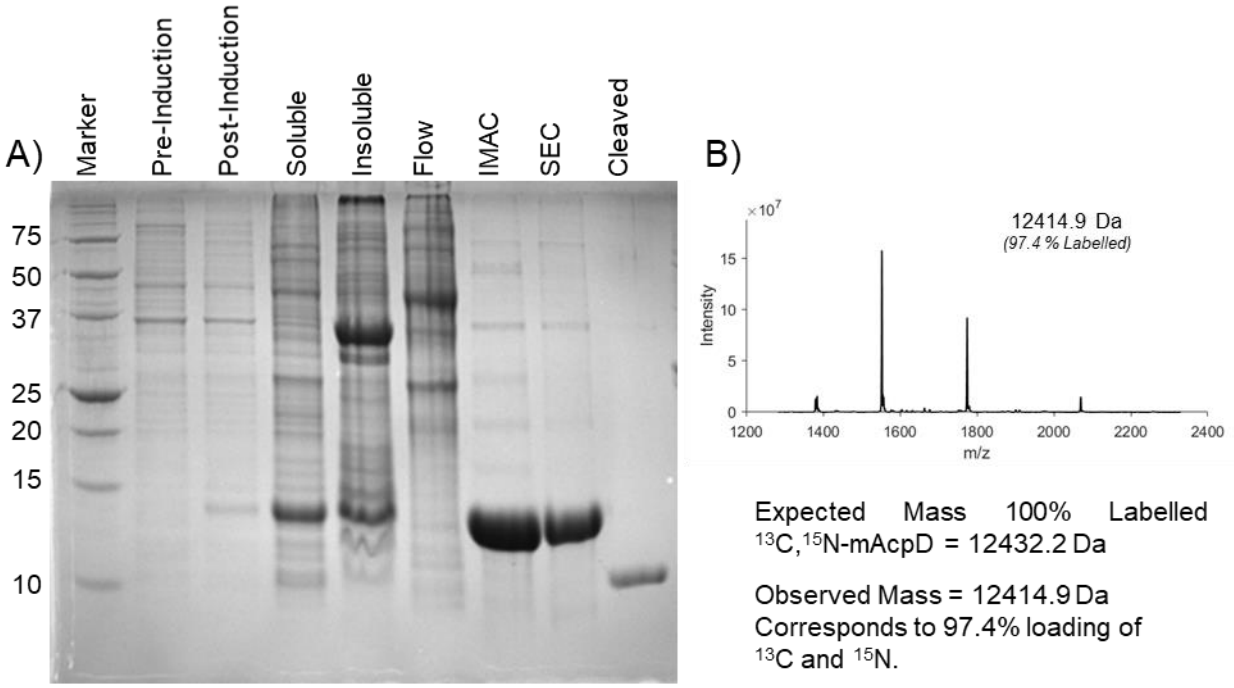


Figure 20: Characterisation of $^{13}C,^{15}N$-mAcpD A) SDS-PAGE gel showing the various stages of $^{13}C,^{15}N$-mAcpD purification. B) ESI-MS showing purified cleaved $^{13}C,^{15}N$-mAcpD. The mass of protein corresponds to approximately 97.4% labelling of ^{13}C and ^{15}N into the backbone of mAcpD.

$^{13}C,^{15}N$-mAcpD in the *apo*-form was added to 25 mM sodium phosphate buffer (pH = 6.0) with 10% D_2O to give a protein concentration of 200 µM. This was added to Shigemi NMR tube and run on Varian VNMRS 600 MHz spectrometer. The backbone and sidechain atoms were assigned using standard triple resonance experiments by Dr Williams as is described elsewhere.[86]

The NMR structure of mAcpD revealed the core of the protein comprised by a 4 α-helix bundle typical of bacterial type II ACPs (Figure 21). Helix 3 is shorter and perpendicular to helix 1, 2 and 4 as shown in Figure 21A. The structure agrees with secondary structural predictions using modelling software.[36] The conserved serine residue, which is key for binding phosphopantetheine, was found at the base of helix 2.

Partial helicity was found in the extended C-terminal domain (helix 5) and limited interactions with the core of the protein was observed. Analysis of the 20 lowest energy structures all showed flexibility in the C-terminal extension when compared to the rigid 4 helix core of the protein. This shows that helix 5 is not likely to be involved in stabilisation of the core 4-helix bundle of mAcpD. Instead, one possibility is that this helix may interact with potential protein partners and facilitating their interaction with mAcpD. In addition to this, the exposed faces of helices 2 and 3 project a negatively charged/ polar face which may also interact with downstream protein partners. This negative charge is provided by the Asp and Glu residues labelled in Figure 21B.

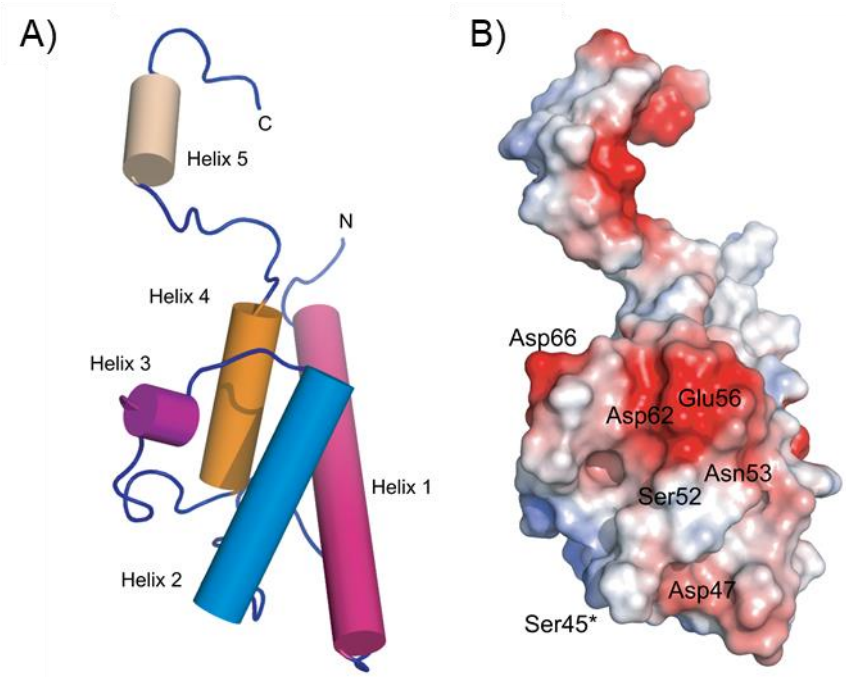


Figure 21: Three-dimensional solution structure of mAcpD A) Structure of mAcpD showing 4 helices (1-4) which make up the core of the protein. B) Surface charge map of mAcpD – negative charge in red and positive charge shown in blue. Key residues are highlighted, Ser45 is the conserved serine residue. Helix 2 and 3 have a predominantly negative charge due to the highlighted residues.

3.1.4 *In vitro* assays with mAcpD C-terminal mutant

The NMR structure showed the C-terminal extension to be a flexible region which does not associate with the core four helix bundle of mAcpD. Secondary structure analysis of mAcpD using JPred4 server is shown in Figure 22A and this was used in conjunction with mAcpD structure to predict the beginning of the C-terminal region. [89] Note, Helix 3 of the core four helix bundle of mAcpD is located between residues 59 and 74 but was not predicted helical by Jpred4 based on its small size. According to this analysis, it was concluded that the C-terminal extension begins at residue 92 of mAcpD WT. A mutant plasmid was designed lacking the C-terminal extension of mAcpD by removal of residues 92 onwards to produce mAcpD_Truncated (mAcpD_T) plasmid. The sequence of this plasmid is shown in Figure 22B. This plasmid was synthesised by ThermoFisher GeneArt. The ability of mAcpD_T to complement mAcpD in the production of 3HP was tested.

Figure 22: A mutant of mAcpD was designed lacking the C-terminal extension. A) Secondary structural prediction of mAcpD using JPred4 server. Regions in red correspond to predicted α-helical secondary structure. C-terminal extension starts at residue 92 onwards. B) Sequence of the mAcpD_T plasmid lacking the C-terminal extension. The residues shown in yellow have been removed to form mAcpD_T plasmid.

mAcpD_T was expressed under standard growth conditions (Methods) and purified by IMAC followed by S100 size exclusion chromatography to yield ~12 mg/L of *apo*-mAcpD_T. This protein species was characterised by SDS-PAGE gel to confirm purity (Figure 23A). Further analysis of mAcpD_T by ESI-MS showed two main species corresponding *apo*-mAcpD_T (observed; 13368 Da, expected; 13367 Da) and phosphogluconoylated *apo*-mAcpD_T (observed; 13545 Da, expected; 13545 Da, Figure 23B and C). Phosphogluconoylation of the polyhistidine tag is a spontaneous post-translational modification of proteins expressed in *E. coli* and leads to a +178 Da mass increase of his-tagged proteins as is reported elsewhere. [90] Removal of the polyhistidine tag will remove this mass discrepancy and produce a single protein species. Therefore, mAcpD_T was incubated with TEV protease overnight (0.5 µM TEV per mg mAcpD_T). The cleavage reaction was monitored after 16 hours by ESI-MS. This showed complete removal of the polyhistidine tag. The cleaved protein was further purified by negative IMAC and desalted to produce cleaved, *apo*-mAcpD_T in 7 mg/L yield as confirmed by ESI-MS (Observed; 10215 Da, expected; 10211 Da, Figure 24)

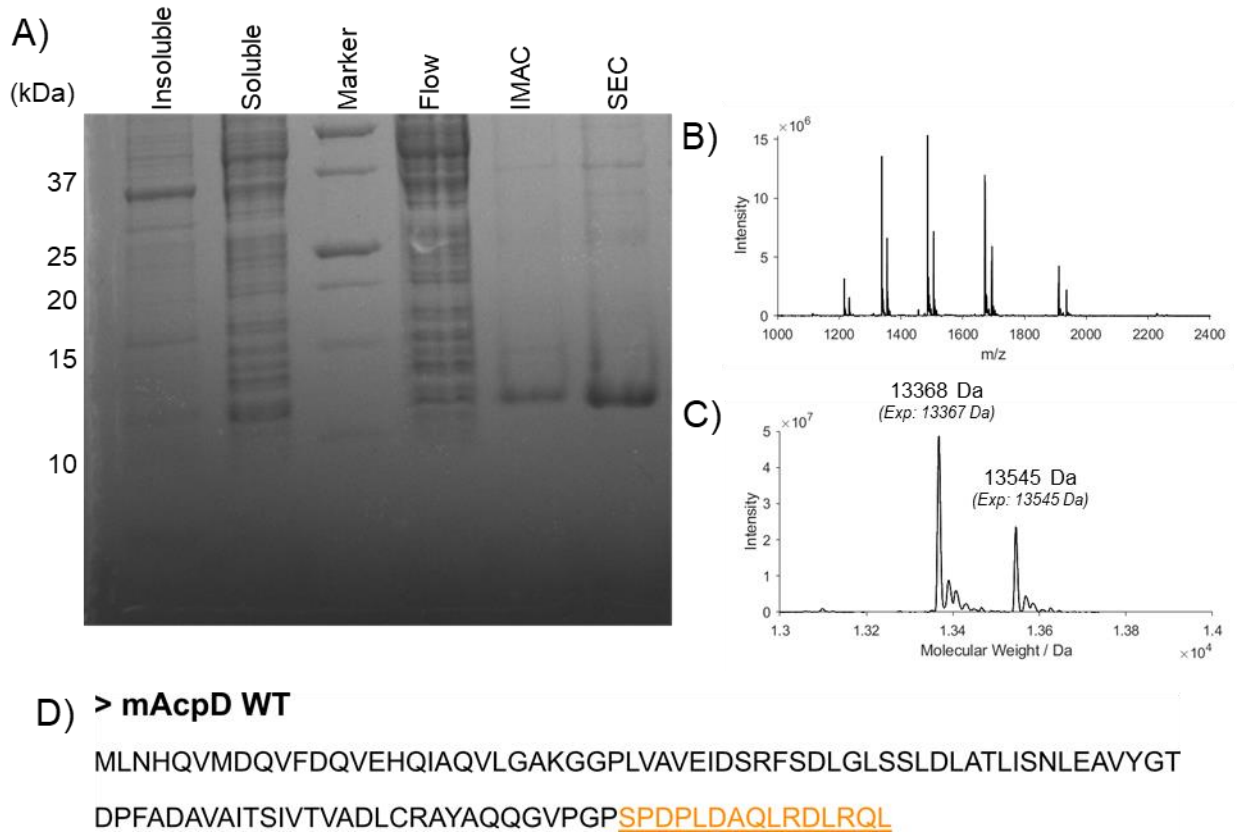


Figure 23: Characterisation of mAcpD_T A) SDS-PAGE gel showing the stages of mAcpD_T purification. B) ESI-MS analysis showing the charge states of purified non-cleaved apo-mAcpD_T. C) Deconvolution of purified non-cleaved apo-mAcpD_T. D) The sequence of wild type mAcpD showing the C terminal extension which has been deleted to give the mAcpD_T mutant (shown in yellow).

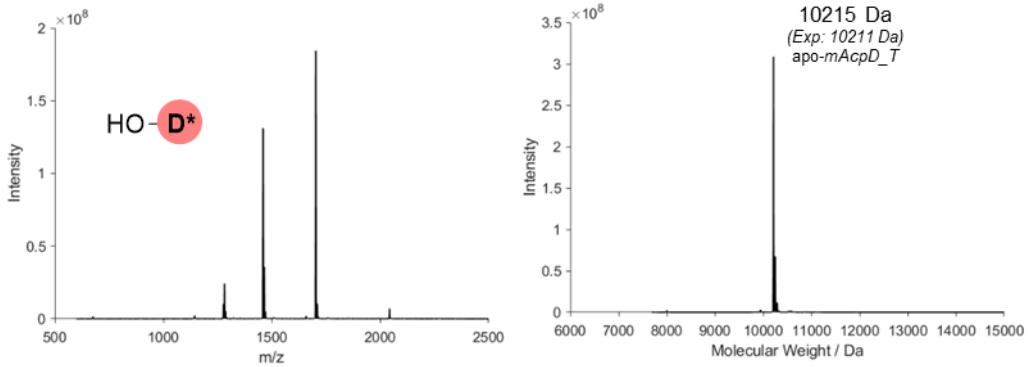


Figure 24: ESI-MS showing formation of cleaved apo-*cleaved mAcpD_T mutant (shown as D* in the image). Charge states (left) and subsequent deconvolution (right) are shown.*

mAcpD_T was purified entirely in the inactive *apo*-form. ACPs are often converted to *holo* form due to interaction with *E. coli* holo-ACP synthase. This E coli protein has been previously observed to have a broad substrate specificity to attach the 4-phosphopantethine group to non-endogenous ACP domains.[91] This was not observed in this case and so conversion to the *holo*-form was necessary to have functional ACP for *in vitro* assays. This ACP conversion was done using purified phosphopantetheinyl

transferase enzyme MupN which has previously been shown to convert mupirocin ACPs to their *holo* forms.[71] This reaction was initiated by the addition of MupN (5 µM) to cleaved *apo*-mAcpD_T (100 µM), $MgCl_2$ (10 mM) and Coenzyme A (1 mM). The assay was monitored by ESI-MS after 2 hours. After this time, the charge states corresponding to *apo*-mAcpD_T were consumed, and a new species was observed. This new species had a mass consistent with *holo*-mAcpD_T (observed; 10554 Da, expected; 10551 Da, Figure 25). This confirmed the successful formation of *holo*-mAcpD_T.

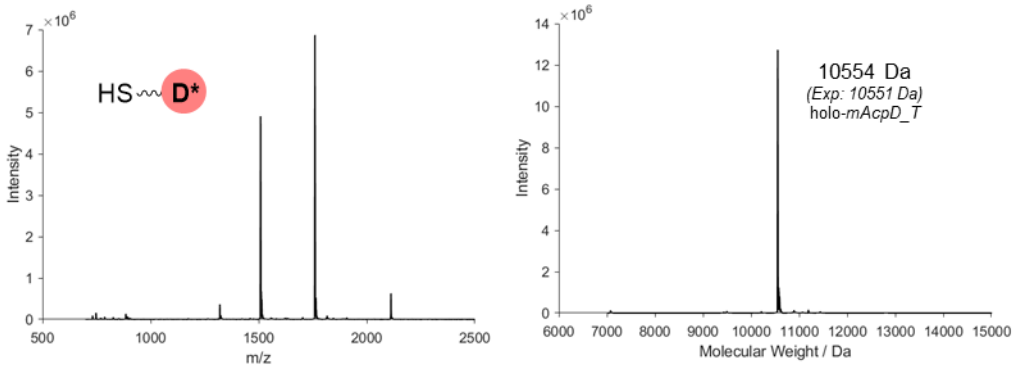


Figure 25: ESI-MS showing the complete formation of holo-mAcpD_T after incubation with MupN and cofactors at room temperature for 2 hours. Charge states (left) and subsequent deconvolution (right) are shown.

To test the ability of mAcpD_T to complement mAcpD in the production of 3HP $\mathbf{10_{D*}}$, *in vitro* assays were conducted. Note, D* corresponds to the version of mAcpD lacking the C-terminal extension (mAcpD_T). The assay was initiated by the addition of *holo*-mAcpD_T (100 µM) to malonyl CoA (1 mM), ATP (1 mM) and MupQ (5 µM). After 1 hour, ESI-MS showed 30% conversion to bis-thioester intermediate $\mathbf{12_{D*}}$ (observed; 11391 Da, expected; 11386 Da, Figure 26B). The remaining 70% consisted of unreacted starting material *holo*-mAcpD_T (observed; 10554 Da, expected; 10551 Da). Next, MupS (5 µM) and NADH (1 mM) was added and monitored by ESI-MS after a further 2 hours. After this time, changes were observed to the deconvolution of mAcpD_T. A single species was now observed with a mass consistent with $\mathbf{10_{D*}}$ the expected 3HP product (observed; 10627 Da, expected; 10623 Da). Ppant ejection assay of this species produced a mixture of ions. Approximately 20% of the Ppant spectrum showed the formation of 3HP species (observed; 333.17 Da, expected; 333.15 Da) confirming the formation of $\mathbf{10_{D*}}$. The remaining 80% was observed with a mass consistent with *holo*-mAcpD (observed; 261.15 Da, expected; 261.13 Da). These assays have confirmed that a mutant lacking the C-terminal extension of mAcpD is able to interact with MupQ and MupS produce 3HP. Therefore, this C-terminal extension is not expected to be required for interaction between mAcpD and MupQ/MupS.

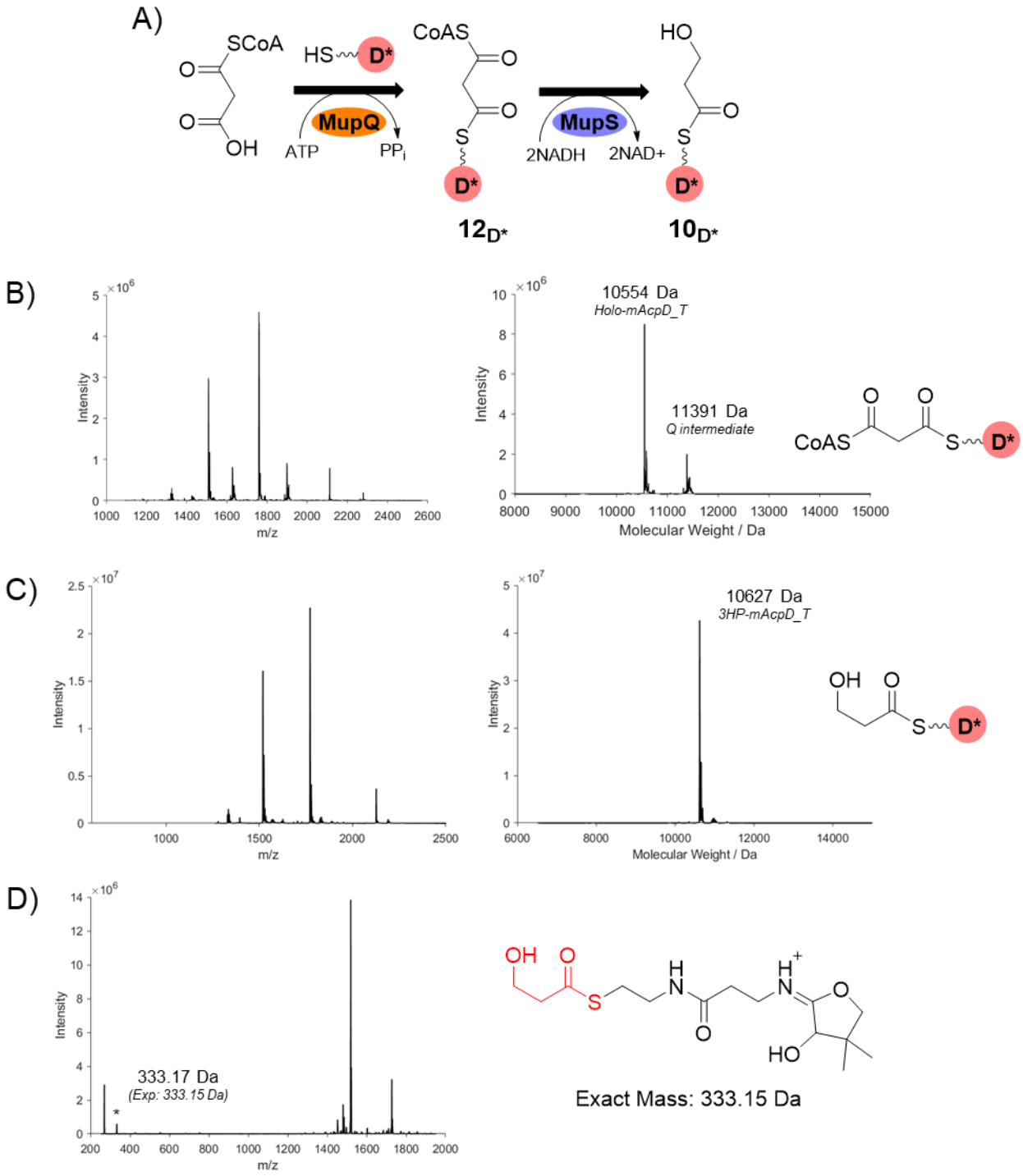


Figure 26: The generation of 3HP using mutant mAcpD_T (D). A) Simplified reaction scheme showing the conversion of malonyl CoA to 3HP-mAcpD_T* ***10D**** *. B) ESI-MS spectra showing the charge states (left) and subsequent deconvolution (right) showing formation of Q intermediate after 1 hour. C) The formation of mAcpD_T 3HP species 2 hours after addition of MupS to Q intermediate. D) Ppant ejection assay of 3HP-mAcpD_T species (left). Asterisked in the ion corresponding to 3HP-mAcpD_T The chemical structure and exact mass of 3HP Ppant ion is shown (right).*

The assays were monitored by ESI-MS and detailed kinetic analysis of the rates of 3HP formation for both mAcpD and mAcpD_T have not been conducted. It is possible that the C-terminal extension helps facilitate the interaction between MupQ/ MupS and mAcpD by increasing the rate of turnover. To provide further evidence for whether the C-terminal extension of mAcpD is required for in interaction with MupQ and MupS, ^{1}H-^{15}N HSQC NMR titrations can be measured using ^{15}N-mAcpD and unlabelled MupQ and MupS partner proteins. The changes to the chemical shift of amino acid residues of mAcpD on addition of each cognate partner protein can potentially provide evidence about the site and nature of the interaction.

3.1.5 Interaction of mAcpD with MupQ/ MupS using ^{1}H-^{15}N HSQC spectroscopy

To characterise the interaction between mAcpD and partner proteins (MupQ and MupS) a chemical shift mapping approach was undertaken. ^{15}N-mAcpD was prepared in minimal media as described in Methods and a ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectra was recorded at either 600 or 700 MHz (courtesy of Dr Chris Williams). Using a spectrum of only mAcpD in conjunction with a series of spectra acquired in the presence of increasing amounts of partner protein (MupQ or S) allows the interaction of the ACP and partner protein to be monitored. Changes to the spectra can involve chemical shift perturbations (CSPs) or peak broadening depending on the strength and kinetics of the interaction. Conversely a lack of changes can be an indication that there is little or no interaction.

3.1.5.1 NMR titration of ^{15}N-mAcpD with MupQ

HSQC titration experiments were conducted using ^{15}N-mAcpD and MupQ. A control ^{1}H-^{15}N HSQC was recorded at 700 MHz with *apo*-^{15}N-mAcpD (100 μM), malonyl CoA (1mM) and ATP (1 mM) in sodium phosphate buffer (20 mM sodium phosphate, 100 mM NaCl, 10% D_2O pH 8.0). FAAL enzymes, of which MupQ is a member, can elicit two conformations during reaction.[92] The adenylation-competent conformation (A-state) facilitates the binding of acyl substrate and subsequent adenylation to produce activated acyl-AMP intermediate. [93] The second conformation is the thioesterification-competent conformation (T-state). During these titrations, addition of ATP (1 mM), malonyl CoA (1 mM) and *holo*-mAcpD was conducted to ensure MupQ would exhibit the reactive T-state. [94,95]

A series of titrations were conducted with increasing ratios of MupQ to mAcpD. The ratios of 1:1, 1.5:1 and 3:1 ratio of MupQ to mAcpD were used (as outlined in Table 1). NMR sample preparation was completed in conjunction with Dr Winter and subsequent NMR analysis was completed by Dr Winter.

Ratio (mAcpD to MupQ)	^{15}N-mAcpD Concentration / µM	MupQ Concentration / µM
1:0	100	0
1:1	60	60
2:3	60	90
1:3	30	90

Table 1: A table showing the various ratios and concentrations used during NMR titrations of mAcpD and MupQ.

Comparison of the amide region of the 1:3 HSQC of mAcpD:MupQ to the mAcpD control showed significant chemical shift changes. Figure 27A shows the amide region of HSQC spectrum of ^{15}N-mAcpD before addition of MupQ and Figure 27B shows the spectrum after addition of 3 equivalents of MupQ. Changes are clearly observed between the two spectra. This is exemplified by looking at Figure 27C which shows a larger image of the region of 7.3 to 8.1 ^{1}H ppm and 116.0 to 120.0 ^{15}N ppm with the control spectrum in blue and MupQ spectrum in red. This region shows the tractability of the CSPs observed for residues Gln9, Gln13 and Gln20. The CSP for all residues were superimposed over the structure of mAcpD with significant CSP shown in green (< 0.02 ppm) and 0.02-0.01 ppm CSPs shown in blue (Figure 27D, conducted by Dr Winter). The CSPs with a with a shift less than 0.01 ppm were not considered significant and are shown in red. Visualising the data in this way shows that the majority of the significant CSPs are located on helix 1 of mAcpD and the loop region between helix 1 and helix 2. A histogram shows the CSP in more detail confirming the location of CSPs observed (Figure 27E).

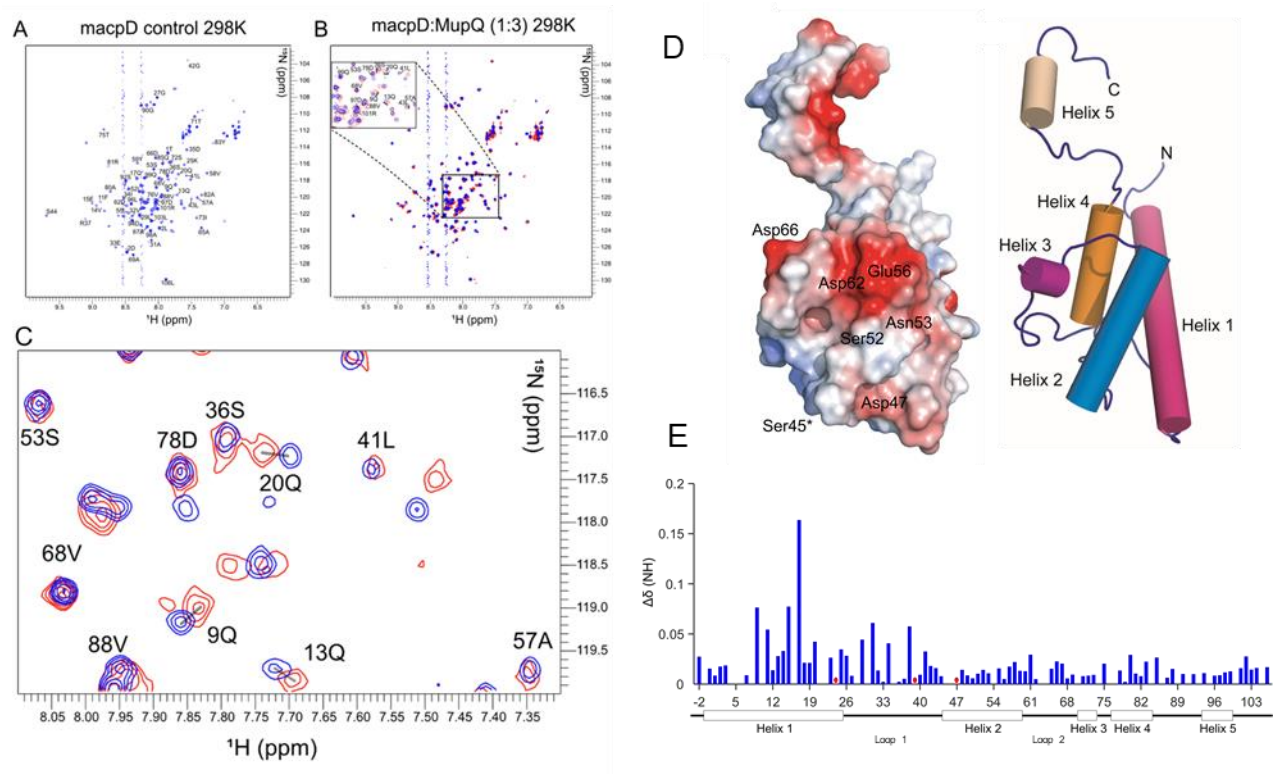


Figure 27: The NMR titration of ^{15}N-mAcpD with MupQ. A) HSQC NMR spectrum showing the amide region of the ^{15}N-mAcpD prior to addition of MupQ. B) Superimposition of control spectrum (blue) and spectrum after addition of 3 equivalents of MupQ (red). C) Spectrum showing the CSP observed between 7.3 to 8.1 ^{1}H ppm and 116.0 to 120.0 ^{15}N ppm. D) Surface charge distribution on mAcpD with conserved serine residue highlighted with asterisk. E) A histogram showing the location and nature of CSPs observed. Red dots indicate broadened residues and those with a CSP of 0 were either proline residues or not visible at pH 8.0.

Analysis of the number and nature of CSPs observed and confirmed that there is a definite positive interaction between ^{15}N-mAcpD and *holo*-MupQ. Both proteins have been shown to interact through experimental assays, so this is to be expected. The residues of mAcpD which have undergone the largest CSPs are located on helix 1 and loop 1. It is the polar residues that have shown the largest CSPs such as Gln9, Gln13, Glu15, Gln17 and Gln20 (Figure 27E). It is not surprising that polar residues are implicated in mAcpD-MupQ interaction as these residues are likely to be solvent exposed making them ideally positioned to interact. [36] A series of hydrophobic residues consisting of Leu29, Ala31 and Phe38 located on loop 1 also show smaller but significant CSPs. No significant CSPs were located on the C-terminal extension of mAcpD providing evidence that this region is not strongly involved in MupQ interaction.

The location of interacting polar residues on helix 1 is surprising. Although no common ACP binding motif has been identified, in FAS and PKS systems, helix 2 is generally thought to be implicated in ACP-protein interactions whereas helix 1 has typically been observed to have a structural role.[96] This said, the loop region of ACPs have previously been implicated in interaction with MupQ homologues. MupQ is a unique member of the 'acetyl-CoA synthetase' superfamily of proteins.[95] Closest

homologues to MupQ are fatty acyl-AMP ligase (FAAL) enzymes such as TamA and MicC from the biosynthesis of tambjamine [97] and micacocidin [98] respectively. FAAL domains from these PKS systems accept long acyl chains and deliver them to ACPs to initiate biosynthesis. MupQ therefore acts in a similar way but interacts with a correspondingly more polar malonyl-CoA delivering it to terminal thiol of mAcpD thereby initiating 9HN biosynthesis. [86] Non ribosomal peptidyl synthases (NRPS) adenylation domains are phylogenetically related to FAAL enzymes (such as MupQ) with both domains being class I adenylate-forming enzymes. The interactions between NRPS adenylation domains and their respective peptidyl carrier protein (PCP) have been well characterised. Within the enterobactin NRPS system, [99] the interaction between adenylation domain EntE and aryl carrier protein (ArCP) containing EntB has been studied. Mutation of residues located on ArCP of EntB were successful in picking out interacting residues. [100] This showed that polar residues located on loop 1 and helix 2 were implicated in binding. This is similar to the MupQ – mAcpD interaction in which polar residues located on helix 1 and loop 1 were found to interact the strongest. Therefore, it seems likely that loop 1 and helix 1 is the region on mAcpD which is responsible for interaction with MupQ.

The residues responsible for ACP docking withinMupQ have been predicted based on similarity to other members of the ANL superfamily. The residues Glu277, Gly319, Asp337, Arg351 and Lys455 are universally conserved among members of the MupQ superfamily. [101] Based on the location of these important residues, a putative binding pocket has been predicted for MupQ. [86] This binding pocket consists of predominantly positively charged residues Arg201, Arg232 and Lys379. These positively charged residues are the correct polarity to favourably interact with the negatively charged residues on helix 1, loop 1 of mAcpD such as Gln9, Gln13, Glu15, Gln17 and Gln20 on mAcpD.

3.1.5.2 NMR Titration of ^{15}N-mAcpD with MupS

NMR titrations were conducted with increasing concentrations of MupS. The concentrations and ratios of ACP and partner protein are shown in Table 2.

Ratio (mAcpD to MupS)	^{15}N-mAcpD Concentration / µM	MupS Concentration / µM
1:0	100	0
1:1	50	50
1:2	50	100

Table 2: A table showing the various ratios and concentrations used during the NMR titration of mAcpD and MupS

As with the MupQ titration, an observable interaction was expected between mAcpD and MupS. When MupS was added to ^{15}N-mAcpD in a 2:1 ratio a number of significant changes to the ^{1}H-^{15}N HSQC were observed (Figure 28B). Unlike the titration with MupQ, the majority of the ^{1}H-^{15}N correlations were broadened and no longer visible. This was indicative of significant interaction between mAcpD and MupS. Analysis of the correlations that remained visible revealed that these included residues S92,

D94, L96, A98, R101 and L103. These residues are all located in the C-terminal region of mAcpD and their persistence in the complex between the two proteins suggests little interaction with MupS. This corroborated the results of *in vitro* assays which showed successful production of 3-hydroxypropanoate attached to mAcpD_T mutant catalysed by MupS.

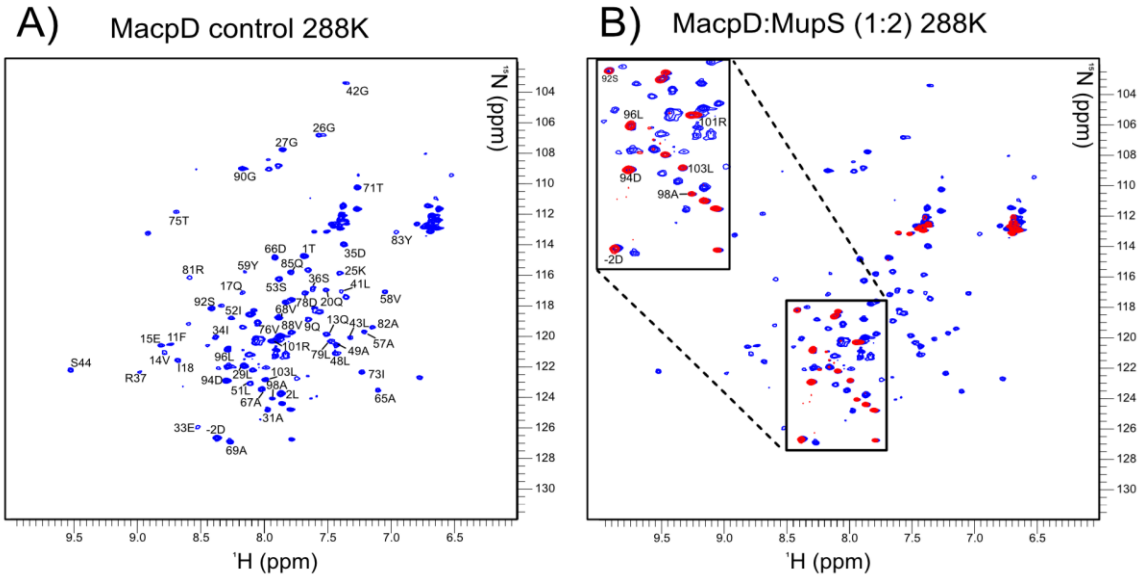


Figure 28: The HSQC NMR titration of ^{15}N-mAcpD and MupS. A) Figure showing the HSQC spectrum of amide region of ^{15}N-mAcpD prior to addition of MupS. B) Figure showing the HSQC spectrum of amide region of ^{15}N-mAcpD after addition of 2 equivalents of MupS (red) superimposed over the control spectrum (blue). The amide residues which remain are S92, D94, L96, A98, R101 and L103, these are exclusively located on the C-terminal extension of mAcpD.

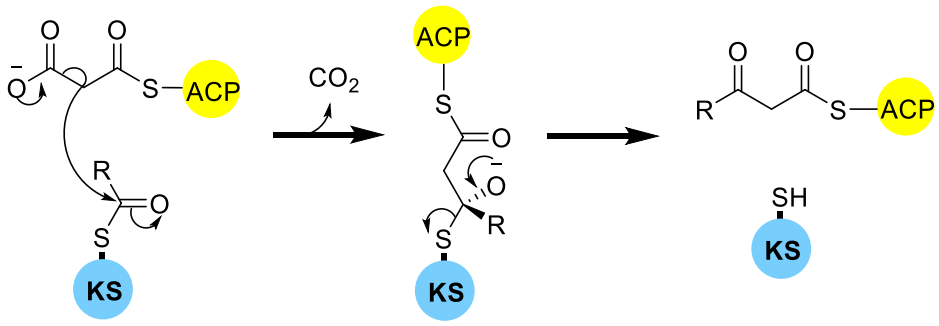
The HSQC NMR titrations and *in vitro* assays conducted in this section provide evidence that the C-terminal extension is not required for interaction between mAcpD with MupQ /MupS. One possible suggestion is that the C-terminal extension may be involved with interaction with another protein partner further downstream in the pathway which will be tested in later sections.

The next goal of this project was to understand the domains responsible for further processing of the mAcpD-3HP moiety in the overall biosynthesis of the fatty acid fragment, 9HN. This will lead to the elucidation of the discrete enzymatic steps required to form 9-hydroxynonanoic acid (9HN) from 3HP.

3.2 Biosynthetic Steps Following 3HP Formation

At the outset of this project, the precise chemical steps responsible for the formation of 9-hydroxynonanoic acid (9HN) were not known. Typically, fatty acid biosynthesis proceeds through the recursive use of a conserved series of catalytic domains, a KS, KR, DH and ER. The biosynthetic gene cluster (BCG) of mupirocin, however, does not contain a single discrete set of these enzymes that are attributable to this sole function. The picture is further complicated by the requirement for incorporation of an unusual starter unit 3HP into the full length 9HN. [86] 3HP is expected to undergo three rounds of fatty acid chain extension and reduction to produce 9HN although this has not been proven experimentally. The precise timing of the extension is also uncertain and may occur before or after esterification with monic acid.

Each chain elongation step in fatty acid biosynthesis proceeds *via* a conserved mechanism known as a decarboxylative Claisen condensation (DCC) catalysed by a ketosynthase (KS) domain. During this reaction, the acyl carbon chain length is extended by two carbon atoms (Scheme 8). Investigation of the BGC revealed three enzymes which contain KS-like sequence homology which are candidates for these elongations: the modular MmpB_KS, the mini module MmpF and the discrete MupB. It was therefore possible that either one or all these domains might play a role in biosynthesis of 9HN. Bioinformatic analysis of MmpF, MmpB_KS and MupB was initially conducted to explore potential roles of each of these catalytic domains.



Scheme 8: The carbon-carbon bond forming mechanism of decarboxylative Claisen condensation (DCC) catalysed by the KS. Loss of carbon dioxide and concerted formation of a carbon-carbon bond occurs between ACP-bound malonyl and the growing acyl chain.

3.2.1 Bioinformatic analysis of MmpF, MmpB_KS and MupB

3.2.1.1 Analysis of the KS domain of MmpF

Within the biosynthetic cluster of mupirocin, the mmpF gene is co-located immediately downstream of mAcpD, mupQ and mupS genes making it a good candidate for downstream processing of the starter

unit mAcpD-3HP (Figure 29). The first KS enzyme is expected to condense this mAcpD-3HP moiety. The protein, MmpF, consists of a condensing KS, an α/β fold docking domain and an α/β hydrolase domain. The grouping together of a KS and acyl transferase (AT) domain in this way is homologous to the KS-AT di-domain within mammalian FAS and type 1 PKS systems suggesting the hydrolase domain may possess AT activity. [102,103,104]

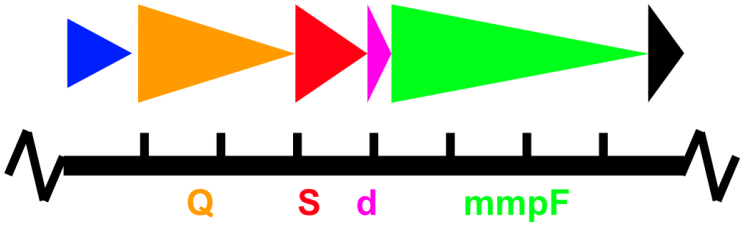


Figure 29: A representation of the organisation of the regions of DNA within the mupirocin gene cluster highlighting the relative locations of of mupQ (orange), mupS (red), mAcpD (d, pink) and mmpF (green) genes. The arrows indicate direction the genes are read.

Within type II FAS systems, three KS enzymes are observed; KAS I, II and III which correspond to *E. coli* FAS enzymes of FabB, FabF and FabH respectively.[73] Typically, FabB and FabF are involved in elongation of the fatty acid chain and condense an acyl-ACP with malonyl-ACP. FabH is involved in initiation of the FAS cycle and will accept an acetyl-CoA moiety to the active site cysteine before catalysing condensation with malonyl-ACP. [18] The KS domain of MmpF consists of a Cys-His-His active site, which is typical of elongating KSs (FabB/FabF) rather than the Cys-His-Asn typically observed in initiating KSs. [105]

Elongating KSs such as MmpF adopt a ping-pong mechanism. First, the active site cysteine of KS accpets the acyl group being extended followed by nucleophilic attack at the cysteine thioester to produce the condensed product. Overall, the KS region of MmpF is a good candidate for one or more rounds of fatty acid chain extension during 9HN biosynthesis.

3.2.1.2 The inactivity of MmpF_AT region

The mupirocin BGC is a *trans*-AT PKS system – where individual modules (such as those observed in MmpA and MmpD) lack intrinsic AT domains. The priming of ACP domains with malonyl is instead catalysed by *trans*-acting MmpC which contains the required AT functionality. [106] The presence of an AT domain on MmpF is therefore seemingly surplus to requirement.

The 'AT' domain of MmpF shows strong homology with 'AT' domain of OzmQ from the biosynthesis of the oxazolomycins. [107] The oxazolomycins are a family of antibiotics produced by many *Streptomyces* species. [108] Many similarities exist between the biosynthetic pathway of mupirocin and oxazolomycin. Both systems are examples of *trans*-AT PKS systems with AT function provided by didomain AT (OzmM in oxazolomycin and MmpC in mupirocin). Both systems also utilise the

incorporation of an unusual starter unit during their biosynthesis. In mupirocin, this starter unit is 3HP and in oxazolomycin an unusual formylglycyl moiety is used (Figure 30). The unusual starter unit in oxazolomycin is condensed by KS enzyme OzmQ which shows genetic similarity to MmpF from mupirocin.

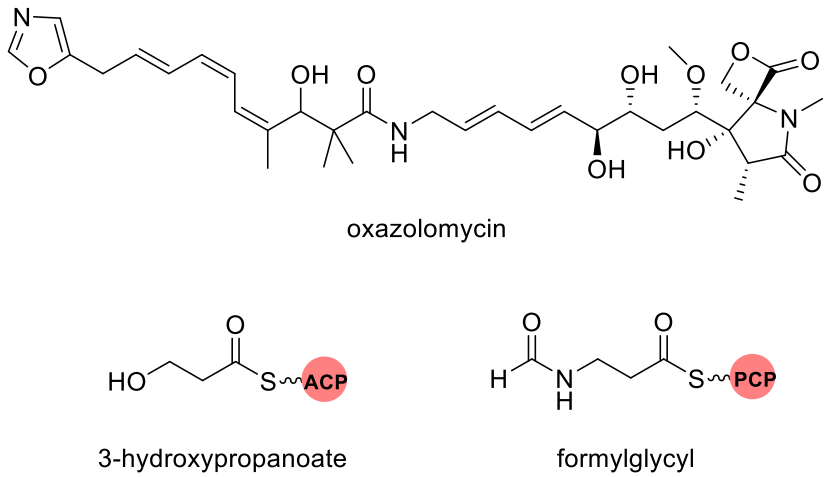


Figure 30: The chemical structure of oxazolomycin. The unusual starter units utilised in biosynthesis of mupirocin (3-hydroxypropanoate) and oxazolomycin (formylglycyl) is shown.

The crystal structure of OzmQ and related *trans*-AT KS domains were solved, and structural analysis conducted by Lohman *et al*. [109] The AT domain of OzmQ was found to consist of only 131 residues which is much shorter than the observed AT domain of a typical KS-AT system. A comparison of OzmQ can be made to KS-AT di-domain of CurL which contains a functional AT module from the *cis*-AT PKS responsible for the biosynthesis of antitubulin natural product curacin A. [110] The function of the CurL module is chain elongation and tailoring of the resultant diketone to produce the C-13 methoxy functionality on curacin A. The crystal structure of the KS-AT di-domain of CurL was elucidated (PDB accession: 4MZ0) and the AT domain characterised. This confirmed the AT domain of CurL was 329 amino acids in length, containing twice as many residues as the 'AT' domain of OzmQ and nine times the length as 'AT' region of KS-DD-AT module MgsF_KS4 from the iso-migrastatin polyketide synthase. [111] The shorter than expected AT domains found within OzmQ_KS and MgsF_KS4 shows they are not functional within their respective BGCs.[109] The AT domain of MmpF is 155 residues long which is comparable to the 131 residue length of OzmQ and far shorter than the active AT domain of CurL. This provides evidence that the 'AT' domain of MmpF is also likely to be inactive.

Lohman et at suggests that the reason for the inactive AT domains observed in BGC of various *trans*-AT KS domains throughout nature is evolution caught in the act. *trans*-AT systems are expected to have evolved from their *cis*-AT counterparts. As a result, some *trans*-AT KSs still contain non-functional remnants of this evolutionary action. [20,109] In other words, these AT fragments evolutionary

intermediates between the type I PKS and AT-less type I PKS architectures. These AT domains are compared to the full length KS-AT didomain in Figure 31.

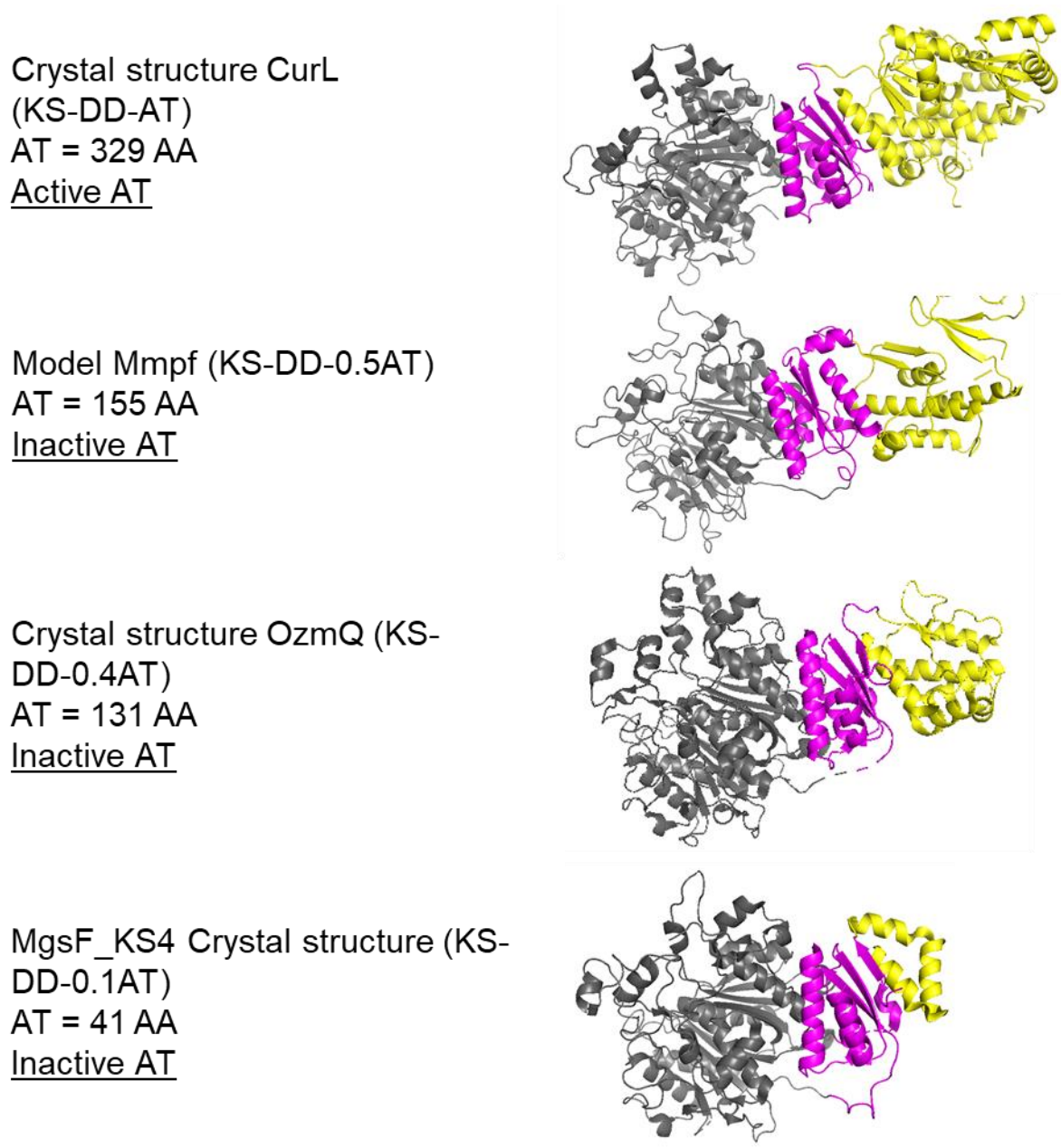


Figure 31: Comparison of KS-dd-AT domain proteins showing inactive AT domain on MmpF and other selected KS enzymes. KS domains are shown in grey, docking domain (dd) shown in pink and AT domains shown in yellow.

3.2.1.3 Bioinformatic Analysis of the KS domain of MmpB

MmpB is a multidomain FAS containing KS, DH, KR, ACP (tri-domain) and TE covalently bound in a type I system. All the domains required for fatty acid chain extension and reduction are present on MmpB apart from AT domain (expected to be provided by *trans*-acting MmpC). The structure of MmpB_KS was modelled using I-Tasser and C164, H299 and H339 were found to make up the active site of the enzyme (Figure 32). This is the same active site as is observed in MmpF and is indicative of

an elongating KS. This provides evidence the enzyme is active and may provide one or multiple rounds of chain extension during 9HN construction. MmpB also contains at thioesterase domain (TE) which is expected to release the fully formed 9HN from its cognate ACP. The presence of TE domain implies that at least the final round of chain extension will be catalysed by MmpB_KS.

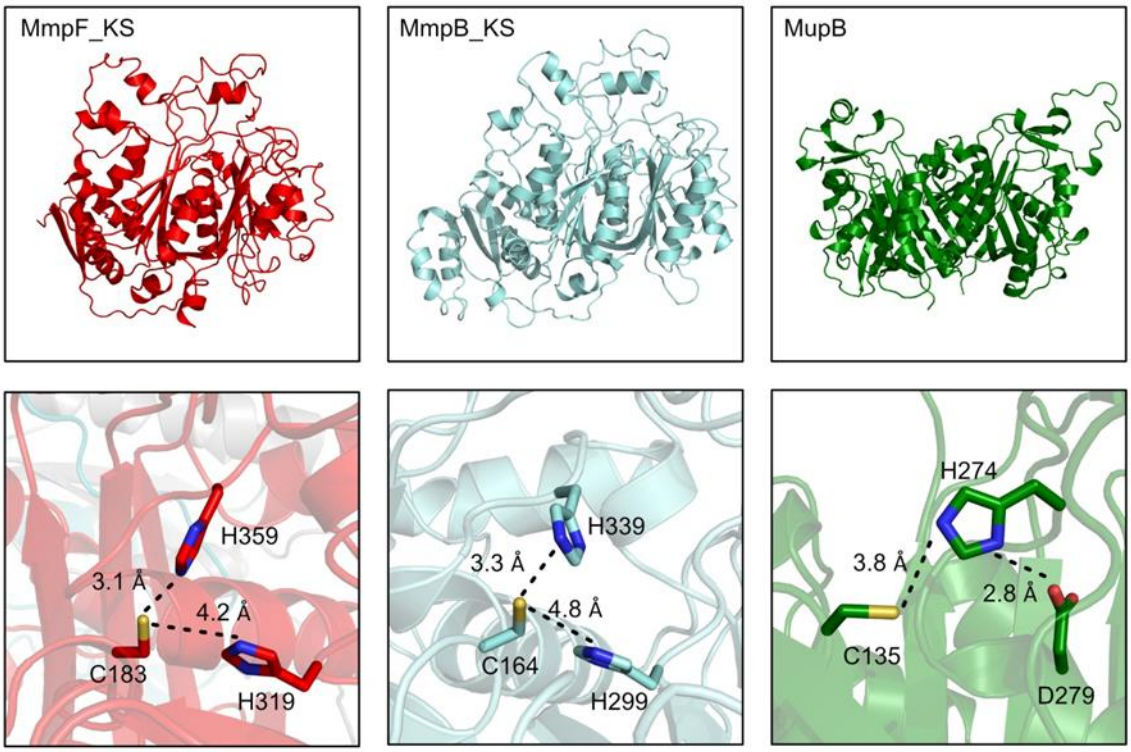
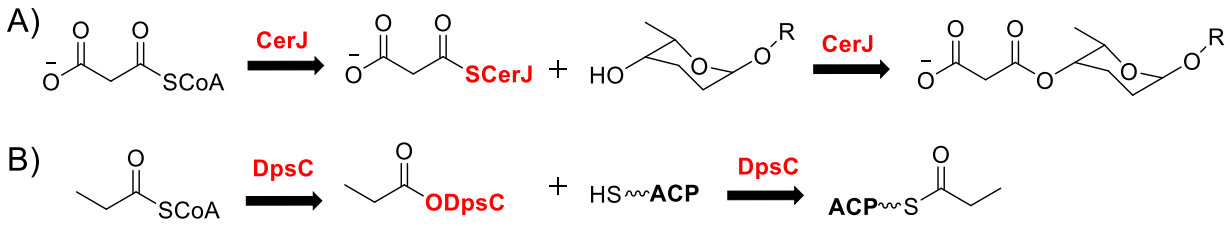


Figure 32: Comparison of the three KS enzymes which may be involved in 9HN assembly. Structures of MmpF_KS and MmpB_KS were modelled using iTasser, whereas the crystal structure of MupB has been solved. Catalytic sites of MmpF and MmpB_KS are consistent with KAS I/II elongating enzymes. The active site of MupB is consistent with a subclass of KAS III enzymes which catalyse unusual acyl transferase activity and may instead catalyse the esterification of 9HN with monic acid.

3.2.1.4 Bioinformatic Analysis of MupB

Analysis of MupB showed an active site triad consisting of Cys-His-Asn residues that is clearly differentiated from MmpF and MmpB_KS. An Nuc-His-Asn active site (where Nuc is a nucleophilic Cysteine or Serine residue) is observed in enzymes such as DspC and CerJ of which MupB shows similarity to (20.8% and 23.4% percent identity respectively). [112,113] The 'CerJ' subclass of KS enzymes has been shown to undergo unusual acyl transferase (AT) activity. To this end, CerJ is expected to accept acyl moieties and catalyse their linkage to other acyl moieties as shown in Scheme 9. This shows the ability of CerJ to accept malonate via its nucleophilic cysteine followed by transfer to hydroxyl group of six-membered ring. In an analogous mechanism, MupB is predicted to interact with pre-monic acid and transfer this to the hydroxy group fatty acid moiety of unknown length. This moiety may be the 3-carbon mAcpD-3HP which would imply direct interaction between mAcpD and MupB. It is also

possible that this transferase activity may occur after complete formation of 9HN, in which case MupB will not be expected to interact with mAcpD.



Scheme 9: The unusual acyl transferase activity of CerJ-type KS enzymes. A) The scheme for AT activity of CerJ enzyme. B) The scheme showing the AT-activity of DpsC enzyme.

The first aim of this work is to determine the enzyme which will further process 3-hydroxypropanoate attached to terminal thiol of mAcpD **10$_D$** (3HP-mAcpD). Bioinformatic analysis points towards MmpF_KS or MmpB_KS as the enzymes likely to be responsible for this elongation step. Whereas MmpB contains all the catalytic domains responsible for DCC and processing of polyketide intermediate, MmpF lacks the tailoring KR, DH and ER domains. Some or all of this activity may be provided by *trans*-acting enzymes MupD and MupE. Analysis of MupD and MupE will be conducted later in this work.

3.2.2 Phylogenetic analysis of KS domains

Study of the evolutionary history (phylogeny) of polyketide synthases has illuminated distinct differences between *trans*-AT PKS and *cis*-AT PKSs. This is due to the domain architecture of each PKS having evolved in separate ways. [20] *cis*-AT systems are expected to have evolved from gene duplication of individual modules, followed by diversification of these domains over many years.[41] In contrast, *trans*-AT systems have evolved from extensive horizontal gene transfer between bacteria.[114] This means that KS domains from *trans*-AT systems will cluster together based on substrate specificity as opposed to *cis*-AT KSs which cluster tightly into natural product specific clades. [115,116] This principle can be used to extract information about KS domains from *trans*-AT systems. For the *trans*-AT KS domains (such as MmpF_KS and MmpB_KS), phylogenetic analysis may reveal important information about substrate specificity which may in turn shed light on their roles within the pathway.

In previous work by Nguyen *et al*. a comprehensive phylogenetic analysis of KS domains from many *trans*-AT systems was conducted including selected KS domains from mupirocin and thiomarinol. [42] MmpB_KS was included in this analysis and found to be a member of the IX clade, showing selectivity for enone substrates. MmpF_KS, TmpF_KS (thiomarinol homologue of MmpF_KS) and the second KS domain of TmpB_KS2 were not included in this initial analysis by Nguyen et al. Thiomarinol is a BGC with many shared features with mupirocin and contains many proteins with strong homology to mupirocin counterparts. [56]

A new phylogenetic tree was therefore constructed containing 61 KS domains including domains initially not-included in the analysis by Nguyen *et al*; MmpF_KS, TmpF_KS and TmpB_KS2 plus MmpF_KS homologues NosB_KS and OzmQ_KS. [107,117]

The phylogenetic tree was constructed using the Maximum-likelihood method using Mega X software. [118] The various *trans*-AT KS proteins were first aligned using ClustalW.[119] The Jones-Taylor-Thornton (JTT) model was then used to construct the phylogenetic tree. Initial phylogenetic analysis by Nguyen was conducted using the Bayesian model [120] but other methods (such as JTT used in this analysis) were also found to produce comparable results.

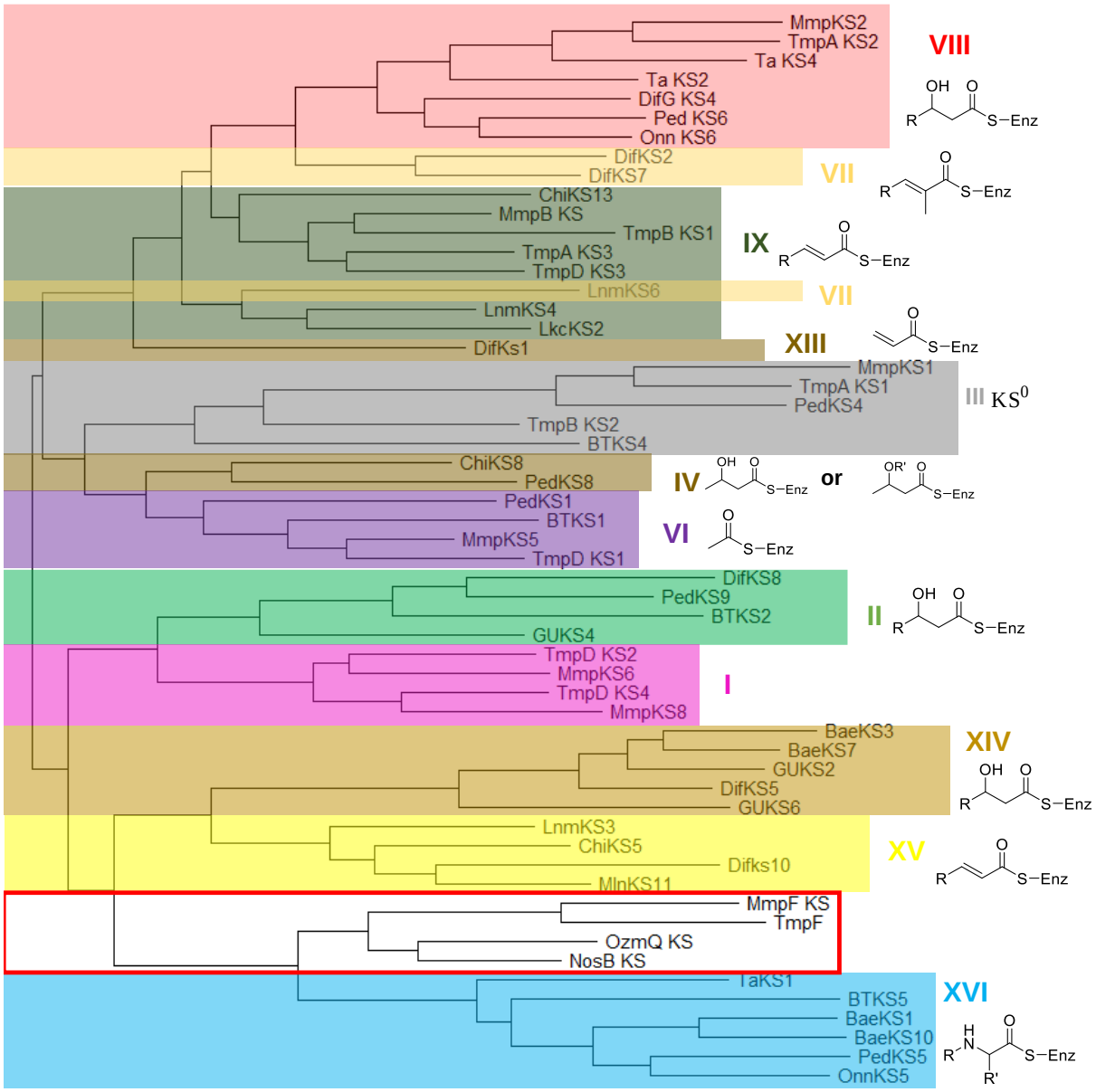


Figure 33: Phylogenetic tree created using Jones-Taylor-Thornton (JTT) model. This phylogenetic tree shows the evolutionary relatedness of many selected KS domains, including those from biosynthesis of mupirocin and thiomarinol. Clade containing MmpF_KS is highlighted with red border.

The phylogenetic tree produced in this work was successful in reproducing the clades initially reported by Nguyen.[42] Only one KS domain was found to clade incorrectly, the KS domain from biosynthetic cluster of Leinamycin, Lnm_KS6.[121] This KS domain was found to be a member of clade VII by Nguyen *et al* but was found to be clade IX in this analysis. This is expected to be due to the propensity of Lnm_KS6 to clade with Lnm_KS4, another KS from Leinamycin. Apart from Lnm_KS4, all KS domains from the initial analysis have been found to group into the same clades as predicted by Nguyen et al.

3.2.2.1 MmpF_KS and TmpF_KS clade with hybrid NRPS/PKS domains

The clades from Nguyen analysis were labelled with roman numerals based on the substrates they would typically bind. It was expected that the KS domain responsible for further condensation of β-hydroxy 3HP would align with KS domains responsible for turnover with β-hydroxy substrates. These KS domains were observed in clades II, IV and VIII. Unexpectedly, this was not the case for MmpF or MmpB_KS. MmpF_KS and TmpF_KS were both found to clade with NosB_KS and OzmQ_KS in a group of KS domains closest in terms to phylogeny to Clade XVI, thought to accept *N*-acyl amino acid substrates. The *N*-acyl amino acid KS domains from Clade XVI have been shown to be from *cis*-AT type I PKSs rather than *trans*-AT type I PKSs.[109] A similar phylogenetic analysis of OzmQ_KS found that the separate clade containing MmpF_KS and TmpF_KS occurs due to recent incorporation of these proteins into their respective BGCs rather than substrate specificity. Therefore no conclusions about the substrate specificity of MmpF_KS can be drawn from this analysis.

As was observed in the initial analysis conducted by Nguyen *et al*, MmpB_KS was found to be a member of IX clade expected to accept enone substrates. The homologue of MmpB_KS from the structurally related thiomarinol gene cluster, TmpB_KS1 was also a member of clade IX. The second KS domain from TmpB (TmpB_KS2) was confirmed to be a KS^0 enzyme (clade III).

One of the amino acid residues that has shown to be important to determining substrate binding of KS domains is the residue immediately preceding the active site cysteine (so called X-Cys residue). This residue is proposed to play an important role in the binding of acyl moiety to the KS. If the X-Cys residue is sterically bulky then β-substituted acyl moieties are prevented from entering the binding pocket preventing enzyme turnover. [122] Within the first condensation reaction during 9HN biosynthesis, the proposed substrate, 3HP, contains a β-hydroxy group. Whereas MmpF contains an X-Cys residue of an alanine, MmpB_KS contains a bulky methionine residue in the equivalent X-Cys position. Due to the bulky β-hydroxy group contained within mAcpD-3HP, this provides evidence that MmpF might be geared to catalyse the DCC with the unusual 3HP starter unit.

Confirmation of which KS domain is likely to catalyse the first round of DCC could involve a daunting number of combinations involving MmpF, MmpB_KS and MupB. To reduce this uncertainty, a series

of ^{1}H-^{15}N HSQC NMR titration experiments were first conducted with mAcpD and each of these candidate KS domains. It was hypothesised that the identification of protein-protein interactions *in vitro* between the ACP and the correct partner might provide an indication of which KS domain binds 3HP-mAcpD during biosynthesis of 9HN *in vivo*. The first step towards this goal was to express and purify MmpF and the excised MmpB_KS domain from MmpB. A sample of MupB was kindly provided by Nahida Akter (University of Bristol).

Expression and Purification of MmpF and MmpB_KS

The gene corresponding to MmpF was amplified into pET28a vector [123] from genomic DNA of *P. flourescens* using gene-specific primers by PCR (Primers in Methods). Expression under standard conditions using 100 mL LB media per 500 mL non-baffled flask was successful in producing soluble protein as confirmed by SDS_PAGE. Protein was subsequently purified by IMAC and further purified by SEC (S200 Column) using high salt (0.5 M) and 10% glycerol buffer with a yield of 2 mg/L. ESI-MS and SDS-PAGE gel confirm MmpF was purified to homogeneity (Figure 34A) and ESI-MS confirmed the observed mass was in excellent agreement with the expected mass (observed; 84839 Da, expected; 84842 Da, Figure 34B). A second species was also observed in 30% yield with a mass of 85017 Da. This species corresponds to a mass increase of 178 Da consistent with spontaneous α-N-6-phosphogluconoylation of the polyhistidine tag. Due to poor yield and low solubility of protein, however, the polyhistidine tag was not removed by incubation with protease. Analytical size exclusion was conducted with an S200 column and based on calibration curve (see Methods) showed MmpF to be a monomer in solution (Figure 34C).

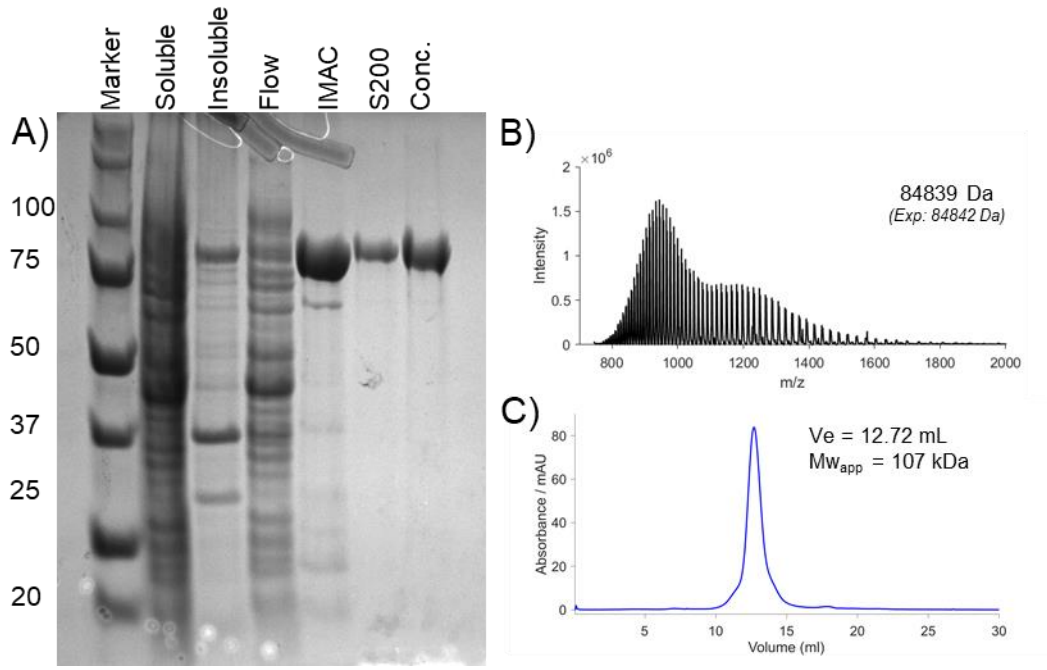


Figure 34: Characterisation of MmpF A) SDS-PAGE gel showing purification of MmpF. Conc. sample shows MmpF after purification by IMAC followed by size exclusion chromatography and subsequent spin concentration. B) ESI-MS of purified MmpF. C) Analytical SEC trace of purified MmpF. Calculated molecular weight shows that MmpF is a monomer in solution (Mw_{app}= 107 kDa, Mw = 84.839 kDa).

Attention was turned to the expression and purification of the KS module of MmpB. This KS is situated at the N-terminal of the 221 kDa MmpB. Within the predicted bounds of this KS, sequence homology and three-dimensional structure prediction suggested an N-terminal region comprised of 24 amino acid residues would be flexible and not form part of the KS domain three-dimensional structure so was removed from the MmpB_KS sequence (Figure 35). In addition, a cysteine residue located near the C-terminus of the sequence was mutated to an alanine to potentially avoid covalent dimerization. This gene construct was designed by Dr Winter and was synthesised by Thermofisher GeneArt and inserted into a pET151_D vector.

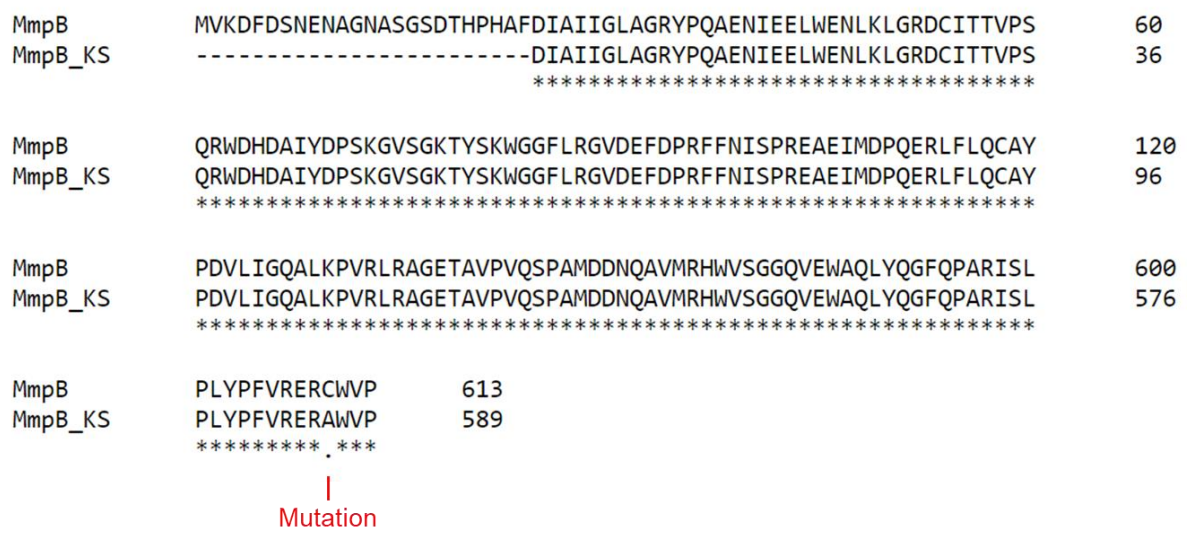


Figure 35: Alignment of selected regions of wild-type MmpB amino acid sequence and MmpB_KS amino acid sequence. The alignment shown corresponds to 0-96 and 516-589 amino acid of the MmpB_KS plasmid. A 24 amino acid disordered region was removed from the N-terminus of wild type MmpB to produce MmpB_KS plasmid. In addition, C-terminal cysteine residue was mutated to an alanine.

The plasmid containing MmpB_KS gene was expressed in *E. coli* followed by purification *via* IMAC and SEC using an S200 column. The yield of this protein post-purification was 3 mg/mL. Purified protein was characterised by SDS-PAGE to confirm purity (Figure 36A). The protein was located between 50 kDa and 70 kDa marker bands as expected due to the size of MmpB_KS. ESI-MS of the pure protein produced a species with good agreement with the expected mass of MmpB_KS (observed, 67011 Da; expected, 67020 Da, Figure 36B). Analytical size exclusion chromatography using an S200 column confirmed that MmpB_KS was a dimer in solution (Mw_{app}= 105 kDa, Figure 36C).

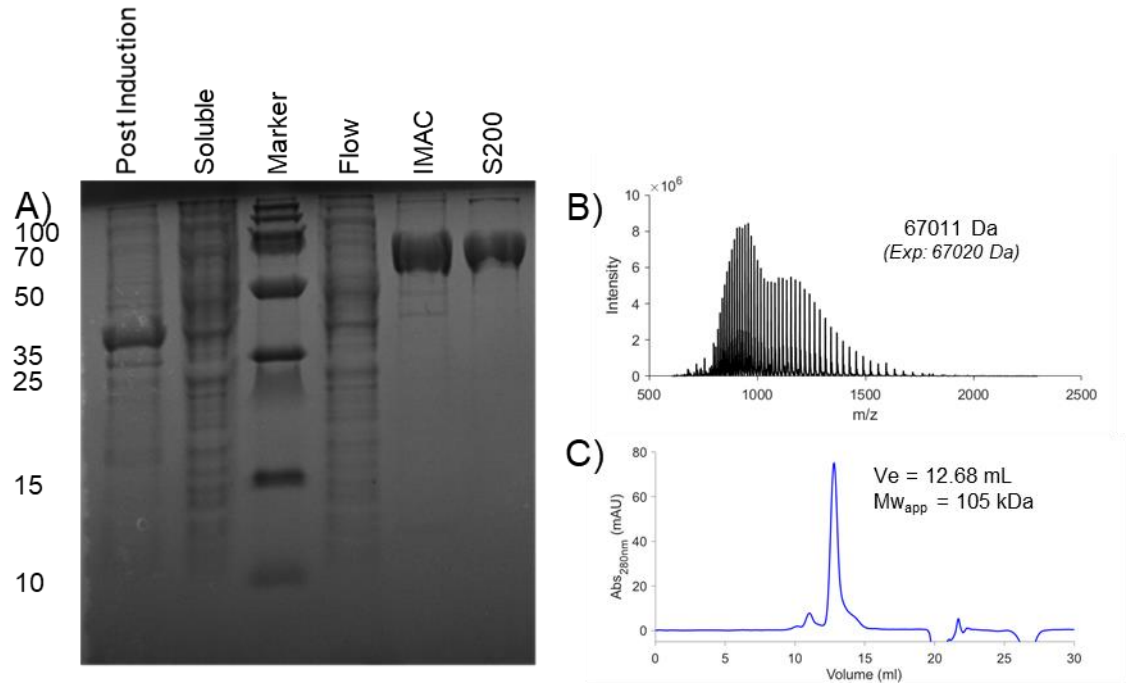


Figure 36: Characterisation of MmpB_KS. A) SDS-PAGE gel showing purification steps for MmpB_KS. The S200 fraction shows Mmp after purification by IMAC followed by size exclusion chromatography. B) ESI-MS of purified MmpB_KS. Observed mass of the predominant species was 9 Da away from the expected mass of MmpB_KS. C) Analytical SEC trace of purified MmpB_KS using analytical S200 column. Calculated molecular weight shows that MmpB_KS is a dimer in solution (Mw_{app}= 105 kDa, Mw = 67020 Da).

3.2.3 NMR titrations of ^{15}N-mAcpD with KS enzymes

NMR titration was conducted with ^{15}N-*apo*-mAcpD and MupB. A series of titrations were conducted with increasing ratios of MupB: mAcpD. For this work, ratios of 1:1, 2:1 and 3:1 of MupB: mAcpD were used. Analytical size exclusion chromatography found that MupB was a dimer in solution and therefore the ratios used for the titration will refer to the dimeric form of MupB. A 1:3 ratio of mAcpD to dimeric MupB consists of 80 µM mAcpD and 240 µM of dimeric MupB (or 480 µM of monomeric MupB). NMR samples were run on the 600 MHz NMR spectrometer. Increasing ratios of MupB were added to ^{15}N-mAcpD and a ^{1}H-^{15}N HSQC experiment was used to monitor chemical shift perturbations.

Ratio (mAcpD to MupB)	15N-mAcpD Conc. / µM	MupB Conc. / µM
1:0	80	0
1:1	80	80
1:2	80	160
1:3	80	240

Table 3: A table showing the concentrations of each component of the NMR titration. MupB concentration is that of the dimeric form. All titrations were made up to 200 µL in NMR Buffer and contained 10% D_2O

Figure 37 compares the ^{1}H-^{15}N HSQC spectrum of only ^{15}N-mAcpD (Figure 37A) with the ^{1}H-^{15}N HSQC spectrum of ^{15}N-mAcpD after addition of three equivalents of MupB (Figure 37B). The assignments shown in Figure 37A were calculated based on previous ^{1}H-^{15}N HSQC spectra of ^{15}N-mAcpD of which the spectrum shown in Figure 37A agrees well. This spectrum is the negative control experiment, the baseline with which to measure the interaction with MupB. This NMR assignment, as with all NMR assignments shown in this work, was completed by Dr Winter.

As the ratio of MupB was increased, the amide region of ^{15}N-mAcpD spectra did not show any significant chemical shift perturbations. The spectrum shown in Figure 37B is very similar to the negative control spectrum. Although some line broadening is clearly discernible, the overall spectral perturbation is minor compared to similar experiments involving mAcpD and known interaction partners MupQ and MupS.[86] It was concluded that this was indicative of a non-binding interaction.

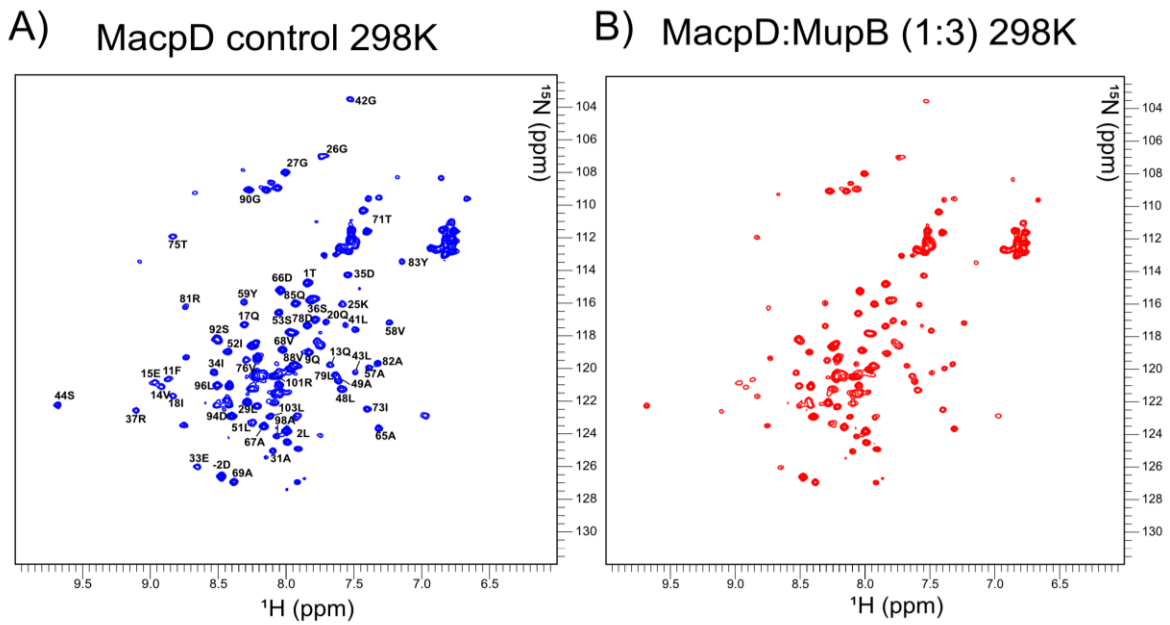


Figure 37: A) HSQC spectra of ^{15}N-mAcpD (amide region) recorded at 298K on 600 MHz NMR spectrometer B) HSQC spectra of ^{15}N-mAcpD (amide region) after the addition of 3 equivalents of MupB (dimer). No tractable NMR shift changes were observed showing that mAcpD and MupB do not interact in vivo.

The NMR titration provided preliminary evidence that MupB is not the KS domain which interacts mAcpD-3HP *in vivo*. Of the two remaining candidates, MmpF and MmpB_KS, MmpB_KS was tested next for interaction with *apo*-^{15}N-mAcpD. NMR titrations of *apo*-^{15}N-mAcpD and MmpB_KS were acquired on the 700 MHz NMR spectrometer at 15 °C. A series of titrations were conducted with increasing ratios of MmpB_KS: mAcpD; 1:2, 1:1 and 2:1 respectively. The ratios and concentrations of each NMR titration is shown in Table 4. NMR sample preparation and analysis were completed in conjunction with Dr Winter.

Ratio (MacpD to MmpB_KS)	^{15}N-MacpD Concentration / µM	MmpB_KS Concentration / µM
1:0	100	0
2:1	100	50
1:1	100	100
1:2	50	100

Table 4: Ratios and concentrations utilised during MmpB_KS NMR titrations

Again, a control ^{1}H-^{15}N HSQC spectrum of *apo*-^{15}N-mAcpD was acquired at 15 °C prior to addition of MmpB_KS. This control spectrum compared well to previous spectrum recorded at 25 °C is shown in Figure 38A. This control was compared to the spectrum recorded after addition of two equivalents of MmpB_KS (Figure 38B). Again, no significant chemical shift perturbations were observed in the amide region of the ^{15}N-mAcpD spectrum after addition of MmpB_KS. This is indicative of a non-binding interaction between MmpB_KS and *apo*-^{15}N-mAcpD

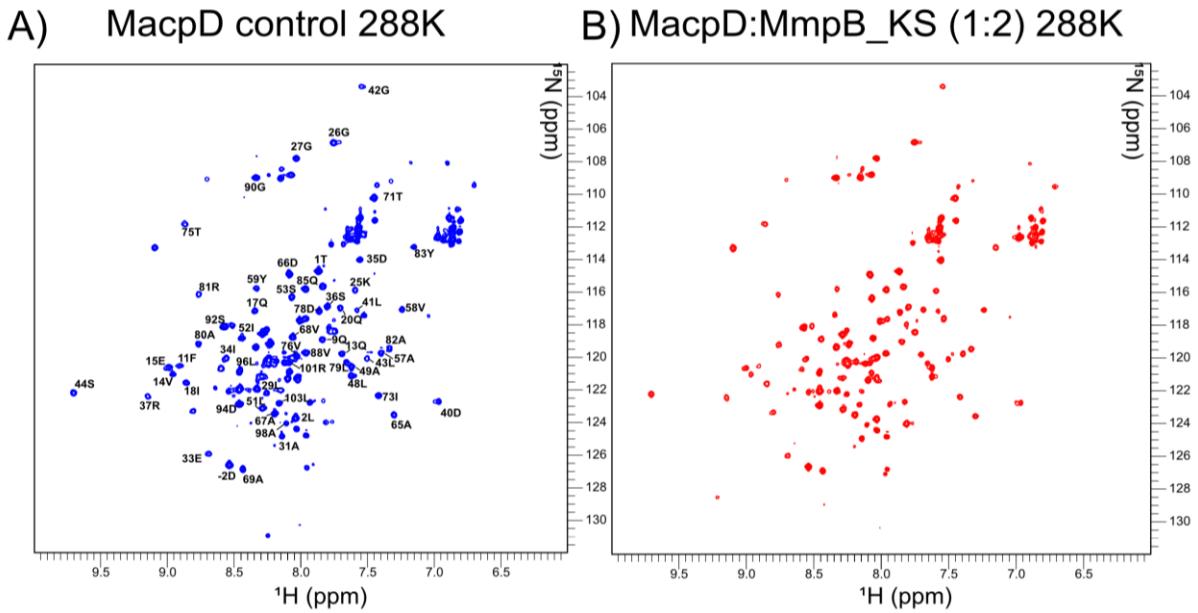


Figure 38: A figure showing the HSQC spectrum of ^{15}N-mAcpD before and after addition of MmpB_KS. A) HSQC spectra of amide region of ^{15}N-mAcpD prior to addition of MmpB_KS. B) HSQC spectra of ^{15}N-mAcpD (amide region) after the addition of 2 equivalents of MmpB_KS. Sample run on NMR by Dr Williams and subsequent NMR analysis conducted by Dr Winter.

This again provided strong evidence that MmpB_KS does not interact with mAcpD and that MmpB_KS does not process 3HP-mAcpD. One caveat is that *apo*-ACP was utilised which did not bear either the phosphopantetheine sidearm are or the 3HP substrate and these may be necessary to ensure an

interaction. However, our previous observation is that *apo*-ACP is sufficient to observe these interactions due to significant interaction between apo-mAcpD and MupQ/MupS.

The final KS to be tested for its ability to interact with mAcpD was MmpF. NMR titrations were conducted on the 600 MHz NMR spectrometer at 15 °C. The ratios and concentrations of the proteins used in the NMR titrations are given in Table 5.

Ratio (MacpD to MmpF)	^{15}N-MacpD Concentration / µM	MmpF Concentration / µM
1:0	180	0
1:1	80	80
1:2	80	160
1:3	60	180
1:4	40	160

Table 5: A table showing the ratios of mAcpD to MmpF used in the NMR titrations.

The control spectrum of *apo*-^{15}N-mAcpD was consistent with previous ^{1}H-^{15}N HSQC spectra and is shown in Figure 39A. When this control was compared to the spectrum recorded after addition of four equivalents of MmpF, significant chemical shift perturbations are observed (Figure 39B). After addition of 4 equivalents of MmpF (Figure 39B) a mixture of peak broadening and chemical shift perturbations were observed. The main changes observed were weakening of mAcpD chemical shifts on addition of MmpF. These were not observed for equivalent titrations with MmpB_KS and MupB. These changes to the spectrum are strongly suggestive of protein-protein interactions between MmpF and mAcpD. The specific amino acids of mAcpD which undergo were observed for the amino acid residues situated within the 4-helix bundle of mAcpD and did not include residues in the unstructured C-terminal extension. Specific residues from the C-terminal extension; Ser92, Asp94, Leu96, Ala98, Arg101 and Leu103 are shown in Figure 39B. These residues, like the remaining residues on the C-terminal helix did not undergo weakening on addition of MmpF which implies they interact less. An overlay of the control spectrum and the 1:4 ratio spectrum is shown in the Supporting Information (Figure S 1).

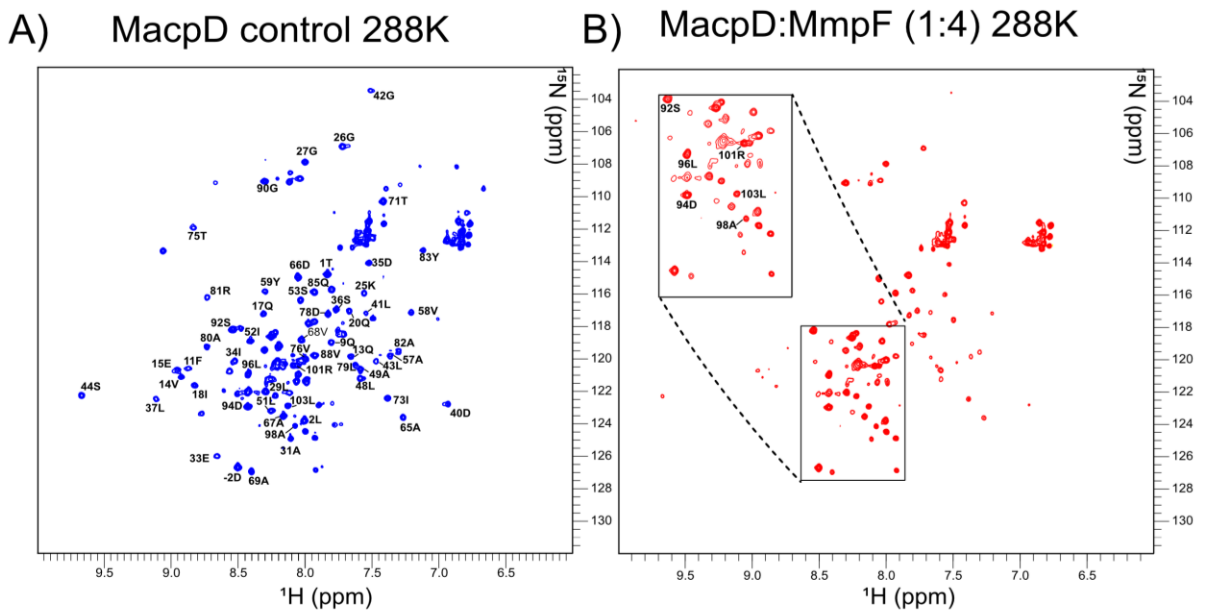
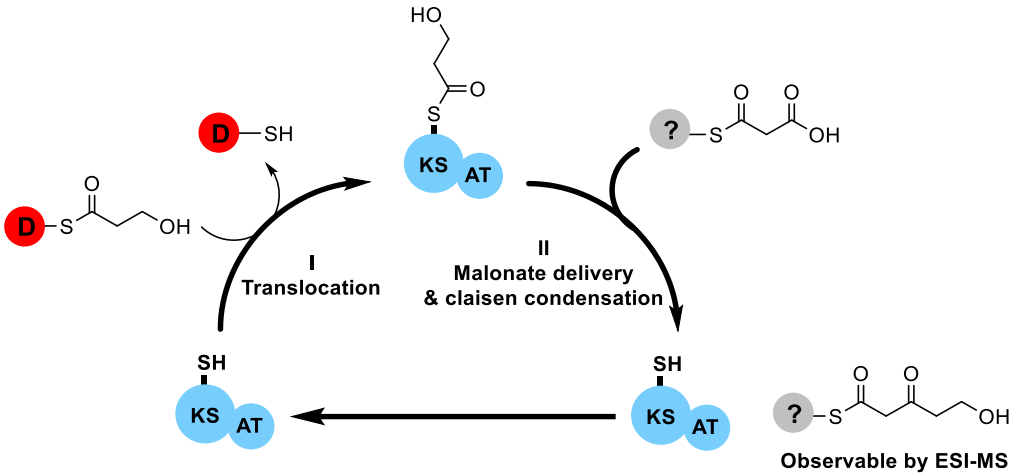


Figure 39: A) The amide region of HSQC spectra of ^{15}N-mAcpD recorded at 288K. B) HSQC spectra of 15N-mAcpD after 4 equivalents of MmpF added. A series of peak disappearances and peak weakening is indicative of interaction between mAcpD and MmpF. Residues that did not show perturbations were located on C-terminal extension of mAcpD (zoomed).

3.2.4 *In-vitro* reconstitution of 3-carbon to 5-carbon condensation

The data collected during the NMR titrations provides evidence that MmpF interacts with the nascent 3HP-mAcpD *in vivo.* This would mean that the next step in the biosynthesis of 9HN involves MmpF catalysed elongation of 3HP-mAcpD. This is consistent with the role of MmpF homologues.[124,125] The active site cysteine of MmpF will bind an acyl moiety when delivered by the correct ACP during the translocation step. Due to the KASI/II like nature of MmpF, ACP-bound substrates are expected to translocate most efficiently and 3HP should be transferred to MmpF by mAcpD (Scheme 10). After successful translocation, MmpF will catalyse the 2-carbon extension utilising malonyl delivered by another ACP to produce a β-ketothiol ester product. Within the biosynthetic cluster of mupirocin, two stand-alone ACPs remain unassigned and are candidates to deliver a malonyl group to MmpF; mAcpA and mAcpB. After the decarboxylative Claisen condensation (DCC) has taken place, the 2-carbon extended diketone will be covalently attached to the terminal thiol of the phosphopantetheine arm of the ACP.



Scheme 10: A scheme showing the proposed reaction of MmpF and 3HP-mAcpD utilising malonyl-ACP. MmpF is proposed to bind 3-carbon alcohol during the translocation step. Addition of malonyl, delivered by the correct ACP (either mAcpA or mAcpB – identity unknown) will facilitate KS-catalysed elongation of the acyl chain producing a diketone intermediate that is observable by protein mass spectrometry.

3.2.5 *In-vitro* translocation of MmpF with pantetheine substrates

Before attempting to fully constitute a multi-enzyme cascade with authentic substrates, a series of model studies were performed. Initially, the ability of MmpF to self-load simple acetyl and propionyl pantetheine substrates was tested. In the *in vivo* biosynthetic pathway, 3HP will be delivered *via* the terminal thiol of mAcpD. To mimic the phosphopantetheine arm of mAcpD, a series of synthetic acyl groups covalently attached to a pantetheine moiety was used. Acetyl-pantetheine **14** and propionyl-pantetheine **15** were chosen due to their similarity to 3HP (Figure 40) and synthesised by Angus Weir. Attempts to synthesise the pantetheine version of the natural substrate, 3HP-Pant was unsuccessful due to propensity to degrade during a key deprotection step as is outlined elsewhere (Angus Weir, Thesis, University of Bristol)

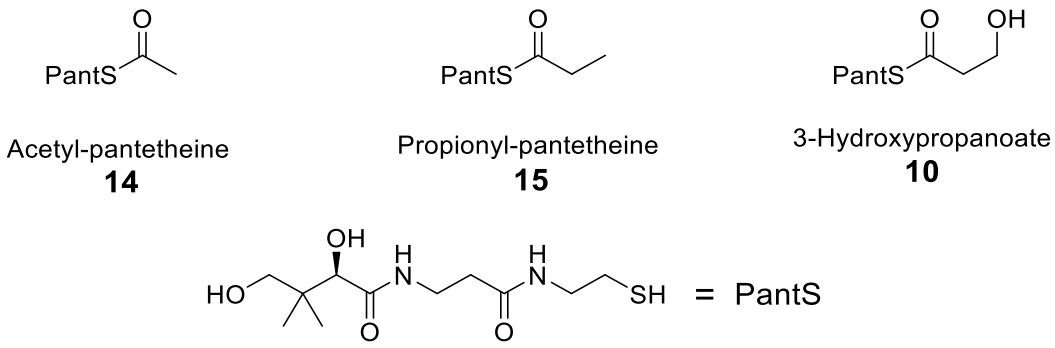


*Figure 40: The pantetheine substrates which will be used to test in vitro translocation of MmpF. Acetyl **14** and propionyl **15** pantetheine were chosen due to their similarity to natural substrate 3-hydroxypropanoate. The pantetheine arm of these substrates has been abbreviated to PantS for simplicity.*

ESI-MS analysis of MmpF prior to addition of pantetheine analogues showed two main peaks; 84852 Da and 85031 Da corresponding to non-phosphogluconoylated and phosphogluconoylated forms of MmpF (expected mass of non-phosphogluconoylated MmpF: 84842 Da, Figure 41A). To initiate the translocation reaction with the first substrate, MmpF (20 µM) was added to an excess of **14** (200 µM). Assays were monitored by ESI-MS after 3 hours at room temperature. After this time, consumption of MmpF and formation of new species at 84895 Da was observed (Figure 41B). This new species constitutes a mass increase of 43 Da from control spectrum of MmpF and is consistent with translocation of acetyl moiety to active site cysteine of MmpF (expected mass increase of acetyl-MmpF; + 42 Da). 100% conversion to acetyl form was observed after 3 hours. This confirmed the ability of MmpF to covalently bind **14**.

The ability of MmpF to translocate propionyl-pantetheine **15** was also tested. Reaction was initiated by addition of an excess of **15** (200 µM) to MmpF (20 µM) and incubated at room temperature. After 3 hours, consumption of MmpF and formation of new species with mass of 84909.5 Da was observed (Figure 41C). This new species constitutes a mass increase of 57.5 Da from the control spectrum of MmpF and is consistent with propionyl-MmpF (prop-MmpF) being formed (expected mass increase of propionyl-MmpF; + 56 Da). In both experiments the phosphogluconoylated forms of MmpF showed identical mass shifts.

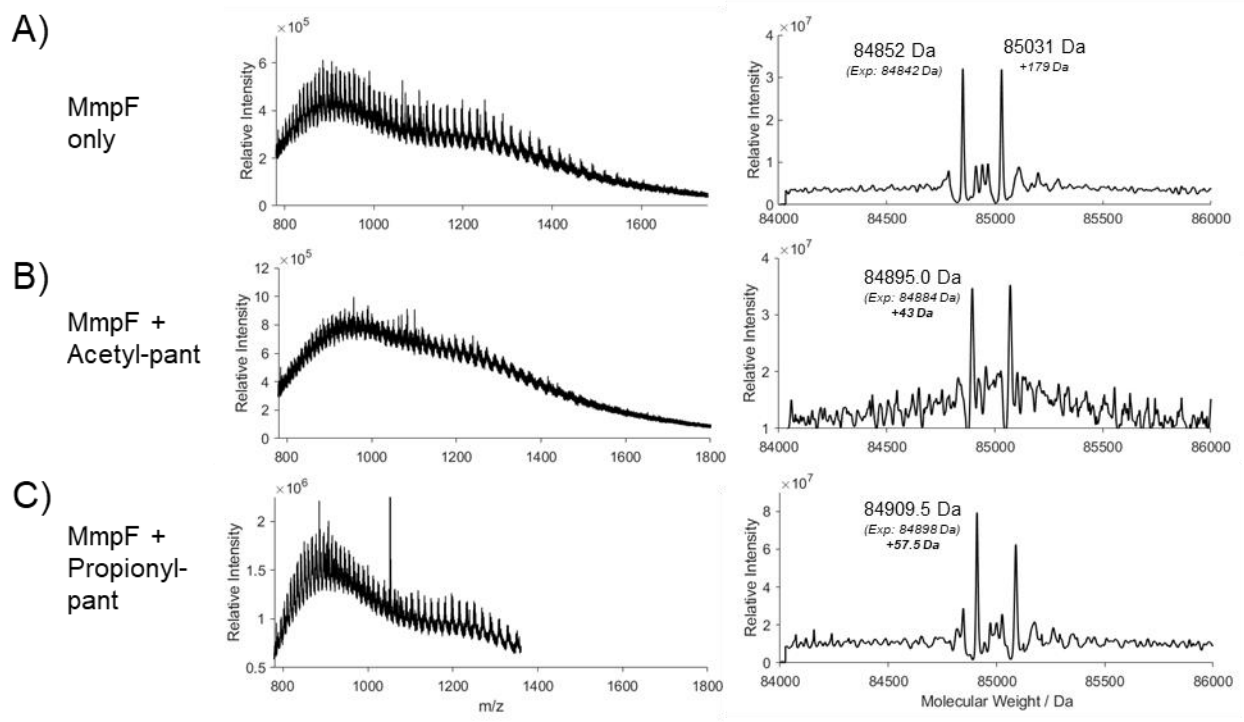


Figure 41: The translocation of acetyl-pantetheine and propionyl-pantetheine to MmpF A) ESI-MS spectrum and subsequent deconvolution of only MmpF. B) Charge states and subsequent deconvolution 3 hours after addition

*of excess **14** to MmpF. C) Charge states and subsequent deconvolution 3 hours after addition of excess **15** to MmpF.*

The successful translocation of MmpF with **14** and **15** provides further evidence for its function as the KS enzyme responsible for the first round of fatty acid chain extension with 3HP. Preparations were made to conduct the condensation reaction *in vitro* with the required malonyl-ACPs.

3.2.6 Expression and purification of mAcpA and mAcpB

The two ACPs potentially able to deliver malonyl to MmpF are mAcpA and mAcpB. The gene corresponding to mAcpA was synthesised by ThermoFisher GeneArt and inserted into a pET151_D plasmid (Supporting Information). The cloning of mAcpB from the genomic DNA of *P. fluorescens* by PCR was conducted using primers given in Supplementary Information. The successful PCR conditions are given in the Methods section. The initial expression and purification protocol for both proteins were conducted by Dr Winter. First, mAcpA was expressed in LB media under standard conditions (Methods). mAcpA protein was purified in the *apo*-form only by IMAC and his-tag subsequently removed from protein by overnight incubation with TEV protease. The free polyhistidine tag was removed by negative IMAC. Subsequent purification by SEC using a S100 column yielded 10 mg/mL of soluble protein as confirmed by SDS-PAGE analysis (Figure 42A). ESI-MS confirmed the mass of this protein to be in good agreement with the expected mass (Observed; 11887 Da, Expected; 11889 Da, Figure 42B). Analytical SEC using an S75 column confirmed mAcpA to be a monomer in solution (Ve = 13.8 mL, Mw_{app} = 10.8 kDa, Figure 42C).

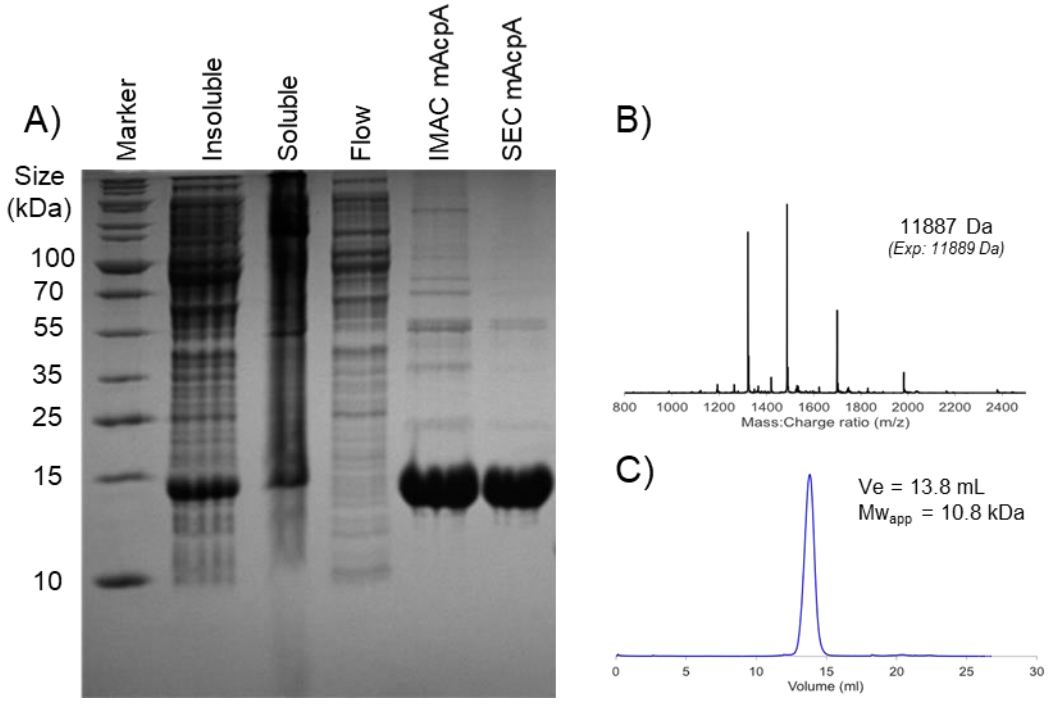


Figure 42: Expression and purification of mAcpA A) SDS-PAGE gel showing the stages of expression and purification of mAcpA. B) Characterisation of cleaved mAcpA by ESI-MS. A single series of charge states

correspond to the expected mass of cleaved apo-mAcpA (observed; 11887 Da, expected; 11889 Da). C) Analytical size exclusion trace of cleaved, apo-mAcpA shows monomeric state in solution.

The gene corresponding to mAcpB was synthesised by ThermoFisher GeneArt into pOPINF plasmid (Supporting Information). mAcpB protein was found to be insoluble under standard *E. coli* growth conditions (as outlined in Methods). Soluble mAcpB was produced by utilising ArcticExpress (D3) Competent Cells (Arctic cells).[126] These cells provide an *in vivo* approach to increasing yield of previously insoluble protein utilising low temperature cell cultivation. These cells contain the low temperature chaperonins Cpn60 and Cpn10 to facilitate proper folding of protein during low temperature expression.[127] Low temperature cultivation leads to production of soluble protein but at reduced yields. [128]

Expression of mAcpB yielded < 1 mg/L of soluble protein. Purification was conducted by IMAC followed by SEC using an S100 column to yield 1 mg/L of soluble protein. The purity of mAcpB was confirmed by SDS-PAGE gel (Figure 43A). The mass of mAcpB observed by ESI-MS was 11371 Da, this is 2 Da away from the expected mass of *apo*-mAcpB (expected mass; 11373 Da, Figure 43B). Analytical size exclusion of *apo*-mAcpB using analytical S75 column showed it to be a monomer in solution (Ve = 13.3 mL, Mw_{app}= 12.9 kDa)

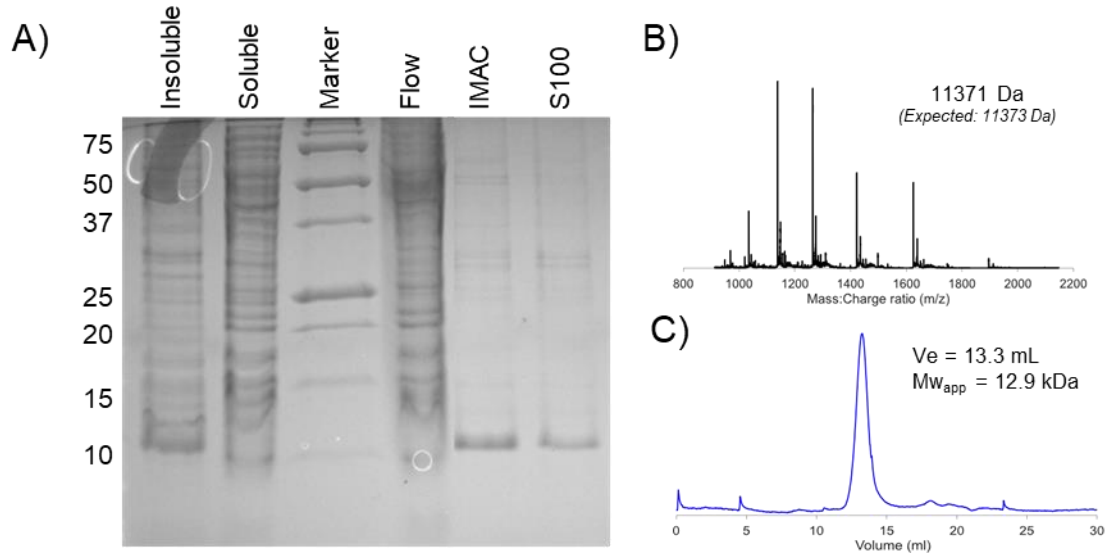


Figure 43: Expression and purification of mAcpB A) SDS-PAGE gel showing the stages of expression and purification of mAcpB. B) Characterisation of apo-mAcpB by ESI-MS. A single series of charge states correspond to the expected mass of apo-mAcpB (observed; 11371 Da, expected; 11373 Da). C) Analytical size exclusion trace of apo-mAcpB shows monomeric state in solution. Expression and purification of mAcpB was conducted by Dr Winter and is reproduced here.

3.2.7 *In-vitro* enzymatic generation of malonyl-ACPs

Purified, soluble *apo*-mAcpB (100 µM) was converted to malonyl-mAcpB in a one-pot chemoenzymatic reaction utilising MupN (5 µM), ATP (1 mM), $MgCl_2$ (10 mM) and malonyl CoA (1

mM). Assays were conducted at room temperature and monitored by ESI-MS after 2 hours. After this time, the charge states corresponding to *apo*-mAcpB were consumed and a new species was observed with a mass consistent with malonyl-mAcpB (observed; 11797.1 Da, expected; 11798.7 Da, Figure 44A).

The *apo*-form of mAcpA was malonylated under identical reaction conditions utilising MupN, malonyl CoA, CoA mix enzymes and their relevant cofactors (conditions same as for malonyl-mAcpB formation). Assays were again conducted at room temperature and monitored by ESI-MS over time. The formation of malonyl-mAcpA took longer when compared to malonyl-mAcpB formation. This is consistent with the conversion of *apo*-mAcpA to *holo*-mAcpA which took substantially longer than equivalent mAcpB reaction. After 8 hours at room temperature, followed by 16 hours at 4°C, the majority species (80% conversion) was found to be malonyl-mAcpA (observed; 12313.0 Da, expected, 12315.0 Da). A lesser amount (20%) of *holo*-mAcpA was also observed. Addition of up to 2 equivalents of PPTase enzyme and cofactors was unsuccessful in forming 100% malonyl-mAcpA.

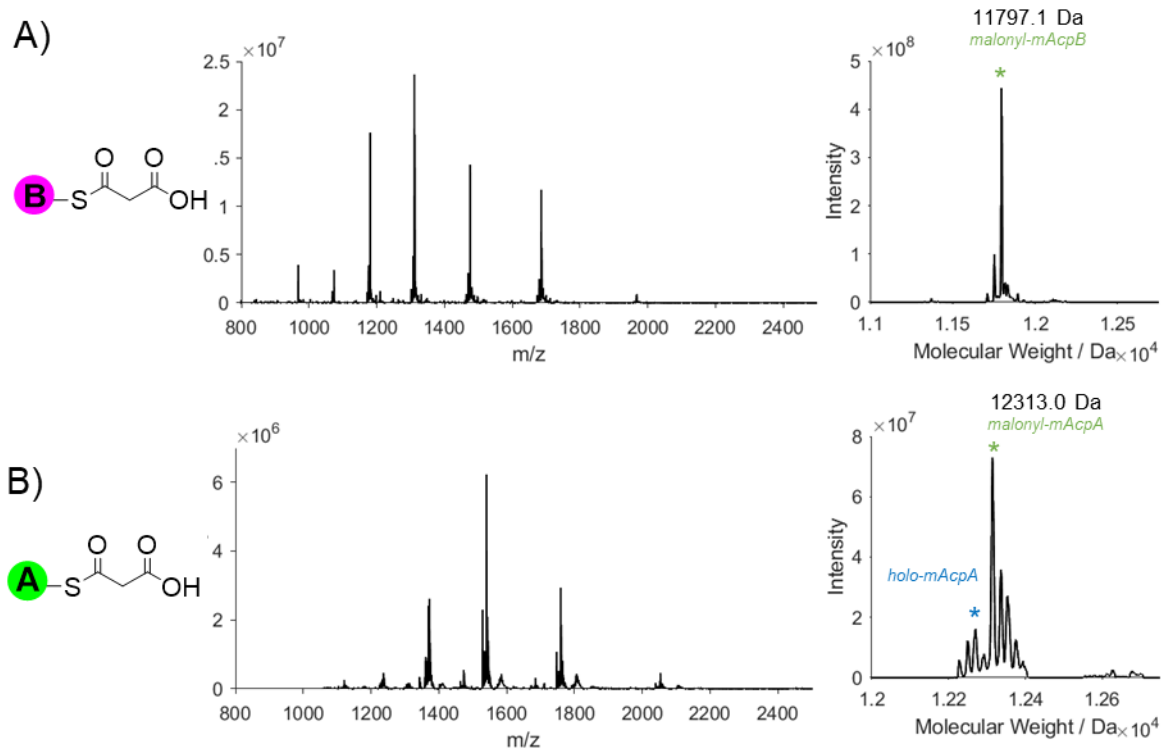


Figure 44: The formation of malonyl-ACPs for use in MmpF chain extension reaction. A) The charge states (left) and subsequent deconvolution (right) of mAcpB-malonyl formation assay after 2 hours at room temperature. B) The charge states (left) and subsequent deconvolution (right) of mAcpB-malonyl formation assay after 8 hours at room temperature, followed by 16 hours at 4°C. A mixture of malonyl-mAcpA (80%) and holo-mAcpA (20%) was observed after this time.

3.2.8 *In-vitro* MmpF condensation assay with propionyl starter unit

After successful formation of malonyl-mAcpB, incubation with propionyl-MmpF was conducted. Propionyl-MmpF was formed by incubating MmpF (20 µM) with excess propionyl-pantetheine **15** (200 µM) at room temperature until complete transthiolation was observed by ESI-MS. Complete transthiolation of propionyl moiety to MmpF was observed after 2 hours.

The reaction scheme showing the proposed condensation reaction is shown in Figure 45A. The reaction was initiated by addition of malonyl-mAcpB (100 µM) to propionyl-MmpF (20 µM) and incubated at room temperature. The reaction was monitored by ESI-MS at 30 minute and 2 hour time points and the analysis was focussed on mAcpB region of the spectrum. Successful condensation is expected to form a 5-carbon β-keto thiolester intermediate $\mathbf{16_B}$ bound to mAcpB.

After 2 hours, significant changes were observed with clear consumption of malonyl-mAcpB. A new species was observed with a mass consistent with $\mathbf{16_B}$ (observed; 11811.4 Da, expected; 11810.1 Da, Figure 45C). Further fragmentation of this species delivered a characteristic pantetheine ion of 359.12 Da, consistent with $\mathbf{16_B}$ (expected mass: 359.16 Da, Figure 45D). The formation of $\mathbf{16_B}$ clearly demonstrated successful MmpF-catalysed condensation between the propionyl moiety and the mAcpB-delivered malonyl.

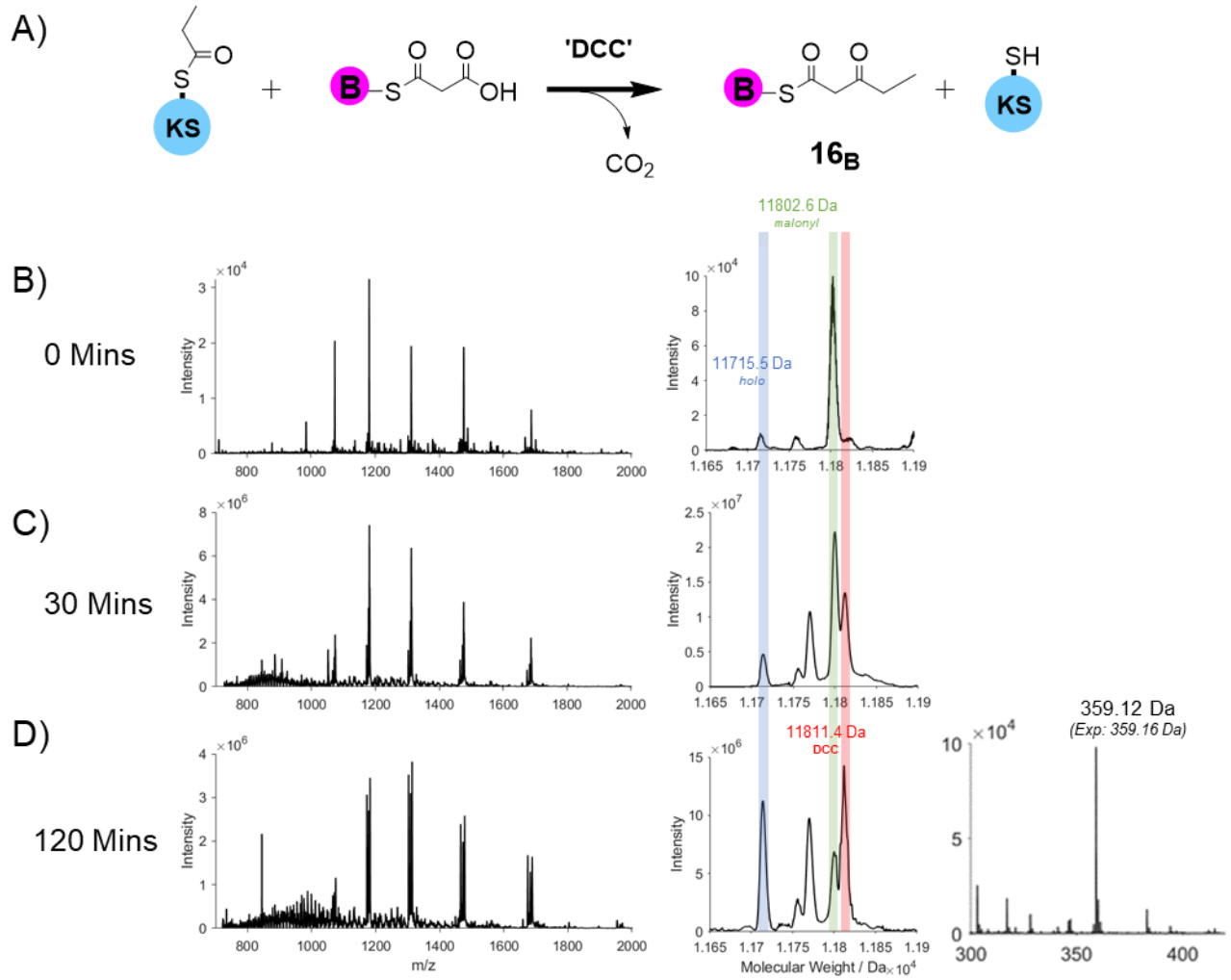
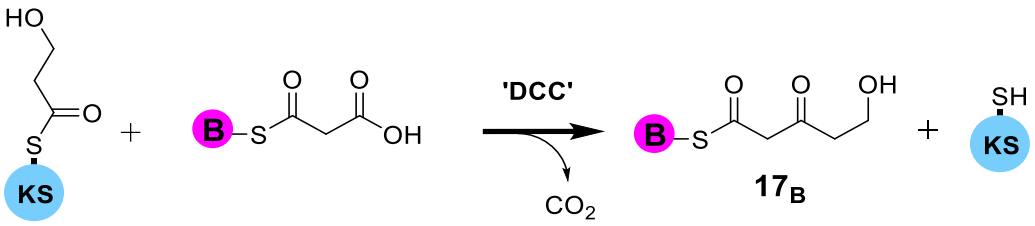


*Figure 45: The addition of malonyl-mAcpB to propionyl-MmpF leads to chain extending DCC reaction. A) Reaction scheme showing the condensation reaction of MmpF with propionyl. B) Charge states (left) and subsequent deconvolution (right) of malonyl-mAcpB prior to addition of MmpF-propionyl. C) Charge states (left) and subsequent deconvolution of mAcpB region after 2 hours (right). Blue shading corresponds to holo-mAcpB, green corresponds to unreacted malonyl-mAcpB and red corresponds to condensation product **16**. D) Ppant ejection of condensation species (left). Chemical structure and exact mass of Ppant ion of **16**.*

In addition to formation of **16$_B$**, however, a further mAcpB species were observed. Propionyl-mAcpB **15$_B$** (observed; 11769.4 Da, expected; 11768.8 Da) and *holo*-mAcpB (observed; 11714.5 Da, expected; 11712.7 Da) were also produced in significant quantities (Figure 45C). The formation of *holo*-mAcpB occurs as the diketone intermediate **16$_B$** does not stably reside at the active site of mAcpB. Literature studies of the KS from LovF from the biosynthesis of lovastatin corroborates this finding. [129] During this work, acetoacetyl-LovF species produced by acylation of the ACP was found to readily degrade to *holo*-LovF, even in the presence of substantial molar excess of acetoacetyl-CoA. Hydrolysis at the thioester position was proposed to be the cause of this degradation of diketone intermediate. A similar mechanism is expected to be responsible for *holo*-mAcpB observed in these assays. A transfer of propionyl to the mAcpB is also occurring to form **15$_B$**. This could form due to self-loading of mAcpB

with propionyl moiety or by transfer of propionyl from MmpF to mAcpB which has been observed for other KS enzymes. [112,130]

3.2.9 *In-vitro* MmpF condensation assay with 3HP-mAcpD



Scheme 11: The reaction scheme showing the condensation reaction catalysed by MmpF with 3HP. Decarboxylative Clasien condensation occurs with 3HP and malonyl-mAcpB catalysed by MmpF, producing 5-carbon dihydroxy intermediate attached to mAcpB ***17B.***

The next aim of the project was to demonstrate KS-catalysed DCC with an authentic 3HP substrate attached to mAcpD. The proposed scheme for the decarboxylative Claisen condensation with 3HP substrate is shown in Scheme 11. As previously discussed, chemical synthesis to produce 3HP-pantetheine was ultimately unsuccessful. Therefore, an enzymatic approach was undertaken to replicate the production of 3HP-mAcpD. The following proteins were expressed and purified; MupN, mAcpD, mAcpB, MupQ, MupS and MmpF. The proteins mAcpD, MupN, MupQ and MupS were all required to form mAcpD-3HP in a previously characterised assay.[86] The MmpF and mAcpB were produced for the condensation reaction itself. All purifications were conducted within three days and according to previously successful protocol to ensure all proteins were newly purified and functional.[86] SDS-PAGE characterisation confirmed the purity of each protein (Figure 46C). mAcpD was purified entirely in the *apo*-form and was converted to *holo* utilising endogenous PPT*ase* MupN and cofactors. Complete formation of *holo*-mAcpD was observed by ESI-MS after 2 hours and following this, addition of malonyl CoA (1 mM), MupQ (5 µM), ATP (1 mM), MupS (5 µM) and NADH (1 mM) was successful in producing mAcpD-3HP **10D** as confirmed by ESI-MS (observed; 15337.2 Da, expected; 15334.9 Da, Figure 46A) and subsequent ejection ion (observed; 333.26 Da, expected; 333.15 Da, Figure 46B).

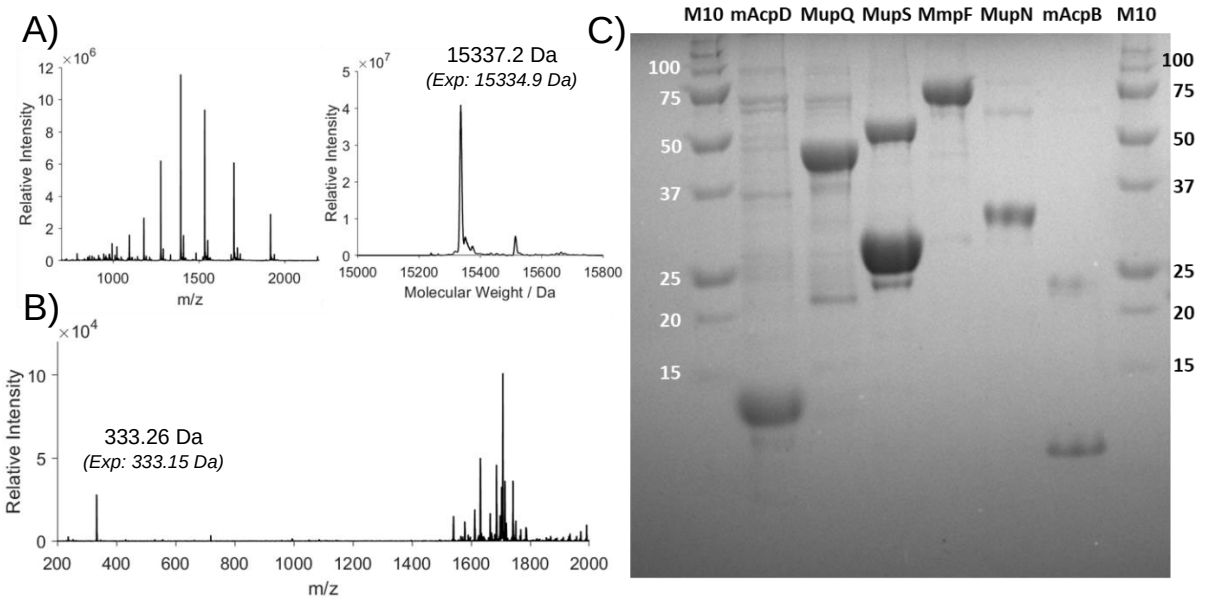


Figure 46: Figure showing formation of mAcpD-3HP and proteins required for MmpF chain extension assay. A) The charge states (left) and subsequent deconvolution (right) showing formation of mAcpD-3HP species. B) Pantetheine ejection spectrum showing formation of 3HP Ppant ion. C) SDS-PAGE gel showing the proteins purified for use in these assays. All proteins are sufficiently pure and concentrated. MupS was observed in both monomeric and dimeric forms on SDS-PAGE gel.

To test the ability of MmpF to undergo DCC with natural substrate 3HP, MmpF (20 µM) was incubated with **10$_D$** (120 µM) and the reaction monitored by ESI-MS. For clarity, **10$_D$** corresponds to 3-hydroxypropanoate (3HP) bound via the terminal thiol to wild-type ACP domain mAcpD. The large molecular weight of MmpF meant that despite an envelope of charge states being observed, the signal to noise ratio was too poor to allow accurate mass determination. The assay was therefore monitored by analysis of the ACP charge states only. After 30 mins, a new mAcpD species corresponding to *holo*-mAcpD was produced in 40% yield (observed; 15263.8 Da, expected; 15262.8 Da, Figure 47E). This is consistent with translocation of 3HP to MmpF, thereby forming *holo*-mAcpD. Following this, the condensation reaction was then initiated by addition of malonyl-mAcpB (20 µM). After a further 30 minutes, analysis of the mAcpB deconvolution showed a new species with a mass consistent with condensation product **17$_B$** (observed; 11830.3 Da, expected; 11825.8 Da, Figure 47B). Further fragmentation of this species gave a pantetheine ejection ion consistent with being derived from **17$_B$** (observed; 375.25 Da, expected; 375.16 Da, Figure 47C). As was observed with previous condensation reaction using propionyl starter unit, a significant amount of *holo*-mAcpB was also produced. The ESI-MS data confirmed that MmpF can catalyse DCC with mAcpD-3HP and malonyl-mAcpB.

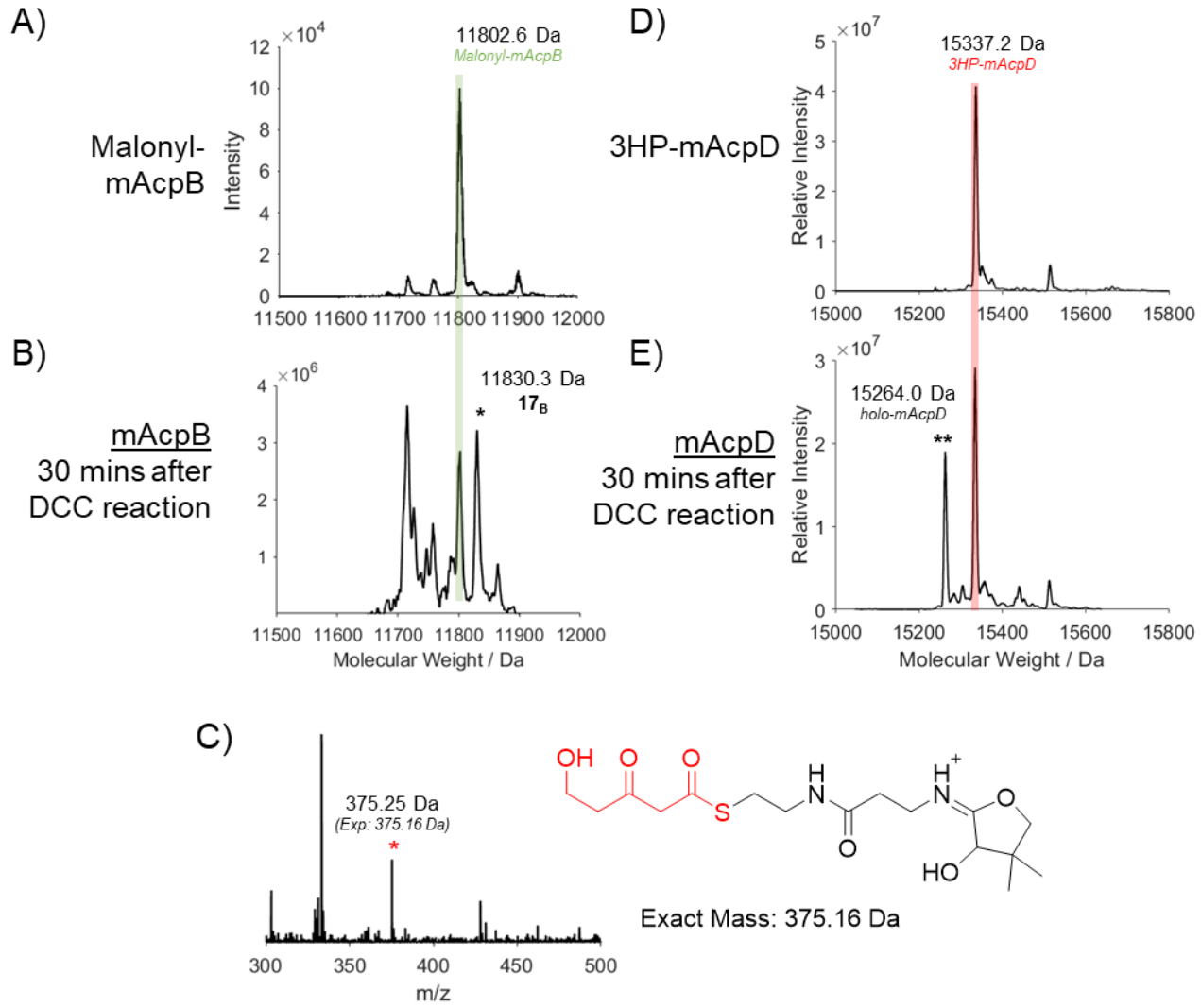
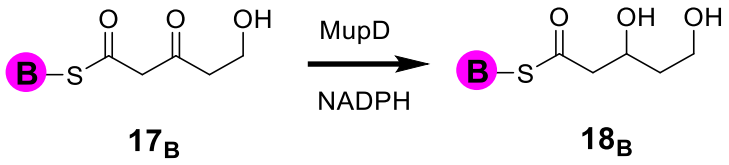


Figure 47: A figure showing the condensation reaction of 3HP-MmpF and malonyl-mAcpB. A) Deconvolution of malonyl-mAcpB prior to addition of MmpF and mAcpD-3HP. B) Deconvolution of the mAcpB region of the spectrum 30 minutes after initiation of DCC reaction, ***17B*** *species is asterisked. C) Further fragmentation of* ***17B*** *gives characteristic pantetheine ejection ion (left). Chemical structure and expected mass of this ejection ion (right). D) Deconvolution of the mAcpD region of the spectrum prior to addition of MmpF and malonyl-mAcpB. E) Deconvolution of the mAcpD region of the spectrum 30 mins after initiation of condensation reaction.*

Further optimisation of reaction conditions was unsuccessful in transforming a greater proportion of malonyl-mAcpB to **17B**. In addition, future attempts to form the condensation product were unsuccessful under identical conditions. Therefore, alternative ways to optimise the reaction were considered.

Previous studies of the kinetics of FAS from *E. coli* has shown interactivity between enzyme components is key to controlling pathway output. [131] Therefore, one method of increasing formation of **17B** may be to add downstream enzyme components to the DCC reaction. MupD is the proposed ketoreductase (KR) enzyme which is expected to act on the nascent **17B** to produce 3,5-dihydroxypentanoyl species **18B** (Scheme 12). The addition of MupD to the condensation reaction may

therefore assist with the turnover of the reaction. Therefore, attention was turned to expression and purification of ketoreductase enzyme MupD.



17_B 18_B

*Scheme 12: The reduction of diketone intermediate **17_B** to produce **18_B** via the action of KR enzyme MupD.*

After the action of MmpF, the next biosynthetic steps are expected to be the step-wise reduction of the diketone intermediate **17** to the fully-saturated acyl chain if a normal pattern of fatty biosynthesis is followed. This comprises a three-step process involving a ketoreducase (KR), dehydratase (DH) and enoyl reductase (ER) domains. MupD and MupE are free standing domains proposed to act as KR and ER respectively whereas a free-standing DH domain has not been identified.

3.2.10 Bioinformatic analysis of MupD

MupD is a member of the short-chain dehydrogenase (SDR) superfamily of proteins which consist of many of NAD(P)(H)-dependent oxidoreductases.[132] MupD shows specific homology to FabG-type enzymes of which *Sinorhizobium meliloti* FabG is the archetype. [133] This class of proteins catalyse the reduction of β-ketoacyl-ACP substrates to β-hydroxyacyl-ACP products using a conserved Ser-Tyr-Lys active site. The active site lysine residue acts to stabilise the deprotonated form of tyrosine thereby enhancing its role as an acid/base catalyst. Serine acts to stabilise and polarise the carbonyl substrate group and the asparagine residue located near the active site facilitates the proton transfer to active site lysine as shown in Figure 48. [134,135] SDR proteins bind either NADH or NADPH and specificity for the latter is dictated by the presence of a basic residue within the Gly-rich motif at the N-terminus of these proteins. [136] Sequence analysis of MupD shows the presence of a basic arginine residue (Arg-12) within this gly-rich motif indicating NADPH specificity. Further to this, crystal structures of related MupD homologues, such as *Brassica napus* KR domain, have been isolated with NADPH bound. [137]

The enzymatic reduction of a ketone to an alcohol is a stereospecific reaction leading to the formation of one enantiomerically pure product. MupD is a "B-type" KR due to the presence of HXAXXXXD motif. B-type KR domains catalyse the formation of stereospecific D-β-hydroxy product (the R-enantiomer).[138] The structure of MupD was modelled (Figure 48) based on similarity to 2-R-hydroxypropylthioethanesulfonate dehydrogenase to which MupD shares 26% sequence identity. [139]

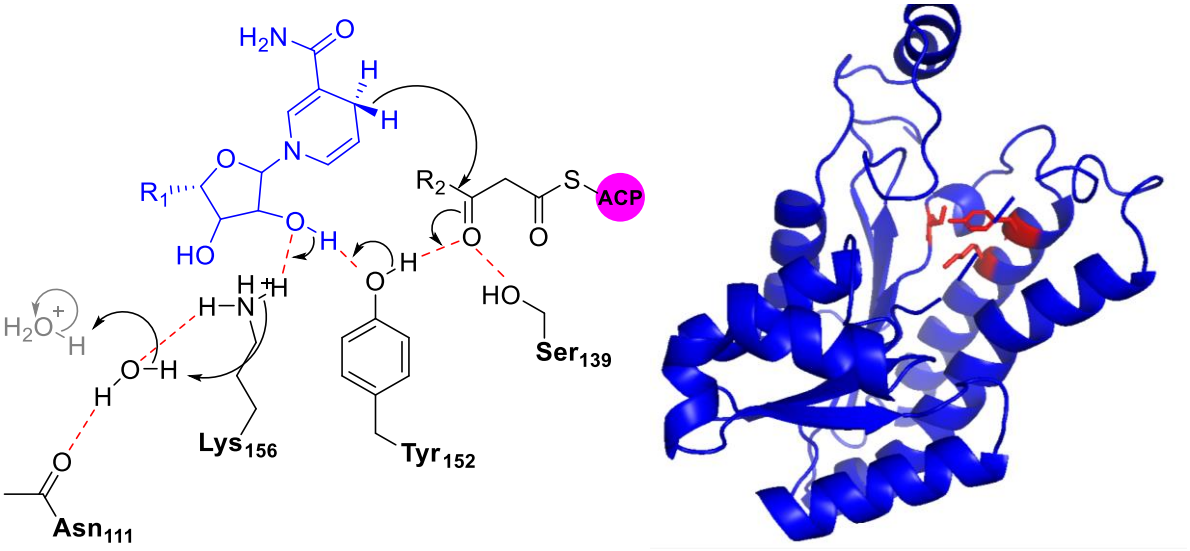


Figure 48: Postulated mechanism of MupD involving active site triad and asparagine, NADPH and diketone intermediate bound to terminal thiol of mAcpB. Catalysis is initiated by proton transfer from Tyr-152 hydroxyl to substrate carbonyl followed by hybrid transfer to alpha-carbon of diketone. A proton relay is formed involving 2'OH ribose of NADPH, Lys-156 side chain and Asn-111 bound H_2O. Proton relay finishes with proton abstraction from solvent H_2O. NADPH is converted to NADP+ in this process. R_1 group corresponds to remainder of NADPH structure. R_2 group corresponds to the growing fatty acid, for natural substate of MupD reaction $R_2 = CH_2CH_2OH$. The modelled structure of MupD is shown on the right. Active site triad are highlighted in red.

3.2.11 Expression and purification of MupD

The amino acid sequence corresponding to MupD was identified using the original gene annotation of *P. fluorescens*.[78] Primers were designed to clone MupD are given in the methods section. The PCR conducted with these primers was successful and gene fragments of mupD were ligated into pOPINF vector producing a viable plasmid as confirmed by Sanger sequencing. Initial attempts to express MupD under standard conditions (NEB BL21 cells, LB media, 0.25 mM IPTG) produced insoluble protein when purified and analysed by SDS-PAGE. The expression protocol was repeated utilising Arctic Cell Expression. Following this, the protein was purified by IMAC and SEC using a S100 column. This purification was successful in producing 1 mg/L of soluble protein as confirmed by SDS-PAGE analysis (Figure 49A). ESI-MS of purified MupD confirmed identity of this species (observed; 28031 Da, expected; 28033 Da, Figure 49B). Analytical size exclusion chromatography (utilising an S200 column) showed that MupD was a trimer in solution. This is unusual for KR enzymes which are typically tetramers. [140]

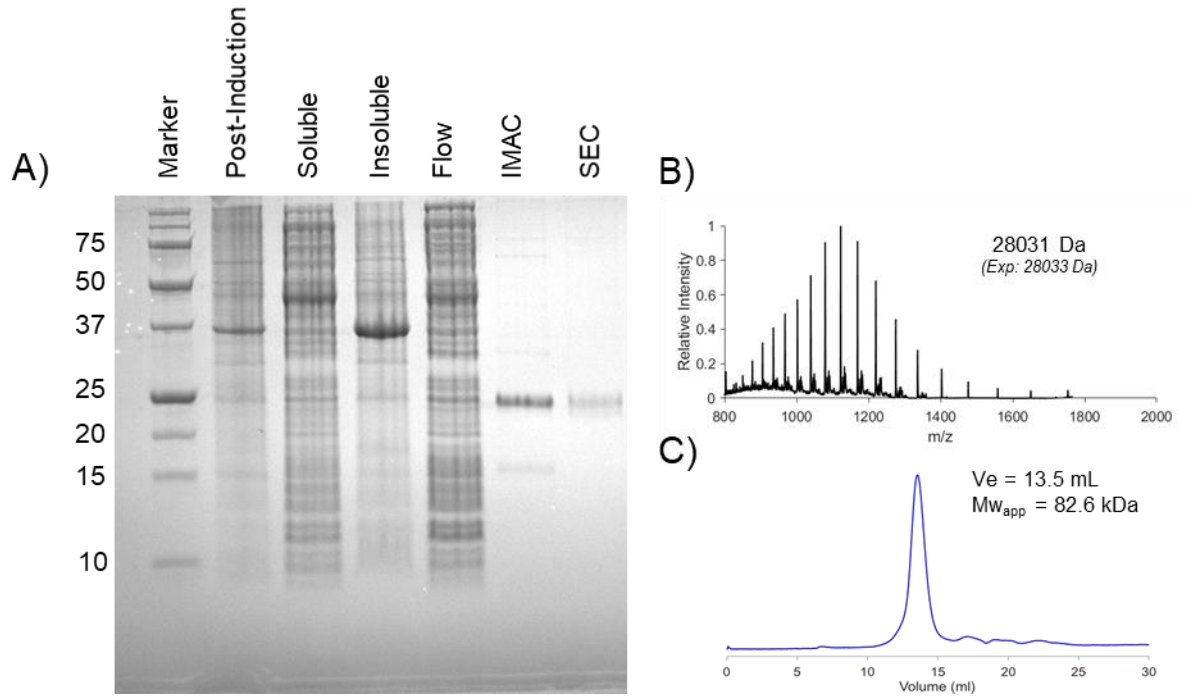
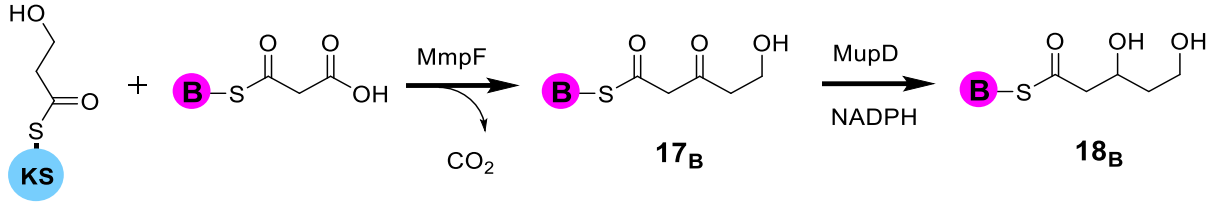


Figure 49: Expression and purification of MupD. A) SDS-PAGE gel showing the different stages of MupD expression using Arctic cells. A single band of pure MupD protein was observed after IMAC and size exclusion chromatography. B) ESI-MS spectrum of purified MupD. A single series of charge states showed a mass consistent with MupD (observed; 28031 Da, expected; 28033 Da). C) Analytical size exclusion chromatography trace of MupD, this shows a trimeric state in solution (Mw_{app}= 82.6 kDa).

3.2.12 *In-vitro* MmpF condensation assay with 3HP-mAcpD and downstream KR



*Scheme 13: A reaction scheme showing the MmpF chain extension reaction with natural substrate 3HP and malonyl-mAcpB. Chain extension reaction will be catalysed by MmpF and produce a 5-carbon diketone intermediate **17$_B$** attached to mAcpB. Resulting diketone will be reduced to dihydroxy intermediate **18$_B$** catalysed by KR enzyme MupD and reducing cofactor NADPH.*

The condensation reaction involving MmpF and the natural substrate, 3HP-mAcpD, was repeated in the presence of the proposed KR enzyme MupD. The proposed reaction is summarised in Scheme 13. The assay was initiated by addition of MmpF (15 µM) to mAcpD-3HP (37 µM), malonyl-mAcpB (50 µM), MupD (15 µM) and NADPH (1 mM). The reaction was monitored by ESI-MS after 30 minutes and 16 hours incubation at room temperature. After 30 minutes, the mAcpB region of the spectrum showed consumption of malonyl-mAcpB and formation of a new mAcpB species with a mass consistent with the 5-carbon dihydroxypenanoyl species **18$_B$** (observed; 11827.1 Da, expected: 11827.8 Da, Figure

50B). Further fragmentation of this species gave the expected characteristic ejection ion (observed; 377.15 Da, expected: 377.17 Da, Figure 50B). This confirmed that condensation has taken place and that MupD catalyses the downstream KR reaction due to the complete production of **18$_B$** at the expense of **17$_B$**. The major species observed was *holo*-mAcpB even after only 30 minutes (observed; 11711.2 Da, expected; 11712.7 Da) and **18$_B$** was present at ~40% the level of *holo*-mAcpB.

After 16 hours, the presence of **18$_B$** diminished whereas *holo*-mAcpB increased further suggesting slow hydrolysis of **18$_B$** over time (Figure 50C). The hydrolysis of mAcpB to produce *holo*-mAcpB may be facilitated by interaction with MmpF. Previous studies have shown that when an ACP bearing an acyl moiety interacts with its cognate partner, exposure of the acyl moiety occurs. [141] The interaction of mAcpB and MmpF may therefore increase the amount of *holo*-mAcpB formed. In addition, the chemical structure of **18$_B$** contains a hydroxyl at the C-5 position. This is perfectly placed for intramolecular attack of the thioester carbonyl to produce a stable six membered lactone further facilitating hydrolysis.

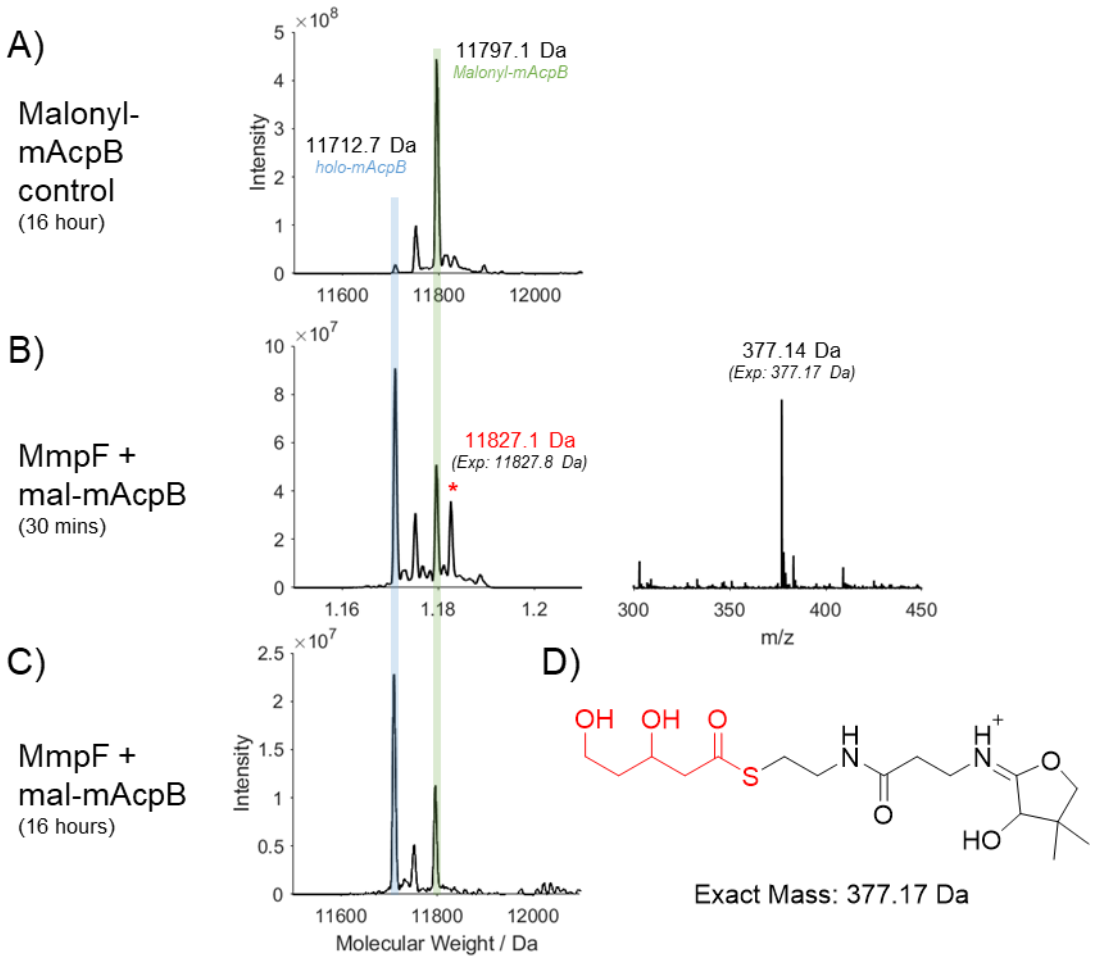


Figure 50: The MmpF DCC reaction using malonyl-mAcpB and 3HP-mAcpD in the presence of MupD and NADPH. A) Malonyl-mAcpB incubated under assay conditions without addition of any enzymes or cofactors. B) The deconvolution of the mAcpB region 30 minutes after initiation of DCC reaction. Asterisked species is condensation

*product **18B** (left). Ppant. ejection of 18 gives ejection ion of 377.15 Da confirming identity of this species (right). C) The deconvolution of the mAcpB region 16 hours after initiation of DCC reaction. D) Chemical structure of the pantetheine ejection ion of dihydroxypenanoyl intermediate **18B**.*

Analysis of the mAcpD region of the ESI-MS trace after 30 minutes showed consumption of mAcpD-3HP and its replacement with a new species with mass consistent with *holo*-mAcpD (observed; 15260.4 Da, expected; 15262.8 Da, Figure 52D) which made up 90% of the mAcpD species observed. This is consistent with translocation of 3HP to MmpF.

To confirm that MmpF is catalytically responsible for the formation of **18B**, a mutant of MmpF lacking the active site cysteine (Cys183) was constructed. This plasmid was known as MmpFC183A and was constructed by PCR mutagenesis of the existing MmpF pET28a vector using a modified PCR procedure (Primers and PCR protocol in Methods). Successful PCR was confirmed by DNA gel of resulting PCR product. Purification of PCR product and ligation into pET28a plasmid was conducted, full Sagner Sequencing of resulting plasmid confirmed that a single mutation of the active site cysteine to an alanine had taken place.

The same expression protocol used to purify MmpF was found to be successful for MmpFC183A. Purification of MmpFC183A was conducted by IMAC, followed by SEC utilising the S200 column. High salt (0.5M) and glycerol (10%) was required to limit precipitation of protein (as was observed for wild type MmpF). Characterisation of MmpFC181A was conducted by SDS_PAGE (Figure 51A) which confirmed the production of pure protein. ESI-MS analysis of this purified sample showed a mass of 84801 Da consistent with the expected mass (expected; 84810 Da).

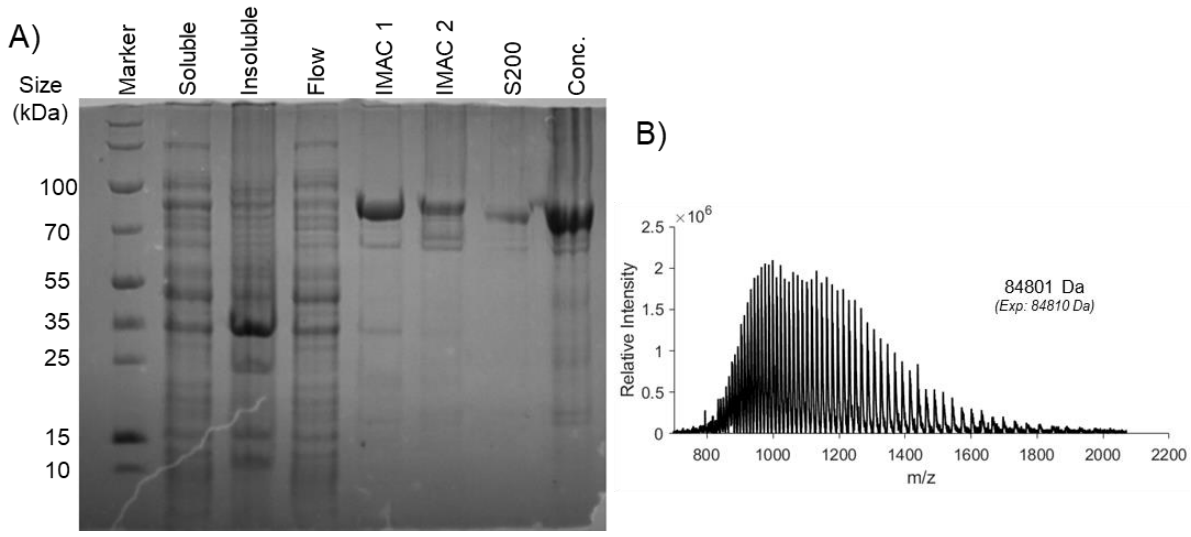


Figure 51: Characterisation of MmpFC183A mutant. A) SDS-PAGE showing the various stages of MmpFC183A purification. B) ESI-MS analysis of MmpFC183A shows an observed mass of 84801 Da which is 9 Da away from expected mass of 84810 Da.

The negative control assay was initiated by addition of MmpFC183A (15 µM) to mAcpD-3HP (37 µM), malonyl-mAcpB (50 µM), MupD (15 µM) and NADPH (1 mM) at room temperature. After 30 minutes, ESI-MS analysis of mAcpB region of spectrum showed that the major species was malonyl-mAcpB (observed; 11797.1 Da, expected; 11797.8 Da, Figure 52B). The comparison of this spectrum to only malonyl-mAcpB is shown in (Figure 52A and B). No major change was observed between the two spectra confirming no formation of **18$_B$** and therefore no condensation had taken place. Further to this, analysis of the mAcpD region of the spectrum after this same time showed mAcpD-3HP as the predominant species (observed; 15334.7 Da, expected; 15334.9 Da, Figure 52C). This is consistent with a lack of translocation of 3HP to MmpF due to the lack of an active site cysteine required for condensation.

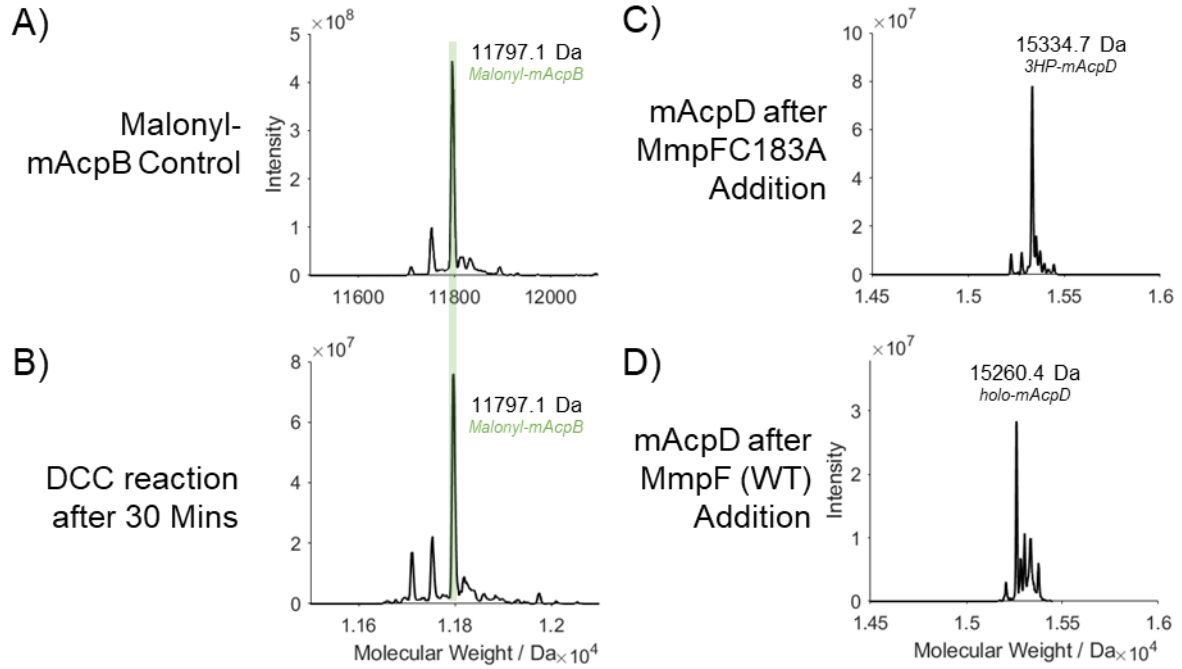


*Figure 52: The condensation reaction conducted with malonyl-mAcpB, 3HP-mAcpD and MmpFC183A in the presence of MupD and NADPH. A) The spectrum of malonyl-mAcpB prior to addition of any enzymes or cofactors. B) The deconvolution of the mAcpB region of the spectrum 30 minutes after initiation of DCC reaction. No formation of **18$_B$** is observed. C) The deconvolution of the mAcpD region of the spectrum 30 minutes after initiation of the reaction. D) For comparison, the mAcpD region of spectrum 30 minutes after initiation of WT MmpF condensation reaction.*

The data confirmed that MmpF can catalyse the key carbon chain extending reaction with mAcpB delivering a malonyl group and mAcpD delivering 3HP. In addition to this, the role of MupD has also been confirmed. MupD utilises NADPH to reduce **17$_B$** to form reduced species **18$_B$**.

3.2.13 *In-vitro* MmpF condensation assay with 3HP-mAcpD_T and downstream KR

In earlier work, a mAcpD_T mutant lacking the C-terminal extension was found to undergo reaction with MupQ or MupS as confirmed by ESI-MS. Further to this, NMR titrations between mAcpD and

MmpF showed residues located on the C-terminal extension do not interact significantly with MmpF. This suggested that the C-terminal extension might not be required successful translocation of 3HP to MmpF. To test this, the transfer assays were repeated using mAcpD_T-3HP **10$_{D*}$** in place of full length mAcpD. **10$_{D*}$** (37 µM) was incubated with MmpF (15 µM), malonyl-mAcpB (50 µM), MupD (15 µM) and NADPH (1 mM) at room temperature. After 30 minutes, ESI-MS showed the consumption of malonyl-mAcpB and the formation of a new species with mass consistent with **18$_B$** (observed; 11826.8 Da, expected; 11827.8 Da, Figure 53B). Ppant ejection confirmed this species to be **18$_B$** (observed; 377.15 Da, expected; 377.17 Da, Figure 53B). This confirmed condensation occurs with mAcpD_T delivered 3HP. Smaller yields of **18$_B$** were observed for mAcpD_T when compared to assays conducted with WT mAcpD after the same time period (30 minutes). As with the mAcpD wild-type assay, the dominant mAcpB species after 30 minutes of reaction was *holo*-mAcpB.

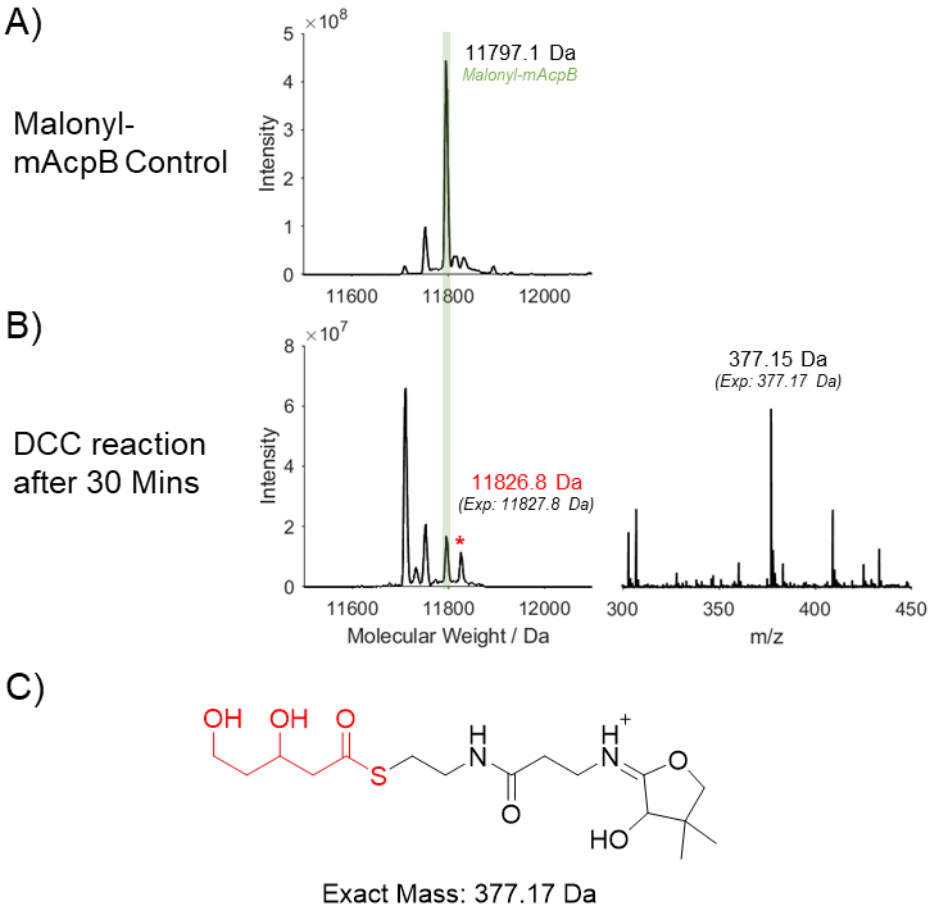


Figure 53: A figure showing the MmpF DCC reaction using malonyl-mAcpB and 3HP-mAcpD_T in the presence of MupD and NADPH. A) The deconvolution of a control spectrum of malonyl-mAcpB before addition of any enzymes or cofactors. B) The deconvolution of the mAcpB region of the spectrum 30 minutes after DCC assay was initiated. ***18$_B$*** *signified by asterisk (left). Subsequent Ppant ejection assay of* ***18$_B$*** *(right). C) The chemical structure and expected mass of Ppant ion formed after fragmentation of* ***18$_B$****.*

Analysis of the mAcpD_T region of the spectrum showed that *holo*-mAcpD_T (observed; 10552.8 Da, expected; 10551 Da) made up 33% of the species present, with the remainder comprised of unreacted starting material 3HP-mAcpD_T (observed; 10624.7 Da, expected; 10623 Da). This represented a slight reduction of hydrolysis when compared to the wild type mAcpD after the same time period. Analysis of the spectrum showed addition of sodium (+ 22 Da) and addition of potassium (+ 38 Da) adducts to mAcpD_T species. The addition of salt adducts to ACPs grown utilising *E. coli* has been observed previously and will not alter the reaction. [142]

An additional assay was conducted, this time utilising mAcpD_T and MmpFC183A. mAcpD_T-3HP (37 µM) was incubated with MmpFC183A (15 µM), malonyl-mAcpB (50 µM), MupD (15 µM) and NADPH (1 mM) at room temperature. After 30 minutes, analysis of the mAcpB region of the spectrum showed no formation of **18$_B$** as expected. The mAcpD_T region of the spectrum showed only unreacted 3HP-mAcpD_T (observed; 10625.3 Da, expected; 10623 Da), consistent with no translocation of 3HP.

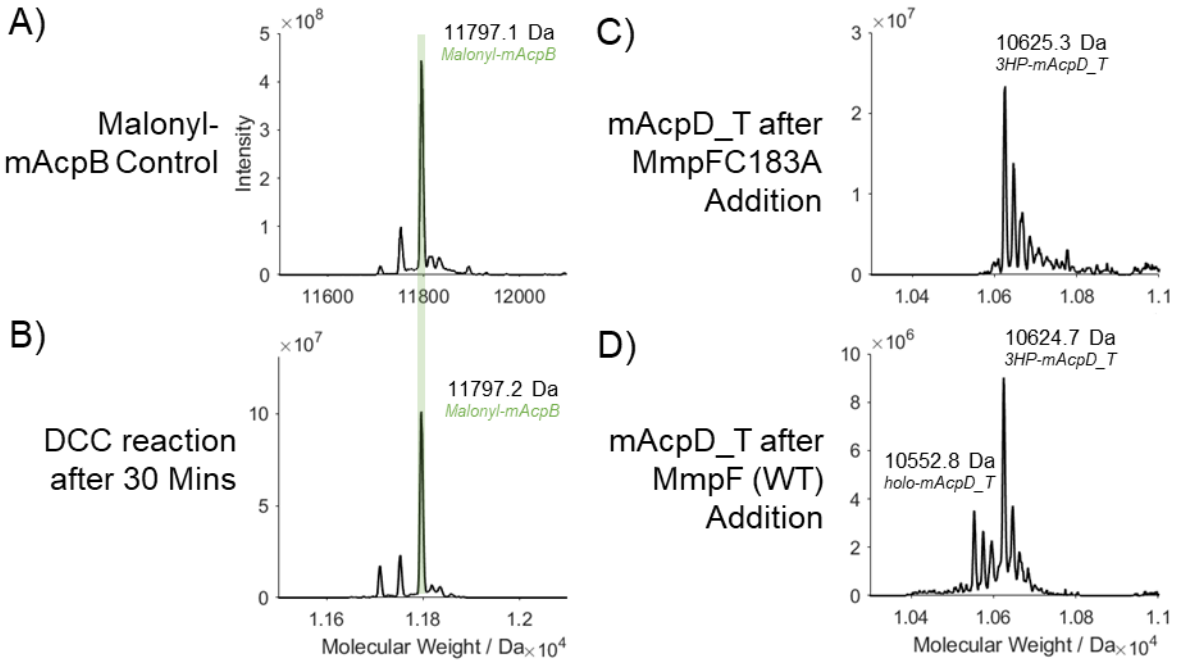


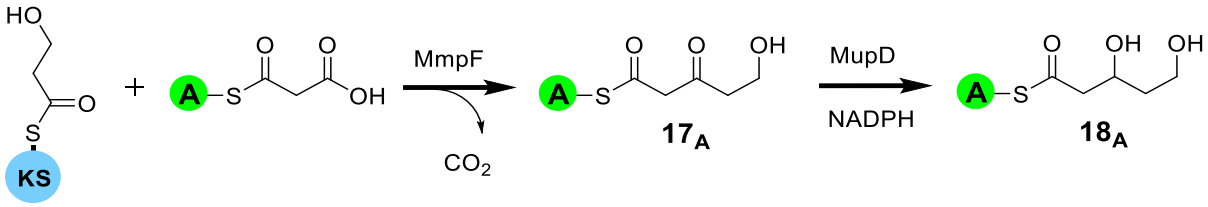
Figure 54: A figure showing the outcome of the MmpFC183A + malonyl-mAcpB condensation reaction with authentic substrate 3HP attached to mutant mAcpD_T. A) Deconvolution of the mAcpB region of the spectrum prior to addition of MmpF_C183A. B) Deconvolution of the mAcpB spectrum 30 minutes after addition of MmpF_C183A and 3HP. No DCC product is observed. C) Deconvolution of mAcpD_T region of the spectrum after addition of MmpF_C183A. D) Deconvolution of mAcpD_T region of the spectrum after addition of MmpF.

These results provide the clearest evidence yet for the role of MmpF as the KS enzyme involved in condensing 3HP-mAcpD. In addition to this, mAcpB is confirmed as an ACP that can deliver a malonyl to MmpF to trigger condensation. Following condensation, mAcpB will deliver **17$_B$** to MupD which

reduces this to form **18$_B$**. This provides new evidence for the sequence of biosynthetic steps which lead to the production of 9HN.

Also in this work, the ability of mAcpD_T to deliver 3HP to MmpF was tested. Although no in-depth kinetic analysis was conducted, mAcpD_T mutant was evidently slower at translocating 3HP to MmpF. After 30 minutes, complete translocation of 3HP from mAcpD to MmpF was observed by ESI-MS. Under identical conditions, only 33% of 3HP underwent translocation using a mutant of mAcpD lacking this C-terminal extension (mAcpD_T). This provides evidence that the C-terminal extension of mAcpD provides a beneficial interaction with MmpF. This said, the C-terminal extension is not needed for interaction with any of MupQ, MupS or MmpF.

3.2.14 *In-vitro* MmpF condensation with malonyl-mAcpA and downstream KR



*Scheme 14: Scheme showing the MmpF DCC reaction utilising malonate delivered by mAcpA in the presence of MupD and NADPH. Diketone intermediate **17$_A$** will be formed which is reduced by MupD to produce the dihydroxypentanoyl intermediate **18$_A$**.*

It is not known whether malonyl transfer to MmpF is unique to mAcpB. The condensation reaction with MmpF was therefore repeated using the remaining unassigned discrete ACP, mAcpA. Initially, it was anticipated that an equivalent functional role to mAcpB would not be observed, and this would serve as a useful control experiment. To test this, malonyl-mAcpA was generated using MupN and malonyl CoA. The mAcpA was successfully converted to *holo* although the sample contained a small fraction of unmodified *apo*-mAcpA. This was not expected to interfere with the reaction beyond some low-level inhibition so further separation was not attempted. [96] In addition, *apo*-/malonyl-mAcpA adducts with Na^+ (observed; 12336 Da, expected; 12337 Da) and K^+ (observed; 12353 Da, expected; 12353 Da) were observed and presumably derived from the assay buffers (Figure 55A).

The MmpF DCC assay was initiated by addition of malonyl-mAcpA (50 µM) to MmpF (15 µM) and mAcpD-3HP (37 µM) and was conducted in the presence of MupD (15 µM) and NADPH (1 mM). After 30 minutes, ESI-MS analysis of the mAcpA region of the spectrum surprisingly showed formation of a new species with mass consistent with **18$_A$** (observed; 12344.7 Da, expected; 12344.5 Da, Figure 55B). Further fragmentation of this species yielded a characteristic pantetheine ejection ion of 377.16 Da (exp; 377.17 Da, Figure 55B) confirming formation of **18$_A$**.

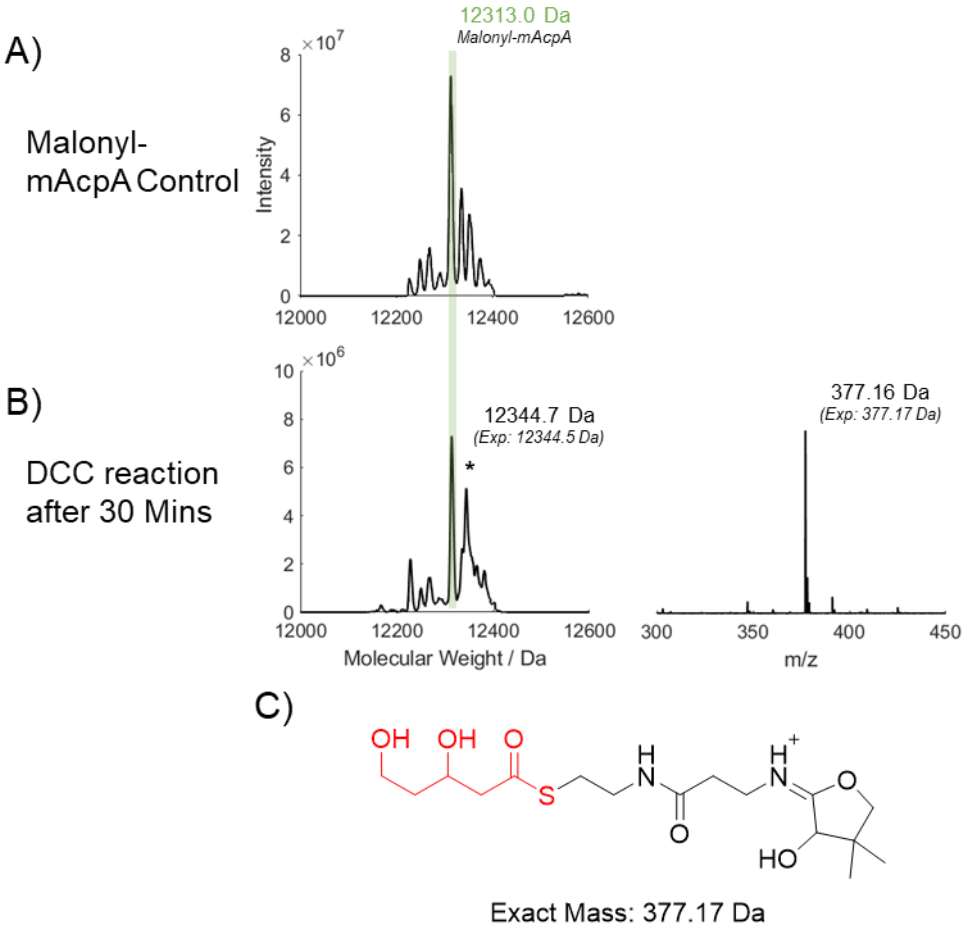


*Figure 55: A figure showing the MmpF DCC reaction utilising mAcpA delivered malonate. A) The deconvolution of malonyl-mAcpA prior to addition of any enzymes or cofactors. B) The deconvolution of mAcpA, 30 minutes after initiation of DCC reaction (left) and subsequent Ppant ejection ion (right). Condensation product **18** is shown with an asterisk. C) Chemical structure and expected mass of the ejection ion formed from fragmentation of **18**.*

In total, the reaction was monitored over a full time course including 30 minutes, 3 hours and 16 hour time points (Figure 56). After 3 hours, **18**$_A$ was observed but in smaller amounts (observed; 12344.1 Da). This was consistent with previous assays which showed that condensation product undergoes hydrolysis over time. However, after this time**,** a new species was also observed with a mass of 12358.0 Da. After 16 hours, the product **18**$_A$ was completely lost and this stable new species predominated, being produced in over 50% yield (observed; 12358.9 Da, Figure 56D). Pantetheine ejection of this new species produced a 391.17 Da ion (Figure 39E) confirming that this intermediate was bound to the phosphopantetheine arm. The gain of 14 Da was consistent with a moiety whereby a saturated carbon is replaced by a carbonyl group. One possibility consistent with this is the new species 3,5-dioxopentanoyl-mAcpA (DOP-mAcpA) **20**$_A$ (expected; 391.15 Da; observed 391.17 Da). This could be formed from condensation involving two malonyl moieties to produce the tri-ketone species **19**$_A$ followed by MupD-catalysed reduction to produce **20**$_A$ (shown in the scheme in Figure 56A). The structure of malonyl is very similar to 3HP, the only difference is extra carbonyl functionality at C-3 so

its elongation by MmpF is not surprising. **20_A** was not observed during equivalent elongation reaction with malonyl-mAcpB. It therefore seems that mAcpA has the unique ability to transfer malonyl to active site cysteine of MmpF. Following this, condensation of KS-bound malonate and ACP-bound malonate occurs to produce **19_A** which is reduced by MupD to **20_A**. mAcpB must lack the ability to transfer malonyl to MmpF therefore no production of **20_B** is observed with equivalent mAcpB assay.

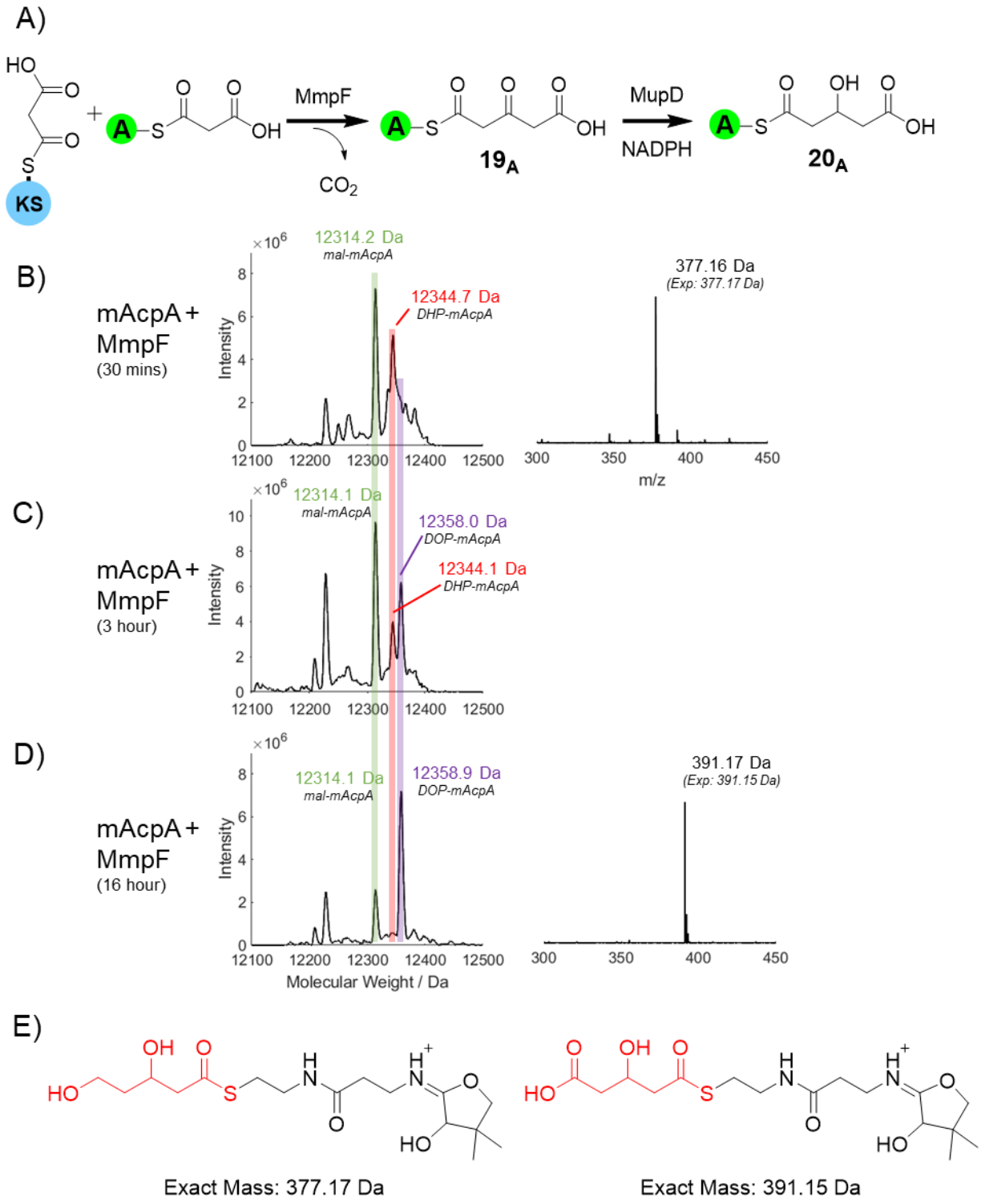


*Figure 56: Comparison of different time points of malonyl-mAcpA + MmpF condensation reaction. DHP-mAcpA is the expected condensation product **18A**. DOP-mAcpA is the 3,5-dioxopentanoyl-mAcpA (DOP-mAcpA) **20A** formed from condensation of two malonyl moieties followed by reduction. A) Scheme showing the formation of **20A** from condensation of two malonyl units. B) Deconvolution of mAcpA region of spectrum 30 minutes after initiation of condensation reaction (left). Ppant ejection of 12344.7 Da species (right). C) Deconvolution of mAcpA region of spectrum 3 hours after initiation of condensation reaction. D) Deconvolution of mAcpA region of spectrum 16 hours*

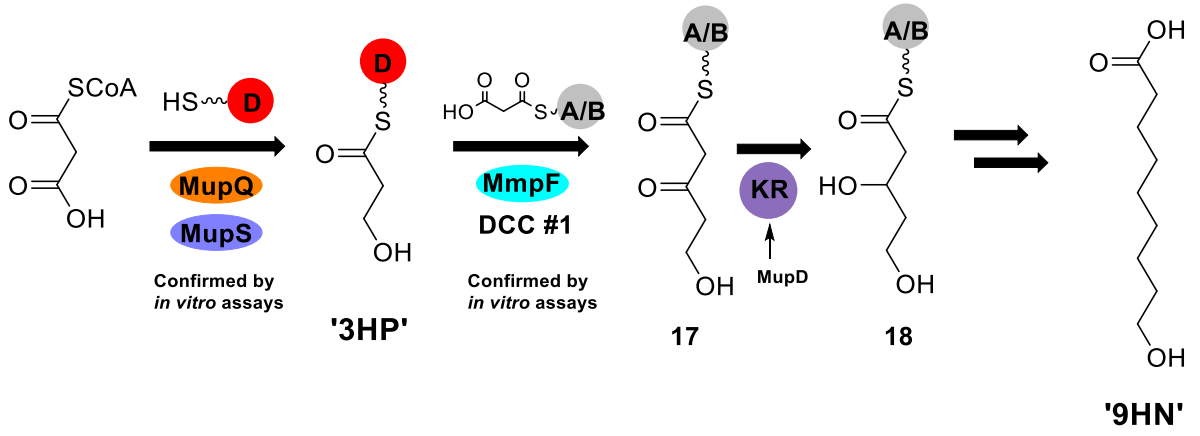
after initiation of condensation reaction (left). Ppant ejection of 12358.9 Da species (right). E) Chemical structure and expected masses of Ppant ions of ***18$_A$*** *and* ***20$_A$.***

This result confirmed that malonyl-mAcpA is a substrate for acylated MmpF but suggests a model in which both mAcpA and mAcpB can perform the first elongation step of the fatty acid extension. Tandem ACPs have been observed in many biosynthetic systems including difficidin, [84] lankacidin [143] and mupirocin itself. Within mupirocin BGC, an additional tandem ACP is found in the final module responsible for construction of monic acid backbone of MmpA. Inactivation of each of these ACPs in turn reduced mupirocin production without abolishing it. This suggests tandem ACPs act in parallel to increase overall pathway flux. [144] Tandem ACPs seemingly occur in regions of the pathway where the presence of a single ACP would cause malonylation of the ACP to be a rate limiting step. It seems that mAcpA and mAcpB might act in parallel to ensure condensation of mAcpD-3HP with malonyl-ACP is not the rate limiting step.

The results of work conducted in this section propose the following model of the early steps of 9HN biosynthesis shown in Scheme 15. Starter unit 3HP **10** is delivered to MmpF which catalyses condensation reaction with malonyl delivered by mAcpA or mAcpB. Following this, 5-carbon 1,3-diketone intermediate **17** is reduced to 5-carbon 3,5-dihydroxy intermediate **18** by the action of KR enzyme MupD. These biochemical steps have been shown by the expression and purification of mAcpA, mAcpB, MmpF, MmpB_KS, MupB, MupD and MupE. HSQC NMR titrations and *in vitro* assays monitored by ESI-MS have confirmed the formation of **18** from malonyl CoA. Further to this, MmpF has been shown to turnover DCC with non-natural propionyl moiety to produce an alternative 5-carbon diketone intermediate **16**. The ability of MmpF to catalyse condensation reaction with non-natural propionyl moiety may provide evidence that supply of precursor is the likely determinant of fatty acid composition. [145]

Furthermore, the delivery of 3HP to MmpF (translocation step) has been tested by using mutant of mAcpD lacking the usual C-terminal extension (mAcpD_T). In the absence of detailed kinetic analysis, mAcpD_T is slower to deliver 3HP to MmpF providing a potential role for this unusual structural feature.

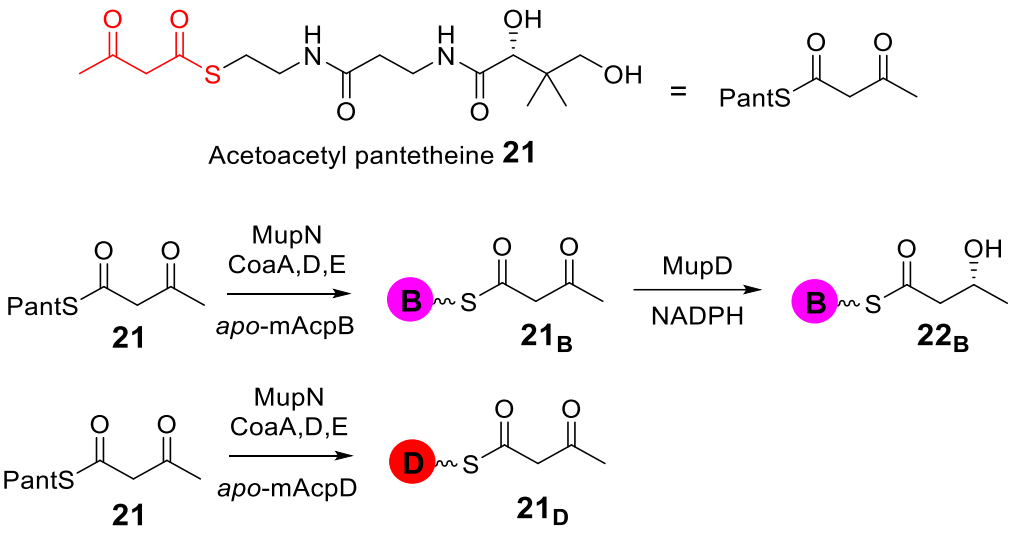
This chapter will conclude with further assays with proposed KR enzyme MupD and mimics of the natural substrate **18**.



Scheme 15: A scheme showing the proposed early steps in the biosynthesis of 9HN. The roles of mAcpD, MupQ, MupS, MmpF, mAcpA /mAcpB and MupD within the biosynthesis of 9HN have been elucidated.

3.2.15 MupD promiscuity with Acetoacetyl Mimic of Natural Substrate

The ability of MupD to act *in vitro* with simplified non-natural β-ketothiolester substrates was tested using acetoacetyl-pantetheine **21** (AcAc-pant). This substrate was synthesised by Angus Weir using the method reported by Dr Walker.[146] The reduction of **21** is expected to proceed in a stereospecific manner to produce a R-β-hydroxy product **22** due to presence of HXAXXXXD motif on MupD as discussed in Chapter 2. MupD has been shown to turnover with substrate attached to either mAcpA or mAcpB, mAcpB was chosen for these assays as it has been shown to accept the AcAc moiety more efficiently. The reduction of $\mathbf{21_B}$ by MupD is shown in Scheme 16. Substrate **21** was loaded onto mAcpB to produce AcAc-mAcpB $\mathbf{21_B}$. **21** was also loaded on to an ACP which is not expected to interact with MupD; mAcpD to produce $\mathbf{21_D}$. The ability of $\mathbf{21_B}$ and $\mathbf{21_D}$ to react with MupD was tested as is shown in Scheme 16.



*Scheme 16: The chemical structure of substrate **21** and the proposed reaction scheme for MupD catalysed ketoreduction using **21** attached to cognate (mAcpB) and non-cognate (mAcpD) ACPs.*

Starting with *apo*-ACPs, AcAc-Pant was transferred on to the active site serine of mAcpB and mAcpD in separate one-pot reactions utilising MupN (5 µM), ATP (1 mM), $MgCl_2$ (10 mM), **21** (1 mM), CoA mix enzymes (3 µM of each of CoaA, CoaD and CoaE) and *apo*-ACP (100 µM). These assays were conducted at room temperature and monitored by ESI-MS after 2 hours.

For formation of Acac-mAcpB **21$_B$**, after 2 hours *apo*-mAcpB was consumed and two new species were observed with masses consistent with **21$_B$** (observed; 11799.5 Da, expected; 11796.8 Da, 65% yield) and *holo*-mAcpB (observed; 11714.5 Da, expected; 11712.7 Da, 35% yield) as shown in Figure 57A. Major species produced Ppant ejection ion of 345.24 Da which confirms successful derivatisation of mAcpB. A small amount of acetyl adduct was also observed (observed; 11758.0 Da, expected; 11754.7 Da).

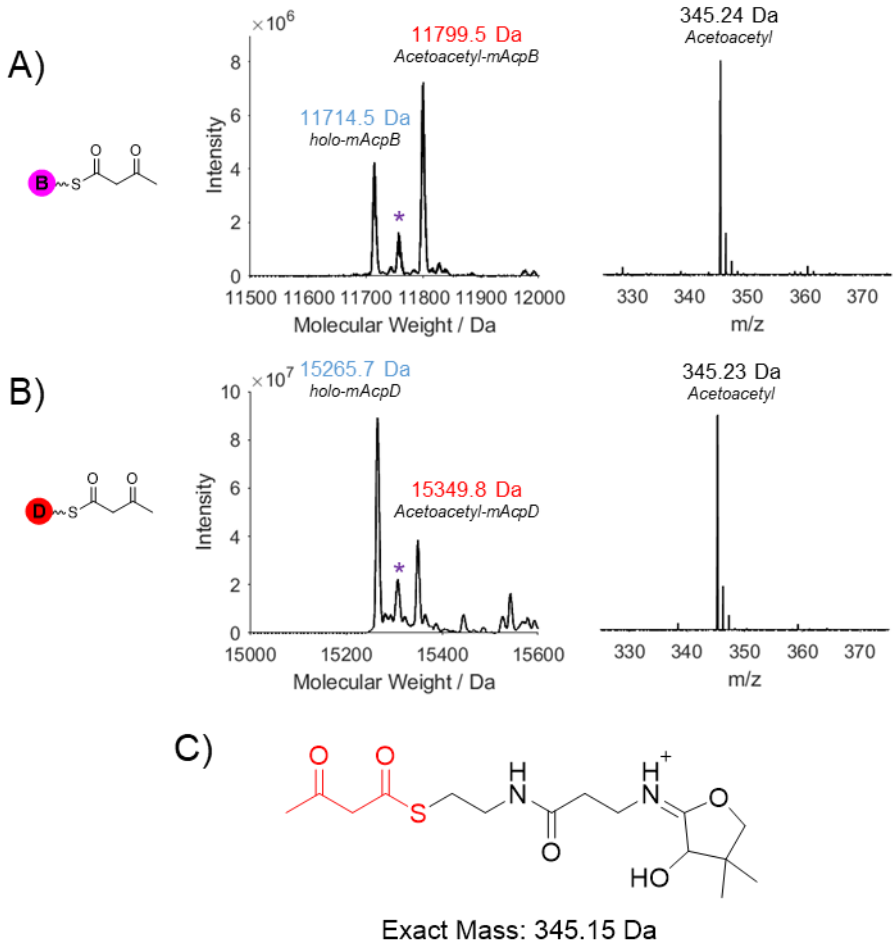
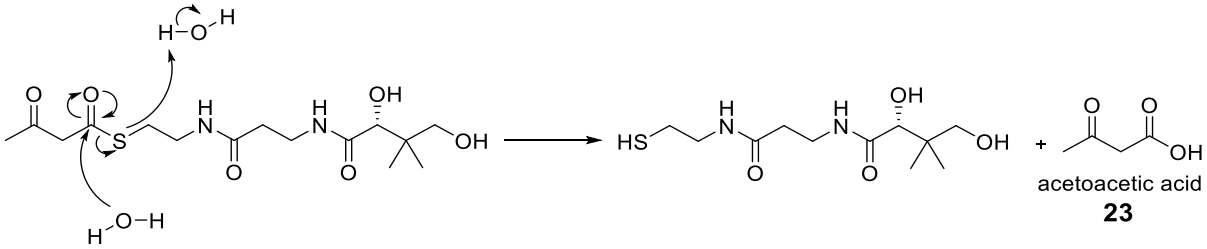


*Figure 57: Spectra showing the formation of acetoacetyl-mAcpB **21$_B$** and acetoacetyl-mAcpD **21$_D$**. A) The deconvolution of the mAcpB region of the Acetoacetyl transfer reaction after 2 hours (left). Further fragmentation of this species gives Ppant ion of 345.24 Da (right). Acetyl-mAcpB species is asterisked. B) The deconvolution of mAcpD after 2 hours of Acetoacetyl transfer reaction (left). Further fragmentation of AcAc-mAcpD gives Ppant ion of 345.23 Da (right). C) Chemical structure of acetoacetyl Ppant ion.*

The formation of *holo*-mAcpB was not expected as previous acyl transfer reactions led to complete formation of the acyl-ACP species. One explanation for the formation of *holo*-mAcpB would be the instability of the starting material **21**. Despite storage of **21** at -20 °C, the C-1 ketone may react with residual water to produce unfunctionalised pantetheine and acetoacetic acid **23** (Scheme 17). This unfunctionalised pantetheine will also be transferred to terminal thiol to produce *holo*-ACP. This was confirmed by high performance liquid chromatography (HPLC) analysis of **21** which showed the presence of unfunctionalised pantetheine. Other than low level competitive inhibition, *holo*-mAcpB is not expected to influence the KR reaction.



*Scheme 17: The proposed breakdown of acetoacetyl-pantetheine **21**. Nucleophilic attack by water occurs at the C-1 position of AcAc-pant causing the elimination of pantetheine and the production of acetoacetic acid **23**. Unfunctionalised pantetheine will lead to formation of holo-mAcpB in the assay with MupN.*

AcAc-pantetheine **21** was loaded on to mAcpD under identical conditions. After 2 hours, formation of **21$_D$** (observed; 15349.8 Da, expected; 15346.8 Da, 33% yield) and *holo*-mAcpD (observed; 15265.7 Da, expected; 15262.8 Da) were observed (Figure 57B). Further fragmentation of **21$_D$** produced Ppant ion of 345.23 Da which confirmed successful derivatisation of mAcpD.

3.2.15.1 KR activity of MupD with acetoacetyl-mAcpB

To test whether MupD could reduce **21$_B$** to the corresponding 3-hydroxybutyrl species **22$_B$** (as outlined in Scheme 16), *in vitro* assays were conducted. The reaction was initiated by addition of **21$_B$** (80µM) to MupD (20 µM) and NADPH (200 µM), incubated at room temperature and monitored by ESI-MS. The mass increase for a successful reduction was 2.02 Da which is a small change compared to the mass of the ACP. Therefore, the reaction was monitored via the pantetheine ejection ion of the predominant acyl-ACP species.

After 30 minutes, the major acyl-ACP species showed a mass of 11800.6 Da (expected for 3-hydroxybutyrl; 11798.8 Da). Further fragmentation of this species gave a pantetheine ejection ion of 347.25 Da (expected for 3-hydroxybutyrl Ppant; 347.16 Da) confirming complete formation of reduced product **22$_B$** (Figure 58A) after 30 minutes.

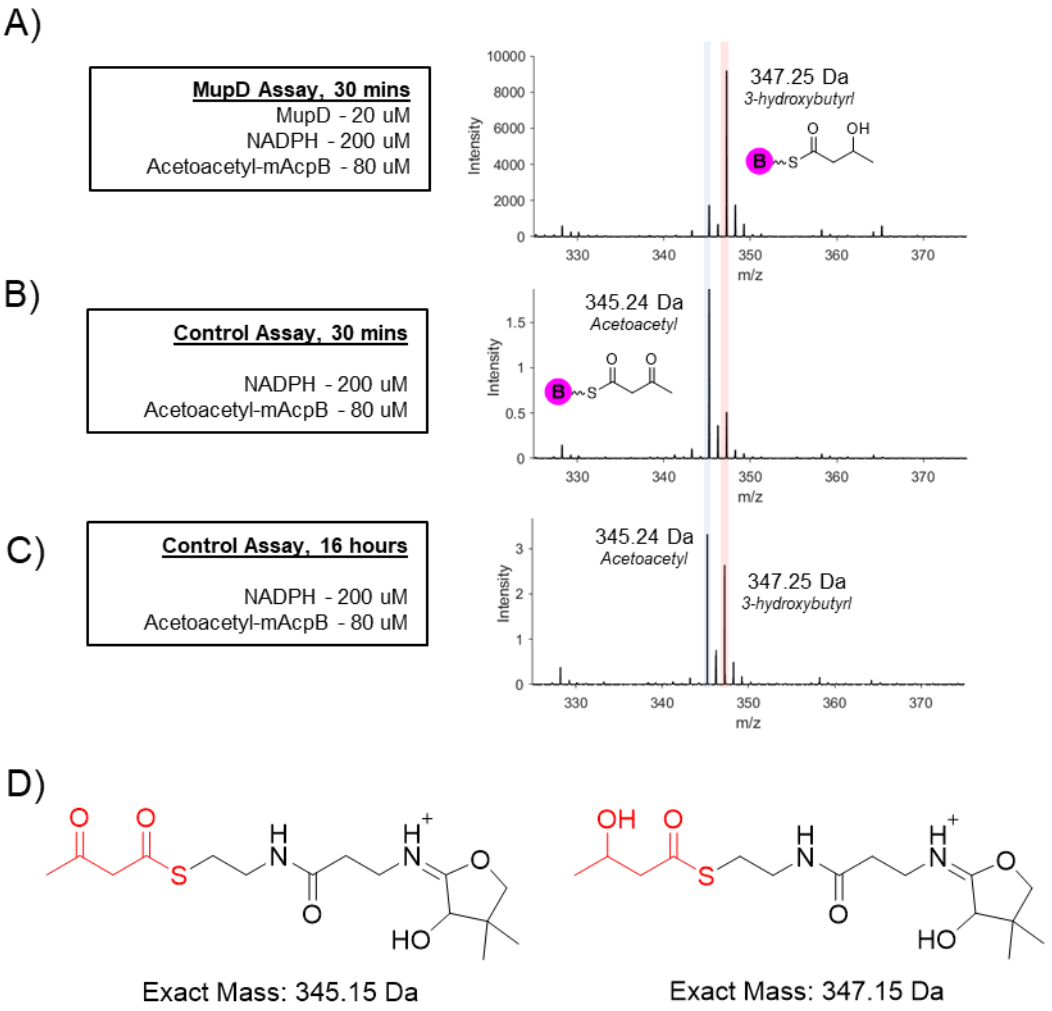


*Figure 58: The reduction of AcAc-mAcpB by ketoreductase enzyme MupD. A) Pantetheine ejection spectrum of the predominant acyl-mAcpB species 30 minutes after addition of MupD and NADPH. B) Pantetheine ejection spectrum of the predominant acyl-mAcpB species of the control reaction (lacking MupD) after 30 minutes. C) Pantetheine ejection spectrum of the predominant acyl-mAcpB species of the control reaction (lacking MupD) after 16 hours C) D) Chemical structures and expected masses of the Ppant ions of **21** and **22**.*

A control assay in the absence of MupD was initiated by addition of AcAc-mAcpB (80 µM) to NADPH (200 µM) at room temperature. After 30 minutes, a major acyl-ACP species with a mass of 11799.6 Da was observed (expected for acetoacetyl; 11796.8 Da). Ppant ejection produced ion of 345.24 Da derived from $\mathbf{21_B}$ only (Figure 58B). The 2-fold excess of NADPH present presumably generated ~15% of the reduced product $\mathbf{22_B}$ after 30 minutes versus almost complete conversion in the presence of MupD. After 16 hours, Ppant ejection assay showed a 55% to 45% mixture of $\mathbf{21_B}$ and $\mathbf{22_B}$ (Figure 58C). This is surprising but shows that reduction can take place in the absence of MupD. Despite this, the reduction is clearly accelerated by the addition of MupD confirming that MupD is responsible for the KR activity observed.

3.2.15.2 KR activity of MupD with acetoacetyl-mAcpD

The ACP selectivity of the MupD reaction was tested by incubation of **21$_D$** with MupD. The reaction was initiated by addition of **21$_D$** (80 µM) to NADPH (1 mM) and MupD (20 µM), incubated at room temperature and monitored by ESI-MS after 30 minutes and 4 hours. Pantetheine ejection of the major acyl-mAcpD species showed only unreacted **21$_D$** for both time points. After 4 hours the deconvolution of mAcpD showed acyl-mAcpD species with a mass of 15349.7 Da (expected; 15346.8 Da). Further fragmentation of this species gave a Ppant ion of acetoacetyl only (observed; 345.23 Da, expected; 345.15 Da, Figure 59B) consistent with no formation of **22$_D$**.

No sign of reduction was observed with **21$_D$**. This shows the important role played by mAcpB in turnover of the reaction. Monitoring the reaction by ESI-MS based methods do not allow for elucidation of the stereochemistry of the resulting alcohol produced. It is expected that a R-β-hydroxy product will be formed – this may be further tested using NMR analysis but is beyond the scope of this project.

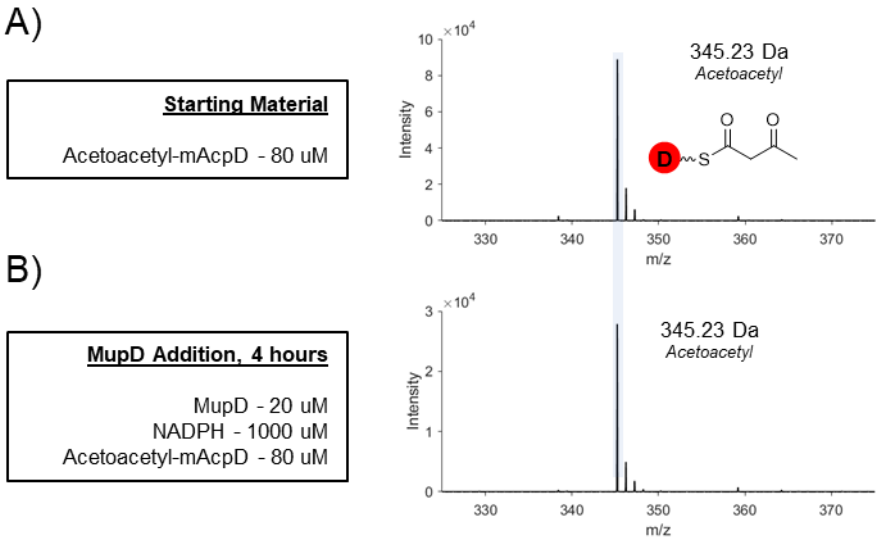
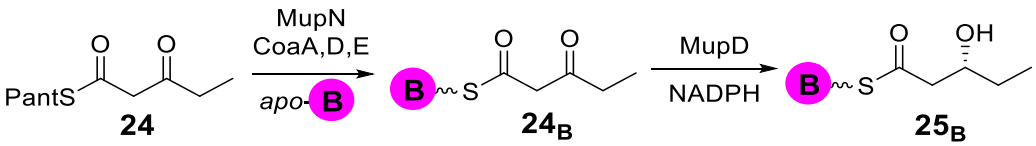


Figure 59: The addition of MupD and NADPH to AcAc-mAcpD. A) Pantetheine ejection spectrum of AcAc-mAcpD starting material. B) Pantetheine ejection spectrum of acyl-mAcpD species 4 hours after addition of MupD and NADPH. Unreacted starting material is observed in both cases (observed; 345.23 Da, expected; 345.15 Da).

3.2.16 Ketoreductase activity of MupD with 3-oxopentanoyl moiety



Scheme 18: The reaction scheme for loading 3-oxopent-pantetheine onto mAcpB. Reduction of the C-3 carbonyl will occur leading to the formation of the C-3 hydroxy substrate tethered to terminal thiol of mAcpB. The reduction is catalysed by MupD and NADPH cofactor.

The ability of MupD to catalyse the reduction of an elongated 5-carbon variant, 3-oxopentanoyl pantetheine (3-oxopent) **24** was tested (Scheme 18). **24** was synthesised by Angus Weir (unpublished work, Angus Weir, University of Bristol) and differs only from the natural substrate by lacking a C-5 hydroxyl group.

3-oxopent pantetheine **24** was transferred onto active site serine of mAcpB via the same protocol used for the formation of **21$_B$**. The assay was initiated by addition of MupN (5 µM) to ATP (1 mM), $MgCl_2$ (10 mM), **24** (1 mM), CoA mix enzymes (3 µM of each of CoaA, CoaD and CoaE) and *apo*-mAcpB (100 µM) and was monitored by ESI-MS after 2 hours. Again, two new species were observed with masses consistent with 3-oxopent-mAcpB (3OP-mAcpB) **24$_B$** (observed; 11810.0 Da, expected; 11810.7 Da) and *holo*-mAcpB (observed; 11712.1 Da, expected; 11712.7 Da, Figure 60A). The **24$_B$** was produced in 80% yield, with the remaining 20% consisting of *holo*-mAcpB. Fragmentation of the major species produced a characteristic Ppant ion for **24$_B$** (observed; 359.15 Da, expected; 359.16 Da, Figure 60A) confirming successful derivatisation.

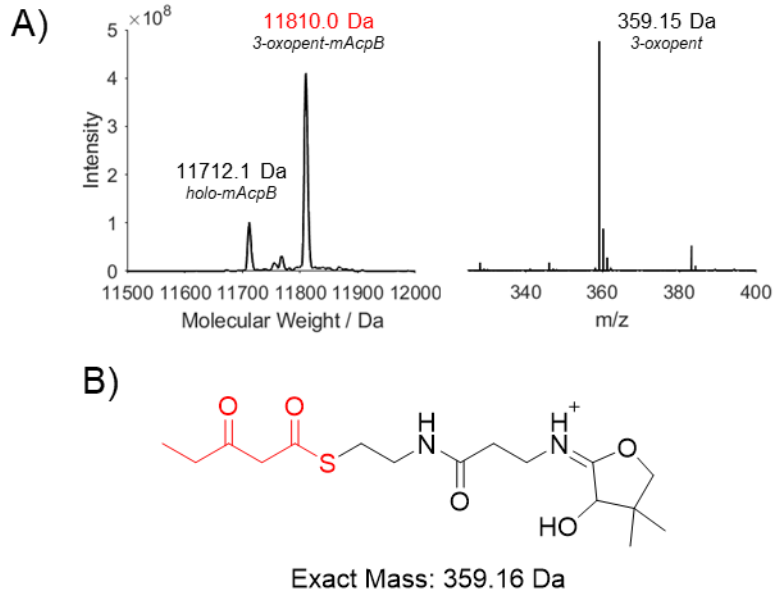


Figure 60: The formation of 3-oxopent-mAcpB. A) The deconvolution of mAcpB region of spectrum after 2 hours and subsequent pantetheine ejection assay of the predominant 3OP-mAcpB species. B) The chemical structure and expected mass of the 3-oxopent-mAcpB Ppant ion.

Reduction was assayed by addition of MupD (20 µM) to **24$_B$** (100 µM) in presence of NADPH (1 mM) as monitored by ESI-MS. After 1 hour, a new species with a mass consistent with the expected 3-hydroxypent-mAcpB **25$_B$** product was produced (observed; 11812.5 Da, expected; 11812.8 Da, Figure 61). This was confirmed by pantetheine ejection to produce the characteristic ion with a mass of 361.16 Da (expected mass; 361.18 Da).

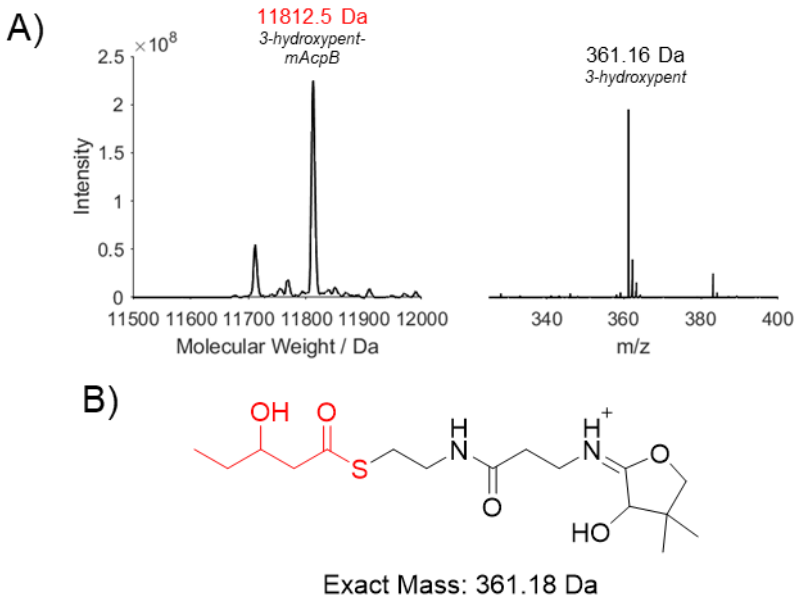
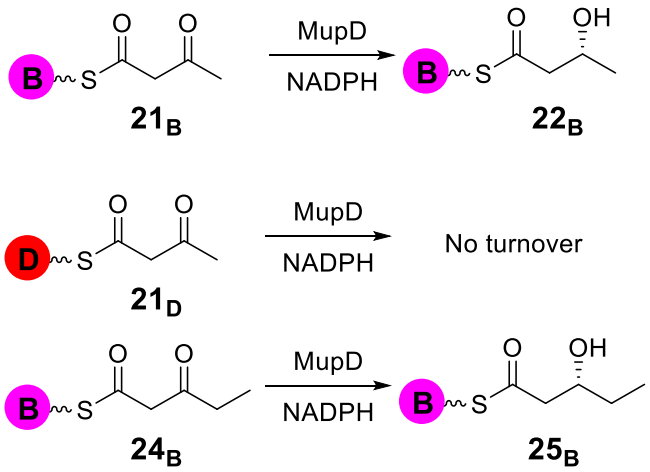


*Figure 61: The reduction of 3-oxopent-mAcpB **24B** by the KR enzyme MupD. A) The deconvolution of mAcpB species 1 hour after addition of MupD to 3-oxopent-mAcpB (left) and subsequent Ppant ejection ion (right). B) The chemical structure and expected mass of the Ppant ion of 3-hydroxypent-mAcpB **25B**.*

These assays with mimics of the natural substrate of MupD show that MupD can accept multiple substrates when attached to mAcpB. The inability of mAcpD to turnover with MupD confirms that specific interactions between mAcpB and MupD facilitate the reaction. The interaction of KR and ACPs has previously been shown to be multivalent, consisting of distinct contributions from both the ACP and the acyl group undergoing reduction. [147] This work confirms the important role played by the ACP in this interaction.



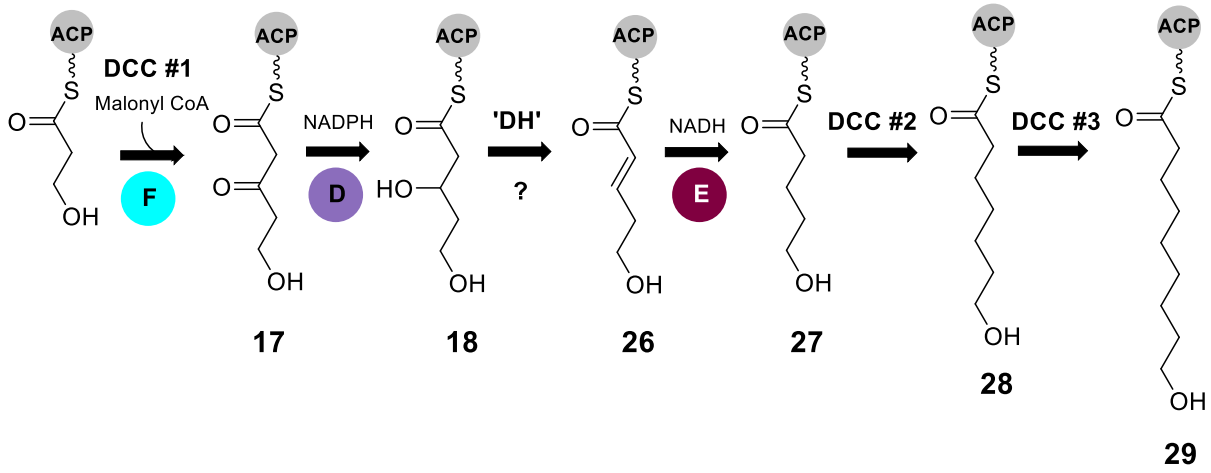
Scheme 19: Scheme showing the outcome of assays involving MupD conducted in this section.

When considered in the context of 9HN biosynthesis, assays with MupD and the proposed natural substrate were found to be successful when delivered by both mAcpB and mAcpA. This provides further evidence that both ACPs are involved in delivering malonyl during the first condensation

reaction. The ability of each ACP to deliver malonyl during subsequent condensations steps will be tested in later Chapters. The work in this chapter has provided useful information about the proposed sequence of events in the early part of 9HN biosynthesis.

3.3 Reductive Tailoring Following MmpF-Catalysed Condensation

During polyketide and fatty acid biosynthesis, tailoring modifications following KS-catalysed condensation are responsible for conversion of the diketone moiety to acyl moieties with varying degrees of saturation. These tailoring reactions are catalysed by three enzymes; the KR, DH and ER. This section will focus on understanding the enzymes responsible for ER and DH tailoring during the biosynthesis of 9HN. The proposed tailoring steps are shown in Scheme 20. MmpF catalysed condensation (DCC #1) will produce a 5-hydroxy β-ketothiolester intermediate **17**. Work in previous chapters has shown that MupD will provide the initial KR tailoring activity to produce 3,5-dihydroxythiolester **18** but further tailoring steps are yet to be demonstrated experimentally. **18** is expected to undergo dehydration to produce **26** catalysed by a yet unknown DH enzyme. Following this, reduction catalysed by MupE is expected to yield the saturated 5-hydroxy thiolester intermediate **27**, but this is yet to be proven experimentally.[48] **27** will then undergo two further condensation and tailoring cycles to produce 7-hydroxy thiolester **28** and 9-hydroxythiolester **29** utilising either MmpF for second elongation step followed by MmpB_KS for the final step or the MmpB_KS for both steps.

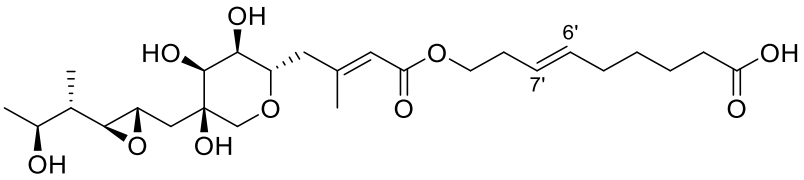


*Scheme 20: Simplified reaction scheme showing the proposed biosynthesis of 9HN. F, D and E represent the proteins MmpF, MupD and MupE respectively. MmpF catalyses the first chain extending reaction (DCC #1) to produce diketone intermediate **17**. In vitro assays have confirmed that MupD is KR enzyme which catalyses reduction of **17** to dihydroxy intermediate **18**. The next step is proposed to be dehydration of **18** to produce enone intermediate **26** and subsequent reduction of this enone to produce fully saturated 5-carbon fatty acid **27**.*

In this chapter, the ER and DH tailoring reactions were individually analysed and experimentally verified utilising simple mimics of the expected natural substrates. The use of simple ketothiolester mimics of this natural substrate attached to cognate and non-cognate ACPs will initially allow the ACP specificity of these reactions to be tested. MupE was studied first as this was known as a discrete ER enzyme within mupirocin BGC. No equivalent *trans*-acting DH enzyme had yet been identified and this will be addressed later.

3.3.1 Bioinformatic Analysis of MupE

MupE is a further discrete enzyme whose function in the mupirocin BGC has not yet been defined. Limited functional analysis concluded this protein to have enoyl reductase activity. [78,48] This was consistent with accumulation of Mupirocin E with a double bond between C-6' and C-7' when *mupE* gene was knocked out in *P. fluorescens* (Figure 62).[81] The presence of only a single point of desaturation on this knockout would suggest MupE acts on **26** only, requiring the action of an additional ER domain. As MmpB lacks a modular ER domain, this activity may be derived from either MmpC, another biosynthetic pathway or primary metabolism. Further subsequent gene knockouts have also shown, somewhat confusingly, that Mupirocin E can also be isolated from *mupD* (KR) knockout experiments. [148] As with most of the mupirocin pathway, the genetic approach sheds little light on the precise action and timing of these enzymes beyond their broad involvement in fatty acid chain biosynthesis.



Mupirocin E

Figure 62: Chemical structure of gene knockout product mupirocin E. This compound contains C-6' to C-7' alkene functionality and has been shown to accumulate after knockout of MupE gene within biosynthesis of mupirocin.

MupE is a member of the Medium-Chain dehydrogenase (MDR) superfamily of proteins. [149] This superfamily has been shown to encompass a plethora of proteins with a broad range of functions and reactions.[150] Furthermore, the sequence of MupE contains structural features which is indicative of Zinc-alcohol dehydrogenase-like (Zn-ADH-like) domain. MupE contains a conserved series of cysteine residues (C_{91}-C_{94}-C_{97}-C_{105}) implicated in the binding of structural zinc. This zinc-binding site has also been observed in zinc-dependent alcohol dehydrogenase from *Thermus thermophilus* HB8, the structure of which has been solved (PDB accession: 2EIH). MupE shares 38% identity with this protein and the structure of MupE has been modelled using this as a template.

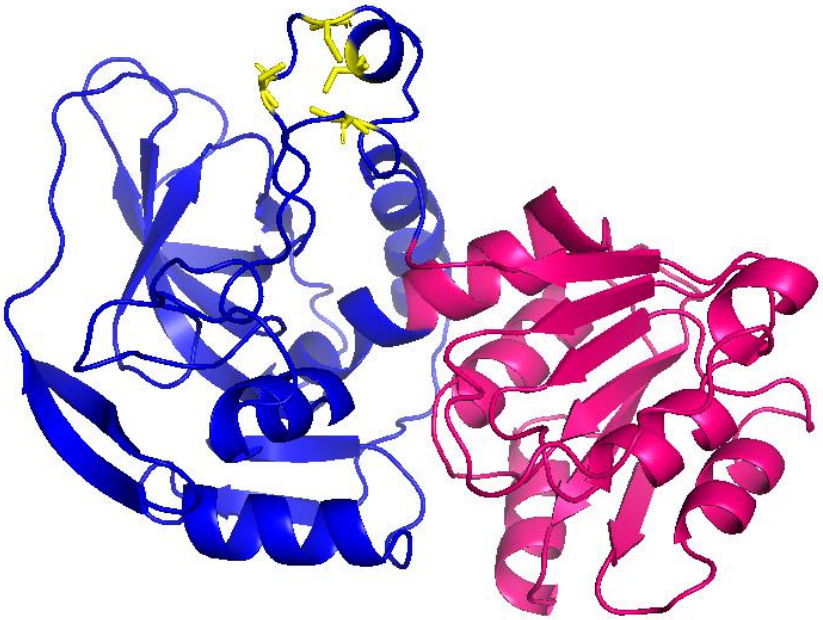
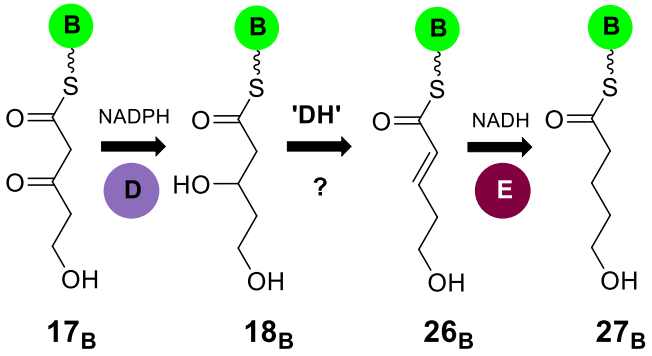


Figure 63: The structure of MupE protein modelled from resemblance to Zn-Dependent Alcohol Dehydrogenase from Thermus Thermophilus HS8 (38% identity, PDB Accession 2EIH). The structure consists of catalytic region (shown in blue) containing binding site for catalytic zinc (shown in yellow) and a NAD(P)H binding site (shown in pink).

The modelled structure of MupE consists of two distinct regions – the first is the catalytic domain. This domain is responsible for binding and modification of substrates. This is shown in blue in Figure 63 with the binding site of structural zinc shown in yellow. The region from residue 151 to 288 is responsible for NAD(P)H binding and is shown in pink. Bioinformatic analysis of homologues of MupE has shown that MupE is likely to show specificity for NADH rather than NADPH. [151] Sequence analysis of MupE was unable to definitively confirm the function of this protein. Further analysis of residue 1-150 of the catalytic domain identified homologues consisting of alcohol dehydrogenase enzymes such as L-threonine 3-dehydrogenase from *Thermus thermphilus* HB27 [152] and 2-deoxy-scyllo-inosamine (DOIA) dehydrogenase from *Streptomyces rimosus subsp. paromomycinus* – both surprisingly responsible for the interconversion of alcohols to ketone. BLAST search of MupE showed similarity to both alcohol dehydrogenase enzymes and PKS enoyl reductase enzymes.

Sequence analysis of members of the MDR superfamily was conducted by Pliess *et al* in 2008.[153] This analysis was able to group each MDR member into families based on sequence identity and shared structural features. MupE was included in this analysis and was placed in family MDR 20. None of the members of MDR 20 have been characterised by structure determination or functional studies. MupE is a member of the MDR 20.3 subclass, of which MupE is the only member so limited information can be attained about its proposed function within the mupirocin biosynthetic gene cluster.

Typical type II FAS Enoyl-ACP reductase enzymes are members of the Short-Chain Dehydrogenase (SDR) family.[18] During fatty acid construction in *E. coli*, ER activity is provided by FabI domain.[154] Analysis of the *trans*-ER domain from the iterative type I PKS used to form lovastatin (LovC) shows this is also a member of the MDR superfamily. Therefore, a sensible hypothesis based on sequence homology and functional data is that MupE should act on **26$_B$** arising from the first MmpF catalysed decarboxylative condensation (Scheme 21)



*Scheme 21: Reaction scheme showing the tailoring of 5-carbon thiolester intermediate **17$_B$** with MupD, and unknown DH and MupE. **17$_B$** will be reduced by proposed KR enzyme MupD to produce dihydroxy intermediate **18$_B$**. Following this **18$_B$** undergoes dehydration catalysed by an unknown DH domain to produce enone intermediate **26$_B$**. **26$_B$** will be further reduced to form 5-carbon fully saturated acyl chain **27$_B$** catalysed by ER domain MupE.*

3.3.2 Expression and purification of MupE

To test this function *in vitro*, MupE was successfully expressed utilising the same arctic growth conditions that yielded functional MupD. The MupE gene was cloned by PCR using primers given in the methods section. This gene was subsequently ligated into pOPINF ready for expression. MupE was purified by IMAC followed by SEC using a S100 column and was successful in producing 1.5 mg/L of soluble protein as confirmed by SDS-PAGE analysis (Figure 47A). ESI-MS of MupE showed a single series of charge states with a mass consistent with MupE (observed; 38591 Da, expected; 38596 Da) (Figure 47B). An analytical SEC trace of MupE (conducted with S75 column) indicated a monomeric state in solution with an elution volume of 14.5 mL consistent with a predicted molecular weight of 50.3 kDa. This monomeric state was consistent with other MDR superfamily ER domains from *trans*-AT biosynthetic pathways such as LovC from the biosynthesis of lovastatin. [155,156]

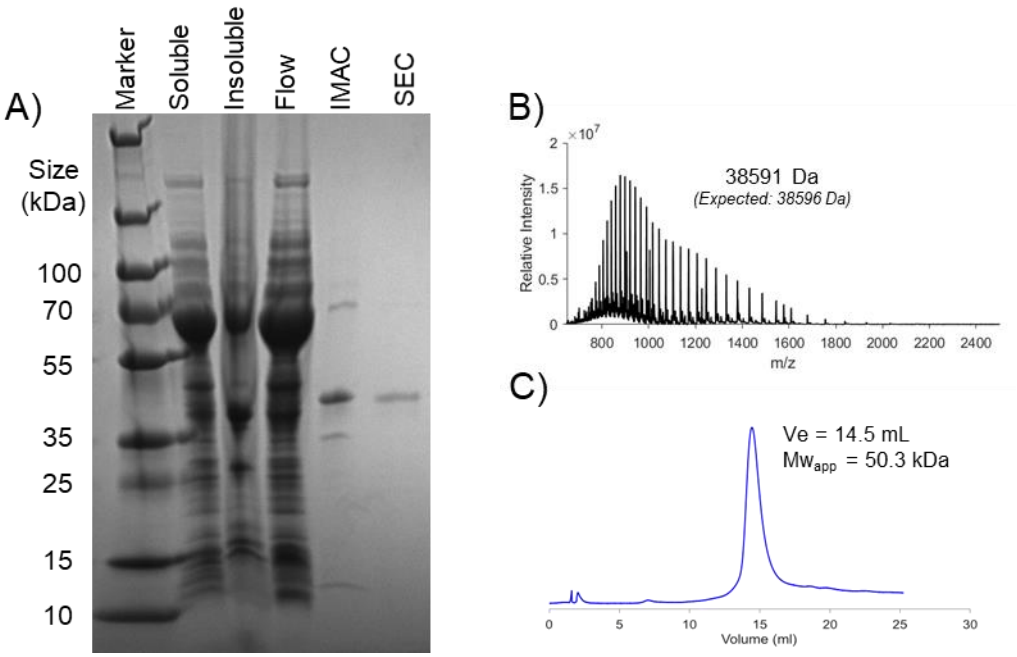


Figure 64: Expression and purification of MupE A) SDS-PAGE gel showing the different stages of MupE purification using Arctic cells. B) ESI-MS spectrum of purified MupE. A single series of charge states showed a mass consistent with MupE (observed; 38591 Da, expected; 38596 Da). C) Analytical size exclusion chromatography trace of MupE.

3.3.3 MupE activity with natural substrate mimics

The activity of MupE was tested by incubation with a series of acyl-pantetheine substrates containing the enone functional group. The substrates crotonyl-pantetheine (crot-pant) **30** and 3-methylbutenoyl-pantetheine (3mb-pant) **31** were synthesised by Angus Weir for use in *in vitro* assays. **30** contained a 4-carbon backbone compared to the 5-carbon backbone and C-5 hydroxyl group of the proposed natural substrate (Figure 65). Substrate **31** also possessed a 4-carbon backbone but contained a β-methyl branch which differentiated it from **30**. The strategy would be to transfer **30** and **31** to mAcpB (to produce $\mathbf{30_B}$ and $\mathbf{31_B}$) and incubate with MupE in an *in-vitro* reconstitution of the proposed biosynthetic pathway.

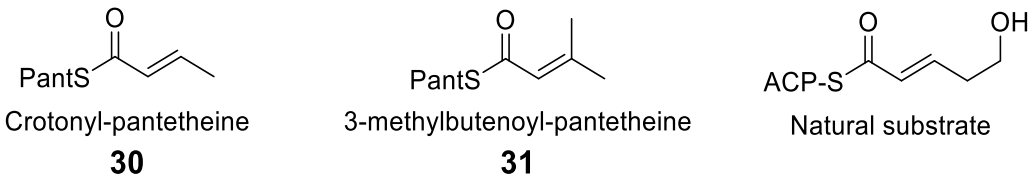
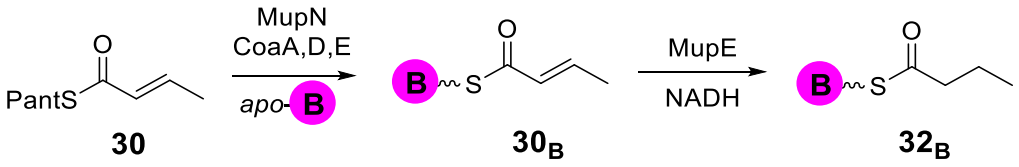


*Figure 65: Synthesised pantetheine analogues **28** and **29** used to test the function of MupE enzyme compared to the proposed natural substrate.*

3.3.3.1 MupE activity with crotonyl-mAcpB

Functional assays were initially conducted with MupE and **30**. The substrate of MupE is expected to be delivered by either mAcpB or mAcpA. In this work, the crotonyl substrate was transferred to the terminal thiol of mAcpB. If MupE is the expected ER enzyme, reduction of crot-mAcpB is expected to occur as shown in Scheme 22.



*Scheme 22: The reaction scheme showing the loading of crotonyl-pant **30** on to mAcpB to produce **30B**. This will be followed by incubation with MupE to produce butyryl-mAcpB **32B**.*

First, crotonyl-pantetheine was loaded onto mAcpB utilising MupN and cofactors. The assay was initiated by addition of MupN (5 µM) to ATP (1 mM), $MgCl_2$ (10 mM), **30** (1 mM), CoA mix enzymes (3 µM of each of CoaA, CoaD and CoaE) and *apo*-mAcpB (100 µM). The reaction was monitored by ESI-MS after 1 hour at room temperature. After this time, *apo*-mAcpB was consumed and a new species with a mass consistent with **30B** (observed; 11780.2 Da, expected; 11780.7 Da) was observed (Figure 66A). The major mAcpB species observed was produced in 100% yield and further Ppant ejection confirmed successful formation of **30B** (observed; 329.14 Da, expected; 329.15 Da).

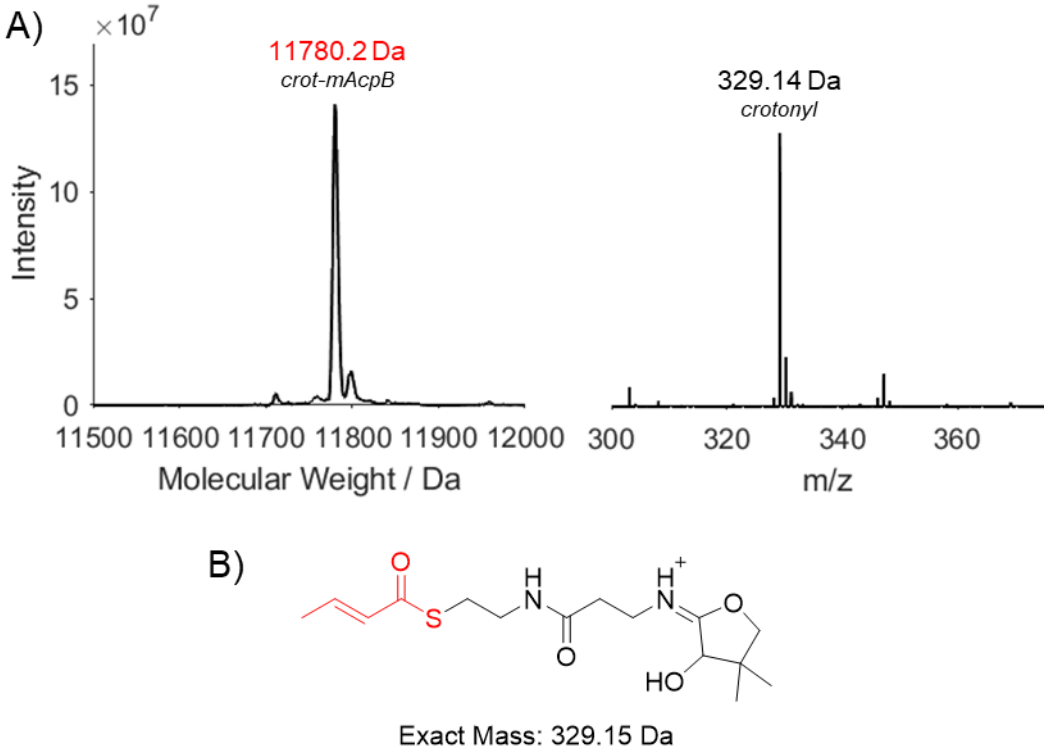


*Figure 66: The formation of crotonyl-mAcpB utilising MupN and crot-pant **28** after 1 hour. A) Deconvolution of the mAcpB region of the spectrum (left) and subsequent Pantetheine ejection of major species (right). B) Chemical structure and expected mass of the Ppant ion of crot-mAcpB.*

The reduction of crotonyl-mAcpB **30$_B$** by MupE was assayed by addition of MupE (20 µM) to **30$_B$** (80 µM) and NADH (200 µM) and monitored by ESI-MS over 30 minute, 2 hour and 16 hour timepoints. After 30 minutes, a single species with a mass of 11784.0 Da was observed. Pantetheine ejection of this species showed a mixture of 75% crotonyl (observed; 329.23 Da, expected; 329.15 Da) and 25% butyryl derivatives (observed; 331.25 Da, expected; 331.17 Da) consistent with partial reduction (Figure 67A). After 2 hours, the mixture percentage increased to 60% butyryl (observed; 331.25 Da) with 40% crotonyl (observed; 329.24 Da) as shown in Figure 67B. After 16 hours, however, the main species showed a mass of 11785.0 Da (Figure 68C). Further fragmentation of this species produced a single Ppant ion with a mass consistent with only the butyryl species (observed; 331.25 Da). The evolution of the Ppant ejection spectrum over time is shown in Figure 67. This confirms that MupE can reduce **30$_B$** to the corresponding **32$_B$** moiety *in vitro*. Although kinetic analysis was not conducted, it can be concluded that ER activity catalysed by MupE is slower than KR activity catalysed by MupD due to complete reduction observed after 16 hours. This is slower than the 30 minutes required to completely reduce the acetoacetyl moiety by MupD.

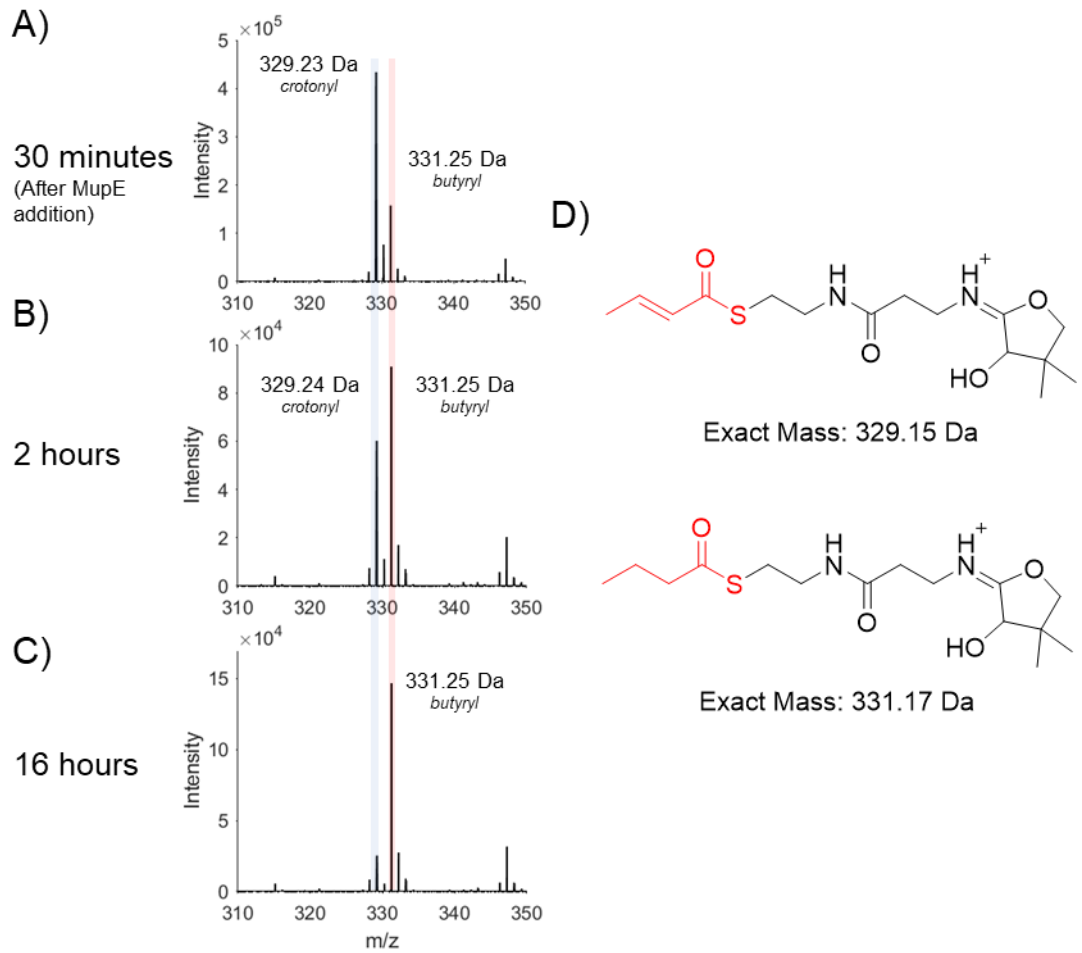


Figure 67: The timepoints of the reduction of crot-mAcpB catalysed by MupE. A) Pantetheine ejection assay of major mAcpB species 30 minutes after addition of MupE to crot-mAcpB. B) Pantetheine ejection assay of major mAcpB species 2 hours after addition of MupE to crot-mAcpB. C) Pantetheine ejection assay of major mAcpB species 16 hours after addition of MupE to crot-mAcpB. D) The chemical structures and expected masses of the crotonyl and butyryl pantetheine ejection ions.

To confirm that the ER activity observed was due to the action of MupE, a negative control assay was conducted. In this assay, **30$_B$** (80 μM) was added to NADH (200 μM) in the absence of MupE. The reaction was conducted concurrently with a successful MupE reduction assay and was monitored by ESI-MS over time. After 16 hours, mAcpB deconvolution showed a single series of charge states which was consistent with unreacted **30$_B$** (observed; 11783.2 Da, expected; 11780.7 Da, Figure 68B). Further fragmentation of the 11783.2 Da species from the 16 hour control assay (lacking MupE) produced a single Ppant ion consistent with no reduction having taken place (observed; 329.23 Da, expected; 329.15 Da).

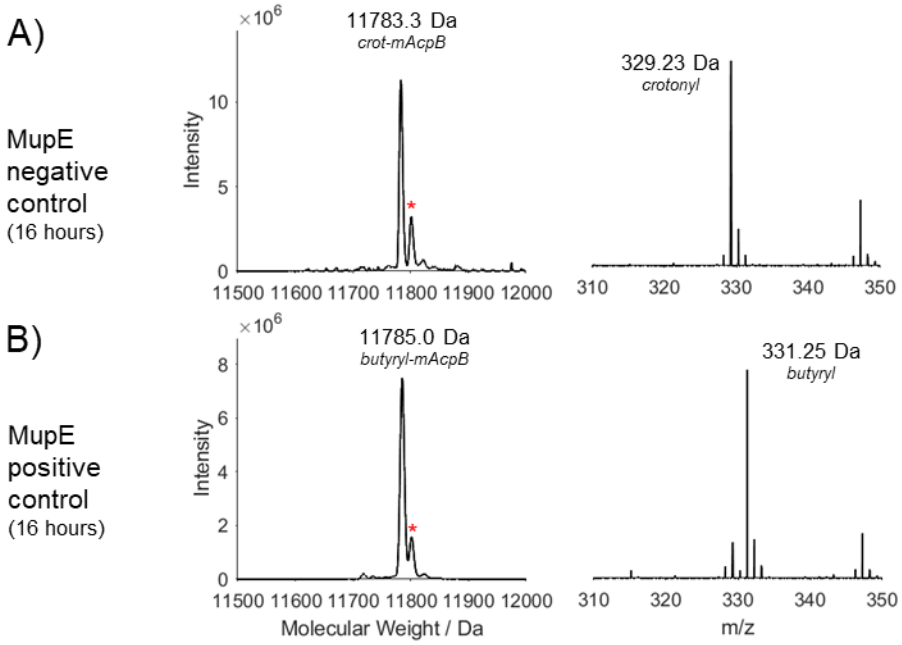
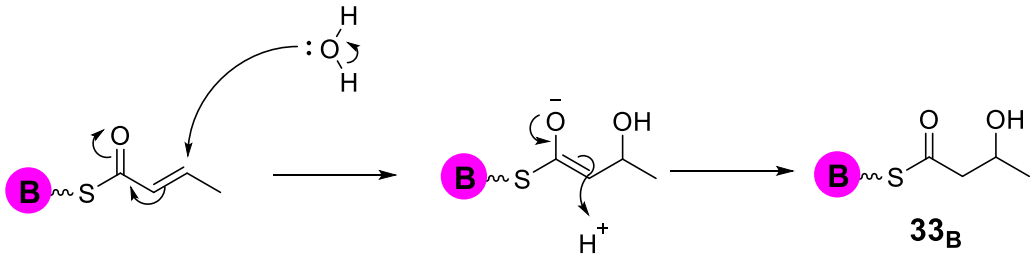


*Figure 68: Summary of the MupE + crotonyl-mAcpB assays. A) The deconvolution and subsequent pantetheine ejection ion of the MupE negative control reaction after 16 hours. **33$_B$** species is asterisked. B) The deconvolution and subsequent pantetheine ejection ion of the MupE negative control reaction (containing only crot-mAcpB and NADH) after 16 hours.*

A small amount of a species with a mass of 11801.2 Da was observed in the deconvolution of both the MupE positive control and the MupE negative control reactions after 16 hours. This corresponds to a mass increase of approximately + 86 Da from the observed mass of *holo*-mAcpB (Figure 68A and B). This species did not give an ejection ion, presumably due to low abundance of the parent species. This species may be formed by the 1,4 addition of water to crotonyl species which will produce the 4-carbon β-hydroxy **33$_B$** species shown in Scheme 23. The formation of this species is greatest for the negative

control reaction where the substrate persists for longer periods of time. The formation of **33$_B$** in small amounts means it is not likely to effect the MupE catalysed reduction.



Scheme 23: The side reaction of crot-mAcpB which produces the β-hydroxy thioloester intermediate shown.

3.3.3.2 MupE activity with crotonyl-mAcpD

To test the effect of changing the ACP on the rate of reaction, **30** was loaded on to the terminal thiol of non-cognate mAcpD to produce **30$_D$**. The reaction was initiated by addition of MupN (5 µM) to ATP (1 mM), $MgCl_2$ (10 mM), **30** (1 mM), CoA mix enzymes (3 µM of each of CoaA, CoaD and CoaE) and *apo*-mAcpD (100 µM). After 2 hours, 100% formation of **30$_D$** was confirmed by ESI-MS (observed; 15334.4 Da, expected; 15330.8 Da, Figure 69A) and Ppant ejection (observed; 329.24 Da, expected; 329.15 Da).

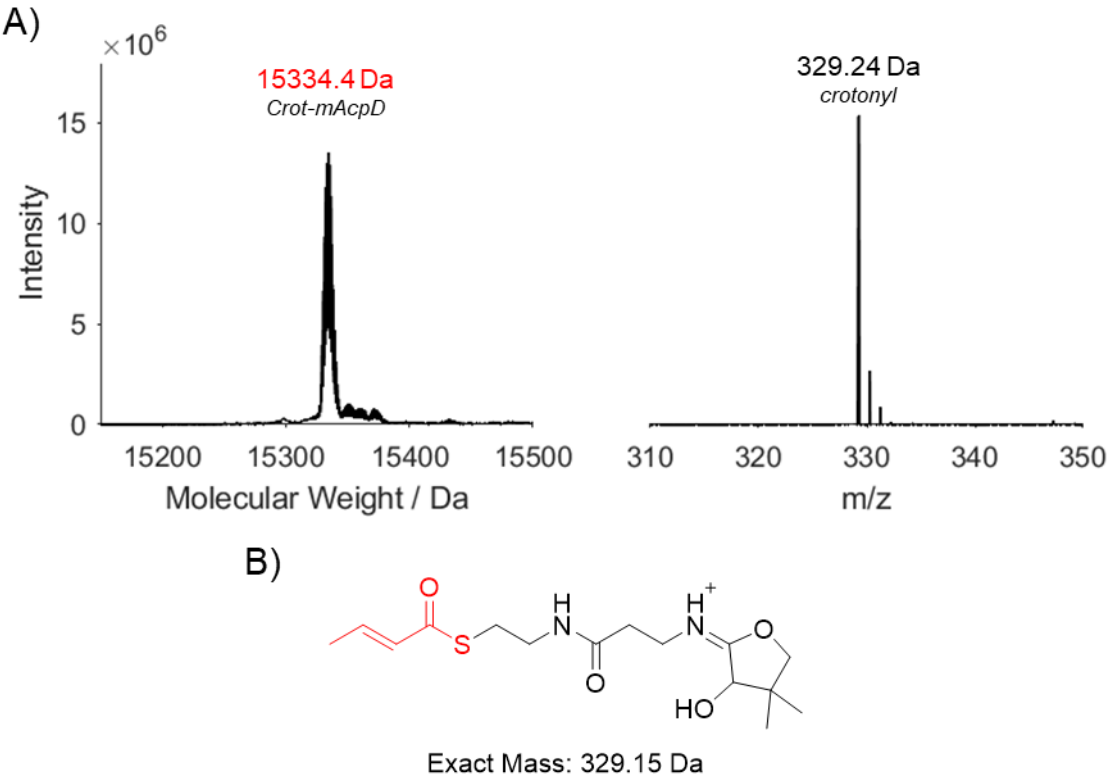


Figure 69: The formation of crotonyl-mAcpD utilising MupN and CoaMix enzymes. A) After 2 hours, deconvolution of the mAcpD region of the spectrum and Ppant ejection assay of predominant species. B) Chemical structure and expected mass of the crot-mAcpD ejection ion.

The reduction of **30$_D$** was assayed by addition of MupE (20 µM) to NADH (200 µM) and **30$_D$** (80 µM). After 16 hours, a species with a mass of 15334.0 Da was observed. Ppant ejection of this species produced a mixture of pantetheine ejection ions (Figure 70B) that consisted of ~ 60% of the crot-mAcpD ion (observed; 329.24 Da, expected; 329.15 Da) and ~ 30% of ions derived from butyryl-mAcpD (observed; 331.25 Da, expected; 331.17 Da). This was consistent with partial reduction and therefore some ability of MupE to recognise this ACP.

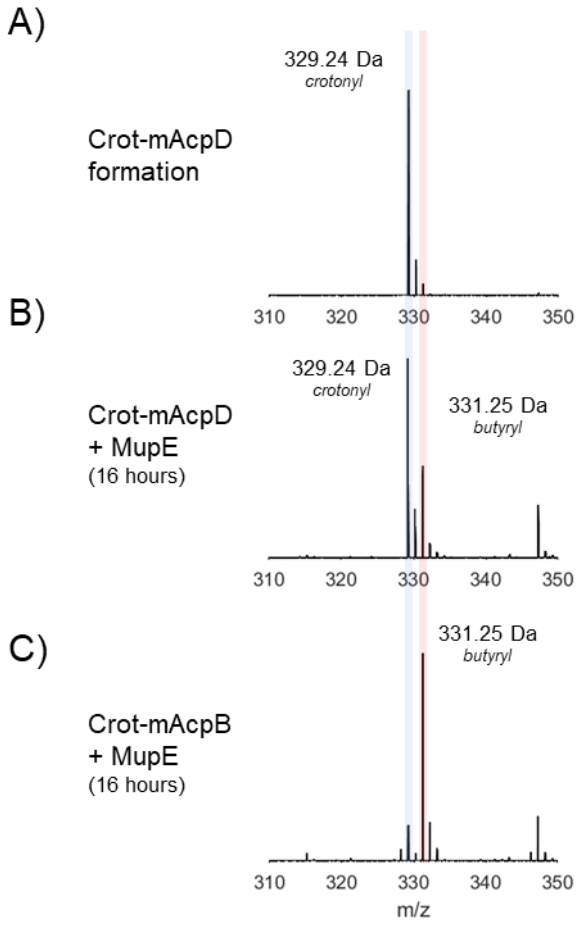


Figure 70: The pantetheine ejection spectra of crot-mAcpD + MupE assay. A) Pantetheine ejection ion of crot-mAcpD only. B) Fragmentation of the major species 16 hours after addition of MupE and NADH to crot-mAcpD. C) Fragmentation of the major species 16 hours after addition of MupE and NADH to crot-mAcpD.

The observation that MupE catalyses complete reduction of crot-mAcpB after 16 hours, but only partial reduction of mAcpD after the same period of time confirms the reduction of **30** is accelerated by protein-protein interactions between mAcpB and MupE. For the equivalent assay involving MupD and non-cognate Acac-mAcpD no reduction was observed after 16 hours. For both MupD and MupE catalysed reductions, mAcpD bound substrates show a slower rate of reaction compared to mAcpB bound substrates. This is because mAcpD lacks the specific protein-protein interactions that exist between mAcpB and its cognate partners MupD and MupE.

3.3.3.3 MupE assays with β-branched enone

The ability of MupE to catalyse reduction of β-branched enone substrate was subsequently tested. The purpose of this assay was to test the specificity of MupE to turnover with a substrate containing a non-natural β-methyl branch. The substrate 3mb-pant **31** was loaded on to mAcpB to produce **31$_B$** utilising previous conditions involving MupN. After 1 hour, *apo*-mAcpB was consumed and a new species with mass consistent with **31$_B$** was produced in 100% yield (observed; 11797.5 Da, expected; 11794.7 Da, Figure 71A). Further fragmentation of this species gave a Ppant ion of 343.26 Da confirming the formation of **31$_B$** (expected; 343.17 Da).

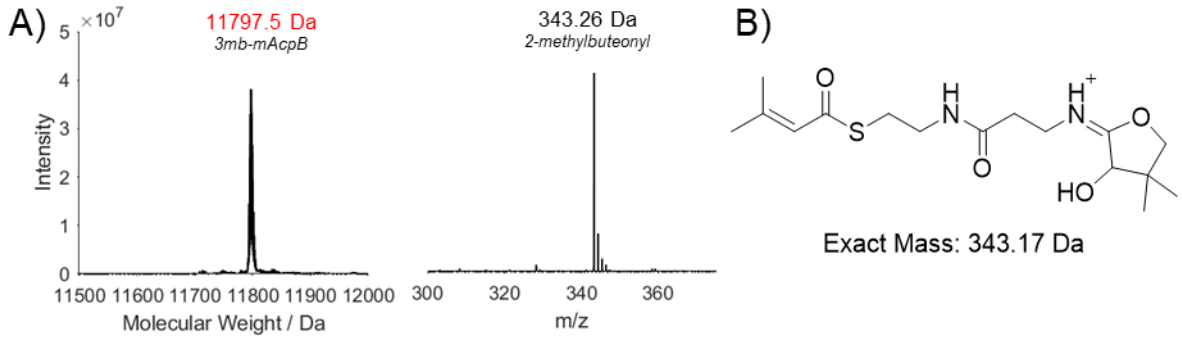


*Figure 71: The formation of **31$_B$** utilising MupN and cofactors. A) Deconvolution of the mAcpB region of the spectrum 1 hour after initiation reaction (left) and subsequent Ppant ejection ion (middle). B) Chemical structure and expected mass of the Ppant ion of **31$_B$***

The reduction of **31$_B$** was initiated by addition of **31$_B$** (80 μM) to NADH (1 mM) and MupE (20 μM) and monitored after 30 minutes, 4 hour and 16 hours incubation at room temperature. After all time points, analysis by ESI-MS showed no changes to the mAcpB region of the spectrum consistent with no reduction having taken place. After 16 hours, the main species had a mass consistent with **31$_B$** (observed; 11797.5 Da, expected; 11794.7 Da) as shown in Figure 72A. Further fragmentation of the main species gave a Ppant ion consistent with no reduction having taken place (observed; 343.26 Da, expected for non-reduced; 343.17 Da, expected for reduced; 345.18 Da, Figure 72B).

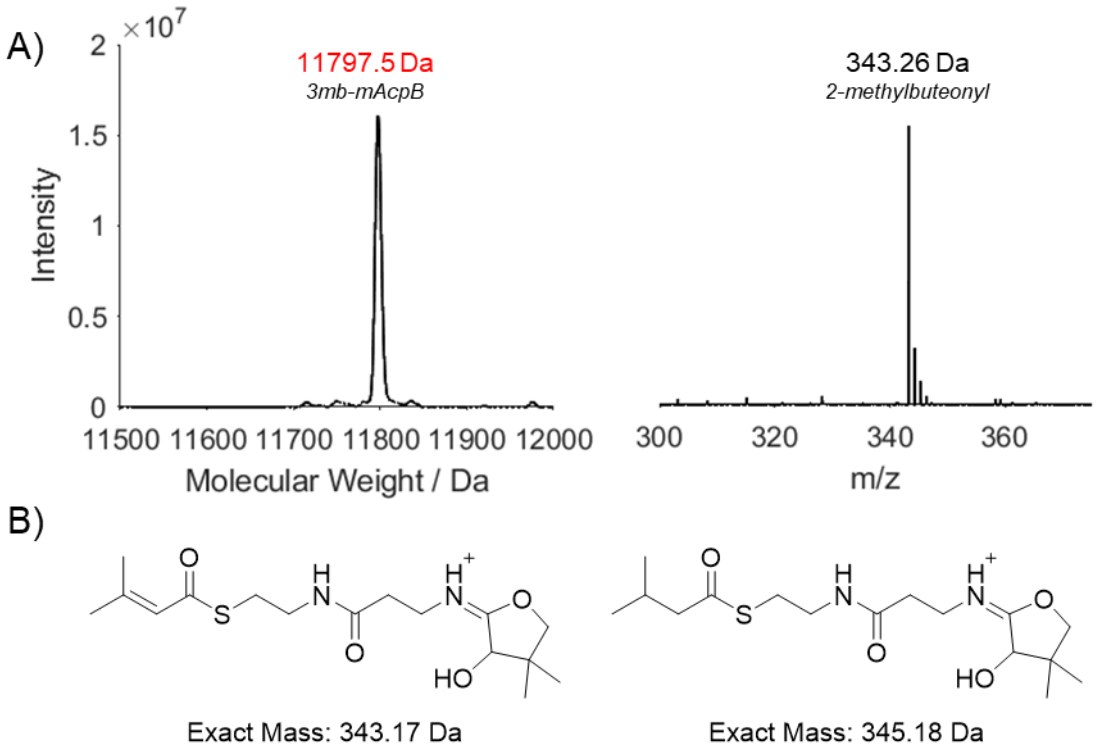


Figure 72: The addition of MupE and NADH to 3mb-mAcpB after 16 hours. A) Deconvolution of the mAcpB region of the spectrum and subsequent Ppant ejection assay of the major species. B) Chemical structures and expected masses of 3mb-mAcpB Ppant ion and reduced intermediate.

The data observed shows that MupE is not able to reduce **31$_B$** *in vitro*. The methyl group at the β-position presents steric bulk located on the double bond of the enone undergoing reduction may disrupt the binding of the substrate to the active site or the catalysis. The exact position and nature of binding interactions are unknown due to the absence of structural characterisation of MupE. Literature analysis of PKS ER enzymes shows that during typical ER mechanism of action, the β-position is positioned close to the C-4 of the nicotinamide ring of NADH to allow hydride transfer to take place. [156] This hydride transfer is key to reduction. It is proposed that the β-methyl group present on **31** disrupts this key interaction, thereby stopping hydride transfer (and subsequent reduction) from taking place.

3.3.4 Identity of DH tailoring enzyme

The DH enzyme from the MmpF elongation, reduction cycle is currently unknown as no *trans*-acting DH enzymes exist in the BGC of mupirocin. Initially reanalysis of the putative functions of the proteins in the mupirocin BGC was conducted. This confirmed that a total of four DH domains were present in the total mupirocin BGC. Three of these DH are located within MmpD – the multidomain PKS involved in the synthesis of the polyketide backbone of the monic acid moiety. The final DH domain was located within MmpB. The feasibility of each of these DH domains to act with MupD and MupE in the tailoring of MmpF elongated intermediates will be discussed.

Like mupirocin, another *trans*-AT PKS whose biosynthetic pathway is lacking a DH domain is iso-migrastatin. [157] Work by Shen *et al*. used SNAC mimics to show that the missing DH activity within the Migrastatin cluster was provided by a dual acting DH domain (MsgF_DH10). This unique DH domain was found to provide activity both within its associated module (module 10) and also so called 'long-range activity' to an upstream module (module 4). [158] This long-range activity was found to be contingent on an unusual 23 AA segment located within the DH motif (HXXXGXXXXP motif) of this enzyme. A mutant of MsgF_DH10 lacking this 23 AA sequence was no longer able to act in a long-range manner. The authors hope that future mechanistic studies will shed a light on why this segment is crucial to long-range activity.

Of the DH domains found within MmpD, MmpD_DH2 and MmpD_DH3 are both proposed to act to produce enone products during monic acid construction. If these DH domains were to act in a long-range manner, they would be expected to contain the 23 AA segment present in MsgF_DH10. Therefore, sequence alignment of *trans*-AT PKS DH domains including MsgF_DH10, MmpB_DH, MmpD_DH1, MmpD_DH2 and MmpD_DH3 was conducted. This revealed that none of the mupirocin DH domains contain this additional motif (Figure 73) and therefore on this basis, MmpD_DH2 and MmpD_DH3, which are known to have a defined role in the polyketide processing, are not likely to possess a further dual activity.

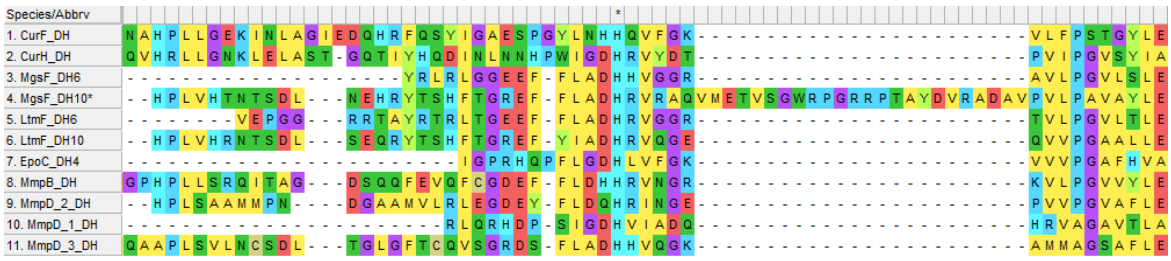


Figure 73: Alignment of selected trans-AT PKS DH domains including MsgF_DH10, MmpB_DH, MmpD_DH1, MmpD_DH2 and MmpD_DH3. MsgF_DH10 contains a 23 amino acid dual action motif which is lacking in all mupirocin DH domains.

The remaining candidates for DH tailoring activity are MmpD_DH1 (from MmpD module 1) and MmpB_DH. Interestingly, module 1 of MmpD is not expected to require DH activity due to the colinearity rule predicting a β-hydroxy intermediate. This implies that MmpD_DH1 may act elsewhere in the pathway. Hotherstall et al. predicted that D_DH1 may be inactive due to lacking Glycine and Proline in the DH motif.[148] Analysis of MmpD_DH3 shows this also lacks the terminal Proline in the DH motif despite having a proposed function in the pathway. Further to this, functional studies of DH enzymes have in fact shown that only the histidine residue within the DH motif is required for catalytic activity (playing direct role in dehydration of substrate) [159] but that the rest of the DH motif is not always universally conserved. [160] Another important residue for function in DH enzymes is the catalytic aspartic acid residue located within the DXXXQ/H motif. MmpD_DH1 contains all residues of this DXXXQ/H motif (including catalytic aspartic acid) and is therefore expected to be functional and may be a candidate for missing DH activity.

The final domain, MmpB_DH, is also a candidate for missing DH activity. This domain contains both the DH and DXXXQ/H motifs which predict active site residues. Furthermore, MmpB_DH is expected to be directly involved in 9HN construction due to its inclusion within MmpB.

3.3.4.1 Expression and Purification of DH enzymes

The two DH domain candidates (MmpB_DH and MmpD_DH1) were expressed and purified. Gene boundaries were confirmed using the BLAST database and secondary structure prediction using I-TASSER server. Comparison of MmpB_DH and MmpD_DH1 gene fragments from MmpB and MmpD respectively was translated into protein sequences and compared to typical DH enzymes from BLAST database. The sequences of MmpB_DH and MmpD_DH1 are outlined in the methods section.

MmpD_DH1 was cloned from genomic DNA of *P. fluorescens* using specifically designed primers (see Methods section for details). Successful PCR was confirmed by 1% agarose gel and the purified PCR band was subsequently ligated into a pOPINF vector. Successful ligation was confirmed by Sanger sequencing of the resulting plasmid. MmpD_DH1 was expressed under standard conditions (See Methods section for details). Purification via IMAC and SEC using an S200 column was successful yielding 5 mg/L soluble protein as confirmed by SDS-PAGE gel (Figure 74A). Analysis of this pure protein yielded a dominant series of charge states giving an observed mass of 29164.8 Da which was

within 1.2 Da of the expected mass (expected; 29166.0 Da). Analytical size exclusion chromatography (using S75 column) showed MmpD_DH1 to be a monomer in solution (Mw_{app} = 26.4 kDa)

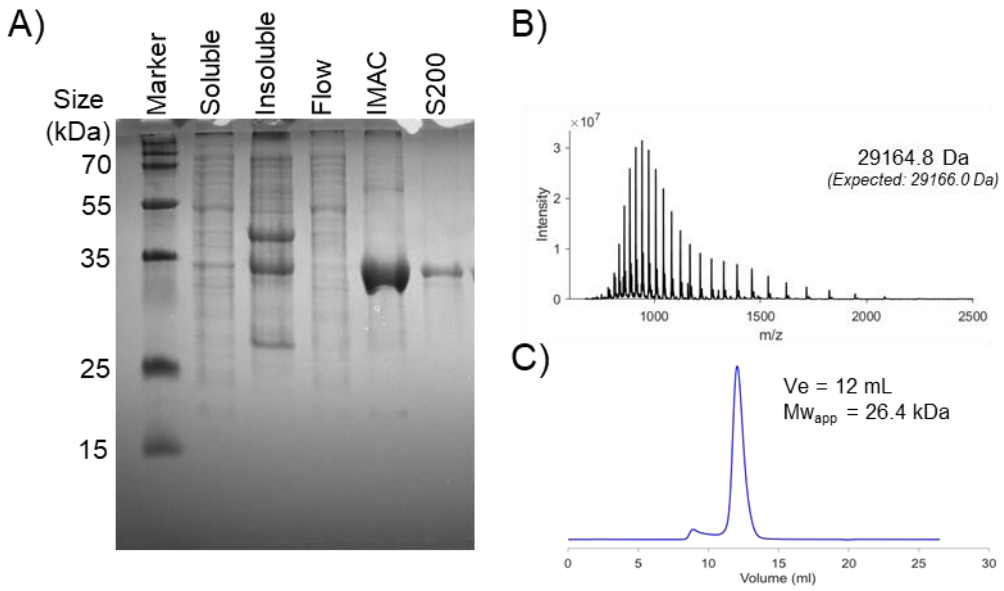


Figure 74: The expression and purification of MmpD_DH1. A) SDS-PAGE gel showing the stages of expression and purification of MmpD_DH1. B) ESI-MS analysis of purified MmpD_DH1 showed a series of charge states giving a molecular weight of 29164.8 Da (expected; 29166.0 Da). C) Analytical size exclusion trace of MmpD_DH1.

The MmpB_DH plasmid was synthesised by ThermoFisher GeneArt into a pET151_D vector. Plasmid design was led by Dr Winter. The protein was expressed under standard conditions outlined in the methods section. Purification via IMAC and SEC using an S200 column was successful yielding 2.5 mg/L soluble protein as confirmed by SDS-PAGE gel (Figure 75A). A single band was observed running slightly higher than expected on the gel between the 35 kDa and 55 kDa markers. ESI-MS analysis of the purified MmpB_DH yielded a series of charge states which gave an observed mass of 35362.7 Da, (expected; 35364.9 Da) confirming this was the correct protein. Analytical SEC trace (S200 column) of MmpB_DH again revealed the discrete domain to be a monomer in solution.

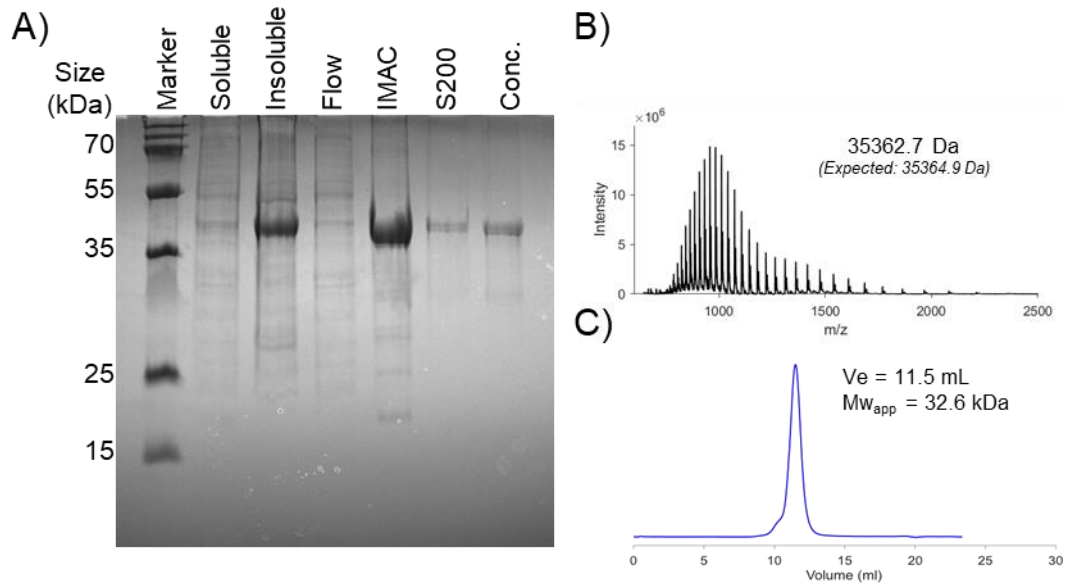


Figure 75: The expression and purification of MmpB_DH. A) SDS-PAGE gel showing the stages of expression and purification of MmpB_DH. B) ESI-MS analysis of purified MmpB_DH showed a single series of charge states giving an observed molecular weight of 35362.7 Da (expected; 35364.9 Da). C) Analytical size exclusion trace of MmpB_DH.

The ESI-MS samples of both DH enzymes showed two charge envelopes. For MmpB_DH, 70% was observed in the expected form at 35362.7 Da and 30% was observed in the α-N-gluconoylated form (mass increase of 178.1 Da, Figure 76). For MmpD_DH1 the amount of α-N-gluconoylated form was less with around 15% showing a mass increase of 178.2 Da with the remaining 85% showing the expected mass.

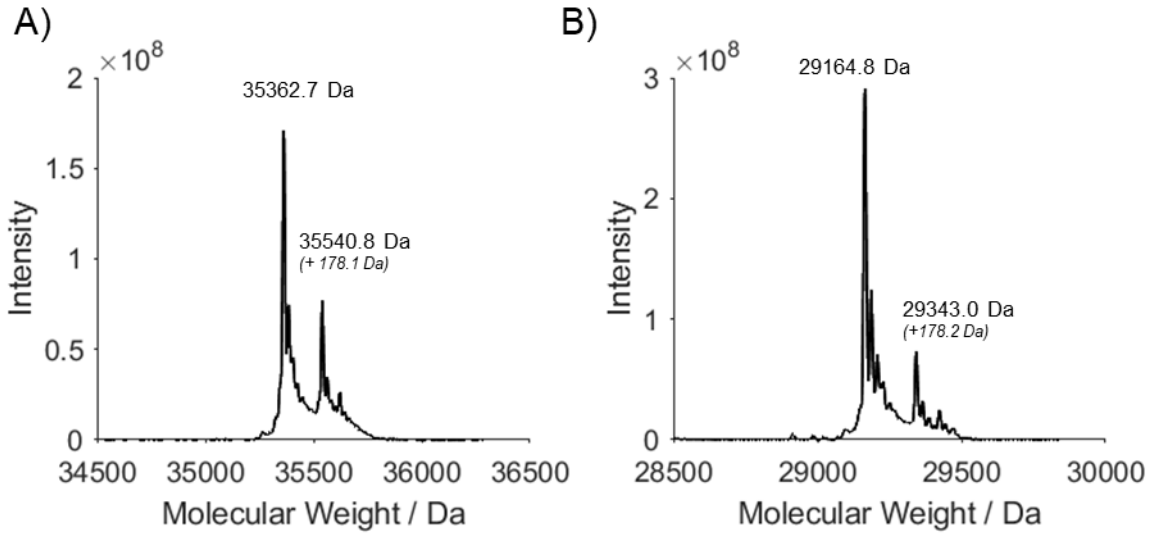


Figure 76: Deconvolutions of purified samples of MmpB_DH and MmpD_DH1. A) Deconvolution of MmpB_DH B) Deconvolution of MmpD_DH1.

3.3.5 *In-vitro* assay with MmpB_DH and MmpD_DH1

A series of *in vitro* assays were conducted to test whether each DH enzyme could catalyse the dehydration of mimics of the proposed natural substrate. The β-hydroxy substrates used to test the function of DH enzymes were 3-hydroxybutyryl **34** and the natural substrate, 3,5-dihydroxypentanyl **18**. These substrates were produced *in vitro* from their respective diketone precursors (**21** and **17**) utilising the KR enzyme MupD.

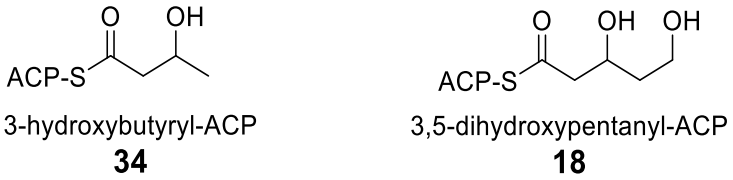
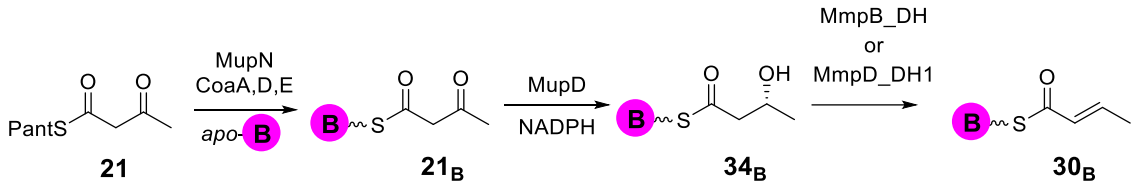


Figure 77: The substrates which will be used to test the function of MmpB_DH and MmpD_DH1. The chemical structure of the proposed natural substrate for tailoring DH is also shown.

3.3.5.1 DH Assays with 3-hydroxybutyryl-mAcpB



*Scheme 24: Reaction scheme showing the formation of the 3-hydroxybutyryl-mAcpB intermediate **34B** from acetoacetyl-pantetheine **21**. **34B** was then used as a substrate to test the DH activity of MmpB_DH and MmpD_DH1.*

The scheme showing the proposed action of the DH enzymes with **34B** is shown in Scheme 24. **21B** had been previously prepared (assay conditions described previously) as confirmed by ESI-MS (observed; 11795.5 Da, expected; 11796.8 Da, Figure 78A). Pantetheine ejection generated the expected Ppant ion confirming **21B** formation (observed; 345.14 Da, expected; 345.15 Da). To this new species was added MupD (20 µM) and NADPH (1 mM), and the reaction was monitored by ESI-MS. After 1 hour, a species was observed with a mass consistent with 3HB-mAcpB **34B** (observed; 11797.8 Da, expected; 11798.8 Da, Figure 78B). Further fragmentation of this species gave the expected ejection ion mass (observed; 347.16 Da, expected; 347.16 Da) confirming the formation of **34B**. The ACP-bound **34B** will be the starting material used to test the action of both DH enzymes.

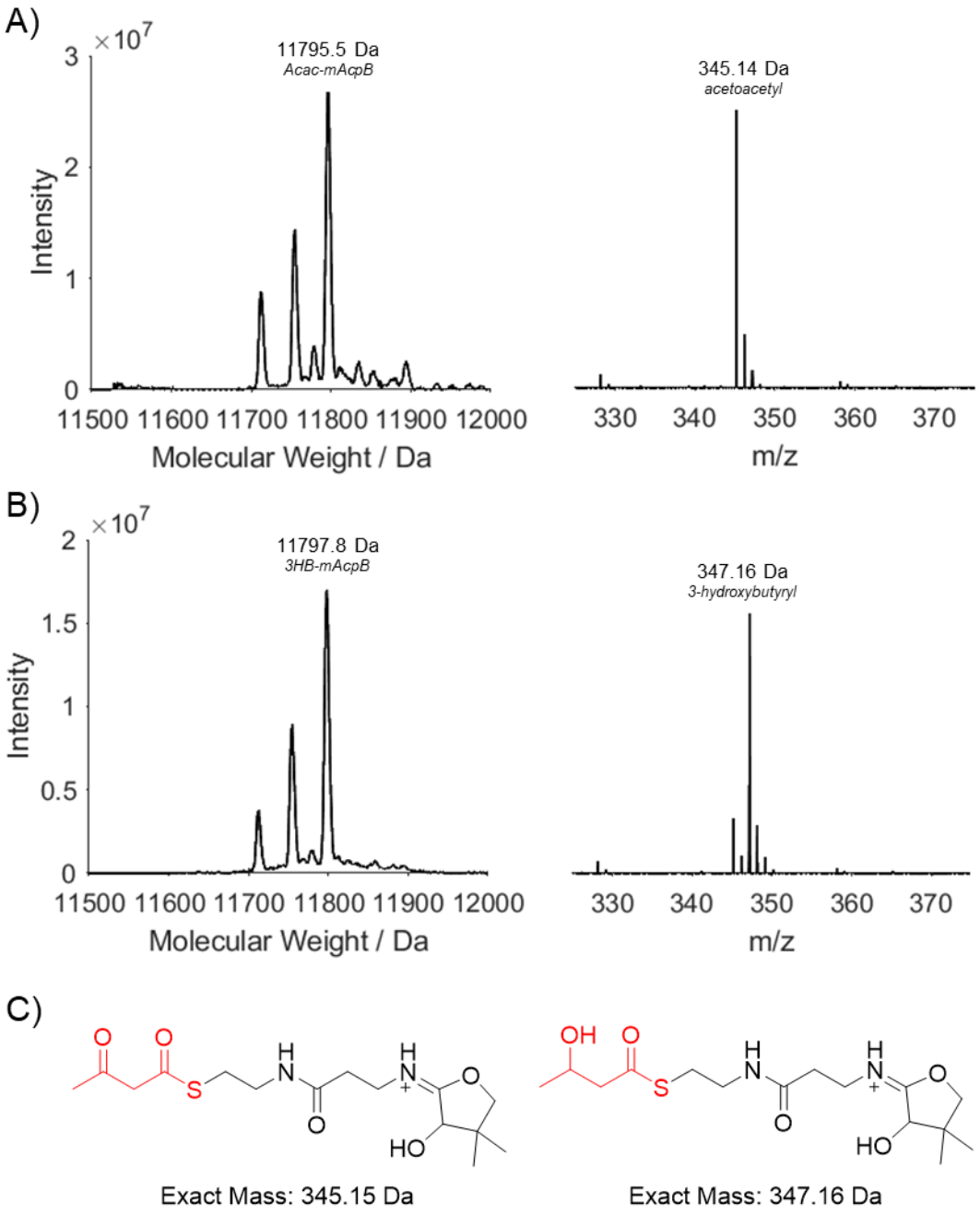


*Figure 78: The formation of mAcpB bound substrates for DH assays. A) Deconvolution showing the formation of AcAc-mAcpB **21B** after 2 hours. B) Deconvolution of the mAcpB region 1 hour after addition of KR enzyme MupD. Mass of the major species and subsequent Ppant ion confirm the formation of 3HB-mAcpB **34B**. C) Chemical structures of the acetoacetyl and 3-hydroxybutyryl Ppant ions formed.*

After the formation of **34B** this species was split into two separate assays. For assay 'BH 4C', MmpB_DH (20 µM) was added to **34B** (70 µM) to initiate the dehydration reaction. The assay was monitored by ESI-MS after 30 minutes, 4 hours and 16 hours incubation at room temperature. After 30 minutes, a new species with a mass of 11779.5 Da was detectable at low levels (~ 5%) corresponding to the expected dehydration product crotonyl-mAcpB **30B** (expected; 11780.7 Da, Figure 79B). The amount of **30B** continued to increase after 4 hours and 16 hours. After 16 hours, 20% of the mAcpB

species was observed in the dehydrated form (observed; 11779.4 Da, expected; 11780.7 Da, Figure 79C). Further fragmentation of this species gave a mixture of ejection ions. Within this mixture, small amounts of the **34$_B$** ion (observed; 329.14 Da, expected; 329.15 Da) and a larger amount of the **30$_B$** ion (observed; 347.15 Da, expected; 347.16 Da) could be seen. This shows that a small amount of dehydration activity has taken place after 16 hours.

In a separate assay, a sample of **34$_B$** was incubated at room temperature for 16 hours without the addition of DH. No formation of **30$_B$** was observed under these conditions (Figure 79A). This confirmed that formation of **30$_B$** does not occur spontaneously and is likely due to the action of a specific dehydratase.

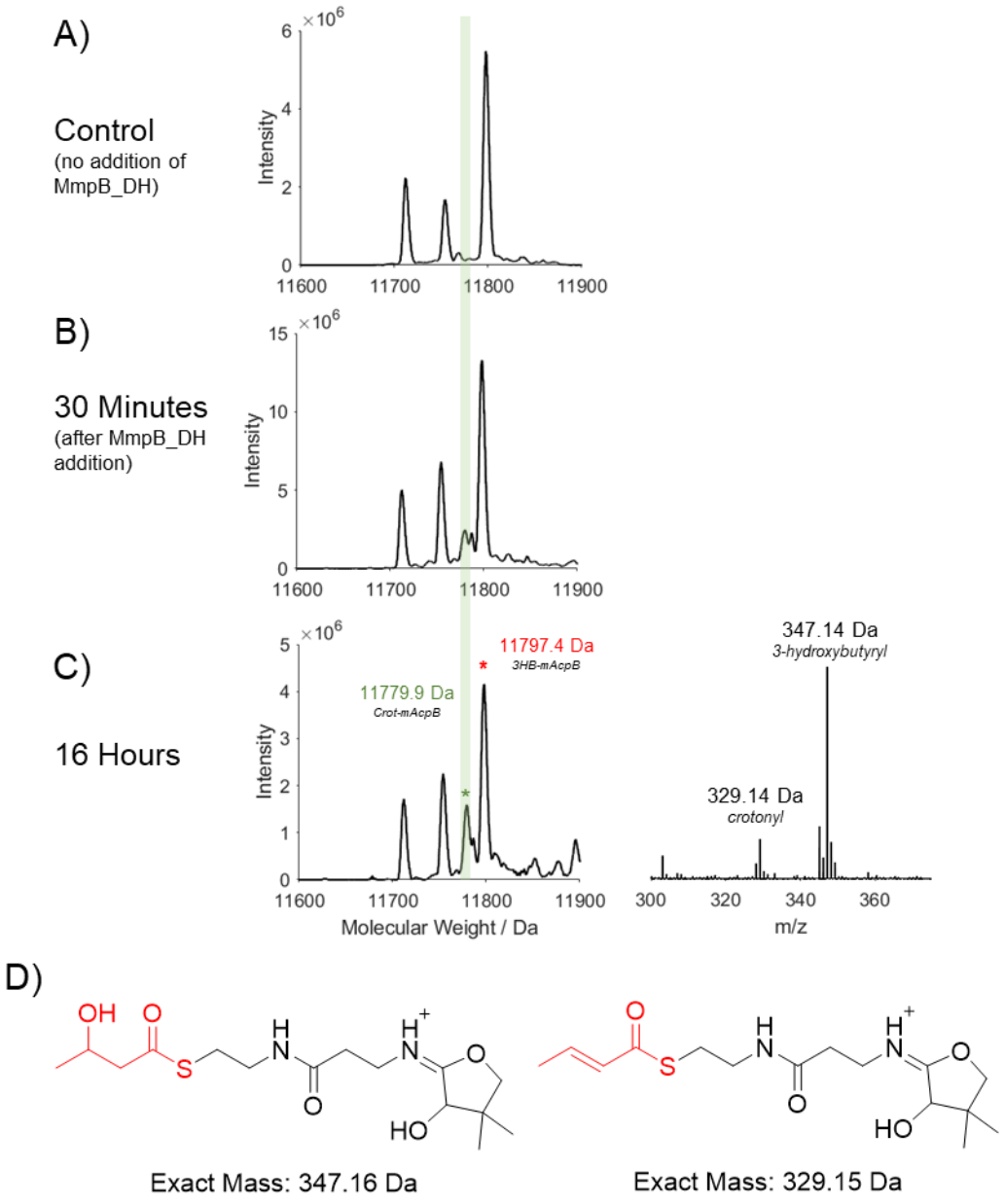


Figure 79: The addition of 3-hydroxybutyryl-mAcpB to the MmpB_DH assay. A) The deconvolution of the mAcpB region of the control reaction (prior to addition of MmpB_DH). B) Assay BH 4C, 30 minutes after addition of MmpB_DH. C) Assay BH 4C, 16 hours after initiation of reaction. Further fragmentation of this species showed a

mixture of 3HB ion (observed; 347.14 Da) and crot ion (329.14 Da) corresponding to starting material and dehydrated product respectively. D) The chemical structures and expected masses of the Ppant. ions observed.

The dehydration activity of MmpB_DH had been tested; the next step was to test the dehydration activity of the remaining DH enzyme MmpD_DH1. The remaining sample of **34$_B$** was incubated with MmpD_DH1 (20 µM) to make up assay 'DH 4C'. After 30 minutes, a small amount of dehydrated intermediate **30$_B$** was observed (observed mass; 11797.6 Da, expected mass; 11780.7 Da). This amounted to approximately 5% of the mAcpB species present. After 4 hours and 16 hours, no further increase in the amount of the dehydrated product was observed. Over all the time periods tested, the major species observed was the unreacted starting material **34$_B$.** Further fragmentation of **30$_B$** was unsuccessful in isolating significant amounts of the Ppant ion due to the small amount of intermediate produced.

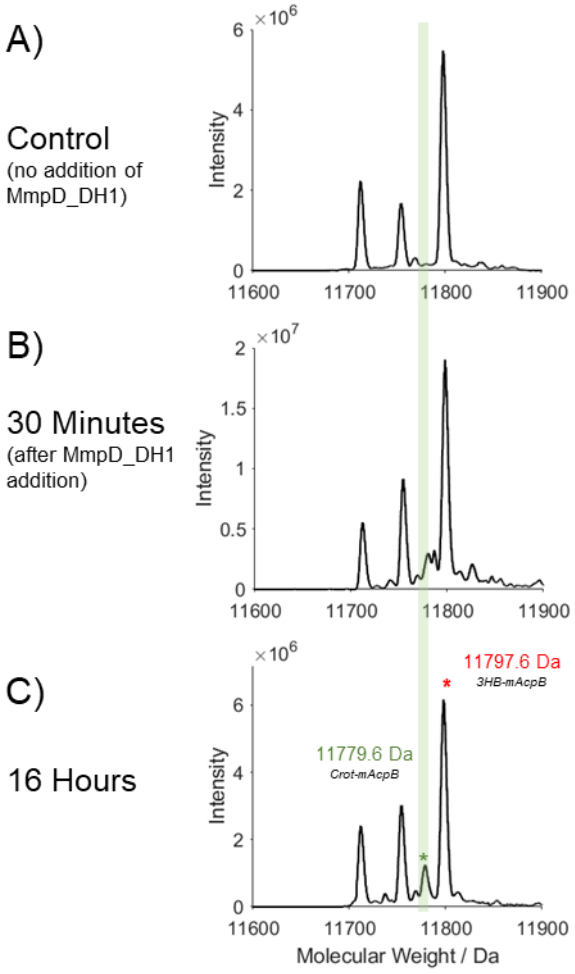


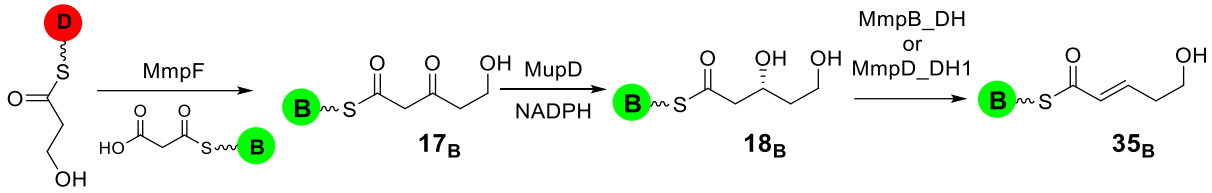
*Figure 80: The addition of 3-hydroxybutyryl-mAcpB to MmpD_DH. A) The deconvolution of mAcpB region of the control reaction (no addition of MmpD_DH). B) The deconvolution of mAcpB region of assay BH 4C, 30 minutes after addition of MmpD_DH. C) The deconvolution of mAcpB region of Assay BH 4C, 16 hours after initiation of reaction. Peak corresponding to dehydrated product crot-mAcpB **30$_B$** is marked with green asterisk, peak corresponding to unreacted starting material is marked with red asterisk.*

A small amount of DH activity was observed for both enzymes with non-natural substrate delivered by mAcpB after 16 hours. The majority of the mAcpB species in both cases was unreacted starting material. Both DH enzymes also showed a low amount of product formation after a long time period (16 hours). The observation of low-level DH activity by both enzymes implies this activity is not specific. Functional studies of PKS DH enzymes has been reported in the literature. One such example was the kinetic analysis of the PikDH2 domain from the biosynthesis of the polyketide macrolide antibiotic pikromycin. [161] The substrate specificity of PikDH2 was tested by modifying the stereocenters of the natural substrate and monitoring the effect on enzyme turnover. This showed a strict substrate specificity for the native stereochemistry and chemical structure. [162] This has been similarly observed in other PKS DH domains such as RifDH10 from rifamycin biosynthesis [163] and BorDH2 from borrelidin biosynthesis. [164] Therefore, the lack of DH activity of MmpB_DH and MmpD_DH1 may be due to the difference between **34** and the proposed natural substrate. As a result, this assay was repeated with the proposed natural substrate; 3,5-dihydroxypentanyl **18** attached to mAcpB. This substrate will be produced *in vitro* from 3HP-mAcpD utilising MmpF, malonyl-mAcpB, MupD and their respective cofactors.

Further to the substrate arguments made above, the kinetics of the *in vitro* assay was also considered. Kinetic analysis of PikDH2 domain reported in the literature found this dehydration to be reversible, producing only 50% yield of *trans*-olefin product. [162] In the biosynthetic pathway, this equilibrium is pushed toward dehydration by downstream module activities. Within mupirocin biosynthesis, MupE is expected to act after dehydration has taken place. Therefore, addition of MupE to the dehydration reaction may drive the reaction towards completion. Therefore, MupE (and NADH) will also be added to the assay when it is repeated with natural substrate **18_B**.

3.3.5.2 DH Assays with 3,5-dihydroxypentanyl-mAcpB

The production of 3,5-dihydroxypentanyl-mAcpB **18_B** was conducted according to conditions outlined in experimental detail elsewhere in this thesis (as described in Chapter 2). In brief terms, the production of 3HP-mAcpD with > 95% yield was confirmed by ESI-MS after 2 hours. Following this, 3HP-mAcpD was added to MmpF, malonyl-mAcpB, MupD to produce significant amounts of **18_B**. To **18_B** was added MupE and each DH enzyme to test whether dehydration could be observed. The proposed reaction scheme for each DH is shown in Scheme 25.



*Scheme 25: A reaction scheme showing the in vitro preparation of 3,5-dihydroxypentanyl-mAcpB **18B** to test the function of MmpB_DH and MmpD_DH1 enzymes. Intermediate **18** is proposed to be the natural substrate of the DH tailoring enzyme.*

The *in vitro* assay with natural substrate was conducted with freshly purified enzymes and newly synthesised substrates. The 'B Nat' Assay was initiated by addition of 3HP-mAcpD (37 µM) to malonyl-mAcpB (50 µM), MmpF (15 µM), MupD (15 µM), NADPH (1 mM), MupE (15 µM), NADH (1 mM) and MmpB_DH (20 µM). Previous preparations of **18B** had shown that hydrolysis of this species to produce *holo*-mAcpB readily occurred over longer reaction times. As a result, the 'B Nat' assay was monitored by ESI-MS after 30 minutes and 2 hours only. After 30 minutes, multiple species were observed in the deconvolution of the mAcpB region of the spectrum, as shown in Figure 81. The main species observed was *holo*-mAcpB (observed; 11711.4 Da, expected; 11712.7 Da) and significant amounts of unreacted malonyl-mAcpB was also seen (observed; 11797.5 Da, expected; 11798.7 Da). The expected mass of the 5-hydroxypent-2-enoyl-mAcpB **35B** was 11809.8 Da; no amount of this species was detected. Instead, a significant amount of **18B** was observed (observed; 11826.8 Da, expected; 11827.8 Da). This is consistent with no tailoring dehydration activity being catalysed by MmpB_DH with this substrate. A further new species was observed with a mass consistent with 3HP-mAcpB (observed; 11785.7 Da, observed; 11784.8 Da). This was unexpected but is presumably formed due to the transfer of 3HP from MmpF to *holo*-mAcpB. Monitoring the reaction after 2 hours showed no major changes when compared to 30 minute time point, both showed no formation of **35B**.

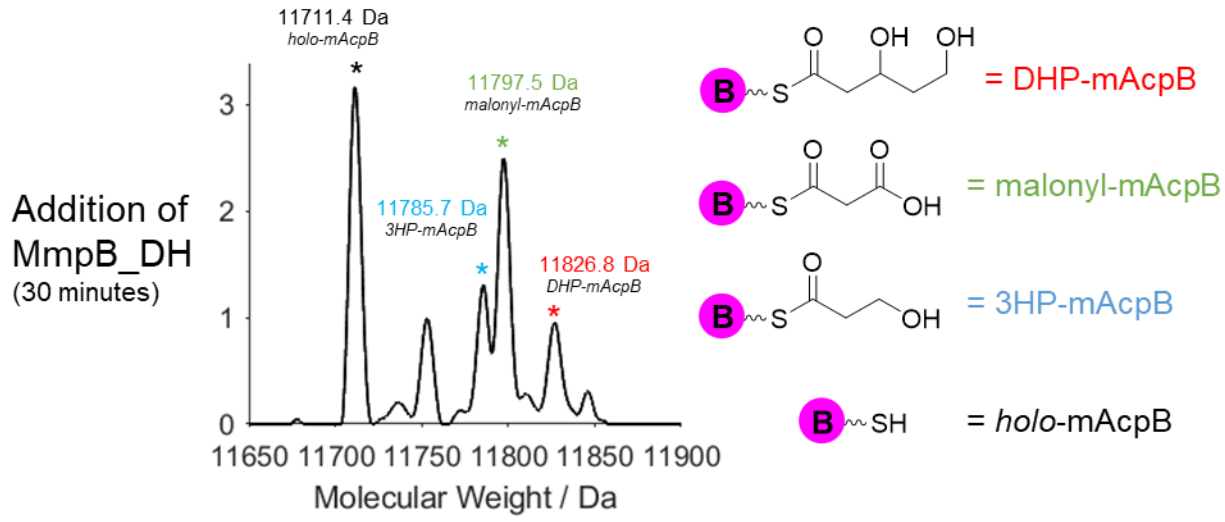


Figure 81: The B Nat assay consisting of addition of MmpB_DH to MmpF, MupD, MupE, malonyl-mAcpB and 3HP-mAcpD after 30 minutes. Each of the mAcpB species observed is shown in a coloured asterisk. A key to the different ACP species is also given.

A second assay was conducted with **18$_B$** and MmpD_DH1. This was called 'D Nat' assay and was conducted under identical conditions to B Nat Assay but with the alternative DH enzyme MmpD_DH1. The assay was initiated by addition of 3HP-mAcpD (37 µM) to malonyl-mAcpB (50 µM), MmpF (15 µM), MupD (15 µM), NADPH (1 mM), MupE (15 µM), NADH (1 mM) and MmpD_DH1 (20 µM). After 30 minutes, a mixture of mAcpB species were observed as is shown in Figure 82. The major species observed was *holo*-mAcpB (observed; 11711.2 Da, expected; 11712.7 Da). Acetyl-mAcpB (observed; 11752.8 Da, expected; 11754.7 Da) and malonyl-mAcpB (observed; 11797.4 Da, expected; 11798.7 Da) was also seen. No sign of the dehydrated product **35$_B$** was observed but 20% of the mAcpB species was 3,5-dihydroxypentanyl-mAcpB **18$_B$** (observed; 11827.2 Da, expected; 11827.8 Da). The reaction was also monitored after 2 hours; this spectrum was very similar to that recorded after 30 minutes and showed no formation of **35$_B$** and a smaller amount of **18$_B$** consistent with hydrolysis from the terminal thiol of mAcpB.

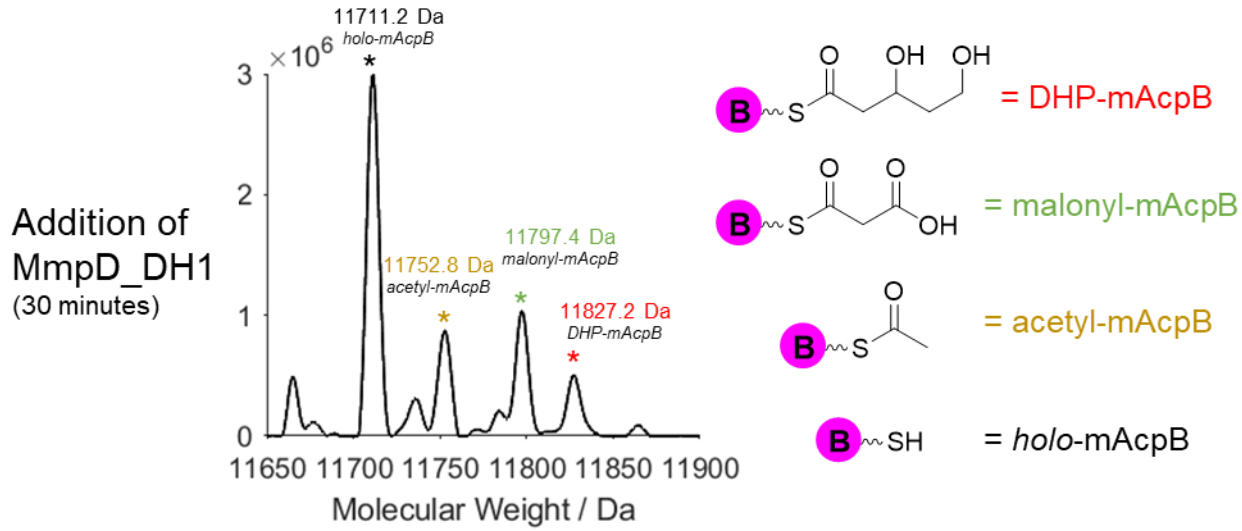


*Figure 82: The D Nat assay consisting of addition of MmpD_DH to MmpF, MupD, malonyl-mAcpB and 3HP-mAcpD after 30 minutes. Formation of 3,5-dihydroxypentanyl-mAcpB (DHP-mAcpB) **18** was observed but no formation of dehydrated product 5-hydroxypent-2-enoyl-mAcpB **19** was observed. This shows no activity of MmpD_DH with **18**.*

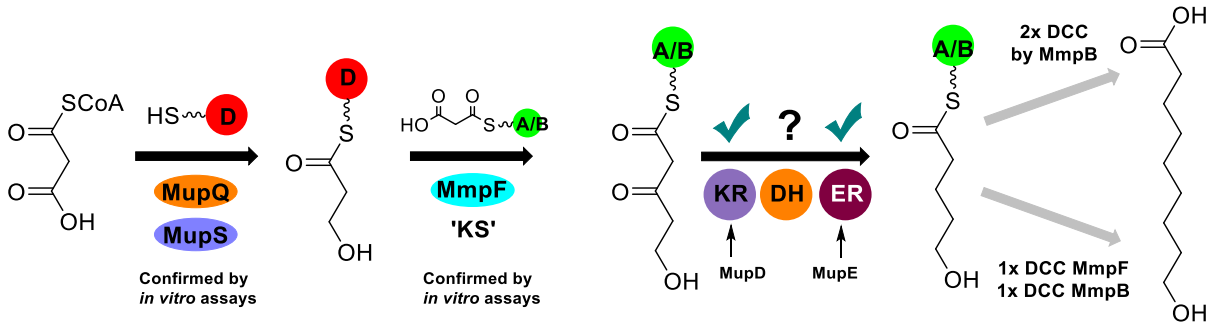
Both assays (B Nat and D Nat) showed formation of β-hydroxy substrate $\mathbf{18_B}$, but no formation of the dehydration intermediate $\mathbf{35_B}$ or the downstream reduced 5-hydroxypenanyl-mAcpB intermediate (formed after the action of MupE). This confirmed that the condensation reaction catalysed by MmpF, followed by reduction catalysed by MupD, were both successful but subsequent dehydration was not. It is therefore possible that neither MmpB_DH or MmpD_DH1 provide tailoring DH activity at this point in the pathway or that these excised domains have not been expressed in their optimal functional form.

Without any further candidates for this DH activity, it is possible that the DH activity is simply absent in this BGC. This is not a feature unique to the biosynthesis of mupirocin and is observed in many other *trans*-AT PKSs. For example, the enacyloxins are a group of polyketide antibiotics isolated from the proteobacterium *Frateuria* sp. W-315. The BGC encoding enacyloxin biosynthesis was isolated and found to consist of a hybrid *cis*-/*trans*-AT PKS.[165] The current biosynthetic model for enacyloxin production predicts that modules 6 and 9 lack the activity of a DH and KR domain respectively. The authors do not speculate as to how this missing DH or KR activity is provided.

Another example of BGC missing DH activity is observed in the biosynthesis of the group of macrolide antibiotics known as sorangicins, which are produced by myxobacterium *Sorangium cellulosum*.[166] The proposed biosynthetic model predicts a *cis*-configured double bond is produced by module 5. This is unusual as this module contains only a KS and di-domain ACP.[167] Module 5 therefore lacks both the KR and DH activity required to produce the observed double bond. This is surprising as no *trans*-acting KR or DH are observed in the BGC of the sorangicins.

The reasons for the lack of DH functionality in the mupirocin pathway are unknown, although this pathway is not unique in lacking specific DH functionality (as described above). Despite this, no consensus has been reached on how the required DH activity is provided in these systems. It is possible that dehydration may occur in modules which are not a part of their respective BGCs at a post-PKS stage. For mupirocin, it is possible that a FabA or FabZ like enzyme from *P. fluorescens* fatty acid biosynthesis (primary metabolism) may be responsible for this missing DH activity.

3.3.6 Chapter 3 Summary



Scheme 26: A scheme showing the proposed steps in the biosynthesis of 9HN. The KS domains responsible for further condensation reactions are not yet known and are shown in grey.

The work performed is this chapter has focused on understanding the remaining DH and ER tailoring steps expected to occur after MmpF catalysed condensation. The activity of proposed ER enzyme MupE was tested with crotonyl **30** and 2-methylbutenoyl **31** substrates attached to the cognate ACP; mAcpB. Successful reduction was observed after 16 hours with $\mathbf{30_B}$ but no reduction was observed for $\mathbf{31_B}$. This provides the first experimental evidence of the role of MupE as proposed ER enzyme within the biosynthesis of 9HN. Further to this, loading of the crotonyl moiety on to non-cognate mAcpD led to a reduction in product formation after the same time period. This shows the ACP selectivity of this reaction is important, but not as important as in the MupD KR reaction.

The ability of two DH enzymes from separate modules of mupirocin biosynthesis to provide tailoring DH activity were tested. MmpB_DH and MmpD_DH1 were expressed and purified. Assays were conducted with 4-carbon β-hydroxy substrates, which showed small amounts of enzyme turnover. Assays were repeated with the natural substrate of the reaction in the presence of the downstream enzyme (MupE). This showed the pathway has stalled at the β-hydroxy intermediate $\mathbf{18_B}$ showing that these DH enzymes are not expected to provide this missing DH activity.

3.4 Remaining Condensation Reactions During 9HN Biosynthesis

The first elongation step during biosynthesis of 9HN has been shown to be catalysed by MmpF (Chapter 2). In total, two further elongation reactions are expected to occur leading to the formation of 9HN **15**. Two KS enzymes from the mupirocin BGC are candidates for these elongation reactions; MmpF and MmpB_KS. In this section, the identity of the enzyme responsible for the second and third elongation reactions will be determined through *in vitro* assays with synthetic pantetheine analogues.

3.4.1 MmpB_KS Translocation Assays

MmpB has been shown to be required for normal PA A production through gene knockout experiments. When the mmpB gene is knocked out from a plasmid containing all other mupirocin BGC genes, complete abolishment of PA A production is observed (unpublished work). As the mmpB gene is surplus to requirement for monic acid synthesis, it is expected to be involved in the biosynthesis of fatty acid chain of 9HN.

To understand more about the proposed substrate specificity of the KS domain of MmpB (MmpB_KS) a series of *in vitro* assays were conducted with purified MmpB_KS. A series of pantetheine substrates of differing acyl chain lengths were chemically synthesised (Figure 83). These included 4-hydroxybutarate (4HB) **36**, 5-hydroxypentanoate (5HP) **37**, 6-hydroxyhexanoate (6HH) **38**, 7-hydroxyheptantoate (7HH) **39** and 9-hydroxynonanoate (9HN) **40**. This set included odd carbon lengths that the MmpB_KS might encounter and therefore transfer (**37**, **39**) and the final 9HN odd carbon length that it should not reasonably encounter (**40**). The two even carbon lengths would not be expected to be on pathway intermediates but would test promiscuity around any odd carbon chain lengths that were selected for. All of the pantetheine substrates were synthesised by Angus Weir apart from **37** which was synthesised by Sbu Mbatha. As with the synthesis of previous pantetheine analogues, these compounds were synthesised utilising a key pantetheine coupling reaction. Detailed experimental procedures describing synthetic methods to produce these compounds are given elsewhere. [146,168]

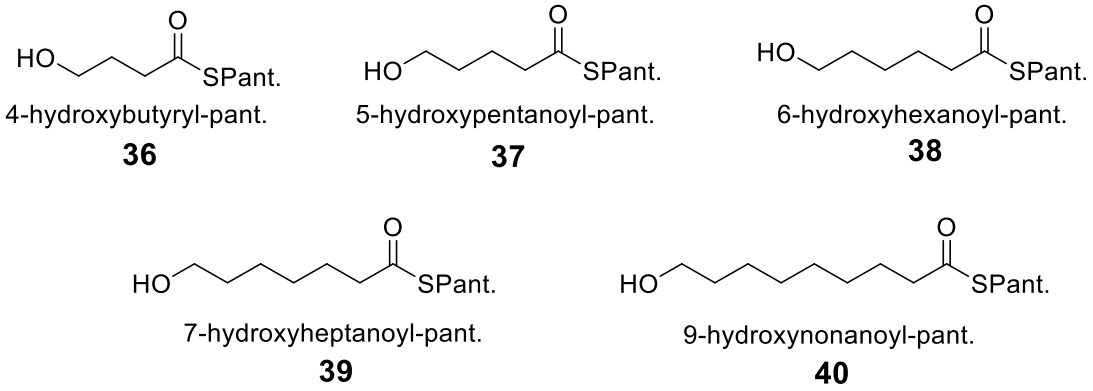


Figure 83: The pantetheine substrates used to test the ability of MmpB_KS to translocate different length acyl moieties.

To test the ability of MmpB_KS to catalyse a transthiolation reaction from the pantetheine substrates, direct self-loading assays were conducted. MmpB_KS (20 μM) was added to an excess of each acyl-pantetheine substrate (1 mM) and the mass of the MmpB_KS was monitored by ESI-MS after 4 hours (Figure 84). The MmpB_KS alone was analysed by ESI-MS prior to addition of substrate. This showed a single series of charge states which correspond to *apo*-MmpB_KS (observed; 67006 Da, expected; 67021 Da, Figure 84A).

The smallest acyl group tested was 4HB **36** (Figure 84B). After 4 hours, the self-loading assay with **36** showed no sign of translocation to MmpB_KS. Only *apo*-MmpB_KS (observed; 67007 Da) was observed. The next acyl group to be tested was 5-carbon 5HP **37** (Figure 84C). This was added to MmpB_KS and left to incubate at room temperature for 4 hours. After this time, analysis by ESI-MS showed the predominant species was again *apo*-MmpB_KS (observed; 67022 Da, expected; 67021 Da). Around 20% of MmpB_KS species observed showed a mass increase of +276 Da (observed; 67298 Da). This +276 Da adduct will presumably occur due to binding of pantetheine to the MmpB_KS at the active site cysteine. Irrespective of this, 5HP showed no signs of translocation to MmpB_KS after 4 hours.

The incubation of MmpB_KS with the longer 6HH-pantetheine **38** for 4 hours also produced *apo*-MmpB_KS as the predominant species (observed; 67007 Da; Figure 67D). Despite this, a new species was also observed corresponding to a mass increase of +114 Da (observed; 67121 Da) in approximately 30% yield. This mass increase was consistent with successful transfer of 6HH to MmpB_KS (expected mass increase: 114 Da). This shows that MmpB_KS successfully binds **38**.

The penultimate acyl moiety to be tested was 7HH **39**. After 4 hours of incubation with MmpB_KS, the predominant species observed was *apo*-MmpB_KS (observed; 67006 Da) making up 45% of the species observed. Successful binding of 7HH was observed with 35% yield (observed; 67135 Da) due to a mass increase of +129 Da from the *apo*-MmpB_KS form (expected mass increase +128 Da). This confirmed MmpB_KS was able to bind 7HH, possibly with slightly better efficiency than the shorter 6HH substrate. 25% of the MmpB_KS was observed in the +276 Da form (observed; 67282 Da) consistent with formation of a pantetheine adduct.

The final acyl-pantetheine moiety to be incubated with MmpB_KS over 4 hours was 9HN **40**. After this time, ESI-MS analysis of the charge states of MmpB_KS showed that no translocation of 9HN was observed and only the *apo*-MmpB_KS species was observed (observed; 67007 Da).

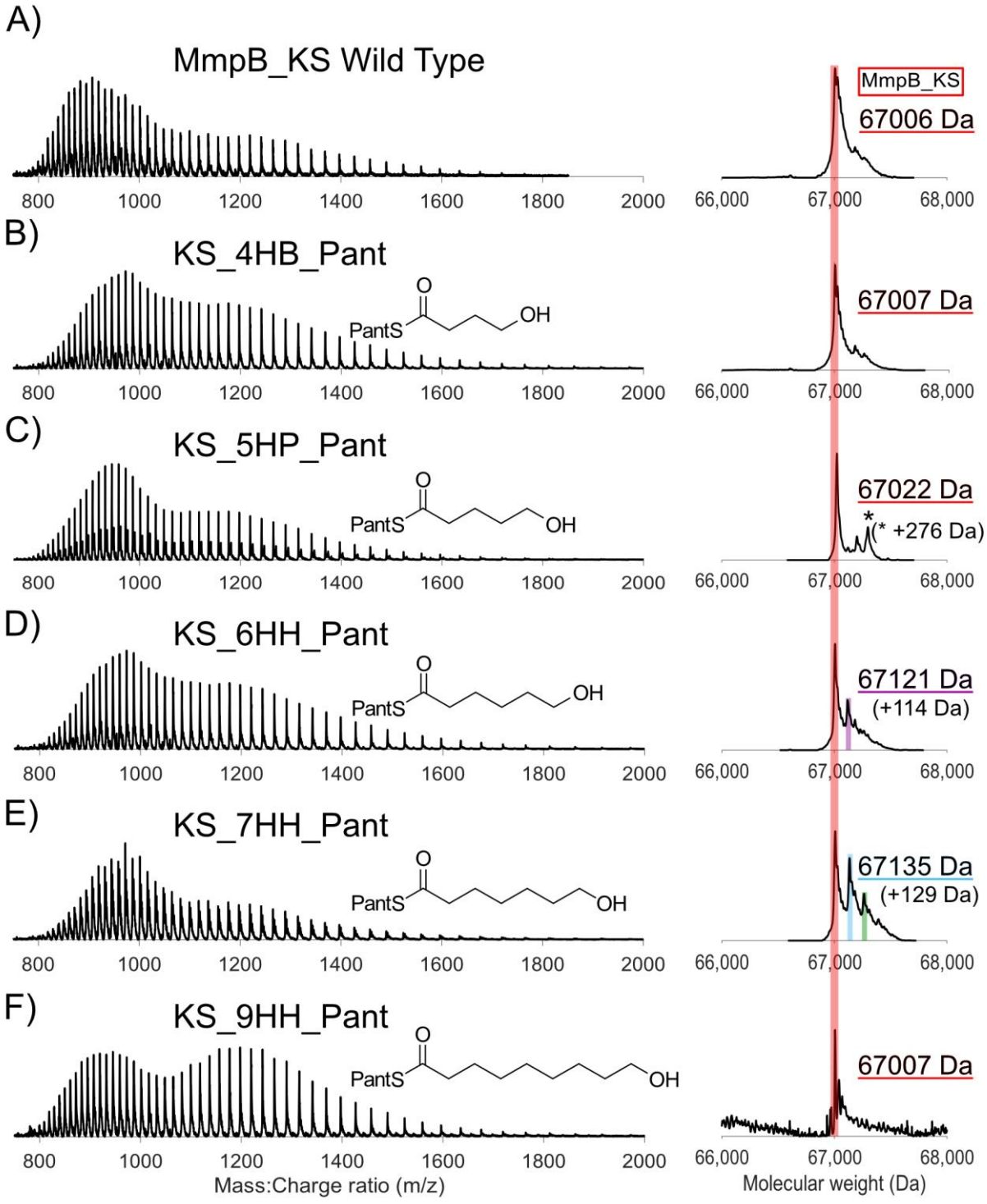
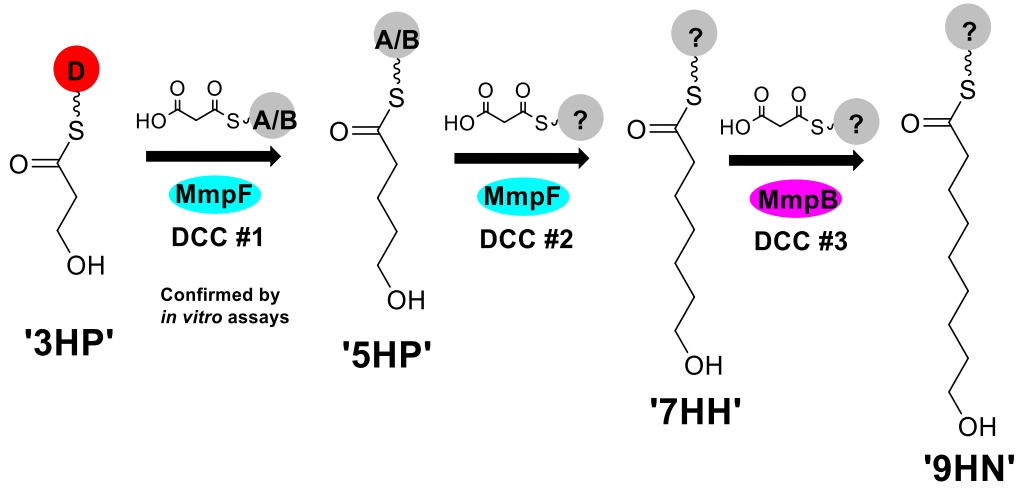


Figure 84: The charge states and subsequent deconvolution of MmpB_KS with hydroxy fatty acid pantetheine analogues of increasing carbon-carbon chain length after 4 hours. Red bar corresponds to apo-MmpB_KS. A) MmpB_KS after no addition of pantetheine substrate. B) MmpB_KS after addition of 4HB. C) MmpB_KS after addition of 5HP. D) MmpB_KS after addition of 6HH, translocated species is shown in purple. E) MmpB_KS after addition of 7HH, translocated species is shown in light blue, +276Da adduct is shown in green. F) MmpB_KS after addition of 9HN.

The acyl loading reactions conducted with MmpB_KS provide useful information about the substrate specificity of this enzyme. The experimental data appears to show preferential binding of 6HH and 7HH. Despite this, no transfer of 5HP was observed. The failure to load 5HP suggested that the MmpB_KS does not perform the 5- to 7-carbon elongation step whereas the increasing selectivity for 7HH indicated that MmpB_KS might catalyse a single 7- to 9- carbon elongation only. In the context of the overall biosynthesis of 9HN, a working hypothesis is that MmpF catalyses the 3-carbon to 5-carbon (DCC #1) and 5-carbon to 7 carbon (DCC #2) elongation reactions but that the final 7-carbon to 9-carbon (DDC #3) elongation is catalysed by MmpB_KS as is shown in Scheme 27.



Scheme 27: Scheme showing the KS enzymes proposed to catalyse each elongation step within biosynthesis of 9HN. In this scheme, the tailoring KR, DH and ER reactions are not shown. MmpF has been shown to catalyse the first elongation reaction with unusual starter unit 3HP. MmpF is then proposed to catalyse the second elongation reaction also. The final elongation reaction is proposed to be catalysed my MmpB_KS.

3.4.2 MmpB_KS 7-Carbon to 9-Carbon Condensation Reaction

To confirm that MmpB_KS catalyses the final condensation reaction during the biosynthesis of 9HN, *in vitro* assays were conducted with MmpB_KS primed from **39** and a malonyl donor ACP. This final elongation reaction will involve the action of two ACPs. The first ACP will deliver the 7HH moiety to the active site of MmpB_KS. This might be one or both mAcpA and mAcpB depending on if they are each able to undergo DCC #2 with MmpF, which has not yet been shown. The second ACP will deliver the chain extending malonyl unit. A tri-domain ACP (ACP567) is encoded for within MmpB, any or all of which may deliver malonate to MmpB_KS although there has been some previous suggestion of functional inequivalence.[144] Each of the individual ACP domains of MmpB_ACP had previously been synthesised as separate entities by another member of this group (Nahida Akter, University of Bristol, Thesis). pET151 plasmids containing each of the individual ACP domains was therefore available for use. The expression and purification of these proteins are outlined elsewhere.

3.4.3 Formation of malonyl-ACPs

Each ACP within the MmpB tri-domain was termed ACP5, ACP6 and ACP7 respectively. Malonyl was transferred to each of these ACPs using *apo*-ACP (100 µM) to MupN (5 µM), malonyl CoA (1 mM) and $MgCl_2$ (10 mM). For the ACP5 malonyl transfer reaction, a new species was observed after 4 hours a with a mass of 12795.3 Da (90% yield, Figure 85A). This species is consistent with the expected mass of malonyl-ACP5 (expected; 12799.9 Da) confirming successful derivatisation.

For the ACP6 transfer reaction, after 4 hours the malonyl species made up 80% of the species present (observed; 13542.2 Da, expected; 13544.7 Da, Figure 85B). As observed in other malonyl formation reactions, small amounts of *holo*-ACP6 (observed; 13455.4 Da, expected; 13458.7 Da), sodium adduct of malonyl-ACP6 (observed; 13563.2 Da) and acetyl adduct (observed; 13583.7 Da) of malonyl-ACP6 were also observed.

For the ACP7 transfer reaction, successful formation of malonyl-ACP7 was confirmed after 4 hours. The major species observed was malonyl-ACP7 (observed; 12987.4 Da, expected; 12991.1 Da, Figure 85C). Small amounts of *holo*-ACP7 (observed; 12901.5 Da, expected; 12905.1 Da), sodium adduct of malonyl-ACP7 (observed; 13009 Da) and acetyl adduct of malonyl-ACP7 (observed; 13028.5 Da) were also observed. All malonyl-ACP species were desalted and stored at -20 °C prior to downstream use.

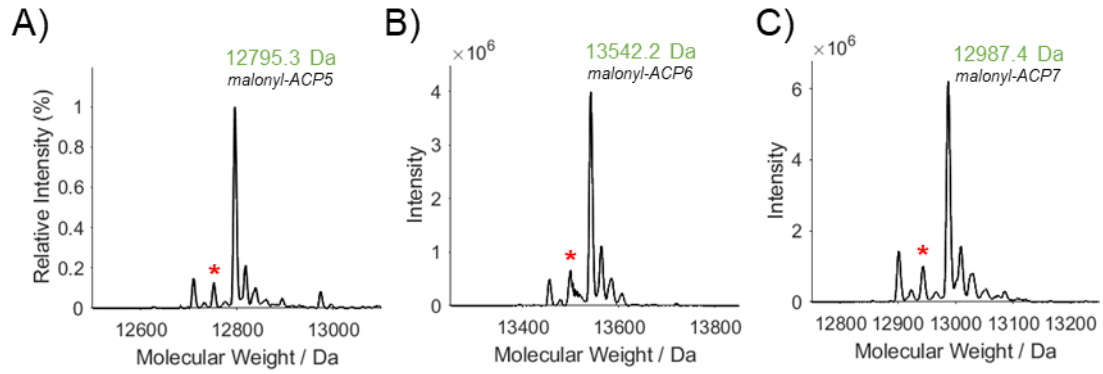
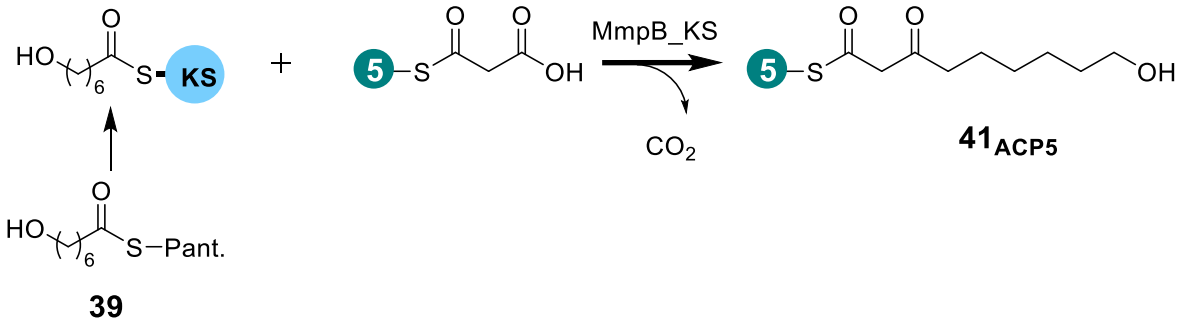


Figure 85: The formation of malonyl-ACP5, malonyl-ACP6 and malonyl-ACP7 4 hours after initiation of reaction. Acetyl species are shown by a red asterisk. A) The deconvolution of the ACP5 malonyl transfer reaction initiation shows predominantly malonyl-ACP5. B) The deconvolution of the ACP6 malonyl transfer reaction. C) The deconvolution of the ACP7 malonyl transfer reaction.

3.4.4 MmpB_KS Condensation with Malonyl-ACP5 and 7-hydroxyheptanoate



Scheme 28: The proposed reaction scheme showing the final C-7 to C-9 elongation reaction catalysed by MmpB_KS. MmpB_ACP5 is simplified to '5' in this diagram.

MmpB_KS was assayed for its ability to catalyse the key condensation reaction using **39** and malonyl-ACP5 substrates. Successful condensation would be confirmed by ESI-MS through the generation of 9-hydroxy-3-oxonanyl-ACP5 (9HO-ACP5) **41**$_{\mathbf{ACP5}}$ (Scheme 28).

The reaction was initiated by addition of MmpB_KS (20 µM) to excess of **39** (3.3 mM) and malonyl-ACP5 (80 µM) and monitored by ESI-MS after 30 minutes and 2 hours. Pleasingly, after 30 minutes a new species (~ 10%) was observed with a mass consistent with 9HO-ACP5 **41**$_{\mathbf{ACP5}}$ (observed; 12884.0 Da, expected; 12884.0 Da, Figure 86A). Concurrently a similar proportion of *holo*-ACP5 was also observed presumably arising from either loss of this product or MmpB_KS catalysed hydrolysis of the malonyl group (observed; 12712.7 Da, expected; 12713.9 Da). The major species observed after 30 minutes (75% yield) was unreacted malonyl-ACP5 (observed; 12799.6 Da, expected; 12799.9 Da).

After 2 hours the major ACP5 species observed was **41**$_{\mathbf{ACP5}}$ (observed; 12883.6 Da, Figure 86B) produced in 35% yield. Ppant ejection generated a 431.23 Da ion providing confirmation of the formation of **41**$_{\mathbf{ACP5}}$ (expected: 431.22 Da, Figure 86C). After this time, increasing spectral complexity was apparent and a further new species was produced in about 10% yield with a mass consistent with 7HH-ACP5 (observed; 12838.6 Da, expected; 12842.0 Da) formed from self-loading of 7HH to terminal thiol of *holo*-ACP5 produced by hydrolytic side reactions. Malonyl-ACP5 (observed; 12796.6 Da), *holo*-ACP5 (observed; 12711.1 Da) and acetyl-ACP5 (observed; 12752.8 Da) were also produced in yields of 30%, 15% and 10% respectively. The formation of **41**$_{\mathbf{ACP5}}$ and subsequent Ppant ion confirmed that successful condensation has taken place.

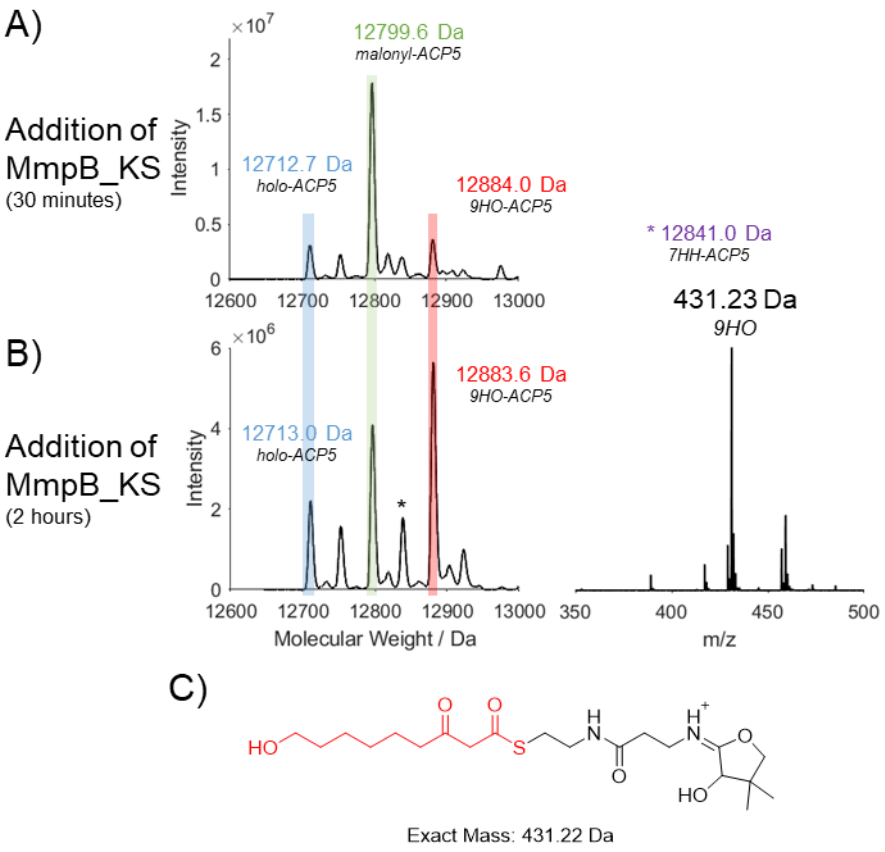


Figure 86: The observed spectra at different time points for the MmpB_KS + 7HH-Pant + mal-ACP5 assay. Malonyl-ACP5 is shown in green, holo-ACP5 is shown in blue, condensation product is shown in red and 7HH-ACP5 is shown in purple. A) The deconvoluted spectrum 30 minutes after initiation of assay. B) The deconvoluted spectrum 2 hours after initiation of assay (left). Ppant ion of 9HO-ACP5 species (right). C) Chemical structure and expected mass of the Ppant ion produced from fragmentation of 9HO-ACP5.

A control reaction was set up to ensure that DCC activity observed was solely due to the action of MmpB_KS. The control assay was conducted under identical conditions but using MmpB_KS which had been denatured by heating to 100 °C for 10 minutes prior to its inclusion in the assay. The reaction was initiated by addition of boiled MmpB_KS (20 µM) to malonyl-ACP5 (80 µM) and excess of **39** (3.3 mM). After 2 hours the assay was monitored by ESI-MS (Figure 87). No sign of 9HO-ACP5 **41**$_{\mathbf{ACP5}}$ was observed which is consistent with non-functional MmpB_KS. The major species observed was malonyl-ACP5 (observed; 12798.9 Da, expected; 12799.9 Da, 65% yield) but again a minor *holo*-ACP5 component was detectable (observed; 12712.9 Da, expected; 12713.9 Da, 15% yield) suggesting that the malonyl-ACP5 had a weak tendency to self-hydrolyse. Similarly, approximately 20% of the ACP5 species present was 7HH-ACP5 (observed; 12841.1 Da, expected; 12842.0 Da) suggesting that ACP self-loading was responsible for 7HH-ACP5 formation rather than transfer via MmpB_KS. This control reaction confirmed that in the absence of a functional MmpB_KS no condensation takes place. Therefore, the formation of **41**$_{\mathbf{ACP5}}$ in previous assay must be due to the action of MmpB_KS.

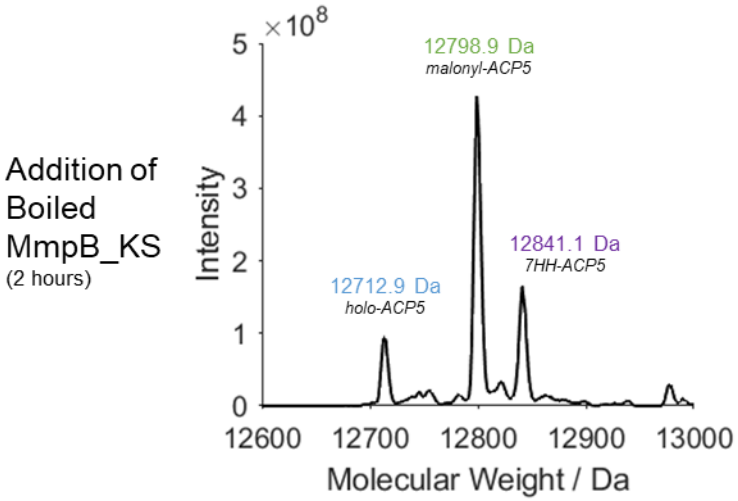


Figure 87: Negative control reaction conducted with boiled MmpB_KS, mal-ACP5 and 7HH-pant. After 4 hours, no sign of 9HO-ACP5 was observed.

3.4.5 MmpB_KS Condensation with Malonyl-ACP6 and Malonyl-ACP7

The assay was repeated with the two remaining ACPs which make up MmpB_ACP; ACP6 and ACP7. First, a condensation assay was repeated with with malonyl-ACP6. This will test the ability of ACP6 to deliver malonyl to MmpB_KS. The reaction was initiated by the addition of MmpB_KS (20 µM) to malonyl-ACP6 (80 µM) and an excess of **39** (3.3 mM) and monitored by ESI-MS after 4 hours. After this time, a new species was observed with a mass consistent with 9-hydroxy-3-oxanoyl-ACP6 (9HO-ACP6) **41**$_{\mathbf{ACP6}}$ (observed; 13625.9 Da. Expected; 13628.8 Da). In addition to **41**$_{\mathbf{ACP6}}$ observed in 30% yield, the major species (35 % yield) was malonyl-ACP6 (observed; 13541.0 Da, expected; 13544.7 Da). Smaller amounts of *holo*-ACP6 (observed; 13455.0 Da, expected 13458.7 Da, 10 % yield), 7HH-ACP6 (observed; 13584.5 Da, expected; 13586.8 Da, 10% yield) were also observed – consistent with ACP5 assays. This assay confirms that ACP6 can also deliver malonyl to MmpB_KS.

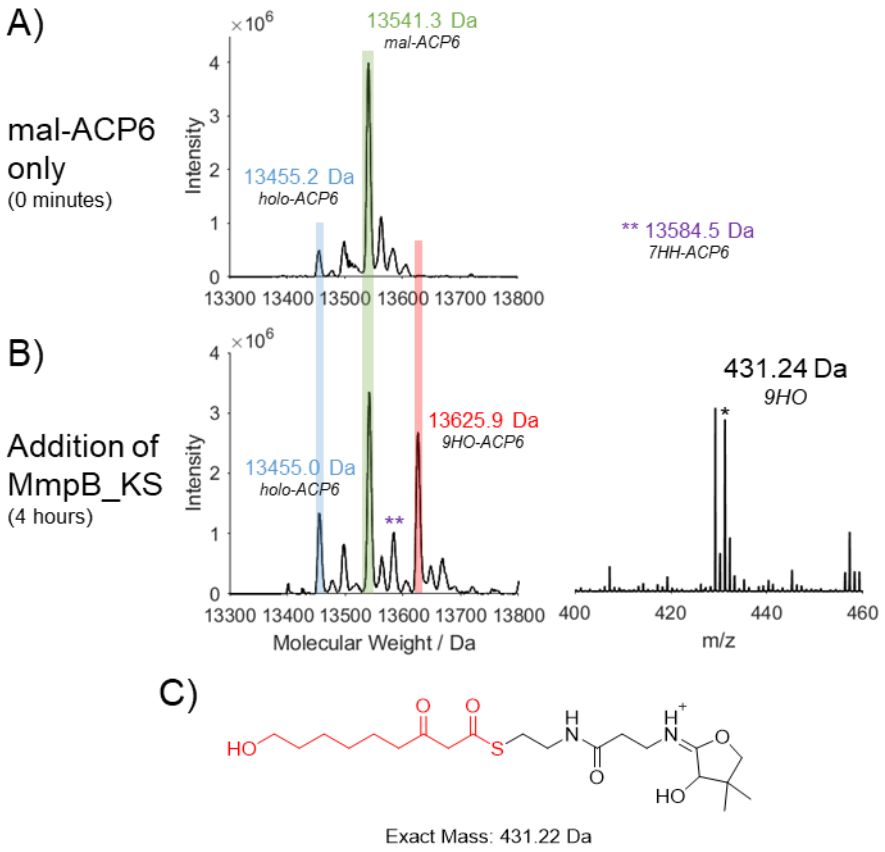


Figure 88: The comparison of mal-ACP6 spectrum with the MmpB_KS, 7HH-Pant, mal-ACP6 assay after 4 hours. Malonyl-ACP6 is shown in green, holo-ACP6 is shown in blue and condensation product is shown in red. A) No formation of DCC product is observed for mal-ACP6 deconvolution. B) After 4 hours a significant amount of 9HO-ACP6 is formed (obs; 13625.9 Da, exp; 13628.8 Da) and 431.24 Ppant ion is observed (asterisked) confirming successful condensation reaction. C) The chemical structure of the Ppant ion of 9HO-ACP6.

The assay was repeated with the final ACP; malonyl-ACP7. After 4 hours, 9HO-ACP7 **41**$_{ACP7}$ was observed (observed; 13072.2 Da, expected; 13075.2 Da) in 30% yield confirming condensation had taken place as shown in Figure 89B. Further fragmentation of **41**$_{ACP7}$ species gave expected Ppant ion confirming successful condensation (observed; 431.24 Da, expected; 431.22 Da). The majority species observed in the deconvolution of ACP7 was unreacted starting material; malonyl-ACP7 (observed; 12988.0 Da, expected; 12991.1 Da, 35% yield). Minor amounts of *holo*-ACP7 (observed; 12901.8 Da, expected; 12905.1 Da) and 7HH-ACP7 (observed; 13030.0 Da, expected; 13033.2 Da) were also observed which is consistent with the ACP5 and ACP6 condensation assays.

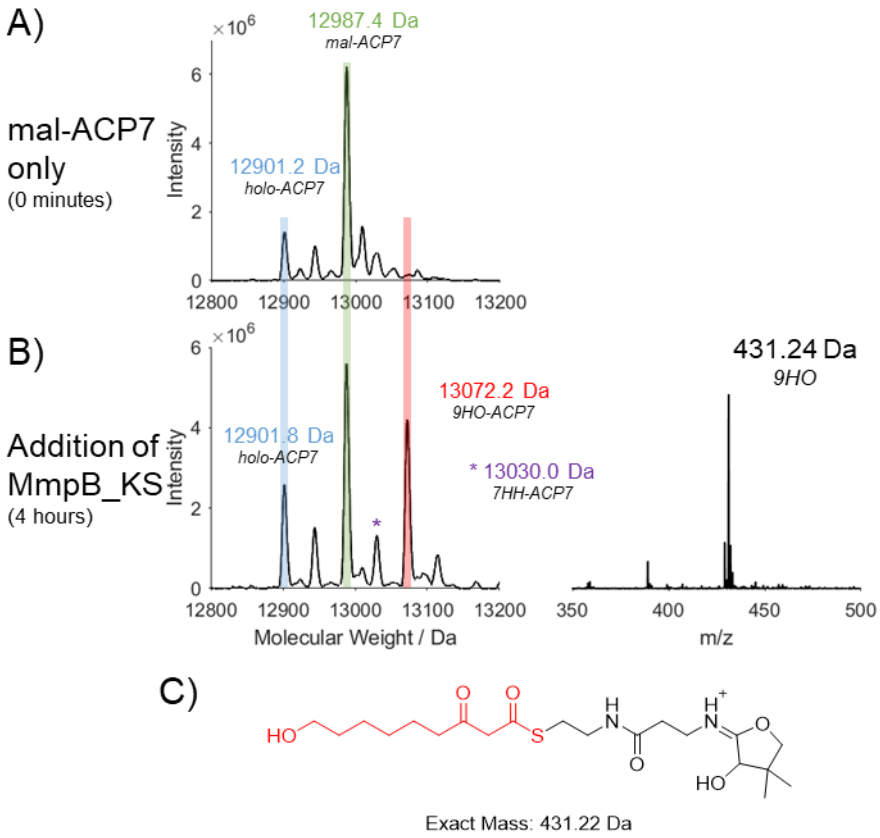


Figure 89: The comparison of mal-ACP7 spectrum with the MmpB_KS, 7HH-Pant, mal-ACP7 assay after 4 hours. A) The deconvolution of the spectrum 4 hours after the initiation of malonyl-ACP7 formation reaction. B) The deconvolution of ACP7 region of spectrum (left) and Ppant ejection ion of 9HO-ACP7 species (right). C) The chemical structure of the Ppant ion of 9HO-ACP7

Further fragmentation of the **41**$_{ACP6}$ and **41**$_{ACP7}$ species showed two Ppant ions of similar intensity, with the higher mass corresponding to the expected mass of the 9HO Ppant ion (observed; 431.24 Da, expected; 431.22 Da) and the second peak showing a loss of 2 Da (observed; 429.22 Da). This was also present as a very minor species for a previous batch of **41**$_{ACP5}$ (Figure 90). This was attributed to an oxidation of one of the hydroxyl groups to carbonyl group, arising from the input of collisional induced dissociation energy during the phosphopantetheine assay rather than any other functional significance.

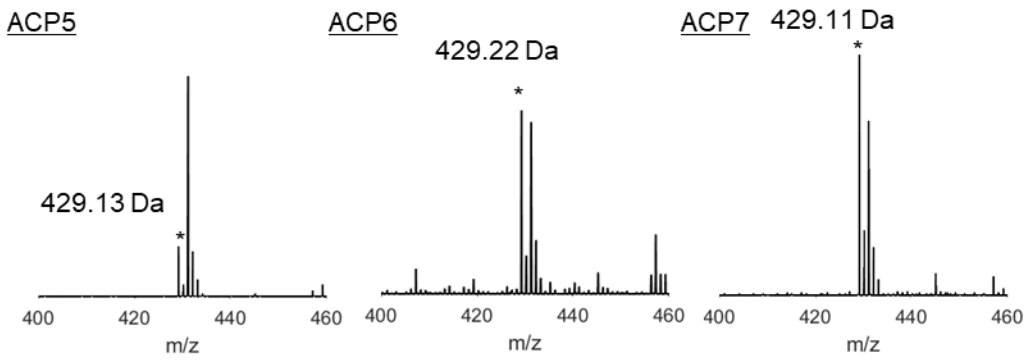


Figure 90: The pantetheine ejection ions of 9HO-ACP5, 9HO-ACP6 and 9HO-ACP7 taken at different time points. A mixture of 429 Da species and 431 Da species is observed in all cases.

The assays described above have shown that MmpB_KS catalysed condensation of **39** can occur with malonyl delivered by any of the individual excised MmpB_ACPs. Next, the ability of mAcpA and mAcpB to delvier malonyl to MmpB_KS will be tested. First, the condensation assay was repeated using malonyl-mAcpA. This assay was initiated by addition of MmpB_KS (20 µM) to **39** (3.3 mM) and malonyl-mAcpA (80 µM) and monitored by ESI-MS after 30 minutes for signs of DCC product; $\mathbf{41_A}$. After this time, no sign of the DCC product was observed in the mAcpA region of the spectrum (Figure 91B). The main species observed was the unreacted malonyl-mAcpA (observed; 12312.5 Da, expected; 12315.0 Da) which is consistent with no condensation reaction having taken place.

Next, the assay was repeated with malonyl-mAcpB. This assay was initiated by the addition of MmpB_KS (20 µM) and **39** (3.3 mM) to malonyl-mAcpB (80 µM). After 30 minutes, again no sign of the condensation product $\mathbf{41_B}$ was observed (Figure 91C). The major species observed was unreacted malonyl-mAcpB (observed; 11798.2 Da, expected; 11798.7 Da). This confirms that condensation also does not occur when malonyl is delivered by mAcpB.

To confirm condensation is still able to take place with this batch of MmpB_KS and over shorter timescales a positive control was set up concurrently. This assay consisted of MmpB_KS (20 µM), **39** (3.3 mM) and malonyl-ACP5 (80 µM). ESI-MS analysis after 30 minutes showed a species with a mass consistent with $\mathbf{41_{ACP5}}$ (observed; 12884.0 Da, expected; 12884.0 Da, Figure 91A) as expected.

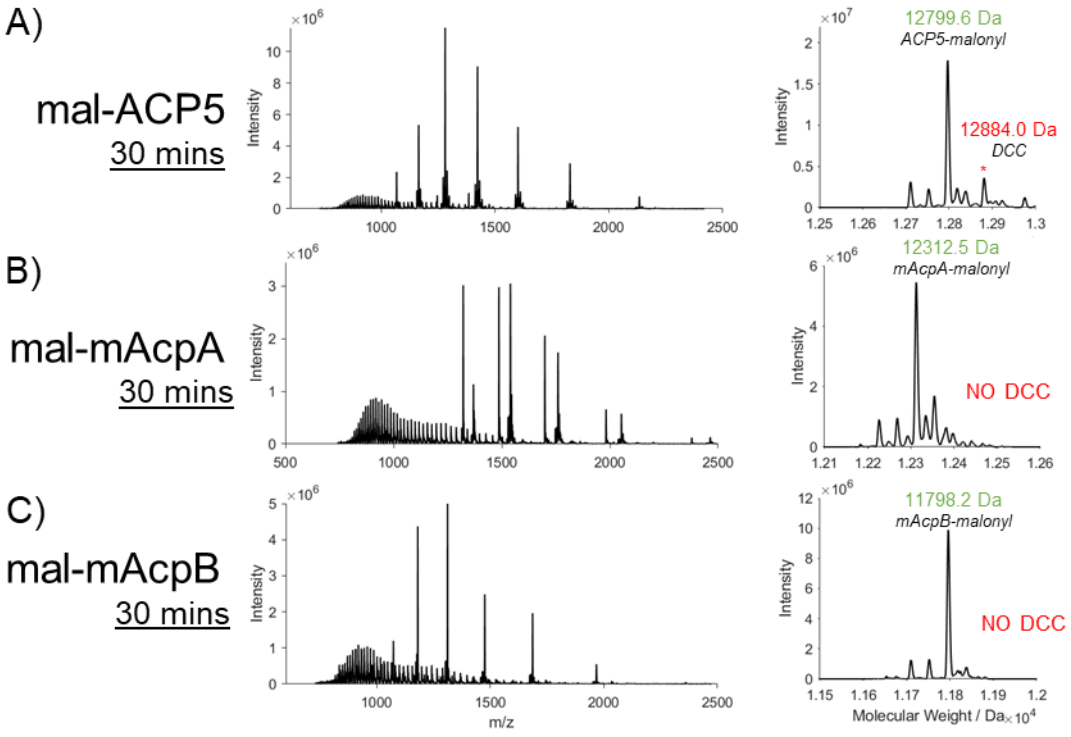


Figure 91: Comparison of malonyl delivery to MmpB_KS by different ACPs. A) Charge states and subsequent deconvolution showing the malonyl-ACP5 reaction after 30 minutes. B) Charge states and subsequent deconvolution showing equivalent reaction with malonyl-mAcpA after 30 minutes. C) Charge states and subsequent deconvolution showing equivalent reaction with malonyl-mAcpA after 30 minutes.

In conclusion, the assays conducted in this section have confirmed that MmpB_KS can catalyse condensation of 7HH to form the corresponding 9-carbon moiety. This reaction has been successful utilising each of the ACP domains from MmpB_ACP; ACP5, ACP6 and ACP7 to deliver malonyl to the KS. The MmpB_KS also appears to select against mAcpA or mAcpB as expected. These results make sense in the overall scheme of the pathway as mAcpA and mAcpB are expected to deliver malonyl to MmpF only.

By utilising specifically designed *in-vitro* assays which reconstitute the biosynthetic pathway of mupirocin, MmpF has been confirmed as the KS that catalyses the first condensation reaction (3-carbon to 5-carbon) and that MmpB_KS will catalyse the final condensation reaction (7-carbon to 9-carbon). The obvious question remains is which KS catalyses the remaining 5-carbon to 7-carbon condensation reaction. Due to the inability of MmpB_KS to translocate 5HP-Pantetheine **37** moiety, the 5-carbon to 7-carbon condensation is expected to be catalysed by MmpF. This will be tested later in this chapter. For now, the MmpB_KS 7-carbon to 9-carbon assays will be repeated with ACP analogues of 7HH. This will provide more evidence about the sequence of events during 9HN biosynthesis.

3.4.6 Formation of ACP-7HH intermediates

A model in which 7HH is elongated by MmpB_KS means that either mAcpA or mAcpB (or both) may transfer this species to the KS. The delivery of 7HH to MmpB_KS was tested by repeating the MmpB_KS catalysed condensation reactions but with 7HH-mAcpA and 7HH-mAcpB. The 7HH-ACP analogues were prepared by addition of MupN (5 µM) to *apo*-mAcpA (100 µM), ATP (1 mM), $MgCl_2$ (10 mM), CoaMix enzymes (3 µM each of CoaA, CoaD and CoaE) and **39** (1 mM). The reaction was monitored after 16 hours by ESI-MS. After this time, a new species was observed with a mass consistent with 7HH-mAcpA (observed; 12354.0 Da, expected; 12357.6 Da, Figure 92A). This new species was produced in 65% yield. Ppant ejection assay of this species produced a single species with the mass of the 7HH Ppant ion (observed; 389.14 Da, expected; 389.21 Da) confirming the formation of 7HH-mAcpA. Minor amounts of *apo*-mAcpA (observed; 11886.7 Da, expected; 11889.5 Da, 20% yield) and *holo*-mAcpA (observed; 12226.8 Da, 10% yield) were observed also. This is consistent with yields observed in previous functionalisation assays involving mAcpA.

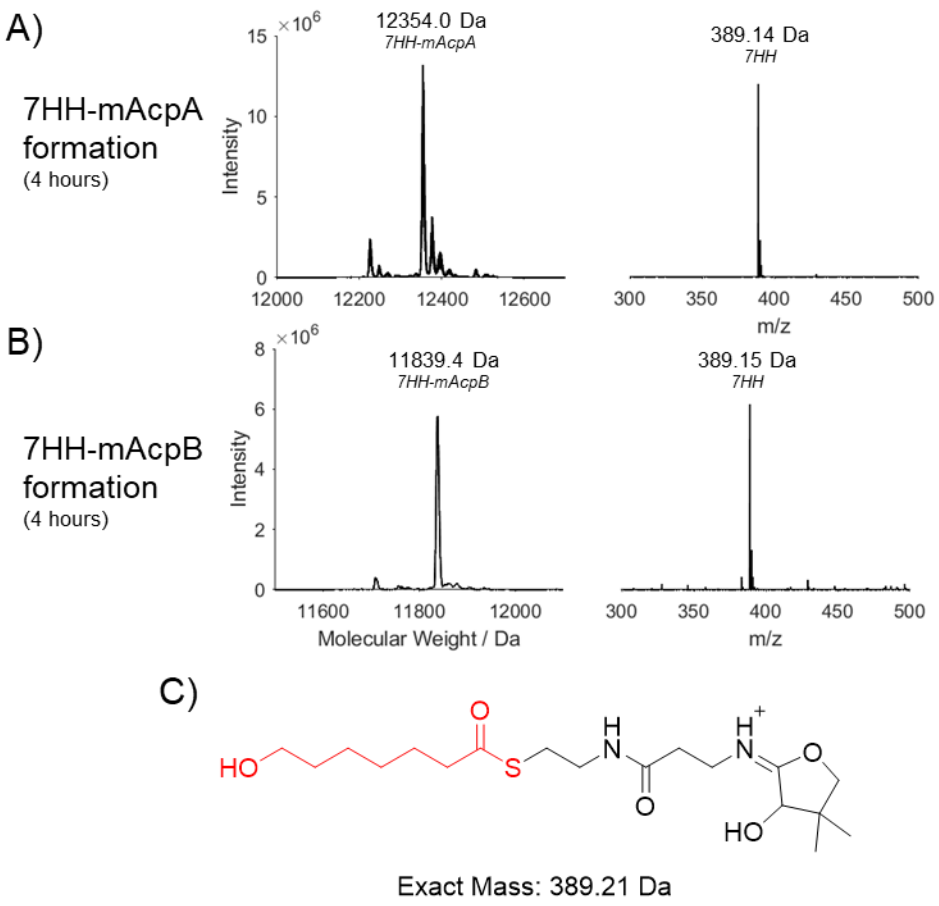
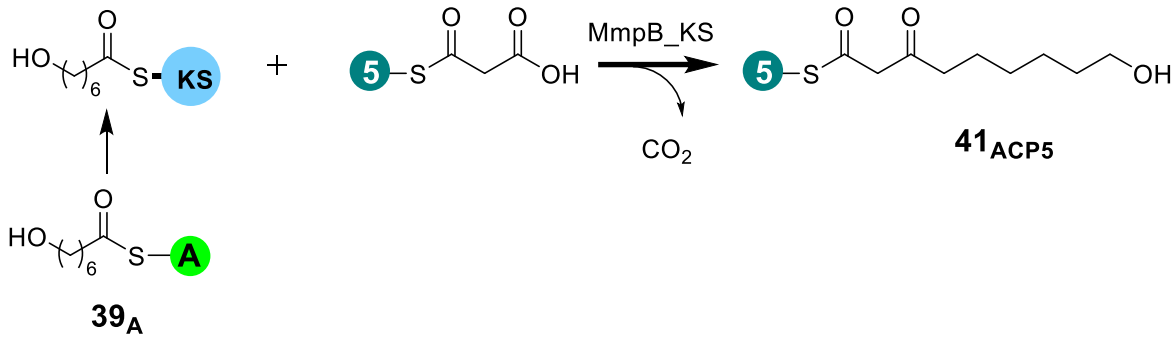


Figure 92: The loading of 7HH on to mAcpA, mAcpB. A) The deconvolution and subsequent Ppant ejection ion 4 hours after initiation of 7HH-mAcpA formation reaction. B) The deconvolution and subsequent Ppant ejection ion 4 hours after initiation of 7HH-mAcpB formation reaction. C) The chemical structure and expected mass of 7HH Ppant ion.

The formation of 7HH-mAcpB was conducted under identical conditions but utilising *apo*-mAcpB. After 4 hours, a species with a mass consistent with 7HH-mAcpB was produced with 100% yield (observed; 11839.4 Da, expected; 11840.8 Da, Figure 92B). Ppant ejection assay of this species produced a characteristic ejection ion of 7HH (observed; 389.15 Da, expected; 389.21 Da). This confirms successful derivatisation of mAcpB with 7HH. 7HH-ACPs were desalted and stored at -20 °C prior to downstream use.

3.4.7 7HH-ACP MmpB_KS condensation assays



Scheme 29: The scheme showing the MmpB_KS condensation reaction using 7HH delivered by mAcpA.

The proposed MmpB_KS condensation reaction with **39$_A$** is shown in Scheme 29. This condensation assay was initiated by addition of MmpB_KS (20 µM) to mAcpA-7HH (40 µM) and Malonyl-ACP5 (40 µM) and monitored by ESI-MS after 2 hours. This revealed ~25% of a new species corresponding to the expected condensation product 9HO-ACP5 **41$_{ACP5}$** (observed; 12880.6 Da, expected; 12884.0 Da, Figure 93A) with concomitant formation of *holo*-mAcpA (observed; 12226.2 Da, expected; 12229.5 Da, 35%). Ppant ejection produced an ejection ion of 431.15 Da (expected; 431.22 Da), confirming **41$_{ACP5}$** formation. Other species were also observed with *holo*-ACP5 (observed; 12710.3 Da, expected; 12713.9 Da, 25%) and unreacted malonyl-ACP5 (observed; 12796.4 Da, expected; 12799.9 Da, 50%). Unreacted 7HH-mAcpA was observed (observed; 12354.2 Da, expected; 12357.6 Da, 30% yield) as well as malonyl-mAcpA (observed; 12312.2 Da, expected; 12315.0 Da, 35% yield) in comparable yield to *holo*-mAcpA. Formation of malonyl-mAcpA might result from transfer of malonyl from ACP5 to the terminal thiol of *holo*-mAcpA generated during the assay. This behaviour has been reported previously for type II aromatic PKS ACPs and may explain the observed.[169] Overall, this data shows that mAcpA can successfully deliver 7HH to MmpB_KS and facilitate condensation.

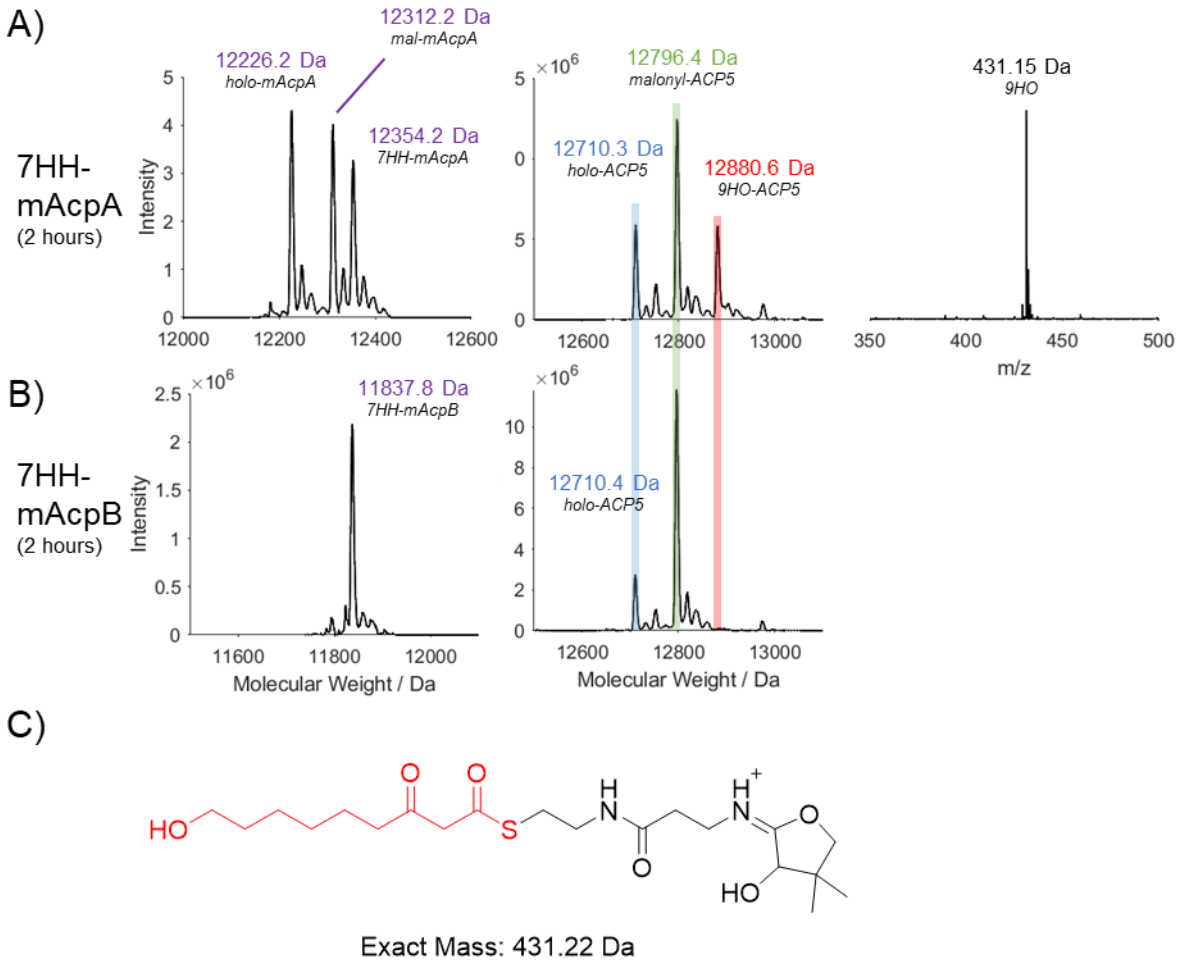


Figure 93: A figure showing the 7HH-ACP + MmpB_KS condensation assays. All spectra recorded 2 hours after initiation of respective assays. Malonyl-ACP is shown in green, holo-ACP shown in blue and condensation intermediate is shown in red. A) Deconvolution of mAcpA region of spectrum (left), ACP5 region of spectrum (middle) and Ppant ejection spectrum showing fragmentation of 9HO-ACP5 species (right). B) Deconvolution of mAcpB region (left) and ACP5 region of spectrum (middle). C) Chemical structure and expected mass of 9HO Ppant ion.

When repeated with 7HH-mAcpB, the equivalent reaction showed no sign of condensation product **41**$_{\mathbf{ACP5}}$. Instead, the predominant species observed was unreacted malonyl-ACP5 (observed; 12796.6 Da, expected; 12799.9 Da, 85% yield). A minor amount of *holo*-ACP5 was also observed (observed; 12710.4 Da, expected; 12713.9 Da, 15% yield) but this was consistent with amounts observed prior to initiation of reaction (see Figure 85A). Analysis of the mAcpB region of the spectrum showed 100% of unreacted 7HH-mAcpB remained (observed; 11837.8 Da, expected; 11840.8 Da).

The above results show that mAcpA can successfully deliver 7HH to MmpB_KS to facilitate condensation reaction. Surprisingly, when 7HH is delivered by mAcpB, condensation is not achieved. This assay was repeated multiple times with different batches of mAcpB and the same result was observed. The finding that only mAcpA can deliver 7HH to MmpB_KS may point towards only mAcpA delivering the chain extending malonate during the second condensation reaction (5-carbon to 7-

carbon). This will be tested in later sections by conduction of condensation reaction with 5HP-Pant **37** substrate.

3.4.8 MmpB_KS Condensation with cognate malonyl-ACP and 6-hydroxyhexanoate

Self-loading assays conducted with MmpB_KS and 6HH-Pant **38** showed that binding of 6HH occurs after 4 hours. To test whether MmpB_KS can catalyse condensation with **38** *in vitro* assays were designed. The condensation reaction was conducted using malonyl delivered by cognate ACP5. The reaction was initiated by addition of MmpB_KS (20 µM) to **38** (3.3 mM) and malonyl-ACP5 (80 µM) and monitored by ESI-MS after 3 hours. After this time, the ACP5 region of the spectrum showed that the major species observed was unreacted malonyl-ACP5 (observed; 12796.6 Da, expected; 12799.9 Da) as shown in Figure 94A. This result confirms that no condensation occurs with non-natural **38**. A positive control reaction was conducted concurrently with MmpB_KS, **39** and malonyl-ACP5. This showed formation of $\mathbf{41}_{\mathbf{ACP5}}$. This confirms that lack of condensation with **38** is due to substrate incompatibility.

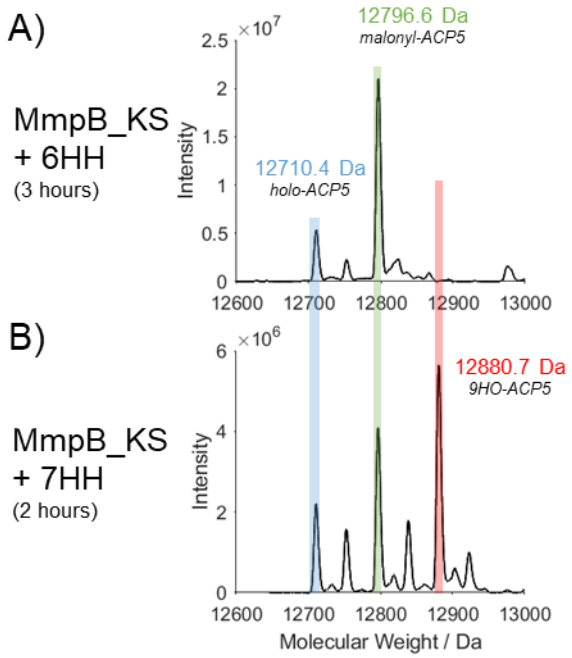
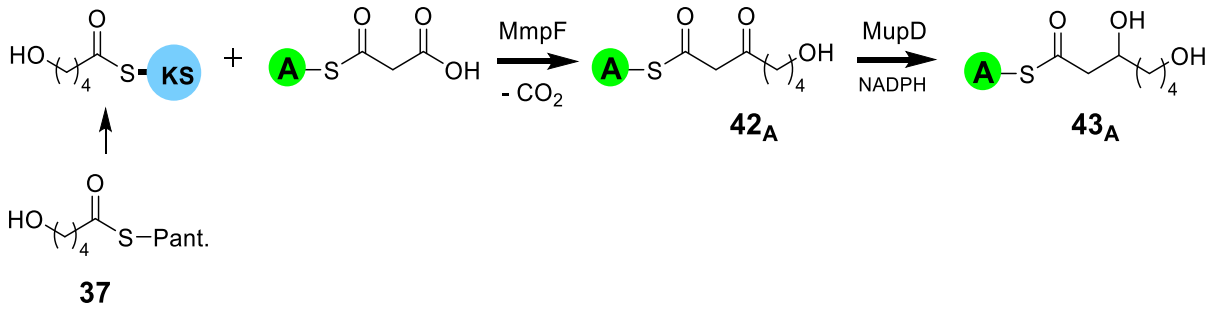


Figure 94: The condensation of 6HH using MmpB_KS and malonyl-ACP. A) Deconvolution 3 hours after initiation of 6HH condensation reaction. B) Positive control reaction conducted with 7HH and MmpB_KS and malonyl-ACP5.

The ability of MmpB_KS to bind 6HH but not catalyse subsequent condensation is not entirely surprising. Analysis of KS-catalysed condensation reactions in the literature shows that this is expected. There are two stages to a KS-catalysed condensation reaction, the binding of the extender unit (translocation) and subsequent elongation. Previous studies have shown that higher specificity exists for the elongation step rather than initial binding of the acyl moiety. [170] Therefore, MmpB_KS may be able to bind **38** but subsequent elongation does not occur.

3.4.9 Identity of KS domain which catalyses 5HP condensation

Despite the ability of both mAcpA and mAcpB to participate in the first round of fatty acid chain extension, only mAcpA is able to transfer the 7HH group to the MmpB_KS for the third and final elongation. By definition this would also suggest that only mAcpA should participate in the second round of chain elongation, so it correctly bears the 7HH moiety. Therefore, to further explore the roles of mAcpA and mAcpB, *in vitro* assays with 5HP-Pant **37**, MmpF and malonyl-ACPs were conducted. MupD was included in this reaction to facilitate the condensation reaction.



Scheme 30: A scheme showing the condensation of 5HP with malonyl-mAcpA catalysed by MmpF in the presence of MupD.

Starting with **37** (Scheme 30), an assay was initiated by addition of MmpF (10 µM) to malonyl-mAcpA (80 µM), **37** (3.3 mM), MupD (15 µM) and NADPH (1 mM). After 3 hours, starting material malonyl-mAcpA was consumed and a new species with a mass consistent with 7,3-dihydroxyheptanyl-mAcpA **43$_A$** (DHH-mAcpA) was observed (observed; 12372.6 Da, expected; 12373.6 Da, 35% conversion, Figure 95A). Pantetheine ejection of this species generated a 405.23 Da ion that confirmed **43$_A$** had been produced (expected mass; 405.21 Da). In addition to this species, *holo*-mAcpA (observed; 12228.2 Da, 35%) and *holo*-mAcpA with pantetheine adduct (observed; 12504.1 Da, 15%) were also observed. A further species with a mass of 12358.8 Da was also observed. This species is unique to the MmpF condensation reaction using mAcpA delivered malonyl and has been previously observed in the 3-carbon to 5-carbon condensation reaction. This is 3,5-dioxopentanoyl-mAcpA (DOP-mAcpA) **20$_A$** species, appeared to be formed from condensation of two malonyl units as is discussed in previous sections.

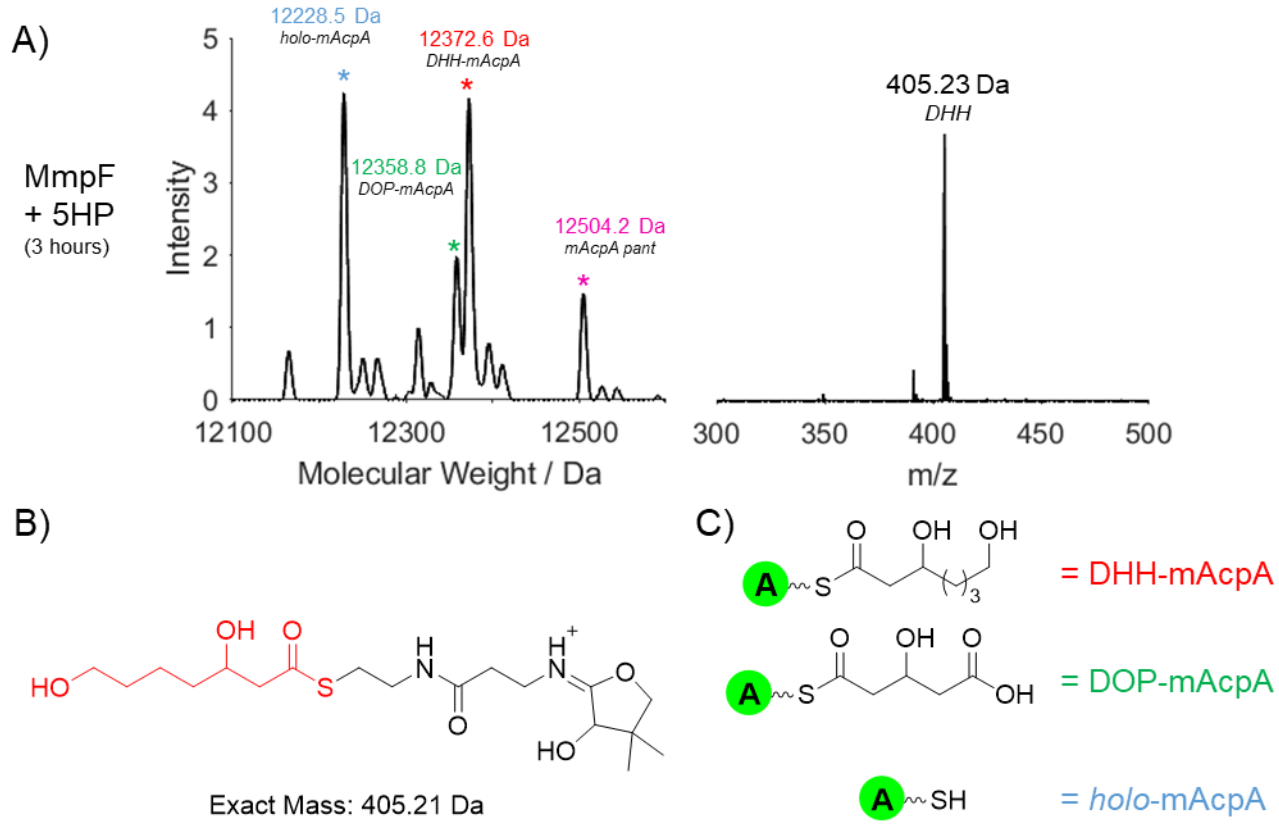


Figure 95: The condensation of 5HP-Pant catalysed by MmpF with malonyl-mAcpA, MupD and NADPH after 3 hours. A) The deconvolution of the mAcpA region of the spectrum (left). The Ppant ejection spectrum after fragmentation of 12372.6 Da species (right). B) The chemical structure and expected mass of DHH-mAcpA Ppant ion. C) The chemical structures of some of the mAcpA bound species observed.

A negative control reaction was set up under identical conditions but using MmpF_C183A mutant instead of MmpF. Reaction was inititated by addition of MmpF_C183A (10 µM) to malonyl-mAcpA (80 µM) and **37** (3.3 mM) in the presence of MupD (15 µM) and NADPH (1 mM). After 3 hours, the malonyl-mAcpA species was observed in small amounts (10% yield, Figure 96A). The major species observed were *holo*-mAcpA (observed; 12228.2 Da, 30% yield), acetyl-mAcpA (12269.8 Da, 30% yield) and the pantetheine adduct of *holo*-mAcpA (12504.2 Da, 30% yield). No sign of $\mathbf{43_A}$ (or $\mathbf{20_A}$) was observed which confirms that MmpF_C183A is unable to catalyse condensation of **37**.

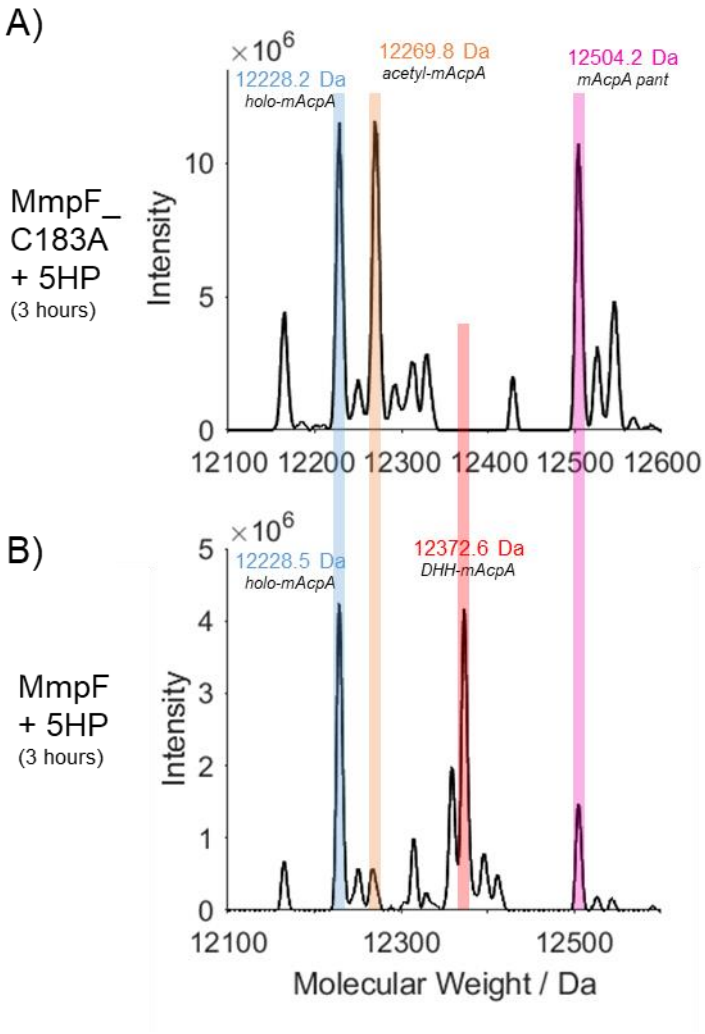
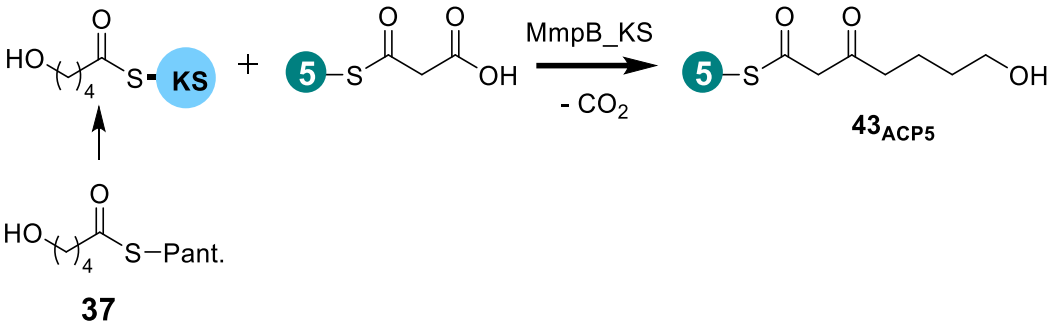


Figure 96: Comparison of positive and negative control reactions of the MmpF, condensation reaction with 5HP substrate after 3 hours. A) The deconvolution of mAcpA region of the negative control reaction involving MmpF_C183A, malonyl-mAcpA, 5HP, MupD and NADPH. B) For comparison, the deconvolution of the positive control reaction utilising functional MmpF.

These above assays have confirmed that MmpF can undergo condensation of 5HP moiety utilising malonyl delivered by mAcpA. The 5HP condensation assay with repeated with MmpB_KS and its cognate malonyl-ACP (malonyl-ACP5). Successful condensation is shown in Scheme 31 and would lead to the formation of **43**$_{\mathbf{ACP5}}$ species.



Scheme 31: A scheme showing the condensation of 5HP-Pant catalysed by MmpB_KS.

The assay was initiated by addition of MmpB_KS (10 µM) to malonyl-ACP5 (80 µM) and **37** (3.3 mM) and reaction monitored by ESI-MS after 3 hours. After this time, unreacted malonyl-ACP5 remained the major species (observed; 12799.8 Da, expected; 12798.7 Da, 40% yield) (Figure 97) with no sign of **43**$_{ACP5}$ observed. This is consistent with no condensation having taken place. Further ACP5 species were also observed with mass increases corresponding to +276.1 Da (observed; 12988.5 Da, 30% yield) which is pantetheine adduct of ACP5 which has been described previously. The lack of condensation is consistent with MmpB_KS only catalysing the 7-carbon to 9-carbon condensation reaction (as expected).

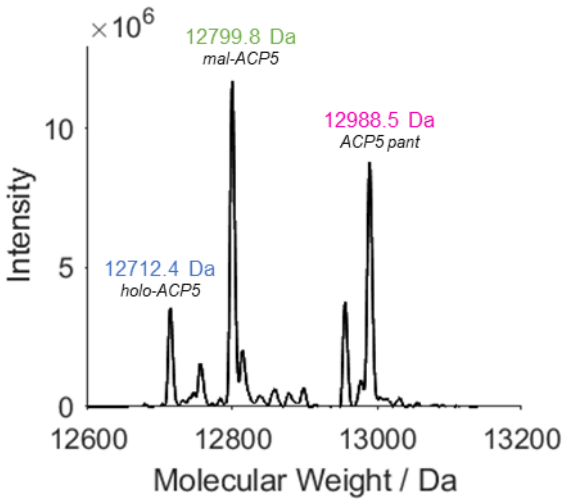
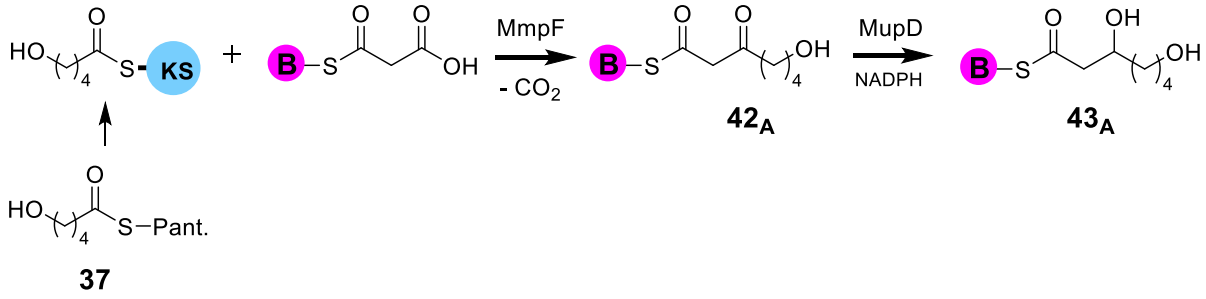


Figure 97: The addition of MmpB_KS (10 µM), 5HP-Pant (3.3 mM) and mal-ACP5 (80 µM) after 3 hours. This figure shows the deconvolution of the ACP5 region of the spectrum.

3.4.10 Condensation of MmpF with 5-hydroxypentanoyl pantetheine using Malonyl-mAcpB



*Scheme 32: A scheme showing the condensation of 5HP **37** with malonyl-mAcpB catalysed by MmpF in the presence of MupD.*

Finally, the MmpF condensation reaction with **37** was repeated using malonyl delivered by mAcpB (Scheme 32). The reaction was initiated by addition of MmpF to malonyl-mAcpB (80 µM) and **37** (3.3 mM) in the presence of MupD (15 µM) and NADPH (1 mM). After 3 hours, the main species observed was unreacted malonyl-mAcpB (observed; 11798.8 Da, expected; 11798.7 Da, 60% yield, Figure 98). Approximately 20 % *holo*-mAcpB was observed (observed; 11713.3 Da) as well as the pantetheine adduct of *holo*-mAcpB (observed; 11989.6 Da, 20%). No signal corresponding to the condensation intermediate $\mathbf{43_B}$ could be detected after 3 hours.

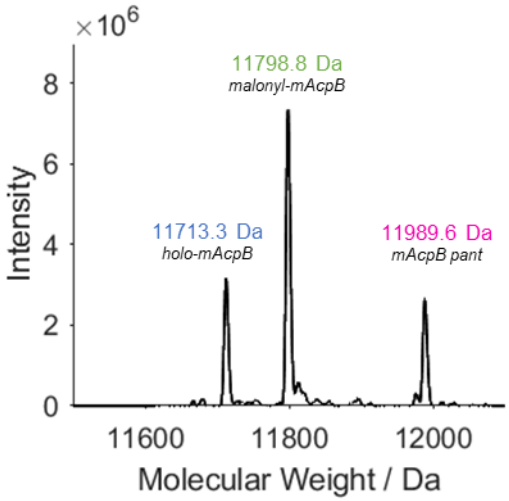


Figure 98: The MmpF + 5HP + malonyl-mAcpB reaction after 3 hours. The deconvolution of the mAcpB region of the spectrum is shown.

The assays conducted in this section have shown that only MmpF can catalyse the 5-carbon to 7-carbon condensation reaction (DCC #2). Further to this, only malonyl-mAcpA has been shown to turnover with MmpF. No condensation is observed with equivalent malonyl-mAcpB reaction. This tells us two things. Firstly, that these reactions are strictly controlled by the ACPs – no reaction is observed with non-cognate ACPs. Secondly, this confirms the sequence of condensation events during 9HN biosynthesis.

MmpF must catalyse the first and second condensation reactions and the final step will be catalysed by MmpB_KS. Following the assays conducted in this section, the selectivity of each of mAcpA and mAcpB to deliver 5HP to MmpF will be tested through *in-vitro* assays. This will provide further information about which ACP will carry 5HP during the second condensation reaction.

3.4.11 Condensation reaction with 5HP-ACP substrates

In light of the apparent activity of mAcpB in only DCC #1, the obvious question was what does this ACP do next when loaded with 5HP species? To test the ability of mAcpB (and mAcpA) to transfer **37** back to MmpF and initiate DCC #2, a series of ACP analogues derivatised with 5HP were constructed. Assay conditions were as follows, *apo*-ACP (120 µM), MupN (5 µM), 5HP-Pantetheine (1 mM), ATP (1 mM), $MgCl_2$ (10 mM) and CoAMix (3 µM each). For the mAcpB 5HP transfer reaction, after 4 hours a species consistent with 5HP-mAcpB $\mathbf{37_B}$ was formed (observed; 11811.8 Da, expected; 11812.8 Da, Figure 99A). Ppant ejection produced an ion with a mass of 361.18 Da, confirming formation of 5HP-mAcpB (expected; 361.18 Da).

The formation of 5HP-mAcpA $\mathbf{37_A}$ was conducted under identical conditions with using *apo*-mAcpA but yielded very minor amounts of derivatised ACP after 4 hours. After 16 hours, 30% $\mathbf{37_A}$ was produced (observed; 12328.7 Da, expected; 12329.5 Da) and confirmed by Ppant ejection (observed 361.18 Da; expected 361.18 Da). *apo*-mAcpA and *holo*-mAcpA were also present (~ 35% each). Optimisation of reaction conditions (addition of extra MupN, addition of extra CoAMix Enzymes, addition of extra cofactors) were all unsuccessful in producing greater yields of $\mathbf{37_A}$.

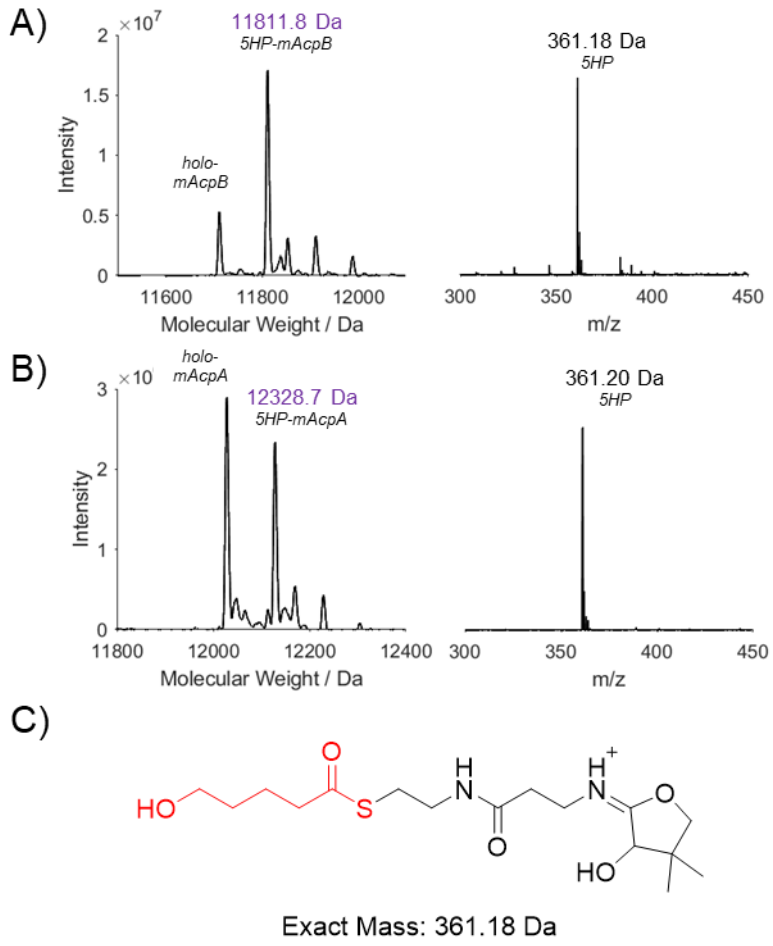
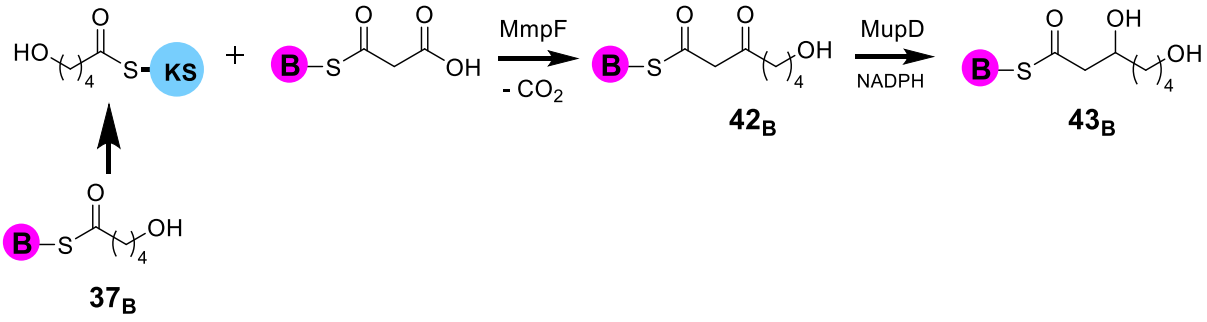


Figure 99: The formation of 5HP-ACPs. A) The deconvolution (left) and subsequent Ppant ejection spectrum (right) 4 hours after initiation of 5HP-mAcpB formation reaction. B) The deconvolution (left) and subsequent Ppant ejection spectrum (right) 4 hours after initiation of 5HP-mAcpA formation reaction. C) The chemical structure and expected mass of the Ppant ejection ion.

3.4.12 Condensation reaction with 5HP-ACP, MmpF and malonyl-mAcpB



Scheme 33: The MmpF + 5HP condensation reaction utilising 5HP delivered by mAcpB

The condensation reaction was first conducted with 5HP-mAcpB **37$_B$**. The reaction was initiated by addition of MmpF (10 µM) to **37$_B$** (40 µM), malonyl-mAcpA (40 µM), MupD (15 µM) and NADPH (1 mM). After 3 hours the reaction was monitored by ESI-MS but there was no indication of condensation products **42$_B$** or **43$_B$** (Figure 100). The major species were *holo*-mAcpA (observed;

12229.0 Da) and unreacted malonyl-mAcpA (observed; 12314.9 Da). This is consistent with no condensation having taken place.

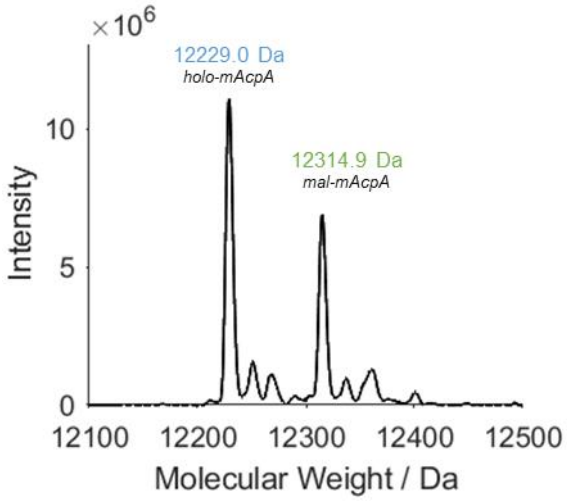
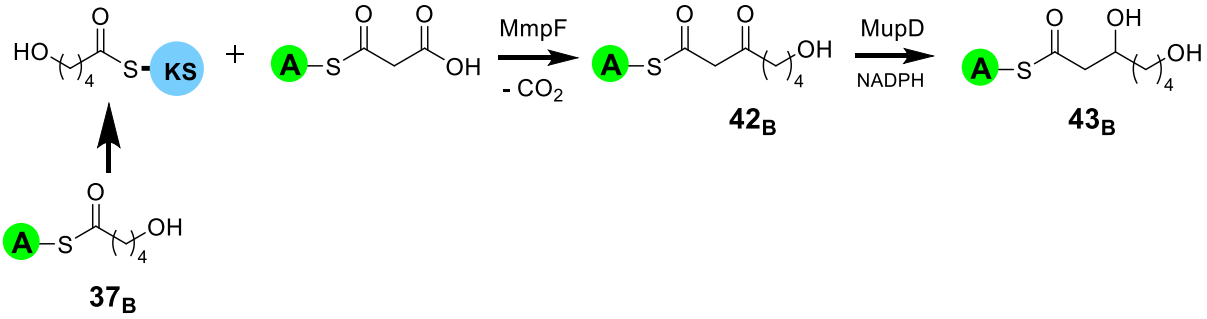


Figure 100: The addition of 5HP-mAcpB to MmpF in the presence of malonyl-mAcpA, MupD and NADPH after 3 hours. The deconvolution of the mAcpA region of the spectrum is shown.

3.4.13 Condensation reaction with 5HP-ACP, MmpF and malonyl-mAcpA



Scheme 34: The MmpF + 5HP condensation reaction utilising 5HP delivered by mAcpA

The assay was repeated under identical conditions using **37$_A$**. As the assay description reaction states, this batch of **37$_A$** also contains *apo*-mAcpA and *holo*-mAcpA which is also present in downstream spectra. This mixture of species means it is likely that the actual concentration of **37$_A$** is lower than the 40 µM added to this assay. 3 hours after the initiation of the condensation reaction, analysis of mAcpA region of the spectrum showed consumption of **37$_A$** and production of a new species with a mass consistent with condensation product **43$_A$** (observed; 12372.6 Da, expected; 12373.6 Da, Figure 101A). The yield of condensation product was low (~ 6%) but successful Ppant ejection confirmed formation of **43$_A$** (observed 405.24 Da; expected mass; 405.21 Da, Figure 101C). *holo*-mAcpA (observed; 12228.5 Da) and unreacted malonyl-mAcpA (observed; 12314.5 Da) made up the majority of the mixture. In addition, 6 % of 5HP-mAcpA (observed; 12329.5 Da) and 10% of DOP-mAcpA **20$_A$** product was also

produced in 10% yield (observed; 12358.4 Da, as described previously). The small amount of condensation product formed may be due to the relatively small amount of 5HP-mAcpA mixture present (40 µM) when compared to the equivalent 5HP-Pant reaction (3.3 mM of substrate used).

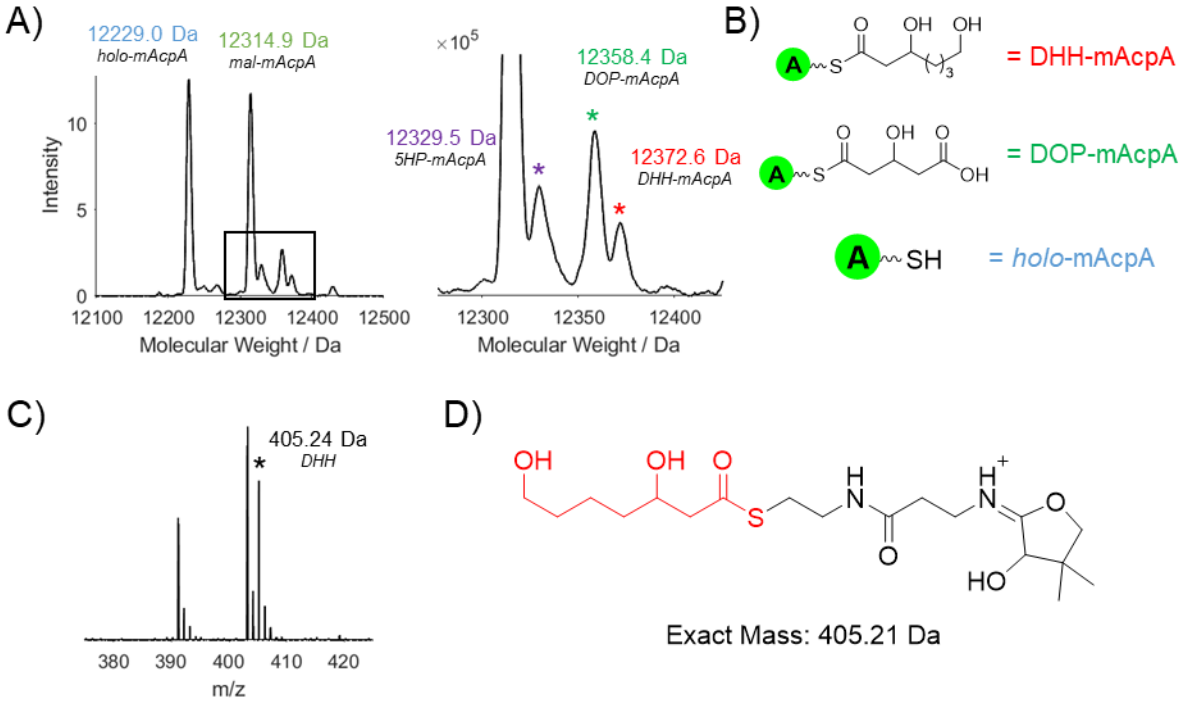


Figure 101: The addition of 5HP-mAcpA to MmpF and malonyl-mAcpA (in the presence of MupD and NADPH) after 3 hours. A) Deconvolution of mAcpA region of the spectrum showing formation of condensation product (left). Zoomed in to region within black square (right). B) A key showing the identity of the abbreviated species. DHH-mAcpA is the condensation product and is labelled with a red asterisk. C) Ppant ejection spectrum of DHH-mAcpA. D) Chemical structure and expected mass of DHH-mAcpA Ppant ion.

These assays provide evidence that only mAcpA can deliver 5HP to MmpF to catalyse the 5-carbon to 7-carbon condensation. The equivalent assay involving mAcpB delivered 5HP showed no turnover under identical conditions. mAcpB appears to only take part in the first condensation reaction. There are many possibilities for this unusual finding. The initial condensation step may be rate limiting and a second ACP may increase metabolic flux through this part of the pathway. This model however would require transfer of the 5HP from mAcpB back to mAcpA. Further kinetic characterisation of these condensation reactions should be conducted to ascertain this. If this is true, overproduction of MmpF and/or mAcpA may provide an interesting avenue to increase the overall production of mupirocin within this BGC.

The above model would suggest that mAcpA and mAcpB act in parallel for the first condensation reaction. This is surprising due to the low similarity between the two ACP domains. mAcpA and mAcpB both share a similarity of only 22.5% identity. This is unusual for tandem ACP domains which typically have much greater similarity. [171] There do exist some examples of tandem ACPs with lower

percent similarity such as the di-domain ACP from the leinamycin biosynthetic pathway.[172] This di-domain ACP is found within the lnmJ gene and both ACP domains only share 20% sequence identity. The expected reason for this sequence divergence is that each ACP of the didomain is expected to interact with a different protein partner, although this is yet to be proven experimentally. Therefore, the low sequence homology of mAcpA and mAcpB as di-domains is not unprecedented.

If mAcpB is recruited for the first condensation reaction, but not the second condensation then a transfer mechanism to remove the acyl intermediate from mAcpB is required. There would presumably be two ways this transfer reaction could occur. The first scenario would involve spontaneous transfer of the acyl moiety from 5HP-mAcpB to *holo*-mAcpA. This has not been directly tested in this work, although incubation of 5HP-mAcpB in the presence of a mixture of *holo*-mAcpA and malonyl-mAcpA (in the presence of MmpF, MupD and NADPH) has been tested. This showed no formation of 5HP-mAcpA providing tentative evidence that spontaneous transfer does not take place.

The second possibility is that another protein from the BGC of mupirocin may be responsible for the transfer of 5HP moiety from mAcpB to mAcpA. Although most of the enzymes within the BGC of mupirocin now have assigned functions, a small number remain without a role. One such enzyme is MupL which is a member of the alpha/beta hydrolase superfamily of proteins.[173] This superfamily of proteins has a diverse range of catalytic activities.[174] It is therefore possible that MupL is responsible for the transfer of the 5-carbon acyl group from mAcpB to mAcpA prior to the second condensation reaction. Future work should seek to test this by expressing and purifying this protein and conducting *in vitro* assays.

4 Conclusions and Future Work

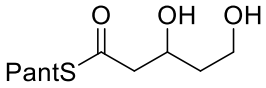
The results of the assays conducted in this work allow a hypothesis to be presented regarding the precise chemical steps of 9HN construction. The work done in this project has involved the use of a variety of molecular biology techniques to monitor the 9HN biosynthetic pathway *in vitro*. The KS enzymes, ACPs and subsequent tailoring enzymes expected to be utilised during 9HN biosynthesis were all cloned, expressed, and purified. This led to the reconstitution of 9HN biosynthetic pathway and its interrogation using a series of rationally designed assays. These assays have involved the use of both ACP and non-ACP bound substrates to extract information about the role of each biosynthetic enzyme within the pathway. The majority of the work completed has relied on analysis of *in vitro* experiments by protein mass spectrometry. This technique has proven to be a very robust at analysing ACP bound species when coupled with the tandem mass spectrometry (pantetheine ejection assay) which allows high mass accuracy analysis of the terminal thiol of the ACP species. The complementation of ESI-MS with ^{15}N-HSQC NMR titrations has allowed for a hypothesis for the biosynthesis of the key steps of 9HN biosynthesis to be proposed which will be outlined below.

The genesis of this pathway involves the production of a 3HP moiety from malonyl CoA utilising MupQ, MupS and mAcpD.[86] Following this, a series of ^{15}N-HSQC NMR titrations and *in vitro* assays have confirmed that MmpF catalyses the first chain extending condensation reaction utilising ACP delivered malonyl. Experimental assays have shown that the first condensation step can take place with malonyl delivered by either mAcpA or mAcpB. Following this, MupD was confirmed as the KR enzyme responsible for downstream tailoring of this nascent acyl intermediate. The role of MupD as a KR enzyme had been previously reported in the literature but was yet to be experimentally proven.[48] Like its predecessor, MupD was shown to turnover with both mAcpA and mAcpB bound substrates thereby confirming its role. By utilising 4-carbon and 5-carbon mimics of the natural substrate, MupD was also shown to have a broad substrate specificity. The ACP selectivity of this enzyme was found to be tighter, with successful reduction occurring with mAcpB bound substrates but not with non-cognate mAcpD.

Further work on enzymes proposed to tailor the outgoing MmpF condensed acyl group was conducted. MupE was confirmed to have ER activity due to turnover with a 4-carbon crotonyl mimic of the natural substrate attached to mAcpB. Equivalent assay with 4-carbon substrate containing a β-methyl branch was unsuccessful in turning over under identical conditions, pointing to MupE not tolerating this substrate. Again, strong ACP selectivity was observed, reduced turnover shown with crotonyl moiety attached to non-cognate mAcpD. Future work should seek to confirm the ability of MupE to turnover with mAcpA bound substrates. This was attempted in this work but was foiled by an inability of mAcpA to accept the crotonyl moiety consistently. Further structural characterisation of MupE by X-ray crystallography would provide more evidence for the ACP and substrate selectivity of this enzyme.

Further to this, not much is known about the structure of MupE and its homologues – MupE sits within a subclass of Medium Chain Dehydrogenase enzymes which contains at least 47 proteins (MDR20 – see Persson *et al*), none of which have been structurally characterised.[175] This provides more impetus to structurally characterise MupE.

Following this, attempts were made to discover the identity of the missing DH enzyme from the tailoring cycle. Two DH enzymes (MmpB_DH and MmpD_DH1) were assayed for their ability to act as the required DH with natural substrate, but both enzymes remained unreactive. Future work should focus on the production of a synthetic version of the natural substrate which will allow this assay to be conducted without the need for multiple enzymatic components further simplifying the assay and therefore the MS spectra obtained. A synthetic mimic such as 3,5-dihydroxypentanyl pantetheine which could be subsequently transferred to the relevant ACP would be suffice.



3,5-dihydroxypentanyl pantetheine

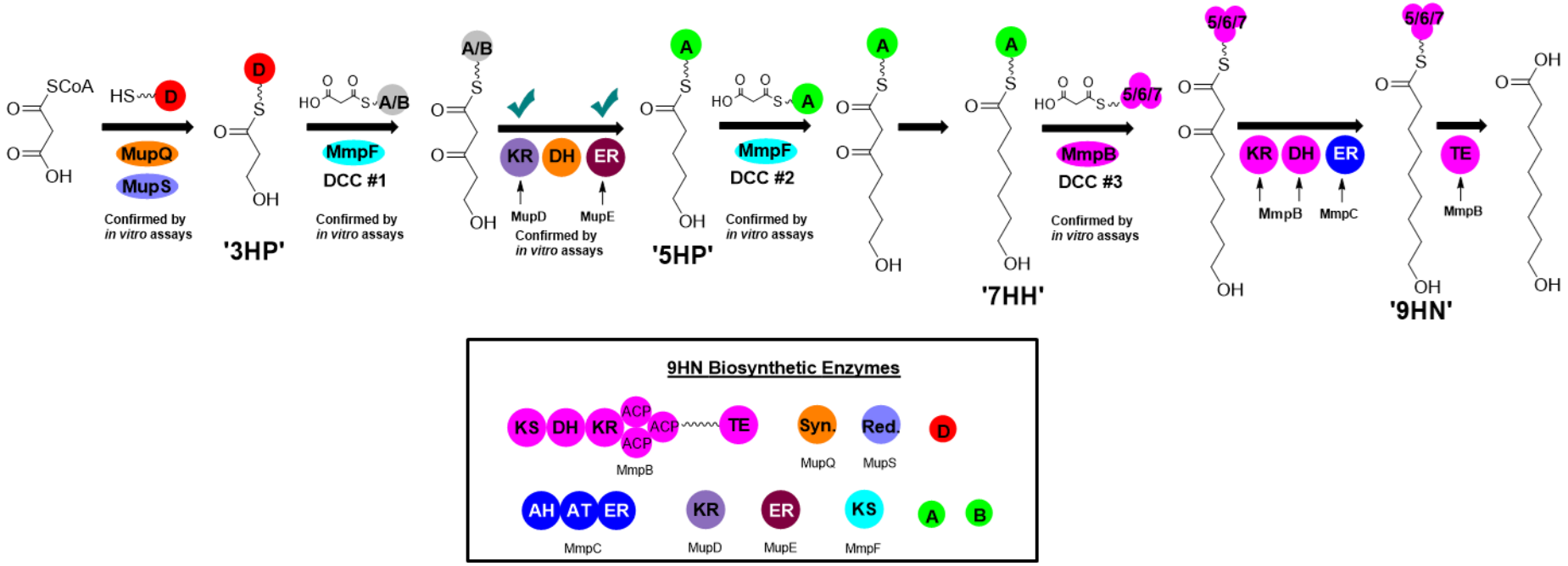
In addition, further work should seek to understand why mupirocin and other BGC seem to lack intrinsic DH domains. A series of DH domains from primary metabolism (such as FabA and FabZ homologues) should be prepared and their ability to provide the required DH activity tested.

Next in this work, the KS enzyme MmpB_KS was assayed for its ability to acylate a range fatty acid pantetheine substrates of increasing chain length. 6-carbon and 7-carbon substrates were found to acylate most efficiently providing evidence that MmpB_KS may catalyse the final (7-carbon to 9-carbon) condensation reaction in 9HN biosynthesis. This was confirmed by *in vitro* assays with MmpB_KS and 7HH-Pant utilising malonyl delivered by cognate ACPs MmpB_ACP5/6/7. mAcpA and mAcpB were found to be unable to deliver malonate during this elongation reaction confirming that specific ACP-KS interaction is key for turnover of condensation. The ability of ACPs to deliver 7HH to MmpB_KS was tested to ascertain which ACP holds the acyl chain prior to final condensation. mAcpA was found to deliver 7HH to MmpB_KS but no condensation product was observed for equivalent reaction with mAcpB delivered 7HH. This suggested that only mAcpA may deliver 7HH to MmpB_KS.

Finally, the ability of MmpB_KS and MmpF to catalyse the remaining condensation reaction with a 5HP substrate and their cognate ACPs was conducted. This showed that only MmpF with malonyl-mAcpA was able to catalyse condensation. This provided further proof that two rounds of condensation are catalysed by MmpF followed by the final condensation occurring with MmpB_KS as is shown in Scheme 35. This means that only mAcpA will hold the acyl chain after the 5-carbon to 7-carbon condensation reaction prior to transfer to MmpB_KS. The ACP selectivity of acylation step with 5HP

was tested. Only mAcpA was found to deliver 5HP to MmpF to facilitate condensation. Equivalent reaction with 5HP-mAcpB showed no condensation product after the same period of time. This provides further evidence that mAcpB does not deliver malonyl to the second condensation reaction.

In future work, the KS domain of MmpF should be structurally determined using X-ray crystallography. This structure should be used for a basis for point mutations to ascertain the nature of the MmpF-ACP selectivity observed in this work. The use of a chloroacryl probe has previously been used covalently attach ACPs to KS enzymes by exploiting the nucleophilicity of the active site cysteine of the KS.[176] This has been shown to allow the crystallisation of ACP and KS di-domains by X-ray crystallography. [177] Crystallographic studies of di-domains of MmpF-mAcpA, MmpF-mAcpB and MmpF-mAcpD using this chloroacryl probe may provide more evidence about the binding interfaces of each domain. This should be complemented with further with ^{15}N-HSQC titrations involving MmpF and each ACP to further understand the nature of these interactions.



Scheme 35: The scheme showing biosynthesis of 9HN utilising three multidomain enzymes (MmpF, MmpB and MmpC), four individual catalytic domains (MupQ, MupS, MupD and MupE) and three individual ACP domains (mAcpA, mAcpB and mAcpD) labelled A, B and D respectively.

4.1 Insights into thiomarinol fatty acid construction

The insights gained from studying the biosynthetic pathway of mupirocin can be applied to related biosynthetic clusters. One such BGC has been isolated from *Pseudoalteromonas* and is responsible for construction of the thiomarinol antibiotics.[85] As has been previously discussed, the major metabolite, thiomarinol A has a very similar structure to pseudomonic acid A. Both compounds contain fatty acid regions of slightly different chain lengths. In thiomarinol, the fatty acid region is 8-carbons long compared to the 9-carbon chain found within mupirocin. The BGC of thiomarinol has been sequenced and is very similar to the mupirocin BGC. [56] Therefore, conclusions reached about fatty acid construction in mupirocin can be applied to construction of 8-hydroxyoctanoate (8HO) moiety in thiomarinol.

The thiomarinol gene cluster contains many homologues of proteins from the mupirocin BGC. Many thiomarinol proteins are homologues of those implicated in 9HN construction within mupirocin including: tAcpA, tAcpB, tAcpD, TmlQ, TmlS, TmpF, TmpB, TmlL and TmpC (Tmp = thiomarinol multifunctional protein). The domain architecture of TmpB is also slightly different compared to MmpB (Figure 102). The first module of TmpB matches that observed in MmpB but with a quad-ACP as opposed to the tridomain-ACP observed in MmpB. TmpB also contains an extra module consisting of a KS^0-ACP bound to a TE domain. The notable absences from thiomarinol BGC include TmlD and TmlE, homologues of MupD and MupE respectively.

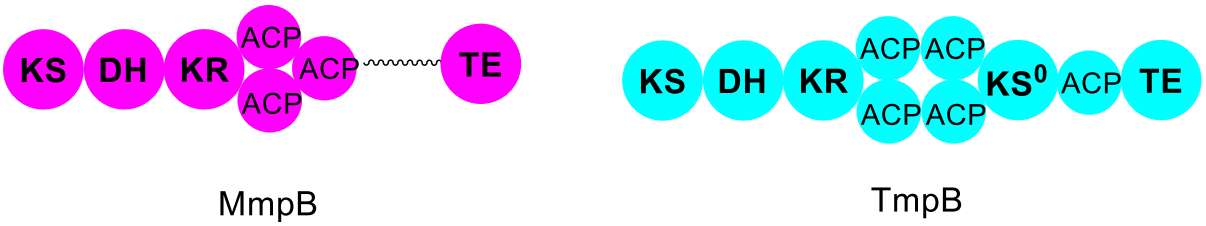
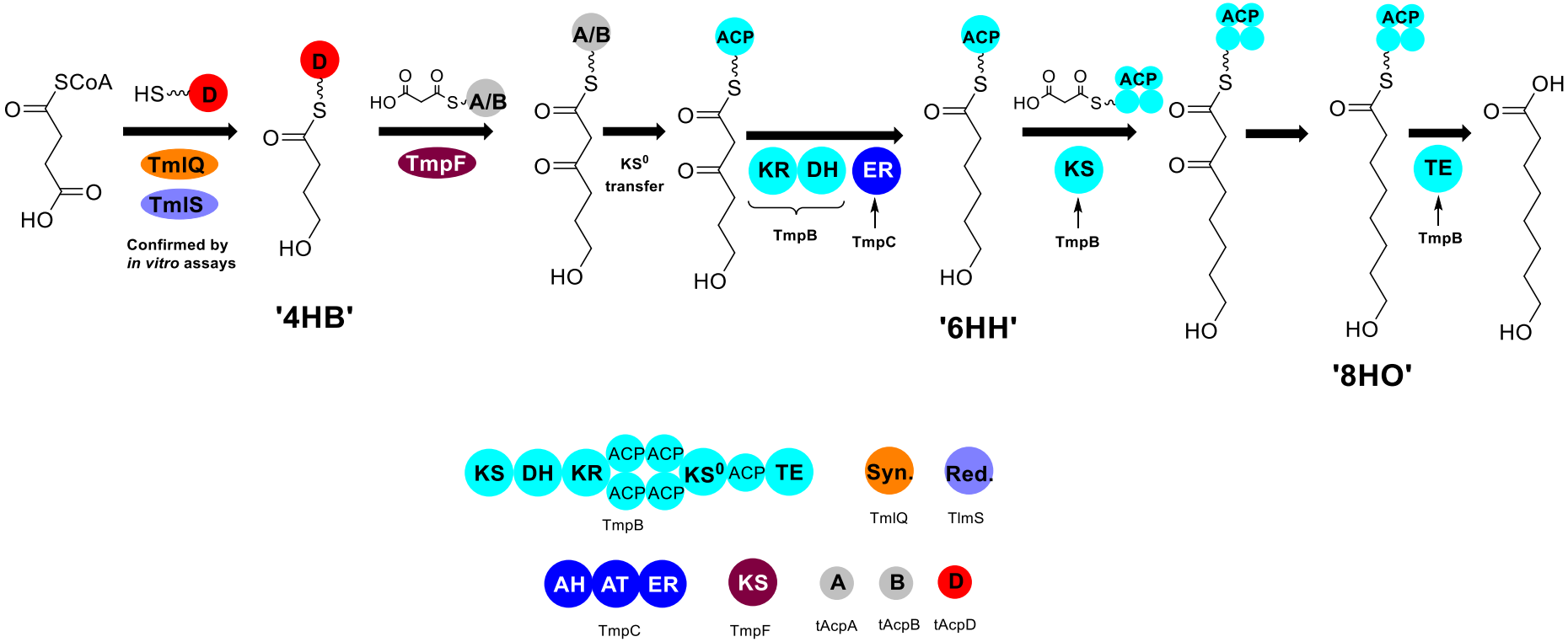


Figure 102: The domain architecture of MmpB (pink) and TmpB (light blue). The waved line on MmpB represents a flexible linker.

The early stages of the biosynthesis of thiomarinol have been elucidated. [86] In a manner analogous to the formation of 3HP *via* the action of mAcpD, MupQ and MupS, the equivalent 4-carbon moiety of 4-hydroxybutyrate (4HB) is produced *via* the action of tAcpD, TmlQ and TmlS. 4HB is expected to undergo two rounds of condensation and tailoring reactions to produce 8HO. Based on the work conducted in this thesis, a scheme for 8HO biosynthesis can be proposed. The first round of condensation is expected to be catalysed by TmpF, with 4HB delivered by tAcpD and malonyl delivered by both tAcpA and tAcpB in a process analogous to that observed in mupirocin. Future work should seek to confirm this by expression and purification of the proteins and subsequent *in vitro* assays monitored by ESI-MS. Due to the absence of TmlD and TmlE and presence of non-condensing KS within TmpB, it makes sense that transfer of the 6-carbon diketone to KS^0 takes place next. This is

consistent with previous studies of TmpB_KS^0 which has shown it is functional.[178] The acyl moiety will then be transferred to the single ACP located within TmpB. This will allow tailoring of the 6-carbon diketone moiety to occur with TmpB_KR, TmpB_DH and TmpC_ER. Following this, condensation is expected to occur catalysed by TmpB_KS with malonyl delivered by the quad domain ACP to produce the 8-carbon diketone. This will be followed by tailoring again catalysed by TmpB_KR, TmpB_DH and TmpC_ER to produce the 8HO moiety which will then be released by TmpB_TE. This is shown in Scheme 36.

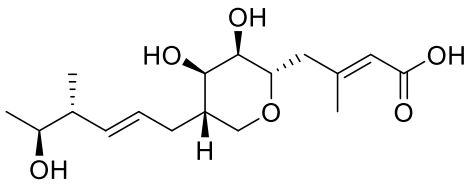


Scheme 36: Scheme showing proposed biosynthesis of fatty acid region of thiomarinol.

4.2 Further Work on the Biosynthetic Pathway of Mupirocin.

Future work should seek to understand more about ACP selectivity of the 3-carbon to 5-carbon (DCC #1) and 5-carbon to 7-carbon (DCC #2) condensation reactions. A mechanism is required to transfer the acyl chain from mAcpB to mAcpA after DCC #1 and subsequent tailoring. MupL is an enzyme which does not have a role and may be a candidate for this transfer reaction. Previous attempts to clone, express and purify MupL were unsuccessful due to the insoluble nature of this protein. Future work should seek increase the solubility of this protein. Growth of this plasmid using Arctic expression cells may help to this end. Following this, *in-vitro* assays should be conducted with 5HP-mAcpB and *holo*-mAcpA to see if this MupL can catalyse this transfer reaction.

This work has focussed on understanding the chemical steps responsible for 9HN formation during mupirocin biosynthesis. Work continues on understanding the remaining chemical steps that are still unknown within mupirocin biosynthesis. For example, MupB is expected to act as the enzyme which will join the 9HN moiety to the polyketide derived fragment in an esterification reaction. This reaction is yet be experimentally shown and is the basis of future work being conducted on this pathway. Work is underway to produce a complex synthetic mimic (Figure 103) of the polyketide moiety which can be used to test the function of MupB through *in vitro* assays monitored by ESI-MS.



44

Figure 103: The proposed natural substrate of MupB

The timing of esterification has been ascertained due to attachment of a simple mimic of the proposed substrate to different ACPs. Therefore, showing the esterification of 9HN and monic acid catalysed by MupB is an important role for the remainder of the work on this pathway.

The assays conducted in this work have made use of a chemoenzymatic method to transfer acyl moieties to the terminal thiol of ACPs through one-pot *in-vitro* assays. This technique to load ACPs works due to the capability of MupN-type PPT*ase* enzymes to transfer CoA moieties with a broad substrate specificity. [169,179] Within the *in vivo* biosynthetic pathway, MupN is responsible for converting ACPs from *apo*-form to active *holo*-form only. The loading of *holo*-ACPs with the required malonyl is expected to be catalysed by AT enzyme. Due to the *trans*-AT nature of the mupirocin pathway, this is expected to be provided by a single protein, MmpC which contains a di-domain AT. This has yet to be proven experimentally so future work should seek to show this conversion *in vitro*.

5 Methods

The protocols described in parts of this section are general molecular biology protocols and are not unique to this report (See M Rowe Masters Thesis and M Rowe 2nd Year Report) - as is standard practice for molecular biology theses. This should be considered when analysing the Turnitin report. Specific methodology for expression and purification protocols are given in later section.

5.1 General Protocols

5.1.1 Buffer Compositions

Column Buffer A: Trizma Base (50 mM), NaCl (0.5 M), glycerol (10%), pH=8.0

Column Buffer B: Trizma Base (50 mM), NaCl (0.5 M), imidazole (0.8 M), glycerol (10%), pH=8.0

Cathode Buffer: Trizma Base (200 mM) pH=8.9

Anode Buffer: Trizma Base (100 mM), tricine (100 mM), SDS (3.5 mM) pH=8.25

Gel Stain: Coomassie Brilliant Blue (3 mM), EtOH (7.7 M), AcOH (1.75 M)

Gel Loading Buffer: Trizma Base (780 mM), mercaptoethanol (142 mM), SDS (35 mM), glycerol (10%), bromophenol blue (140 mM), pH=6.8

Tris Buffer (for assays): Trizma Base (50 mM), NaCl (100 mM) pH=8.0

Tris Buffer + Glycerol: Trizma Base (50 mM), NaCl (100 mM), glycerol (10%), pH=8.0

NMR Buffer: 20 mM Sodium Phosphate, 100 mM NaCl, pH=8.0

LB Media: Per 1L deionised water was added tryptone (10g), yeast extract (5g) and NaCl (5g). pH modified to 7.4. Sterilised before use.

M9 Media: Per 1L deionised water was added Na_2HPO_4 (6g), KH_2PO_4 (3g), NaCl (0.5g). Buffered to pH=7.4. Sterilised before use.

Analytical SEC Buffer: Trizma Base (25 mM), NaCl (150 mM), pH=8.5.

5.1.2 General Points about Protein Expression

All media and glassware used were autoclaved (130˚C for 15 mins) prior to use to maintain sterility during growth and expression. The sterility required for protein expression and purification was achieved using a Medical Air Technology LTD BioMat Class II microbiological safety cabinet or a Bunsen flame. All surfaces were sterilised before use using 70% ethanol and Lipsol detergent.

For overexpression of proteins using *E. coli*, LB media was used unless otherwise specified. Specific protocol for expression of each protein is given in later sections. General protocol for purification of each protein is given in the results section of this thesis.

Centrifugation carried out in this project was carried out by a Sorvall RC 6 Plus Centrifuge at 4 ˚C unless otherwise specified.

5.1.3 Production of Agar Plates

LB Agar was produced by formation of 100 mL LB Media and addition of 1.5g of Agar. This was subsequently autoclaved. Following this, solid LB agar was melted by gentle heating and allowed to cool to 50 ˚C – 55 ˚C. Under sterile conditions, a 1:1000 dilution of antibiotic was added to agar and distributed evenly between 4-5 petri dishes which were left to cool (under sterile conditions) until the agar had solidified. Care was taken when adding the antibiotic to ensure agar was not too hot as to degrade the antibiotic.

5.1.4 Cell Transformation (Non Arctic)

To 20 μL of High-Efficiency *E. Coli* Cells was added 1 μL of plasmid DNA with gentle agitation. This was left at 4°C for 30 minutes. Following this, cells were heated to 42°C in a water bath for 30 seconds, followed by placing on ice (4°C) for 5 minutes. Under sterile conditions, 200 μL of preheated SOC Outgrowth Media was added. The resulting cell suspension was kept at 37°C with 180 RPM shaking for 1 hour. 200 μL of the resulting cell suspension was added to a freshly prepared 100 mm LB Agar petri dish containing appropriate antibiotic and spread evenly using a recently sterilised spreading rod. Agar plates containing cell suspension were placed in a 37°C incubator for 16 hours until colonies were observed.

5.1.5 Plasmid Isolation for Miniprep

Cell transformation and subsequent pre culture was performed as is described elsewhere. Following this, the cells were pelleted by centrifugation (6000 RPM, 10 minutes) and the plasmid DNA purified using a GenElute Spin Miniprep kit (Sigma) using Eppendorf MiniSpin Plus Microcentrifuge. The resulting plasmid DNA was checked by Sanger Sequencing (GeneWiz) to confirm successful Miniprep.

5.1.6 Plasmid Isolation for PCR Purification

Plasmid containing PCR fragment (see Plasmid Ligation) was Pre-Cultured overnight (stirring at 37 ˚C) with 10 mL of cells in a 50 mL sterile Falcon Tube. The cells were pelleted by centrifugation (6000 RPM, 5 minutes) and the plasmid DNA purified using a GenElute Spin Miniprep kit (Sigma) using Eppendorf MiniSpin Plus Microcentrifuge. The successful construction of the plasmids was confirmed by Sanger sequencing (GeneWiz).

5.1.7 Protein Expression (Non-Arctic)

Cell transformation was conducted as described above. An isolated colony was chosen and added to sterile LB media (50 mL) and antibiotic (1:1000 dilution of 1 mg per mL stock, 50 µL) was added. This was incubated overnight stirring (200 RPM) at 37 ˚C. LB media containing antibiotic (1:1000 dilution, 200 µL) was inoculated with pre-culture solution (2 mL) and incubated (37 ˚C, 200 RPM) until Optical Density at 600nm reached between 0.7-1.0. Once this was achieved, IPTG (50 mL, 250 µM) and antibiotic (1:1000 dilution, 50 µL) was added to each flask and incubated overnight (16 hours) at a reduced temperature (16 ˚C, 200 RPM). The cell pellet was harvested by centrifugation (7000 RPM, 10 minutes) and suspended in resuspension buffer (20 mL). The cell pellet was flash frozen in liquid N_2 and stored at -20 °C until required.

5.1.8 Arctic Cell Transformation and Expression

Arctic cell was chosen for proteins which had proven to be insoluble under standard protein expression conditions. Cell transformation and expression was conducted as follows:

ArcticExpress (DE3) Competent Cells were stored at -80 ˚C until required. 100 µL of cells were allowed to thaw on ice. 2 µL of diluted XL10-Gold β-mercaptoethanol was added to the cells with gentle agitation. Cells were left on ice for 10 minutes. 3 µL of plasmid DNA was added to cells and incubated on ice for 30 minutes. After this, cells were heat-pulsed by adding to 42 ˚C water bath for 20 seconds before placing at 4 ˚C for a further 5 minutes. To the cells was added 900 µL SOC media (pre-heated to 42 ˚C) under sterile conditions. Cells were then placed stirring at 37 ˚C for 1 hour (225 RPM). After this time, 200 µL of cells were spread onto LB agar plate containing the required antibiotic. Remaining 800 µL cells can were spun down, 600 µL SOC removed, vortexed and remaining 200 µL added to LB agar plates. Agar plates were left at 37 ˚C for 16 hours until colonies were observed.

Each isolated colony was added to 10 mL sterile LB containing 1:1000 appropriate antibiotic (for selection of the expression plasmid) and 1:1000 gentamicin. Pre-culture tubes are incubated at 37 ˚C, 200 RPM for 24 hours. After this time, successful pre-culture flasks (10 mL) are added to 100 mL sterile LB (non-baffled flasks – no antibiotic added at this stage). This is left at 30 ˚C for 3 hours. After this time, temperature is reduced to 11 ˚C and left for a further hour after which 0.5mM IPTG is added to each flask and left for a further 24 hours. After this time, cells are harvested, frozen in liquid N_2 and stored at -20 ˚C until required for purification.

5.2 Protein Purification Protocols

5.2.1 Protein Extraction

Cell Pellet was thawed on ice. *E. Coli* cells were disrupted by ultrasonic sound waves (VCX130 Vibra-Cell, Sonics) until a uniform, translucent liquid was observed (12-15 minutes). This suspension was

centrifuged (15000 RPM, 15 minutes) to isolate the soluble and insoluble components. The soluble component was collected for subsequent purification. Samples of both soluble and insoluble components were kept for subsequent SDS-PAGE Analysis (so-called 'Insol.' and 'Sol.' samples).

5.2.2 Immobilised Metal Ion Affinity Chromatography (IMAC) Purification Protocol

His-tagged protein was purified via Immobilized Metal Ion Affinity Chromatography (IMAC). A HisTrap HP Nickel Column (5 mL, GE Healthcare) was cleaned with 5 mL NaOH (0.5 M) and 15 mL HPLC grade water prior to use. Column was charged with 10 mL $NiSO_4$ (0.5 M) and equilibrated with running buffer (95% Column Buffer A: 5% Column Buffer B). The injection loop of the FPLC was cleaned with NaOH (0.5 M) and water. After soluble protein had been extracted from insoluble components (see above), soluble supernatant was added to FPLC through the injection loop followed by injection onto the Nickel Column. Once injected, non-binding proteins were washed away with 5 column volumes of running buffer until UV absorbance reached baseline levels. Non-binding proteins were collected, and sample taken for SDS-PAGE analysis (so-called 'Flow' sample). His-tagged protein was then selectively eluted using an increasing imidazole gradient to a maximum of 0.8M and collected in 1 mL fractions. These fractions were placed immediately on ice to hinder precipitation of the protein. Fractions were analysed by SDS-PAGE to confirm the fractions which contained protein of interest. Fractions containing protein were pooled together for downstream SEC purification and characterisation.

5.2.3 Protein Desalt

For pure proteins eluting from IMAC (confirmed by SDS-PAGE) subsequent purification was curtailed and removal of imidazole (desalting procedure) was undertaken using a desalting column. An injection loop was cleaned with 5 mL NaOH (0.5 M) and 15 mL HPLC grade water. A HiPrep 26/10 desalting column (50 mL, GE Healthcare) was added to the FPLC and equilibrated with the required desalt buffer (100 mL). Required IMAC fractions were combined and injected into the column. Protein elution was confirmed by a rise in 280nm absorbance and protein was collected in a sterile 50 mL falcon tube. The collection was stopped prior to the elution of small molecules and salt, indicated by a peak in conductivity. A sample of protein was taken and characterised by SDS-PAGE and ESI-MS.

5.2.4 Size Exclusion Chromatography

For larger proteins and proteins not found to be pure after IMAC, size-exclusion chromatography (SEC) was conducted as a further purification step. The injection loop was washed with the required desalt buffer (10 mL) to remove remnant supernatant**.** One of a series of size exclusion columns were used to purify the protein of interest depending on its molecular weight and oligomeric state in solution. The columns available for use were S75, S100 and S200 (GE Healthcare) and the chosen column was

equilibrated with required running buffer (400 mL). Protein sample was injected onto the column. Protein elution was signified by an increase in 280nm absorbance at the predetermined elution volume based on size and oligomeric state of protein. The protein was collected in 7 mL fractions prior to downstream concentration and further use.

5.2.5 Protein Concentration

After protein purification, fractions containing the required protein were determined by SDS-PAGE and ESI-MS. These fractions were combined and concentrated using spin concentrators (Pierce Protein Concentrator Polyethersulfone, 20 mL) with the required molecular weight cut off (MWCO) depending on size of protein. Concentrators with 3k, 10k, 30k and 100k MWCO were available for use. During concentration, the mg/mL of the protein was periodically checked using UV spectrophotometer until the required concentration was achieved. At this point the protein was aliquoted into the required volume, flash frozen in liquid N_2 and stored at -80°C until required.

5.3 Protein Characterisation

5.3.1 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Samples for analysis were collected at various points during the protein expression and purification protocol (see above). The samples were prepared by addition of equal volume Gel Loading Buffer (10 μL) and heating to 95 °C for 10 minutes. The electrophoresis gel was constructed in house and consisted of Separating (5 mL) and Loading (2.5 mL) regions. SDS-Marker (10 μL) was added to the first well and a small amount of each prepared gel sample (7 μL) was added to each well. The gel was operated between 110-150V for 50-70 minutes until the stain front had reached the bottom of the gel. Cathode and Anode buffer (see Buffers) was used to keep the pH constant and provide the mobile ions required to carry the current. Once complete, the PAGE gel was submerged in Coomassie Blue stain for at least one hour to ensure adequate staining. Following this, excess stain was removed, and de-staining was achieved by heating while submerged in water for approximately 15 minutes. Protein bands were detected using an E-gel imager (Life Technologies).

5.3.2 Electrospray Ionisation Protein Mass Spectrometry

5.3.2.1 Zip-Tip Sample Preparation

Protein samples were prepared using C4 ZipTip Pipette Tips (Merck Millipore). The final buffer consisted of 0.2% Formic acid within a 7:3 ratio of acetonitrile to water. 10 μL of protein was loaded into the Synapt G2-Si (Waters). Formic acid (10 μL) and Acetonitrile (10 μL) was often added to the sample before analysis to denature the protein and ensure better spray of the protein species.

5.3.2.2 Methanol-Chloroform Protein Preparation

For larger molecular weight proteins (> 30 kDa) it was more effective to prepare the mass spectrometry sample using methanol-chloroform extraction. The protein (20 µL) was added methanol (60 µL) and chloroform (15 µL) and vortexed well. Deionised Water (45 µL) was added, and precipitation observed. Following this, the sample was vortexed well and centrifuged at full speed (15k RPM) for 1 minute. The upper aqueous layer was removed, and remaining pellet was resuspended in methanol (45 µL). Prior to further vortexing and centrifugation for 2 minutes. The resulting supernatant was discarded and a 1:1 ratio of acetonitrile to water in 0.2% Formic acid (20 – 50 µL, depending on concentration of sample) was added to resuspend the precipitated protein. After resuspension, sample was centrifuged for 5 mins at full speed to remove remaining precipitate. 10 µL of this protein was loaded into the Synapt G2-Si (Waters).

5.3.3 Analytical Size Exclusion Chromatography

The monomeric state of the protein in solution was confirmed by analytical size exclusion chromatography. Two columns were available for use, Analytical S75 or Analytical S200. The required column was selected based on the size and predicted monomeric state of the protein. Prior to use, the required column was calibrated periodically with several known protein standards. The chosen column was equilibrated with Analytical SEC Buffer (25 mM Tris, 150 mM NaCl, pH=8.5). The protein sample was made up to 1 mg/mL using Tris Buffer and injected into the column. Special care was taken to ensure no air entered the column or FPLC system during injection. The elution of the protein was signalled by an increase in 280nm absorbance at a specific elution volume.

5.3.4 DNA Gel Electrophoresis

DNA Gel electrophoresis was conducted to confirm the successful PCR had taken place. Firstly, 100 mL of 1% agarose gel solution was liquified by gentle heating, and Midori Green DNA Stain (2-4 µL) was added. The resulting mixture was cast in a gel tray with the addition of an appropriate comb. The gel was placed in an electrophoresis tank and immersed in TAE buffer (containing Tris base, Acetic acid and EDTA). A mixture of PCR product sample and 5x DNA Loading Buffer Blue (1:1, 4 µL, Bioline) was added to each well, and HyperLadder 1 KB (4 µL, 43 Bioline) was added to each analysis row. Electrophoresis was performed at 200 V for 15-25 minutes and DNA bands were detected on the resulting gel via a UV-transilluminator (Essential, UVITEQ Cambridge).

5.3.5 Plasmid Ligation

Once PCR products had been confirmed to contain DNA region of interest and subsequently purified, they were ready to ligate into appropriate vector.

For ligation into pOPINF plasmid, DNA fragments were added to pre-linearised pOPINF vector according to the following reaction conditions; 3 µL of DNA, 4 µL dH_20, 1 µL linearised pOPINF vector, 1 µL Takura 5x Infusion HD Enzyme Premix. The reaction was left until complete (20 minutes, 50 ˚C).

Ligation into pre-linearised pET28a vector was conducted using the following reaction conditions; 2 µL of DNA, 2 µL of cut pET28a, 1 µL of Takura 5x Infusion HD Enzyme Premix. This was incubated at 50 ˚C for 20 minutes prior to transformation into NEB 5α cells, pre-culture and mini-prep protocol followed by Gene Sequencing (GeneWiz).

5.4 Protein Expression and Purification Details

Plasmids for the growth of MupD, MupE, MmpD_DH1 were ligated into a pOPINF vector (carbenicillin resistance). MmpF was ligated into pET28a vector (kanamycin resistance).

Cells were induced with 250 µM IPTG once OD_{600} reached 0.7-1 to ensure sufficient expression of protein. Cell growth was conducted at 37°C followed by induction at 16°C for 16 hours.

5.5 Polymerase Chain Reaction

The primers used in this project were designed in house and synthesised by Sigma Aldrich (Supplementary Information). Primer stock solutions (100 µM) were prepared using Tris Buffer and working solutions (25 µM) were prepared by further dilution in distilled water. PCR reactions were conducted using either Q5 polymerase or KAPA polymerase.

All PCR reactions were performed using a peqSTAR 96x Universal Gradient and 200 µL PCR tubes. To confirm PCR had been successful, PCR product was analysed by 1% agarose gel electrophoresis. PCR reactions were initially conducted on a 50 µL scale over a range of temperatures (48-66 °C) and successful temperature recorded. PCR was then repeated at a single successful temperature. For each specific gene fragment, PCR was conducted under the following conditions (with the polymerase shown in brackets).

5.5.1 MmpF (KAPA):

1) Initial denaturation at 95 °C for 3 mins; 2) Denaturation at 98 °C for 20 s; 3) 60 °C for 15 seconds 4) Extension at 72 °C for 2 mins 30 s. Step 2-4 were repeated 25 times prior to a final extension at 72 °C for 10 mins.

5.5.2 MupD and MupE (KAPA):

1) Initial denaturation at 95 °C for 3 mins; 2) Denaturation at 98 °C for 20 s; 3) 60 °C for 15 seconds 4) Extension at 72 °C for 1 minute. Step 2-4 were repeated 35 times prior to a final extension at 72 °C for 10 mins.

The successful PCR DNA fragments were ligated into pre-linearised plasmid as per the conditions listed; Cut plasmid 5 µL, 5x T7 infusion enzyme 1 µL, PCR Product 3 µL, 10x buffer 1 µL.

5.5.3 MmpD_DH1 (Q5)

1) Initial denaturation at 98 °C for 30 s; 2) Denaturation at 98 °C for 5-10 s; 3) Temp range (55 – 72 °C) for 15 seconds 4) Extension at 72 °C for 45 s. Step 2-4 were repeated 35 times prior to a final extension at 72 °C for 2 mins. All temperatures were found to produce a band on the PCR gel.

5.5.4 PCR Mutagenesis to produce MmpFC183A (KAPA)

5% DMSO was added to PCR mixture. Template DNA used was MmpF_pET28a.

1) Initial denaturation at 95 °C for 3 mins; 2) Denaturation at 98 °C for 20 s; 3) 60 °C for 15 seconds 4) Extension at 72 °C for 6 mins 30 s. Step 2-4 were repeated 25 times prior to a final extension at 72 °C for 10 mins.

5.6 Protein Expression and Purification Details

In this report, all proteins discussed herein contain a polyhistidine tag unless removal was specified. All the protocols described were conducted in a sterile environment to minimise cross-contamination. Standard protein expression protocol has been outlined previously. Specifics of protein expression and purification for each protein are described here:

5.6.1 Expression of mAcpD

Transformation was conducted with 25 µL BL21 (NEB) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.6L scale using LB media within non-baffled flasks; cells were grown at 37°C 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 µM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.2 Expression of MupN

Transformation was conducted with 20 µL BL21 (Novagen) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 2L scale using LB media within non-baffled

flasks; cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 μM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.3 Expression of mAcpD_T

Transformation was conducted with 20 μL BL21 (Novagen) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 2L scale using LB media within non-baffled flasks; cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 μM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.4 Expression of mAcpB

Transformation conducted with 100 μL Agilent ArcticExpress (DE3) Competent Cells. Protocol as described in 'Arctic growth protocol' section.

5.6.5 Expression of mAcpA

Transformation conducted with 100 μL Agilent ArcticExpress (DE3) Competent Cells.

5.6.6 Expression of MupD

Transformation conducted with 100 μL Agilent ArcticExpress (DE3) Competent Cells.

5.6.7 Expression of MupE

Transformation conducted with 100 μL Agilent ArcticExpress (DE3) Competent Cells.

5.6.8 Expression of MmpF

Transformation was conducted with 20 μL BL21 (Novagen) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.4L scale using LB media with 100mL LB per non-baffled flask. The cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 μM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.9 Expression of MmpF_C183A

Transformation was conducted with 20 μL BL21 (Novagen) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.4L scale using LB media with 100mL LB per non-baffled flask. The cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 μM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.10 Expression of MmpB_KS

Transformation was conducted with 20 µL T7 high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.6L scale using LB media with 200mL LB per non-baffled flask. The cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 µM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.11 Expression of MmpB_DH

Transformation was conducted with 25 µL BL21 (NEB) high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.6L scale using LB media with 200mL LB per non-baffled flask. The cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 µM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested.

5.6.12 Expression of MmpD_DH1

Transformation was conducted with 25 µL T7 high-efficiency cells. Pre-culture was left overnight with stirring at 37°C. Growth was conducted on 1.6L scale using LB media with 200mL LB per non-baffled flask. The cells were grown at 37°C, 200 RPM until required OD_{600} reached. At this point, cells were induced with 250 µM IPTG and temperature dropped to 16°C. Cells were left shaking for 16 hours after which time the cells were harvested

5.7 Expression of Isotopically Enriched Protein:

5.7.1 Expression of ^{15}N-^{13}C-mAcpD

Transformation was conducted with 25 µL BL21 (Novagen) high-efficiency cells. Pre-culture using 100mL LB media was left overnight at 37°C (200 RPM). The cells were pelleted (spun at 6000 RPM for 10 minutes), resuspended in M9 media (20 mL), pelleted again and resuspended in M9 (20 mL). 1 mL of this preculture was added to 100mL of M9 Buffer in non-baffled flask (total volume of growth 2L). 2 grams ^{13}C-D-Glucose, 1 gram 15N-ammonium chloride, 1 ml MgSO4 (1M), 1 ml trace metal solution and 100 mg carbenicillin was added per 1L of M9 solution (solutions were sterile filtered prior to use). Flasks were incubated at 37°C with shaking (220 RPM) until OD_{600} of 0.7 was reached. At this point cells were induced with 1 mM IPTG per 100 ml flask. Cells where harvested after 16 hours.

5.7.2 Expression of ^{15}N-mAcpD

Transformation was conducted with 20 µL BL21 (Novagen) high-efficiency cells. Pre-culture using 100mL LB media was left overnight with shaking at 37°C. 1 mL of preculture was added per 100 mL LB. Cells were grown at 37°C to an OD of 2 with a total volume of 2L. Once OD was reached, cells were pelleted (6000 RPM for 10 minutes), resuspended in M9 (400 ml), pelleted again and resuspended in M9 (50 mL). 10mL of the cell suspension was added to 100 mL M9 (total 5 flasks, pre-warmed to 37°C). Per 100 mL M9 was added carbenicillin (100 mg/L), 100 µL $MgSO_4$ (1M), 2.5 mL 50/52, 100 µL trace metal solution and ^{15}N-ammonium chloride (0.4 g). Flasks incubated at 37°C for 1 hour until temperature was dropped to 16°C and induced with 500 µM IPTG. Cells were left for 16 hours and harvested.

6 References

1 H. F. Ji, X. J. Li and H. Y. Zhang, *EMBO Rep.*, 2009, **10**, 194–200.

2 R. S. Solecki, *Science*, 1975, **190**, 880–881.

3 D. J. Newman, G. M. Cragg and K. M. Snader, *Nat. Prod. Rep.*, 2000, **17**, 215–234.

4 S. A. Selim, M. E. Adam, S. M. Hassan and A. R. Albalawi, *BMC Complem. and Altern. M.*, 2014, **14**, 179–186.

5 R. J. Huxtable and S. K. W. Schwarz, *Mol. Interv.*, 2001, **1**, 189–191.

6 J. v. Pham, M. A. Yilma, A. Feliz, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park and Y. J. Yoon, *Frontiers in Microbiology*, 2019, **10**, 1404.

7 E. Patridge, P. Gareiss, M. S. Kinch and D. Hoyer, *Drug Discov. Today*, 2016, **21**, 204–207.

8 A. G. Atanasov, S. B. Zotchev, V. M. Dirsch and C. T. Supuran, *Nat. Rev. Drug Discov.*, 2021, **20**, 200–216.

9 Poppy Capsule Drop Of Water Wet , https://pixabay.com/photos/poppy-capsule-drop-of-water-wet-4684046/, (accessed June 30, 2021).

10 F. Sato and K. Matsui, *Engineering the biosynthesis of low molecular weight metabolites for quality traits (essential nutrients, health-promoting phytochemicals, volatiles, and aroma compounds)*, Academic Press, 2012.

11 D. Voet and J. G. Voet, *Biochemistry*, Wiley, 4th edn., 2010.

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

13 Y.-W. Chin, M. J. Balunas, H. B. Chai and A. D. Kinghorn, *AAPS J.*, 2006, **8**, E239–E253.

14 J. P. Anderson, C. A. Gleason, R. C. Foley, P. H. Thrall, J. B. Burdon and K. B. Singh, *Funct. Plant Biol.*, 2010, **37**, 499–512.

15 J. Barnes, *Medicinal Natural Products*, Wiley, Chichester, 3rd edn., 2010, vol. 3.

16 B. S. Moore and C. Hertweck, *Nat. Prod. Rep.*, 2002, **19**, 70–99.

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

18 K. Finzel, D. J. Lee and M. D. Burkart, *ChemBioChem*, 2015, **16**, 528–547.

19 D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 14234–14243.

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

21 R. Al-Dhelaan, P. S. Russo, S. E. Padden, A. Amaya, D. W. Dong and Y. O. You, *ACS Chem. Biol.*, 2019, **14**, 304-312.

22 S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science*, 1991, **252**, 675–679.

23 F. Schluenzen, R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath and F. Franceschi, *Nature*, 2001, **413**, 814–821.

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

25 S. Smith and S. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.

26 C. Hertweck, *Angew. Chem. Int. Ed.*, 2009, **48**, 4688–4716.

27 G. A. Zornetzer, B. G. Fox and J. L. Markley, *Biochemistry - US*, 2006, **45**, 5217–5227.

28 A. Chen, R. N. Re and M. D. Burkart, *Nat. Prod. Rep.*, 2018, **35**, 1029–1045.

29 N. Singh, S. Wakil and J. Stoops, *J. Biol. Chem.*, 1984, **259**, 3605–3611.

30 A. Witkowski, A. Joshi and S. Smith, *Biochemistry - US*, 1996, **35**, 10569–10575.

31 A. K. Joshi, A. Witkowski and S. Smith, *Biochemistry - US*, 1997, **36**, 2316–2322.

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 Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *P. Natl. Acad. Sci. USA*, 2006, **103**, 11124–11129.

35 C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Nat. Prod. Rep.*, 2007, **24**, 162–190.

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

37 G. J. Williams, *Curr. Opin. Struc. Biol.*, 2013, **23**, 603–612.

38 A. Gallo, M. Ferrara and G. Perrone, *Toxins*, 2013, **5**, 717–742.

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

40 R. v O'Brien, R. W. Davis, C. Khosla and M. E. Hillenmeyer, *J. Antibiot.*, 2014, **67**, 89–97.

41 H. Jenke-Kodama, T. Börner and E. Dittmann, *PLOS Comp. Biol.*, 2006, **2**, 1210–1218.

42 T. A. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.

43 A. Baader and C. Garre, *Corresp. Bi. Schweiz Aerzte*, 1887, **17**, 597–597.

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

45 B. Sir, E. B. Chain and G. Mellows, *J. Chem. Soc., Perk. Trans. 1*, 1977, **3**, 294–309.

46 R. G. Alexander, J. P. Clayton, K. Luk, N. H. Rogers and T. J. King, *J. Chem. Soc.* , 1978, **1**, 561–565.

47 R. Sutherland, R. J. Boon, K. E. Griffin, P. J. Masters, B. Slocombe and A. R. White, *Antimicrob. Agents Ch.* , 1985, **27**, 495–498.

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

49 T. Coates, R. Bax and A. Coates, *J. Antimicrob. Chemoth.*, 2009, 64, 9–15.

50 M. J. Sugden, PhD Thesis, University of Bristol, 1992.

51 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.

52 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.

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

54 M. J. B. Brown, L. M. Mensah, M. L. Doyle, N. J. P. Broom, N. Osbourne, A. K. Forrest, C. M. Richardson, P. J. O'hanlon and A. J. Pope, *Biochemistry - US*, 2000, **39**, 6003–6011.

55 G. Eriani, M. Delarue, O. Poch, J. Gangloff and D. Moras, *Nature*, 1990, **347**, 203–206.

56 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.

57 P. Liras, *Appl. Microbiol. Biot.*, 2014, **98**, 1023–1030.

58 B. Li, W. J. Wever, C. T. Walsh and A. A. Bowers, *Nat. Prod. Rep.*, 2014, **31**, 905–923.

59 Z. D. Dunn, W. J. Wever, N. J. Economou, A. A. Bowers and B. Li, *Angew. Chem. Int. Edit.*, 2015, **54**, 5137–5141.

60 M. H. Medema and A. Osbourn, *Nat. Prod. Rep.*, 2016, **33**, 951–962.

61 M. J. Byrne, N. R. Lees, L. C. Han, M. W. van der Kamp, A. J. Mulholland, J. E. M. Stach, C. L. Willis and P. R. Race, *J. Am. Chem. Soc.*, 2016, **138**, 6095–6098.

62 S. Basler, S. Studer, Y. Zou, T. Mori, Y. Ota, A. Camus, H. A. Bunzel, R. C. Helgeson, K. N. Houk, G. Jiménez-Osés and D. Hilvert, *Nat. Chem.*, 2021, **13**, 231–235.

63 D. A. Hopwood, *PLoS Biol.*, 2004, **2**, 166–169.

64 C. Llor and L. Bjerrum, *Ther. Adv. Drug Saf.*, 2014, **5**, 229–41.

65 J. O'Neill, *Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations.*, 2014, https://amr-review.org/.

66 U. Hofer, *Nat. Rev. Microbiol.*, 2019, **17**, 3.

67 J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Nat Prod Rep.*, 2014, **31**, 61–108.

68 J. Beld, K. Finzel and M. D. Burkart, *Chem. Biol.*, 2014, **21**, 1293–1299.

69 R. J. Cox, J. Crosby, O. Daltrop, F. Glod, M. E. Jarzabek, T. P. Nicholson, M. Reed, T. J. Simpson, L. H. Smith, F. Soulas, A. E. Szafranska and J. Westcott, *J. Chem. Soc. Perk. Trans. 1*, 2002, **14**, 1644–1649.

70 P. Dall Aglio, C. J. Arthur, C. Williams, K. Vasilakis, H. J. Maple, J. Crosby, M. P. Crump and A. T. Hadfield, *Biochemistry - US*, 2011, **50**, 5704–5717.

71 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.

72 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. Chem.-Ger. Ed.*, 2019, **131**, 12576–12580.

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

74 K. M. Clarke, A. C. Mercer, J. J. la Clair and M. D. Burkart, *J. Am. Chem. Soc.*, 2005, **127**, 11234–11235.

75 D. T. Wagner, J. Zeng, C. B. Bailey, D. C. Gay, F. Yuan, H. R. Manion and A. T. Keatinge-Clay, *Structure*, 2017, **25**, 1045–1055.

76 S. K. Piasecki, C. A. Taylor, J. F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, D. R. Siegel and A. T. Keatinge-Clay, *Chem. Biol.*, 2011, **18**, 1331–1340.

77 M. Jenner, J. P. Afonso, C. Kohlhaas, P. Karbaum, S. Frank, J. Piel and N. J. Oldham, *Chem. Commun.*, 2016, **52**, 5262–5265.

78 A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–430.

79 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, **8**, 2040–2042.

80 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, **9**, 1179–81.

81 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. Chem. Int. Edit.*, 2017, **56**, 3930–3934.

82 J. Wu, J. Hothersall, C. Mazzetti, Y. O'Connell, J. A. Shields, A. S. Rahman, R. J. Cox, J. Crosby, T. J. Simpson, C. M. Thomas and C. L. Willis, *ChemBioChem*, 2008, **9**, 1500–1508.

83 F. M. Martin and T. J. Simpson, *J. Chem. Soc. Perk. T. 1*, 1989, **1**, 207–209.

84 X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Sussmuth and R. Borriss, *J. Bacteriol.*, 2006, **188**, 4024–4036.

85 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.

86 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.

87 M. Rowe, Masters Thesis, University of Bristol, 2017.

88 F. Hemmerling, K. E. Lebe, J. Wunderlich and F. Hahn, *ChemBioChem*, 2018, **19**, 1006–1011.

89 A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, *Nucleic Acids Res.*, 2015, **43**, 389–394.

90 K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. Lemotte, A. S. McColl, A. v. Kamath and J. G. Stroh, *Anal. Biochem.*, 1999, **267**, 169–184.

91 J. Crosby, D. H. Sherman, M. J. Bibb, W. Peter Revill, D. A. Hopwood and T. J. Simpson, *BBA-Prot. Struct. M.*, 1995, **1251**, 32–42.

92 R. Wu, A. S. Reger, X. Lu, A. M. Gulick and D. Dunaway-Mariano, *Biochemistry-US*, 2009, **48**, 4115–4125.

93 H. K. D'Ambrosio and E. R. Derbyshire, *ACS Chem. Biol.*, 2020, **15**, 17–27.

94 G. Jogl and L. Tong, *Biochemistry*, 2004, **43**, 1425–1431.

95 K. Priyadarshan and R. Sankaranarayanan, *J. Indian I. Sci.*, 2018, **98**, 261–272.

96 S. E. Evans, C. Williams, C. J. Arthur, E. Płoskoń, P. Wattana-amorn, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and M. P. Crump, *J. Mol. Biol.*, 2009, **389**, 511–528.

97 P. M. Marchetti, V. Kelly, J. P. Simpson, M. Ward and D. J. Campopiano, *Org. Biomol. Chem.*, 2018, **16**, 2735–2740.

98 H. Kage, M. F. Kreutzer, B. Wackler, D. Hoffmeister and M. Nett, *Chem. Biol.*, 2013, **20**, 764–771.

99 M. P. Schmitt and S. M. Payne, *J. Bacteriol.*, 1991, **173**, 816–825.

100 C. A. Mitchell, C. Shi, C. C. Aldrich and A. M. Gulick, *Biochemistry-US*, 2012, **51**, 3252–3263.

101 A. M. Gulick, *ACS Chem. Biol.*, 2009, **4**, 811–827.

102 F. J. Asturias, J. Z. Chadick, I. K. Cheung, H. Stark, A. Witkowski, A. K. Joshi and S. Smith, *Nat. Struct. Mol. Biol.*, 2005, **12**, 225–232.

103 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.

104 T. Robbins, Y.-C. C. Liu, D. E. Cane and C. Khosla, *Curr. Opin. Struc. Biol.*, 2016, **41**, 10–18.

105 J. G. Olsen, A. Kadziola, P. Von Wettstein-Knowles, M. Siggaard-Andersen and S. Larsen, *Structure*, 2001, **9**, 233–243.

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

107 C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen and Z. Deng, *J. Biol. Chem.*, 2010, **285**, 20097–20108.

108 B. CL, M. MG and T. AL, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 4081–4086.

109 J. R. Lohman, M. Ma, J. Osipiuk, B. Nocek, Y. Kim, C. Chang, M. Cuff, J. Mack, L. Bigelow, H. Li, M. Endres, G. Babnigg, A. Joachimiak, G. N. Phillips and B. Shen, *P. Natl. Acad. Sci. USA*, 2015, **112**, 12693–12698.

110 Z. Chang, N. Sitachitta, J. v. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *J. Nat. Prod.*, 2004, **67**, 1356–1367.

111 B. Zhang, D. Yang, Y. Yan, G. Pan, W. Xiang and B. Shen, *Appl. Microbiol. Biot.*, 2016, **100**, 2267–2277.

112 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.

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

114 J. Piel, D. Hui, N. Fusetani and S. Matsunaga, *Environ. Microbiol.*, 2004, **6**, 921–927.

115 D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure*, 2014, **22**, 444–451.

116 A. Ginolhac, C. Jarrin, P. Robe, G. Perrière, T. M. Vogel, P. Simonet and R. Nalin, *J. Mol. Evol.*, 2005, **60**, 716–725.

117 D. Hoffmann, J. M. Hevel, R. E. Moore and B. S. Moore, *Gene*, 2003, **311**, 171–180.

118 T. K. Kumar S., Stecher G., Li M., Knyaz C., *Mol. Biol. Evo.*, 2018, **35**, 1547–1549.

119 J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res.*, 1994, **22**, 4673–4680.

120 J. P. Huelsenbeck and F. Ronquist, *Bioinformatics*, 2001, **17**, 754–755.

121 Y.-Q. Cheng, G.-L. Tang and B. Shen, *J. Bacteriol.*, 2002, **184**, 7013–7024.

122 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chem.-Ger. Edit.*, 2013, **125**, 1181–1185.

123 A. H. Rosenberg, B. N. Lade, C. Dao-shan, S. W. Lin, J. J. Dunn and F. W. Studier, *Gene*, 1987, **56**, 125–135.

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

125 T. Robbins, J. Kapilivsky, D. E. Cane and C. Khosla, *Biochemistry-US*, 2016, **55**, 4476–4484.

126 M. Jerpseth, B. Jerpseth, L. Briester and A. Greener, *Strategies*, 1998, **11**, 3–4.

127 M. Ferrer, T. N. Chernikova, M. M. Yakimov, P. N. Golyshin and K. N. Timmis, *Nat. Biotechnol.*, 2003, **21**, 1266–1267.

128 C. H. Schein, *Nat. Biotechnol.*, 1989, **7**, 1141–1149.

129 M. J. Meehan, X. Xie, X. Zhao, W. Xu, Y. Tang and P. C. Dorrestein, *Biochemistry-US*, 2011, **50**, 287.

130 J. T. Tsay, W. Oh, T. J. Larson, S. Jackowski and C. O. Rock, *J. Biol. Chem.*, 1992, **267**, 6807–6814.

131 A. Ruppe and J. M. Fox, *ACS Catal.*, 2018, **8**, 11722–11734.

132 K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, J. Hedlund and H. Jörnvall, *Cell. Mol. Life Sci.*, 2008, **65**, 3895–3906.

133 J. W. Campbell and J. E. Cronan, *Annu. Rev. Microbiol.*, 2001, **55**, 305–332.

134 C. Filling, K. D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jörnvall and U. Oppermann, *J. Biol. Chem.*, 2002, **277**, 25677–25684.

135 A. Koumanov, J. Benach, S. Atrian, R. Gonzàlez-Duarte, A. Karshikoff and R. Ladenstein, *Proteins*, 2003, **51**, 289–298.

136 A. M. Lesk, *Curr. Opin. Struc. Biol.*, 1995, **5**, 775–783.

137 M. Fisher, J. T. M. Kroon, W. Martindale, A. R. Stuitje, A. R. Slabas and J. B. Rafferty, *Structure*, 2000, **8**, 339–347.

138 S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E. Detelich and A. T. Keatinge-Clay, *Proteins*, 2014, **82**, 2067–2077.

139 A. M. Krishnakumar, B. P. Nocek, D. D. Clark, S. A. Ensign and J. W. Peters, *Biochemistry-US*, 2006, **45**, 8831–8840.

140 Allen C. Price, Yong-Mei Zhang, Charles O. Rock and Stephen W. White, *Biochemistry-US*, 2001, **40**, 12772–12781.

141 T. Sztain, T. G. Bartholow, J. A. Mccammon and M. D. Burkart, *Biochemistry*, 2019, **58**, 3557–3560.

142 A. Brown, V. Affleck, J. Kroon and A. Slabas, *FEBS Lett.*, 2009, **583**, 363–368.

143 S. Mochizuki, K. Hiratsu, M. Suwa, T. Ishii, F. Sugino, K. Yamada and H. Kinashi, *Mol. Microbiol.*, 2003, **48**, 1501–1510.

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

145 K. H. Choi, R. J. Heath and C. O. Rock, *J. Bacteriol.*, 2000, **182**, 365–370.

146 Paul Walker, PhD Thesis, University of Bristol, 2019.

147 L. Moretto, S. Vance, B. Heames and R. W. Broadhurst, *Chem. Commun*, 2017, **53**, 11457–11460.

148 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.

149 B. Persson, J. Hedlund and H. Jörnvall, *Cell. Mol. Life Sci.*, 2008, **65**, 3879–3894.

150 H. Riveros-Rosas, A. Julian-Sanchez, R. Villalobos-Molina, J. P. Pardo and E. Pina, *Eur. J. Biochem.*, 2003, **270**, 3309–3334.

151 C. Ceccarelli, Z. X. Liang, M. Strickler, G. Prehna, B. M. Goldstein, J. P. Klinman and B. J. Bahnson, *Biochemistry-EU*, 2004, **43**, 5266–5277.

152 A. Henne, H. Brüggemann, C. Raasch, A. Wiezer, T. Hartsch, H. Liesegang, A. Johann, T. Lienard, O. Gohl, R. Martinez-Arias, C. Jacobi, V. Starkuviene, S. Schlenczeck, S. Dencker, R. Huber, H. P. Klenk, W. Kramer, R. Merkl, G. Gottschalk and H. J. Fritz, *Nat. Biotechnol.*, 2004, **22**, 547–553.

153 M. Knoll and J. Pleiss, *Prot. Sci.*, 2008, **17**, 1689–1697.

154 H. Bergler, S. Fuchsbichler, G. Högenauer and F. Turnowsky, *Eur. J. Biochem.*, 1996, **242**, 689–694.

155 L. Hendrickson, C. R. Davis, C. Roach, D. K. Nguyen, T. Aldrich, P. C. McAda and C. D. Reeves, *Chem. Biol.*, 1999, **6**, 429–439.

156 B. D. Ames, C. Nguyen, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W-H Li, J. C. Vederas, Y. Tang and S.-C. Tsai, *PNAS*, 2012, **109**, 11144–11149.

157 J. W. Seo, M. Ma, T. Kwong, J. Ju, S. K. Lim, H. Jiang, J. R. Lohman, C. Yang, J. Cleveland, E. Zazopoulos, C. M. Farnet and B. Shen, *Biochemistry-US*, 2014, **53**, 7854–7865.

158 B. Zhang, Z. Xu, Q. Teng, G. Pan, M. Ma and B. Shen, *Angew. Chem. Int. Ed.*, 2017, **56**, 7247–7251.

159 T. Moriguchi, Y. Ebizuka and I. Fujii, *ChemBioChem*, 2008, **9**, 1207–1212.

160 A. K. Joshi and S. Smith, *J. Biol. Chem.*, 1993, **268**, 22508–22513.

161 J. D. Kittendorf and D. H. Sherman, *Bioorg. Med. Chem.*, 2009, **17**, 2146.

162 Y. Li, G. J. Dodge, W. D. Fiers, R. A. Fecik, J. L. Smith and C. C. Aldrich, *J. Am. Chem. Soc.*, 2015, **137**, 7003–7006.

163 D. Gay, Y.-O. You, A. Keatinge-Clay and D. E. Cane, *Biochemistry*, 2013, **52**, 8916–8928.

164 O. Vergnolle, F. Hahn, A. Baerga-Ortiz, P. F. Leadlay and J. N. Andexer, *ChemBioChem*, 2011, **12**, 1011–1014.

165 E. Mahenthiralingam, L. Song, A. Sass, J. White, C. Wilmot, A. Marchbank, O. Boaisha, J. Paine, D. Knight and G. L. Challis, *Chem. Biol.*, 2011, **18**, 665–677.

166 H. Irschik, R. Jansen, K. Gerth, G. Hofle, H. Reichenbach, G. Gesellschaft and B. Forschung, *J. Antibiot.*, 1987, **40**, 7–13.

167 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840–1849.

168 A. Weir, PhD Thesis, University of Bristol, 2021.

169 C. J. Arthur, A. E. Szafranska, J. Long, J. Mills, R. J. Cox, S. C. Findlow, T. J. Simpson, M. P. Crump and J. Crosby, *Chem. Biol.*, 2006, **13**, 587–596.

170 M. Jenner, J. P. Afonso, H. R. Bailey, S. Frank, A. Kampa, J. Piel and N. J. Oldham, *Angew. Chem-Ger. Edit.*, 2015, **127**, 1837–1841.

171 L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chem. Int. Edit.*, 2011, **50**, 2795–2798.

172 G. L. Tang, Y. Q. Cheng and B. Shen, *Chem. Biol.*, 2004, **11**, 33–45.

173 M. Nardini and B. W. Dijkstra, *Curr. Opin. Struc. Biol.*, 1999, **9**, 732–737.

174 C. Li, M. Hassler and T. D. H. Bugg, *ChemBioChem*, 2008, **9**, 71–76.

175 J. Hedlund, H. Jörnvall and B. Persson, *BCM Bioinformat.*, 2010, **11**, 534–550.

176 A. S. Worthington, G. H. Hur, J. L. Meier, Q. Cheng, B. S. Moore and M. D. Burkart, *ChemBioChem*, 2008, **9**, 2096–2103.

177 J. C. Milligan, D. J. Lee, D. R. Jackson, A. J. Schaub, J. Beld, J. F. Barajas, J. J. Hale, R. Luo, M. D. Burkart and S. C. Tsai, *Nat. Chem. Biol.*, 2019, **15**, 669–671.

178 J. Connolly, PhD Thesis, University of Birmingham, 2017.

179 C. J. Arthur, C. Williams, K. Pottage, E. Płoskoń, S. C. Findlow, S. G. Burston, T. J. Simpson, M. P. Crump and J. Crosby, *ACS Chem. Biol.*, 2009, **4**, 625–636.

7 Supplementary Information

7.1 Protein sequences

Gene_Plasmid	**Sequence**
mAcpD_ pET151_D	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTLNHQVMDQVFDQVEHQ IAQVLGAKGGPLVAVEIDSRFSDLGLSSLDLATLISNLEAVYGTDPFADAV AITSIVTVADLARAYAQQGVPGPSPDPLDAQLRDLRQL
mAcpD_T_ pET151_D	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMLNHQVMDQVFDQVEH QIAQVLGAKGGPLVAVEIDSRFSDLGLSSLDLATLISNLEAVYGTDPFADA VAITSIVTVADLCRAYAQQGVPGP
mAcpA_pOPINF	MAHHHHHHSSGLEVLFQGPMNPERRNMYMEEIYTFVVSTLASSCKVQP GDIEPTTNLFADLGIDSVDFLDAVFCIEKHYDIRIPVGQWMSAVNEGNAAM TDYFVMEHFVAQIAARAAASA
mAcp3a_ pET151_D	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMAPLAAKAAPVVPVADD ECAQFLRQSLAAMLYCEPGQIRDGSRFLELGLDSVIAAQWIREINKHYQL KIPADGIYTYPVFKAFTQWVGTQLQ
MupD_pOPINF	MAHHHHHHSSGLEVLFQGPMRRQVVVTGGGRGIGASISTALVQAGFVVL IVGHRGISAGVALCEELDARYGAGSARWVGCDLGDALAVEGLCRSLEAE DIYGLVHCAGVSQDSLAAVADLDTARHMMQVNFWAFVQLYQALVRPMS MRRAGRILCIGSVAADFGMKGNALYASSKAALRGFVRSSMTEVASRAITI NCIEPGFIETALLDRYAEQRSAIDRRIPAGRYGSPEEVAGLAAFLFSEAGS YINGQHLTVDGGLSRSMAG
MmpF_pET28a	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFVSQALESHLG DLAVVGLACRFPGAPDQRVFWENLRQGRESIVDLDDQQLSAAGVTREQ WSEPGYIRRSPVLEGIEWFDARFFGVSAREAILLDPQHRLLLEVGYEALE DAGHVKAAQAGVTGVYACMGGITTSYLHQFGDRLDPLHHETASLVHQG NDKDFLATRLSFKLNLTGPSMTVQTA**C**SSSLVALHLACNALRLGEVDTAL VGAAAIRIPHHTGYPLGQSPLLSRDGRCCPFSSDASGTLFGSGVASVIVR RHADAIRDGDHVYALIKSSHVNNDGAMKIGYTATSVPGQAKAMVRAITLA RANARQISYVECHGTATSLGDPLEIKALEKAFRLDTQDHGFCAIGSVKSNI GHLEQAAGIASLIKVALMLKHDTLVPSLNFTTPNPRIDFEHSPFRVSQDTR VWSQALEQPADTARLAAINCLGIGGTNVFSVLQSVAPALVRGTDAAVDVP ICLSAATREQLRQYLLRFAHFVRDSGPIDRLALAHTINISRSAHRERFAGVL KAGIDVDRFFEEAAHSVLDAPATFTPRLVYYLCEQPQHLDDASRQAWLA APRYQRMADDYAECLGRLSSTPQTVEAAARCAELAFEVALYQQLVRWG MVFEAVIDRGYGHWINRCLQLGCGASSALDASPEWLSAFVQEGQGAAQ LASPGWPDSAQVAVLVVSRQAWAVTVPSSGRCVHLAEEMLTFADVERF ACDVVQAGISFDWLDYYRLVKAQKLSLPTYPFARQRYWPVED

MmpF_C183A_ pET28a	MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFVSQALESHLG DLAVVGLACRFPGAPDQRVFWENLRQGRESIVDLDDQQLSAAGVTREQ WSEPGYIRRSPVLEGIEWFDARFFGVSAREAILLDPQHRLLLEVGYEALE DAGHVKAAQAGVTGVYACMGGITTSYLHQFGDRLDPLHHETASLVHQG NDKDFLATRLSFKLNLTGPSMTVQTA**A**SSSLVALHLACNALRLGEVDTAL VGAAAIRIPHHTGYPLGQSPLLSRDGRCCPFSSDASGTLFGSGVASVIVR RHADAIRDGDHVYALIKSSHVNNDGAMKIGYTATSVPGQAKAMVRAITLA RANARQISYVECHGTATSLGDPLEIKALEKAFRLDTQDHGFCAIGSVKSNI GHLEQAAGIASLIKVALMLKHDTLVPSLNFTTPNPRIDFEHSPFRVSQDTR VWSQALEQPADTARLAAINCLGIGGTNVFSVLQSVAPALVRGTDAAVDVP ICLSAATREQLRQYLLRFAHFVRDSGPIDRLALAHTINISRSAHRERFAGVL KAGIDVDRFFEEAAHSVLDAPATFTPRLVYYLCEQPQHLDDASRQAWLA APRYQRMADDYAECLGRLSSTPQTVEAAARCAELAFEVALYQQLVRWG MVFEAVIDRGYGHWINRCLQLGCGASSALDASPEWLSAFVQEGQGAAQ LASPGWPDSAQVAVLVVSRQAWAVTVPSSGRCVHLAEEMLTFADVERF ACDVVQAGISFDWLDYYRLVKAQKLSLPTYPFARQRYWPVED
MupD_pOPINF	MAHHHHHHSSGLEVLFQGPMRRQVVVTGGGRGIGASISTALVQAGFVVL IVGHRGISAGVALCEELDARYGAGSARWVGCDLGDALAVEGLCRSLEAE DIYGLVHCAGVSQDSLAAVADLDTARHMMQVNFWAFVQLYQALVRPMS MRRAGRILCIGSVAADFGMKGNALYASSKAALRGFVRSSMTEVASRAITI NCIEPGFIETALLDRYAEQRSAIDRRIPAGRYGSPEEVAGLAAFLFSEAGS YINGQHLTVDGGLSRSMAG
mAcpB_pOPINF	AHHHHHHSSGLEVLFQGPMEINVAQQTVSESIRELKLPGWETASIETETH LQGELGIDSLHKLILLTRLQERSNFQFAQLNNEAYKFDTVGDLVNLLVAHG
MupE_pOPINF	MAHHHHHHSSGLEVLFQGPMRALQLVDKQRLALRELVLPARPAPGHAR VRMAFMTLNHLDLFGYRGMAFARRTLPVTVGAEGAGFVQDVGPGSDPE LIGQAMAIYPSVFCGCCPACLAGRENLCADSPGVMGFHCDGVAAEYVDV PLRMLVPIPAGVPMDMAACAPVTFATVHHMLIDNARLMSGETLLVHAGG SGIGSIAIRLARHLGVTVFATVGRADKVERARALGAEYVIDYREERFLRVIR EQTARRGVDVVFEHVGADTWEQSLLSLARGGRLVICGSTTGVEAPTNLM HLFNQQIHIHGSFGAPLSGLRDALALIQQGVLPVIDCVMPVAGFESGLQRL RERDVFGKIVLQF
MmpB_KS_ pET151_D	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTDIAIIGLAGRYPQAENIEE LWENLKLGRDCITTVPSQRWDHDAIYDPSKGVSGKTYSKWGGFLRGVD EFDPRFFNISPREAEIMDPQERLFLQCAYHVLEDAGYTRQSLSAKGRVGV YVGVQYTEYTAFSTAQTLVAALPASIANRVSFFCDFRGPSLTLDSM**C**SSS LTTIHLACQSLRSGESEYAIAGGVNVSIHPNKYLLHAQGRFASSVGCCKTF GQGGDGYVPAEGVGAVLLKPLPQAIADGDRIHAVIKGSSINHGGRATGFT VPKSSAQAVAVRSALHQAGLASSDLSYIEAHGTGTALGDLIEIEGLRTVFE ADGFEPQTCAIGSIKSNIGHCESAAGIAGLTKVLLQLKHRQLVPSLHSTQL

	NRNIDFAGSPFHVQQTLEPWLPKGAADAVQPRRAGISSFGAGGSNAHLV VEEYSLQAQAHTHVPTLPSDVLIVLSARTHERLQAVAALLLGRLEQTPLAC DLAGLHDLAYTLQVGREALDARAAFTAQSVQVLKERLVALADGAQHPDV LIGQALKPVRLRAGETAVPVQSPAMDDNQAVMRHWVSGGQVEWAQLY QGFQPARISLPLYPFVRERAWVP
MmpB_DH_pET151_D	MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTPGDGPHPLLSRQITAGD SQQFEVQFCGDEFFLDHHRVNGRKVLPGVVYLEMALAALSRVSGQADL ALNSILWVRPFECFNDPVTLTLAVEQARKGSGWEFEFFAWLTREPGGEP VRQVYCRGQASVPAEPAPLAHLDVNALAALAQGAASTTEFYRHFAAVGI EYGPGHQGLHALRGGAEQVLGHVRLPDFLKPGDEFRLHPCLLDSALQAA MGFTTGSESSQGPQVPFSLEQLTLGRSRRTPAWVRLTAHTSGGGRGSF AFDVEVFDAQGEPCVHLHGLRSRVMARAPAPP
MmpD_DH_pOPINF	MAHHHHHHSSGLEVLFQGPRLQRHDPSIGDHVIADQHRVAGAVTLAVAA RAWGALPAGGWCASQVAWLQALDVPLQGALLRVSLRPASAGATDFEVR LEGLVHPLCSGTLAPGQTDEVRRDVHALRRAHPDAWLTREVLYEQLRNV GMAYGSTYRVLQDSHLQGDQSLLALVVVPAGAADDLPFSPAAVDAAFQA AVLLAIRQSGEQARLGFSVGQLQLSAPLPERYWVHVSAATTGAQQCRFN IEWIAEDGRVCATAKDFVLKAAPVQQEQ

Table S 1: The sequences of proteins expressed during this project.

7.2 PCR Primers

Primer	Sequence
MmpF_FOR_pET28a	TGGGTCGCGGATCCGAATTCGTGAGCCAGGCCCTAGAG
MmpF_REV_pET28a	CAAGCTTGTCGACGGAGCTCTCAGTCTTCAACTGGCCAAT
MupD_FOR_pOPINF	AAGTTCTGTTTCAGGGCCCGATGCGTAGGCAGGTAGTCGT
MupD_REV_pOPINF	ATGGTCTAGAAAGCTTTACTAGCCGGCCATGCTG
MupE_FOR_pOPINF	AAGTTCTGTTTCAGGGCCCGATGCGGGCGCTTCAA
MupE_REV_pOPINF	ATGGTCTAGAAAGCTTTATTAGAATTGCAATACGATCCTGCC
mAcpB_FOR_pOPINF	AAGTTCTGTTTCAGGGCCCGATGGAAATCAACGTGGCG
mAcpB_REV_pOPINF	ATGGTCTAGAAAGCTTTACTAGCCATGCGCTACCAG
MmpF_C183A_FOR	GTGCAGACCGCGGCGTCAAGTTCGCTG
MmpF_C183A_REV	CAGCGAACTTGACGCCGCGGTCTGCAC
MmpF_C183A_FOR2	ACCGTGCAGACCGCGGCGTCAAGTTCGCTG
MmpF_C183A_REV2	CAGCGAACTTGACGCCGCGGTCTGCACGGT
MmpD_DH1_For_pOPINF	AAGTTCTGTTTCAGGGCCCGCGCCTGCAACGGCATGAC
MmpD_DH1_Rev_pOPINF	ATGGTCTAGAAAGCTTTATTATTGCTCTTGTTGGACTGGGGC

Table S 2: PCR Primers designed and used during this project (NOTE: 5' to 3')

7.3 mAcpD-MmpF HSQC Peak Weakening

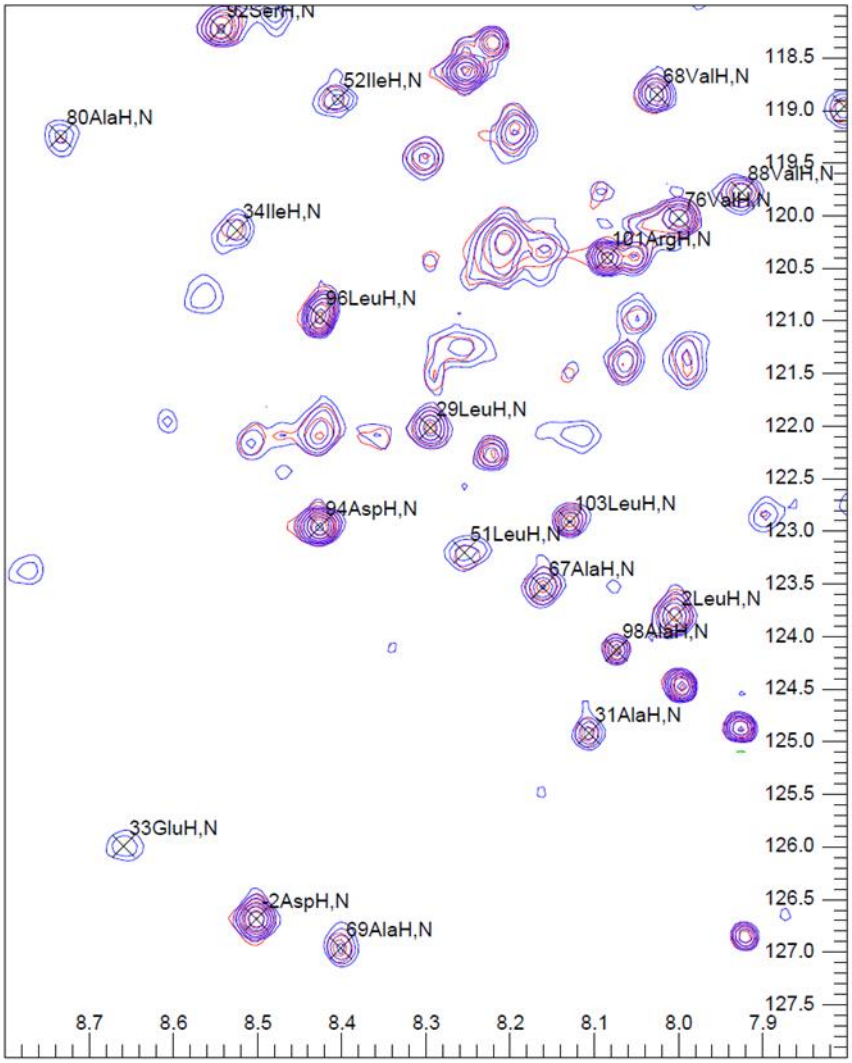


Figure S 1: The peak broadening observed in the mAcpD:MmpF, 1:4 HSQC NMR titration. Shown in blue are the peaks from ^{15}N-apo-mAcpD control. In red are the shifts after addition of MmpF. Many interacting peaks appear weaker. Conversely, residues from the C-terminal extension of mAcpD are not weakened implying no interaction.

7.4 Analytical SEC Calibration Curves

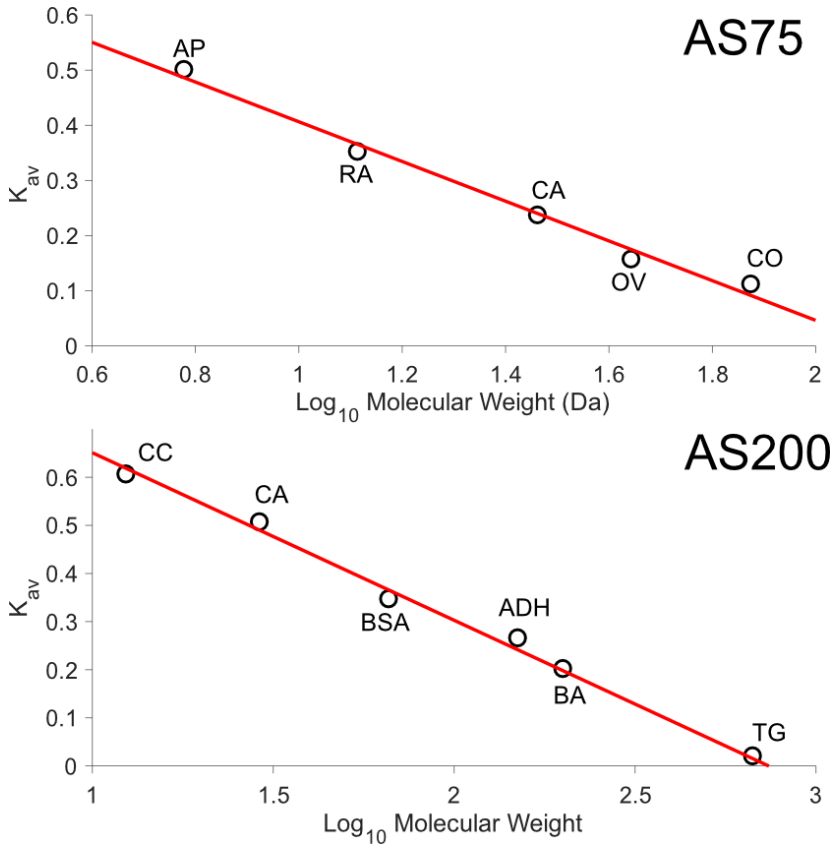


Figure S 2: Calibration curves for analytical SEC columns. Calibrants are as follows; AP = Aprotinin, RA = Ribonuclease A, CA = Carbonic Anhydrase, OV = Ovalbumin, CO = Conalbumin, CC = Cytochrome C, BSA = Bovine Serum Albumin, BA = β-amylase, TG=Thyroglobulin.